

ORSAY

N° D'ORDRE : 2770

UNIVERSITÉ DE PARIS SUD
U.F.R. SCIENTIFIQUE D'ORSAY

THÈSE

présentée
en vue d'obtenir le grade de

DOCTEUR EN SCIENCES
DE L'UNIVERSITÉ PARIS XI

SPÉCIALITÉ : Mathématiques

par

Nicolas BOUSQUET

Analyse bayésienne de la durée de vie de composants industriels

Soutenue publiquement le 19 décembre 2006 devant le jury composé de :

M.	Gilles CELEUX	INRIA	Directeur
M.	Pascal MASSART	Université Paris XI	Président
M.	Christian P. ROBERT	Université Dauphine	Rapporteur
M.	Nozer D. SINGPURWALLA	Université G. Washington	Rapporteur
M.	Jean-Michel MARIN	INRIA	Examineur
M.	André LANNOY	IMdR	Examineur
M.	François BILLY	EDF R&D	Invité

“Un expert est une personne qui a fait toutes les erreurs qui peuvent être faites dans un domaine très restreint.”

Niels Bohr

“Si le monde explose, la dernière voix audible sera celle d’un expert disant que la chose est impossible.”

Peter Ustinov, *L’amour des quatre colonels*

REMERCIEMENTS

Mes tout premiers et mes plus chers remerciements vont bien évidemment à Gilles Celeux qui, depuis le DEA, m'a donné l'élan et la motivation au jour le jour et m'a permis d'enrichir mes connaissances comme mes analyses avec son humour féroce, son café sans sucre, quelques réparties cinglantes et son grand stylo rouge. De nombreux mots ne suffiraient pas à dire le respect et la gratitude que j'éprouve pour lui, mais je me tairai car il m'accuse toujours d'en écrire trop. Tant pis.

Merci à Christian P. Robert et many thanks to Nozer D. Singpurwalla, dont les articles et ouvrages respectifs m'ont été considérablement utiles pendant ce travail de recherche, de par leur clarté et leur portée scientifique; eux-mêmes n'ont jamais hésité à répondre à mes questions ou m'aiguiller vers d'autres sources. C'est un honneur pour moi qu'ils aient accepté de rapporter cette thèse.

Au sein d'EDF, je voudrais adresser tous mes remerciements à ceux qui m'ont fait confiance dès le premier jour et ont survécu aux avalanches de rapports : François Billy, Emmanuel Remy et Florent Josse. Merci également à André Lannoy pour ses encouragements, ses conseils et corrections, ainsi que pour son érudition de l'expertise industrielle, érudition qu'il n'a pas hésité à faire partager. La chance de pouvoir travailler en collaboration avec des industriels confrontés à des problèmes pratiques importants et à l'écoute du travail de recherche, est une opportunité que tous vous m'avez permis de saisir.

Je ne peux pas certainement pas oublier le rôle prépondérant de Jean-Michel Marin dans le suivi et la correction de ce travail. Ses idées à la minute et ses combats incessants contre mes approximations en font l'ange gardien de ma thèse! Merci également à lui pour ses présentations sans respiration et sa rigueur qui m'ont beaucoup appris. Par ailleurs, je voudrais remercier Henri Bertholon, qui porte une lourde responsabilité dans le déclenchement de cette thèse. Lui le premier, il a su trouver les encouragements nécessaires pour descendre dans l'arène. J'ai énormément apprécié pouvoir travailler avec lui.

Il me faut aussi exprimer toute ma reconnaissance à l'équipe de Probabilités et Statistiques d'Orsay qui m'a accueilli, au sortir de mon école d'ingénieur grenobloise. En particulier, mes remerciements vont à Pascal Massart, Estelle Kuhn et Gilles Stoltz pour leurs mots rassurants à mon arrivée. Merci en outre à Pascal d'avoir bien voulu présider mon jury de thèse. Sans compter tous les thésards de l'équipe SELECT, des Stats et Probas, et des mathématiques appliquées ou non, qui m'ont intégré avec la gentillesse qui les caractérise. Ils se reconnaîtront (en particulier ceux du bureau 16, passés et présents). Un grand merci à notre formidable secrétaire, Valérie Blandin-Lavigne, sans qui ce document ne serait qu'un infâme ramassis de notes enchevêtrées et illisibles, et à Patrick Jakubowicz et Yves Misiti qui ont enduré avec courage mes démêlés informatiques. Le tableau ne serait pas complet sans un petit mot pour les trois profs qui, par leur sens de la pédagogie et de l'exemple qui tue, m'ont peu à peu aiguillé vers les rails des statistiques appliquées : je remercie Olivier François, Jean-Louis Soler et Olivier Gaudoin.

Un merci tout spécial à Guillemette, Ismaël B., Stefano, Laurent Z. et Bertrand, pour tant de choses qui j'espère vont se perpétuer! Et je n'oublie pas les autres, Mathieu, Romain, Aurélien, Sophie, Laurent T. et son fanzine préféré, Sylvain et l'autre Nico du bureau, pour les franches rigolades et les autres bons moments passés ensemble. Sans oublier Robin Hobb, George R.R. Martin, le cerisier du Japon, François G., son thermomètre et sa Corse, Ferréol R., Mathieu D. et le blog du Deg, Pierre & Steph et le Parc des Princes. Bonne chance à Manel pour cette même route!

Enfin, mes pensées vont à mes amis proches et ma famille, qui m'ont stimulé, soutenu et permis de prendre le recul nécessaire au déroulement d'un travail de si longue haleine ... et à Marie, qui donc vécu toute la thèse à mes côtés, impasses comme avancées, avec courage, humour et amour. Merci infiniment à toi.

Sommaire

1	Enjeux industriels et modélisation statistique	21
1.1	Introduction	21
1.2	Particularités et difficultés du contexte industriel	22
1.3	Description et interprétation des modèles statistiques	23
1.3.1	Modèle exponentiel	23
1.3.2	Modèle de Weibull	23
1.3.3	Modèle à risques concurrents	24
1.4	Formalisation statistique du cadre d'étude	25
1.4.1	Variable de durée de vie	25
1.4.2	Phénomène de censure (données incomplètes)	25
1.4.3	Structure à données manquantes	26
2	Limites des méthodes d'inférence classiques pour l'application industrielle	27
2.1	Introduction	28
2.2	Descriptif des principales méthodes d'estimation	29
2.2.1	Méthodes empiristes	29
2.2.2	Maximisation de la vraisemblance	31
2.3	Intervalles de confiance et incertitude d'un estimateur	35
2.3.1	Intervalles de confiance théoriques	36
2.3.2	Procédures de bootstrap pour l'estimation d'une incertitude	38
2.4	Indicateurs décisionnels de l'analyste industriel	44
2.5	Applications numériques	46
2.5.1	Présentation des données et estimations des paramètres	46
2.5.2	Application au modèle exponentiel	47
2.5.3	Application au modèle de Weibull	50
2.5.4	Évaluation des méthodes par facteur d'erreur	52
2.6	Conclusion	53
2.7	Annexes	54
3	Une modélisation de l'expertise industrielle en durabilité	57
3.1	Introduction	58
3.2	Modélisation non informative	61
3.2.1	Motivations	61

3.2.2	Application aux modèles de durée de vie	62
3.3	Une modélisation informative des paramètres de Weibull	65
3.3.1	Motivations	65
3.3.2	La modélisation	66
3.3.3	Traduction de l'expertise quantitative	73
3.3.4	Calibration de l'expertise quantitative	76
3.3.5	Recalibration finale de l'expertise	81
3.3.6	Caractéristiques effectives de la modélisation	85
3.4	Analyse <i>a posteriori</i>	86
3.4.1	Consistance et convergence	87
3.4.2	Gain d'information	89
3.4.3	Étude de sensibilité	91
3.5	Indicateurs de taille des données observées	93
3.5.1	Taille effective des données	93
3.5.2	Un indicateur approximatif	94
3.6	Annexes	97
3.6.1	Preuves et justifications	97
3.6.2	Cas d'une expertise paramétrique	104
4	Measuring agreement between prior and data	107
4.1	Motivation	108
4.2	Previous works	109
4.2.1	An intuitive tool	110
4.2.2	A marginal viewpoint	110
4.3	The DAC criterion	111
4.3.1	The formal definition	111
4.3.2	Main features	113
4.3.3	Difficulties of computation	113
4.4	Examples of ideal cases	114
4.4.1	Discrete models	114
4.4.2	A continuous bounded model	116
4.5	An intrinsic adaptation of DAC	117
4.5.1	Introducing the intrinsic heuristic	117
4.5.2	Adapting DAC	120
4.5.3	A real example with the Weibull distribution	124
4.6	Prior calibration with DAC	126
4.6.1	Application to the exponential distribution	126
4.6.2	Application to the Weibull distribution	127
4.7	A possible improvement	129
4.8	Conclusion	131
4.9	Appendix	131
4.9.1	Propositions and proofs	131
4.9.2	Kullback-Leibler divergences	135
4.9.3	Posterior priors	136

5	Méthodes de calcul bayésien	139
5.1	Introduction	140
5.2	Algorithmes d'échantillonnage	142
5.2.1	Méthodes MCMC	142
5.2.2	Méthodes d'échantillonnage préférentiel	145
5.3	L'algorithme BRM : une étude critique	147
5.3.1	Définition de la méthode	147
5.3.2	Difficultés et précautions	148
5.3.3	Une densité instrumentale de référence	149
5.3.4	Comparaisons	149
5.4	L'algorithme Population Monte-Carlo	152
5.4.1	Définition	152
5.4.2	Application formelle au modèle $\mathcal{B}(\eta_0, \eta_1, \beta)$	157
5.4.3	Applications numériques	159
5.5	Conclusions	164
6	Conclusions et perspectives	165
6.1	Principaux résultats, difficultés et motivations	165
6.2	Un modèle de défaillance à la sollicitation	166
6.2.1	Définition formelle	167
6.2.2	Nature et vraisemblance des données	167
6.2.3	Objectifs de recherche	168
6.3	Une méthodologie de calibration	169
6.3.1	Une modélisation d'un <i>a priori</i> subjectif	169
6.3.2	Objectifs de recherche	170
6.4	Une proposition d'échantillonnage préférentiel non adaptatif	171
6.5	Annexes	174
6.5.1	Minimisation hiérarchique	174
6.5.2	Simulation d'importance pour la loi de Weibull	175
	ANNEXES GÉNÉRALES	176
A	The competing risk model $\mathcal{B}(\eta_0, \eta_1, \beta)$	177
A.1	Introduction	178
A.2	Characteristics of the \mathcal{B} distribution	179
A.3	Estimating the parameter of the \mathcal{B} distribution	181
A.3.1	Maximum Likelihood Estimation	181
A.3.2	The EM algorithm	182
A.3.3	Bayesian inference through importance sampling	184
A.4	Assessing the failure distribution	188
A.5	Numerical experiments	189
A.5.1	Estimations and tests on simulated data	189
A.5.2	Estimation and tests on prospective data	190
A.5.3	Analysis of vehicles failures	192
A.6	Discussion	194
A.7	Appendix: existence of a consistent root of the likelihood equations	194

B	An industrial Bayesian analysis with Weibull distributions	197
B.1	Introduction	198
B.2	Data and parameters	199
B.2.1	The statistical context	199
B.2.2	Meaning of Weibull parameters	199
B.3	The prior modelling	200
B.3.1	The prior distribution of the shape parameter	200
B.3.2	The prior distribution of the scale parameter	201
B.4	Prior calibration	204
B.4.1	Default prior calibration	205
B.4.2	Recalibration and consensus	207
B.5	Posterior computation	210
B.6	A numerical example	211
B.7	Conclusion	214
.1	Appendix : proof of proposition 17	214
C	Tables de données (retour d'expérience)	215
C.1	Durées de vie de composants nucléaires (circuit d'eau secondaire)	215
D	Glossaire	217
D.1	Glossaire des termes statistiques	217
D.2	Glossaire des termes d'ingénierie	218
E	Bibliographie	221

Présentation générale

Au sein d'une entreprise industrielle comme EDF, la maîtrise des risques techniques qui peuvent survenir au cours du fonctionnement d'une installation est une préoccupation majeure. Des stratégies de maintenance sont mises en place pour permettre l'exploitation de matériels et préserver la sécurité des personnes, des biens et de l'environnement. Entretien, réparation, remplacement : ces opérations indispensables doivent, dans le respect de ces exigences de sécurité, être conduites sous une double contrainte de performance et de moindre coût. Nécessairement, une maintenance optimale est reliée à une prévision de la durée de vie des composants de ces systèmes, prévision qui se fonde en premier lieu sur une analyse fine des données de retour d'expérience (ou REX). Plus précisément, l'étude de la durée de vie de tout système industriel se scinde en deux alternatives distinctes, qu'il est important de ne pas confondre et qui nécessite un travail important de validation de ce REX.

- La *fiabilité* est l'aptitude de ce système à accomplir la fonction à laquelle il est dévolu, dans des conditions d'utilisation données et pendant un intervalle de temps donné (normes NF X 60-500 et 50-120). L'ingénieur fiabiliste cherche à déduire de l'évolution temporelle de cette fiabilité une future stratégie de maintenance. Le type de REX utilisé est *fonctionnel*.
- La *durabilité* est l'aptitude de ce système à accomplir cette fonction, dans des conditions d'utilisation et de maintenance données, jusqu'à ce qu'un état limite soit atteint (norme EN 13-306, 2001). On peut grossièrement résumer la démarche de l'ingénieur durabiliste à l'estimation de la durée de vie *restante*, lorsque le système est en cours de fonctionnement. Une telle étude cherche à prendre en compte la modification des conditions d'exploitations, le renouvellement des matériaux, etc. pour en déduire le temps de bon fonctionnement du système. La durabilité est donc focalisée sur le patrimoine et l'investissement de l'industriel, et est étudiée avec un REX de *remplacement*.

Les ingénieurs utilisent alors, de façon croissante, des méthodes statistiques pour prévoir le comportement des composants industriels, que ce soit sous l'angle fiabiliste ou durabiliste. La difficulté ou l'impossibilité d'obtenir ce type de résultat par des résolutions d'équations de la physique en est une raison importante. Une solution consiste à modéliser la distribution des temps de défaillance de ces composants par une loi statistique. En définissant de façon formelle des grandeurs compréhensibles par les ingénieurs, telles que la probabilité de défaillance ou la durée de survie après un temps fixé (par exemple un temps de maintenance), puis en les estimant au moyen du REX, les statisticiens fournissent des outils de prise de décision pour la maintenance des composants. En vertu de normes

de sécurité et de calculs technico-économiques, une stratégie de rénovation ou de remplacement peut être établie.

Quelles sont alors les principales causes de défaillance d'un composant ? Nous considérerons que le mécanisme de dégradation¹ dû au vieillissement représente l'origine la plus probable d'une panne. Peuvent s'y ajouter diverses causes accidentelles, telles qu'une erreur humaine ou une fabrication défectueuse². De nombreux autres contextes industriels se révèlent adaptés à ce cadre d'étude ; ainsi en est-il, par exemple, de l'industrie automobile. La plus célèbre des modélisations statistiques du vieillissement est la loi de Weibull, celle de la défaillance par accident étant la loi exponentielle. Un modèle mettant en compétition ces deux risques de défaillances a été proposé par Bertholon (2001).

Ce souci de modélisation et de prévision prend un relief particulier lorsque les composants étudiés appartiennent à des systèmes complexes, très coûteux à tester et potentiellement dangereux. Les politiques de maintenance passées ont conduit à éviter les défaillances, et les REX disponibles sont alors de taille faible et contiennent des données parfois hétérogènes et souvent censurées, qui correspondent plus à des durées de fonctionnement avant arrêt définitif qu'à des durées de vie réelles³. Les composants nucléaires appartiennent clairement à cette catégorie.

Très rapidement, l'appel aux seules données de REX ou d'essais ne suffit pas pour prévoir précisément le comportement des composants, et il est nécessaire de faire appel à une connaissance annexe de la durée de vie du matériel. Cette connaissance est en général apportée par des experts industriels, qui peuvent se subdiviser en plusieurs catégories : les exploitants (qui connaissent le fonctionnement global d'une installation particulière), les fournisseurs (qui cherchent à respecter un cahier des charges sur des projets précis), les ingénieurs matériaux et les physiciens (qui connaissent les propriétés des matériaux et de leurs interactions), les ingénieurs systèmes (qui tiennent compte des conditions d'exploitation), etc. Difficilement quantifiable, cette connaissance est donc essentiellement subjective et doit s'additionner avec précaution à la connaissance objective apportée par les données de REX. Statistiquement parlant, cette possibilité est offerte par les techniques d'inférence bayésienne.

Ce travail de thèse est essentiellement méthodologique. En s'inscrivant à la suite de nombreuses études de modélisation et d'analyse de modèles de durée de vie, il a pour objectif de fournir un ensemble d'outils bayésiens permettant à l'ingénieur d'estimer de façon satisfaisante les grandeurs typiques de la fiabilité et de la durabilité industrielles. Par "satisfaisante", nous entendons avec une précision suffisante au regard des normes de sécurité. Nous donnons également à ce terme le sens de "compréhensible". Notre démarche a un but didactique, celui de sensibiliser l'ingénieur aux difficultés de l'approche bayésienne en lui proposant une gamme d'indicateurs mesurant les quantités d'informations objective et subjective utilisées. Nous tentons donc, dans ce travail, d'éclaircir au maximum les aspects techniques de l'analyse bayésienne (un glossaire expliquant quelques termes statistiques typiques mais peu explicites est proposé en ce sens, à la fin du document ; il rappelle également quelques définitions normatives du vocabulaire de la fiabilité et de la durabilité). Les principales contributions

¹Voir Glossaire p. 219. Dans la suite du document, tout rappel au Glossaire sera signifié par *.

²Cependant, avant le traitement statistique, le REX est censé avoir été soumis à un traitement rigoureux, qui tend à éliminer les données de défaillance reflétant des causes autres que les phénomènes généraux que l'on cherche à modéliser

³Paradoxe d'une discipline comme la fiabilité qui, à l'instar de la médecine, n'a d'autre but que de se rendre inutile.

de ce travail sont les suivantes.

1. *Une analyse bayésienne des modèles utilisés dans le cadre d'expertise proposé par EDF* (chapitre 3 ; un article reprenant les principaux points de ce chapitre est fourni en Annexes B). Nous construisons une modélisation *a priori*, pour les paramètres des familles de lois de Weibull, qui tient compte des points suivants :

1. pour chaque paramètre, son sens *qualitatif* (lié au comportement du composant, par exemple son vieillissement) ou *quantitatif* (lié à des valeurs représentatives de la durée de vie) ;
2. la connaissance experte disponible, qualitative et quantitative ;
3. la corrélation entre paramètres.

En particulier, cette modélisation *a priori* est suffisamment souple pour être calibrée par un analyste bayésien de façon compréhensible. Celui-ci juge l'expert quantitatif comme un fournisseur de données "fictives", au travers d'un hyperparamètre a correspondant à la taille de ces données. Nous proposons alors une règle de calibration par défaut, en fixant a en fonction de la précision de l'information que l'expert indique. Un ensemble de questions ouvertes est proposé afin d'améliorer cette calibration. Enfin, *via a*, l'information de l'expertise peut être comparée à l'information apportée par les données réelles (le REX), au moyen d'indicateurs de taille. On peut donc repérer sinon tempérer les cas où, *a posteriori*, l'estimation reste majoritairement subjective.

2. *La définition et l'étude d'un critère d'éloignement, noté DAC, entre données de REX et information d'expertise* (chapitre 4). Le cadre plus général de cette contribution est celui d'une étude bayésienne concrète, où la connaissance *a priori* est majoritairement subjective. Sa modélisation, nécessairement arbitraire, peut se révéler très éloignée de l'information objective apportée par les données, et témoigner, par exemple, d'une modélisation *a priori* médiocre ou d'une hétérogénéité de l'information apportée par l'expert vis-à-vis de la nature des données (interrogé sur un composant \sum_1 , son avis peut plutôt porter sur un composant \sum_2 proche de \sum_1 , placé dans des conditions de fonctionnement différentes). Cet éloignement, que l'on perçoit parfois intuitivement, doit être détecté avant l'inférence et l'acceptation des résultats *a posteriori*.

La prévision de la durée de vie industrielle est particulièrement concernée par ce type de problème. Par exemple, les données de REX peuvent correspondre à la durée de vie de composants dont l'arrêt d'exploitation n'est pas vraiment imputable au vieillissement mais plutôt à des opportunités de remplacement, des difficultés d'exploitation, des problèmes de composant ou des modifications de conception. Certaines d'entre elles peuvent se révéler peu représentatives de la durée de vie des composants actuellement utilisés. Pour des raisons d'évolution technique ou de différence de composant, un expert de maintenance peut fournir, par exemple, un avis à la fois très informatif et très optimiste vis-à-vis de données de temps de défaillance.

Dans les faits, donc, les connaissances *a priori* et objective relatives à une question de *fiabilité* doivent être cohérentes, sinon il subsiste un problème de validation de données ou d'expertise. Lorsqu'elles sont relatives à une question de *durabilité*, ce type de problème apparaît moins essentiel : il s'agit de prévoir une durée de vie restante, après une stratégie de maintenance qui tient compte de l'évolution technique. Vis-à-vis du REX disponible, l'expertise peut donc paraître optimiste sans

qu'un conflit effectivement détecté soit à prendre en compte par l'analyste bayésien. La figure 1 permet d'illustrer brièvement cette idée déjà exprimée que le contexte de l'étude modifie la pertinence des outils statistiques. Cependant, un critère mesurant la divergence entre ces connaissances reste important pour valider le travail d'investigation de l'expert.

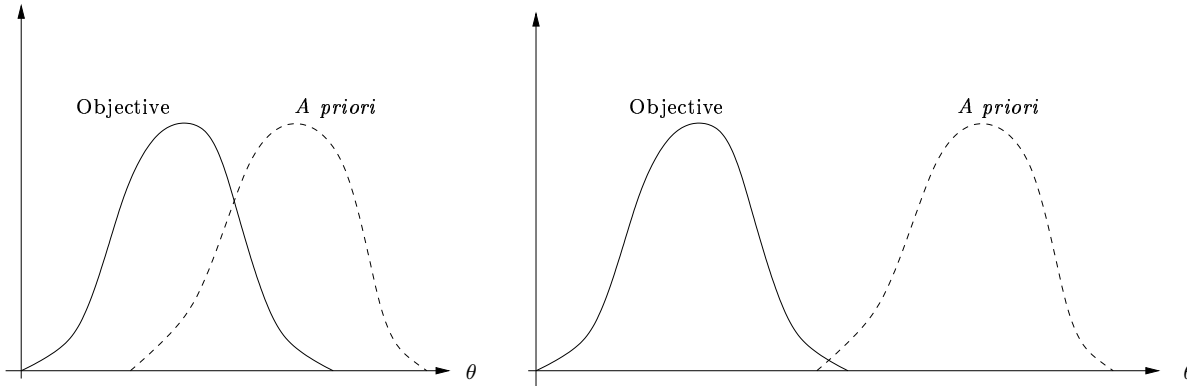


FIG. 1 – Représentations simplifiées de distributions objective (fournie par la connaissance des données de REX) et *a priori* sur un paramètre d'intérêt θ . À gauche, les deux connaissances restent cohérentes que le point de vue soit fiabiliste ou durabiliste. À droite, elles sont incohérentes d'un point de vue fiabiliste. D'un point de vue durabiliste, un tel éloignement n'est pas rédhibitoire puisque la vision de la durée de vie est prévisionnelle.

Le critère DAC peut également être utilisé comme un outil de calibration par défaut, lorsque l'analyste bayésien, recueillant et modélisant l'avis de l'expert, sait ne pas disposer d'information *crédible* sur l'incertitude de cet expert. Il s'agit donc de proposer, typiquement, une valeur objective de la variance *a priori*, d'autant plus large que l'écart entre expertise et données est grand.

En introduction à ces deux contributions, le chapitre 1 offre une définition générale du cadre statistique et des modèles auxquels nous nous intéressons. Notons en Annexes A un article coécrit en cours de thèse sur le modèle à risques concurrents proposé par Bertholon (2001). Cet article sera fréquemment cité au cours de l'étude. Le chapitre 2 se veut fournir un panorama des méthodes *fréquentistes* usuelles, qui sont des méthodes d'estimation objectives des modèles statistiques (fondées uniquement sur l'information apportée par les données de REX). L'objectif est de mesurer la limite de ces méthodes en termes de précision d'estimation. Nous proposons par exemple l'emploi conjoint de plusieurs algorithmes de *bootstrap* pour s'assurer que l'appel aux techniques bayésiennes est nécessaire. Méthodologiquement parlant, cette vérification nous paraît utile, étant donné la difficulté que peut revêtir le travail préparatoire à l'inférence bayésienne qu'est la calibration *a priori*.

Le chapitre 5 est essentiellement un récapitulatif des principales techniques de calcul bayésien. Typiquement, on cherche à approximer une expression intégrale correspondant à une moyenne *a posteriori*. Nous nous focalisons en particulier sur des algorithmes d'*échantillonnage préférentiel*, préférés

aux méthodes MCMC traditionnelles. L'algorithme BRM (*Bayesian Restoration Maximization*, Bacha 1996), utilisé dans de récents logiciels de fiabilité, fait l'objet d'une étude critique particulière qui conduit à le rejeter pour des raisons théoriques et pratiques. Le très récent et puissant algorithme PMC (*Population Monte Carlo*, Cappé *et al.* 2004, Celeux *et al.* 2006) est par ailleurs détaillé dans un but didactique, et des exemples d'étude sont proposés, pour les schémas bayésiens émanant des chapitres précédents. Là encore, selon le but à atteindre, nous fournissons une conclusion méthodologique.

Le chapitre 6, consacré à quelques perspectives de recherche, clôt ce travail. Nous nous focalisons en particulier sur le choix des hyperparamètres d'une distribution *a priori* arbitraire, en conservant l'idée pratique que l'analyste bayésien ne peut fournir sur l'expert qu'un jugement limité, exprimable par une "taille" d'échantillon fictif (comme a au chapitre 3). Cette taille est donc l'unique paramètre modifiable par l'analyste en fonction de sa connaissance de l'expert (par exemple *via* son taux de réussite passé). Nous proposons donc quelques pistes en ce sens.

Au final, nous espérons avoir obtenu un certain nombre d'outils, d'indicateurs et de méthodes facilitant le travail de l'ingénieur fiabiliste ou durabiliste, lorsqu'il doit chercher à inférer sur les modèles classiques de durée de vie, avec pour sources d'information des données de REX en petit nombre et censurées ainsi qu'une connaissance experte parfois forte, mais dont il a du mal à mesurer l'incertitude. Bien que nous utilisions des exemples typiques des études d'EDF, le cadre de cette thèse reste donc général (voire, pour le chapitre 4, ouvert à tout problème de modélisation bayésienne subjective). À l'avenir, d'autres modèles de durée de vie sont susceptibles d'être étudiés de façon similaire, qui intègrent par exemple des processus de défaillances à la sollicitation.

Ajoutons, en conclusion de cette présentation générale, que ce travail de thèse est complété par des programmes informatiques qui devraient, dans un proche avenir, aboutir à un logiciel accueillant et maniable par un ingénieur, dont la connaissance statistique serait limitée, et qui n'effectuerait une analyse bayésienne qu'occasionnellement. Les indicateurs, critères, algorithmes et calculs inférentiels proposés dans notre travail y sont codés (y compris les procédures d'estimation fréquentistes), en langages C et R/Splus.

General presentation

The main issues raised by the estimation of lifetime statistical models used in industrial reliability are censoring and the sample size of *feedback experience data* (FED). Many studies have to deal with homogeneous, small-sized, censored failure times which have to be integrated into Bayesian procedures with informative prior distributions. This way of dealing with statistical inference has been especially followed by EDF (Électricité de France) for predicting failures on nuclear material.

This thesis work has been motivated by such difficulties encountered in industry, and has been for the most part produced in collaboration with EDF. It is essentially methodological and proposes some recipes and tools of statistical analysis for the industrial analyst, about the integration of subjective expert opinion in the Bayesian estimation of decision-making models.

Indeed, various experts are liable to give information about the lifetime of a system or a component, for instance component producers, operators, service engineers. They offer information that is difficult to quantify and should be added very carefully to the objective information provided by the FED.

More precisely, we are interested in modelling the lifetime of an industrial component \sum which is submitted to aging. The two best-known lifetime models are the exponential and the Weibull distributions. A competing risk model has been studied as an alternative to the Weibull distribution. Its definition and its frequentist and Bayesian estimations are the subject of an article by Bertholon, Bousquet & Celeux (2006), accepted by *Lifetime Data Analysis* and provided in Appendix A. On several examples, numerical tests show that this model improves the representation of the lifetime as compared with the Weibull distribution. Besides, it allows a suitable modelling of the reliability of a two-component series system.

Note that a constant worry of our work is to be the most didactic possible, in order to sensitize the engineers to the main procedures and the difficulties of the Bayesian subjective analysis. In addition to this article, the main contributions of our work are following.

1. *The elicitation of a prior class for the Weibull (and poly-Weibull) distributions* (Chapter 3 in French, Appendix B in English). This work extends the work of Berger & Sun (1993) and Bacha (1996) and proposes an alternative method to the Weibull prior elicitation of Singpurwalla & Song (1986). Our prior elicitation takes into account the available expert knowledge (in the case of EDF). This knowledge is essentially in terms of lifetime, and is not formulated directly on

the parameters. It refers to some characteristics of the prior marginal (predictive) distribution. We give a quantitative or qualitative sense to the parameters. Then we elicit a prior about the quantitative opinion, of which a hyperparameter takes the sense of a size of a *virtual sample* which reflects the expert uncertainty.

Thus we can easily modulate the prior by means of discussion between the Bayesian analyst and the expert, since the size is a simple and understandable indicator of information. Default choices are proposed in this way, as starting points for the discussion. Some indicators of the effective size of the observed censored data are given to locate this subjective information with respect to the objective data information.

From an industrial point of view, the characteristics of the complete prior are studied, showing its relative appropriateness. A main feature is that the global prior sets out a correlation between the parameters and simplifies the posterior computation. Finally we focused on the posterior distribution, studying the consistency and convergence properties; then, from a sensitivity analysis, we propose some default prior choices and advices to let the posterior distribution remain the most stable possible with respect to small prior changes. Finally, this prior is easy to use and understandable by an industrial analyst without advanced statistical knowledges; that was a main motivation of the chapter.

2. *The definition and the study of the DAC criterion* (Chapter 4). This chapter is in English. DAC is a criterion which defines a statistical agreement between a prior information and the data information. The setting of this work is more general than reliability and concerns any Bayesian setting where a subjective prior can be far from the data information. However, this is especially relevant in industrial Bayesian analysis. For instance, lifetime data can reflect the past and the expert opinions can reflect the future, for reasons of technical evolution or material divergence; there is a risk to obtain a time incoherency between the two sources of information. Other difficulties are listed through real and simulated examples.

Note however that the notion of reliability is different from *durability*, which is the study of the remaining (or residual) lifetime of any working system, in given conditions of use and maintenance. Reliability's aim is to represent the complete lifetime process to establish, in the future, some strategies of use and maintenance. Thus, from a durabilist viewpoint, an engineer can naturally take into account the technical evolution and a possible prior-data conflict is no more relevant. This can be intuitively illustrated by Figure 2. Then the context of the study must clearly be defined to know if DAC gives an useful information.

Besides, the DAC criterion can be used as a tool of default prior calibration, when no *credible* information about the real expert uncertainty is available to the Bayesian analyst. Typically, when a prior central opinion is given, DAC can be used to propose an objective value for the prior variance, which increases with the difference between the expert opinion and the location of the data.

Then DAC appears to be an alternative and more powerful approach to a procedure proposed by Evans & Moshonov (2005a) very recently, for detecting a conflict between prior and data.

Some other features make us consider this criterion as a practical method in the toolkit of the Bayesian analyst.

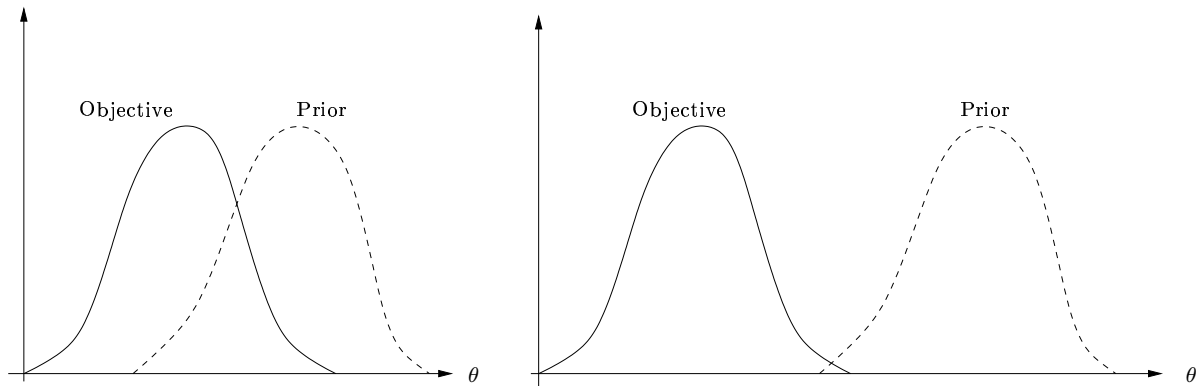


FIG. 2 – Ideal view of the probabilistic distributions on a parameter of interest θ yielded by the objective knowledge of data and the prior knowledge from an expert. Left, both knowledges remains coherent from reliability and durability viewpoints. Right, they are only incoherent from a reliability viewpoint. From a durability viewpoint, such a discrepancy is not unreasonable because the sight of the lifetime is forward-looking.

Chapters 1 and 2 (in French) of the thesis are introductory chapters, summarizing the main features of the industrial context, the statistical setting and the precision defects of the frequentist procedures. Chapter 5 (in French) is essentially a summary of the main Bayesian computational techniques that are necessary for the other chapters. Typically, we are looking to approximate an integral expression corresponding to a posterior mean. Especially, we focus our attention on *importance sampling* algorithms, which can be preferred to traditional MCMC methods in our context of highly-censored small samples. We make a critical analysis of the *Bayesian Restoration Maximization* (BRM) algorithm (Bacha 1996, Bacha & Celeux 1996), which is used in several recent reliability softwares. We reject it for theoretical and practical defects and consider the very recent and powerful *Population Monte Carlo* (PMC) algorithm (Cappé *et al.* 2004, Celeux *et al.* 2006). Its construction is detailed and some applications are done, using the prior schemes considered in previous chapters. According to the aim of the Bayesian analyst, a methodological conclusion is given.

Chapter 6 is dedicated to some future avenues of research and concludes this thesis. In particular, we propose to adapt our methodology to a shock model for assessing component aging reliability, which is of interest for numerous industrialists. This work should continue the preparatory work of Celeux & Rodionov (2002). A second avenue is to propose a methodology of subjective prior elicitation which generalizes the use of virtual data. For this we propose to consider priors that can be written under the form (or close to the form) of *expected pseudoposterior priors*.

The tools and methods presented in this thesis are completed by programs (in C/R/Splus) that should be integrated into a software package for the occasional use of an engineer with limited statistical knowledge.

Notations

La définition des notations suivantes sera rappelée à leur première occurrence dans le document, et elles seront réutilisées par la suite sans rappel obligatoire. D'une manière générale, les variables aléatoires (v.a.) seront notées en majuscules, les réalisations de ces variables en minuscules. Rappelons qu'un Glossaire est disponible à la fin de ce travail (p. 217), expliquant certains termes statistiques et d'ingénierie. Y réfèrent les termes suivis d'un * dans le texte.

VARIABLES

X	variable modélisant la durée de vie (en général) d'un composant industriel Σ
Y	variable observée résultant de la stratégie de censure de X
C	variable modélisant la censure de X
<i>i.i.d.</i>	indépendantes et identiquement distribuées (en parlant des réalisations d'une v.a.)
REX	échantillons d'observations de Y (retour d'expérience industriel)

MODÈLES PARAMÉTRIQUES ET DISTRIBUTIONS DE PROBABILITÉ

$\mathcal{M}(\theta)$	modèle statistique de comportement, de paramètre $\theta \in \Theta$
Θ	espace paramétrique
$\mathcal{E}(\eta)$	modèle exponentiel
$\mathcal{W}(\eta, \beta)$	modèle de Weibull
$\mathcal{B}(\eta_0, \eta_1, \beta)$	modèle de compétition Weibull contre exponentiel (Bertholon)
$\mathcal{N}(\mu, \sigma^2)$	modèle normal de moyenne μ et de variance σ^2
$\mathcal{U}[a, b]$	modèle uniforme sur $[a, b]$
P_θ	mesure de probabilité conditionnelle à θ
$h_{\mathcal{M}}(x \theta)$	taux de défaillance du modèle $\mathcal{M}(\theta)$
$f(x \theta)$	densité de X sachant le paramètre θ
$F(x \theta)$	fonction de répartition de X sachant le paramètre θ
$S(x \theta)$	fonction de survie (ou <i>fiabilité</i>) de X ($= 1 - F(x \theta)$) sachant le paramètre θ
$\pi(\theta)$	densité du paramètre θ s'il est supposé aléatoire

SYMBOLES MATHÉMATIQUES

$\xrightarrow{\mathcal{L}}$	convergence en loi
$\xrightarrow{p.s.}$	convergence presque sûre
$\text{Supp}(f)$	support de la fonction f
$\mathbb{1}_{\{A\}}$	indicatrice de l'évènement A
\log	fonction logarithme népérien (ln)

PARAMÈTRES

η	paramètre d'échelle (modèles $\mathcal{E}, \mathcal{W}, \mathcal{B}$)
$\lambda (= 1/\eta)$	paramètre d'échelle inverse
β	paramètre de forme (modèles \mathcal{W}, \mathcal{B})
$\mu (= \eta^{-\beta})$	paramètre de simplification (modèles \mathcal{W}, \mathcal{B})

SIGLES ET ABRÉVIATIONS

EMV/MLE	estimateur du maximum de vraisemblance (français/anglais)
DIP	test d'unimodalité d'Hartigan (chapitre 3), de statistique notée Dip
DAC	critère de proximité entre <i>a priori</i> et données (chapitre 4)
NEF	<i>Natural Exponential Family</i> , ou famille exponentielle naturelle (chapitre 4)
MTS	<i>Minimal Training Sample</i> , ou <i>échantillon minimal d'entraînement</i> (chapitres 4 et 6)

ALGORITHMES

NR	algorithme de descente de type Newton-Raphson (chapitre 2)
EM	algorithme <i>Expectation Maximization</i> (chapitre 2)
SEM	algorithme <i>Stochastic Expectation Maximization</i> (chapitre 2)
BRM	algorithme <i>Bayesian Restoration Maximization</i> (chapitre 5)
PMC	algorithme <i>Population Monte Carlo</i> (chapitre 5)

Enjeux industriels et modélisation statistique

1.1 Introduction

Nous cherchons à modéliser la durée de vie d'un composant industriel Σ . Pour ce faire, nous la supposons représentable par une variable aléatoire X suivant une loi statistique $\mathcal{M}(\theta)$ dont θ est le paramètre à estimer. Ce choix de modélisation paramétrique est intéressant de par la possibilité d'établir une stratégie de maintenance préventive* (par exemple, lorsque la probabilité de défaillance estimée est la plus forte) ou de prévoir des comportements futurs, telles que la survie ou la durée de vie que Σ peut encore espérer après un certain temps (par exemple un temps d'arrêt décidé auparavant). Chacun des aspects fiabilistes et durabilistes de la sûreté de fonctionnement* peut donc en être bénéficiaire.

L'enjeu d'une telle démarche est bien entendu une amélioration de la connaissance du comportement des matériels. À la clé, il s'agit de vérifier le bien-fondé des dates de maintenance ou de remplacement, voire de les rectifier. Par ailleurs, le respect de normes de sécurité est essentiel ; il est donc indispensable d'obtenir la meilleure précision possible sur ces estimations, afin de concilier fiabilité et modération de l'effort industriel d'entretien et de réparation.

Nous proposons d'abord une description succincte du contexte industriel dans lequel se place ce travail de thèse. Si ce dernier a été réalisé en collaboration avec EDF, le contexte de l'étude revêt un cadre beaucoup plus large, qui est celui des composants industriels soumis à du vieillissement et des défaillances par accident, et pour lesquels les coûts de remplacement, de maintenance et d'essai sont élevés. En second lieu, nous introduisons les modèles statistiques utilisés tout au long de ce travail, dont nous cherchons à estimer les paramètres. Vis-à-vis des caractéristiques des données EDF disponibles, des modèles trop complexes doivent être abandonnés. Enfin, nous concluons ce chapitre introductif par la formalisation mathématique du contexte de l'étude.

1.2 Particularités et difficultés du contexte industriel

Un grand nombre de données de défaillance de Σ est souvent requis pour obtenir des estimations fiables du paramètre θ , de quantités fonction de θ et des intervalles de confiance portant sur ces quantités. De tels nombres dépassent malheureusement le nombre de retours d'expérience (REX) disponibles, lorsque Σ appartient à un système industriel complexe pour lequel les coûts de relevé ou d'essai sont très lourds. Les composants de centrales nucléaires en sont un très bon exemple.

Par ailleurs, de par leur rareté, ces données hétérogènes sont parfois regroupées afin de constituer des échantillons de travail. Différences de matériaux, de conception, de type d'emplacement géographique, de conditions d'exploitation : ces données non identiquement distribuées rendent l'estimation plus douteuse, et il est parfois judicieux d'en ôter certaines de l'échantillon. Une liste célèbre, regroupant des âges d'apparition de fuites causées par la fatigue thermique, a par exemple été recensée par Atwood *et al.* (1999) et utilisée par Bertholon (2001) pour l'estimation d'un modèle de durée de vie à risques concurrents. Les données ont été relevées sur des centrales REP dans le monde entier, aux caractéristiques parfois différentes. Ajoutons à cela la possibilité que certaines durées de vie très courtes correspondent plutôt à des temps de rodage de Σ , ou *déverminage*, et ne sont pas représentatifs de la durée de vie de Σ .

Enfin, beaucoup de données de REX sont *censurées à droite*. Elles correspondent à des dates d'arrêt de Σ et non à des défaillances réelles. Dues à des politiques de maintenance ou de remplacement ou simplement à l'arrêt de la collecte des données, ces durées de vie indiquent uniquement que les temps de défaillance réels qu'elles masquent leur sont supérieurs. Dans le reste de documents, nous parlerons alors de *données incomplètes*.

Cependant, dans notre cadre d'étude, les données sont généralement soumises à un traitement particulier avant d'être mises à disposition de l'analyste statisticien. La justesse et la pertinence des données sont examinées avec soin, en particulier lorsqu'elles proviennent d'expériences coûteuses. Là encore, l'objectif de l'étude est pris en compte : fiabiliste ou durabiliste ? Certaines données anciennes peuvent être ôtées car peu représentatives du fonctionnement actuel (point de vue durabiliste). Des renseignements qualitatifs accompagnent le plus souvent les tableaux de REX ; dans le cas d'EDF, une date d'arrêt ou de défaillance peut potentiellement être fournie avec

- la mention de la durée de vie active en unités de fonctionnement ;
- le mode de défaillance* ou une explication de l'origine de l'arrêt ou de la défaillance ;
- les tâches de maintenance possiblement réalisées sur l'exemplaire du composant considéré.

Nous considérons donc dans ce travail que les données disponibles ont fait l'objet d'un contrôle qualitatif, et que les problèmes fondamentaux que pose leur traitement sont dus à la faible taille de l'échantillon et la proportion élevée de censures. Sur certains jeux de données reconstitués, cependant, nous indiquerons qualitativement quelles mesures pourraient être entreprises pour les "épurer". Dans la section suivante, nous introduisons les modèles à estimer. La dimension du vecteur des paramètres reste faible (de 1 à 3), permettant de ne pas accroître la difficulté de l'estimation. Par la suite, nous fournissons un cadre statistique pour entreprendre d'inférer sur ces paramètres dans ce contexte industriel.

1.3 Description et interprétation des modèles statistiques

Les modèles $\mathcal{M}(\theta)$ considérés sont définis par leur taux de défaillance* $h_{\mathcal{M}}(x|\theta)$. Ce taux de défaillance est une mesure de la probabilité instantanée d'une défaillance du système au temps x :

$$h_{\mathcal{M}}(x|\theta) = \frac{dP_{\theta}(x < X < x + dx)}{dx}$$

Il permet de définir la probabilité de défaillance avant x (la fonction de répartition F)

$$P_{\theta}(X \leq x) = F_{\mathcal{M}}(x|\theta) = 1 - \exp\left(-\int_0^x h_{\mathcal{M}}(u|\theta) du\right).$$

qui elle-même induit la représentation en densité de probabilité de la loi de comportement

$$f_{\mathcal{M}}(x|\theta) = \frac{dP_{\theta}(X < x)}{dx} = h_{\mathcal{M}}(x|\theta) \exp\left(-\int_0^x h_{\mathcal{M}}(u|\theta) du\right).$$

Nous allons en particulier nous intéresser à trois modèles emboîtés, dont les deux premiers sont certainement les plus utilisés en durabilité et fiabilité. Une référence majeure en la matière est le livre de Lawless (1982). Nous focalisons notre étude sur un composant \sum , soumis à une voire deux causes de défaillance, et non à un système multicomposants. Les taux de défaillance sont donc des fonctions simples de x , de type polynôme ou exponentiel.

1.3.1 Modèle exponentiel

Sachant $\lambda > 0$, le modèle exponentiel $\mathcal{E}(\lambda)$, de densité

$$f(x|\lambda) = \lambda \exp(-\lambda x) \mathbb{1}_{\{x \geq 0\}}$$

modélise la durée de vie d'un système \sum soumis à des défaillances *accidentelles* de par son taux de défaillance constant $h_{\mathcal{E}}(x) = \lambda$. La durée de vie moyenne est l'espérance $E[X] = \eta = 1/\lambda$.

1.3.2 Modèle de Weibull

Sachant $(\eta, \beta) > 0$, le modèle de Weibull $\mathcal{W}(\eta, \beta)$ a pour densité

$$f(x|\eta, \beta) = \frac{\beta}{\eta} \left(\frac{x}{\eta}\right)^{\beta-1} \exp\left\{-\left(\frac{x}{\eta}\right)^{\beta}\right\} \mathbb{1}_{\{x \geq 0\}}.$$

Son taux de défaillance $h_{\mathcal{W}}(x) = (\beta/\eta)(x/\eta)^{\beta-1}$ est croissant - et modélise donc un \sum qui vieillit - quand $\beta > 1$. Il est décroissant - et modélise un \sum qui rajeunit - quand $\beta < 1$. Il est exponentiel quand $\beta = 1$. On note aussi $\lambda = 1/\eta$, et la paramétrisation $\mu = \lambda^{\beta}$ sera beaucoup utilisée dans le reste de ce document.

S'il y a vieillissement, celui-ci s'accélère quand $\beta > 2$. En effet, la vitesse du vieillissement peut être perçue par la dérivée du taux de défaillance $h'_{\mathcal{W}}(x)$, qui est alors positive. Ainsi, la connaissance de β offre une indication de nature *qualitative* sur le comportement de \sum , puisqu'il permet de distinguer un composant qui vieillit fortement d'un composant bien entretenu, dont les défaillances sont surtout accidentelles ($\beta \simeq 1$), quelle que soit la valeur du paramètre d'échelle η . Ce dernier est quant à lui homogène à X , alors que β est sans dimension, et correspond au 63^{ième} percentile de la distribution, soit

$$P(X < \eta) = 0.63.$$

Le livre de Dodson (2006) constitue une référence intéressante puisqu'il fournit une revue complète des propriétés du modèle et de son emploi dans de nombreuses études de fiabilité. Notons que les modèles gamma et lognormal, de densités respectives

$$f(x|\alpha, \gamma) = \frac{\gamma^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\gamma x) \mathbb{1}_{\{x \geq 0\}},$$

$$f(x|m, \sigma) = \left(x\sigma\sqrt{2\pi}\right)^{-1} \exp\left\{-\frac{\log^2(x/m)}{2\sigma^2}\right\} \mathbb{1}_{\{x \geq 0\}}$$

sont eux aussi très utilisés, soit pour modéliser le vieillissement, soit pour modéliser le temps moyen mis pour réparer un système en panne (MTTR). Cependant, le sens des paramètres est bien moins évident.

REMARQUE 1. *En fiabilité des composants, la constatation d'un rajeunissement est plus souvent la conséquence d'une période de déverminage dans la vie de Σ ou d'une maintenance préventive* de Σ qu'un phénomène physique réel. Très majoritairement, nous considérons dans ce travail des Σ qui vieillissent.*

1.3.3 Modèle à risques concurrents

Une variable aléatoire X suivant le modèle à risques concurrents $\mathcal{B}(\eta_0, \eta_1, \beta)$ étudié originellement par Bertholon (2001) est définie par

$$X = \min(E, W)$$

où

$$E \sim \mathcal{E}(1/\eta_0),$$

$$W \sim \mathcal{W}(\eta_1, \beta)$$

indépendamment l'une de l'autre. Le taux de défaillance de ce modèle consiste en la somme des taux de défaillance exponentiel et de Weibull :

$$h_B(x|\eta_0, \eta_1, \beta) = \eta_0 + \frac{\beta}{\eta_1} \left(\frac{x}{\eta_1}\right)^{\beta-1} \mathbb{1}_{\{x \geq 0\}}.$$

Ainsi, lorsque $\beta \geq 1$, ce modèle permet de modéliser un Σ soumis à du vieillissement mais également susceptible de tomber en panne par accident. L'article de Bertholon, Bousquet & Celeux (2006), placé en Annexes A, offre un résumé des propriétés de cette distribution et étudie l'estimation des paramètres dans un cadre durabiliste (par maximisation de la vraisemblance et inférence bayésienne).

Dans cet article, le vieillissement de l'objet d'étude est perçu comme la principale cause de défaillance. Ainsi, on suppose $\beta > 1$ et $\eta_0 > \eta_1$; ce qui implique que la durée de vie moyenne avant accident est plus longue que la durée de vie moyenne avant défaillance par usure. On montre, *via* des tests numériques et des mesures de grandeurs prédictives, que cette modélisation fait sens dans des cas où, par habitude, seule la loi de Weibull est utilisée. L'intérêt est donc d'améliorer fortement la prédiction du comportement de Σ .

REMARQUE 2. Notons que pour ces deux derniers modèles, nous faisons toujours l'hypothèse que le vieillissement (ou le rajeunissement) commence au temps 0. Autrement dit, ces modèles ne comportent pas de terme de position.

Notons cependant que dans le strict cadre d'EDF, ce modèle est apparu peu utile au cours de notre étude ; en effet, comme précisé auparavant, les données EDF sont généralement triées au sens où l'on connaît, de façon assez précise, leur origine. D'autres données industrielles ont permis d'utiliser ce modèle avec succès, en particulier des durées de vie de moyens de transport (SNCF, entreprises de location de voitures).

1.4 Formalisation statistique du cadre d'étude

1.4.1 Variable de durée de vie

Les données sont définies sur un espace mesurable (χ, \mathcal{A}) , que l'on munit d'une mesure dominante ν positive. Dans notre étude, \mathcal{A} est la tribu borélienne* $\mathcal{B}(\chi)$ et ν est (le plus souvent) la mesure de Lebesgue. Dans le cadre de la durée de vie, $\chi = \mathbb{R}^+$.

On note alors X une variable aléatoire définie sur l'espace probabilisé $(\chi, \mathcal{A}, \mathcal{M}_\theta : \theta \in \Theta \subset \mathbb{R}^d)$. On notera plus simplement $X \sim \mathcal{M}(\theta)$ et $\mathcal{M}(\theta)$ sera désignée sous le terme de *distribution* (de probabilité) de X . Cette variable aléatoire sera associée en général à une durée de vie, excepté au chapitre 4 où le cadre de l'étude s'élargit (et ne concerne plus seulement la durabilité). La densité de X par rapport à ν sera notée $f(x|\theta)$ et sa fonction de répartition $F(x|\theta)$. La fonction de survie (ou *fiabilité*) sera désignée par

$$S(x|\theta) = P_\theta(X > x) = 1 - F(x|\theta).$$

Un échantillon $\mathbf{x}_n = (x_1, \dots, x_n)$ i.i.d. est alors la représentation de $\mathbf{X}_n = \{X_n\}$, une suite de variables aléatoires indépendantes et identiquement distribuées de $(\chi, \mathcal{A}, \mathcal{M}_\theta)$. Dans les applications, on notera souvent par commodité

$$x_1, \dots, x_n \stackrel{i.i.d.}{\sim} \mathcal{M}(\theta).$$

1.4.2 Phénomène de censure (données incomplètes)

Le phénomène de censure peut se traduire comme suit. On définit $\mathbf{Y}_n = \Psi(\mathbf{X}_n, \mathbf{C})$ comme la suite de variables aléatoires réellement observées, non identiquement distribuées, où $\Psi(\cdot, C)$ représente un mécanisme de censure à droite transformant une donnée complète x_i en une observation partielle y_i . D'une façon générale, on notera \mathbf{y}_n les échantillons de données disponibles (REX ou simulations). On spécifie deux types de mécanismes :

1. *Censure de type I progressive*. Il s'agit du mécanisme prépondérant dans cette étude, qui correspond à une durée d'observation prédéfinie. Soit $\mathbf{C} = \{C_n\}$ une suite de valeurs fixes et indépendantes de \mathbf{X}_n . On définit alors $Y_i = \min(X_i, C_i) \forall i \in \{1, \dots, n\}$. Lorsque ces censures sont uniques, on parlera simplement de Type I. D'un point de vue industriel, le REX de Type I correspond à des observations provenant d'installations mises en service au même moment. Le

REX de type I progressif (ou *échelonné*) est plus réaliste, au sens où ces installations sont mises en service à des dates diverses.

2. *Censure de type II.* On définit un nombre fixe $r \in \{1, \dots, n\}$ indépendamment de \mathbf{X}_n . On définit alors $\mathbf{C} = \{X_r^*\}$ le singleton restreint à la statistique d'ordre r de X . Puis $Y_i = X_i^* \forall i \in \{1, \dots, r\}$ et $Y_j = X_r^* \forall j \in \{r+1, \dots, n\}$. Ce type de censure aléatoire caractérise en majorité des essais de laboratoire, que l'on stoppe au bout de la $r^{\text{ième}}$ défaillance constatée.

1.4.3 Structure à données manquantes

La structure des données observables induit donc en général l'existence d'un ensemble de *données manquantes*, noté le plus souvent \mathbf{z} , et dont la densité conditionnellement à l'échantillon observé sera notée $k(\mathbf{z}|\theta, \mathbf{y}_n)$. Cet ensemble diffère de façon fondamentale selon les modèles.

1. Dans le cas d'un modèle soumis à une source unique de défaillance (exponentiel et Weibull), ces données manquantes sont les valeurs de défaillance censurées, soit les données incomplètes. Ainsi $\mathbf{z} \sim \mathcal{M}(\theta)$.
2. Lorsque le modèle est à risques concurrents, chaque donnée *i.i.d.* observée est attribuée à l'un des modèles en concurrence. Elle censure donc des réalisations provenant d'autres modèles. Ainsi, dans le cas du modèle $\mathcal{B}(\eta_0, \eta_1, \beta)$, une valeur de défaillance observée, d'origine accidentelle (exponentielle), "cache" une donnée de défaillance de vieillissement (de Weibull), et réciproquement. Pour ce type de modèle, les données manquantes sont également les origines des défaillances. Bien évidemment, les données de censure ne permettent pas de pouvoir distinguer la source de défaillance qui aurait produit la donnée inconnue.

Le terme *données manquantes* (*missing data*) est souvent introduit dans les articles scientifiques pour inclure l'un ou l'autre de ces processus dans une certaine généralité. Dans la suite de ce travail, on tentera d'utiliser au maximum le terme *incomplet* pour éviter les confusions.

Chapitre 2

Limites des méthodes d'inférence classiques pour l'application industrielle

Table locale

2.1	Introduction	28
2.2	Descriptif des principales méthodes d'estimation	29
2.2.1	Méthodes empiristes	29
2.2.2	Maximisation de la vraisemblance	31
2.3	Intervalles de confiance et incertitude d'un estimateur	35
2.3.1	Intervalles de confiance théoriques	36
2.3.2	Procédures de bootstrap pour l'estimation d'une incertitude	38
2.4	Indicateurs décisionnels de l'analyste industriel	44
2.5	Applications numériques	46
2.5.1	Présentation des données et estimations des paramètres	46
2.5.2	Application au modèle exponentiel	47
2.5.3	Application au modèle de Weibull	50
2.5.4	Évaluation des méthodes par facteur d'erreur	52
2.6	Conclusion	53
2.7	Annexes	54

2.1 Introduction

L'objectif de ce chapitre est de présenter un aperçu global des différentes techniques d'inférence dites *classiques* traditionnellement mises en œuvre pour estimer les paramètres des modèles présentés au chapitre 1. Par *classiques*, nous entendons les méthodes qui se fondent sur une approche non *bayésienne* de la Théorie de la Décision, au sens où le paramètre θ est supposé fixe, et où seule l'information provenant des données est prise en compte. Nous renvoyons au chapitre 2 du livre de Robert (2001), qui présente en détail les bases de la Théorie de la Décision et permet d'appréhender clairement les différentes approches possibles de cette théorie. Au sens communément utilisé, l'approche considérée ici est qualifiée de *fréquentiste*¹, puisqu'elle se fonde intrinsèquement sur la répartition en fréquence des données, répartition interprétée par la fonction $F(x|\theta) = P(X < x|\theta)$.

Cette présentation succincte se veut critique vis-à-vis de l'emploi de ces méthodes dans le contexte particulier d'échantillons \mathbf{y}_n de faible taille et censurés. La plupart des études portant sur l'impact de la censure sur l'estimation de quantités fiabilistes ou durabilistes se placent dans des conditions où le nombre de censures évolue parallèlement avec le nombre de données n , et où la validité des intervalles de confiance est discutée asymptotiquement. En témoigne par exemple l'article de Harder (1990) sur la validité de tels intervalles construits à partir de données exponentielles fortement censurées ($n = 50$ au minimum). De nombreux articles comme celui de Tingley & Field (1990) abordent le problème d'échantillons beaucoup plus proches des conditions réelles rencontrées dans l'industrie (à partir de $n = 5$), mais sans introduire de censure. Ainsi, les conditions optimales d'utilisation de ces techniques, dans un cadre fréquentiste, ne sont-elles jamais réalisées dans notre étude ; or les limites d'application de ces méthodes restent floues. Il importe donc de fournir des outils permettant de juger la validité de telle ou telle méthode d'estimation, et de les appliquer sur des exemples typiques d'échantillons industriels.

Une première partie est consacrée aux méthodes d'estimation de θ proprement dites.

- On rappelle tout d'abord quelques méthodes, qualifiées d'*empiristes* car fondées uniquement sur la valeur des données et non sur une compréhension fine de leur structure. Ces techniques servent généralement de méthodes préalables, au sein de la communauté industrielle, à la fois pour opérer un choix de modèles de comportement - sinon le confirmer - et procéder à une estimation assez grossière. Elles restent d'utilisation très limitée dans notre contexte d'étude.
- Par la suite, nous faisons le rappel de méthodes classiques de maximisation de la vraisemblance. Elles s'appuient sur la structure manquante des données et paraissent beaucoup plus aptes à fournir des estimations raisonnables de θ . Ces méthodes feront l'objet d'une attention critique dans le reste du chapitre.

La seconde partie de ce chapitre sera focalisée sur la construction d'intervalles de confiance pour les estimations précédentes, et d'indicateurs acceptables permettant de juger la validité de ces méthodes dans le contexte industriel d'EDF.

Enfin, quelques tests numériques sont présentés, permettant d'illustrer les critiques importantes

¹terme traditionnel en statistiques, mais impropre en langue française ; le terme correct est *fréquentielle*.

faites sur les méthodes classiques d'estimation. Peu ou pas adaptées à nos données, elles se révèlent sources de confusions et d'inexactitudes. D'autres techniques doivent alors être employées pour améliorer l'inférence, qui nécessitent de modifier notre cadre d'analyse.

2.2 Descriptif des principales méthodes d'estimation

2.2.1 Méthodes empiristes

À disposition des ingénieurs en durabilité et fiabilité, un certain nombre de méthodes sont disponibles, fondées en général sur une approche intuitive du problème. Elles utilisent la répartition empirique des valeurs - sans mettre à profit leur structure éventuellement censurée - pour proposer des estimations de θ , en reliant cette répartition à la répartition fréquentiste. Ainsi, une représentation empirique nommée *graphe de Weibull* est couramment utilisée. Elle a l'avantage de présenter sur un graphe unique les principales tendances du composant \sum au cours du temps. Elle permet la détection d'un vieillissement ou d'un rajeunissement, d'une évolution linéaire du taux de défaillance (ni rajeunissement, ni vieillissement) ou encore d'une compétition entre plusieurs modes de défaillance. Par des ajustements linéaires, cette représentation peut fournir des valeurs de θ .

2.2.1.1 Graphes de Weibull

L'échelle des abscisses du graphe de Weibull est $\log x$ quand celle des ordonnées est $Q(x) = \log(\log S(x))$ où $S(x)$ est la fonction de survie. Connaissant les données (y_1, \dots, y_n) , on peut en fournir l'estimation empirique sans biais de Kaplan-Meier

$$S_n(x) = \prod_{i \in \{1, \dots, n\}, y_i \leq x} \left(\frac{n-i}{n-i+1} \right)^{\delta_i}$$

où $\delta_i = 1$ si y_i n'est pas censuré et 0 sinon. Voir également Lo *et al.* (1989) pour une redéfinition de cet estimateur, ainsi que Bitouzé *et al.* (1999) pour une mesure de la concentration de cet estimateur vis-à-vis de la distribution réelle, dans un cadre non asymptotique. En comparant $S_n(x)$ au tracé de la survie théorique $S(x|\theta)$, fonction d'un choix de modèle et d'une ou plusieurs valeurs de θ , la démarche empirique constitue un bon préalable au choix de modèle.

Si les données suivent le modèle de Weibull $\mathcal{W}(\eta, \beta)$, leur répartition suit approximativement la droite d'équation $Q(x) = \beta \log x - \beta \log \eta$, de coefficient directeur β . Un ajustement linéaire se justifie donc pour estimer les paramètres. Si jamais les données proviennent du modèle à risques concurrents $\mathcal{B}(\eta_0, \eta_1, \beta)$, l'équation $Q(x) = \log(x/\eta_0 + (x/\eta_1)^\beta)$ est celle d'une courbe convexe.

Pour illustrer ces deux comportements, le graphe de Weibull de données de durée de vie provenant de Bertholon *et al.* (2006) est tracé sur la figure 2.1. L'ajustement linéaire à la répartition empirique des données ne tient visiblement pas compte des premières défaillances, à la différence du tracé d'une distribution $\mathcal{B}(\eta_0, \eta_1, \beta)$, qui paraît bien plus adaptée. En utilisant ce modèle, la compétition entre types de défaillance est mise en évidence par le "coudé" de la distribution, aux alentours de $\log x \simeq 3.8$.

Temps de déverminage. Un temps de déverminage potentiel peut être détecté par l'utilisation des graphes de Weibull. Il transforme en effet les droites empiristes de Weibull en courbes concaves jusqu'au

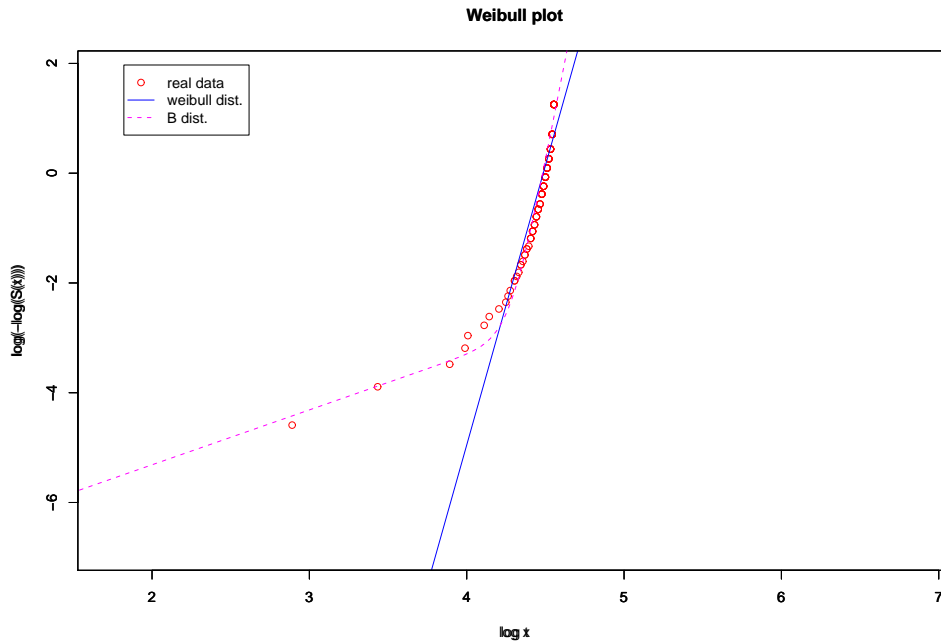


FIG. 2.1 – Graphes de Weibull de données prospectives de durée de vie humaine.

log-temps où commence véritablement le vieillissement (ou le rajeunissement) de Σ .

Limitations. Cette démarche est cependant extrêmement limitée. En effet, l'estimateur de Kaplan-Meier, s'il est sans biais, est sensible à la faible taille et à la censure de l'échantillon (voir ainsi Peterson, 1977, et Chen, Hollander & Landberg, 1982) ; la présence d'un point atypique, par exemple en queue de distribution (déverminage oublié, donnée censurée à droite en fin d'échantillon, etc.), perturbe considérablement l'ajustement linéaire.

Références. Les méthodes graphiques restent plus des préalables aux tests classiques d'adéquation de modèle (par exemple Kolmogorov-Smirnov, Cramer-von Mises) que des méthodes d'estimation. De nombreuses références sont disponibles pour l'application de ces méthodes comme préalable au choix de modèle. Dans le contexte de la durée de vie, citons Lawless (1982, chap. 6), D'Agostino & Stephens (1986) et Bertholon (2001, pages 22-27).

2.2.1.2 Méthodes de moments et quantiles

Une autre façon simple de fixer les paramètres des modèles est d'estimer empiriquement les moments ou les quantiles de la distribution de $X \sim \mathcal{M}(\theta)$. Le paramètre η du modèle de Weibull est ainsi le 63^{ème} percentile de la distribution. Cependant, ce type d'estimation est soumis à plusieurs difficultés :

- l'expression des quantiles ou des moments n'est pas forcément explicite ; c'est le cas avec le modèle \mathcal{B} à risques concurrents (Bertholon *et al.* 2006). Par ailleurs, l'estimation empirique des quantiles est faussée par la présence de censure.
- les intervalles de confiance de tels estimateurs empiriques sont généralement asymptotiques (et donc non adaptés à notre contexte industriel).

2.2.1.3 Méthodes des moindres carrés

Les méthodes des moindres carrés ont pour objectif d'estimer θ en minimisant une expression telle que

$$\phi = \sum_{i=1}^n [\psi^*(X_i^*) - \psi(X_i^*|\theta)]^2$$

où X_i^* est la statistique d'ordre i de l'échantillon, $\psi(x|\theta)$ est typiquement une fonction d'évolution de X (densité, fonction de répartition,...) et $\psi^*(x)$ est par exemple un estimateur non paramétrique de $\psi(x|\theta)$. Canfield & Borgman (1975) ont proposé le choix $\psi(x|\theta) = \log S(x|\theta)$ où $S(x|\theta)$ est la fonction de survie, dont la survie empirique est un estimateur sans biais disponible, indépendamment du choix de modèle paramétrique.

Ces méthodes présentent certaines difficultés, à commencer par le grand nombre de solutions à rejeter car n'appartenant pas à Θ (paramètres d'échelle négatifs, etc.). Il est donc nécessaire de leur rajouter de nombreuses contraintes. Beaucoup d'auteurs ont émis à leur sujet de fortes restrictions d'emploi dans des cas de faible taille et censurés (Lawless, 1982, pages 332-333) ; par exemple, Friedman & Gertsbakh (1980) indiquent que l'estimation des moindres carrés d'un modèle \mathcal{B} à risques concurrents - où l'un des paramètres est fixé - est bien plus biaisée que l'estimation du maximum de vraisemblance (voir § 2.2.2.1). Indiquons cependant qu'une procédure de moindres carrés pour des données manquantes proposée par Healy & Westmacott (1956) s'apparente à un algorithme EM (voir § 2.2.2.2). On trouvera plus de précision sur ce sujet dans McLachlan & Krishnan (1997, pages 51-56).

2.2.1.4 Conclusion

Si les méthodes graphiques peuvent aider à la validation du choix d'une famille de modèles, les méthodes d'estimation par régression ou par moments ne sont pas utilisables dans notre cadre d'étude, d'autant plus qu'elles n'utilisent qu'une partie de l'information transmise par l'échantillon. Les méthodes de maximisation de la vraisemblance, élaborées en fonction de la structure des données, promettent à première vue d'être beaucoup plus efficaces.

2.2.2 Maximisation de la vraisemblance

2.2.2.1 Principe

La vraisemblance $\mathcal{L}(\mathbf{y}_n; \theta)$ correspond à la densité de la probabilité jointe des données $\mathbf{y}_n = (y_1, \dots, y_n)$ de provenir du modèle \mathcal{M} ; elle représente donc toute la connaissance objective sur θ apportée par la structure des données. En divisant l'échantillon \mathbf{y}_n en données non censurées \mathbf{x}_r et censures à droite \mathbf{c}_{n-r} comme au § 1.4, on obtient, dans un cas de censure de type I progressif :

$$\mathcal{L}(\mathbf{y}_n, \theta) = \left[\prod_{i=1}^r f(x_i|\theta) \right] \left[\prod_{j=1}^{n-r} S(c_j|\theta) \right].$$

Dans le cas de censure de type II, on obtient

$$\mathcal{L}(\mathbf{y}_n, \theta) = f(x_1^*, \dots, x_r^*|\theta) (S(x_r^*|\theta))^{n-r}$$

où x_i^* est la statistique d'ordre i du sous-échantillon \mathbf{x}_r (voir par exemple Gaudoin 2002 pour des écritures plus explicites). On décrit dans le tableau 2.1 la vraisemblance observée de données censurées

selon les trois modèles utilisés (là encore pour une censure de type I progressif).

exponentiel $\mathcal{E}(\eta)$	$\eta^{-r} \exp\left(-\sum_{i=1}^n y_i/\eta\right)$
Weibull $\mathcal{W}(\eta, \beta)$	$\left[\beta^r \eta^{-r} \prod_{j=1}^r x_j^{\beta-1}\right] \exp\left(-\sum_{i=1}^n (y_i/\eta)^\beta\right)$
$\mathcal{B}(\eta_0, \eta_1, \beta)$	$\left[\prod_{j=1}^r \left(1/\eta_0 + (\beta/\eta_1)(x_j/\eta_1)^{\beta-1}\right)\right] \exp\left(-\sum_{i=1}^n y_i/\eta_0 - \sum_{i=1}^n (y_i/\eta_1)^\beta\right)$

TAB. 2.1 – Vraisemblance des données censurées (type I progressif) pour les trois modèles étudiés.

Le maximum de vraisemblance $\hat{\theta}_{ML}$ est la solution des équations de vraisemblance $\partial l(\mathbf{y}_n, \theta)/\partial \theta = 0$ où $l(\mathbf{y}_n, \theta)$ est la log-vraisemblance. Sous des conditions de régularité atteintes pour les trois modèles, $\hat{\theta}_{ML}$ est unique et tel que

$$\sqrt{n} \left(\hat{\theta}_{ML} - \theta_o \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N} \left(0, I^{-1}(\theta_o) \right)$$

où θ_o est la vraie valeur (inconnue) du paramètre θ et $I(\theta)$ est la matrice d'information de Fisher du modèle $\mathcal{M}(\theta)$ considéré :

$$I(\theta) = -\mathbf{E}_\theta \left[\frac{\partial^2 f(x|\theta)}{\partial \theta_i \partial \theta_j} \right]_{i,j \in \{1, \dots, d\}}.$$

Des informations plus précises sur cette optimalité asymptotique, dans un cadre de données manquantes et pour les modèles utilisés dans ce document, sont présentées dans Hoadley (1971), Bacha (1996, pages 58-63), Bertholon *et al.* (2006).

2.2.2.2 L'algorithme EM et son dérivé stochastique

Lorsqu'il n'y a pas de solution explicite au problème de maximisation, l'emploi d'une procédure numérique d'estimation est nécessaire. L'algorithme EM (*Expectation Maximization*; Dempster, Laird & Rubin, 1977) est certainement la procédure d'estimation la plus utilisée dans le cadre des modèles à données manquantes. Elle a donné lieu à l'élaboration de nombreux algorithmes dérivés améliorant ses performances.

L'accélération des algorithmes EM et de ses dérivés, les nombres d'itérations, la stabilité de l'estimation vis-à-vis de l'initialisation des algorithmes, la sensibilité de celle-ci à la régularité de la vraisemblance ont fait l'objet de très nombreuses recherches, impossibles à lister même dans le domaine restreint de la durabilité. Si le livre de McLachlan & Krishnan (1997) constitue une référence majeure sur l'emploi généralisé des algorithmes EM, un travail important sur l'emploi d'EM et de ses dérivés sur les modèles de Weibull est aussi présenté dans la thèse de Bacha (1996, chap. 1 et 2). Une revue générale de la construction et des propriétés asymptotiques du dérivé stochastique SEM est présentée dans Nielsen (2000). Nous faisons ici le rappel des caractéristiques essentielles des deux principales méthodes.

Procédure EM. Elle consiste à maximiser l'espérance de la vraisemblance complète du modèle conditionnellement aux données observées et à une valeur de θ_p courante. On choisit une initialisation θ_0 . L'itération $p \rightarrow p + 1$ s'écrit

1. Étape E : sachant θ_p , on calcule l'espérance conditionnelle

$$\theta \rightarrow \mathcal{Q}(\theta|\theta_p) = \mathbb{E}[l(\mathbf{x}_n, \theta)|\mathbf{y}_n, \theta_p].$$

2. Étape M : on détermine $\theta_{p+1} = \arg \max_{\theta \in \Theta} \mathcal{Q}(\theta|\theta_p)$.

La suite $\{\theta_p\}$ ainsi produite fait croître la vraisemblance et converge vers $\hat{\theta}_{ML}$, moyennant quelques hypothèses. En général la convergence dépend fortement de l'initialisation θ_0 (des améliorations sont possibles en itérant EM sur un grand nombre d'initialisations), et la méthode n'évite pas le piège de maxima locaux de la vraisemblance et peut donc produire des résultats éloignés de $\hat{\theta}_{ML}$ (Wu, 1983).

Procédure SEM. Le dérivé stochastique de EM, l'algorithme SEM (*Stochastic Expectation Maximization*; Celeux & Diebolt 1984, 1992) a alors été conçu pour surmonter ces difficultés (originellement dans un contexte de mélanges de lois de probabilité). Son principe est de construire une chaîne de Markov homogène ergodique, convergente vers une loi stationnaire unique de moyenne $\hat{\theta}_{ML}$, en remplaçant les données manquantes par des perturbations aléatoires.

On choisit une initialisation θ_0 . L'itération $p \rightarrow p + 1$ est

1. Étape E : pour $i \in \{1, \dots, n\}$ on calcule la densité conditionnelle des données manquantes $k_i(X|Y = \mathbf{y}_i, \theta_p)$.
2. Étape S : pour $i \in \{1, \dots, n\}$ on simule des valeurs $\tilde{X}_i \sim k_i(X|Y = \mathbf{y}_i, \theta_p)$.
3. Étape M : on maximise la log-vraisemblance complétée courante :

$$\theta_{p+1} = \arg \max_{\theta \in \Theta} \sum_{i=1}^n \log f(\tilde{X}_i|\theta).$$

On propose alors $\hat{\theta}_{SEM} = \sum_{i=m+1}^P \theta_i / (P - m)$, $m \ll P$ avec m et P suffisamment grands, comme estimation de $\hat{\theta}_{ML}$. Bacha *et al.* (1998, page 49) conseillent de toujours choisir $P > 300$ et $m > 100$. Selon la dimension de θ , plusieurs essais sur P et m sont nécessaires avant d'obtenir la stabilité de la chaîne de Markov. Cette méthode intuitive mais difficile à étudier théoriquement, a fait l'objet de nombreux tests (Bacha 1996, Rodionov 2005) qui ont permis de constater qu'en règle générale elle améliore la précision et le réalisme de l'estimation de $\hat{\theta}_{ML}$ (Celeux & Diebolt 1994, Diebolt & Ip 1996).

Cependant, l'étape stochastique de simulation de l'information manquante paraît susceptible de "noyer" l'information présente, si la censure est plus importante que les données réelles et que l'échantillon est de faible taille (la valeur de simulation courante θ_{p+1} risquant d'être très erronée). Rodionov (2005) indique par exemple que l'application au modèle de Weibull reste acceptable jusqu'à $n = 20$, pour un taux de censure ne dépassant pas 10%. Malheureusement peu d'essais sur des échantillons industriels (de l'industrie nucléaire) sont venus conforter cette idée. Ce qui fait l'élégance et l'intérêt pratique de SEM dans un cadre général risque d'en faire une méthode à la précision floue (et à l'apport

minime par rapport à EM), dans notre cadre d'étude particulier.

2.2.2.3 Application aux modèles utilisés

On rappelle ci-après les points importants de la maximisation de vraisemblance des modèles utilisés. Singpurwalla *et al.* (1999) constitue une référence pratique pour les modèles exponentiels et de Weibull. Pour le modèle à risques concurrents $\mathcal{B}(\eta_0, \eta_1, \beta)$, on pourra se référer à Bertholon (2001), Bousquet (2003) et Bertholon *et al.* (2006), dont les travaux d'estimation s'inscrivent dans le prolongement des études de Giesbrecht & Kempthorne (1976) et Friedman & Gertsbakh (1980). En particulier, on peut prouver que, pour ces deux derniers modèles, que les données soient censurées ou non, les équations de vraisemblance admettent une solution unique et consistante.

- *Modèle exponentiel.* Soit $T = \sum_{i=1}^n x_i$. Alors $\hat{\eta}_{ML} = T/r$ et est sans biais si $r = n$ (pas de censure).
- *Modèle de Weibull.* $\hat{\theta}_{ML} = (\hat{\eta}_{ML}, \hat{\beta}_{ML})$ est solution des équations de vraisemblance

$$\frac{1}{\beta} = \frac{\sum_{i=1}^n x_i^\beta \log x_i}{\sum_{i=1}^n x_i^\beta} - \frac{\sum_{i=1}^r \log y_i}{r}, \quad \eta = \frac{\left(\sum_{i=1}^n x_i^\beta\right)^{1/\beta}}{r^{1/\beta}}.$$

Cette solution est traditionnellement approchée par des méthodes de descente type Newton-Raphson. L'estimateur obtenu est noté $\hat{\theta}_{NR}$.

- *Modèle à risques concurrents $\mathcal{B}(\eta_0, \eta_1, \beta)$.* $\hat{\theta}_{ML} = (\hat{\eta}_{0,ML}, \hat{\eta}_{1,ML}, \hat{\beta}_{ML})$ est la solution des équations de vraisemblance

$$\begin{aligned} \eta_0 &= \frac{\sum_{i=1}^r x_i}{\sum_{i=1}^r \tilde{p}_E(x_i)}, \\ \frac{1}{\beta} &= \frac{\sum_{i=1}^n (y_i)^\beta \ln(y_i)}{\sum_{i=1}^n (y_i)^\beta} - \frac{\sum_{i=1}^r \tilde{p}_W(x_i) \ln(x_i)}{\sum_{i=1}^r \tilde{p}_W(x_i)}, \\ \eta_1 &= \left(\frac{\sum_{i=1}^n (y_i)^\beta}{\sum_{i=1}^r \tilde{p}_W(x_i)} \right)^{\frac{1}{\beta}}, \\ \text{avec } \begin{cases} \tilde{p}_E(x_i) &= h_E(x_i)/[h_B(x_i)], \\ \tilde{p}_W(x_i) &= 1 - \tilde{p}_E(x_i), \end{cases} \end{aligned}$$

$\tilde{p}_E(x_i)$ et $\tilde{p}_W(x_i)$ étant les probabilités pour une donnée non censurée x_i de provenir de l'un des modèles en compétition. Les solutions de ces équations sont atteintes au bout d'une procédure itérative, où l'origine de chaque donnée est réactualisée puis où l'on procède à deux maximisations exponentielle et de Weibull séparées. $\hat{\theta}_{ML}$ peut alors être estimé par $\hat{\theta}_{EM}$ et $\hat{\theta}_{SEM}$. Voir l'article de Bertholon *et al.* (2006), fourni en Annexes A, pour plus de précision.

2.2.2.4 Difficultés et limites expérimentales

À la suite des études de Bain & Engelhardt (1991), Bacha *et al.* (1998) indiquent plusieurs améliorations possibles de l'estimateur du maximum de vraisemblance. En particulier, ils offrent une tabulation de coefficients (estimés par simulations de Monte-Carlo) permettant d'éliminer le biais de l'estimateur du paramètre de forme β du modèle de Weibull en fonction de la taille et des caractéristiques de censure de l'échantillon.

Enfin, après de nombreux tests sur des échantillons simulés et réels, et en se référant aux travaux de Lannoy & Procaccia (1994), ces auteurs préconisent l'utilisation des méthodes de maximisation de la vraisemblance pour les échantillons de Weibull dont les caractéristiques sont présentées dans le tableau ci-dessous. Le sigle NR désigne l'algorithme de descente de Newton-Raphson (soit EM pour le modèle de Weibull).

taux de censure τ	taille n de l'échantillon			
	$6 < n \leq 20$	$20 \leq n \leq 40$	$40 \leq n \leq 80$	$n \geq 80$
$\tau = 0\%$	NR corrigé	NR	NR	NR
$\tau \leq 50\%$	SEM	SEM	SEM	SEM
$50\% \leq \tau \leq 75\%$	SEM	SEM	NR	NR
$\tau \geq 75\%$	SEM	SEM	SEM/NR	NR

TAB. 2.2 – Stratégies de choix de la méthode de maximisation de la vraisemblance pour un échantillon de Weibull.

Les caractéristiques listées dans le tableau 2.2 restent cependant des moyennes et une comparaison de plusieurs méthodes de maximisation de la vraisemblance est toujours nécessaire pour évaluer leur pertinence, lorsque les données utilisées sont réelles. Entre deux estimations de θ , on choisira celle pour laquelle la valeur de la *vraisemblance observée* est la plus élevée.

Par ailleurs, une estimation ponctuelle doit nécessairement être accompagnée d'une indication permettant de mesurer la reproductibilité de ce résultat. Ainsi, l'analyste doit proposer une mesure de l'incertitude de cette estimation. Dans la suite de ce chapitre, nous faisons tout d'abord quelques rappels sur les intervalles de confiance*. Puis nous introduisons un critère nommé *facteur d'erreur*, traditionnellement utilisé par les industriels pour estimer si la précision d'une estimation est acceptable. Enfin, nous appliquons ces outils à quelques échantillons exponentiels et de Weibull supposés représentatifs des études industrielles qui ont motivé ce travail de thèse.

2.3 Intervalles de confiance et incertitude d'un estimateur

Lorsque θ est supposé fixé, bien plus qu'une estimation ponctuelle de ce paramètre, la connaissance d'un intervalle de confiance sur θ ou sur une fonction d'intérêt de θ (comme la durée de vie moyenne ou la survie en une valeur de temps fixée) est une aide à la prise d'une décision de maintenance. Lorsqu'une estimation ponctuelle $\hat{\theta}$ de θ est nécessaire (c'est-à-dire en l'absence de résultat théorique indépendant d'une estimation), obtenir une mesure de l'incertitude de cette estimation est indispensable. Le souhait interne à l'entreprise de se fixer des barrières techniques nécessite alors de rejeter des méthodes d'estimation trop imprécises.

Idéalement, un intervalle de confiance est produit par une méthode de *fonction pivotale**, indépendamment d'une estimation ponctuelle de θ . Il est alors dit *théorique*. En ce sens, il constitue un outil pratique puisqu'il évacue les difficultés du choix de la méthode d'estimation. Le premier paragraphe de cette section est consacré à l'établissement de tels intervalles. Cependant, son usage reste limité : l'établir nécessite de connaître explicitement la loi d'une statistique *ancillaire*, (c'est-à-dire dont la loi est indépendante de θ), ce qui est très limitatif.

REMARQUE 3. *Une autre difficulté posée par les intervalles de confiance est de requérir une observance stricte des hypothèses statistiques : l'échantillon \mathbf{y}_n doit effectivement correspondre à un ensemble de réalisations de la loi $\mathcal{M}(\theta)$, ce qui est délicat voire impossible à vérifier dans les cas industriels qui nous préoccupent, où les échantillons contiennent des données censurées difficile à manier. Ce défaut peut être tempéré dans certains cas, au travers d'études empiriques (voir § 2.3.1).*

Dans notre contexte industriel de tels intervalles, lorsqu'ils sont disponibles, sont généralement très larges, comme nous l'illustrerons dans la suite du chapitre. Ils fournissent des limites raisonnables pour la recherche d'une estimation $\hat{\theta}$ de θ , et témoignent de l'imprécision certaine de toute tentative d'estimation de θ basée sur le respect des hypothèses statistiques classiques.

Lorsque ces intervalles théoriques ne sont pas constructibles ou le sont avec difficulté (comme dans le cas Weibull), une méthode alternative est traditionnellement utilisée. Si l'on note \mathbf{Y}_n l'ensemble des variables aléatoires associées à l'échantillon observé \mathbf{y}_n , on peut chercher à estimer la variance $\hat{\sigma}^2$ d'un estimateur $\hat{\theta} = \hat{\theta}(\mathbf{Y}_n)$. Reproduire si possible les conditions de l'expérimentation menant au REX, au travers de la reconstitution de M échantillons de mêmes caractéristiques que \mathbf{y}_n , fait sens pour obtenir une distribution d'échantillonnage de cet estimateur. Empiriquement, on peut alors obtenir une estimation de $\hat{\sigma}^2$. En faisant par exemple l'hypothèse d'une répartition gaussienne de cette distribution d'échantillonnage (légitimée sous certaines conditions par le théorème central limite), on peut obtenir un intervalle de confiance autour de l'estimé $\hat{\theta}(\mathbf{y}_n)$.

La reproduction de données proches du REX est donc une alternative valable pour pallier l'absence des intervalles de confiance théoriques. C'est notamment l'enjeu des méthodes de *bootstrap*, présentées au § 2.3.2. Dans cette partie, nous proposons plusieurs possibilités pour reproduire ces conditions d'expérimentations le plus fidèlement possible.

2.3.1 Intervalles de confiance théoriques

Nous fournissons ici des intervalles de confiance théoriques pour les paramètres des modèles exponentiel et de Weibull, dans un contexte non asymptotique et censuré, construits d'après des méthodes de fonctions pivotales*. Si le premier modèle, très bien connu, a fait l'objet de nombreuses études, les récents résultats disponibles sur le second restent employés de manière marginale. Aucun résultat n'est disponible sur le modèle à risques concurrents $\mathcal{B}(\eta_0, \eta_1, \beta)$, de par sa complexité.

2.3.1.1 Modèle exponentiel

On suppose que les données non censurées \mathbf{x}_r sont des réalisations de la variable aléatoire $X \sim \mathcal{E}(\eta)$. On considère les censures de type I (fixe) et de type II (aléatoire). Les résultats suivants sont issus de

la connaissance du comportement en loi de la variable

$$2 \sum_{i=1}^r x_i/r \sim \chi_{(2r)}^2.$$

On trouvera par exemple dans Lannoy (1995, pages 83-84) des précisions et des exemples concrets sur la justification d'encadrements comparables lorsque seule une mesure du temps écoulé T pendant l'expérience est disponible (soit lorsqu'aucune valeur de défaillance passée n'est connue).

Censure de type I. On pose $T = \sum_{i=1}^n y_i$. On note $\chi_{(2r),\alpha}^2$ le quantile d'ordre α de la loi $\chi_{(2r)}^2$ (soit $P(T \sim \chi_{(2r)}^2 \leq \chi_{(2r),\alpha}^2) = \alpha$). Selon Sundberg (2001), l'encadrement non asymptotique le plus efficace de la valeur du paramètre η est formulé dans la proposition de Cox (1953) : on admet comme approximation $2r(\hat{\eta}/\eta) \rightsquigarrow \chi_{(2r+1)}^2$. On obtient l'intervalle de confiance $1 - \alpha$ pour η

$$\left[\frac{2T}{\chi_{(2r+1),1-\alpha/2}^2}, \frac{2T}{\chi_{(2r+1),\alpha/2}^2} \right]. \quad (2.1)$$

Censure de type II. Ici $T = \sum_{i=1}^r x_i + (n-r)x_r$. Sous une hypothèse d'indépendance des y_i , un intervalle de confiance $1 - \alpha$ pour η , non asymptotique, également considéré comme le meilleur en général dans l'étude de Sundberg (en se basant sur le caractère éventuellement censuré et en faible nombre des données), est

$$\left[\frac{2T}{\chi_{(2r),1-\alpha/2}^2}, \frac{2T}{\chi_{(2r),\alpha/2}^2} \right]. \quad (2.2)$$

2.3.1.2 Modèle de Weibull

On suppose que les données non censurées \mathbf{x}_r sont des réalisations de la variable aléatoire $\mathcal{W}(\eta, \beta)$. On dispose seulement de résultats non asymptotiques exacts sur des données censurées de type II (et de type II progressif, non traité ici).

Censure de type II. [Wu, 2002]. On suppose que $r > 2$. Une aire de confiance jointe $1 - \alpha$ (non asymptotique) pour les paramètres (η, β) est déterminée par les inégalités suivantes :

$$\begin{aligned} \varphi \left(x_1, \dots, x_r, F_{((1+\sqrt{1-\alpha})/2)(2r-2,2)} \right) < \beta < \varphi \left(x_1, \dots, x_r, F_{((1-\sqrt{1-\alpha})/2)(2r-2,2)} \right), \\ \left(\frac{2 \sum_{i=1}^n y_i^\beta}{\chi_{(2r), (1-\sqrt{1-\alpha})/2}^2} \right)^{1/\beta} < \eta < \left(\frac{2 \sum_{i=1}^n y_i^\beta}{\chi_{(2r), (1+\sqrt{1-\alpha})/2}^2} \right)^{1/\beta} \end{aligned}$$

où $F_{\alpha(\delta_1, \delta_2)}$ est le quantile d'ordre α de la loi de Fisher-Snedecor de degrés de libertés δ_1 et δ_2 (i.e. $P(X \rightsquigarrow F_{(\delta_1, \delta_2)} \geq F_{\alpha(\delta_1, \delta_2)}) = \alpha$), et $\varphi(x_1, \dots, x_r, t)$ est la solution unique en β de l'équation

$$\frac{\sum_{i=1}^r x_i^\beta - nx_1^\beta}{n(r-1)x_1^\beta} = t \quad \text{ou encore} \quad \sum_{i=1}^r \left(\frac{x_i}{x_1} \right)^\beta = n(1 + t(r-1)).$$

Cette solution est facilement estimable à l'aide d'une méthode de Newton-Raphson. On déduit de ce résultat des intervalles de confiance $1 - \alpha$ pour chacun des paramètres. Dans la pratique, la mise en oeuvre de la méthode reste cependant très coûteuse.

Bibliographie. Une célèbre méthode d'approximation développée empiriquement permet de construire des intervalles de confiance pour β sachant $\hat{\beta}_{ML}$ (toujours pour le type II) ; on suppose qu'il existe h et g deux fonctions des caractéristiques n et r de l'échantillon telles que

$$g \frac{\beta}{\hat{\beta}_{ML}} \sim \chi_h^2.$$

Lawless (1982, pages 157-159) fournit des tabulations de g et h . Par ailleurs, l'ouvrage de Lawless propose une vision large des problèmes inhérents à la construction d'intervalles de confiance pour des modèles de durée de vie et constitue une référence importante de ce domaine. Enfin, d'autres méthodes, très souvent utilisées dans l'industrie (par exemple dans Perroud & Moureau, 2004), fournissent des intervalles de confiance bilatéraux sur les deux paramètres en nécessitant le calcul de l'information de Fisher en $\hat{\theta}_{ML}$. L'utilisation de la méthode d'Ip (voir § 2.3.2.5) est alors conseillée. Présentant une sensibilité élevée à la présence de censure, ces méthodes restent généralement restreintes à des échantillons de grande taille.

2.3.2 Procédures de bootstrap pour l'estimation d'une incertitude

2.3.2.1 Présentation

Les procédures de *bootstrap*, introduites par Efron (1982) et depuis couramment utilisées (Tingley & Field 1990), permettent de pallier l'absence ou de proposer une alternative à la construction théorique d'intervalles de confiance. Comme indiqué précédemment, la validité de ces derniers peut être remise en question, au sens où les données de REX utilisées respectent peu les conditions d'indépendance, de distributivité et de type de censure obligatoirement supposées au § 2.3.1. Contrairement à cette construction théorique, ces procédures sont liées à un estimateur $\hat{\theta} = \hat{\theta}(\mathbf{Y}_n)$, dont elles estiment la variance $\hat{\sigma}^2$.

Une fois l'observé $\hat{\theta}(\mathbf{y}_n)$ obtenu par maximisation de la vraisemblance², le schéma général du bootstrap consiste à simuler un grand nombre M d'échantillons de caractéristiques semblables à \mathbf{y}_n , qui fournissent à leur tour M estimations $\hat{\theta}_i$ du maximum de vraisemblance. L'estimateur empirique

$$\hat{\sigma}_M^2 = \frac{1}{M-1} \sum_{i=1}^M \left(\hat{\theta}_i - \frac{1}{M} \sum_{i=1}^M \hat{\theta}_i \right)^2$$

offre ainsi une estimation de la variance de l'estimateur $\hat{\theta}(\mathbf{Y}_n)$. Les justifications de ce schéma (en particulier la consistance asymptotique) sont précisées par exemple dans Van der Waart (1998, pp. 326-331). Dans les prochains paragraphes, nous résumons et formulons quelques propositions sur la reconstitution de ces échantillons. Par ailleurs, une méthode alternative consiste à estimer la matrice de covariance de l'estimateur $\hat{\theta}(\mathbf{Y}_n)$ en inversant une estimation de la matrice d'information de Fisher. Celle-ci est présentée au § 2.3.2.5.

Une fois produit l'estimateur $\hat{\sigma}_M$, plusieurs méthodes permettent de construire des intervalles de confiance pour θ (et pour toute fonction de θ). Nous résumons ici un éventail de méthodes présentées et discutées dans Efron & Tibshirani (1998, pages 158-176).

²Notons que cette maximisation de la vraisemblance peut là encore être approximée par plusieurs méthodes (EM, SEM, etc.). L'estimateur $\hat{\theta}(\mathbf{Y}_n)$ étant bien entendu indépendant de cette méthode, nous proposons d'utiliser les estimations bootstrap produites par l'ensemble des méthodes disponibles afin de ne pas sous-estimer cette incertitude. Entre plusieurs estimations, on choisira celle qui maximise la vraisemblance observée de l'échantillon bootstrap.

Loi des grands nombres. Si $\hat{\sigma}_M$ est produit par l'utilisation d'un grand nombre d'échantillons bootstrappés, on utilise en général l'approximation $(\theta - \hat{\theta})/\hat{\sigma}_M \sim \mathcal{N}(0, I_p)$ où I_p est l'identité dans \mathbb{R}^p . Un intervalle de confiance pour θ , bilatéral et de seuil α , est alors

$$\left[\hat{\theta}(\mathbf{y}_n) - u_{\alpha/2} \hat{\sigma}_M, \hat{\theta}(\mathbf{y}_n) + u_{\alpha/2} \hat{\sigma}_M \right] \quad (2.3)$$

où u_α est le quantile d'ordre α de la loi $\mathcal{N}(0, 1)$, soit $P(X \sim \mathcal{N}(0, 1) \leq u_{\alpha/2}) = \alpha/2$.

“t-studentisation”. L'approximation (2.3) n'est cependant valide que lorsque n est grand (typiquement, $n \geq 30$). Lorsque n est petit, l'approximation de Student fournit de meilleurs résultats : $(\theta - \hat{\theta})/\hat{\sigma}_M \sim St_{n-1}$ où St_{n-1} représente la loi de Student avec $n - 1$ degrés de liberté, élargie à la dimension p . En notant $t_{n-1}^{(\alpha)}$ le quantile d'ordre α de St_{n-1} , on obtient un intervalle de seuil α

$$\left[\hat{\theta}(\mathbf{y}_n) - t_{n-1}^{(\alpha/2)} \hat{\sigma}_M, \hat{\theta}(\mathbf{y}_n) - t_{n-1}^{(1-\alpha/2)} \hat{\sigma}_M \right]. \quad (2.4)$$

“t-bootstrapping”. Afin d'éviter toute approximation sur le comportement de la variable aléatoire $(\theta - \hat{\theta})/\hat{\sigma}_M$, une méthode simple est de générer pour chaque échantillon bootstrappé d'estimation $\hat{\theta}_i$ la valeur $z_i = (\hat{\theta}_i - \hat{\theta})/\hat{\sigma}^i$ où $\hat{\sigma}^i$ est l'écart-type bootstrap de $\hat{\theta}_i$ (cette méthode nécessite donc un double bootstrap). On remplace alors $t_{n-1}^{(\alpha)}$ dans (2.4) par le quantile empirique d'ordre α de la distribution des (z_1, \dots, z_M) . Cette méthode a cependant le défaut de fournir des résultats parfois erratiques et très influencés par des groupes de données atypiques, ce qui constitue un handicap sérieux au traitement de données industrielles en faible nombre.

Fractiles empiriques des estimations bootstrap. Enfin, les fractiles ou quantiles empiriques de la distribution des $(\hat{\theta}_1, \dots, \hat{\theta}_M)$ permettent de fournir des zones de confiance comparables aux intervalles précédemment définis. D'après les tests numériques d'Efron & Tibshirani (1998, chapitre 13), le recouvrement de l'intervalle est généralement bon à condition que le modèle paramétrique soit correct. Remarquons que dans la stricte acceptation du terme “intervalle de confiance”, cette définition semble la plus valide : sur M estimations possibles, $100(1 - \alpha)\%$ sont situées entre les quantiles empiriques de seuil $\alpha/2$ et $1 - \alpha/2$.

Cependant, il est très clair que cette définition doit être modulée : tout d'abord, la méthode d'estimation produit un *biais* estimé par

$$b_M(\hat{\theta}) = \frac{1}{M} \sum_{i=1}^M \hat{\theta}_i - \hat{\theta}(\mathbf{y}_n) \quad (2.5)$$

qui implique que ces fractiles empiriques doivent être débiaisés vis-à-vis de l'estimation initiale $\hat{\theta}(\mathbf{y}_n)$, en leur ôtant la quantité $b_M(\hat{\theta})$. Ce biais sera graphiquement mis en évidence dans les exemples clôturant ce chapitre. Enfin, cette méthode a tendance à sous-estimer l'incertitude réelle sur les quantités estimées, puisqu'elle ne remet pas en cause le modèle. Cette méthode est donc extrêmement dépendante de la justesse du modèle paramétrique ainsi que du *plan d'échantillonnage** du bootstrap.

2.3.2.2 Types de bootstrap

Deux types de bootstrap peuvent être considérés. Le bootstrap **non paramétrique** simule de nouveaux échantillons en piochant aléatoirement des données dans \mathbf{y}_n avec remise. La justification de

cette technique repose sur la convergence asymptotique de l'estimateur empirique de la fonction de répartition. L'intérêt de cette procédure est d'éviter la spécification éventuelle d'un processus de censure généralement inconnu. C'est aussi un défaut dans la mesure où ce processus de censure, dans les études industrielles, est le fruit d'une stratégie réfléchie qui n'est pas forcément indépendante du composant \sum dont on mesure la durée de vie. Selon Meeker & Escobar (1998, pages 206-208), cette technique n'est valide que pour des échantillons non censurés de taille en moyenne supérieure à 10 (au sens où elle donne des résultats proches de l'approche paramétrique dans des conditions de simulation idéales).

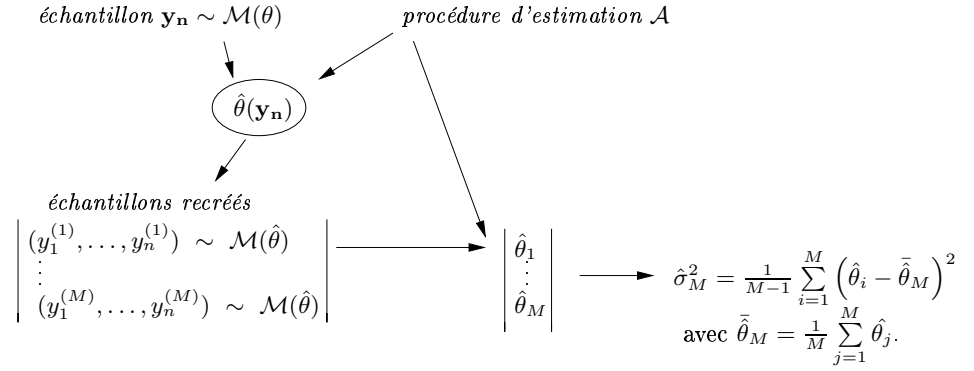
Porter *et al.* (1997) ont recensé les principales propriétés et défauts des intervalles de confiance de type "t-bootstrapping" construits de façon non paramétrique ; ils sont fortement sensibles à la taille et les éventuelles dissymétries de la distribution empirique des données. Lei & Smith (2003), plus récemment, ont présenté une étude empirique des principales procédures de bootstrap non paramétrique ainsi qu'une revue des principales publications sur le sujet. En proposant plusieurs méthodes d'amélioration de ces procédures, ils témoignent cependant de leur faible rentabilité sur de petits échantillons, pour lesquels les intervalles de confiance produits restent fragiles et parfois fortement erronés vis-à-vis des intervalles théoriques disponibles.

EXEMPLE 1. Soit l'échantillon exponentiel censuré de type I progressif $\mathbf{y}_n^0 = (\mathbf{x}_r^0, \mathbf{c}_{n-r}^0)$ avec $n = 9$, $r = 6$, $\mathbf{x}_r^0 = (63, 109, 140, 21, 58, 67) \sim \mathcal{E}(\eta_o = 100)$ et $\mathbf{c}_{n-r} = (50, 100, 150)$ les valeurs de censure. Le maximum de vraisemblance est $\hat{\eta} = 126.33$. Le bootstrap non paramétrique appliqué à l'échantillon \mathbf{y}_n^0 consiste à simuler M nouveaux échantillons $\{\mathbf{y}_n^i = (\mathbf{x}_r^i, \mathbf{c}_{n-r}^i)\}_{i \in \{1, \dots, M\}}$ où $\{x_1^i, \dots, x_r^i\} \subset \{63, 109, 140, 21, 58, 67\}$ avec répétitions et $\{c_1^i, \dots, c_{n-r}^i\}$ peut être alternativement tiré de $\{50, 100, 150\}$ avec répétition ou alors choisi de manière conservative ($\mathbf{c}_{n-r}^i = \mathbf{c}_{n-r}^0 \forall i \in \{1, \dots, M\}$). ■

Récemment, plusieurs travaux se sont attachés à développer des approches non paramétriques pour des problèmes d'estimation de petits échantillons. Ainsi, Porter *et al.* (1997)

Le bootstrap **paramétrique** (Fig.2.2), préférable dans le cadre de petits échantillons *i.i.d.*, recrée intégralement des échantillons selon le modèle $\mathcal{M}(\hat{\theta})$. La difficulté est alors de reproduire le processus de censure sous-jacent. Les cas de censure aléatoire (type II) et de valeur unique fixe (type I) ne présentent pas de difficultés, mais des données de censure échelonnées dans \mathbf{y}_n (type I progressif) sont problématiques.

Traditionnellement, on se contente d'inclure intégralement les censures dans l'échantillon bootstrappé, ou d'en inclure la moyenne, la médiane ou le maximum (Castanier 1997). Les échantillons ainsi bootstrappés ne correspondent pas aux réalisations d'une variable aléatoire mise en compétition avec un phénomène de censure. Une méthode tentant de reproduire les conditions d'obtention des données est donc proposée, qui sera comparée par la suite à ces méthodes traditionnelles.


 FIG. 2.2 – Schéma du bootstrap paramétrique pour l'estimation de $\text{Var}(\hat{\theta}(\mathbf{y}_n))$.

EXEMPLE 2. (suite de l'ex.1) *Le bootstrap paramétrique appliqué à l'échantillon \mathbf{y}_n^0 consiste à simuler M nouveaux échantillons $\{\mathbf{y}_n^i = (\mathbf{x}_r^i, \mathbf{c}_{n-r}^i)\}_{i \in \{1, \dots, M\}}$ où $\mathbf{x}_r^i \sim \mathcal{E}(\hat{\eta} = 126.33)$ et \mathbf{c}_{n-r}^i provient d'une spécification du phénomène de censure. Si l'échantillon était de type II, on choisirait $\mathbf{c}_{n-r}^i = (x_r^i, \dots, x_r^i)$ ($n-r$ fois la dernière valeur de défaillance). Dans ce cas particulier de type I progressif, sans autre connaissance de l'obtention des valeurs de censure, on propose par exemple $c_j^i = \sum_{k=1}^{n-r} c_k^0 / (n-r) \forall j \in \{1, \dots, n-r\}, \forall i \in \{1, \dots, M\}$ (censure bootstrap constante égale à la moyenne des censures initiales). ■*

2.3.2.3 Une reproduction du processus de censure (bootstrap paramétrique)

On suppose que plusieurs valeurs de censures différentes au sein d'un échantillon peuvent être considérées comme des réalisations d'une variable aléatoire C , dont on va estimer empiriquement la répartition conditionnellement à la taille des données disponibles. Par un procédé d'inversion de cette répartition, on simulera un échantillon de censure de taille n que l'on mettra en concurrence avec un échantillon de taille n simulé d'après $\hat{\theta}$. Les conditions de l'expérience aboutissant aux données \mathbf{x}_n seront alors supposées recréées. Notons qu'une démarche similaire de simulation de la censure selon une loi statistique sous-jacente, présentée dans Doss & Chiang (1994), permet la consistance des estimateurs bootstrap.

Estimateur de Kaplan-Meier (EKM) pour la répartition des censures. On note $c_1^* < \dots < c_q^*$ ($q \leq n-r$) la suite ordonnée par ordre croissant des q valeurs différentes de censure. On note f_1, \dots, f_q les fréquences relatives de ces valeurs. Elles vérifient $\sum_{j=1}^q f_j = (n-r)/n$. On pose également c_{min} et c_{max} deux valeurs de censures définies par $c_{min} \leq c_1^* < \dots < c_q^* \leq c_{max}$. L'estimateur de Kaplan-Meier de la fonction de répartition \mathcal{F}_C de la distribution de C se définit par

$$\begin{cases} P(C < c_{min}) &= F_0 = 0, \\ P(C < c_1^*) &= F_1 = f_1, \\ P(C < c_2^*) &= F_2 = f_1 + f_2, \\ &\vdots \\ P(C < c_q^*) &= F_q = r/n \\ P(C < c_{max}) &= F_{q+1} = 1. \end{cases}$$

Si les propriétés de consistance de l'EKM pour de grands échantillons sont bien connues (une vue générale en est présentée par Peterson 1977), il est nécessaire de supposer, dans ce cas de petite

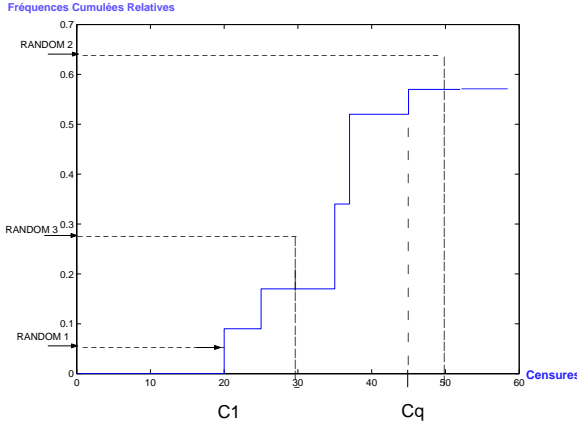


FIG. 2.3 – Fonction de répartition optimiste de la censure.

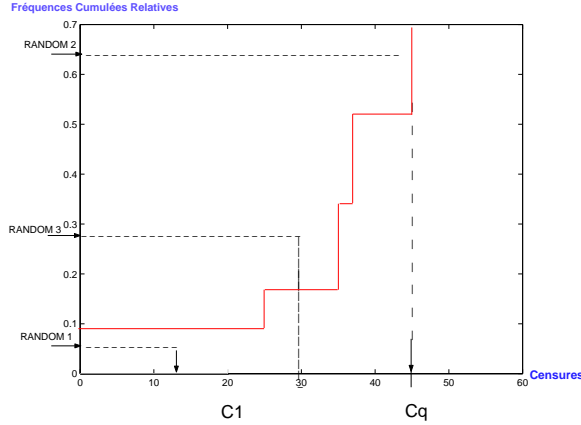


FIG. 2.4 – Fonction de répartition pessimiste de la censure.

taille, que les valeurs de censure observées sont véritablement des réalisations *i.i.d.* d'une variable aléatoire pour obtenir un biais faible et la consistance des simulations par inversion de \mathcal{F}_C (voir Chen *et al.* 1982). Le processus de censure n'a pas de raison particulière d'être aléatoire, mais on le relie à un processus aléatoire inconnu susceptible de fournir les réalisations (c_1, \dots, c_{n-r}) . Un échantillon $\tilde{\mathbf{c}}_n = (\tilde{c}_1, \dots, \tilde{c}_n)$ de censures peut alors être reconstruit de deux manières différentes.

1 - censure optimiste. On choisit $c_{min} = c_1$ et $c_{max} > c_q$ (cas extrême $c_{max} \rightarrow \infty$). La fonction de répartition empirique prend la forme de la figure 2.3. Ce choix est considéré comme optimiste car il laisse la possibilité à de grandes valeurs de l'échantillon bootstrap d'être non censurées. Cette manière d'inverser la répartition des censures a tendance à sous-évaluer l'incertitude sur la valeur finale des paramètres. Un choix de c_{max} raisonnablement grand est par exemple $c_{max} = 2c_q$, qui suffit à produire rapidement des échantillons bootstrap dont le taux de censure diffère de moins de 10% du taux de censure de l'échantillon initial.

2 - censure pessimiste. On choisit $c_{min} = 0$ et $c_{max} = c_q$. La fonction de répartition empirique prend la forme de la figure 2.4. Ce choix est *a contrario* considéré comme pessimiste car il interdit de grands temps de défaillance non censurés. On peut ainsi le considérer comme le plus raisonnable des deux.

Les fréquences présentées sur les figures 2.3 et 2.4 sont égales aux rapports cumulés du nombre de censures égales sur la taille de l'échantillon. Leur somme ne vaut donc pas 1, puisqu'il faut prendre aussi en compte les valeurs non censurées de l'échantillon initial. Lors de l'inversion de cette fonction de répartition, on peut tirer des aléas supérieurs à la fréquence maximale cumulée des censures (RANDOM 2), ainsi que des aléas inférieurs à la fréquence minimale cumulée (RANDOM 1). La méthode d'inversion est présentée ci-dessous.

Pour $i = 1, \dots, n$

1. On simule $U_i \sim \mathcal{U}[0, 1]$.
2. Soit j tel que $U_i \in [F_j, F_{j+1}]$.

- Si $0 < j < q$, on simule $\tilde{c}_i \sim \mathcal{U}[c_j, c_{j+1}]$.
- Si $j = q$, on simule $\tilde{c}_i \sim \mathcal{U}[c_q, c_{max}]$.
- Si $j = 0$, on simule $\tilde{c}_i \sim \mathcal{U}[c_{min}, c_1]$.

Après avoir simulé un échantillon de censure $\tilde{\mathbf{c}}_n = (\tilde{c}_1, \dots, \tilde{c}_n)$, puis un échantillon non censuré, $\tilde{\mathbf{x}}_n \sim \mathcal{M}(\hat{\theta})$, on construit l'échantillon bootstrappé $\{\tilde{y}_i\} = \{\min(\tilde{x}_i, \tilde{c}_i)\}$, $i \in \{1, \dots, n\}$. Ce faisant, la censure et les défaillances simulées sont mises en concurrence. Enfin, on fait le choix de ne conserver que les échantillons censurés dont le taux de censure est très peu éloigné du taux de censure de l'échantillon réel (moins de 10% de différence) pour améliorer l'évaluation de l'algorithme sur le type de données à utiliser. Cette technique peut amener à rejeter certaines configurations de censure, et il est donc nécessaire de simuler de nombreux échantillons selon $\mathcal{M}(\hat{\theta})$ pour atteindre le nombre d'échantillons bootstrappés conseillé par de nombreux auteurs (entre 2000 et 5000 en général selon Meeker & Escobar 1998).

Définir le processus de censure comme aléatoire est risqué, car la petite taille de l'échantillon n'offre qu'une estimation très imprécise de sa fonction de répartition. En particulier, la variance de ce processus est peu visible. Un résultat est cependant disponible pour le modèle exponentiel, placé en annexe de ce chapitre.

2.3.2.4 Conditions et difficultés d'utilisation du bootstrap paramétrique

L'inversion de la censure et la sélection d'échantillons de taux de censure constant ou peu variable est un facteur très limitant de l'algorithme de bootstrap, dont le coût est déjà lourd. Le bootstrap paramétrique censuré n'est ainsi utilisable que pour des méthodes automatiques rapides, telles que EM ou SEM.

Dans le cas d'une procédure optimiste, on propose de choisir un écart maximal de pourcentage de censure entre l'échantillon initial et les échantillons de bootstrap de l'ordre de 10%. Pour cela, on a constaté que $4\,000 \leq M \leq 10\,000$ suffit en général à conserver plus de 3 000 échantillons correctement bootstrappés. Dans le cas pessimiste, on conseille $M \simeq 10\,000$ ce qui permet d'obtenir plus de 2 500 échantillons bootstrappés, de taux de censure de moins de 5% de différence avec le taux initial. Notons cependant que la stabilisation des écarts-types bootstrap peut être assurée sur certains exemples, lorsque la dimension p du modèle est petite, avec quelques centaines d'échantillons bootstrappés (par exemple pour les modèles exponentiel et de Weibull).

Après tests, les méthodes optimistes et pessimistes fournissent des estimations des écarts-types qui diffèrent peu, sauf dans le cas particulier d'échantillons de taille très faible ($n \leq 15$) fortement censurés, où la méthode pessimiste apporte des résultats plus satisfaisants. La censure étant plus contraignante, l'écart-type pessimiste sera en moyenne plus grand, ce qui est appréciable. En effet, les procédures de bootstrap paramétrique ne remettent pas en cause le modèle paramétrique utilisé et ont généralement tendance à sous-estimer l'incertitude réelle sur $\hat{\theta}$. Nous conseillons donc son utilisation.

2.3.2.5 Un cas particulier de bootstrap : la méthode d'Ip (*Monte Carlo Bootstrap*)

La structure des données précisée au § 1.4 permet d'envisager l'estimation par Monte-Carlo de la matrice d'information de Fisher $\mathcal{I}_{obs}(\hat{\theta}|\mathbf{y}_n)$ des données observées \mathbf{y}_n . L'inversion de cette matrice

fournit une estimation de la matrice de covariance de $\hat{\theta}$. L'estimation de $\mathcal{I}_{obs}(\hat{\theta}|\mathbf{y}_n)$, proposée par Ip (1994), repose sur l'utilisation de l'identité de Louis (1982).

$$\mathcal{I}_{obs}(\hat{\theta}|\mathbf{y}_n) = \mathcal{I}_c(\hat{\theta}|\mathbf{y}_n) - \mathcal{I}_m(\hat{\theta}|\mathbf{y}_n)$$

où $\mathcal{I}_m(\hat{\theta}|\mathbf{y}_n)$ est l'information de Fisher des données manquantes et $\mathcal{I}_c(\hat{\theta}|\mathbf{y}_n)$ est l'information de Fisher des données complètes. L'idée de Ip, précisée notamment par Castanier (1997), est de reconstituer M sous-échantillons de données manquantes $(\mathbf{z}_{\mathbf{n}-\mathbf{r}}^{(i)})_{i \in [1, \dots, M]} \sim \mathcal{M}(\hat{\theta})$ qui complètent les données non censurées \mathbf{x}_r (en cela, la méthode s'apparente à une méthode de bootstrap paramétrique conditionnel). Un estimateur naturel de $\mathcal{I}_{obs}(\hat{\theta}|\mathbf{y}_n)$ est

$$\hat{\mathcal{I}}_{obs,M}(\hat{\theta}|\mathbf{y}_n) = - \left[\frac{1}{M} \sum_{i=1}^M \frac{\partial^2 l_c(\theta; \mathbf{x}_r, \mathbf{z}_{\mathbf{n}-\mathbf{r}}^{(i)})}{\partial \theta \partial \theta'} + \frac{1}{M} \sum_{i=1}^M \frac{\partial l_c(\theta; \mathbf{x}_r, \mathbf{z}_{\mathbf{n}-\mathbf{r}}^{(i)})}{\partial \theta} \frac{\partial l_c(\theta; \mathbf{x}_r, \mathbf{z}_{\mathbf{n}-\mathbf{r}}^{(i)})}{\partial \theta'} \right]_{\theta=\hat{\theta}} \quad (2.6)$$

où $l_c(\theta; \mathbf{x}_n)$ est la log-vraisemblance complète d'un échantillon *i.i.d.* \mathbf{x}_n selon $\mathcal{M}(\theta)$.

Le gros avantage de cette méthode est d'être moins coûteuse que le bootstrap paramétrique censuré. Elle introduit cependant une difficulté qui est d'assurer des reconstitutions d'échantillon selon une estimation $\hat{\theta}$ en simulant au-dessus de la censure. Le rejet d'un tel tirage peut être particulièrement long. De plus, si l'estimation $\hat{\theta}$ du maximum de vraisemblance est mauvaise, $\hat{\theta}$ ne peut plus être solution des équations de vraisemblance, la simplification (2.6) n'est pas toujours possible et $\mathcal{I}_{obs}(\hat{\theta}|\mathbf{x}_n)$ doit être estimée par (Castanier, 1997)

$$\tilde{\mathcal{I}}_{obs,M}(\hat{\theta}|\mathbf{y}_n) = \hat{\mathcal{I}}_{obs}(\hat{\theta}|\mathbf{y}_n) + \frac{1}{M^2} \left\{ \left[\sum_{i=1}^M \frac{\partial l_c(\theta; \mathbf{x}_r, \mathbf{z}_{\mathbf{n}-\mathbf{r}}^{(i)})}{\partial \theta} \right] \left[\sum_{i=1}^M \frac{\partial l_c(\theta; \mathbf{x}_r, \mathbf{z}_{\mathbf{n}-\mathbf{r}}^{(i)})}{\partial \theta'} \right] \right\}_{\theta=\hat{\theta}}^T.$$

On peut ainsi s'attendre, pour des modèles dont les équations de vraisemblance n'ont pas de solutions explicites (comme Weibull et \mathcal{B}), cette caractéristique allant de pair avec l'accroissement de la dimension de θ , à rencontrer des difficultés dans l'utilisation de la méthode d'Ip, aboutissant à des estimations erronées.

2.4 Indicateurs décisionnels de l'analyste industriel

Au su d'une estimation $\hat{h}_{\mathbf{y}_n} = h(\hat{\theta}(\mathbf{y}_n))$ d'une fonction d'intérêt $h(\theta)$ et d'une mesure de son incertitude, l'analyste industriel doit juger si la méthode utilisée est pertinente. En particulier, quelques indications importantes sont fournies par les grandeurs suivantes :

- *le biais de l'estimateur ponctuel*³ $\hat{h}_{\mathbf{Y}_n} = h(\hat{\theta}(\mathbf{Y}_n))$; il peut être estimé par le biais du bootstrap en utilisant la formule (2.5) appliquée à la fonction $h(\theta)$, en sus des corrections fournies en références au § 2.2.2.4. Si ce biais est important vis-à-vis de la valeur de l'estimé $\hat{h}_{\mathbf{y}_n}$, il indique que le maximum de vraisemblance, en tant qu'estimateur ponctuel, peine à modéliser correctement le comportement statistique des données.

³Notons qu'une abondante littérature scientifique s'est fait l'écho des difficultés d'estimation du biais, en particulier Dempster, Laird & Rubin (1977), Meng & Rubin (1991), Jennrich & Jamshidian (1996) et McLachlan & Krishnan (1997). La détection du biais est le sujet de nombreux travaux commandés par les organismes fixant les règles nationales et internationales de sûreté (voir par exemple Rodionov 2005).

Cette estimation du biais varie donc selon le type de bootstrap utilisé. Les applications numériques montreront que le bootstrap censuré proposé auparavant (avec simulation intégrale de la censure) offre une estimation du biais généralement supérieure (en valeur absolue) au bootstrap de type I, classiquement utilisé. Si l'on peut spécifier le processus de censure, il est donc intéressant d'utiliser le bootstrap censuré pour détecter la surestimation ou la sous-estimation systématique d'une méthode vis-à-vis d'un type d'échantillon donné. Le bootstrap de type I apparaît comme un choix conservatif pour l'estimation des paramètres de forme dans une étude fiabiliste; en effet, ceux-ci sont en général surestimés, ce qui implique une sous-estimation de la durée de vie d'un composant. D'un point de vue fiabiliste, un tel résultat n'est pas forcément réhibitoire.

- *la proximité des intervalles de confiance autour de l'estimateur*; une forte divergence (en particulier avec les intervalles théoriques, s'ils existent) indique la faible reproductibilité de ce plan d'échantillonnage.
- *le facteur d'erreur* de la distribution d'échantillonnage autour de l'estimé $h(\hat{\theta})_{y_n}$. Le *facteur d'erreur* ρ est un outil très simple, couramment utilisé par les industriels pour caractériser la précision de cette méthode d'estimation. Bien que plusieurs définitions existent (Lannoy 1995, pages 85-87, Bonneville & Billy 2006), la réglementation d'EDF des études probabilistes de sûreté (EPS) préconise d'utiliser la formule *facteur d'erreur* sur $h(\theta)$ comme

$$\rho(h) = \sqrt{\max\left(\frac{q_{50\%}}{q_{5\%}}, \frac{q_{95\%}}{q_{50\%}}\right)} \quad (2.7)$$

où q_α est le quantile d'ordre α de la distribution d'échantillonnage⁴. Tacitement, les ingénieurs R&D d'EDF considèrent qu'une estimation de $h(\theta)$ est trop imprécise si $\rho(h) \geq 3$.

Cependant, ce facteur est fondé originellement sur la manipulation de distributions symétriques (et idéalement gaussiennes). Si nous tentons de l'appliquer à n'importe quel type de distribution quelque peu dissymétrique - ce qui est parfaitement susceptible de se produire lorsque la fonction d'intérêt $h(\theta)$ est non linéaire, contractante ou a des propriétés de dilatation - cette règle empirique paraît fragile. Après des discussions avec les ingénieurs d'EDF, nous avons retenu comme règle d'admettre que la méthode d'estimation est acceptable si $\rho(h) \geq 2.5$. Ainsi, ce facteur permet de ne considérer que des distributions raisonnablement équilibrées, aux queues légères. Il agit comme un détecteur de *concentration* de l'estimation.

Dans la suite de ce travail, le calcul de ce facteur d'erreur est réservé aux deux premiers moments de la variable aléatoire X représentant la durée de vie. Ils constituent donc un score permettant de juger si les répercussions de l'erreur d'estimation de θ sont acceptables ou non pour l'estimation de $E[X]$ et de $\text{Var}[X]$.

⁴Cette définition est donc applicable dans un contexte autant fréquentiste que bayésien.

2.5 Applications numériques

Ces applications numériques ont pour but d'illustrer, à travers le traitement d'échantillons de durées de vie typiques des études EDF, la méthodologie présentée tout au long de ce chapitre, et de montrer l'insuffisance des méthodes fréquentistes d'estimation. Pour chaque échantillon, nous comparons plusieurs méthodes d'estimation des paramètres, des intervalles de confiance et d'estimation du biais. Méthodologiquement parlant, les informations qui en émanent sont des outils d'aide à la décision, permettant de valider ou de rejeter l'emploi de ces méthodes d'estimation dans un cadre industriel.

2.5.1 Présentation des données et estimations des paramètres

Les données utilisées dans toute cette partie sont issues de simulations présentées dans Billy *et al.* (2004). Exemples typiques d'échantillons industriels, fortement censurés et de faible taille, ils vont permettre d'appliquer les méthodes de maximisation de la vraisemblance pour l'estimation des paramètres et de leur incertitude, et d'en déduire un verdict sur la validité de leur emploi. Ils illustrent une démarche générale de test sur un grand nombre d'échantillons similaires.

On présente également l'estimation du maximum de vraisemblance (notée EMV) des paramètres. L'EMV n'est explicite que pour le modèle exponentiel ; autrement elle est calculée par NR (Newton-Raphson) ou par SEM. L'estimation qui maximise la *vraisemblance observée* est mise en gras. On ne fournira pas ici des résultats d'estimation sur un ou plusieurs jeux du modèle de compétition $\mathcal{B}(\eta_0, \eta_1, \beta)$. Une étude comparative présentée dans Bertholon *et al.* (2006) a montré la forte sensibilité de ce modèle à la censure et aux faibles tailles d'échantillon.

Le tableau 2.3 présente tout d'abord les jeux *a1bis* et *b1bis*, censurés à droite de taille 10. Ces données sont simulées selon une loi exponentielle de paramètre $\eta_0 = 8760$ (heures, soit 1 an de fonctionnement). Leurs EMV, présentées dans le tableau 2.4, très éloignées malgré leur différence minimale en pourcentage de censure, illustrent à quel point une estimation fréquentiste à partir d'échantillon de faible taille peut y être sensible.

Le tableau 2.3 présente aussi les jeux *A215* et *B215*, censurés à droite de taille 10 également. Ils sont issus d'une loi de Weibull de paramètres $\eta_0 = 8760$ heures (1 an) et $\beta_0 = 2.15$. Leurs EMV sont également présentés dans le tableau 2.4. Si l'EMV de η reste stable, celle de β varie de façon importante. Remarquons que l'estimation SEM, si elle peut sembler meilleure pour le jeu *B2125*, ne peut pas être retenue car la vraisemblance observée qu'elle induit n'est pas la plus haute.

modèle	jeu	% censure	défaillances réelles	valeurs de censures
<i>exponentiel</i>	<i>a1bis</i>	70%	671, 802, 1517	4380, 4380, 4818, 5256, 5256, 6132, 6132
	<i>b1bis</i>	60%	143, 230, 1323, 3856	4380, 4818, 4818, 5256, 5694, 6132
<i>Weibull</i>	<i>A215</i>	70%	3388, 2881, 2652	6132, 6132, 5256, 5256, 4818, 4380, 4380
	<i>B215</i>	60%	4380, 1791, 1611, 1291	6132, 5694, 5296, 4818, 4818, 4380

TAB. 2.3 – Jeux de données exponentiel et de Weibull.

modèle	jeu	EMV (explicite/calculé par NR)	EMV (calculé par SEM)
<i>exponentiel</i>	<i>a1bis</i>	$\hat{\eta}_{ML} = \mathbf{13\ 114}$	
	<i>b1bis</i>	$\hat{\eta}_{ML} = \mathbf{9\ 162}$	
<i>Weibull</i>	<i>A215</i>	$\hat{\eta}_{NR} = \mathbf{8\ 687}$, $\hat{\beta}_{NR} = \mathbf{1.94}$	$\hat{\eta}_{SEM} = 7\ 800$, $\hat{\beta}_{SEM} = 2.60$
	<i>B215</i>	$\hat{\eta}_{NR} = \mathbf{8\ 752}$, $\hat{\beta}_{NR} = \mathbf{1.21}$	$\hat{\eta}_{SEM} = 8\ 172$, $\hat{\beta}_{SEM} = 1.44$

TAB. 2.4 – EMV des jeux de données (NR=Newton-Raphson). En gras est placée l’estimation maximisant la vraisemblance observée.

Il faut cependant remarquer que les temps de censures proposés, dans tous les échantillons, sont situés bien au-delà des plus hautes valeurs de temps de défaillance ; ainsi, le vieillissement du système menace d’être occulté (l’estimation de β tend donc naturellement vers 1 pour le modèle de Weibull). La stratégie de simulation de ces échantillons de REX respecte peu la démarche qualitative réelle que nous avons présenté au § 1.2, qui consiste à “élaguer” les données de REX en cherchant à percevoir la pertinence de chacune d’entre elles. Le but est ici de tester la sensibilité des méthodes d’estimation sur des données dont on ignore délibérément le contexte d’étude (fiabiliste ou durabiliste).

2.5.2 Application au modèle exponentiel

On décide tout d’abord de calculer les intervalles de confiance théoriques donnés par les expressions (2.1) et (2.2) de l’EMV des jeux *a1bis* et *b1bis* (tableau 2.3). Ce faisant, on considère donc la censure comme de type I ou de type II. Parallèlement, on propose l’utilisation d’un bootstrap traditionnel de type I (où la censure bootstrappée est la moyenne des censures de l’échantillon initial). De même, on présente les résultats de la méthode d’Ip. Enfin, on compare ces résultats à ceux que produit le bootstrap censuré (optimiste et pessimiste, voir § 2.3.2.3). Ces méthodes de bootstrap permettent d’obtenir l’écart-type $\hat{\sigma}_M$ sur $\hat{\eta}_{ML}$ puis les intervalles de confiance selon les procédures décrites au § 2.3.2.1.

Le tableau 2.6 présente les intervalles de confiance obtenus pour l’EMV $\hat{\eta}_{ML} = 13\ 114$ dans le cas du jeu *a1bis*. Ceux-ci sont présentés de façon plus lisible sur la figure 2.5. Sur cette figure, les intervalles de confiance construits d’après les fractiles bootstrap ne sont pas débiaisés. Parallèlement, le jeu *b1bis* est traité sur la figure 2.6. Le plan d’échantillonnage de *b1bis* diffère très peu de celui de *a1bis*. Seule une valeur de défaillance réelle remplace une censure, alors que la somme des temps reste semblable dans les deux cas. Une rapide comparaison entre les deux figures permet de noter la forte sensibilité des résultats à l’ajout d’une valeur de censure (ce qui justifie la précaution du choix des données).

Biais de l’estimation. L’EMV est biaisée vis-à-vis de la vraie valeur η_o , en la surestimant de façon importante dans le cas du jeu *a1bis* : on a $\hat{\eta}_{ML} - \eta_o = 4354$. Ce fort décalage illustre le peu de représentativité que possède l’EMV. Ce biais est effectivement détecté par la moyenne des estimations bootstrap (pour lesquelles on utilise 1000 échantillons), mais il est mieux estimé lorsqu’on utilise la méthode d’Ip et les estimations bootstrap où la censure est intégralement simulée (cf. tableau 2.5). Dans le cas du jeu *b1bis*, ce biais effectif est cependant surestimé, ce qui témoigne d’un mauvais comportement de la méthode d’estimation : pour un ensemble d’échantillons de caractéristiques semblables à *b1bis*, la surestimation constatée par la méthode est plus importante que celle constatée sur

b1bis. L'estimation $\hat{\eta}_{ML} = 9\,162$, proche de $\eta_o = 8\,760$, est en fait très peu reproductible sur ce type d'échantillon (en supposant que le plan d'échantillonnage est correct).

méthode	estimation du biais
bootstrap type I	2860
bootstrap pessimiste	3700
bootstrap optimiste	3575

TAB. 2.5 – Estimation du biais d'estimation du paramètre exponentiel η pour le jeu *a1bis*.

Validité des intervalles de confiance. Pour le jeu *a1bis*, seules trois méthodes offrent un intervalle de confiance [5% – 95%] contenant la véritable valeur $\eta_o = 8760$ du paramètre : les méthodes pivotales de Cox-Sundberg (§ 2.3.1.1) et la méthode d'Ip (§ 2.3.2.5). Les premières offrent des intervalles de confiance très grands mais qui semblent susceptibles d'être un peu améliorés, étant donné qu'ils ne prennent pas en compte la véritable distribution des valeurs de censure. La méthode d'Ip fournit ici un résultat plus réaliste. On peut noter, malgré leur biais à décaler, que les intervalles de confiance construits d'après les fractiles bootstrap sont abusivement restreints, ce qui indique que le plan d'échantillonnage retenu reste inadéquat (voire inaccessible).

	$\hat{\sigma}_M$	<i>approx. normale</i>	<i>approx. Student</i>	<i>quantiles boot.</i>	<i>théorique</i>
boot. type I	898	[11669 , 14559]	[11504 , 14725]	[13160 , 15962]	
boot. optimiste	862	[11696 , 14532]	[11534 , 14695]	[14911 , 16976]	
boot. pessimiste	864	[11692 , 14536]	[11530 , 14699]	[13987 , 17894]	
méthode d'Ip	2498	[9005 , 17223]	[8535 , 17694]		
type I (2.1)					[5593 , 36306]
type II (2.2)					[6249 , 48115]

TAB. 2.6 – Estimations de $\hat{\sigma}$ et des intervalles [5%-95%] autour de $\hat{\eta}_{ML} = 13114$ pour le jeu *a1bis* (M=2000). Les intervalles produits par fractiles bootstrap doivent être soustrait d'un biais proche de 3000.

	$\hat{\sigma}_M$	<i>approx. normale</i>	<i>approx. Student</i>	<i>quantiles boot.</i>	<i>théorique</i>
boot. type I	732	[7956 , 10368]	[7818 , 10506]	[8906 , 11325]	
boot. optimiste	735	[7954 , 10370]	[7816 , 10508]	[10986 , 13437]	
boot. pessimiste	742	[7941 , 10383]	[7802 , 10522]	[9942 , 12355]	
méthode d'Ip	1663	[6492 , 11832]	[6187 , 12137]		
type I (2.1)					[4332 , 22044]
type II(2.2)					[4726 , 26823]

TAB. 2.7 – Estimations de $\hat{\sigma}$ et des intervalles [5%-95%] autour de $\hat{\eta}_{ML} = 9162$ pour le jeu *b1bis* (M=2000). Les intervalles produits par fractiles bootstrap doivent être soustrait d'un biais estimé en moyenne à 2500.

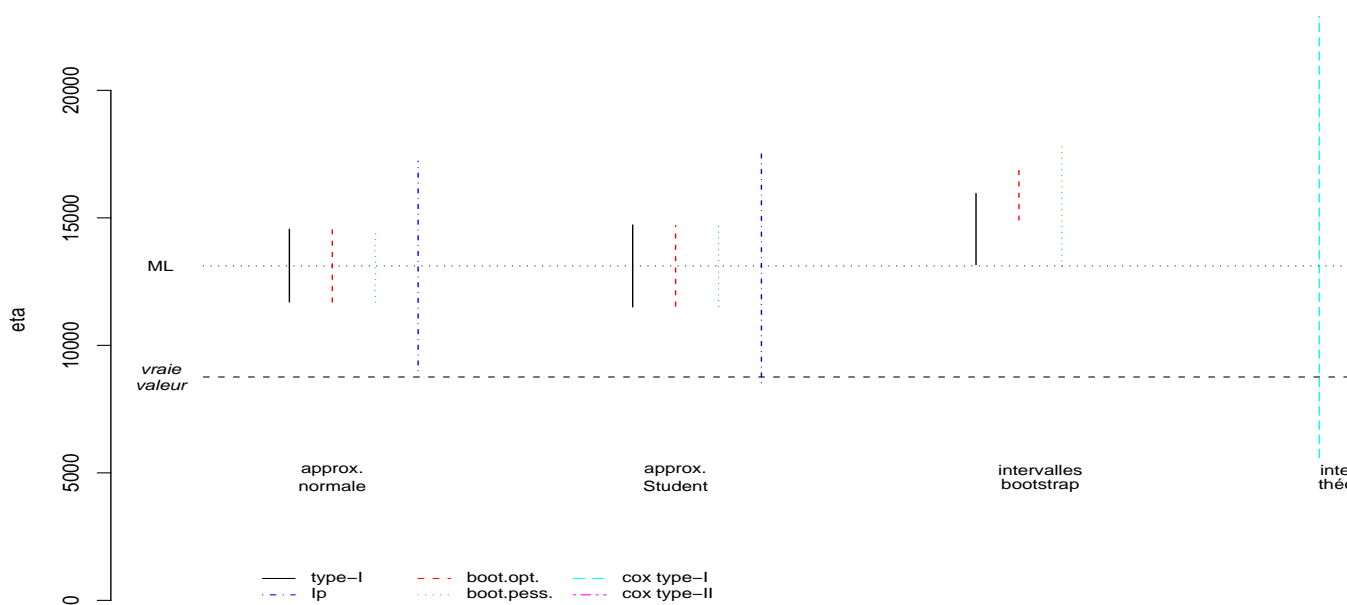


FIG. 2.5 – Comparaison des intervalles de confiance [5% – 95%] autour de $\hat{\eta}_{ML} = 13\ 114$ pour le jeu *albis*.

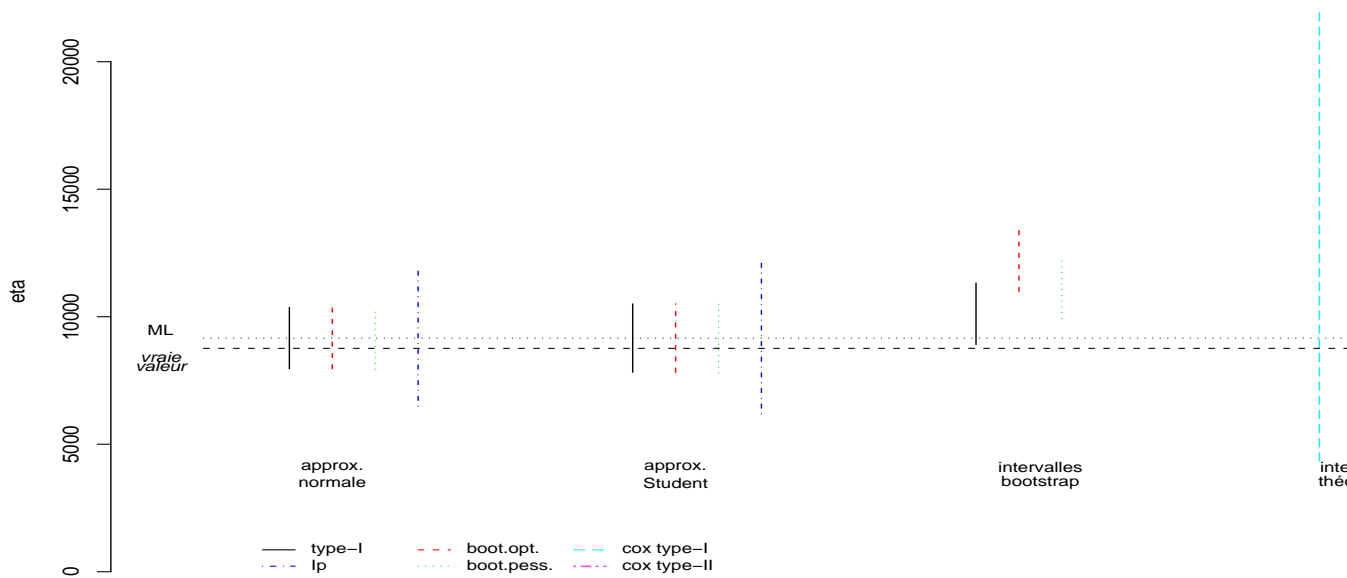


FIG. 2.6 – Comparaison des intervalles de confiance [5% – 95%] autour de $\hat{\eta}_{ML} = 9\ 162$ pour le jeu *bibis*.

2.5.3 Application au modèle de Weibull

Sur les estimations NR du maximum de vraisemblance des jeux de Weibull $A215$ et $B215$ (tableau 2.3), on procède aux mêmes comparaisons des écart-types et des intervalles de confiance sur (η, β) qu'au § 2.5.2. On ne présente que les résultats provenant du bootstrap de type I, de la méthode d'Ip et du bootstrap pessimiste. Les résultats sont placés dans le tableau 2.8.

Application au jeu $A215$. La figure 2.7 permet de visualiser plus facilement les intervalles de confiance du tableau 2.8, mesurant l'incertitude sur l'estimation $(\hat{\eta}_{ML}, \hat{\beta}_{ML}) = (8687, 1.94)$. Cependant, une façon plus réaliste de représenter la zone de confiance jointe sur $(\hat{\eta}, \hat{\beta})_{ML}$ et $(\hat{\eta}, \hat{\beta})_{SEM}$ est de représenter l'*enveloppe convexe* des estimations bootstrap de type I et pessimiste. Entourant 90% des estimations bootstrap, cette enveloppe est présentée sur la figure 2.8.

Les intervalles de confiance contiennent en général le couple de valeurs réelles $(\eta_o, \beta_o) = (8760, 2.15)$. En termes de taille d'intervalle, le bootstrap de type I apporte sur η une précision meilleure que le bootstrap pessimiste, et inversement sur β . D'une manière générale, ces intervalles sont plus grands que pour le modèle exponentiel. Leur élargissement est consécutif à l'augmentation de la dimension de θ sur des données de même taille et de même taux de censure que les données exponentielles précédentes⁵. Notons cependant que la méthode d'Ip fournit des intervalles de confiance significativement plus précis que les autres (ce qui est très visible sur la figure 2.7).

Application au jeu $B215$. Parallèlement au cas précédent, l'enveloppe convexe des estimations bootstrap (figure 2.9) permet d'apprécier l'incertitude de l'estimation. De nouveau, une faible diminution du pourcentage de censure permet cependant d'améliorer drastiquement la précision de l'estimation (notamment celle de β); on vérifie ainsi la présence de $(\eta_o, \beta_o) = (8760, 2.15)$ dans la plupart des intervalles de confiance. On constate que la méthode d'Ip sous-estime l'incertitude sur l'estimation de $\theta = (\eta, \beta)$; les intervalles de confiance ne contiennent pas la vraie valeur du paramètre. Une tendance naturelle de l'estimation fréquentiste du modèle de Weibull est cependant de surestimer un paramètre en sous-estimant l'autre (tendance que l'on peut remarquer en notant la forme des zones de confiance jointe des figures 2.8 et 2.9.

	$\hat{\sigma}_M$		<i>approx. Student</i>		<i>quantiles boot.</i>	
	η	β	η	β	η	β
boot. type I	3096	3.01	[3010 , 14364]	[0.00 , 7.48]	[5972 , 14278]	[1.00 , 7.25]
boot. pessim.	5034	1.04	[0.00 , 17916]	[0.03 , 3.85]	[7535 , 21581]	[0.77 , 3.50]
méthode d'Ip	862	0.17	[7106 , 10268]	[1.62 , 2.26]		
	η	β	η	β	η	β
boot. type I	2942	0.88	[3358 , 14147]	[0.0 , 2.82]	[6316 , 14672]	[0.62 , 3.26]
boot. pessimiste	3642	0.69	[2075 , 15430]	[0.0 , 2.47]	[7156 , 18464]	[0.59 , 2.74]
méthode d'Ip	1302	0.23	[6365 , 11140]	[0.8 , 1.63]		

TAB. 2.8 – Estimations de Newton-Raphson pour les jeux $A215$ (haut) et $B215$.

⁵On a par ailleurs remarqué numériquement que les intervalles construits d'après les estimations SEM sont en règle générale moins précis que les intervalles construits d'après les estimations NR : le rajout de bruit par SEM est préjudiciable à la précision de l'estimation et à la validité des intervalles de confiance.

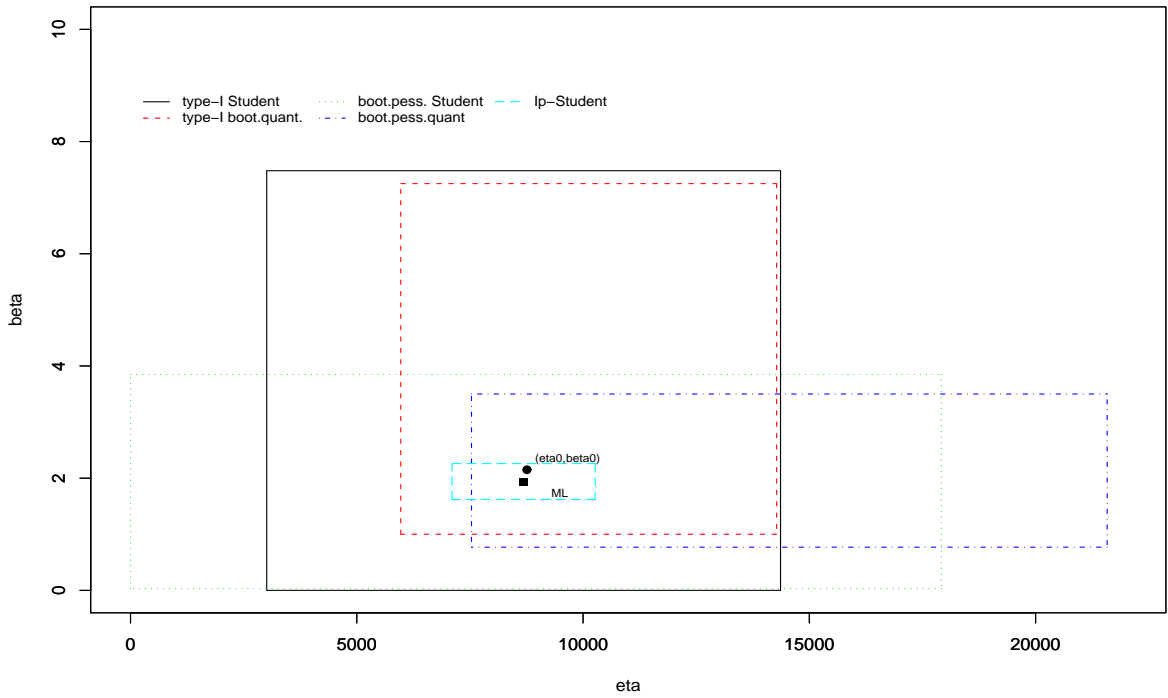


FIG. 2.7 – Comparaison des intervalles de confiance à [5% – 95%] autour de $(\hat{\eta}_{NR}, \hat{\beta}_{NR}) = (8687, 1.94)$ pour le jeu *A215*.

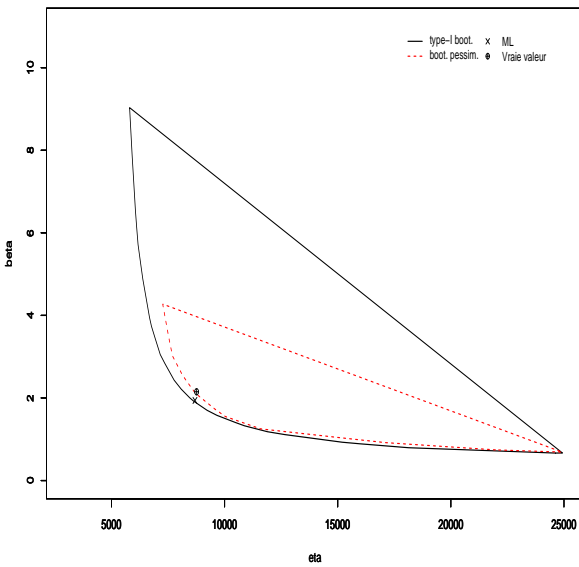


FIG. 2.8 – Zone de confiance jointe à 90% pour $(\hat{\eta}_{NR}, \hat{\beta}_{NR}) = (8687, 1.94)$ (jeu *A215*) estimée par le bootstrap de type I et le bootstrap pessimiste.

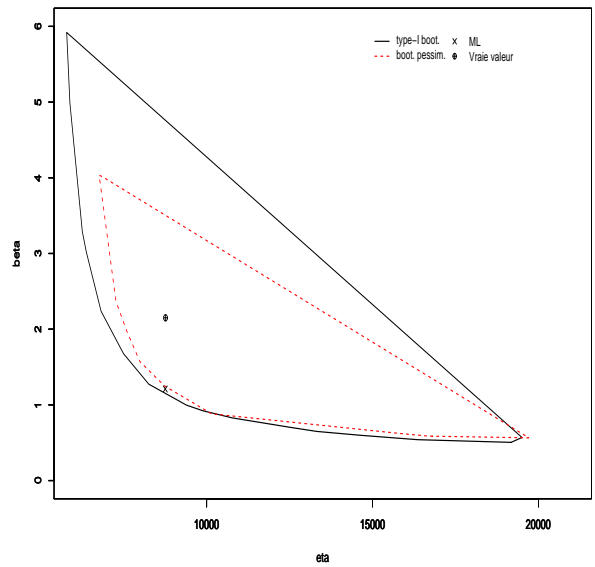


FIG. 2.9 – Zone de confiance jointe à 90% pour $(\hat{\eta}_{NR}, \hat{\beta}_{NR}) = (8752, 1.21)$ (jeu *B215*) estimée par le bootstrap de type I et le bootstrap pessimiste.

2.5.4 Évaluation des méthodes par facteur d'erreur

Le biais des estimations, la taille des intervalles de confiance ainsi que leur forte sensibilité à la censure amènent à douter fortement du bien-fondé de l'utilisation des méthodes fréquentistes. Les imprécisions de ces méthodes se répercutent de façon significative sur la modélisation de la variable de durée de vie $X \sim \mathcal{M}(\hat{\theta})$ et des fonctions d'intérêt de θ , caractérisant l'évolution de X et indispensables à la prise de décision. Comme nous l'avons introduit auparavant, un moyen simple de visualiser cette imprécision est de calculer le *facteur d'erreur* (3.10) de quelques fonctions d'intérêt.

On présente sur le tableau 2.9 les différentes estimations bootstrap du facteur d'erreur $\rho_{E(X)}$ de l'espérance de la durée de vie, pour les différents jeux proposés, et pour les méthodes de maximisation de la vraisemblance (directe, NR ou SEM pour les modèles exponentiels et de Weibull). Le tableau 2.10 présente de même les estimations du facteur d'erreur $\rho_{Var(X)}$ de la variance de cette durée de vie.

	<i>Modèle exponentiel</i>				<i>Modèle de Weibull</i>			
	Jeu <i>a1bis</i>		Jeu <i>b1bis</i>		Jeu <i>A215</i>		Jeu <i>B215</i>	
boot. type I	1.05	1.06	1.09	1.09	1.50	1.12	1.51	1.25
boot. type II	1.89	1.93	1.78	1.78	1.69	1.34	2.04	1.63
boot. optimiste	1.05	1.06	1.06	1.07	1.53	1.16	1.60	1.28
boot. pessimiste	1.06	1.07	1.07	1.08	1.56	1.19	1.59	1.35

TAB. 2.9 – Estimations bootstrap (EM/SEM) du facteur d'erreur $\rho_{E(X)}$ de l'espérance de la durée de vie (M=4000).

	<i>Modèle exponentiel</i>				<i>Modèle de Weibull</i>			
	Jeu <i>a1bis</i>		Jeu <i>b1bis</i>		Jeu <i>A215</i>		Jeu <i>B215</i>	
boot. type I	1.13	1.17	1.21	1.21	10.38	5.27	7.14	4.69
boot. type II	3.63	3.80	3.21	3.24	9.37	6.48	11.47	6.88
boot. optimiste	1.13	1.16	1.15	1.17	5.94	2.97	6.61	3.59
boot. pessimiste	1.14	1.18	1.19	1.19	7.13	3.61	6.92	5.54

TAB. 2.10 – Estimations bootstrap (EM/SEM) du facteur d'erreur $\rho_{Var(X)}$ de la variance de la durée de vie (M=4000).

Si l'estimation de $E(X)$ provenant de l'estimation $\hat{\theta}$ est admissible d'une manière générale, il en va tout autrement de celle de $Var(X)$, en tout cas pour le modèle de Weibull. Au-delà de ces résultats, les méthodes fréquentistes classiques sont visiblement peu efficaces lorsqu'on cherche à estimer θ à partir de données en faible nombre et censurées, sur des modèles possédant plus d'un paramètre.

2.6 Conclusion

L'application des méthodes classiques d'estimation et d'encadrement des paramètres, des fonctions d'intérêt et des quantités observables des modèles de durée de vie est extrêmement sensible à la censure et à la taille des données disponibles. L'imprécision de telles méthodes peut être détectée par l'estimation d'intervalles de confiance, eux-mêmes le plus souvent construits *via* des méthodes de bootstrap. Ces méthodes de bootstrap ont vocation à répéter la procédure d'estimation sur des échantillons dont l'origine est supposée reproduire les conditions réelles de l'expérience. Par ailleurs, l'écart entre l'estimation et les valeurs centrales d'une distribution bootstrap permettent d'estimer le biais de la méthode, voire de le corriger.

La difficulté de reproduire le plan d'échantillonnage, c'est-à-dire les conditions de l'expérience, la taille des données et la présence de censures rend très incertaine l'aide à la décision à partir des données de REX seules. Des ordres de grandeur de taille acceptable sont fournis par des études bibliographiques (voir § 2.2.2.4) et peuvent diriger la méthodologie employée par les ingénieurs, mais il est toujours intéressant de dérouler des méthodes d'estimation classiques sur des exemples réels ou proches du réel et d'évaluer leur bien-fondé, au cas par cas. La possibilité d'erreur d'estimation pouvant entraîner de lourdes conséquences de coût ou de sécurité (selon l'objectif de rentabilité ou de fiabilité que l'on se fixe), d'autres techniques d'estimation doivent être mises à contribution.

Les méthodes bayésiennes permettent ainsi de régulariser et d'améliorer la précision des estimations et des intervalles de confiance à partir d'une connaissance *a priori* du comportement de Σ . Cette connaissance peut revêtir une forme objective, soit parce qu'elle provient de données anciennes, soit parce qu'elle indique une information certaine (par exemple, des interdits physiques).

Par ailleurs, la présence de données de censure dues à la maintenance témoigne souvent d'une connaissance *experte* du système Σ . Un exemple typique est la connaissance du vieillissement de Σ sans que l'on observe réellement des défaillances dues à l'usure du matériel, Σ étant arrêté et remplacé préventivement. Ce type de connaissance prospective est subjectif et peut permettre, en étant incorporé dans une étude statistique, d'améliorer considérablement la connaissance totale de Σ .

La nécessité de l'utilisation d'une information objective ou subjective, dans le contexte industriel qui est le nôtre, constituera le socle des chapitres suivants. Ainsi, le chapitre 3 propose tout d'abord une introduction aux techniques d'estimation bayésienne dans notre contexte de durabilité et de fiabilité. Il sera majoritairement consacré à la construction de modélisations *a priori* susceptibles de représenter efficacement cette connaissance experte, dans le cadre d'étude voulu par EDF ; modélisations ensuite utilisées pour l'estimation *a posteriori* des modèles de durée de vie présentés au chapitre 1. Le chapitre 4 traitera d'un problème auquel l'inférence bayésienne subjective peut typiquement se heurter, celui de la disparité entre connaissances objective et subjective. L'importance donnée à une connaissance *a priori* et sa calibration vis-à-vis de la connaissance objective des données disponibles seront des points d'étude majeurs de ces deux chapitres.

2.7 Annexes

Supposons avoir un estimateur fréquentiste $\hat{\theta}$ de θ . Nous utilisons une procédure de bootstrap pour estimer la variance V de cet estimateur. Dans le cadre du bootstrap “censuré” que nous proposons, nous simulons des données de censure en inversant la fonction de répartition des censures observées, relative à l'ensemble des données observées. L'impact de la variance de ces données aléatoires sur l'estimation de V est difficile à percevoir, excepté pour le modèle exponentiel, comme l'indique la proposition suivante.

PROPOSITION 1. (*Modèle exponentiel*). Soit un échantillon de taille n $\mathbf{y}_n = (\mathbf{x}_r, \mathbf{c}_{n-r})$ où $\mathbf{x}_r = (x_1, \dots, x_r) \sim \mathcal{E}(\eta_o)$ et $\mathbf{c}_{n-r} = (c_1, \dots, c_{n-r})$ un ensemble de censure à droite de type I progressif. Soit $\hat{\eta}_{ML}$ le maximum de vraisemblance. Soit σ_c^2 la variance inconnue du processus de censure dans la création des échantillons bootstrap. On suppose pouvoir simuler M échantillons bootstrap de taux de censure égal à celui de l'échantillon initial. Alors, lorsque $M \rightarrow \infty$, l'estimateur bootstrap de la variance de $\hat{\eta}_{ML}$ converge vers

$$\frac{(n-r)^2}{r^2} \sigma_c^2 + \frac{\hat{\eta}_{ML}^2}{r}.$$

Preuve. On suppose que chaque échantillon bootstrap $\mathbf{Y}_n^{(j)} = (Y_1^{(j)}, \dots, Y_n^{(j)})$, $\forall j \in \{1, \dots, M\}$, possède le même nombre r de données censurées (on suppose simuler suffisamment d'échantillons pour n'utiliser que ceux qui respectent cette règle). Chaque échantillon bootstrap se décompose entre un sous-échantillon $\mathbf{X}_n^{(j)} \rightsquigarrow \mathcal{E}(\eta_o)$ et un sous-échantillon de censures $\mathbf{C}_{n-r}^{(j)} \rightsquigarrow \mathcal{F}_c$. Soit σ_C^2 la variance (*a priori* inconnue) du processus de censure et e_c son espérance.

Posons $T_j = \sum_{i=1}^n Y_i^j$. Par indépendance des échantillons bootstrap entre eux, les T_j sont i.i.d. On peut alors noter $T_j = A_j + B_j$ avec

$$\begin{aligned} A_j &= \sum_{i=1}^r X_i^{(j)} \rightsquigarrow \mathcal{G}(r, \hat{\eta}_{ML}^{-1}) \\ B_j &= \sum_{i=1}^{n-r} C_i^{(j)}. \end{aligned}$$

Les A_j et B_j sont i.i.d., et indépendants entre eux. La variance de B_j est alors $(n-r)^2 \sigma_C^2$ et son espérance vaut $(n_r) e_c$. La variance bootstrap de $\hat{\eta}_{ML}$ s'écrit alors

$$\begin{aligned} \text{Var}[\hat{\eta}_{ML}] &= \frac{1}{r^2 M(M-1)} \sum_{j=1}^M \left(MT_j - \sum_{i=1}^M T_i \right)^2 \\ &= \frac{1}{r^2 M^2 (M-1)} \sum_{j=1}^M \left[\left(MA_j - \sum_{i=1}^M A_i \right) + \left(MB_j - \sum_{i=1}^M B_i \right) \right]^2 \\ &= \frac{1}{r^2 M^2 (M-1)} \sum_{j=1}^M \left[\left((M-1)A_j - \sum_{i=1, i \neq j}^M A_i \right) + \left((M-1)B_j - \sum_{i=1, i \neq j}^M B_i \right) \right]^2 \end{aligned}$$

Posons $U_j = (M-1)A_j$ et $V_j = \sum_{i=1, i \neq j}^M A_i$. Ces deux variables sont indépendantes, et

$$\begin{aligned} U_j &\sim \mathcal{G}(r, [(M-1)\hat{\eta}_{ML}]^{-1}), \\ V_j &\sim \mathcal{G}(r(M-1), \hat{\eta}_{ML}^{-1}). \end{aligned}$$

Posons également $W_j = (M - 1)B_j$ et $Q_j = \sum_{i=1, i \neq j}^M B_i$. Ces deux variables sont indépendantes entre elles, et indépendantes de U_j et V_j . Alors

$$\begin{aligned} \mathbb{E}[W_j] &= \mathbb{E}[Q_j] = (M - 1)(n - r) e_c, \\ \text{Var}(W_j) &= (M - 1)^2(n - r)^2\sigma_c^2, \\ \text{Var}(Q_j) &= (M - 1)(n - r)^2\sigma_c^2. \\ \mathbb{E}[U_j] &= \mathbb{E}[V_j] = (M - 1)r \hat{\eta}_{ML}, \\ \text{Var}(U_j^2) &= (M - 1)^2r \hat{\eta}_{ML}^2, \\ \text{Var}(V_j^2) &= (M - 1)r \hat{\eta}_{ML}^2. \end{aligned}$$

On en déduit

$$\begin{aligned} \text{Var}[\hat{\eta}_{ML}] &\stackrel{M \rightarrow \infty}{\simeq} \frac{1}{r^2 M^2 (M - 1)} \sum_{j=1}^M [M(M - 1)r \hat{\eta}_{ML} + M(M - 1)(n - r)^2\sigma_c^2] \\ &\rightarrow \frac{(n - r)^2}{r^2} \sigma_c^2 + \frac{\hat{\eta}_{ML}^2}{r}. \quad \square \end{aligned}$$

Chapitre **3**

Une modélisation de l'expertise industrielle en durabilité

Table locale

3.1	Introduction	58
3.2	Modélisation non informative	61
3.3	Une modélisation informative des paramètres de Weibull	65
3.4	Analyse <i>a posteriori</i>	86
3.5	Indicateurs de taille des données observées	93
3.6	Annexes	97

3.1 Introduction

Paradigme bayésien et apport d'information extérieure

À la suite des constatations du chapitre 2, il est légitime de faire appel, en sus de données \mathbf{y}_n faiblement informatives, à l'ajout d'une connaissance autre du comportement du composant Σ , si elle existe. On espère alors, à condition de pouvoir modéliser raisonnablement cette connaissance, en tirer une précision suffisante dans l'estimation de fonctions d'intérêt de θ . Une façon courante d'agglomérer ces deux connaissances est de faire appel aux techniques bayésiennes, en donnant au paramètre θ la nature d'une variable aléatoire sur Θ . Une référence majeure en la matière est le livre de Robert (2001). Si l'on peut construire une mesure *a priori* $\pi(\theta)$ selon une connaissance indépendante de \mathbf{y}_n , la mise en œuvre du paradigme bayésien permet de définir une mesure *a posteriori* sur Θ

$$\pi(\theta|\mathbf{y}_n) = \frac{\mathcal{L}(\mathbf{y}_n; \theta) \pi(\theta)}{\int_{\Theta} \mathcal{L}(\mathbf{y}_n; \theta) \pi(\theta) d\theta}$$

qui est une densité de probabilité si $\int_{\Theta} \mathcal{L}(\mathbf{y}_n; \theta) \pi(\theta) d\theta < \infty$. Ce faisant, les fonctions d'intérêt de θ que l'on cherchait auparavant à estimer deviennent à leur tour des variables aléatoires dont on tente, *a posteriori*, d'estimer les grandeurs caractérisantes (espérance, mode, etc.). La connaissance explicite de la densité *a posteriori* $\pi(\theta|\mathbf{y}_n)$, d'une approximation de cette densité ou au minimum la possibilité de simuler en grand nombre des réalisations $(\theta_1, \dots, \theta_M)$ permettent d'estimer des grandeurs *prédictives* de la forme

$$g_i(x|\mathbf{y}_n) = \int_{\Theta} g(x|\theta) \pi(\theta|\mathbf{y}_n) d\theta$$

qui sont plus simplement accessibles à l'ingénieur habitué à travailler sur des grandeurs observables. Offrir un caractère "raisonnable" à cette modélisation, selon des indicateurs les plus objectifs possibles, est l'étape de *calibration*. En peu de mots, elle consiste à choisir l'incertitude sur θ transmise par $\pi(\theta)$. Dans ce chapitre comme dans le chapitre 4 (dans un cadre plus général), nous cherchons à sensibiliser l'analyste industriel aux enjeux importants de cette calibration.

L'objectif du présent chapitre est de présenter en détail le cadre bayésien d'une étude industrielle en durabilité, où les modèles poly-Weibull sont utilisés (regroupant les modèles exponentiel, de Weibull classique et de compétition entre Weibull).

Nous introduisons tout d'abord l'aspect formel de la modélisation bayésienne non informative et nous en présentons quelques caractéristiques exploitables. Cette modélisation permet, dans notre cadre d'étude, de structurer de façon probabiliste la compréhension de fonctions d'intérêt estimées auparavant dans un cadre fréquentiste : la maximisation de la vraisemblance est ainsi perçue comme l'estimation du mode de $\pi^J(\theta|\mathbf{y}_n)$ où π^J modélise une ignorance. Proposée dans le but de régulariser les résultats fréquentistes (en "gommant" par exemple certains maximums locaux de la vraisemblance), cette modélisation ne sera pas particulièrement étudiée dans la suite du chapitre mais interviendra constamment au sein du chapitre 4.

En effet, l'expertise industrielle existe, apporte une véritable information, et s'en priver est dommageable. Matériaux, construction, conditions d'exploitation : de nombreuses connaissances humaines sont mises en œuvre dans l'industrie pour établir des stratégies d'entretien ou de renouvellement des composants. Ces connaissances, si elles s'appuient en général sur des contraintes objectives (caractéristiques physiques des matériaux, par exemple) ou des données anciennes (sur des composants analogues

à \sum), restent relatives à une personne donnée, dans une situation donnée. Les réponses que les experts industriels peuvent apporter à une série de questions de l'analyste bayésien sont entachées d'une subjectivité par essence difficilement modélisable.

On confondra alors dans ce terme - *analyste bayésien* - toute personne qui interroge le ou les experts, recueille leur(s) réponse(s), utilise des critiques sur ces avis d'expert, additionne les informations subjectives aux éventuelles informations objectives disponibles (autres que \mathbf{y}_n), construit une modélisation *a priori* $\pi(\theta)$ et la calibre en fonction des informations obtenues.

L'aspect le plus important de ce chapitre est donc le choix et l'étude d'une famille de densités *a priori* $\pi(\theta)$, spécialement dédiée à l'inférence de Weibull et ce dans le cadre d'expertise sur \sum proposé par EDF. En particulier, ce chapitre se place dans la continuité des travaux de Bacha (1996) et Berger & Sun (1993) en modélisant de façon conditionnelle les deux paramètres $\theta = (\eta, \beta)$ de Weibull et en approfondissant leur approche. Tout au long de cette étude, nous cherchons en effet à dégager un sens clair de cette représentation *a priori*, en nous efforçant de respecter un ensemble de règles de cohérence avec l'expertise et le contexte de l'étude : modéliser le vieillissement de \sum .

Au final, nous aboutissons à un outil flexible et intelligible pour l'analyste bayésien. Nous désirons, *via* ce travail, sensibiliser cet analyste à des aspects importants de la modélisation subjective, l'enjeu principal étant de modérer une expertise qui menacerait d'être trop informative et d'entraîner, *a posteriori*, des résultats trop subjectifs. Ainsi, au cours du texte, nous proposons quelques questions ouvertes et d'autres très focalisées, qui pourraient servir de points de départ au dialogue entre analyste et expert.

REMARQUE 4. *Les principaux points développés dans ce chapitre sont résumés (en anglais) dans un rapport de recherche (Bousquet 2006), disponible en Annexes B. Celui-ci propose par ailleurs une démarche de calibration par défaut de la loi a priori du paramètre de forme β de Weibull. Enfin, un exemple réel, faisant le récapitulatif des aspects méthodologiques proposés dans ce chapitre, est mis en exergue dans ce rapport.*

Vocabulaire et notations

Par la suite (dans ce chapitre comme dans les suivants) nous serons fréquemment amenés à utiliser le vocabulaire courant de la statistique bayésienne.

- Le terme *hyperparamètre* désigne le paramètre d'une mesure $\pi(\theta)$ ou $\pi(\theta|\mathbf{y}_n)$.
- Une mesure *a priori* π est *propre* s'il s'agit d'une densité (de probabilité), i.e. si $\int_{\Theta} \pi(\theta) d\theta < \infty$.
- On dit qu'une mesure π est *informative* si son *entropie relative* (ou différentielle) $\int_{\Theta} \pi(\theta) \log \pi(\theta) d\theta$ est finie. L'entropie relative est un indicateur de l'incertitude d'une mesure, comme la variance, mais cette dernière peut ne pas être définie. On peut donc percevoir une mesure informative comme possédant un indicateur d'incertitude mesurable (et donc calibrable).
- *Changement de paramétrisation*. Dans le reste de ce travail, on sera amené à considérer plusieurs paramétrisations différentes, notamment en raison de la facilité à manier des densités de lois bien connues. Si on connaît une mesure π_{θ} pour $\theta \in \Theta \subset \mathbb{R}^p$, il est aisé de connaître $\pi_{\theta'}$ pour une

autre paramétrisation θ' en considérant la transformation bijective $\Phi : \theta \mapsto \theta'$ de Jacobien

$$J(\theta') = \det \left[\frac{\partial^2 \Phi(\Phi^{-1}(\theta'))}{\partial \theta_i \partial \theta_j} \right]_{i,j \in \{1, \dots, p\}}.$$

Alors $\pi_{\theta'}(\theta') = J^{-1}(\theta') \pi_{\theta}(\Phi^{-1}(\theta'))$.

- *Probabilités fréquentiste et bayésienne.* Soit $t(\theta)$ une fonction d'intérêt de θ . Lorsque θ est perçu comme une variable aléatoire de densité π , on notera $P(t(\theta) \in A)$ la probabilité bayésienne (A étant un évènement fixé dans \mathcal{A}) en opposition à la probabilité fréquentiste $P_{\theta}(t(\theta) \in A)$ où θ est fixé et A est un évènement aléatoire.

3.2 Modélisation non informative

3.2.1 Motivations

Introduit par Jeffreys (1946, 1961), l'emploi de mesures *a priori* non informatives π^J tente d'apporter des réponses à un ensemble de questions légitimes lorsqu'on cherche à modéliser une ignorance. Un exemple très simple illustre la vision générale que développe Bernardo (1997).

EXEMPLE 3. Soit le modèle exponentiel de densité $f(x|\theta) = \theta \exp(-\theta x)$. On choisit $\pi(\theta)$ comme une loi Gamma d'hyperparamètres a et b . Si l'on fixe la moyenne $m = a/b$ en utilisant une connaissance ponctuelle de θ (provenant par exemple d'un expert), on peut augmenter la variance $a/b^2 = m^2/a$ en faisant tendre a vers 0. La densité *a priori* deviendra de plus en plus "plate", reflétant une précision qui s'amenuise jusqu'à rendre inutile la connaissance de m . Le cas limite est évidemment obtenu en $(a, b) = (0, 0)$, mais l'*a priori* est alors réduit à une mesure non intégrable $\pi^J(\theta) \propto \theta^{-1}$. ■

Une façon intuitive de percevoir un *a priori* non informatif est ainsi de le faire correspondre à une limite d'*a priori* informatifs. Il paraît encore plus intuitif de proposer pour θ des mesures uniformes sur Θ (π^J étant alors d'entropie maximale). Cependant, une solution si expéditive aboutit (entre autres) à des *paradoxes de partitionnement* (Kass & Wasserman 1996) lorsqu'on choisit une autre paramétrisation θ' , pour lesquelles les mesures ne sont plus obligatoirement uniformes (il suffit de choisir $\theta' = h(\theta)$ où h n'est pas linéaire). Voir également Robert (2006, § 3.5 pour des précisions et des exemples en français) À moins de choisir une unique paramétrisation θ , le choix d'un π^J non informatif doit prendre en compte certaines invariances du problème d'estimation, et cela d'autant plus qu'au final, c'est moins l'estimation paramétrique que l'estimation marginale (sur X) qui nous intéresse. La mesure de Jeffreys $\pi^J(\theta) \propto \{\det I_\theta\}^{1/2}$, où I_θ est la matrice d'information de Fisher du modèle considéré, est ainsi invariante quel que soit le choix de paramétrisation.

L'article de Kass & Wasserman (1996) constitue une référence majeure en la matière, en introduisant et expliquant les principales règles formelles d'élicitation et les problèmes d'invariance et de paradoxes liés à l'emploi d'*a priori* non informatifs. Étudié par un très grand nombre d'auteurs, l'emploi de ces mesures se révèle précieux pour régulariser de façon objective l'information fréquentiste fournie par la vraisemblance des données. Citons par exemple Hartigan (1964), Bernardo (1979), Berger & Bernardo (1992), Robert (2001, Chap.3 & 7), Zellner (1977, 1996), Yang & Berger (1997), Syversveen (1998), pour leur clarté et leur généralité. Dans une optique d'estimation bayésienne, une règle d'élicitation prévaut de par la validité fréquentiste qu'elle induit sur le recouvrement *a posteriori* (Peers 1965, Ghosh & Mukerjee 1993, Datta & Ghosh 1995, Datta 1996, Ghosal 1999b) : on caractérise π^J par la définition suivante.

DÉFINITION 1. Soit $\mathbf{X}_n = X_1, \dots, X_n \sim \mathcal{M}(\theta)$. Soit $\theta_n(\alpha)$ le quantile d'ordre α *a posteriori* (i.e. $P(\theta \leq \theta_n(\alpha) | \mathbf{X}_n) = \alpha$). La mesure *a priori* π^J est dite d'ordre de recouvrement i et notée *CMP* d'ordre i (*coverage matching prior*) si

$$\underbrace{P_\theta(\theta \leq \theta_n(\alpha))}_{\text{probabilité fréquentiste}} = \underbrace{P(\theta \leq \theta_n(\alpha) | \mathbf{X}_n)}_{\text{probabilité bayésienne}} + \mathcal{O}(n^{-i/2}).$$

En d'autres termes, l'utilisation de π^J permet d'obtenir des régions de confiance *a posteriori* pos-

sédant une validité fréquentiste. La mesure π^J peut alors être légitimement considérée comme non informative puisqu'elle n'apporte *a posteriori* pas plus d'information que les données, à $O(n^{-i/2})$ près.

Une définition plus formelle des CMP, vis-à-vis de n'importe quelle fonction d'intérêt *a posteriori* (et plus seulement θ) et pour n'importe quelle dimension de Θ , est présentée dans Datta (1996). La caractérisation générale de π^J comme solution d'une équation différentielle permet de la connaître explicitement dans des cas simples (en particulier quand $\dim \Theta$ est petite). Voir Ghoshal (1999b) pour des exemples.

3.2.2 Application aux modèles de durée de vie

Modèle exponentiel. Quand $\dim \Theta = 1$, π^J est réduit à la mesure de Jeffreys (CMP d'ordre 2, Welch & Peers, 1963). Pour n'importe quelle paramétrisation θ d'un modèle d'échelle (*scale model*), $\pi^J(\theta) \propto \theta^{-1}$.

Modèle de Weibull. Sun (1997) propose une étude des meilleurs CMP pour le modèle de Weibull. Le meilleur CMP est le *reference prior* de Berger & Bernardo (1992), que l'on fasse le choix d'une hiérarchie (d'un conditionnement) entre les paramètres ou non. L'ordre de recouvrement dépend du choix de paramétrisation. Ainsi $\pi^J(\eta, \beta) \propto (\eta\beta)^{-1}$ est d'ordre 2 mais $\pi^J(\mu, \beta) \propto (\mu\beta)^{-1}$ est réduit à la mesure de Jeffreys d'ordre 1.

Modèle à risques concurrents. Les mesures non informatives standard pour le modèle de Weibull produisent des *a posteriori* impropres pour un modèle de compétition entre Weibull (Berger & Sun 1993). L'utilisation de mesures impropres est donc proscrite. Wang & Ghosh (2000) ont cependant développé une stratégie de construction d'*a priori* non informatifs pour les paramètres identifiables d'un modèle de compétition exponentiel.

3.2.2.1 Cas des données censurées à droite

Les valeurs de censure éventuellement présentes influent sur I_θ et permettent donc d'en déduire la forme des mesures de Jeffreys correspondantes. Cette démarche a notamment été entreprise par De Santis, Mortera et Nardi (2001). On suppose donc avoir $\mathbf{y}_n = (y_1, \dots, y_n)$ valeurs observées parmi lesquelles r valeurs non censurées $\mathbf{x}_r = (x_1, \dots, x_r)$ de densité $f(x|\theta)$ et de fonction de répartition $F(x|\theta)$. On note $\mathbf{c}_{n-r} = (c_1, \dots, c_{n-r})$ les valeurs de censure. Si l'on peut supposer que le processus de censure est aléatoire et spécifier une densité des instants de censure f_c , on note

$$\delta(\theta) = \int_{\mathbb{R}} f_c(x) F(x|\theta) dx.$$

On aboutit alors aux résultats des théorèmes 1 et 2 et des propositions 2 et 3 (preuves en Annexes pp. 97-99).

THÉORÈME 1. (De Santis et al. 2001). *Pour n'importe quelle paramétrisation θ , la mesure de Jeffreys du modèle exponentiel censuré à droite est*

$$\pi_c^J(\theta) \propto \theta^{-1} \sqrt{\delta(\theta)}. \quad (3.1)$$

PROPOSITION 2. *Pour la paramétrisation λ du modèle exponentiel, on obtient :*

1. Censure de type I progressif : $\delta(\lambda) = n - \sum_{i=1}^{n-r} \exp(-\lambda c_i)$;
2. Censure uniforme sur $[0, c_\infty]$: $\delta(\lambda) = 1 - \{1 - \exp(-\lambda c_\infty)\}(\lambda C_\infty)^{-1}$;
3. Censure de type II : $\delta(\lambda) = r/n$ est indépendant de λ . Donc $\pi_c^J(\lambda) = \pi^J(\lambda)$.

THÉORÈME 2. Soit γ la constante d'Euler ($\gamma \simeq 0.57722$). Soient $\gamma_1 = \pi^2/6 + \gamma^2 - 2\gamma > 0$, $\gamma_2 = -2(1 - \gamma) < 0$. On note $\tilde{\delta}(\mu, \beta) = \delta^2(\mu, \beta) + [\delta(\mu, \beta) - 1](\gamma_1 + \gamma_2 \log \mu + \log^2 \mu) + \pi^2/6 - 1$. Alors la mesure de Jeffreys pour la paramétrisation (μ, β) du modèle de Weibull censuré à droite est

$$\pi_c^J(\mu, \beta) \propto (\mu\beta)^{-1} \sqrt{\tilde{\delta}(\mu, \beta)}. \quad (3.2)$$

PROPOSITION 3. Pour le modèle de Weibull, on obtient

1. Censure de type I progressif : $\delta(\mu, \beta) = n - \sum_{i=1}^{n-r} \exp(-\mu c_i^\beta)$;
2. Censure uniforme sur $[0, c_\infty]$: $\delta(\mu, \beta) = 1 - (\beta\mu^{1/\beta})^{-1} \{1 - \Gamma(\beta^{-1}, \mu c_\infty^\beta)\}$ où $\Gamma(x, \alpha)$ est la fonction gamma incomplète ;
3. Censure de type II : $\delta(\mu, \beta) = r/n$.

Quelques remarques peuvent être faites sur les résultats des propositions 2 et 3. Dans les cas de censure de type I progressif et uniforme, une valeur de censure infinie ou nulle entraîne que $\pi_c^J(\theta)$ est l'*a priori* standard $\pi^J(\theta)$. Par définition, une valeur de censure de type II n'est jamais nulle ou infinie. Consistant en la répétition d'une valeur qui correspond à la plus vieille défaillance observée, cette censure n'apporte pas d'information supplémentaire sur θ en sus de cette donnée non censurée.

Selon De Santis *et al.* (2001), l'utilisation de π_c^J permet de diminuer légèrement l'incertitude sur l'estimation de θ vis-à-vis de l'emploi d'une mesure de Jeffreys π^J classique. Par ailleurs, pour le modèle exponentiel, une observation censurée suffit à obtenir un *a posteriori* propre (et donc utilisable pour l'estimation), ce qui n'est pas le cas de π^J . Enfin, π_c^J se révèle posséder un meilleur recouvrement fréquentiste *a posteriori* que π^J , quand le pourcentage de censure augmente dans l'échantillon. D'une façon générale, les auteurs remarquent que la non prise en compte des censures à droite dans π^J favorise *a posteriori* des valeurs observables élevées. Selon la connaissance que l'on a du processus de censure, il est donc intéressant d'utiliser de tels *a priori* non informatifs, en particulier dans un cadre fiabiliste. Leur emploi est cependant peu maniable (on perd par exemple les propriétés de conjugaison pour le modèle exponentiel), notamment dans le cas Weibull.

3.2.2.2 Lois *a posteriori*

Modèle exponentiel. Pour le modèle exponentiel, $\pi^J(\theta|\mathbf{y}_n)$ est explicite par conjugaison. Ainsi

$$\lambda|\mathbf{y}_n \sim \mathcal{G}\left(r, \sum_{i=1}^n y_i\right)$$

où r est le nombre de données non censurées (quel que soit le processus de censure). $\pi^J(\lambda|\mathbf{y}_n)$ est propre si $r \geq 1$. L'espérance *a posteriori* $E[\lambda|\mathbf{y}_n]$ est égale au MLE $\hat{\lambda}_n = r / \sum_{i=1}^n y_i$ (l'espérance de la

paramétrisation η n'étant définie que si $r > 1$). Un intervalle bilatéral de niveau α (dit de *crédibilité*) qui découle de la loi *a posteriori* est

$$\left[\frac{\chi_{(2r),\alpha/2}^2}{2 \sum_{i=1}^n y_i}, \frac{\chi_{(2r),1-\alpha/2}^2}{2 \sum_{i=1}^n y_i} \right]$$

qui correspond très exactement à l'intervalle de confiance fréquentiste en cas de censure de type II (voir Chap.2, p. 36, formule 2.2).

Modèle de Weibull. D'après Sun (1997), si π^J est le *reference prior*, $\pi^J(\mu, \beta | \mathbf{y}_n)$ est propre si $r \geq 2$ et les $y_i > 1$. Une hiérarchie entre paramètres s'établit de façon naturelle. En effet, $\pi^J(\mu, \beta | \mathbf{y}_n) = \pi^J(\mu | \beta, \mathbf{y}_n) \pi^J(\beta | \mathbf{y}_n)$ où

$$\mu | \beta, \mathbf{y}_n \sim \mathcal{G} \left(r, \sum_{i=1}^n y_i^\beta \right), \quad (3.3)$$

$$\pi^J(\beta | \mathbf{y}_n) \propto \frac{\beta^{r-2} \left(\prod_{i=1}^r x_i \right)^\beta \left(\sum_{i=1}^n y_i^\beta \right)^{-r}}{\int_0^\infty \beta^{r-2} \left(\prod_{i=1}^r x_i \right)^\beta \left(\sum_{i=1}^n y_i^\beta \right)^{-r} d\beta}. \quad (3.4)$$

L'utilisation de mesures non informatives pour l'estimation reste limitée, car elle permet au mieux de régulariser des résultats fréquentistes imprécis et de corriger les zones de confiance obtenues par le biais du bootstrap. Dans le cas exponentiel, de par la connaissance explicite d'intervalles de confiance fréquentistes, elles apportent peu d'information en sus des résultats fréquentistes. Cependant, elles restent précieuses pour traduire en termes probabilistes, *a posteriori*, une absence d'expertise. Ainsi, elles interviendront de façon importante au chapitre 4, dans la construction de mesures de références pour la calibration d'*a priori* informatifs subjectifs.

3.3 Une modélisation informative des paramètres de Weibull

3.3.1 Motivations

Dans cette section, nous construisons une modélisation informative de densité π pour le paramètre θ d'un modèle poly-Weibull $\mathcal{M}(\theta)$. Vis-à-vis des principaux travaux passés, les aspects novateurs de ce travail sont les suivants. Nous fournissons une modélisation

1. qui tient compte du type d'information proposée par les experts industriels, en particulier dans le cas d'EDF :
 - (a) une information *qualitative* sur le comportement *prospectif* (futur) de Σ ; celle-ci est reliée à l'estimation du ou des paramètres de forme β . Outre une connaissance objective disponible, permettant de borner le domaine de $\pi(\beta)$, des questions précises sont posées aux experts et des valeurs préférentielles de β peuvent être proposées pour utiliser et améliorer des approches existantes (notamment celle de Bacha 1996).
 - (b) une information *quantitative*, s'exprimant en valeurs représentatives de la durée de vie X de Σ ; il s'agit donc d'une information (*post mortem* lorsqu'elle est directement observable), qui est naturellement reliée à l'estimation d'un paramètre d'échelle θ^1 . Cependant, on déduit cette information sur θ' à partir d'une information marginale sur X , ce qui diffère de nombreuses études précédentes, où l'on supposait connue une information directement sur les paramètres. Notre méthode paraît plus utilisable dans un contexte industriel.
2. qui induit volontairement des corrélations logiques entre les composantes de θ , témoignant de la hiérarchie naturelle entre les paramètres.
3. qui est enfin facilement calibrable par l'analyste bayésien. Nous entendons par là que l'analyste doit pouvoir modérer la "force" de l'expertise par le biais de paramètres compréhensibles et peu nombreux. Ainsi, lorsque sa confiance en l'expertise *quantitative* varie, il peut modifier à loisir un hyperparamètre dont le sens sur \mathcal{N} est celui d'une taille de données "fictives".

Cette modélisation s'efforce également de respecter plusieurs conditions de cohérence entre expertises qualitative et quantitative. Par ailleurs, sa structure paramétrique se révèle être un outil de simplification de l'inférence, simplification qui sera par la suite mise à contribution au chapitre 5 (consacré à la mise en œuvre des techniques inférentielles). Techniquement, nous travaillons quasi exclusivement sur le modèle de Weibull à 2 paramètres, noté $\mathcal{W}(\theta)$. La modélisation devient conjuguée dans le cas simple du modèle exponentiel. Balakrishnan & Basu (1995) ou Singpurwalla *et al.* (1999), comme de nombreux autres auteurs, fournissent une étude très complète de l'emploi du modèle exponentiel dans un cadre à la fois bayésien et fiabiliste. Enfin, les sous-modèles exponentiel et de Weibull du modèle $\mathcal{B}(\eta_0, \eta_1, \beta)$ sont indépendants. Il est donc naturel de choisir deux *a priori* indépendants. Restreindre l'étude au modèle de Weibull à 2 paramètres fait donc sens.

¹ou d'une reparamétrisation de ce paramètre d'échelle

Référence sous-jacente de cette section, le livre de Press (2003) offre une vision très claire des diverses approches de modélisation, d'élicitation et d'implémentation de lois *a priori* subjectives, sans entrer toutefois dans les détails techniques. En cela, il est un complément idéal d'ouvrages comme celui de Robert (2001) ou ceux de la série *Bayesian Statistics* (Bernardo *et al.* (eds) 1980-2003). La modélisation que nous proposons peut enfin être présentée comme une évolution des schémas bayésiens introduits par Singpurwalla & Song (1986), Singpurwalla (1988) et Berger & Sun (1993).

3.3.2 La modélisation

L'absence de loi *a priori* continue conjuguée pour le modèle de Weibull (Soland 1969) est un fait connu, du fait de la présence du paramètre de forme β . Celui-ci agit de fait comme un paramètre "nuisible", en l'absence duquel le modèle devient conjugué. Un grand nombre d'auteurs choisissent en général une densité $\pi(\beta)$ qu'ils calibrent en fonction de l'information qualitative disponible, tout en restant flou sur la définition de cette information. C'est par exemple le cas de Singpurwalla & Song (1986) et Berger & Sun (1993). L'expertise quantitative étant plus accessible, leurs études se concentrent surtout sur le choix de la modélisation d'un paramètre d'échelle (η, λ) .

La modélisation *a priori* de η est souvent proposée de façon indirecte ; par exemple, Singpurwalla & Song (1986) puis Singpurwalla (1988) considèrent une modélisation *a priori* sur la médiane $m = \eta(\log 2)^{1/\beta}$ de Weibull et peuvent en déduire, en supposant m et β indépendants, la densité $\pi(\eta)$. Obtenir à peu de frais une expression explicite (ou même définie à une constante près) de la densité sur les deux paramètres reste cependant rare, et très généralement, une simulation est nécessaire pour obtenir des estimateurs (voir par exemple Kaminskiy & Krivtsov 2005).

Or, obtenir explicitement $\pi(\eta, \beta)$ facilite considérablement l'inférence bayésienne. Mais l'hypothèse d'indépendance des paramètres, si elle simplifie les calculs, reste peu justifiée. Nous proposons alors de construire $\pi(\eta, \beta)$ comme une loi conditionnelle

$$\pi(\eta, \beta) = \pi(\eta|\beta) \pi(\beta).$$

La supposition suivante nous paraît en effet raisonnable : pour indiquer une durée de vie représentative du composant Σ , il est nécessaire d'avoir une idée du comportement futur de Σ (vieillesse, rajeunissement, etc.) ou des conditions d'exploitation qui induisent ce comportement futur. Même si cette idée peut rester imprécise, une telle supposition suggère que la connaissance d'un paramètre d'échelle est conditionnée à celle d'un paramètre de forme, alors que celui-ci représente *intrinsèquement* la connaissance de Σ . Dans la suite, nous nous focalisons

1. premièrement, sur le choix de $\pi(\beta)$; nous réutilisons la modélisation proposée par Bacha (1996).
2. en second lieu, sur le choix de $\pi(\eta|\beta)$ ², en reprenant les idées de Berger & Sun (1993).

Pour chacune de ces deux approches, nous proposons des améliorations, soit en précisant un questionnaire déjà existant, soit en précisant le sens des hyperparamètres.

²ou d'une reparamétrisation de $\pi(\eta|\beta)$

3.3.2.1 L'approche de Bacha (1996) pour le paramètre de forme β

Pour le modèle de Weibull $\mathcal{W}(\eta, \beta)$, on admet qu'il existe une connaissance objective de β . Vu comme coefficient directeur de la vitesse de vieillissement (la dérivée du taux de défaillance $h_W(x)$), le paramètre β a un sens *qualitatif* puisqu'il traduit le comportement *prédictif* de Σ . D'une façon générale, un fort consensus des spécialistes industriels indique que l'on peut borner β sur $[\beta_l, \beta_r] \subset]0, +\infty[$ (Bacha 1996).

Dans notre contexte de durabilité, on fait l'hypothèse d'un vieillissement général de Σ et l'on choisira le plus souvent $\beta_l \geq 1$. En outre, Lannoy & Procaccia (1994), comme d'autres spécialistes de fiabilité industrielle, insistent sur le caractère invraisemblable d'un $\beta > 5$, et cela pour deux raisons principales.

1. Si Σ est un composant réparable, il y a généralement suffisamment de dispersion dans les paramètres influençant la défaillance pour que l'on n'observe jamais un vieillissement aussi accéléré.
2. Si Σ est un composant non-réparable, l'industriel agit afin de contrer les défaillances. Auquel cas le modèle statistique représente plutôt une évolution *contrariée* de la durée de vie de Σ , par le biais d'une *maintenance préventive**.

REMARQUE 5. *Industriellement parlant, l'écart entre $\beta_l = 1$ et $\beta_r = 5$ peut être perçu comme un "infini". Un β proche de 5 correspond à une situation où le vieillissement est si accéléré que l'industriel est confronté à une avalanche de difficultés qui rendent l'exploitation impossible. En pratique, $\beta \simeq 3$ est fortement improbable et plus sûrement, β évolue généralement dans $[1, 2.5]$. Le cas $\beta \simeq 1$ correspond ainsi à un industriel qui a correctement effectué son travail de maintenance, en ne laissant pas vieillir le matériel. Une difficulté majeure des fiabilistes est donc de prévoir l'apparition de ce vieillissement, comme l'expliquent Lannoy et al. (2005).*

Ainsi, nous supposons toujours pouvoir choisir (β_l, β_r) . Bacha (1996) propose alors d'utiliser une loi Bêta sur $[\beta_l, \beta_r]$

$$\beta \sim \mathcal{B}_e(p, q)$$

de densité

$$\pi(\beta) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \frac{(\beta - \beta_l)^{p-1} (\beta_r - \beta)^{q-1}}{(\beta_r - \beta_l)^{p+q-1}} \mathbb{1}_{[\beta_l, \beta_r]}(\beta)$$

car elle est flexible au sens où elle possède suffisamment d'hyperparamètres pour être calibrée à la fois en termes de variance et de valeur centrale, indépendamment du couple (β_l, β_r) , propriété qui intéresse tout particulièrement l'analyste industriel lorsqu'il désire modifier légèrement sa loi *a priori*. En témoigne la maniabilité de ses moments et de son mode ;

$$\begin{aligned} \mathbb{E}[\beta] &= \beta_l + \frac{1}{1 + q/p} (\beta_r - \beta_l), \\ \mathbb{M}_d[\beta] &= \beta_l + \frac{1}{1 + (q-1)/(p-1)} (\beta_r - \beta_l) \quad \text{si } p, q > 1, \\ \text{Var}[\beta] &= \frac{pq}{(p+q+1)(p+q)^2} (\beta_r - \beta_l)^2. \end{aligned}$$

Dans le cas particulier où $p = q = 1$, cette loi est uniforme. Par ailleurs, $E_\pi[\beta]$ ne dépend que du rapport q/p et peut rester constante indépendamment d'une modification de la variance (de même pour le mode qui ne dépend que du rapport $(q-1)/(p-1)$). Enfin, les quelques figures 3.1 à 3.10 illustrent l'éventail des formes que la densité peut emprunter (on choisit ici $[\beta_l, \beta_r] = [1, 5]$). Notons que le cas courant d'un vieillissement léger est modélisé par un choix $1 < p < q$ (cf. Figure 3.11).

REMARQUE 6. *La propriété de log-concavité de la loi Bêta permet d'appliquer l'algorithme de Gibbs proposé par Berger & Sun (1993) dans le cas poly-Weibull pour estimer a posteriori les β_i de $i = 1, \dots, Q$ modèles de Weibull mis en compétition. En effet, les distributions a posteriori sont également log-concaves, ce qui permet l'utilisation de fonctions enveloppantes et l'accélération de l'algorithme. Voir le chapitre 5 pour une définition succincte de l'algorithme de Gibbs.*

Jenkinson (2005) propose un inventaire très complet des différentes techniques de sélection des hyperparamètres de la loi Bêta, en particulier lorsque des valeurs de β ont été observées ou reconstruites (provenant par exemple de régressions sur des bases de données ; voir par exemple Lannoy & Procaccia (2005) ou le site <http://www.barringer1.com>). Bacha (1996, p. 69-74) fournit les grandes lignes d'une méthodologie de calibration s'appuyant sur la spécification d'une valeur β_m choisie par défaut comme le milieu de $[\beta_l, \beta_r]$.

Amélioration. Nous ne retraçons pas ici la méthodologie de Bacha mais suggérons d'y ajouter une question permettant de sélectionner β_m indirectement, les experts pouvant difficilement répondre à une interrogation portant directement sur le paramètre. Demandons aux experts les probabilités p_0 et p_1 pour Σ de tomber en panne avant les temps respectifs x_0 et x_1 , avec $x_1 > x_0$. Alors $p_i = 1 - \exp(-\mu x_i^\beta)$ pour $i = 0, 1$. Ainsi

$$\frac{1-p_1}{1-p_0} = \exp \left\{ -\mu x_0^\beta \left(\left\{ \frac{x_1}{x_0} \right\}^\beta - 1 \right) \right\} = (1-p_0) \left(\frac{x_1}{x_0} \right)^\beta - 1.$$

On en déduit

$$\beta_m = \log \left\{ \frac{\log(1-p_1)}{\log(1-p_0)} \right\} \log^{-1}(x_1/x_0). \quad (3.5)$$

En choisissant plusieurs couples (x_0, x_1) , on obtient un certain nombre de valeurs permettant de conforter ou de rectifier $[\beta_l, \beta_r]$ et β_m . Vis-à-vis de Biernacki *et al.* (1998, p.15) qui proposent, dans cette même optique, de demander un rapport de deux valeurs $(h_W(x_0), h_W(x_1))$ du taux de défaillance de Σ et en déduisent β_m , ce type d'interrogation paraît plus adapté car il fait appel à une connaissance passée observable. D'ailleurs, un spécialiste du domaine confirme l'incapacité manifeste des experts du nucléaire à répondre à la question de Biernacki *et al.* (conversation privée avec André Lannoy, 2006). Comme le souligne Clarotti (1998) qui propose quant à lui une interrogation sous forme de *paris* (on demande à l'expert son pourcentage d'erreur lorsqu'il prédit des grandeurs de la durée de vie), l'appel probabiliste à la connaissance subjective réduit les risques de malhonnêteté intellectuelle.

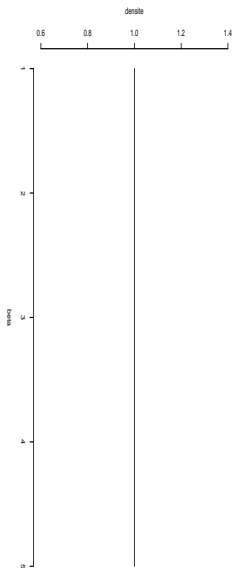


FIG. 3.1 – $(p, q) = (1, 1)$.

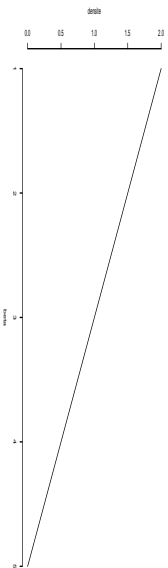


FIG. 3.3 – $(p, q) = (1, 2)$.

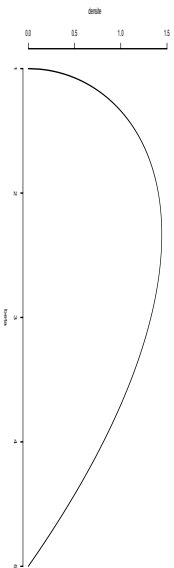


FIG. 3.5 – $(p, q) = (1.5, 2)$.

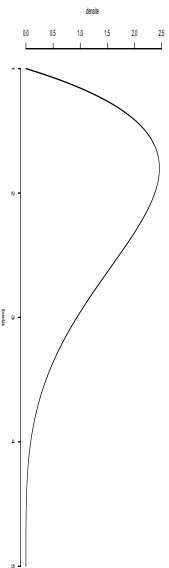


FIG. 3.7 – $(p, q) = (2, 5)$.

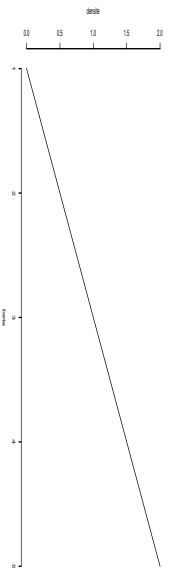


FIG. 3.9 – $(p, q) = (2, 1)$.

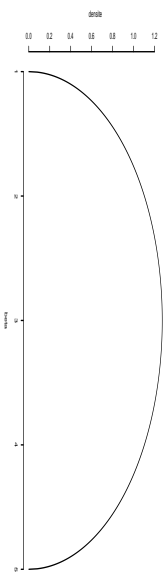


FIG. 3.2 – $(p, q) = (1.5, 1.5)$.

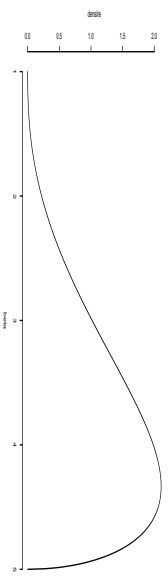


FIG. 3.4 – $(p, q) = (3.5, 1.5)$.

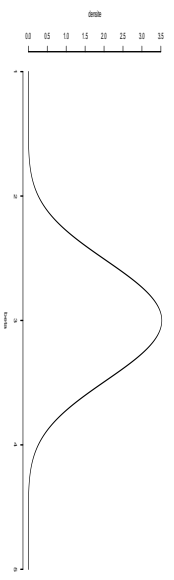


FIG. 3.6 – $(p, q) = (10, 10)$.

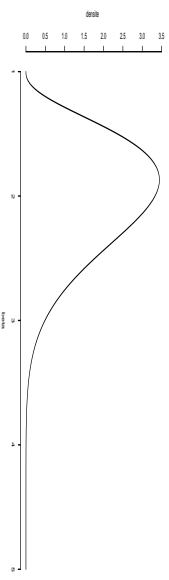


FIG. 3.8 – $(p, q) = (3.5, 10)$.

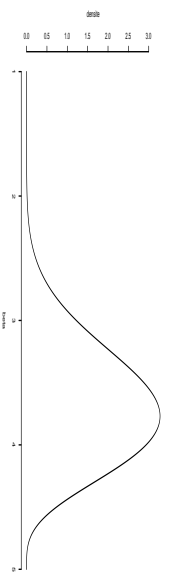


FIG. 3.10 – $(p, q) = (10, 5)$.

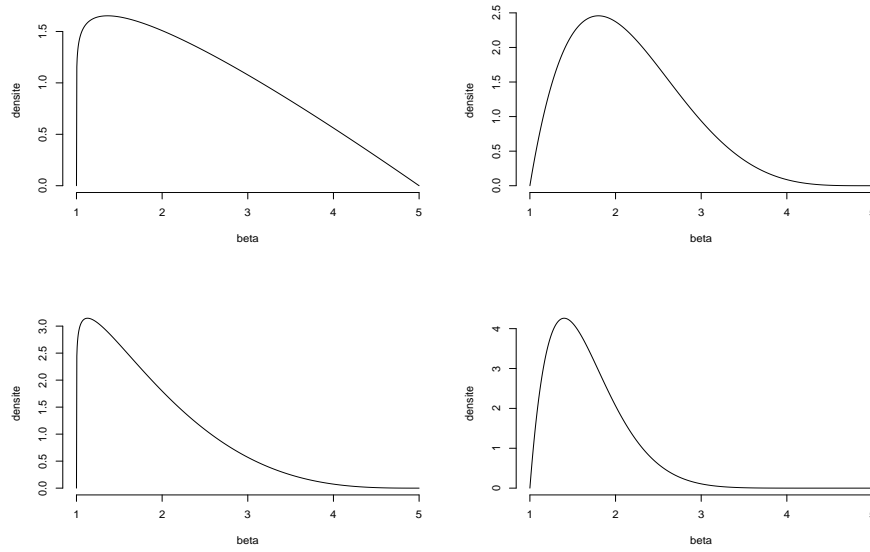


FIG. 3.11 – Formes typiques de densité *a priori* privilégiant un faible vieillissement.

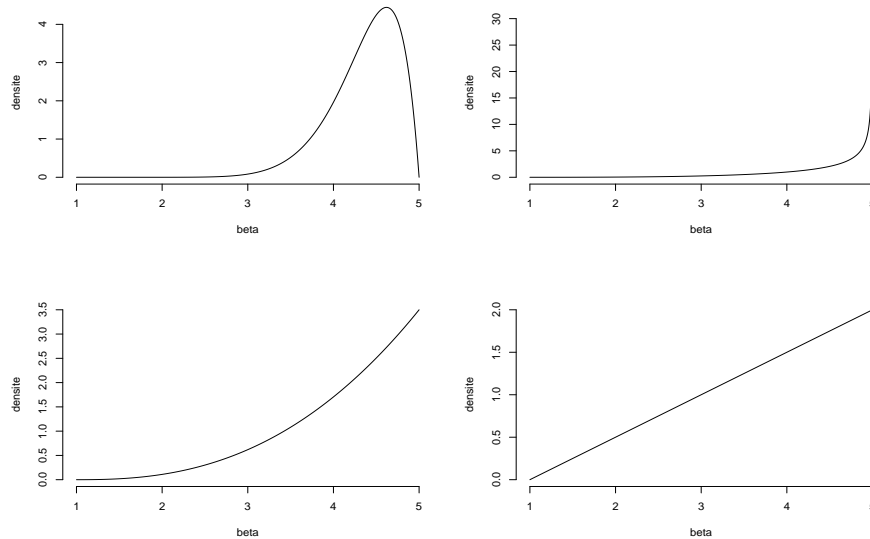


FIG. 3.12 – Formes typiques de densité *a priori* incohérentes avec le choix de β_r .

Cohérence de $\pi(\beta)$. Si π est plus informative qu'une loi uniforme sur $[\beta_l, \beta_r]$, on doit avoir $\pi'(\beta) < 0$ (soit $\pi(\beta)$ décroissante) quand $\beta \rightarrow \beta_r$. Sinon, les valeurs plausibles de β s'étendent au-delà de β_r . Par ailleurs, une décroissance très brutale en des valeurs hautes de β est peu vraisemblable. D'une manière générale, il est nécessaire de vérifier que $\pi(\beta)$ ne favorise pas la borne supérieure β_r (voir figure 3.12). Ajoutons qu'*a posteriori*, une caractérisation similaire de $\pi(\beta|\mathbf{y}_n)$ indique que β_r est mal choisie. Enfin, lorsque n augmente, un MLE $\hat{\beta}_n \notin [\beta_l, \beta_r]$ indique un $\pi(\beta)$ incohérent avec l'information apportée par \mathbf{y}_n et un $\pi(\beta|\mathbf{y}_n)$ menacé d'*inconsistance* (voir § 3.4.1.1).

3.3.2.2 Amélioration de l'approche de Berger & Sun (1993) pour le paramètre d'échelle

Très généralement, les analystes bayésiens supposent *a priori*, pour des raisons de simplicité, que les paramètres de Weibull sont indépendants. Il faut attendre Berger & Sun (1993) pour obtenir, dans un cadre poly-Weibull, une construction délibérément conditionnelle. Les deux auteurs ne précisent pas de choix particulier de $\pi(\beta)$ mais proposent, pour η , d'utiliser une *loi gamma inverse généralisée*.

DÉFINITION 2. $\forall (a, \beta, b) \in (\mathbb{R}^{*+} \times \mathbb{R}^{*+} \times \mathbb{R}^{*+})$, $\forall x \in \mathbb{R}$, on définit la loi gamma inverse généralisée (GIG) par sa densité

$$f(x|a, b, \beta) = \frac{b^a \beta}{\Gamma(a)} \frac{1}{x^{a\beta+1}} \exp\left(-\frac{b}{x^\beta}\right) 1_{[0;+\infty[}(x) \quad (3.6)$$

Son moment d'ordre k est

$$E[X^k] = \frac{b^{k/\beta} \Gamma(a - k/\beta)}{\Gamma(a)} \quad \forall a\beta > k > 0,$$

Son mode est toujours défini et vaut $M_d[X] = \left(\frac{b\beta}{a\beta+1}\right)^{1/\beta}$.

Cette loi *a priori* apparaît comme un outil de simplification de l'inférence. En effet, la densité (3.6) est conjuguée conditionnellement à β (proposition 4) et il suffit donc de connaître ou d'approcher par des moyens numériques le comportement *a posteriori* de β pour obtenir l'ensemble des informations sur les deux paramètres. Par ailleurs, un tel choix permet d'unifier le travail de construction avec le cas exponentiel où β est fixé à 1. Enfin, cette loi est fermée par transformation d'échelle conditionnellement à β (proposition 5), ce qui la rend intéressante pour modéliser des informations *a priori* sur η ou toute reparamétrisation de η conditionnelle à β , comme la médiane, l'espérance ou le mode de la loi de Weibull. Dans la suite de ce chapitre, nous manierons préférentiellement $\pi(\mu|\beta)$ où $\mu = \eta^{-\beta}$, plus simple d'utilisation.

PROPOSITION 4. Soit $\eta|\beta \sim \mathcal{GIG}(a, \beta, b)$. Soit \mathbf{y}_n un échantillon contenant r données non censurées de Weibull et $n - r$ valeurs de censure fixées. En notant $\mu = \eta^{-\beta}$,

$$\begin{aligned} \mu|\beta &\sim \mathcal{G}(a, b), \\ \mu|\beta, \mathbf{y}_n &\sim \mathcal{G}\left(a + r, b + \sum_{i=1}^n y_i^\beta\right), \\ \eta|\beta, \mathbf{y}_n &\sim \mathcal{GIG}\left(a + r, \beta, b + \sum_{i=1}^n y_i^\beta\right) \end{aligned}$$

Preuve. Pour tout $t \geq 0$,

$$\begin{aligned} P(\mu < t) &= 1 - P(\eta < t^{-1/\beta}) \\ &= 1 - \int_0^{t^{-1/\beta}} \frac{b^a \beta}{\Gamma(a)} \frac{1}{x^{a\beta+1}} \exp(-b/x^\beta) dx \\ &= 1 - \int_t^\infty \frac{b^a \beta}{\Gamma(a)} u^{a-1} \exp(-bu) du \quad \text{avec } u = x^{-\beta}. \quad \square \end{aligned}$$

PROPOSITION 5. Soit $X \sim \mathcal{GIG}(a, \beta, b)$. Alors $\forall c > 0$, $cX \sim \mathcal{GIG}(a, \beta, bc^\beta)$.

Preuve. Pour tout $t > 0$,

$$\begin{aligned} P(cX \leq t) &= \int_0^{t/c} \frac{b^a \beta}{\Gamma(a)} \frac{1}{x^{a\beta+1}} \exp\left(-\frac{b}{x^\beta}\right) dx, \\ &= \int_0^t \frac{b^a c^{a\beta} \beta}{\Gamma(a)} \frac{1}{x^{a\beta+1}} \exp\left(-\frac{bc^\beta}{u^\beta}\right) du \quad \text{où } u = cx. \quad \square \end{aligned}$$

Améliorations. Dans l'étude de Berger & Sun (1993), la sélection des hyperparamètres reste seulement au stade théorique. L'expertise est surtout supposée paramétrique (contraintes sur $\theta = (\eta, \beta)$) et les équations intégrales à résoudre sont nombreuses. Par ailleurs, ces auteurs ne proposent pas un *a priori* réellement conditionnel puisque $\pi(\mu|\beta) \equiv \pi(\mu)$. Ce choix reste donc très artificiel. Dans un premier travail, nous avons utilisé cette modélisation et proposé une vision succincte du rôle des hyperparamètres (Bousquet 2005). Cependant, les propriétés structurelles de π indiquent que le choix d'un hyperparamètre b qui soit une fonction déterministe de β plutôt qu'une valeur constante résulte d'une expertise cohérente avec le modèle $\mathcal{W}(\theta)$ et améliore la modélisation en définissant a comme un hyperparamètre de calibration.

En effet, supposons pouvoir modéliser un avis d'expert par un échantillon "fictif" de données observables (non censurées) de Weibull $\tilde{\mathbf{x}}_m = (\tilde{x}_1, \dots, \tilde{x}_m)$ de moyenne x_e . Alors, d'après l'expression (3.3), on a

$$\pi(\theta'|\beta) = \pi^J(\theta'|\beta, \tilde{\mathbf{x}}_m)$$

où π^J est le *reference prior* pour le modèle de Weibull et θ' n'importe quel choix de paramétrisation $\{\eta, \lambda, \mu\}$. Ainsi,

$$\begin{cases} a = m, \\ b = b(a, \beta) = \sum_{i=1}^a \tilde{x}_i^\beta. \end{cases}$$

Le choix $b = b(a, \beta)$ apparaît donc de façon naturelle. Si β est connu et vaut 1 (modèle exponentiel), on obtient $b = ax_e$. L'hyperparamètre a s'interprète comme la taille de l'échantillon fictif. L'analyste bayésien peut donc accorder sa confiance en une certaine expertise quantitative en modulant a .

Idéalement, la loi *a priori* adopte une structure hiérarchique. En supposant $\tilde{\mathbf{x}}_m \stackrel{i.i.d.}{\sim} \mathcal{W}(\eta_e, \beta_e)$ où η_e est une estimation *a priori* de η , on a, sachant $\pi(\beta)$,

$$\begin{aligned} b|\beta &\sim \mathcal{G}(a, \eta_e^{-\beta}), \\ \mu|\beta, b &\sim \mathcal{G}(a, b). \end{aligned}$$

Cependant, cette modélisation est lourde à manier, et le choix d'un b déterministe en fonction de a et β est souhaitable. Selon l'expertise disponible, nous proposons dans la sous-partie suivante des choix de b (proposition 6) qui "miment" le comportement d'une fonction $\beta \mapsto \sum_{i=1}^a \tilde{x}_i^\beta$. On notera par la suite B la distribution de $b = b(\beta)$ (a étant supposé fixé). Ainsi, pour tout événement A de \mathcal{A} dans l'espace probabilisé des données (cf. § 1.4), $P(A|B)$ sera la probabilité de A sachant le choix b .

3.3.3 Traduction de l'expertise quantitative

Nous nous intéressons ici au choix de $b = b(a, \beta)$ en fonction de l'expertise quantitative disponible, portant sur la variable observable X . De nombreux auteurs supposent qu'un expert se prononce directement sur le paramètre d'échelle η ou des fonctions d'intérêt de η . Voir par exemple Bacha (1996), Lijoi (2003) ou Wisse *et al.* (2005). Ce qui est loin d'être évident à obtenir, puisque l'expert n'a en général pas de formation statistique et ne connaît pas le modèle $\mathcal{M}(\theta)^3$. Il s'agit plus sûrement d'un avis "transformé" par l'analyste bayésien.

L'expertise quantitative est en réalité plutôt marginale : un expert perçoit la durée de vie X et en donne une estimation x_e . Dans l'industrie, certains analystes voient en x_e une estimation ponctuelle de η , homogène à X . Agissant par habitude, ces analystes produisent un biais de modélisation d'autant plus important que l'incertitude sur β est grande. Cette confusion fréquente doit être combattue. La valeur x_e est une grandeur caractérisant la distribution marginale M de densité

$$m(x) = \int_{\Theta} \pi(\theta) f_W(x|\theta) d\theta.$$

La spécification de m entraîne alors celle de π , $f_W(x|\theta)$ étant connue. En effet, le théorème de représentation de De Finetti (1974) indique l'existence d'une unique densité π sous des conditions d'échangeabilité des variables $X_1, \dots, X_n, \dots \sim \mathcal{W}(\theta)$ (condition remplie par le choix d'un comportement *i.i.d.*). Cette vision de l'information d'expertise a été utilisée en particulier par Kadane & Wolfson (1980, 1995, 1998) dans des études appliquées. Berger (1985, § 3.5) fournit une revue des principales techniques utilisables. En choisissant $\pi(\theta) = \pi(\theta|\omega)$ dans une famille hyperparamétrisée par ω , des contraintes marginales telles que $E[X] = x_e$ permettent de choisir ω . Par *maximum d'entropie*, on peut également construire une densité π satisfaisant des contraintes de moment, mais de forme arbitraire, parfois difficilement maniable et modifiable, et dont les hyperparamètres risquent d'avoir un sens peu compréhensible.

Nous définissons alors deux types d'expertise quantitative, tels qu'observés dans les problèmes de durabilité et de fiabilité rencontrés par EDF. En Annexes (p. 104), nous proposons parallèlement une modélisation dans le cas où l'information quantitative s'exerce directement sur η .

3.3.3.1 Expertise quantitative minimale

Spécifications. Nous supposons ici qu'un expert, suite à une demande du type "fournissez une grandeur représentative de la durée de vie X du composant Σ " ou "quelle chance a Σ de tomber en panne avant le temps x_e ?", est capable de fournir au minimum l'une des spécifications définies comme suit.

DÉFINITION 3. Soit x_e une valeur de \mathbb{R}^{+*} fournie par une source d'expertise. On nomme spécification de x_e le sens donné à x_e vis-à-vis de M . Cette valeur observable peut être perçue comme l'estimation

1. d'un quantile d'ordre α , i.e. $P(X < x_e) = \alpha \in]0, 1[$;
2. de l'espérance $E[X]$;
3. du mode $M_d[X]$.

³Excepté le modèle exponentiel, sur lequel il peut s'exprimer directement en termes de taux de défaillance.

Lorsque $\alpha \simeq 0.5$, ces trois grandeurs ne sont généralement, dans l'esprit d'un expert, que le reflet d'une même idée de X (cf. Daneshkhah 2004). De façon inconsciente, l'expert aura souvent tendance à donner une appréciation qui se situe au milieu d'un ensemble de valeurs reliées à la durée de vie (par exemple lorsqu'il connaît plusieurs exemplaires identiques ou non de \sum). Ainsi, il fournira plutôt une valeur médiane qu'une valeur moyenne, même si cette dernière est, de son point de vue, plus intuitive. Parmi les auteurs ayant remarqué de semblables comportements dans l'expertise industrielle, financière, démographique, etc., citons Schieren (1993) ou Lannoy & Procaccia (2001, p.23). Ces spécifications aboutissent cependant à des modélisations différentes. Elles doivent donc être précisées au maximum ; un questionnaire en ce sens serait par exemple (pour les spécifications 1 à 3, respectivement) :

1. lorsque l'âge d'un groupe de composants \sum_1, \dots, \sum_P atteint x_e , quelle proportion $1 - \alpha$ des composants est-elle encore en fonctionnement ?
2. la valeur x_e se fonde-t-elle sur une compilation d'évènements passés ?
3. cette valeur correspond-elle plutôt à la valeur de temps de défaillance la plus souvent observée dans le passé ?

Remarquons cependant que la pertinence de ces questions est relative à un certain contexte d'étude : la question 1 est plutôt *durabiliste*, puisque y répondre peut induire la connaissance d'une stratégie de maintenance. La question 3 est clairement *fiabiliste*, puisqu'elle recourt à une vision passée de la durée de vie de composants. La question 2 est mixte, car on ne peut réellement la poser que lorsque le composant est neuf ou lorsqu'il a déjà atteint un certain âge.

REMARQUE 7. *La spécification d'un mode apparaît en pratique comme peu réaliste, en tout cas dans certains domaines spécialisés qui ont servi de cadre à cette étude (l'ingénierie nucléaire). Nous l'avons cependant incluse pour des raisons de généralité.*

Modélisations. Pour chaque spécification, nous choisissons l'expression déterministe de b (symbolisée par sa distribution B) indiquée dans la proposition 6. Pour cela l'expression analytique des spécifications est rendue indépendante des hyperparamètres de $\pi(\beta)$. Supposons par exemple avoir $E[X] = x_e$. Supposons alors connaître $\beta \sim \pi_1$ et $\beta \sim \pi_2$ deux modélisations possibles. Impérativement, nous devons choisir b tel que $E[X|B, \pi_1] = E[X|B, \pi_2] = x_e$.

Dans la sous-partie suivante, la taille fictive a est fixée par l'analyste bayésien. Si l'on décidait de calibrer π uniquement en fonction de l'expertise, à l'instar de Dawid (1982), le couple (a, x_e) minimiserait une fonction de perte entre M et une représentation "idéale" de M (telle que la perçoit l'expert, par exemple par le biais d'un histogramme (van Noortwijk *et al.* 1992, Daneshkhah *et al.* 2006)). Le sens de a serait alors moins clair. Par ailleurs, cette représentation se heurterait aux propriétés structurelles de π : le risque serait élevé d'obtenir des distributions "aux vices cachés" (par exemple très dissymétriques).

EXEMPLE 4. Choisissons un *a priori* conjugué $\mathcal{G}(a, b)$ pour le paramètre λ d'un modèle exponentiel. On suppose que l'expert, interrogé sur le taux de défaillance, indique les estimations $(0.45, 0.5)$ du

mode $\text{Mod}[\lambda] = (a - 1)/b$ et de l'espérance $E[\lambda] = a/b$. Ce faisant, en acceptant sans restriction l'avis de l'expert, on obtient $a = 20$. En percevant a comme une taille fictive, cet expert peut clairement être considéré comme trop informatif. Il est vraisemblable que son idée du mode et de la moyenne est assez confuse, puisqu'il rapproche exagérément ces deux valeurs. Si l'on ne désire pas modifier a mais l'une des deux spécifications, laquelle choisir? ■

PROPOSITION 6. Soit $x_e \in \mathbb{R}^{+*}$, $a > 0$ et $\alpha \in]0, 1]$. On note

$$\begin{aligned} b_1(a, \beta) &= \left((1 - \alpha)^{-1/a} - 1 \right)^{-1} x_e^\beta, \\ b_2(a, \beta) &= \left(\frac{\Gamma(a)}{\Gamma(1 + 1/\beta)\Gamma(a - 1/\beta)} \right)^\beta x_e^\beta, \\ b_3(a, \beta) &= \frac{a\beta + 1}{\beta - 1} x_e^\beta. \end{aligned}$$

On rappelle que B désigne la modélisation induite par le choix de $b(a, \beta)$. Alors, pour tout choix de $\pi(\beta)$ de support $[\beta_l, \beta_r] \subset]0, +\infty[$,

$$\begin{aligned} (i) \quad P(X \leq x_e | B_1) &= \alpha, \\ (ii) \quad E[X | B_2] &= x_e \quad \text{si } a > \beta_l^{-1}, \\ (iii) \quad M_d[X | B_3] &= x_e \quad \text{si } \beta_l > 1. \end{aligned}$$

Dans le cas (iii), le mode x_e de la distribution marginale M est unique et strictement positif. Dans les autres cas, si $\beta_l \geq 1$, M possède au moins un mode strictement positif.

Preuve. Voir Annexes § 3.6.1 (p. 99).

Remarques et propriétés.

1. Dans le cas (i), on a $P(\eta < x_e) = P(U > k_{\alpha, a})$ où $U \sim \chi_{2a}^2$ et $k_{\alpha, a} = 2b_1(\beta)x_e^{-\beta}$: si l'on sait spécifier un quantile sur X , on peut spécifier un quantile sur η (ce qui permet la comparaison avec d'autres modélisations). Par ailleurs, en notant $\ell_{\alpha, a} = k_{\alpha, a}/\chi_{2a}^2(1 - \alpha)$, on a $P(\eta < x_e \ell_{\alpha, a}^{1/\beta} | \beta) = \alpha$, résultat utile lorsque β est fixé ou faiblement variable. Enfin, si $a \gg 1$, si x_e est le quantile 50% (la médiane), $b_1(\beta) \simeq ax_e^\beta / \log 2$. On retrouve ce résultat lorsqu'on adapte la démarche de Singpurwalla & Song (1986) en supposant que l'expert fournit une valeur moyenne de la médiane $m = \eta(\log 2)^{1/\beta}$, avec $m|\beta \sim \mathcal{GIG}(a, \beta, b \log 2)$.
2. Dans le cas (ii), si $\beta_l \geq 1$, l'espérance de η reste proche de x_e : on a $E[\eta]/x_e \in [1, 2/\sqrt{\pi}] \simeq [1, 1.1284]$. En cas de fort vieillissement, si $a > 4$, $b_2(\beta) \simeq ax_e^\beta$ d'après la proposition 9 (Annexes § 9).
3. Dans le cas (iii), la condition $\beta_l > 1$ n'est pas surprenante : une distribution de Weibull est de mode strictement positif dès que le vieillissement est assuré, soit dès que $\beta > 1$. En cas de fort vieillissement, on retrouve $b_3(\beta) \simeq ax_e^\beta$.
4. Dans le cas d'un modèle exponentiel, β est fixé à 1. Avec $\eta \sim \mathcal{IG}(a, b)$, on retrouve, si on fixe $E[X | B_2] = x_e$, $E[\eta] = x_e$. Cependant, il apparaît plus naturel de proposer $b = ax_e$; dans ce cas, x_e n'est plus perçue comme l'espérance de la distribution marginale M mais comme la moyenne de l'échantillon fictif auquel on assimile l'expert. Outre le fait que cette modélisation n'impose

aucune restriction sur a , une possible discussion avec l'expert permettrait d'entériner ce choix plutôt que le choix (ii).

3.3.3.2 Expertise quantitative développée

Par ailleurs, l'interrogation d'un expert peut être orientée de façon à proposer en sus un (ou plusieurs) intervalle de *crédibilité* $[x_{e_1}, x_{e_2}]$ d'ordre $(\alpha_1, \alpha_2) \in]0, 1[$ tel que pour $i = 1, 2$, $P(X < x_{e_i}) = \alpha_i$ (nécessairement, $x_{e_1} < x_{e_2} \Rightarrow \alpha_1 < \alpha_2$). Un cas plutôt fréquent en durabilité est l'expertise de l'intervalle (50%, 90%). Par exemple, la méthode de la *bisection* (Garthwaite *et al.* 2005) propose de considérer une suite de questions du type

1. Pouvez-vous déterminer une valeur médiane $x_{e,0}$ pour X ?
2. Supposez que $X < x_{e_0}$. Pouvez-vous déterminer une nouvelle valeur $x_{e_{-1}} < x_{e_0}$ telle que $P(X < x_{e_{-1}}) = 1/2$? ($x_{e_{-1}}$ est alors perçu comme le quartile inférieur de M).
3. Même question en supposant que $X > x_{e_0}$.

Une autre méthode est de fixer un ensemble d'intervalles sur $\text{Dom}(M)$ et de demander à l'expert la probabilité de défaillance de \sum dans chaque intervalle. Enfin, dans un contexte de fiabilité industrielle, ces quantités s'apparentent souvent aux temps de défaillance (ou de censure) les plus courts et les plus longs jamais observés par l'expert dans le passé. Ainsi, (x_{e_1}, x_{e_2}) prennent le sens de quantiles extrêmes (typiquement, 5%-95%). Si un expert peut ainsi fournir plusieurs spécifications $i = 1, \dots, P$, nous proposons de modéliser son expertise en combinant les modélisations π_1, \dots, π_P émanant de chacune des spécifications qu'il fournit, considérée indépendamment, de la façon suivante :

$$\mu|\beta \sim \mathcal{G} \left(\sum_{i=1}^P P\omega_i a_i, \sum_{i=1}^P b_i(P\omega_i a_i, \beta) \right), \quad (3.7)$$

soit la distribution émanant d'inférences bayésiennes successives sur les échantillons fictifs (de taille pondérées), à partir de la mesure non informative de Jeffreys $\pi^J(\mu|\beta)$. Les $\omega_i > 0$ sont des poids que l'on peut fixer en fonction de la confiance relative que l'on accorde aux spécifications, avec $\sum_{i=1}^P \omega_i = 1$.

Adopter cette démarche de séparation de l'expertise en "expertises atomiques" permet de construire π en conservant les hyperparamètres a_i dédiés à la calibration. Par ailleurs, elle permet d'additionner deux spécifications différentes (mode et quantiles par exemple).

3.3.4 Calibration de l'expertise quantitative

Nous proposons ici un ensemble de règles simples pour fixer la taille fictive a d'un expert unique e , selon la spécification qu'il propose. Ces recettes sont établies à partir de résultats très généraux d'analyse psychologique des experts (voir Daneshkhah 2004 pour une revue) et de l'étude de *prior feedbacks* de la littérature subjectiviste, et nous ne garantissons certainement pas leur validité pour tous les problèmes de fiabilité industrielle, même si nous les considérons appropriées dans le cadre d'étude proposé par EDF.

Par ailleurs, il faut considérer ces règles de calibration comme des approches *par défaut*. Rien ne saurait remplacer, pour approcher a , une discussion entre l'expert et l'analyste bayésien, d'autant

plus parce que ce paramètre est aisément compréhensible par les deux parties. Des questions ouvertes sont régulièrement proposées pour diriger cette discussion, mais sans que nous puissions certifier que l'expert puisse y répondre. D'où l'importance de donner des règles objectives sur lesquelles s'appuyer, préalablement aux résultats de la discussion.

Notons que dans cette étude, nous ne cherchons pas à modifier les valeurs x_e proposées par l'expert, utilisées dans la proposition 6. De notre point de vue, ces valeurs reflètent l'opinion personnelle de l'expert, et nous ne disposons pas d'information objective pour nous permettre de modifier cette opinion⁴. Des auteurs comme Singpurwalla & Song (1986) construisent des modélisations *a priori* aux nombreux hyperparamètres, permettant de modérer l'information d'expertise à la fois en position et en incertitude. Mais dans les faits - en tout cas dans le contexte d'EDF, comme le font remarquer Bacha *et al.* (1998) - calibrer finement l'*a priori* reste inaccessible. Nous considérons donc comme bénéfique de n'avoir qu'un hyperparamètre unique et compréhensible à calibrer pour moduler l'incertitude de l'expert.

Spécification espérance/médiane/mode. Lorsque l'expert spécifie x_e comme une valeur centrale de M , il s'appuie nécessairement sur un certain nombre de réalisations passées de \sum ou de systèmes proches ou en interaction avec \sum . On peut alors choisir a comme le nombre minimal de réalisations nécessaire à cette spécification, et le modifier *via* des questions subsidiaires, portant sur l'expérience passée de l'expert.

Espérance/médiane. Pour définir une moyenne (soit l'estimateur sans biais d'une espérance) ou une médiane, on peut légitimement supposer que l'expert utilise au moins la connaissance de $a = 1$ ou $a = 2$ données passées (respectivement). Supposons que l'expert soit capable d'indiquer un nombre de défaillances passées sur lesquelles il fonde son avis. S'il répond positivement à la question "avez-vous déjà observé des défaillances éloignées de x_e et les incluez-vous dans votre estimation?", il est plus vraisemblable que x_e représente une médiane. Sinon, le a correspondant risque d'être considérablement élevé. Dans ce cas, la question "avez-vous observé une forte concentration de défaillances dans un faible intervalle de temps?" permet de corriger a et x_e .

Mode. Intuitivement, un expert capable de spécifier un mode connaît un nombre plutôt élevé de défaillances passées, dont au moins 2 égales ou proches. Au minimum, il faut $a = 3$ données pour définir la classe modale d'une distribution continue.

Nous représentons sur la figure 3.13 les densités $m(x)$ produites selon chacune de ces spécifications. On choisit ici $x_e = 100$, $[\beta_l, \beta_r] = [1.1, 5]$, $(p, q) = (1.5, 1.5)$ et $a = 2$. En faisant varier a , on indique sur la table 3.1 l'évolution de l'écart-type et la dissymétrie (*skewness*) de M . Plus cette dernière valeur est forte positivement, plus la distribution M privilégie les valeurs les plus hautes de X .

En cas d'hésitation entre les spécifications, celle du mode assure une plus large incertitude et peut constituer un choix par défaut. Cependant, ce choix donne à M la queue la plus lourde quand $x \rightarrow \infty$. La spécification du mode privilégie donc les valeurs de X les plus hautes. Comme nous le précisons au § 3.3.6, cette spécification témoigne d'une connaissance élevée des conditions d'exploitation de \sum

⁴Au contraire, elle nous apparaît comme une richesse de la modélisation bayésienne subjective.

et de la conviction du vieillissement, et nous ne conseillons son utilisation que lorsque la connaissance *a priori* est validée par un large consensus ou des raisons objectives. Ainsi, nous conseillons plutôt de choisir la médiane par défaut, qui assure une plus large incertitude que l'espérance.

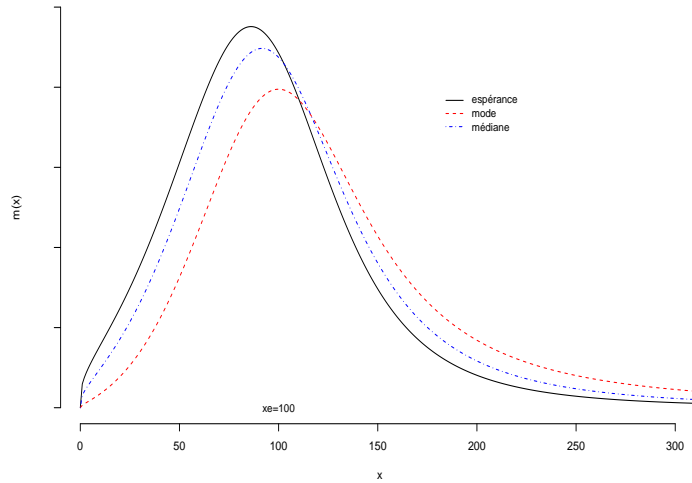


FIG. 3.13 – Densités $m(x)$ de la distribution marginale M

a	espérance		mode		médiane	
	σ	γ	σ	γ	σ	γ
2	67.4	4.5	360.7	31.3	94.2	5.5
3	53.8	2.4	185.0	12.5	66.0	2.7
4	50.0	1.91	151.6	9.8	58.2	1.94
5	48.1	1.45	136.7	7.5	55.8	1.46
10	44.0	1.26	110.7	6.3	50.3	1.13

TAB. 3.1 – Estimations de l'écart-type σ et de la dissymétrie γ de M .

Spécification d'un quantile. Nous faisons l'*hypothèse* suivante : la spécification d'un couple (x_e, α) tel que $P(X < x_e) = \alpha$ induit que l'expert, de façon sous-jacente, connaît autant de réalisations passées qu'il est nécessaire pour définir la précision α ; si $\alpha = 0.5$ (médiane), il a perçu un couple $(x_{e,l}, x_{e,r})$ tel que $x_{e,l} < x_e < x_{e,r}$. Soit $a = 2$. Sous une règle tacite d'*équirépartition* des données fictives, il est alors aisé de dérouler une méthode similaire à la *bisection* introduite plus haut pour évaluer la taille a correspondante. Ainsi, si $\alpha = 10\%$, l'expert est censé avoir perçu $a = 10$ données.

Cette règle reste bien entendu une procédure par défaut. Si l'analyste bayésien a la possibilité de converser avec l'expert (ce qui est généralement le cas), la valeur de a que nous proposons a surtout pour but de faire réagir l'expert, et notamment de lui faire percevoir sa "témérité" lorsqu'il propose un seuil de quantile trop précis. Sous forme d'enchères, nous pensons qu'un consensus est en général atteignable pour fixer une valeur a . La discussion peut être également orientée par des questions du type "le quantile que vous proposez correspond-t-il à une proportion de composants \sum ou proches de \sum tombés en panne dans le passé?".

Nous représentons sur la figure 3.14 les densités $m(x)$ produites selon la spécification de $x_e = 100$ comme quantile de seuil α , pour différentes valeurs de α . On choisit ici $[\beta_l, \beta_r] = [1.1, 5]$, $(p, q) = (1.5, 1.5)$ et $a = 10$.

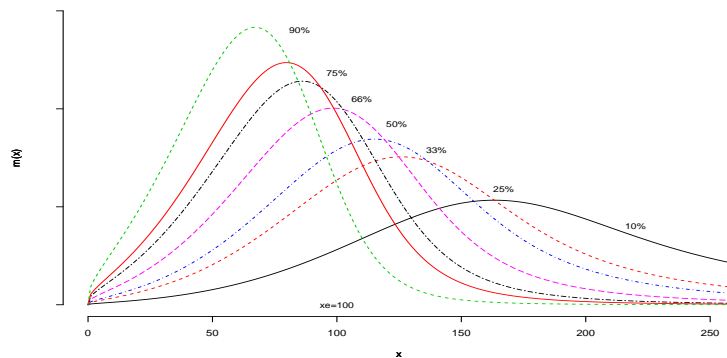


FIG. 3.14 – Densités $m(x)$ de la distribution marginale M (selon α tel que $P(X < x_e) = \alpha$).

Spécification d'un intervalle de crédibilité. Comme indiqué au § 3.3.3.2, nous spécifions séparément la taille équivalente a_i , $i = 1, 2$ de chaque quantile du couple, et nous recombinaisons les spécifications en adoptant la démarche (3.7). Cette recombinaison nécessite cependant une renormalisation des a_i que nous illustrons dans l'exemple suivant.

EXEMPLE 5. Soit $(x_{e,1}, x_{e,2})$ l'intervalle (50%, 95%). Nous supposons que l'expert nécessite $a_1 = 2$ données pour fixer $x_{e,1}$ indépendamment et $a_2 = 20$ données pour fixer $x_{e,2}$ indépendamment. Pour ne pas surestimer la connaissance de l'expert, on doit avoir $2\omega_1 a_1 + 20\omega_2 a_1 \leq 20$. Par défaut, on choisit évidemment $\omega_1 = \omega_2 = 1/2$, soit $a_1 \leq 20/11$ et $a_2 \leq 200/11$. On choisit alors $a_1 = \lfloor 20/11 \rfloor = 1$ et $a_2 = \lfloor 200/11 \rfloor = 18$. ■

Là encore, *via* une interrogation poussée de l'expert et une connaissance technique (objective) de \sum , on peut tenter de préciser quel quantile est le plus digne de confiance. C'est particulièrement

conseillé lorsque les seuils sont extrêmes. Un quantile spécifié à 5% peut correspondre à la date de dernière défaillance de jeunesse (déverminage) de Σ et un quantile à 95% à une date d'arrêt de Σ avant défaillance (soit une censure). Dans l'exemple précédent, imaginons qu'on ait deux fois plus confiance en la spécification de $x_{e,1}$ que celle de $x_{e,2}$. Alors $\omega_1 = 2/3$ et $a_1 = 2$. Incidemment, on définit ω_1 comme la probabilité que l'expert soit plus juste lorsqu'il spécifie l'ordre de $x_{1,e}$ que celui de $x_{2,e}$.

Il est également primordial de vérifier que la spécification de l'intervalle $[x_{e,1}, x_{e,2}]$ est *cohérente* avec le modèle bayésien. On utilise pour cela la formule (3.5). Soit

$$\beta_e = \log \left\{ \frac{\log(1 - \alpha_2)}{\log(1 - \alpha_1)} \right\} \log^{-1}(x_{e,2}/x_{e,1}).$$

Si $\beta_e \notin [\beta_l, \beta_r]$, expertises observable et prédictive sont conflictuelles. Une solution est évidemment d'élargir modérément $[\beta_l, \beta_r]$. Cependant, si $\beta_e \gg \beta_r$, le modèle de Weibull de paramètre de forme β_e est proche d'une loi normale et l'expert est clairement trop informatif. Il est alors nécessaire de pondérer la crédibilité (α_1, α_2) jusqu'à obtenir $\beta_e = \beta_0 \in [\beta_l, \beta_r]$ (par exemple la moyenne). L'algorithme suivant, dérivé de Newton-Raphson, permet de réaliser formellement cette pondération avec le taux de convergence constant

$$\frac{|\alpha_1^{k+1} - \alpha_1^k|}{|\alpha_2^{k+1} - \alpha_2^k|} = \frac{\omega_1}{1 - \omega_1}$$

à l'étape $k \rightarrow k + 1$. Ainsi, si $\omega_1 \rightarrow 1$, α_1 évolue peu au cours de l'algorithme. La justification de la construction est donnée en Annexes § 3.6.1 (p. 100).

Algorithme 1. Pondération de la crédibilité.

1. Soit $0 < \alpha_1^0 < \alpha_2^0 < 1$ et $\alpha_0 = (\alpha_1^0, \alpha_2^0)$. On note $\ell_0 = (x_{e,2}/x_{e,1})^{\beta_0}$, $\ell(\alpha) = \frac{\log(1 - \alpha_2)}{\log(1 - \alpha_1)}$, et $\delta_{-1} = \varepsilon + (1, 1)$ où ε est le choix d'une précision. On fixe $0 < \rho \ll 1$ (typiquement, $\rho \sim 100\varepsilon$).
2. Étape $k = 0, \dots, K$: tant que $\|\delta_{k-1}\| > \varepsilon$,
 - calculer $\delta_k = \begin{pmatrix} (1 - \alpha_1^k) \log(1 - \alpha_1^k) (\ell(\alpha_k) - \ell_0) / \ell(\alpha_k) \\ -(1 - \alpha_2^k) \log(1 - \alpha_2^k) (\ell(\alpha_k) - \ell_0) \end{pmatrix}$;
 - calculer $h_k = -\frac{(1 - \alpha_1^k)}{(1 - \alpha_2^k) \ell(\alpha_k)}$.
 - calculer $\alpha^{k+1} = \begin{cases} \alpha_1^k - \rho h_k (\alpha_2^k - \alpha_1^k), \\ \alpha_2^k - \rho h_k (\alpha_2^k - \alpha_1^k) \frac{\omega_1}{1 - \omega_1} \end{cases}$

EXEMPLE 6. Fixons $\beta_0 = 3$ et $(x_{e,1}, x_{e,2}) = (200, 300)$. En posant $(\alpha_1^0, \alpha_2^0) = (0.05, 0.95)$, on a alors $\beta_e \simeq 10.03$. L'expertise induit donc un vieillissement irréaliste. On fixe $\omega = (1/2, 1/2)$. On trouve alors $\alpha = (0.3, 0.7)$. Si l'on pose maintenant $\alpha_1^0 = 0.25$, on trouve $\beta_e \simeq 5.78$. Le vieillissement induit est encore trop fort. Avec des poids équilibrés, on trouve $\alpha = (0.4, 0.8)$. Si l'on a une forte confiance dans la crédibilité de la borne inférieure, on fixe $\omega = (0.95, 0.05)$. Alors $\alpha = (0.26, 0.64)$. Inversement, avec $\omega = (0.05, 0.95)$, on obtient $\alpha = (0.55, 0.93)$. ■

Une méthode générale d'histogramme. D'une façon générale, autour d'une valeur x_e qu'il fournit, l'expert peut avoir une opinion de la dissymétrie de la distribution de X , dont la répartition peut être appréciée par une méthode d'histogramme (qui n'est qu'un succédané de la méthode de la bisection). Van Noortwijk *et al.* (1992) proposent de segmenter \mathbb{R}^+ en intervalles temporels disjoints, en choisissant les frontières en des valeurs observées réellement (de censure ou non). Sur chaque intervalle, l'expert indique sa probabilité de défaillance, par exemple sous forme de fréquence ou de pari. La répartition et le nombre des données fictives peuvent alors être proposés en fonction de la précision atteinte.

EXEMPLE 7. Soit $(50, 80, 170)$ trois valeurs de défaillance observées et 90 une valeur de censure à droite. On sépare l'espace des temps de défaillance en $D_1 = [0, 90[$ et $D_2 = [90, \infty[$. Soient (n_1, n_2) les nombres de données fictives respectivement dans D_1 et D_2 . Supposons que l'expert indique 66% de chance de tomber en panne dans la période D_1 . On a donc $n_1 = 2n_2$. Séparons D_1 en $D_{1,a} = [0, 50[$ et $D_{1,b} = [50, 90[$.

Indiquons à l'expert que la durée de vie effective est comprise dans D_1 . Supposons alors qu'il indique 25% de chance de tomber en panne dans $D_{1,b}$. À une question similaire sur une partition de D_2 , il est incapable de se prononcer. Si l'on suppose que l'expert ne peut plus répondre à une question portant sur une autre subdivision dans D_1 , on fait l'hypothèse qu'il a "vu" $n_1 = 4$ données dans D_1 . Au final, il a donc perçu $n_1 + n_2 = 6$ données fictives. ■

3.3.5 Recalibration finale de l'expertise

L'étape de recalibration finale correspond au travail critique de l'analyste bayésien, indépendamment de tout contact avec le ou les experts. Une fois construites les modélisations des expertises disponibles, il est nécessaire de les combiner et de borner des tailles a_1, \dots, a_M éventuellement trop grandes vis-à-vis de la connaissance objective apportée par les données \mathbf{y}_n . Nous donnons en ce sens quelques conseils. Enfin, lorsque l'analyste décide d'utiliser le modèle à risques compétitif $\mathcal{B}(\eta_0, \eta_1, \beta)$, nous discutons du problème d'équilibrage de l'information *a priori* fournie aux sous-modèles entrant en concurrence.

3.3.5.1 Consensus entre experts

Lorsque plusieurs expertises e_1, \dots, e_M sont disponibles, deux modélisations peuvent être utilisées. La *combinaison convexe*

$$\pi(\theta) = \frac{\prod_{i=1}^M \pi_i^{\omega'_i}(\theta)}{\int_{\Theta} \prod_{i=1}^M \pi_i^{\omega'_i}(\theta) d\theta},$$

où $\sum_{i=1}^M \omega'_i = 1$, minimise la perte d'information de Kullback-Leibler

$$KL(\pi; \pi_1, \dots, \pi_M | \beta) = \sum_{i=1}^M \omega'_i \int_{\mathbb{R}} \pi(\theta) \log \frac{\pi(\theta)}{\pi_i(\theta)} d\theta$$

et réalise un *consensus* optimal des différentes expertises. Les poids ω'_i représentent l'importance relative des experts. Cooke *et al.* (1988), Budescu & Rantilla (2000), Lannoy & Procaccia (2001)

proposent notamment un ensemble de critères tels que l'expérience ou le pourcentage d'erreur passé pour juger des ω'_i . La méthode Delphi est un exemple célèbre de méthodologie en ce sens. Liisberg (1991) ou Alturazza *et al.* (2004), entre autres, font ainsi de ce consensus la représentation idéale d'un groupe d'experts. En l'appliquant à $\pi(\theta) = \pi(\mu|\beta)\pi(\beta)$, on obtient l'appréciable stabilité

$$\begin{aligned}\mu|\beta &\sim \mathcal{G}\left(\sum_{i=1}^M \omega'_i a_i, \sum_{i=1}^M \omega'_i b_i(a_i, \beta)\right), \\ \beta &\sim \mathcal{B}_e\left(\sum_{i=1}^M \omega'_i p_i, \sum_{i=1}^M \omega'_i q_i\right)\end{aligned}$$

lorsque $\pi_i(\beta)$ est une loi Bêta $\mathcal{B}_e(p_i, q_i)$ définie sur un domaine $[\beta_l, \beta_r]$ commun.

Ce type de modélisation n'est cependant pas, pour O'Hagan (2003, 2005), le meilleur possible. En effet, elle nécessite que les experts soient indépendants entre eux, ce qui est difficile à vérifier. Dans l'industrie, seul le cas de deux experts (fabricant et exploitant⁵ abonde généralement en ce sens. Au-delà de deux experts, l'hypothèse d'indépendance est peu crédible. O'Hagan préfère généralement définir un consensus par le biais de discussions entre experts, et ensuite procéder à la modélisation. Dans notre cas, cette méthode implique de construire un échantillon fictif "consensuel" mais permet de retomber sur le même type de modélisation.

3.3.5.2 Limitations de taille.

Très rapidement, selon les règles présentées auparavant, on peut arriver à de grandes tailles a_1, \dots, a_M d'échantillon fictif (pour M expertises disponibles). Il nous paraît indispensable de borner supérieurement ces tailles en fonction des remarques suivantes. On note $a = \sum_{i=1}^M \omega'_i a_i$ la taille de l'échantillon fictif correspondant à la combinaison des experts.

1. On doit avoir $a < \tilde{n}$ où \tilde{n} est un indicateur correspondant à la taille de l'échantillon i.i.d. apportant la même quantité d'information que les données observées \mathbf{y}_n . Nous étudions au § 3.5 la construction d'un tel indicateur. Cette borne fait sens puisqu'elle rend les résultats *a posteriori* sur $\theta' \in \{\eta, \lambda, \mu\}$ majoritairement dépendants des données réelles. Les a_i doivent alors être renormalisés en

$$a'_i = a_i \frac{a}{\sum_{i=1}^M \omega'_i a_i}.$$

Là encore, utiliser la partie entière d'un tel résultat fait sens si l'on désire conserver des tailles entières. Cependant, certaines expertises risquent alors de disparaître ($a'_i = 0$). On préfère donc conserver $0 < a'_i < 1$ en notant que l'expert e_i apporte alors autant d'information qu'une fraction de donnée *i.i.d.* (soit l'équivalent d'une donnée censurée), ou en comprenant qu'il faut $p > 1$ experts e_1 pour apporter l'équivalent d'une donnée fictive i.i.d.

2. Une abondante littérature est consacrée à la vérification du seuil des quantiles. Beaucoup d'auteurs en fiabilité industrielle s'accordent sur le flou du sens réel des valeurs proposées par les experts. Face aux résultats expérimentaux, celles-ci sont créditées par les experts eux-mêmes de

⁵L'un est souvent optimiste et l'autre pessimiste.

seuils de confiance très souvent surévalués. Copen (1975), Meyer & Booker (1987), Lannoy & Procaccia (2001) parmi d'autres insistent sur le biais important de ces valeurs de seuil. Semble se dégager une règle tacite qui veut que l'on accorde aux experts les ordres maximaux suivants (voir aussi tableau 2.1, Lannoy & Procaccia 2001), qui fournissent des bornes supérieures a_i^* .

expertise	réalité	a_i^*
5%	25%	4
20%	33%	3
25%	40%	2
75%	60%	2
80%	66%	3
95%	75%	4

TAB. 3.2 – Correspondance tacite expertise-réalité entre ordres de quantiles.

3. Si jamais preuve est faite que x_{e_i} (formulée par l'expert e_i) est une valeur reflétant plus la connaissance de temps de censure plutôt que de véritables temps de défaillance, et en particulier en cas de quantile extrême, il est rassurant d'accorder à l'expert l'équivalent informatif d'une seule donnée ($a_i = 1$) voire celui d'une donnée censurée. De tels indicateurs sont là aussi définis au § 3.5. Il s'agit d'un choix par défaut lorsqu'aucune précision n'est disponible pour expliquer la spécification.
4. Une autre procédure de calibration, objective, est possible en utilisant le critère DAC défini et étudié au chapitre 4 : on peut fixer $a = a^*$ telle que $\text{DAC}(a^* | \mathbf{y}_n) = 1$. Ce faisant, si l'expertise est éloignée des données, on s'assure que π modélise une incertitude élevée, ce qui est un état de fait rassurant pour des problèmes industriels. Voir ce chapitre 4 pour des explications plus précises.

Notons enfin que le choix B_1 (cf. proposition 6) est tel que l'on puisse modéliser un quantile d'ordre élevé, par exemple 95%, tout en lui accordant une valeur de taille fictive bien moindre que prévue si l'on suppose que l'expert est correct (ici $a = 20$). On peut faire le même commentaire sur les deux autres modélisations. En résulte une grande flexibilité dans le travail de l'analyste bayésien.

3.3.5.3 Calibration du modèle $\mathcal{B}(\eta_0, \eta_1, \beta)$.

L'emploi du modèle à risques compétitif $\mathcal{B}(\eta_0, \eta_1, \beta)$ nécessite *a priori* que l'expert puisse répondre affirmativement à une question du type "le système \sum risque-t-il d'être soumis à des défaillances accidentelles, en sus du vieillissement?". L'analyste bayésien peut aussi utiliser les outils non paramétriques introduits au chapitre 2 pour appuyer ce choix. Notons par simplicité $\lambda = 1/\eta_0$, $\mu = (1/\eta_1)^\beta$ et $\lambda \sim \mathcal{G}(a', b')$.

On peut supposer qu'un très bon expert fournisse, après sollicitation, une durée de vie moyenne avant défaillance accidentelle, en sus des informations déjà utilisées pour la modélisation de Weibull. Nous devons cependant faire un choix par défaut si aucun expert ne peut répondre à cette question. La difficulté est donc de mettre en compétition $\pi(\lambda)$ et $\pi(\mu, \beta)$ sans favoriser arbitrairement l'un ou l'autre. Le choix des hyperparamètres doit rendre les modèles bayésiens équitables.

Plusieurs auteurs (Dawid & Lauritzen 2000, Leucari & Consonni 2003, Roverato & Consonni 2004, Consonni *et al.* 2004b) ont proposé des formalisations pour donner un sens clair à cette notion d'équitabilité (ou *compatibilité*). Marin (2006) définit $\pi(\lambda)$ et $\pi(\mu, \beta)$ comme compatibles si la distance de Kullback-Leibler entre les marginales

$$\begin{cases} m_W(x) &= \int \int f_W(x|\mu, \beta)\pi(\mu, \beta) d\mu d\beta \\ m_E(x) &= \int f_E(x|\lambda)\pi(\lambda) d\lambda \end{cases}$$

est minimale. En d'autres termes, sachant $\pi(\mu, \beta)$, on sélectionne les hyperparamètres $(a', b')^*$ de $\pi(\lambda)$ tels que

$$(a', b')^* = \arg \min \int_{\mathbb{R}} m_W(x) \log \frac{m_W(x)}{m_E(x)} dx \quad (3.8)$$

pour obtenir la densité $\pi(\lambda)$ la plus compatible. La proposition 7 indique que $(a', b')^*$ existe et est unique sous la condition que, pour $k = 1, 2$,

$$E[X^k|B] = \int_{\mathbb{R}} b^{k/\beta}(a, \beta) \frac{\Gamma(a - k/\beta)}{\Gamma(a)} \Gamma(1 + k/\beta) \pi(\beta) d\beta < \infty,$$

ce qui est vrai si $a > 2/\beta_l$ (pour tout choix de B).

PROPOSITION 7. *Soit $\pi(\beta)$ de support $[\beta_l, \beta_r] \subset]0, +\infty[$. Soit $\pi(\mu|\beta)$ telle que pour $k = 1, 2$, $E[X^k] < \infty$ où E est l'espérance par rapport à M_W . Alors il existe un unique couple $(a', b')^* > 0$ solution de (3.8).*

Preuve : voir Annexes § 3.6.1 (p. 101)

Malheureusement cette proposition induit des lourdeurs de calcul qui rendent la méthode peu pratique, notamment si l'on manie plusieurs *a priori* (par exemple lors d'une étude de sensibilité). Même imprécises, les estimations de (a', b') trouvées induisent en général des *a priori* dont les caractéristiques sont proches du choix

$$\begin{cases} a' &= a, \\ b' &= b(a', \beta = 1) \end{cases} \quad (3.9)$$

qui nous semble le plus cohérent avec la spécification : une même grandeur marginale, obtenue avec un échantillon fictif de même taille, est commune au deux modèles. Il nous paraît le plus satisfaisant car les hyperparamètres conservent un sens compréhensible.

REMARQUE 8. *Nous n'avons pas ici imposé la règle proposé par Bertholon et al. (2006), qui suppose que $\eta_0 > \eta_1$ afin de garantir que le vieillissement est la cause principale des défaillances. Le choix *a priori* devient alors*

$$\pi(\lambda, \mu, \beta) = \pi(\lambda|\mu, \beta)\pi(\mu|\beta)\pi(\beta)$$

avec

$$\begin{aligned} \lambda|\mu, \beta &\sim \mathcal{G}(a', b') \mathbb{1}_{\{\lambda \leq \mu^{1/\beta}\}}, \\ \mu|\beta &\sim \mathcal{G}(a, b(a, \beta)), \\ \beta &\sim \mathcal{B}_e(p, q) \end{aligned}$$

et la minimisation (3.8) nécessite un calcul très coûteux, là encore susceptible de modifier le sens des hyperparamètres. Sans avoir testé la proximité des résultats, nous suggérons simplement de borner la distribution $\mathcal{G}(a', b')$ choisie selon (3.9).

3.3.6 Caractéristiques effectives de la modélisation

L'analyste bayésien doit enfin s'assurer que la construction *a priori* respecte certaines contraintes de cohérence avec les propriétés du modèle de Weibull et le contexte industriel dans lequel l'expertise est faite. Nous indiquons ici deux caractéristiques effectives que celui-ci doit chercher à vérifier.

3.3.6.1 Unimodalité de la distribution $M(X)$

L'unimodalité de M en une valeur strictement positive est un aspect non négligeable de la modélisation. En effet, M doit rester proche de la distribution de Weibull de mode strictement positif (si non exponentielle ou modélisant un rajeunissement). Par ailleurs, l'expert propose une vision de la durée de vie d'un composant \sum , soumis principalement à du vieillissement. En l'absence d'autres sources de défaillance non accidentelles, cette durée de vie possède légitimement un pic unique en une valeur de temps non nulle où la probabilité de défaillance est maximale (cf. figures 3.13 et 3.14). Vérifier cette unimodalité participe de la bonne représentativité du modèle bayésien dans son ensemble.

Dans les cas (i) et (ii) de la proposition 6, nous avons constaté l'unimodalité de M en une valeur strictement positive sur tous les exemples (où β n'est pas connu et fixé à 1) via le test DIP d'Hartigan & Hartigan (1985a). On note, pour $p \gg 1$,

$$\text{Dip}(M_p|B) = \inf_{V \in \mathcal{V}} \sup_x |M_p(x|B) - V|$$

où $M_p(x)$ est la distribution empirique $(1/p) \sum_{i=1}^p \mathbb{1}_{\{X_i \leq x\}}$ et \mathcal{V} est la classe des distributions unimodales (numériquement, des distributions uniformes ou nulles sont utilisées comme références). Si M est unimodale et sous certaines contraintes de régularité, $\sqrt{p} \text{Dip}(M_p|B) \rightarrow 0$ en probabilité. On peut ainsi construire un test de puissance élevée (Hartigan 1985b), qui indique l'unimodalité si la statistique est proche de 0. De nombreux outils informatiques ont par la suite été développés (fournissant par exemple un encadrement du mode). Dans le tableau 3.3, nous indiquons à titre d'illustration l'estimation de $\text{Dip}(M_p|B)$ pour quelques scénarios *a priori*. On utilise 200 valeurs de simulations. Or, sachant p , sous l'hypothèse que M soit unimodale, $P(\text{Dip} < 0.0496) = 0.999$. En constatant que nos valeurs sont toujours largement inférieures à 0.0496, on conclut donc que M n'est pas multimodale (elle est donc unimodale en une valeur strictement positive d'après la proposition 6).

(p, q)	B_1			B_2
	$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$	
(1,1)	0.0156	0.0155	0.0159	0.0192
(1,5)	0.0151	0.0165	0.0152	0.0180
(5,1)	0.0133	0.0221	0.0209	0.0161
(5,5)	0.0190	0.0220	0.0158	0.0171

TAB. 3.3 – Estimations de la statistique de test d'unimodalité $\text{Dip}(M_p|B)$. On choisit *a priori* $\beta \sim \mathcal{B}(p, q)$ sur $[1.1, 5]$, $x_e = 100$ et $a = 1$.

3.3.6.2 Corrélation de (η, β)

La paramétrisation $\theta = (\eta, \beta)$ est la plus aisément compréhensible par l'analyste bayésien. Une modélisation *cohérente* avec le modèle paramétrique $\mathcal{W}(\theta)$ et une connaissance *a priori* fondée sur des données anciennes ou fictives imposerait que la corrélation (et donc que la covariance) du couple (η, β) est négative. En effet, les estimations fréquentistes de η et β évoluent généralement de façon antagoniste (voir le chapitre 2 et Wu 2002). Une aire de crédibilité cohérente avec $\mathcal{W}(\theta)$ prend donc, idéalement, une forme proche de celle de la figure 3.15. La table 3.4 présente les estimations du coefficient de corrélation pour quelques scénarios *a priori*.

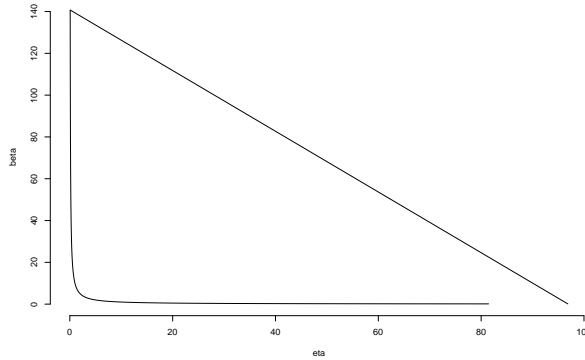


FIG. 3.15 – Enveloppe convexe d'un tirage (η, β) de corrélation négative.

a	espérance (B_2)	mode (B_3)	quantile (B_1)						
			10%	25%	33%	50%	66%	75%	90%
2	-0.007	-0.34	-0.35	-0.26	-0.22	-0.17	-0.08	0.01	0.12
5	-0.015	-0.48	-0.64	-0.59	-0.54	-0.33	-0.02	0.11	0.24
10	-0.022	-0.54	-0.72	-0.70	-0.65	-0.42	0.03	0.28	0.42

TAB. 3.4 – Valeurs de la corrélation $\text{Cov}(\eta, \beta)/(\sigma_\eta\sigma_\beta)$, pour $\beta \sim \mathcal{U}_{[1,5]}$, selon chaque choix de modélisation. Les valeurs négatives indiquent les choix *a priori* les plus cohérent avec le modèle paramétrique de Weibull.

Quand on augmente l'information sur β , ces valeurs de corrélation diminuent rapidement. Expérimentalement, les distributions *a priori* convergent vers des lois normales décorréliées. Cet exemple permet de vérifier que, dans un cadre d'étude plutôt commun en fiabilité, les modélisations de la proposition 6 respectent d'une façon générale cette règle de cohérence. En particulier, la spécification de quantiles d'ordre faible ($\leq 50\%$) et du mode induisent les modélisations les plus cohérentes.

3.4 Analyse *a posteriori*

L'utilisation des techniques bayésiennes permet d'améliorer la précision fréquentiste. L'ingénieur de sûreté industrielle est ainsi particulièrement intéressé par l'estimation de quantités telles que la durée de vie ou la survie. Ces outils lui permettent de prendre une décision de maintenance, de vérification ou de remplacement du composant Σ . La consistance et la convergence de la distribution *a posteriori* de densité $\pi(\theta|\mathbf{y}_n)$ sont des justifications importantes de l'acceptation des résultats inférentiels. Par

ailleurs, mesurer la stabilité de ces quantités vis-à-vis de petites modifications *a priori* permet à l'analyste bayésien de définir des modélisations par défaut, lorsqu'il hésite entre deux spécifications *a priori*.

Dans cette section, nous vérifions donc que $\pi(\theta|\mathbf{y}_n)$ adopte un comportement raisonnable lorsque n grandit. En second lieu, après avoir défini les principales quantités prédictives utiles au fiabiliste, nous proposons une étude (sur des données EDF) montrant que le gain bayésien en information, dans les conditions de calibration conseillées auparavant, permet d'atteindre une précision suffisante pour proposer une décision industrielle (non atteinte par les méthodes d'estimation classiques). Enfin, une étude de sensibilité sur des échantillons simulés vient clore la section et ce chapitre.

3.4.1 Consistance et convergence

Le choix de densités π subjectives requiert la vérification du bon comportement des distributions *a posteriori*. Nous vérifions donc pour Weibull et le modèle à risques compétitifs certaines propriétés de consistance et de convergence de ces distributions, qui autorisent l'emploi des densités π proposées auparavant. L'article de Ghosal (1999a) et le chapitre 1 du livre de Ramamoorthi & Ghosh (2003) sont ici nos principales références. Ce paragraphe ne constitue cependant qu'un rappel, qui participe d'une étude bayésienne complète, mais qui est généralement passé sous silence dans les articles appliqués.

3.4.1.1 Consistance

Une distribution *a posteriori* est consistante si, au fur et à mesure que l'on agrège des données, elle converge vers la mesure dégénérée δ_{θ_0} où θ_0 est la valeur réelle du paramètre. Clairement, il s'agit d'une condition *sine qua non* d'acceptation du résultat inférentiel (les estimateurs bayésiens étant donc garantis consistants).

DÉFINITION 4. Soit $\mathbf{X}_n = X_1, \dots, X_n \sim \mathcal{M}(\theta)$ et $\theta_0 \in \Theta$. La distribution *a posteriori* est dite consistante en θ_0 si pour tout voisinage U de θ_0 ,

$$\int_U \pi(\theta|\mathbf{X}_n) d\theta \xrightarrow[p.s.]{} 1.$$

La consistance *a posteriori* nécessite très peu voire aucune hypothèse sur π . Le théorème de Doob (cf. van der Vaart 1998, p. 149) indique que la suite des distributions *a posteriori* est presque partout consistante lorsque n augmente, en supposant simplement que le modèle $\mathcal{M}(\theta)$ est identifiable⁶ et que les espaces Θ et χ munis de leurs tribus boréliennes* respectives sont euclidiens. Voir le lemme 1 ci-dessous pour l'application aux modèles de Weibull et à risques compétitifs. Le théorème 1.3.4 de Ramamoorthi & Ghosh (2003) renforce cette propriété, en indiquant la consistance en un θ_0 particulier. Il faut supposer que le modèle $\mathcal{M}(\theta)$ respecte les conditions de Wald (1949) et que θ_0 est contenu dans le support de $\pi(\theta)$ (ainsi, on doit logiquement avoir $\beta_0 \in [\beta_l, \beta_r]$ dans notre cas). On obtient la consistance du MLE $\hat{\theta}_n$, soit, $\forall \varepsilon > 0$,

$$P_{\theta_0} \left(|\hat{\theta}_n - \theta_0| \geq \varepsilon \right) \xrightarrow{n \rightarrow \infty} 0.$$

Ces conditions sont vérifiés pour le modèle de Weibull. La consistance du MLE pour les modèles de Weibull et à risques compétitifs est prouvée respectivement dans Bacha (1996) et Bertholon *et al.*

⁶cf. Lexique p. 217.

(2006). Notons que lorsque n est petit, $\hat{\beta}_n$ menace d'être loin de β_0 . Ainsi, le fait d'avoir $\hat{\beta}_n \notin [\beta_l, \beta_r]$ n'est pas un gage d'inconsistance. Si n augmente, il faut cependant vérifier que la suite des $(\hat{\beta}_n)_n$ converge à l'intérieur de $[\beta_l, \beta_r]$.

LEMME 1. Pour $(\mu, \beta) \in \mathbb{R}^+ \times \mathbb{R}^+$, le modèle de Weibull est identifiable. Pour $(\lambda, \mu, \beta) \in \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+ / \{1\}$, le modèle $\mathcal{B}(\lambda, \mu, \beta)$ est identifiable.

Preuve. Soit f_W et f_B les densités respectives des deux modèles. Supposons que pour tout $x \geq 0$, $f_W(x|\mu, \beta) = f_W(x|\mu', \beta')$. Alors

$$\frac{\mu\beta}{\mu'\beta'} x^{\beta-\beta'} = \exp\left(-\mu x^{\beta'} + \mu x^{\beta}\right).$$

C'est en particulier vrai pour $x = 0$. Nécessairement, $\beta = \beta'$ puis $\mu = \mu'$. Supposons maintenant que pour tout $x \geq 0$, $f_B(x|\lambda, \mu, \beta) = f_B(x|\lambda', \mu', \beta')$. Alors

$$\frac{\lambda + \mu\beta x^{\beta-1}}{\lambda' + \mu'\beta' x^{\beta'-1}} = \exp\left(-(\lambda' - \lambda)x - \mu' x^{\beta'} + \mu x^{\beta}\right).$$

En $x = 0$, en supposant $\beta' \neq 1$, $\beta \neq 1$, on obtient $\lambda = \lambda'$, puis on revient au cas précédent. Nécessairement, $(\lambda, \mu, \beta) = (\lambda', \mu', \beta')$. En supposant $\beta' = 1$ et $\beta \neq 1$, on obtient $\lambda + \mu\beta x^{\beta-1} = \lambda \exp(\mu x^{\beta})$ ce qui entraîne $f_B(x|\lambda, \mu, \beta) = \lambda \exp(-\lambda x)$ (modèle exponentiel). Ceci est contradictoire avec $\mu < \infty$. Finalement, on a donc $\beta = \beta' = 1$. Le modèle est alors exponentiel et non-identifiable puisque $\lambda = \lambda'$ ou $\lambda = \mu'$ (resp. $\mu = \mu'$ ou $\mu = \lambda'$). \square

3.4.1.2 Convergence

Lorsque $n \rightarrow \infty$, et que le modèle bayésien est choisi suffisamment régulier, la distribution *a posteriori* peut souvent être approximée par une distribution normale. Le théorème de Bernstein-von Mises indique la convergence, au sens de la distance en variation totale, de la distribution *a posteriori* de la variable réduite $\sqrt{n}(\theta - \hat{\theta}_n)$ vers la distribution normale centrée de densité $N_{(0, I_{\theta_0})}(x)$, de covariance la matrice de Fisher I_{θ_0} . Plusieurs formulations de ce théorème sont disponibles (Ghoshal 1999a). Nous donnons en Annexes § 3.6.1 une vérification des conditions proposées par Ramamoorthi & Ghosh (2003, p. 35) dans le cas Weibull.

THÉORÈME 3. (Bernstein-von Mises). Soit $\mathbf{X}_n = X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathcal{W}(\theta)$ et $\hat{\theta}_n$ le MLE correspondant. Notons θ_0 la vraie valeur du paramètre. Soit π une densité a priori positive et continue en θ_0 . On suppose que $\text{Supp}(\pi(\beta)) = [\beta_l, \beta_r] \subset]0, +\infty[$, que $\beta_0 \in [\beta_l, \beta_r]$ et qu'il existe $p \in \mathbb{N}^*$ tel que $\forall n \geq p$, $\hat{\beta}_n \in [\beta_l, \beta_r]$. Soit $h = \sqrt{n}(\theta - \hat{\theta}_n)$. Alors, quand $n \rightarrow \infty$,

$$\int \left| \pi(s|\mathbf{X}_n) - N_{(0, I_{\theta_0})}(s) \right| ds \xrightarrow{\mathcal{W}(\theta_0)} 0.$$

Preuve. voir Annexes § 3.6.1 (p. 102).

Notons enfin que pour ce modèle, puisque $\int_{\Theta} |\theta| \pi(\theta) d\theta < \infty$ où $\theta = (\mu, \beta)$, on a, en notant

$$\theta_n^* = E[\theta | \mathbf{X}_n],$$

$$\begin{aligned} \sqrt{n}(\theta_n^* - \hat{\theta}_n) &\xrightarrow{\mathcal{W}(\theta_0)} 0, \\ \sqrt{n}(\theta_n^* - \theta_0) &\xrightarrow{\mathcal{L}} \mathcal{N}(0, I_{\theta_0}^{-1}). \end{aligned}$$

3.4.2 Gain d'information

Après avoir défini les principales quantités intéressant le fiabiliste industriel, nous proposons une mesure du *facteur d'erreur* ρ de chacune pour différents type d'estimation, à partir de données EDF. On rappelle que nous avons défini ce facteur d'erreur sur une quantité variable A par

$$\rho(A) = \sqrt{\max\left(\frac{q_{50\%}}{q_{5\%}}, \frac{q_{95\%}}{q_{50\%}}\right)}$$

où q_α est le quantile d'ordre α de la distribution de A , soit la valeur telle que

$$P(A < q_\alpha) = \alpha.$$

Si les méthodes classiques ne permettent guère d'atteindre la précision industrielle désirée ($\rho < 2.5$), les modélisations bayésiennes présentées dans ce chapitre, utilisant des tailles a typiques des procédures de calibration proposées, pallient cette déficience. L'appel aux techniques bayésiennes est donc fructueux.

3.4.2.1 Outils décisionnels

On rappelle quelques notations. Soit \mathbf{y}_n l'échantillon de données. On le suppose contenir un sous-échantillon \mathbf{x}_r de r données i.i.d. de Weibull et un ensemble \mathbf{c}_{n-r} de $n - r$ données de censure fixe (à droite). Soit B la modélisation retenue pour $b = b(\beta)$. *A posteriori*, afin de prendre une décision d'exploitation, les fiabilistes ou durabilistes cherchent à mesurer les quantités suivantes.

1. La *durée de vie moyenne* de \sum

$$E[X|B, \mathbf{y}_n] = \int_{\mathbb{R}} \Gamma(1 + 1/\beta) \left\{ b(a, \beta) + \sum_{i=1}^n y_i^\beta \right\}^{1/\beta} \frac{\Gamma(a + r - 1/\beta)}{\Gamma(a + r)} \pi(\beta|B, \mathbf{y}_n) d\beta$$

qui est définie si $a + r > \beta_l^{-1}$. En pratique cette condition est très peu restrictive, puisqu'en général $r \geq 4$ (pour les données EDF) et β_l reste proche de 1.

2. La *survie* au temps t_0

$$\begin{aligned} S(t_0|B, \mathbf{y}_n) &= P(X > t_0|B, \mathbf{y}_n), \\ &= \int_{\mathbb{R}} \left(1 + \frac{t_0^\beta}{\left\{ b(a, \beta) + \sum_{i=1}^n y_i^\beta \right\}} \right)^{-(a+r)} \pi(\beta|B, \mathbf{y}_n) d\beta. \end{aligned}$$

3. La *durée de vie résiduelle* au temps t_0 (Finkelstein 2006)

$$\begin{aligned} \text{MRTF}(t_0|B, \mathbf{y}_n) &= E[X - t_0 | X > t_0, B, \mathbf{y}_n], \\ &= \frac{1}{S(t_0|B, \mathbf{y}_n)} \int_{t_0}^{\infty} S(x|B, \mathbf{y}_n) dx. \end{aligned}$$

Toutes ces quantités sont donc facilement estimables par exemple par des techniques de Monte-Carlo si l'on peut obtenir un tirage $(\beta_1, \dots, \beta_M)$, où M est aussi grand que possible, qui suive asymptotiquement la densité $\pi(\beta|\mathbf{y}_n)$. En effet,

$$\pi(\beta|B, \mathbf{y}_n) \propto \beta^r \left(\prod_{j=1}^r x_j \right)^\beta (\beta - \beta_l)^{p-1} (\beta_r - \beta)^{q-1} b^a(a, \beta) \left\{ b(a, \beta) + \sum_{i=1}^n y_i^\beta \right\}^{-(a+r)} \mathbb{1}_{[\beta_l, \beta_r]}(\beta)$$

est connue à un facteur près et ne peut être directement simulée, excepté *via* des techniques plutôt coûteuses d'acceptation-rejet (Robert & Casella 2004). Le chapitre 5 est dédié à la présentation et la comparaison de méthodes numériques adéquates. L'incertitude sur ces estimations est alors mesurée par l'écart-type empirique calculé sur le tirage $(\beta_1, \dots, \beta_M)$. Dans l'exemple suivant, on préférera indiquer le *facteur d'erreur* défini au chapitre 2 (formule (3.10)), plus parlant pour les ingénieurs EDF.

3.4.2.2 Exemple

Nous réutilisons l'échantillon *B215* (tableau 2.3 p. 46), dont les données non censurées proviennent du modèle de Weibull de paramètre $(\eta_0, \beta_0) = (8760, 2.15)$. On présente sur le tableau 3.5 l'estimation du facteur d'erreur ρ de $S(t_0)$ et $\text{MRTF}(t_0)$ selon les méthodes classiques de Newton-Raphson (NR) et SEM, puis dans notre cadre bayésien. Rappelons qu'il est défini, pour une variable de distribution A , par

$$\rho(A) = \sqrt{\max\left(\frac{q_{50\%}}{q_5\%}, \frac{q_{95\%}}{q_{50\%}}\right)}$$

où q_α représente le quantile d'ordre α de la distribution A . Dans le cadre bayésien, nous fixons $\beta \sim \mathcal{U}_{[1,5]}$, nous considérons une médiane marginale $x_e = 8500$ et nous faisons évoluer la taille des données fictives a . La survie et la durée de vie résiduelle sont estimées en $t_0 = 9000$. Les tirages sont respectivement produits par bootstrap (cadre fréquentiste) et PMC (cadre bayésien, cf. chapitre 5). Dans les deux cas, les calculs sont effectués sur 10^3 valeurs simulées.

		$S(t_0)$	$\text{MRTF}(t_0)$
fréquentiste	NR	2.86	3.12
	SEM	2.77	2.95
<hr/> a <hr/>			
bayésien	1	2.38	2.62
	2	1.84	2.43
	3	1.32	2.21
	4	1.15	1.64

TAB. 3.5 – Mesures du facteur d'erreur sur les grandeurs fiabilistes.

Pour des valeurs de taille a représentatives des procédures de calibration que nous avons proposées dans la section précédente, on voit clairement que le facteur d'erreur ρ des deux grandeurs fiabilistes atteint la précision minimale voulue ($\rho < 2.5$), que les méthodes fréquentistes permettent à peine (ou pas) d'obtenir. La modélisation de Weibull que nous proposons permet en outre de mesurer le nombre de données fictives nécessaires pour atteindre, *a posteriori*, une précision industrielle sur un problème

donné. On peut par exemple en déduire le nombre d'experts nécessaire pour l'étude. Sur cet exemple, un ou deux experts suffisent.

3.4.3 Étude de sensibilité

L'influence de modifications *a priori* sur les résultats *a posteriori* est perçue *via* la mesure de l'incertitude des grandeurs fiabilistes définies précédemment. Nous considérons les modifications suivantes, les plus probables lors de l'étude de l'analyste industriel :

1. une légère variation des bornes (β_l, β_r) ;
2. une hésitation entre les spécifications espérance/médiane/mode ;
3. une modification de l'ordre α d'un quantile marginal.

Afin de moyenniser les résultats, on utilisera 100 échantillons \mathbf{y}_n de taille $n = 10$, simulés selon la loi de Weibull de paramètres $(\eta_0, \beta_0) = (110, 3)$. Nous les censurons par des données tirées uniformément selon $[110, 130]$ de façon à obtenir un taux de censure moyen de 50%. On fixe $t_0 = 130$.

3.4.3.1 Spécifications espérance/médiane/mode

Nous fixons ici $x_e = 100$ et $a = 2$ ou $a = 4$. Nous comparons les modélisations B_i , $i = 1, 2, 3$ (cf. proposition 6) dans le cas où B_1 spécifie x_e comme une médiane. On pose $\beta \sim \mathcal{B}_e(p, q)$ sur $[\beta_l, \beta_r]$. On décide tout d'abord $\beta_l \sim \mathcal{U}[1 - \varepsilon_l, 1 + \varepsilon_l]$ et on fixe $\beta_r = 5$. On mesure alors l'écart-type des estimations des grandeurs fiabilistes en fonction de l'augmentation de ε_l . Les résultats sont fournis dans le tableau 3.6. Deuxièmement, on adopte la démarche inverse en décidant $\beta_r \sim \mathcal{U}[5 - \varepsilon_r, 5 + \varepsilon_r]$ et en fixant $\beta_l = 1$. On reproduit dans le tableau 3.7 ces mêmes écart-types en fonction de ε_r . Dans les deux cas, (p, q) sont modifiés de façon à conserver $E[\beta] = 3$ et $\text{Var}[\beta] = 1$.

	ε_l	E[X]			$S(t_0)$			MRTF(t_0)		
		B_1	B_2	B_3	B_1	B_2	B_3	B_1	B_2	B_3
$a = 2$	0.05	5.47	5.50	5.44	0.038	0.040	0.040	7.22	7.26	7.25
	0.1	6.21	6.31	6.33	0.051	0.053	0.054	8.27	8.32	8.30
	0.2	6.78	6.91	6.88	0.054	0.055	0.052	8.30	8.37	8.35
	0.5	7.53	8.00	7.78	0.068	0.070	0.077	9.33	9.42	9.39
$a = 4$	0.05	5.45	5.48	5.50	0.040	0.043	0.042	6.33	6.34	6.35
	0.1	5.53	5.89	5.84	0.050	0.053	0.052	7.26	7.28	7.26
	0.2	5.98	6.08	6.12	0.061	0.062	0.064	7.78	7.86	7.82
	0.5	6.37	6.54	6.58	0.062	0.064	0.066	8.08	8.13	8.11

TAB. 3.6 – Estimations des écart-types vis-à-vis de modifications sur β_l , pour chaque choix de modélisation.

La modélisation B_1 (médiane) apparaît comme la plus stable, au sens où les trois grandeurs fiabilistes qui en émanent sont, d'une façon générale, moins sensibles aux variations des bornes (β_l, β_r) . Ce résultat conforte le choix par défaut de B_1 proposé au § 3.3.4 (p. 77) lorsque l'expertise en valeur centrale est délicate à spécifier. Par ailleurs, les modifications de β_l provoquent, pour chacune des modélisations, légèrement moins de perturbations *a posteriori* que des modifications équivalentes sur β_r . C'est un résultat plutôt bénéfique ; en effet, cette borne supérieure est en général moins susceptible d'être modifiée que β_l , qui est liée à la connaissance du vieillissement.

	ε_r	E[X]			S(t ₀)			MRTF(t ₀)		
		B ₁	B ₂	B ₃	B ₁	B ₂	B ₃	B ₁	B ₂	B ₃
a = 2	0.05	5.04	5.22	5.27	0.052	0.054	0.053	7.27	7.35	7.32
	0.1	5.35	5.38	7.02	0.045	0.048	0.045	7.77	7.82	7.83
	0.2	6.99	7.13	7.82	0.071	0.073	0.070	8.38	8.41	8.46
	0.5	7.63	8.21	8.73	0.063	0.065	0.062	9.18	9.23	9.21
a = 4	0.05	5.31	5.50	6.08	0.057	0.058	0.058	6.46	6.49	6.47
	0.1	5.47	5.70	6.12	0.058	0.060	0.060	6.89	7.03	7.10
	0.2	6.17	6.33	6.22	0.066	0.067	0.065	7.41	7.45	7.52
	0.5	7.32	7.44	7.87	0.075	0.082	0.077	8.22	8.34	8.33

TAB. 3.7 – Estimations des écart-types vis-à-vis de modifications sur β_r , pour chaque choix de modélisation.

3.4.3.2 Spécification de quantiles

On suppose ici que l'expert spécifie un quantile $x_e = 100$ d'ordre α . On choisit $a = 2$. Pour quelques valeurs de α , nous calculons l'écart-type σ des estimations fiabilistes de $E[X]$ et $S(t_0)$ en fonction des variations ε_l et ε_r respectives sur β_l et β_r , dans les mêmes conditions que précédemment. Ces résultats sont présentés sur le tableau 3.8.

	5%	10%	25%	50%	75%	90%	95%
ε_l							
0.1	(14.0,0.024)	(9.1,0.035)	(6.8,0.048)	(6.21,0.051)	(6.7,0.061)	(6.9,0.064)	(7.1,0.067)
0.3	(20.8,0.025)	(11.3,0.034)	(7.8,0.049)	(6.7,0.054)	(6.92, 0.056)	(7.11, 0.058)	(7.23, 0.062)
ε_r							
0.1	(16.3,0.023)	(9.3,0.038)	(6.8,0.043)	(5.3,0.045)	(6.3,0.051)	(7.1,0.062)	(7.8,0.068)
0.3	(20.9,0.025)	(12.4,0.041)	(7.1,0.042)	(5.7,0.058)	(6.8,0.062)	(7.4,0.070)	(8.1,0.079)

TAB. 3.8 – Estimations des écart-types fiabilistes vis-à-vis de modifications sur β_l (haut) et β_r (bas). Un doublet correspond à l'écart-type σ sur $(E[X], S(t_0))$. Les pourcentages 5%, ..., 95% correspondent à l'ordre α du quantile $x_e = 100$.

Nous constatons que l'estimation *a posteriori* de la durée de vie, à partir d'une spécification par quantiles extrêmes, est la plus sensible aux modifications des bornes. Ce résultat est donc un argument en faveur de l'étape de recalibration qui propose de corriger l'ordre des quantiles fournis par l'expert, en les modulant autour des principales valeurs (20%, 25%, 33%, 66%, 75%, 80%).

Enfin, notons que la sensibilité de l'indicateur de survie augmente lorsque $\alpha \rightarrow 1$: la spécification de $\alpha > 0.5$ est plus fragile que celle de $\alpha \leq 0.5$ lorsqu'il s'agit de prédire un comportement futur. Ce résultat entraîne le commentaire suivant : lorsque l'expert fournit un intervalle de crédibilité, il est intéressant de préciser plus particulièrement la spécification du quantile supérieur, lorsque l'objectif de l'ingénieur est de prédire le comportement futur de \sum . La question "cette valeur supérieure correspond-elle à une défaillance observée, à une date d'arrêt moyenne, ou est-elle extrapolée à partir de dates d'arrêt antérieures ?" permet de juger du sérieux de cette spécification et d'amorcer une rectification de l'ordre du quantile, puis de la taille équivalente, en prenant en compte les règles proposées au § 3.3.5.2.

3.5 Indicateurs de taille des données observées

Dans cette sous-partie, dans un cadre plus général que le modèle de Weibull, nous nous intéressons au calcul de la taille \tilde{n} de données *i.i.d.* $\mathbf{x}_{\tilde{n}} \sim \mathcal{M}(\theta)$ apportant la même information que les données observées \mathbf{y}_n , qui sont censurées à droite. La valeur de \tilde{n} est en effet un indicateur permettant de limiter la valeur de l'hyperparamètre a , au sein de la procédure de recalibration proposée au § 3.3.5.2. On peut également estimer par le rapport a/\tilde{n} l'importance apportée à la connaissance *a priori* vis-à-vis de l'information objective, ou fixer une valeur de a comme un multiple ou une fraction de \tilde{n} dans des études de sensibilité (voir par exemple Celeux *et al.* 2006a).

Nous définissons tout d'abord la taille effective des données comme celle d'un échantillon *i.i.d.* ayant *a posteriori* l'influence la plus proche de celle des données réelles \mathbf{y}_n . Cette définition nous semble donc légitime dans le contexte bayésien de notre étude. Face aux difficultés de calcul de cette taille effective, nous proposons un indicateur approximant \tilde{n} en considérant que les deux échantillons apportent la même quantité d'information sur le paramètre θ . La pertinence de cet indicateur n'est pas discutée ici sinon au travers d'une comparaison avec la taille effective, par le biais d'exemples sur les modèles exponentiel et de Weibull. À notre idée, il constitue un point de départ potentiel pour des recherches plus en profondeur.

3.5.1 Taille effective des données

Soit π^J un *a priori* non informatif, dont le recouvrement *a posteriori* possède la meilleure validité fréquentiste (soit d'ordre maximal selon la définition 1 du § 3.2.1) pour un échantillon *i.i.d.* De façon similaire, on définit l'*a priori* π_c^J pour l'échantillon observé \mathbf{y}_n , potentiellement censuré. Ainsi π_c^J n'est pas forcément égal à π^J ; par exemple, dans le cas exponentiel, π_c^J est l'*a priori* proposé par De Santis *et al.* (2001). Voir § 3.2.2.1 pour plus de précisions.

On définit la taille effective des données observées par

$$\tilde{n}^* = \arg \min_{\tilde{n}} \mathbb{E}_{X_{\tilde{n}}^*} \mathcal{D}^{J_c, J}(\mathbf{y}_n, \mathbf{x}_{\tilde{n}}) \quad (3.10)$$

avec

$$\mathcal{D}^{J_c, J}(\mathbf{y}_n, \mathbf{x}_{\tilde{n}}) = KL \{ \pi_c^J(\cdot | \mathbf{y}_n) \parallel \pi^J(\cdot | \mathbf{x}_{\tilde{n}}) \}. \quad (3.11)$$

En d'autres termes, l'échantillon $\mathbf{x}_{\tilde{n}}$ minimise la divergence de Kullback-Leibler $\mathcal{D}^{J_c, J}(\mathbf{y}_n, \mathbf{x}_{\tilde{n}})$ et on obtient la taille effective en moyennant sur l'ensemble des $\mathbf{x}_{\tilde{n}}$ possibles. On perçoit donc $\mathbf{x}_{\tilde{n}}$ comme l'échantillon *i.i.d.* virtuel transportant une information dont l'impact *a posteriori* est le plus proche de l'information transportée par \mathbf{y}_n , ce qui fait sens dans le cadre bayésien de notre étude. L'article le plus récent sur le sujet est celui de Lin *et al.* (2006). L'existence et l'unicité de \tilde{n}^* est garanti par la proposition suivante.

PROPOSITION 8. (Lin *et al.* 2006). *Sous les conditions de Wald (Wald 1949) et sous réserve que les densités a posteriori dans (3.11) soient propres, l'échantillon $\mathbf{x}_{\tilde{n}}^*$ existe avec probabilité 1 et est unique.*

L'estimation de \tilde{n} est généralement très coûteuse. Elle peut être considérablement simplifiée lorsque le modèle $\mathcal{M}(\theta)$ admet des statistiques exhaustives. Soit T une statistique exhaustive minimale. Souvent, $\mathcal{D}^{J_c, J}(\mathbf{y}_n, \mathbf{x}_{\tilde{n}})$ est une fonction convexe de \tilde{n} et de $t(\mathbf{x}_{\tilde{n}})$, rendant l'espérance dans (3.10) inutile.

Voir ainsi l'exemple 8. Mais en général, des méthodes d'échantillonnage pondéré (voir chapitre 5) couplées à des méthodes de gradient sont nécessaires, et rendent le calcul considérablement ardu.

EXEMPLE 8. *Modèle exponentiel.*

On suppose que l'échantillon \mathbf{y}_n contient $n - r$ censures de type I progressif c_1, \dots, c_{n-r} . D'après § 3.2.2.1, la densité *a posteriori* provenant du choix de Jeffreys censuré est

$$\pi_c^J(\lambda|\mathbf{y}_n) \propto \lambda^{r-1} \left(n - \sum_{i=1}^{n-r} \exp(-\lambda c_i) \right)^{1/2} \exp \left(-\lambda \sum_{j=1}^n y_j \right)$$

qui est propre pour tout $r \leq n$ quand $n \geq 1$. Notons $\hat{y} = 1/\hat{\lambda}_n$ où $\hat{\lambda}_n$ est le MLE et \bar{x} la moyenne de $\mathbf{x}_{\tilde{n}}$. Alors

$$\mathcal{D}^{J_c, J}(\mathbf{y}_n, \mathbf{x}_{\tilde{n}}) = -\tilde{n} \log \tilde{n} \bar{x} + \log \Gamma(\tilde{n}) - (\tilde{n} - 1) \mathbb{E}_n[\log \lambda] + \tilde{n} \bar{x} \mathbb{E}_n[\lambda] + \mathbb{E}_n[\log \pi_c^J(\lambda|\mathbf{y}_n)]$$

où l'espérance est définie par rapport à $\pi_c^J(\lambda|\mathbf{y}_n)$. Cela mène à la solution unique $\bar{x}^* = 1/\mathbb{E}_n[\lambda]$ et \tilde{n}^* telle que

$$\varphi(\tilde{n}^*) = \log \tilde{n}^* - \Psi(\tilde{n}^*) = \log \mathbb{E}_n[\lambda] - \mathbb{E}_n[\log \lambda] \quad (\geq 0 \text{ d'après l'inégalité de Jensen}),$$

φ étant une fonction bijective sur \mathbb{R}^+ . Une méthode de Newton-Raphson est parfaitement adaptée à l'estimation de \tilde{n}^* avec une bonne précision. Notons que si nous choisissons $\pi_c^J = \pi^J$ (l'*a priori* de Jeffreys standard), nous obtenons $\tilde{n}^* = r$, ce qui minore l'information apportée par les données.

Sur le tableau 3.9 les valeurs de \tilde{n}^* sont calculées et moyennées pour les échantillons simulés selon une loi de paramètre $\lambda_0 = 1/100$. Les données censurées sont choisies uniformément au hasard dans [80, 130]. Les pourcentages de censure et les résultats sont moyennés sur 50 échantillons (pour obtenir des écart-types plus petits que 0.2). Il apparaît pour ces échantillons que la censure a un impact limité sur l'information, puisque \tilde{n}^* reste proche de r .

n	censure %			
	33 %	50 %	75 %	90 %
10	6.16	5.15	2.22	1.48
5	3.42	2.62	1.31	

TAB. 3.9 – Valeurs de la taille virtuelle \tilde{n}^* . ■

3.5.2 Un indicateur approximatif

Pour pallier les difficultés du calcul de \tilde{n}^* , nous proposons l'indicateur approximatif défini de la manière suivante. On considère que les échantillons \mathbf{y}_n et $\mathbf{x}_{\tilde{n}}$ apportent la même information de Fisher (en supposant qu'elle est toujours définie) sur le paramètre θ . L'additivité de cette information permet de décrire simplement l'influence des données censurées sur le paramètre θ . Supposons que \mathbf{y}_n

contienne $n - r$ valeurs de censures à droite $\mathbf{c}_{n-r} = (c_1, \dots, c_{n-r})$. Quand les r données non censurées restantes sont *i.i.d.*, l'information de Fisher apportée par \mathbf{y}_n sur θ est

$$I_\theta(\mathbf{y}_n) = H_D(\theta) + H_C(\theta)$$

où

$$H_D(\theta) = -\mathbb{E}_\theta \left[\sum_{k=1}^n \frac{\partial^2 \log f(y_k|\theta)}{\partial \theta_i \partial \theta_j} \right] \mathbb{E}[r|\theta],$$

$$H_C(\theta) = -\mathbb{E}_\theta \left[\sum_{k=1}^n \frac{\partial^2 \log S(y_k|\theta)}{\partial \theta_i \partial \theta_j} \right] \mathbb{E}[n - r|\theta],$$

avec

$$\mathbb{E}[r|\theta] = \mathbb{E} \left[\sum_{k=1}^n \delta_k | \theta \right]$$

l'espérance du nombre de données non censurées dans l'échantillon. Malheureusement, nous devons avoir une connaissance précise du processus de censure pour calculer cette espérance. En l'absence de cette connaissance - ce qui est souvent le cas, industriellement parlant - nous proposons d'utiliser la distribution empirique des censures observées (voir chapitre 2) si $n - r$ est grand, ou un cas de censure de type I progressif. Nous supposons nécessairement vérifiées les conditions classiques d'existence et de régularité de $I_\theta(\mathbf{y}_n)$ (voir par exemple Clarke & Barron 1990). L'information de Fisher standard

$$I_\theta = -\mathbb{E}_\theta \left[\frac{\partial^2 \log f(x|\theta)}{\partial \theta_i \partial \theta_j} \right].$$

est la quantité d'information moyenne apportée par une donnée *i.i.d.* sur θ . Nous définissons alors l'indicateur de taille effective par par

$$\tilde{n}(\theta) = |I_\theta(\mathbf{y}_n) I_\theta^{-1}|^{1/2}.$$

Visiblement, $\tilde{n}(\theta)$ est invariante par reparamétrisation. Quand \mathbf{y}_n n'est pas censuré, $\tilde{n}(\theta) = n$ pour tout θ . Autrement, nous avons toujours $r \leq \tilde{n}(\theta) \leq n$ où r est le nombre de données non censurées.

La difficulté évidente posée par cette définition est la dépendance en θ de $\tilde{n}(\theta)$. L'information de Fisher est l'information utilisée usuellement dans un cadre fréquentiste, où le paramètre θ est fixé. Afin de construire un indicateur objectif de \tilde{n} , nous proposons de l'estimer au maximum de vraisemblance $\hat{\theta}_n$. Dans nos exemples, l'indicateur fourni par cette méthode est proche de la taille effective \tilde{n}^* définie auparavant, et a pour avantage d'être beaucoup plus facile à calculer.

EXEMPLE 9. *Modèle exponentiel.*

On considère un échantillon exponentiel \mathbf{y}_n contenant r données *i.i.d.* apportant l'information de Fisher $I_\lambda = \lambda^{-1}$, et $n - r$ censures c_1, \dots, c_{n-r} de type I progressif apportant chacune l'information $I_\lambda(c) = \lambda^{-1}(1 - \exp(-\lambda c))$. Alors

$$\tilde{n}(\check{\lambda}) = n - \sum_{i=1}^{n-r} \exp(-\check{\lambda} c_i).$$

n	% de censure			
	33%	50%	75%	90%
10	6.19	5.18	2.26	1.52
5	3.47	2.66	1.34	

TAB. 3.10 – Valeurs de l'indicateur $\tilde{n}(\hat{\lambda}_n)$. ■

En réutilisant les échantillons simulés dans l'exemple 8, on présente sur le tableau 3.10 les valeurs moyennées de $\tilde{n}(\hat{\lambda}_n)$. Elles approximent correctement (en les surestimant légèrement) les tailles effectives \tilde{n}^* présentées dans l'exemple 8.

EXEMPLE 10. *Modèle de Weibull.*

Nous dérivons ici une approximation de $\tilde{n}(\theta)$ pour le modèle de Weibull, utilisable dans les procédures décrites au § 3.3.5.2. Considérons la paramétrisation $\theta = (\mu, \beta)$. Soit $n - r$ valeurs de censure à droite (c_1, \dots, c_{n-r}) , de type I progressif. D'après le théorème 2 et la proposition 3, nous obtenons, en notant $\kappa = 1 + \pi^2/6$,

$$\tilde{n}(\check{\mu}, \check{\beta}) = n - k_1(\check{\mu}, \check{\beta})k_2(\check{\mu}) - \frac{1}{n}k_1^2(\check{\mu}, \check{\beta})(k_2^2(\check{\mu}) - 2\kappa) + o(n^{-1})$$

où

$$\begin{cases} k_1(\mu, \beta) &= \frac{1}{2\kappa} \sum_{i=1}^{n-r} \exp(-\mu c_i^\beta), \\ k_2(\mu) &= \kappa + (\gamma - 1)^2 - 2(1 - \gamma) \log \mu + \log^2 \mu. \end{cases}$$

Sur le tableau 3.11 nous indiquons les estimations du minimiseur de Kullback \tilde{n}^* quand π_c^J est l'a priori de Jeffreys censuré défini au § 3.2.2.1 et π^J est le *reference prior* standard. En parallèle, nous donnons la valeur de l'indicateur approximatif

$$\tilde{n}(\hat{\theta}_n) = \tilde{n}(\hat{\mu}_n, \hat{\beta}_n).$$

Ces résultats sont moyennés sur 50 échantillons de taille variable, simulées selon les paramètres $\eta_0 = (1/\mu_0)^{1/\beta} = 100$ et $\beta_0 = 2$ et censurés de la même façon que dans l'exemple 8. Les deux indicateurs sont visiblement très proches, même si \tilde{n}^* reste là encore légèrement plus faible. Le calcul de \tilde{n}^* est difficile et extrêmement coûteux.

n	% de censure							
	33%		50%		75%		90%	
	\tilde{n}^*	$\tilde{n}(\hat{\theta}_n)$	\tilde{n}^*	$\tilde{n}(\hat{\theta}_n)$	\tilde{n}^*	$\tilde{n}(\hat{\theta}_n)$	\tilde{n}^*	$\tilde{n}(\hat{\theta}_n)$
15	10.2	10.3	7.95	8.05	4.00	4.14	1.82	1.90
10	6.75	6.78	5.10	5.14	3.70	3.75	1.08	1.13
5	3.40	3.42	2.58	2.62	1.28	1.32		

TAB. 3.11 – Valeurs de la taille fictive \tilde{n}^* et de l'indicateur $\tilde{n}(\hat{\theta}_n)$. L'indicateur $\tilde{n}(\hat{\theta}_n)$, très simple à calculer, approxime la taille fictive \tilde{n}^* en la surestimant légèrement. ■

3.6 Annexes

3.6.1 Preuves et justifications

PROPOSITION 2 (*Modèle exponentiel*).

On suppose avoir $\mathbf{y}_n = (y_1, \dots, y_n)$ valeurs observées parmi lesquelles r valeurs non censurées $\mathbf{x}_r = (x_1, \dots, x_r)$ de densité exponentielle $f(x|\lambda)$ et de fonction de répartition $F(x|\lambda)$. On note $\mathbf{c}_{n-r} = (c_1, \dots, c_{n-r})$ les valeurs de censure. Supposons pouvoir spécifier une densité probabiliste des instants de censure f_c ; on note alors

$$\delta(\lambda) = \int_{\mathcal{R}} f_c(x) F(x|\lambda) dx.$$

Alors, on obtient dans les différents cas de censure suivants,

1. *censure de type I progressif* : $\delta(\lambda) = n - \sum_{i=1}^{n-r} \exp(-\lambda c_i)$;
2. *censure uniforme sur $[0, c_\infty]$* : $\delta(\lambda) = 1 - \{1 - \exp(-\lambda c_\infty)\}(\lambda c_\infty)^{-1}$;
3. *censure de type II* : $\delta(\lambda) = r/n$.

Preuve. Les résultats des cas de censure de type I progressif et uniforme proviennent de l'article de De Santis, Mortera et Nardi (2001). Dans un cas de censure de type II, l'unique valeur de censure est la valeur de la la statistique d'ordre r de la partie non censurée de l'échantillon \mathbf{y}_n . Le nombre r est considéré comme indépendant des observations de \mathbf{y}_n . Alors

$$\delta(\lambda) = P\left(F(x|\lambda) \leq \frac{r}{n}\right) = \frac{r}{n}. \quad \square$$

THÉORÈME 2 (*Modèle de Weibull*).

Notons $f_W(x) = \beta \mu x^{\beta-1} \exp(-\mu x^\beta)$ la densité de Weibull et F_W sa fonction de répartition. Soit γ la constante d'Euler ($\gamma \simeq 0.57722$). Soient $\gamma_1 = \pi^2/6 + \gamma^2 - 2\gamma > 0$, $\gamma_2 = -2(1 - \gamma) < 0$. En se plaçant dans les mêmes conditions que pour la proposition 2, on suppose connaître une densité $f_c(x)$ des valeurs de censure. On note alors

$$\delta(\mu, \beta) = \int_{\mathcal{R}} f_c(x) F_W(x|\mu, \beta) dx.$$

On note par ailleurs $\tilde{\delta}(\mu, \beta) = \delta^2(\mu, \beta) + [\delta(\mu, \beta) - 1] (\gamma_1 + \gamma_2 \log \mu + \log^2 \mu) + \pi^2/6 - 1$. Alors la mesure de Jeffreys pour la paramétrisation (μ, β) du modèle de Weibull censuré à droite est

$$\pi_c^J(\mu, \beta) \propto (\mu\beta)^{-1} \sqrt{\tilde{\delta}(\mu, \beta)}.$$

Preuve. La densité d'une variable X de Weibull censurée à droite par la variable C (indépendante de X) est

$$g(X) = [f_W(X)]^{\mathbb{1}_{\{X < C\}}} [1 - F_W(X)]^{\mathbb{1}_{\{X \geq C\}}} = [\beta \mu X^{\beta-1} \exp(-\mu X^\beta)]^{\mathbb{1}_{\{X < C\}}} [\exp(-\mu X^\beta)]^{\mathbb{1}_{\{X \geq C\}}}.$$

La matrice de score associée aux paramètres (μ, β) est alors

$$S_{(\mu, \beta)}(X) = \begin{bmatrix} \frac{1}{\mu} \mathbb{I}_{\{X < C\}} - X^\beta \\ \left(\frac{1}{\beta} + \log X\right) \mathbb{I}_{\{X < C\}} - \mu X^\beta \log X \end{bmatrix}$$

La matrice hessienne sur (μ, β) s'écrit ensuite

$$H_{(\mu, \beta)}(X) = \begin{bmatrix} -\frac{1}{\mu^2} \mathbb{I}_{\{X < C\}} & -X^\beta \log X \\ -X^\beta \log X & -\frac{1}{\beta^2} \mathbb{I}_{\{X < C\}} - \mu X^\beta \log^2 X \end{bmatrix}.$$

La matrice de Fisher obtenue est

$$I_{(\mu, \beta)} = \begin{bmatrix} \frac{1}{\mu^2} P(X < C) & E[X^\beta \log X] \\ E[X^\beta \log X] & \frac{1}{\beta^2} P(X < C) + E[\mu X^\beta \log^2 X] \end{bmatrix}$$

Si $X \rightsquigarrow \mathcal{W}(\mu, \beta)$ alors $\mu X^\beta = Y \rightsquigarrow \mathcal{E}(1)$. Donc $E[X^\beta \log X] = \frac{1}{\beta \mu} E_Y[Y(\log Y - \log \mu)]$. Cette espérance est bien définie car $g(y) = y \exp(-y) \log y$ est intégrable sur \mathbb{R}^+ , pour $\mu > 0$, et

$$\int_{\mathbb{R}^+} y e^{-y} \log y \, dy = \Psi(2) = 1 - \gamma$$

où Ψ est la fonction digamma (dérivé logarithmique de la fonction gamma) et γ la constante d'Euler ($\gamma \simeq 0.57722$). Plus de précisions sont fournies par exemple dans Abramowitz & Stegun (1972, p. 258-259). Alors

$$E[X^\beta \log X] = \frac{1}{\beta \mu} (1 - \gamma - \log \mu).$$

$$\begin{aligned} \text{De plus, } E[\mu X^\beta \log^2 X] &= \frac{1}{\beta^2} E_Y[Y \log^2(Y/\mu)], \\ &= \frac{1}{\beta^2} E_Y[Y \log^2 Y] - \frac{2 \log \mu}{\beta^2} \int_{\mathbb{R}^+} y e^{-y} \log y \, dy + \frac{\log^2 \mu}{\beta^2} \int_{\mathbb{R}^+} y e^{-y} \, dy \\ &= \frac{\text{Var}[Z] + \gamma^2 - 2\gamma}{\beta^2} - \frac{\log \mu}{\beta^2} \{2(1 - \gamma) - \log \mu\} \quad \text{d'après Sun (1997),} \end{aligned}$$

où Z est une variable aléatoire suivant la densité de Gumbel $k(x) = \exp(-x) \exp(-\exp(-x))$ définie sur \mathbb{R} , de variance $\frac{\pi^2}{6}$. Ainsi, en notant $\gamma_3 = (1 - \gamma)^2$,

$$\begin{aligned} E[\mu X^\beta \log^2 X] &= \frac{1}{\beta^2} [\gamma_1 + \gamma_2 \log \mu + \log^2 \mu] \\ \text{puis } \det I_{(\mu, \beta)} &= \left(\frac{1}{\mu \beta}\right)^2 [P^2(X < C) + P(X < C) (\gamma_1 + \gamma_2 \log \mu + \log^2 \mu)] \\ &\quad - \left(\frac{1}{\mu \beta}\right)^2 (\gamma_3 + \gamma_2 \log \mu + \log^2 \mu). \end{aligned}$$

On en déduit sans difficulté l'expression de $\tilde{\delta}(\mu, \beta)$. \square

PROPOSITION 3 (*Modèle de Weibull*).

En réutilisant les notations du théorème 2, on obtient, selon le type de censure,

1. *censure de type I progressif* : $\delta(\mu, \beta) = n - \sum_{i=1}^{n-r} \exp(-\mu c_i^\beta)$;
2. *censure uniforme sur $[0, c_\infty]$* : $\delta(\mu, \beta) = 1 - (\beta\mu^{1/\beta})^{-1} \{1 - \Gamma(\beta^{-1}, \mu c_\infty^\beta)\}$ où $\Gamma(x, \alpha)$ est la fonction gamma incomplète ;
3. *censure de type II* : $\delta(\mu, \beta) = r/n$ est indépendant de (μ, β) .

Preuve. Le cas type I progressif correspond à des données de censure fixes C_1, \dots, C_{n-r} . Donc $P(X < C) = r + \sum_{i=1}^{n-r} F(C_i | \mu, \beta) = n - \sum_{i=1}^{n-r} \exp(-\mu C_i^\beta)$. Le cas de censure uniforme est immédiat en considérant la paramétrisation $U = \mu X^\beta$. Dans le cas type II, de la même façon que dans la preuve de la proposition 2, on obtient simplement que $\delta(\mu, \beta) = r/n$. \square

PROPOSITION 6. Soit (μ, β) la paramétrisation d'un modèle de Weibull. Soit $x_e \in \mathbb{R}^{+*}$, $a > 0$ et $\alpha \in]0, 1]$. On note

$$\begin{aligned} b_1(a, \beta) &= \left((1 - \alpha)^{-1/a} - 1 \right)^{-1} x_e^\beta, \\ b_2(a, \beta) &= \left(\frac{\Gamma(a)}{\Gamma(1 + 1/\beta)\Gamma(a - 1/\beta)} \right)^\beta x_e^\beta, \\ b_3(a, \beta) &= \frac{a\beta + 1}{\beta - 1} x_e^\beta. \end{aligned}$$

On rappelle que B désigne la modélisation induite par le choix de $b(a, \beta)$. Alors, pour tout choix de $\pi(\beta)$ de support $[\beta_l, \beta_r] \subset]0, +\infty[$, en choisissant *a priori*

$$\mu | \beta \sim \mathcal{G}(a, b),$$

on a

$$\begin{aligned} (i) \quad P(X \leq x_e | B_1) &= \alpha, \\ (ii) \quad \mathbb{E}[X | B_2] &= x_e \text{ si } a > \beta_l^{-1}, \\ (iii) \quad M_d[X | B_3] &= x_e \text{ si } \beta_l > 1. \end{aligned}$$

Dans le cas (iii), le mode x_e de la distribution marginale M est unique et strictement positif. Dans les autres cas, si $\beta_l \geq 1$, M possède au moins un mode strictement positif.

Preuve. On rappelle que F_W désigne la fonction de répartition de la loi de Weibull. On a alors

$$\begin{aligned} P(X \leq x_e | B_1) &= M(x_e | B_1), \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} F_W(x_e | \mu, \beta) \pi(\mu | B_1, \beta) \pi(\beta) \, d\mu d\beta, \\ &= 1 - \int_{\mathbb{R}} \left(\frac{1}{1 + \frac{x_e^\beta}{b_1(\beta)}} \right)^a \pi(\beta) \, d\beta, \\ &= 1 - \int_{\mathbb{R}} (1 - \alpha) \pi(\beta) \, d\beta = \alpha. \end{aligned}$$

Ce résultat est aussi valable même si $\pi(\beta)$ n'est pas de support compact. Par ailleurs, en supposant $a > \beta_l^{-1}$,

$$\begin{aligned} E[X|B_2] &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mu^{-1/\beta} \Gamma(1 + 1/\beta) \pi(\mu|B_2, \beta) \pi(\beta) \, d\mu d\beta, \\ &= \int_{\mathbb{R}} \Gamma(1 + 1/\beta) b_2^{1/\beta}(\beta) \frac{\Gamma(a - 1/\beta)}{\Gamma(a)} \pi(\beta) \, d\beta = x_e. \end{aligned}$$

Ce résultat est encore valable si $\text{Supp}(\pi(\beta))$ est seulement borné à gauche par β_l . Enfin, un mode éventuel de M est situé en un point $x \neq 0$ telle que la dérivée $m'(x)$ soit nulle. On a

$$m(x) = \int_{\mathbb{R}} \frac{ab^a}{(b + x^\beta)^{a+1}} \beta x^{\beta-1} \pi(\beta) \, d\beta.$$

Avec $\text{Supp}(\pi(\beta))$ compact, la dérivation sous l'intégrale est sans difficulté et

$$m'(x) = - \int_{\mathbb{R}} \frac{ab^a x^{\beta-2}}{(b + x^\beta)^{a+2}} [x^\beta(a\beta + 1) - b(\beta - 1)] \pi(\beta) \, d\beta. \quad (3.12)$$

Lorsque $a > \beta_l^{-1}$, le choix de $b = b_3$ permet visiblement d'annuler $m'(x_e)$. Par ailleurs, $\forall x > 0$, on a

$$m'(x|B_3) = \int_{\mathbb{R}} \frac{ab_3^a x^{\beta-2}}{(b_3 + x^\beta)^{a+2}} (a\beta + 1) \pi(\beta) \{x_e^\beta - x^\beta\} \, d\beta$$

qui est du même signe que la différence $x_e - x$. Il n'existe donc qu'un seul zéro non nul de $m'(x)$. Supposons que $\beta_l \geq 1$. Pour les choix (i) et (ii) de b , en réutilisant l'expression (3.12), il est aisé de montrer qu'il existe $x_0 < x_e$ tel que pour $0 < x < x_0$, $m'(x) > 0$. On note qu'il existe semblablement $x'_0 > x_e$ tel que $\forall x > x'_0$, $m'(x) < 0$. Cela implique par continuité que $m(x)$ possède au moins un mode compris dans $[x_0, x'_0]$. \square

JUSTIFICATION DE L'ALGORITHME 1.

Le terme δ_k est le rapport de la fonction $\ell(\alpha^k) - \ell_0$ à annuler en α^k sur son gradient. Celle-ci s'annule quand $\beta_e = \beta_0$, avec

$$\beta_e = \log \left\{ \frac{\log(1 - \alpha_2)}{\log(1 - \alpha_1)} \right\} \log^{-1}(x_{e_2}/x_{e_1}).$$

L'algorithme a la forme usuelle

$$\alpha^{k+1} = \alpha^k - \rho_k \delta_k.$$

Le pas adaptatif ρ_k est choisi de façon à assurer $0 < \alpha_1^{k+1} < \alpha_2^{k+1} < 1$ à l'itération $k \rightarrow k+1$. On note que $\delta_k[1] < 0$ et $\delta_k[2] > 0$. Par ailleurs, si $\ell^k > 1$ (ce qui est toujours le cas dans nos applications), $-1 < h_k = \delta_k[1]/\delta_k[2] < 0$. Alors,

$$\alpha_1^{k+1} < \alpha_2^{k+1} \Leftrightarrow \rho_k[2] < \rho_k[1] \frac{\delta_k[1]}{\delta_k[2]} + \frac{\alpha_2^k - \alpha_1^k}{\delta_k[2]} \quad (3.13)$$

Une CNS de (3.13) est donc de choisir $\rho_k[2] = -\rho_k[1] + (\alpha_2^k - \alpha_1^k)/\delta_k[2]$. On désire $\rho_k > 0$ pour augmenter α_1 et diminuer α_2 ($\Rightarrow 0 < \alpha_1^{k+1} < \alpha_2^{k+1} < 1$). C'est chose possible en choisissant

$$\rho_k[1] = \rho(\alpha_2^k - \alpha_1^k)/\delta_k[2].$$

où $0 < \rho < 1$. On choisit alors

$$\rho_k[2] = \gamma \frac{\alpha_2^k - \alpha_1^k}{\delta_k[2]} \left(\rho \frac{\delta_k[1]}{\delta_k[2]} + 1 \right)$$

où $0 < \gamma < 1$, ce qui permet de respecter (3.13). L'étape de Newton-Raphson s'écrit alors

$$\begin{aligned} \alpha_1^{k+1} &= \alpha_1^k - \rho(\alpha_2^k - \alpha_1^k) \frac{\delta_k[1]}{\delta_k[2]}, \\ \alpha_2^{k+1} &= \alpha_2^k - \gamma(\alpha_2^k - \alpha_1^k) \left(\rho \frac{\delta_k[1]}{\delta_k[2]} + 1 \right). \end{aligned}$$

On désire obtenir un rapport de convergence constant au cours des itérations :

$$\frac{|\alpha_1^{k+1} - \alpha_1^k|}{|\alpha_2^{k+1} - \alpha_2^k|} = \frac{\omega_1}{1 - \omega_1}.$$

On obtient alors

$$\gamma = -\rho \frac{\delta_k[1]}{(\rho\delta_k[1] + \delta_k[2])} \frac{\omega_1}{(1 - \omega_1)}$$

puis

$$\begin{aligned} \alpha_1^{k+1} &= \alpha_1^k - \rho(\alpha_2^k - \alpha_1^k) \frac{\delta_k[1]}{\delta_k[2]}, \\ \alpha_2^{k+1} &= \alpha_2^k + \rho(\alpha_2^k - \alpha_1^k) \frac{\delta_k[1]}{\delta_k[2]} \frac{\omega_1}{(1 - \omega_1)}. \end{aligned}$$

On obtient enfin

$$\alpha_2^{k+1} - \alpha_1^{k+1} = (\alpha_2^k - \alpha_1^k) \left\{ 1 + \rho \frac{\delta_k[1]}{\delta_k[2](1 - \omega_1)} \right\}.$$

Plus le terme ρ est petit devant 1, plus l'algorithme sera précis (mais lent). \square

PROPOSITION 7.

Nous rappelons les termes de la proposition. Soit $\pi(\beta)$ de support $[\beta_l, \beta_r] \subset]0, +\infty[$. Soit $\pi(\mu|\beta)$ telle que pour $k = 1, 2$, $E[X^k] < \infty$ où E est l'espérance par rapport à la distribution M_W de densité $m_w(x)$. Soit λ le paramètre d'un modèle exponentiel de densité $f(x|\lambda)$. On place sur λ une densité *a priori* $\pi(\lambda)$ de loi $\mathcal{G}(a, b)$ et on note

$$m_E(x) = \int_0^\infty f(x|\lambda) \pi(\lambda) d\lambda.$$

Alors il existe un unique couple $(a, b)^* > 0$ tel que

$$(a, b)^* = \arg \min \int_{\mathbb{R}} m_W(x) \log \frac{m_W(x)}{m_E(x)} dx.$$

Preuve. Avec $\lambda \sim \mathcal{G}(a, b)$, $(a, b) > 0$ on a

$$m_E(x) = \frac{ab^a}{(b+x)^{a+1}}.$$

Pour tout choix de $\pi(\mu|\beta)$ et $\pi(\beta)$ de support $[\beta_l, \beta_r] \subset]0, +\infty[$, on définit

$$\begin{aligned}\varphi(a, b) &= \int_{\mathbb{R}} m_W(x) \log \frac{m_W(x)}{m_E(x)} dx, \\ &= -H_W - a \log b - \log a + (a+1)E[\log(b+X)] < \infty.\end{aligned}$$

où H_W est l'entropie de la distribution M_W et E est l'espérance par rapport à $m_W(x)$. Alors $\varphi(a, b)$ est continue et deux fois dérivable par rapport à chaque variable, puisque $E[(b+X)^{-k}] < b^{-k} < \infty$ pour $k = 1, 2$. Si $(a, b)^*$ est un extrémum de φ , alors c'est un minimum. La convexité de φ sur $\mathbb{R}^{++} \times \mathbb{R}^{++}$ impose qu'il est unique. En effet,

$$\frac{\partial \varphi(a, b)}{\partial a} = 0 \quad \Rightarrow \quad \frac{a}{b^2} = (a+1)E\left[\frac{1}{b(b+X)}\right]$$

et

$$\begin{aligned}\frac{\partial^2 \varphi(a, b)}{\partial a^2} &= 1/a^2 > 0, \\ \frac{\partial^2 \varphi(a, b)}{\partial b^2} &= \frac{a}{b^2} - (a+1)E\left[\frac{1}{(b+X)^2}\right], \\ &= (a+1)E\left[\frac{1}{(b+X)}\left(\frac{1}{b} - \frac{1}{b+X}\right)\right] > 0\end{aligned}$$

pour $a = a^*$. Notons ensuite que b^* , s'il existe, est solution de l'équation

$$g(b) = E\left[\frac{1}{b+X}\right] \{1 - \log b + E[\log(b+X)]\} - 1/b = 0.$$

On a $g(b) \rightarrow -\infty$ quand $b \rightarrow 0$, $b > 0$. Sous la condition que $E[X^k] < \infty$ pour $k = 1, 2$, on a pour $b \rightarrow \infty$

$$g(b) = \frac{E[X^2] - E^2[X]}{b^2} + o(b^{-2}) = \frac{\text{Var}[X]}{b^2} + o(b^{-2}).$$

Nécessairement, par continuité de g sur \mathbb{R}^{++} , il existe $b^* > 0$ tel que $g(b^*) = 0$. Alors

$$a^* = (E[\log(b^*+X)] - \log b^*)^{-1} > 0.$$

On a alors prouvé l'existence d'un unique couple $(a^*, b^*) > 0$ qui minimise la distance de Kullback entre les marginales. \square

THÉORÈME 3.

Les conditions premières du théorème de Bernstein-von Mises sont la consistance du MLE et des conditions classiques de régularité de la log-vraisemblance $\mathcal{L}(\mathbf{X}_n; \theta)$ (Clarke & Barron 1990). Elles sont usuellement vérifiées pour les modèles réguliers comme Weibull. Nous les rappelons brièvement ci-dessous. Soit $f(x|\theta)$ la densité considérée.

1. l'ensemble $\{x : f(x|\theta) > 0\}$ est le même pour tout $\theta \in \Theta$;
2. $L(\theta, x) = \log f(x|\theta)$ est trois fois différentiable par rapport à θ dans un voisinage $\mathcal{D}_{\theta_0}(\delta) =]\theta_0 - \delta, \theta_0 + \delta[$. On note \dot{L} , \ddot{L} et \dddot{L} les différentielles d'ordres 1, 2 et 3 respectivement. Alors

$$\begin{aligned}\mathbf{E}_{\theta_0}[\dot{L}(\theta_0, X)] &= 0, \\ \mathbf{E}_{\theta_0}[\ddot{L}(\theta_0, X)] &< \infty, \\ \mathbf{E}_{\theta_0}[\ddot{L}(\theta_0, X)] &= -\mathbf{E}_{\theta_0}[\ddot{L}(\theta_0, X)] < 0\end{aligned}$$

et il existe une fonction $M(x)$ telle que $E_{\theta_0}[M(X)] < \infty$ et $\sup_{\theta \in \mathcal{D}_{\theta_0}(\delta)} |\ddot{L}(\theta)| < M(x)$.

Par ailleurs, $\pi(\theta)$ doit être continue et positive en θ_0 . Le support de $\pi(\beta)$ étant compact, les conditions $\beta_0 \in [\beta_l, \beta_r]$ et $\hat{\beta}_n \in [\beta_l, \beta_r]$ pour tout n assez grand sont nécessaires à la consistance de $\pi(\theta|\mathbf{y}_n)$. Au final, nous devons seulement démontrer le lemme suivant, qui constitue la condition principale de la version du théorème de Bernstein-von Mises proposée par Ramamoorthi & Ghosh (2003, p. 35).

LEMME 2. *Soit $Q_n(\theta) = \frac{1}{n}\{L_n(\theta) - L_n(\theta_0)\}$ où $L_n(\theta) = \sum_{i=1}^n \log f(X_i|\theta)$. Alors, $\forall \delta > 0, \exists \varepsilon > 0$ tel que*

$$P_{\theta_0} \left(\sup_{|\theta - \theta_0| > \delta} Q_n(\theta) < -\varepsilon \right) \rightarrow 1.$$

Preuve. Posons $\delta = (\delta_1, \delta_2) > 0$ et $\theta = (\mu, \beta) \in \mathbb{R}^+ \times [\beta_l, \beta_r]$. On a, en maximisant sur μ ,

$$\begin{aligned} \sup_{|\theta - \theta_0| > \delta} Q_n(\theta) &\leq \sup_{|\beta - \beta_0| > \delta_2} Q_n^*(\beta) \\ \text{où } Q_n^*(\beta) &= \log \frac{\beta}{\beta_0 \mu_0} + (\beta - \beta_0) \sum_{i=1}^n \frac{\log X_i}{n} - 1 + \mu_0 \sum_{i=1}^n \frac{X_i^{\beta_0}}{n}. \end{aligned}$$

On a alors, $\forall \beta \in [\beta_l, \beta_r]$, en supposant que $\beta_0 \in [\beta_l, \beta_r]$, et sous la condition que les X_1, \dots, X_n sont i.i.d.,

$$Q_n^*(\beta) \xrightarrow[p.s.]{\mathcal{W}(\theta_0)} Q \left(\frac{\beta}{\beta_0} \right) = \log \frac{\beta}{\beta_0 \mu_0} + (\beta - \beta_0) E_{\theta_0} [\log X] - 1 + \mu_0 E_{\theta_0} [X^{\beta_0}] - \log E_{\theta_0} [X^\beta].$$

On obtient (en utilisant les mêmes types de calcul que la preuve du Théorème 2 p. 63)

$$Q(x) = \log x + \Psi(1)x - \log \Gamma(1+x) - \Psi(1),$$

fonction qui est nulle en $x = 1$ et strictement négative en tout autre point $x > 0$. Pour tout $\delta > 0$, il existe donc $\beta_\delta \in [\beta_l, \beta_r]$, $\beta_\delta \neq \beta_0$, et $\varepsilon' > 0$ tel que, $\forall \beta \in [\beta_0 - \delta, \beta_0 + \delta] \cap [\beta_l, \beta_r]$,

$$Q_n^*(\beta) \leq Q_n^*(\beta_\delta) \xrightarrow[p.s.]{\mathcal{W}(\theta_0)} -\varepsilon'.$$

On en déduit (la convergence presque sûre induisant la convergence en loi)

$$P_{\theta_0} \left(\sup_{|\theta - \theta_0| > \delta} Q_n(\theta) < 0 \right) \geq P_{\theta_0} (Q_n^*(\beta_\delta) < 0) \rightarrow 1$$

d'où, nécessairement, l'énoncé du lemme. \square

PROPOSITION 9. Une propriété des fonctions Gamma. $\forall \beta > 0, \forall n > 0, \forall r > n/\beta, \forall p < 1,$

$$\lim_{r \rightarrow \infty} r^p \left\{ \frac{\Gamma(r - n/\beta)}{\Gamma(r)} r^{-n/\beta} - 1 \right\} = 0.$$

Preuve. On utilise la formule exacte de Stirling (Artin, 1964, p.24).

$$\begin{aligned} \frac{\Gamma(r - n/\beta)}{\Gamma(r) r^{-n/\beta}} &= \frac{\sqrt{2\pi} \left(r - \frac{n}{\beta}\right)^{r-n/\beta-1/2} \exp(-(r - n/\beta) + \mu(r - n/\beta))}{\sqrt{2\pi} r^{r-n/\beta-1/2} \exp(-(r) + \mu(r))} \text{ avec } \mu(r) = \frac{\alpha}{12r} \text{ où } \alpha \in]0, 1[, \\ &= \left(1 - \frac{n}{r\beta}\right)^{r-n/\beta-1/2} \exp(n/\beta + \mu(r(r\beta - n)/n)). \end{aligned}$$

$$\text{Donc } Q_{\beta,n}(r) = \frac{\Gamma(r - n/\beta)}{\Gamma(r) r^{-n/\beta}} - 1 = r^{-1} \frac{n}{\beta} \left[1/2 + \frac{n}{3\beta}\right] + o(r^{-1}).$$

On en déduit la convergence vers 0 lorsqu'on multiplie $Q_{\beta,n}(r)$ par r^p avec $p < 1$. À n et β fixés, l'approximation $Q_{\beta,n}(r) \simeq 1$ est valable rapidement (voir un exemple Figure 3.16), pour $\beta \geq 1$, $n \geq 1$ et à partir de $r \geq 3$. \square

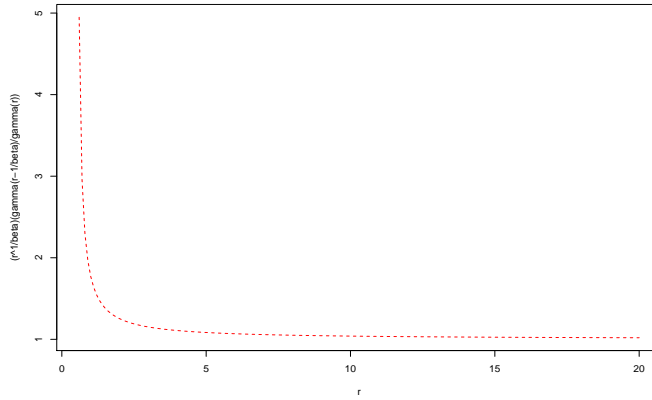


FIG. 3.16 – Évolutions du produit $\frac{r^{1/\beta} \Gamma(r-1/\beta)}{\Gamma(r)}$ en fonction de r , pour $\beta = 2$.

3.6.2 Cas d'une expertise paramétrique

Dans cette sous-partie annexe, nous considérons le cas d'une expertise quantitative qui s'exprime directement sur le paramètre η de Weibull et non plus marginalement sur X .

3.6.2.1 Expertise minimale.

Le cas d'un quantile, d'une espérance ou d'un mode est réglé par la proposition 10.

PROPOSITION 10. *Soit $\eta_e > 0$ et $0 < \alpha < 1$. Notons*

$$\begin{aligned} b_4(a, \beta) &= \eta_e^\beta \chi_{2a}^2(1 - \alpha)/2, \\ b_5(a, \beta) &= (\eta_e \Gamma(a) / \Gamma(a - 1/\beta))^\beta \text{ pour } a > \beta^{-1}, \\ b_6(a, \beta) &= \beta^{-1}(a\beta + 1)\eta_e^\beta. \end{aligned}$$

Alors

- (i) $P(\eta < \eta_e | B_4) = \alpha,$
- (ii) $E[\eta | B_5] = \eta_e,$
- (iii) $M_d[\eta | B_6] = \eta_e$ si $\pi(\beta)$ est de support $[\beta_l, \beta_r] \subset]0, \infty[.$

Preuve. Dans le cas (i), on a $P(\eta < \eta_e) = \alpha \Rightarrow P(\eta < \eta_e | \beta) = \alpha$. Or, si $\mu | \beta \sim \mathcal{G}(a, b)$, $2b\mu | \beta \sim \chi_{2a}^2$. On note $\eta_e(b) = (2b/\chi_{2a}^2(1 - \alpha))^{1/\beta}$. Alors $P(\eta \leq \eta_e(b) | \beta) = \alpha$. On en déduit aisément l'énoncé (i).

Dans le cas (ii), la preuve est immédiate en considérant que $E[\eta] = E[E[\eta | \beta]]$ et en utilisant la formule des moments de la définition 2 p. 71. Dans le cas (iii), il suffit de constater que $\pi(\eta) = E[\pi(\eta | \beta)]$ est dérivable (puisque $\pi(\beta)$ est de support $[\beta_l, \beta_r]$) et d'écrire

$$\pi'(\eta) = \int_{\mathbb{R}} \frac{b^a}{\Gamma(a)} \frac{\exp(-b/x^\beta)}{\eta^{(a+1)\beta+2}} (b\beta - \eta^\beta(a\beta + 1)) \pi(\beta) d\beta.$$

De la même façon que pour la proposition 6 p. 75, il est aisé de noter que le mode $\eta_e > 0$ est unique pour la modélisation B_6 , et que les modélisations B_4 et B_5 possèdent au moins un mode strictement positif. Par ailleurs, on note que la distribution M' de η peut être unimodale en une valeur strictement positive cependant que la distribution M de X n'a pas de mode défini (par exemple quand $a < \beta_l^{-1}$). \square

3.6.2.2 Intervalle de crédibilité.

Dans le cas où un intervalle de crédibilité $[\eta_1, \eta_2]$ d'ordre (α_1, α_2) est disponible, nous suggérons de choisir un $b = b(a, \beta)$ unique comme la solution explicite d'un problème de moindres carrés (proposition 11), qui minimise la distance entre quantiles souhaités et quantiles effectifs. Ainsi l'expression de b réalise un consensus pratique d'utilisation des deux connaissances. L'ajout d'une pondération privilégiant l'un ou l'autre des quantiles est là encore possible. Une telle proposition n'est pas facilement adaptable au cas où les quantiles sont marginaux en X , les quantiles de M n'étant pas explicites.

PROPOSITION 11. Pour $i = 1, 2$, notons $\eta_i(b) = (2b/\chi_{2\alpha_i}^2(1 - \alpha_i))^{1/\beta}$ et $b_i = \eta_i^\beta \chi_{2\alpha_i}^2(1 - \alpha_i)/2$. Soit le problème des moindres carrés

$$b^* = \arg \min_{b \in [b_1, b_2]} \left\{ \left(1 - \frac{\eta_1(b)}{\eta_1}\right)^2 + \left(1 - \frac{\eta_2(b)}{\eta_2}\right)^2 \right\}.$$

Alors b^* existe, est unique et s'écrit $b^* = \left(\frac{b_1^{-1/\beta} + b_2^{-1/\beta}}{b_1^{-2/\beta} + b_2^{-2/\beta}} \right)^\beta$.

Preuve. Sachant β , pour $i \in \{1, 2\}$, notons $\delta_i = \eta_i(b)/b^{1/\beta}$ et effectuons le changement de variable $x = b^{1/\beta}$. La fonction à minimiser sur le domaine $[b_1^{1/\beta}, b_2^{1/\beta}]$ est

$$f(x) = \eta_1^{-2}(\eta_1 - \delta_1 x)^2 + \eta_2^{-2}(\eta_2 - \delta_2 x)^2.$$

Alors $\frac{\partial f}{\partial x}(x)$ est strictement croissante sur $[b_1^{1/\beta}, b_2^{1/\beta}]$ car $\frac{\partial^2 f}{\partial x^2}(x) = 2(\delta_1^2/\eta_1^2 + \delta_2^2/\eta_2^2) > 0$. En remarquant que $b_i^{1/\beta} = \eta_i/\delta_i$, on constate alors que $\frac{\partial f}{\partial x}(b_1^{1/\beta}) \frac{\partial f}{\partial x}(b_2^{1/\beta}) < 0$. La fonction $\frac{\partial f}{\partial x}$ possède donc un zéro unique sur $[b_1^{1/\beta}, b_2^{1/\beta}]$ qui définit le minimum unique de f dans l'intervalle. \square

Measuring agreement between prior and data

Dans ce chapitre, nous définissons et faisons l'étude d'un critère statistique noté DAC qui détecte un conflit entre une connaissance modélisée par un *a priori* π et les données y_n . En effet, l'inférence bayésienne peut se révéler dangereuse si ce conflit n'est pas détecté auparavant. Ces conflits peuvent survenir pour des raisons d'hétérogénéité de l'expertise ou, en fiabilité, à cause de décalages temporels entre données de durée de vie et expertise tenant compte de l'évolution technique. Dans le cadre d'une étude probabiliste de sûreté, l'utilisation du critère DAC fait surtout sens lorsqu'il participe d'une analyse *fiabiliste* non prévisionnelle; plus de précisions à ce sujet sont fournies dans la Présentation générale de la thèse. Voir par ailleurs Billy *et al.* (2006) pour un court article en français reprenant les principaux points développés dans ce chapitre.

Si donner un sens à la notion de conflit entre *a priori* et données ne se justifie pas d'un point de vue théorique, l'intérêt de ce critère est cependant manifeste lors d'une analyse bayésienne concrète, où une connaissance subjective est utilisée. Le critère DAC (pour *data-agreement criterion*) est un outil que l'on perçoit comme complémentaire de la seule approche statistique publiée sur la question (Evans & Moshonov 2005a). Il a cependant des propriétés qui en font un outil plus souple, et qui a en sus des vertus de calibration lorsque peu ou pas d'information *crédible* n'est disponible sur l'incertitude de l'expertise.

Local contents

4.1	Motivation	108
4.2	Previous works	109
4.3	The DAC criterion	111
4.4	Examples of ideal cases	114
4.5	An intrinsic adaptation of DAC	117
4.6	Prior calibration with DAC	126
4.7	A possible improvement	129
4.8	Conclusion	131
4.9	Appendix	131

4.1 Motivation

Among the articles on subjective Bayesian inference for estimating the parameter $\theta \in \Theta$ of a decision-making model $\mathcal{M}(\theta)$, many of them mention the undesirable behavior of the posterior odds when the prior information is badly conditioned. By this term we encompass a wrong information given by expert opinions (or any other source of subjective information) and a prior density $\pi(\theta)$ far from the frequentist confidence region on parameter θ brought by observed data $\mathbf{y}_n = (y_1, \dots, y_n)$. The objective knowledge of the reality yielded by the likelihood $\mathcal{L}(\mathbf{y}_n; \theta)$ can thus be littered by the choice of a wrong prior and the Bayesian analyst can take a posterior decision with unwelcome consequences. In most cases, these articles are centered on a reinforcement of the posterior robustness. However, a first difficulty is just checking that the available knowledge is not self-contradictory, before the inference. Note that in case of highly censored data and small sample sizes, when the Bayesian informative approach is very recommended (Robert 2001), the heterogeneity of the data or a high level of censoring can have the same effect as a wrong prior with correct data. A prior can be well chosen but the posterior threatens to lead to bad decisions. Industrial examples can be taken in Bonneville & Billy (2006).

Then, prior knowledge and real knowledge can have symmetric roles in a conflict. Curiously, this possible conflict seems not to have been much studied although it constitutes, from our point of view, an important preliminary to Bayesian subjective inference. Notice however that in the strict observation of the Bayesian paradigm, the prior distribution and the sampling (parametric) model constitute a whole decision model and it seems not relevant to determine degrees of discrepancy between them from a theoretical point of view. The primary context of such a study is the practical impact of Bayesian estimation by engineers when the data are poorly informative.

Numerous connected works deal rather with robustness problems. Symmetrically, functional structures of the prior and the sampling model have been studied to obtain negligible posterior influence when a possible discrepancy is increasing, knowing one the two sources of information has to be rejected. Main references are De Finetti (1961), Dawid (1982), Hill (1974) Lucas (1993) and especially O'Hagan (1979, 1988, 1990, 2003), Andrade & O'Hagan (2006) and Angers (2000). Some authors as Gelman *et al.* (1996) consider the simultaneous effects of the sampling model and the prior on a possible discrepancy with the data. First, Evans & Moshonov (2005a) have efficiently raise the problem: a primary issue is to check the goodness of the sampling model $\mathcal{M}(\theta)$ with respect to \mathbf{y}_n . Appropriate Bayesian methods have been developed in Bayarri & Berger (2000) in this way. Secondly, if no evidence of error is obtained, we have to check a possible prior-data conflict. In our industrial context, the sampling model is usually well-trying and Bayesian technics are used to overcome precision limits of the frequentist approach.

As we said, few work has been done on the subject. Evans & Moshonov (2005a, 2005b) are the main references dealing precisely with this topic. They compute the P -value of an observed sufficient statistic with respect to its marginal (predictive) distribution. In industrial studies, Idée *et al.* (2001) have suggested to use a Fisher test between prior and empiric measures of uncertainty and an *ad hoc* approach has been proposed by Usureau (2001) without statistical justification. In Section 4.2, first and third approaches are analyzed.

As an alternative approach, we propose the following criterion to detect a possible conflict. Com-

pute the ratio

$$\text{DAC}^J(\pi|\mathbf{y}_n) = \frac{KL(\pi^J(\cdot|\mathbf{y}_n) \parallel \pi)}{KL(\pi^J(\cdot|\mathbf{y}_n) \parallel \pi^J)}$$

where $KL(\pi_1|\pi_2)$ is the Kullback-Leibler divergence between distributions π_1 and π_2

$$KL(\pi_1|\pi_2) = \int_{\Theta} \pi_1(\theta) \log \frac{\pi_1(\theta)}{\pi_2(\theta)} d\theta$$

and π^J is ideally a noninformative prior. DAC can be regarded as a *data-agreement criterion*. If $\text{DAC}^J(\pi|\mathbf{y}_n) \leq 1$, prior and data-given confidence regions for θ are close enough and the prior proposal π is in agreement with the \mathbf{y}_n . Else a conflict is detected. This definition is easily transposable to any function of interest $g(\theta)$. Then the criterion should be computed with the priors on $g(\theta)$. Using the KL-divergence ensures the independence of the result to any bijective reparametrization.

In Section 4.3, we detail the arguments which lead to the choice of this criterion and give some properties for its use when the prior is hierarchically specified or is defined as a combination of priors. The difficulty raised by its definition is its application to general cases. Some ideal cases are firstly treated in Section 4.4. In a more general framework, we need to adjust DAC since π^J can be improper.

Linking our difficulties to problems encountered in Bayesian model selection, we define in Section 4.5 an *intrinsic* form of the criterion. Usual tools of model selection are used, as *minimal training samples* and *posterior priors* (Berger & Perrichi 1996, 2002, Pérez 1998). In this section, we make some comparisons between the ideal and the adjusted form of the criterion. The vision of DAC as a possible tool for the calibration of subjective priors, proposing objective variance bounds or default values, is the subject of Section § 4.6. It is one of the main differences with the procedure proposed by Evans & Moshonov (2005a). Finally, we consider some issues due to the frequent need of adapting DAC and propose some research clues for future work.

Besides, along the sections, examples will put into comparison the results obtained by their method and DAC. Especially, the exponential and Weibull models will be our favorite examples since the detection of conflict between prior and data is of particular interest in reliability. Our contributions and a global comparison with the P -value procedure are summed up in the conclusion section. We think that both methods are complementary and useful since they give, for the Bayesian analyst, a precise view of the prior information in terms of location and uncertainty on the parameter space with respect to the data information. Thus the Bayesian analyst, in particular when subjective priors have been elicited, can check and possibly modify the prior.

4.2 Previous works

Two points of view have been considered in the recent years. The first one is coming from the engineering community and suggests to use a simple rule of thumb while the second approach has a statistical foundation. Firstly, some notations are introduced.

For $n \in \mathbb{N}^*$, let $\mathbf{X}_n = X_1, \dots, X_n \sim \mathcal{M}(\theta)$ be independently and identically distributed real-or-vector-valued random variables in the sample space χ^n with a probability density function (pdf) $f(x|\theta)$ and a survival (reliability) function $S(x|\theta)$. This pdf is defined with respect to a dominating

measure which is usually Lebesgue. However, some of our examples will use discrete measures. Let $\theta \in \Theta$ and denote $d = \dim \Theta$. Let $\mathbf{y}_n = (y_1, \dots, y_n)$ be an observed sample of n data. Usually we will consider that \mathbf{y}_n contains r uncensored i.i.d. data $\mathbf{x}_r = (x_1, \dots, x_r)$ following $\mathcal{M}(\theta)$ and $n - r$ fixed (type-I) right-censored values, denoted $\mathbf{c}_{n-r} = (c_1, \dots, c_{n-r})$. Thus, the observed likelihood can be written as

$$\mathcal{L}(\mathbf{y}_n; \theta) = \prod_{i=1}^r f(x_i | \theta) \prod_{j=1}^{n-r} S(c_j | \theta).$$

4.2.1 An intuitive tool

Usureau (2001) suggests an intuitive vision of the issue, especially in the case where $\mathcal{M}(\theta)$ is a location-scale model. Typically, $\mathcal{M}(\theta)$ is in the natural exponential family (NEF). He defines a convolution product between the prior and the likelihood

$$\begin{aligned} \delta_{\pi, \mathcal{L}}(\omega) &= \int_{-\infty}^{\infty} \pi(\theta - \omega) \mathcal{L}(\mathbf{y}_n; \theta) d\theta, \\ &= \int_{-\infty}^{\infty} \pi(\theta) \mathcal{L}(\mathbf{y}_n; \theta - \omega) d\theta \end{aligned}$$

whose the value increases with the shape similarity of those functions of θ . A maximal value $\delta_{\pi, \mathcal{L}}^m$ is attained in ω^* . When both functions are unimodal, this maximization is realized when the two modes are equalized. Then a “data-agreement” percentage is deduced as $\delta_{\pi, \mathcal{L}}(0) / \delta_{\pi, \mathcal{L}}^m$. Clearly, this criterion has no statistical meaning since it considers implicitly and wrongly that the likelihood function $\theta \mapsto \mathcal{L}(\mathbf{y}_n, \theta)$ is a density on θ . Moreover, a “data-agreement” percentage does not provide a calibrated criterion.

4.2.2 A marginal viewpoint

Evans and Moshonov (2005a) provide a statistical theory and several examples for the detection of data-prior conflict. They consider the marginal prior distribution M_T of a *minimal sufficient statistic* T with sampling density $f_T(t|\theta)$. This distribution has density

$$m_T(t) = \int_{\Theta} \pi(\theta) f_T(t|\theta) d\theta.$$

If the observed statistic $t_o = t(\mathbf{y}_n)$ is a surprising value for M_T , namely when the marginal P -value

$$F_T(t_o) = M_T \{m_T(t) \leq m_T(t_o)\}$$

is extreme, then a conflict is detected. This typically occurs when $F_T(t_o) \leq 0.05$ or $F_T(t_o) \geq 0.95$. In this case, “the data provide little or no support to those values of θ where the prior places its support”. In next example we derive $F_T(t_o)$ for a simple exponential model, when data are censored.

EXAMPLE 11. *Exponential model.*

We suppose to have $\mathbf{y}_n = (\mathbf{x}_r, \mathbf{c}_{n-r})$ where $\mathbf{x}_r = (x_1, \dots, x_r)$ are uncensored i.i.d. data with density $f(x|\lambda) = \lambda \exp(-\lambda x)$ for $x \geq 0$ and \mathbf{c}_{n-r} is a set of $n - r$ censored data. A minimal sufficient statistic is $t(\mathbf{y}_n) = r^{-1} \sum_{i=1}^n y_i$. From Sundberg (2001), we obtain $T \sim \mathcal{G}(p, q\lambda)$ where $p = q = r$ in case of

type-II censure (or in the uncensored case) and $p = r$, $q = r + 1/2$ in case of type-I censure (Cox non-asymptotic approximation). We place on λ a conjugate $\mathcal{G}(a, b)$ prior distribution, where $a, b > 0$. Then

$$F_T(t_o) = \frac{\Gamma(p+a)}{\Gamma(p)\Gamma(a)} \int_0^{\frac{at_o}{b}} \frac{u^{p-1}}{(1+u)^{p+a}} du. \blacksquare$$

However, the sense given to a necessary threshold couple of P -values is not clear. Why should we choose (5%, 95%) quantiles rather than (2%, 98%)? Besides, this check is affected by ancillarity, since the marginal distribution threatens to reflect more the behavior of the sampling model $f_T(\cdot|\theta)$ than the prior distribution. Denoting U an ancillary statistic, namely a statistic whose the distribution does not depend on θ (Ghosh *et al.*, 2005), the authors propose to compute the P -value of the conditional marginal distribution

$$m_T(t|U) = \int_{\Theta} \pi(\theta) f_T(t|\theta, U) d\theta$$

to remove the influence of the sampling model in the check. Clearly, this check works well for NEF models and $F(t_o)$ is often explicit when conjugate priors are used. However, Weibull distributions (for instance) are problematic, since any minimal sufficient statistic is the whole sample. Then the numerical cost of the computation of $F_T(t_o)$ can be high.

4.3 The DAC criterion

4.3.1 The formal definition

Denote π^J a noninformative prior with good frequentist coverage of the Bayes posterior intervals: π^J is noninformative and its posterior density $\pi^J(\cdot|\mathbf{y}_n)$ yields posterior confidence regions with an accurate frequentist validity. Such a prior is said a *coverage matching prior* (CMP). This notion of “non-informativeness” has attracted attention from numerous authors. Note that there exist prior distributions giving optimal estimators according to frequentist criteria (as minimaxity), but such a restriction usually does not lead to unicity when such priors exist, except when invariance structure are required; in this case the right-Haar measure is the appropriate choice. See Robert (2001, Chap. 2 & 8) for more explanations. The formal definition of CMP priors and related references can be found in Kass & Wasserman (1996), Datta (1996) and Ghoshal (1999) for regular and non-regular cases. We remind this definition hereafter.

DEFINITION 5. Let $\theta_n(\alpha)$ be the posterior α -quantile of θ based on the n observations $\mathbf{X}_n = (X_1, \dots, X_n)$, i.e. $P(\theta \leq \theta_n(\alpha)|\mathbf{X}_n) = \alpha$. A prior is said to be an i th order coverage matching prior (CMP) for θ if

$$P_\theta(\theta \leq \theta_n(\alpha)) = P(\theta \leq \theta_n(\alpha)|\mathbf{X}_n) + \mathcal{O}(n^{-i/2}),$$

where the left-hand side is the frequentist probability, where θ is fixed and $\theta_n(\alpha)$ is random.

Ideally, choosing π^J as a CMP with the highest available order is desirable. It can be generally defined as the solution of a partial differential equation. When $d = \dim \Theta = 1$, Welch & Peers (1963) proved that the Jeffreys prior is a second-order matching prior and Clarke (1996) gives a

straightforward defence of this choice. In multiparameters problems, the Berger-Bernardo (1989, 1992) *reference priors* are well known to work better. Especially, the *backward* reference priors are usually matching when interest and nuisance parameters are orthogonal (in the sense of the Fisher information).

Here, we consider implicitly that $\pi^J(\theta)$ can be defined such that $\pi^J(\theta|\mathbf{y}_n)$ is proper or that there are enough available data \mathbf{y}_n . Indeed, there exist models for which that such π^J does not exist, for any n . Competing models are often in this case (Berger & Sun 1993). Then, we suggest to work with each submodel, assuming that priors are independent.

It seems reasonable that π^J , as a noninformative prior, should be considered as in agreement with \mathbf{y}_n . Besides, $\pi^J(\theta|\mathbf{y}_n)$ mimics the likelihood $\mathcal{L}(\mathbf{y}_n; \theta)$. Bernardo (1997) informally proposed to use this density as a benchmark prior for the study of subjective assessments.

In a more formal viewpoint, we propose to consider $\pi^J(\theta|\mathbf{y}_n)$ as a data-dependent prior, which fixes a limit to the acceptable information yielded by an expert (independent from the data). Our idea is to see $\pi^J(\theta|\mathbf{y}_n)$ as the prior density on θ of a *fictitious expert perfectly in agreement with \mathbf{y}_n* . Then we define an *informative regret* between this ideal prior $\pi^J(\cdot|\mathbf{y}_n)$ and the assessed prior π by

$$KL(\pi^J(\cdot|\mathbf{y}_n) \parallel \pi). \quad (4.1)$$

Indeed, $KL(\pi_1 \parallel \pi_2)$ states the regret due to the choice of π_2 when the true distribution is π_1 . Notions of regret with respect to information-theoretic tools as the KL-divergence can be found in Cover and Thomas (1991) and Sweeting, Datta and Ghosh (2005). If π is such that (4.1) is large, $\pi(\theta)$ will be considered as too far from the data information on θ .

Now it is necessary to choose a constant C such that if $KL(\pi^J(\cdot|\mathbf{y}_n) \parallel \pi) > C$, π is declared in conflict with the data. The rationale for choosing C is as follows: π is supposed to be informative. If $KL(\pi^J(\cdot|\mathbf{y}_n) \parallel \pi) > KL(\pi^J(\cdot|\mathbf{y}_n) \parallel \pi^J)$, it cannot mean that π is less informative than π^J which is the less informative prior. Therefore it means that π is far from $\pi^J(\cdot|\mathbf{y}_n)$, which implies that the informations provided by π and \mathbf{y}_n are not in accordance. Consequently, it leads to choose $C = KL(\pi^J(\cdot|\mathbf{y}_n) \parallel \pi^J)$. We obtain finally the normalized criterion

$$DAC^J(\pi|\mathbf{y}_n) = \frac{KL(\pi^J(\cdot|\mathbf{y}_n) \parallel \pi)}{KL(\pi^J(\cdot|\mathbf{y}_n) \parallel \pi^J)} \quad (4.2)$$

and the following rule: π is said *conflicting* with \mathbf{y}_n if $DAC^J(\pi|\mathbf{y}_n) > 1$. The superscript J reflects the careful choice of π^J . In the following, we omit sometimes this notation for simplicity.

Clearly, Definition 4.2 is not restricted to the KL-divergence and another choice in the Ali-Silvey class of information-theoretic measures (Ali and Silvey, 1966) can be made. However, as said before, a very important feature of the KL-divergence is its invariance to reparametrisation. Justifications for using it preferentially can be found in Hartigan (1998) and Sinanović and Johnson (2003) (from both computational and analytic viewpoints). Notice however that our definition is restricted to finite-dimensional models. When the model has an infinite number of parameters, the KL-divergence is unbounded and can not be used.

4.3.2 Main features

1. Following Evans & Moshonov (2005b), we desire to check a potential conflict with the data when the prior is hierarchically specified. A key requirement (and an uncomfortable restriction) for the use of their P-value procedure is the existence of statistics which are ancillary for parts of the parameter. For the DAC criterion, the following proposition makes easy the separate checks of the hierarchical levels and links them to the check of the whole prior. Thus, the agreement of the full prior is not a sufficient condition to obtain the agreement of any hierarchical prior. The opposite is clearly true. The proof is obvious and not reported here. We shall see in Section 4.6 the interest of this feature for prior calibration.

PROPOSITION 12. Hierarchically specified priors. Assume that π and π^J can be written $\pi(\theta) = \pi(\theta_1|\theta_2)\pi(\theta_2)$ and $\pi^J(\theta) = \pi^J(\theta_1|\theta_2)\pi^J(\theta_2)$. Denote $\tilde{\pi}_1(\theta) = \pi(\theta_1|\theta_2)\pi^J(\theta_2)$ and $\tilde{\pi}_2(\theta) = \pi^J(\theta_1|\theta_2)\pi(\theta_2)$. Then

$$DAC(\pi|\mathbf{y}_n) = DAC(\tilde{\pi}_1|\mathbf{y}_n) + DAC(\tilde{\pi}_2|\mathbf{y}_n) - 1. \quad (4.3)$$

2. Suppose to have checked m priors π_1, \dots, π_m . Next proposition indicates that any geometric weighted combination of the priors π_i does not need to be checked if all priors are in agreement with the data. Possibly, some priors can be conflicting while the combination stays in agreement. Such a combination is typically used to obtain a global prior when several independent experts are available. See § 3.3.5.1 for more explanations. The proof is given in Appendix § 4.9.1.

PROPOSITION 13. Geometric combination of multiple priors. Let π_1, \dots, π_m be m priors on Θ , and let $\alpha_1, \dots, \alpha_m$ be m weights such that $0 < \alpha_i < 1$, $\sum_{i=1}^m \alpha_i = 1$. Denote $\pi(\theta) \propto \prod_{i=1}^m \pi(\theta)^{\alpha_i}$. Then

$$DAC^J(\pi|\mathbf{y}_n) \leq \sum_{i=1}^m \alpha_i DAC^J(\pi_i|\mathbf{y}_n).$$

3. *Asymptotic behavior.* DAC could be expected to tend to 1 when n takes large values. Actually, as Hartigan says (1983): “the prior density does not affect the asymptotic distribution of θ in the terms of $O(1)$ or $O(n^{-1/2})$ ”. Under classical regularity conditions on prior densities and likelihood, the asymptotic normality of $\pi^J(\cdot|\mathbf{y}_n)$ should make numerator and denominator of (4.2) take close values. This asymptotic behavior of DAC will be illustrated with an example in the following.

4.3.3 Difficulties of computation

In next section, we shall consider that π^J is proper (i.e., $\int_{\Theta} \pi(\theta) d\theta < \infty$). Typically, it is the case when Θ is discrete or bounded. But usually π^J is improper. In this case, the denominator of DAC is defined up to an unknown additive constant and we need to adapt the criterion. Calling arguments originally used in Bayesian model selection, this adaptation will be the subject of Section 4.5.

4.4 Examples of ideal cases

4.4.1 Discrete models

We begin with two discrete distributions for which π^J is proper. Note that in these two cases, DAC is explicit. The KL-divergences considered here are available in Appendix § 4.9.2. The first example, about a Bernoulli distribution, is taken from Evans & Moshonov (2005a) to compare both approaches. Firstly we use the same numerical values and obtain the same conclusions. Secondly, we show that the P -value procedure can accept priors which are rejected by DAC. Thirdly, in an illustrative purpose, we propose a short study of the asymptotic behavior of DAC. As we guessed, any conflict disappears asymptotically. The second example deals with a multinomial distribution. It reflects the need to choose carefully π^J . In addition, we focus on the evolution of DAC when the prior mean is well centered on the data but the prior variance is modified. A first vision of DAC as a calibration tool appears as a difference with the P -value procedure proposed by Evans & Moshonov.

EXAMPLE 12. *Bernoulli model.*

Let \mathbf{y}_n be an *i.i.d.* n -sample from a Bernoulli distribution $\mathcal{B}_r(\theta)$. We assume on θ a prior Beta distribution $\mathcal{B}_e(\alpha, \beta)$ on $[0, 1]$. Note $\delta_n = \sum_{i=1}^n y_i$. The Jeffreys prior π^J is the $\mathcal{B}_e(1/2, 1/2)$ density. Then $\pi^J(\theta|\mathbf{y}_n) = \mathcal{B}_e(\delta_n + 1/2, n - \delta_n + 1/2)$. In the same conditions as in Evans & Moshonov (2005a, Ex.2), we choose $(\alpha, \beta) = (5, 20)$ (so $E[\theta] = 0.2$) then we generate a sample of size $n = 10$ from the $\mathcal{B}_r(\theta_0 = 0.9)$ distribution. We obtain $\text{DAC}^J(\alpha, \beta|\mathbf{y}_n) = 1.102$ and we conclude to a prior-data conflict. Modifying $\theta_0 = 0.25$, we have $\text{DAC}^J(\alpha, \beta|\mathbf{y}_n) = 0.1026$ which means (as expected) an agreement. Choosing $(\alpha, \beta) = (1, 1)$ (π is thus the uniform prior) leads to the agreement for any Bernoulli dataset. Thus we obtain the same conclusions as Evans & Moshonov (2005a).

We choose now $\theta_0 = 0.7$ and two sizes $n = 10$ and $n = 5$. Fixing the prior standard deviation to 0.2, we are looking for the range of value for the prior mean $\mu_0 = \alpha/(\alpha + \beta)$ such that π is not conflicting. We obtain $\mu_0 \in [0.2, 0.96]$ for $n = 10$ and $\mu_0 \in [0.23, 0.94]$ for $n = 5$ using DAC. By the P -value procedure, the 10%-90% percentile domain is $\mu_0 \in [0.054, 0.928]$ for $n = 10$ and $\mu_0 \in [0.056, 0.925]$ for $n = 5$. The 5%-95% domain remains close to $[0.048, 0.95]$. Thus, DAC is clearly more restrictive.

On the asymptotic side, denoting $0 < \theta_0 < 1$ the true value of the parameter, we obtain the following proposition which confirm that $\text{DAC} \rightarrow 1$ when n takes large values (the proof is given in Appendix § 4.9.1). The convergence is very slow because of logarithmic terms in the asymptotic development of $\text{DAC}^J(\alpha, \beta|\mathbf{y}_n)$. In Table 4.1, mean values of DAC^J are given for several values of n and (α, β) , when $\theta_0 = 0.9$.

PROPOSITION 14. $\forall(\alpha, \beta) > 0$, for any $0 < q < 1$, we have

$$(\log n)^q E \left[\text{DAC}^J(\alpha, \beta|\mathbf{Y}_n) - 1 \right] \xrightarrow{n \rightarrow \infty} 0.$$

Denoting Ψ the digamma function and $A_{\theta_0}(\alpha, \beta) = \Psi(1/2) - \Psi(\beta) + \theta_0 \{\Psi(\beta) - \Psi(\alpha)\}$, the limit value 1 is approached by upper or lower values as follows:

1. if $\alpha = \beta$, the limit is attained by lower (*resp.* upper) values if $\alpha > 1/2$ (*resp.* $\alpha < 1/2$),

n	(α, β)			
	(0.1,0.1)	(0.1,10)	(10,0.1)	(2,2)
10	3.08	3.15	0.46	0.40
10^2	2.30	2.09	0.53	0.62
10^3	1.95	1.80	0.67	0.73
10^5	1.62	1.53	0.78	0.82
10^{10}	1.34	1.29	0.88	0.90
10^{20}	1.18	1.15	0.94	0.96
10^{50}	1.05	1.03	0.98	0.99

Table 4.1: Convergences of DAC^J to 1 when $\theta_0 = 0.9$ (averaged on 100 samples).

2. if $\alpha \neq \beta$, the limit is attained by lower values (*resp.* upper) values if $A_{\theta_0}(\alpha, \beta) < 0$ (*resp.* $A_{\theta_0}(\alpha, \beta) > 0$),
3. if $\alpha \neq \beta$ and $A_{\theta_0}(\alpha, \beta) = 0$, $\lim_{n \rightarrow \infty} P_{\theta_0} \{DAC^J(\alpha, \beta | \mathbf{y}_n) - 1 \leq 0\} = 1/2$.

Condition $A_{\theta_0}(\alpha, \beta) \geq 0$ is equivalent to say that π is accumulating its mass on the bounds of the domain $[0, 1]$, despite the prior mean can sometimes be 0.5 (when $\alpha = \beta < 1/2$). Some examples are displayed on Figure 4.1. A more usual case is when α (*resp.* β) takes high values while β (*resp.* α) remains small. Then the prior mean is very close to 0 or 1. Thus, when n is fixed, these unacceptable priors are always seen conflicting by the DAC criterion. ■

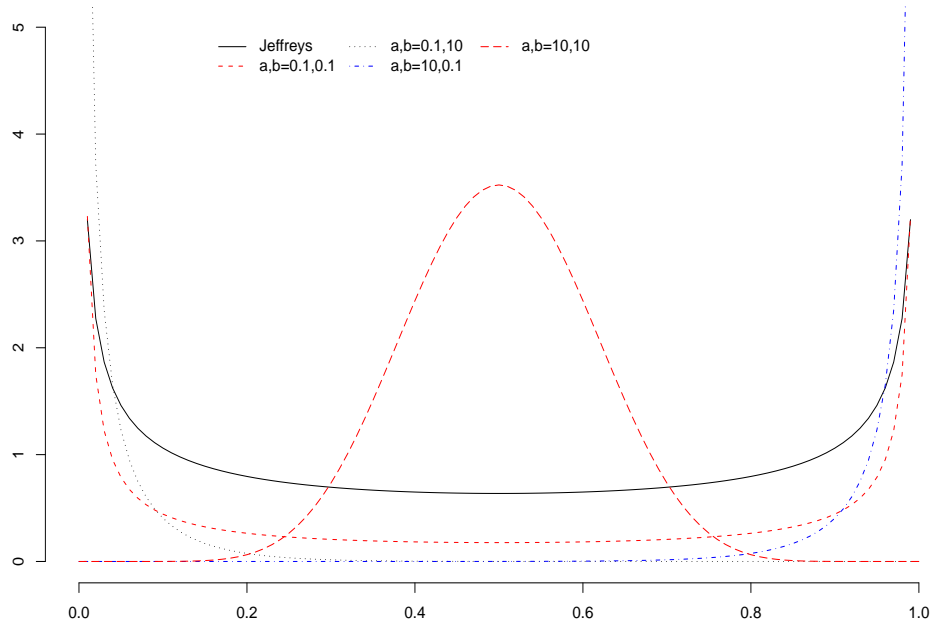


Figure 4.1: $\mathcal{B}_e(\alpha, \beta)$ densities.

EXAMPLE 13. *Multinomial model.*

Let $\mathbf{y}_n = (9, 5, 1)$ be an *i.i.d.* sample from the multinomial distribution $\mathcal{M}(\mathbf{p})$ with $\mathbf{p}_0 = (0.7, 0.25, 0.05)$ the true value of \mathbf{p} . Suppose having chosen $\pi(\mathbf{p})$ as a Dirichlet $Dir(\alpha_1, \alpha_2, \alpha_3)$ density. We consider some triplet values $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ such as $E[\mathbf{p}] = \mathbf{p}_0$, with increasing variance. So that prior

and data informations seem intuitively close. We thus choose $\alpha = \rho(7, 2.5, 0.5)$. The prior variance of \mathbf{p} decreases when ρ increases. We compare on Figure 4.2 the evolutions of DAC^J when π^J is chosen as the multivariate uniform prior and the most appropriate Jeffreys prior $\text{Dir}(1/2, 1/2, 1/2)$ (note however that in this multidimensional case the Jeffreys prior has not the best posterior coverage properties among available noninformative priors ; see Yang & Berger 1997 for alternative choices).

There are small differences between two choices of π^J : Jeffreys and uniform prior means are the same and the Jeffreys variance is quite slightly larger than the uniform variance (less than 10^{-2} upper in L^2 norm). However, using the uniform distribution can lead to refuse vague priors (when $\rho \rightarrow 0$) which are less informative than π^J , unlike the Jeffreys prior. This is a key argument to favor the Jeffreys prior. Secondly, note that data and prior agree only when the prior is not too informative. In this latter case data have a negligible influence on the posterior distribution. This behavior makes a strong difference with the P -value procedure by Evans & Moshonov (2005a), which can never detect any conflict even if π tends to a Dirac prior. This feature of DAC is analysed in the following and takes a special interest for the prior calibration. ■

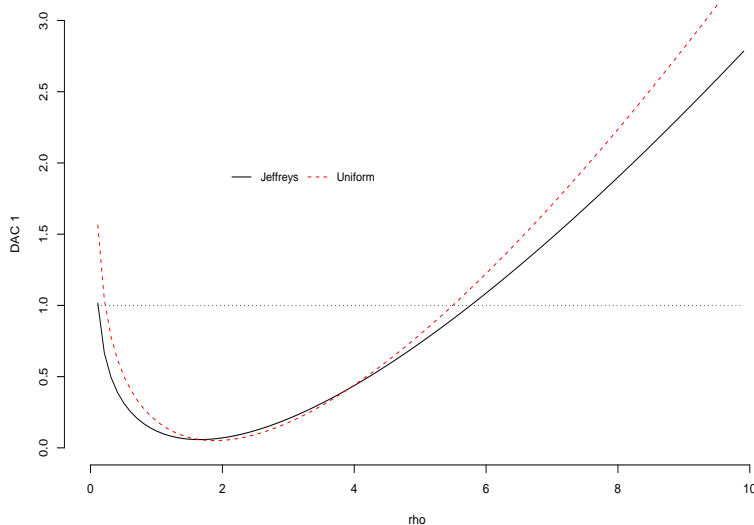


Figure 4.2: Evolution of DAC^J with respect to ρ (the uncertainty hyperparameter of the Dirichlet prior of a multinomial distribution) for the choices $\pi^J =$ multivariate uniform and Jeffreys.

4.4.2 A continuous bounded model

In our opinion, checking for data-prior conflict is of particular interest in industrial Bayesian studies. In a reliability context, the parameter space Θ can often be bounded. Usual arguments are physical constraints. In this case, π^J is proper. Next example illustrates the intuitive behavior of DAC and its stability with respect to the bounds of Θ when π is strongly informative.

EXAMPLE 14. *Location normal model.*

Let \mathbf{y}_n be an uncensored sample from a $\mathcal{N}(\theta, 1)$ distribution with $\theta \in D = [T_l, T_m]$. Denote θ_0 the real value of θ . We place on θ a $\mathcal{N}(\mu_0, \sigma_0^2)$ prior. π^J is here the uniform prior $\pi^J(\theta) = (T_m - T_l)^{-1} \mathbb{1}_D(\theta)$. Then $\pi^J(\theta|\mathbf{y}_n)$ is the $\mathcal{N}(\bar{y}_n, 1/n)$ density restricted on D . See Appendix p. 135 for a simple writing of

DAC^J. Choosing $\theta_0 = 0$, $n = 5$, $T_m = -T_l = 15$ and $\sigma_0 = 1$, we firstly consider the evolution of DAC with respect to μ_0 in Figure 4.3, for several values of σ_0 . A symmetric evolution around θ_0 appears natural, while the length of the agreement domain decreases when σ_0 increases. On Figure 4.4 we modify (T_l, T_m) , shrinking and enlarging symmetrically the prior domain. The agreement domain for μ_0 appears the more sensitive to a modification of the bounds as σ_0 takes high values. ■

4.5 An intrinsic adaptation of DAC

In numerous cases, π^J is improper. Here we focus especially on the general case where π^J is defined up to an unknown multiplicative constant c_j . This constant has an additive effect in the denominator of DAC

$$KL(\pi^J(\cdot|\mathbf{y}_n) \parallel \pi^J) = -\log c_j + KL(\pi^J(\cdot|\mathbf{y}_n) \parallel \pi'^J)$$

where $\pi'^J(\theta) \propto \pi^J(\theta)$. Thus DAC^J is not well defined. In this section, we firstly link this difficulty to an issue which is typically encountered in Bayesian model selection. Some proposals are done to get around this issue, including the *intrinsic heuristic*. Thus we propose an intrinsic version DAC^{AIJ} of DAC^J. This adaptation is compared to the ideal case through examples and a case study involving real Weibull data and industrial expert opinions is considered. The issues raised by the adaptation are highlighted too. A possible improvement will be emphasized in subsection 4.7.

4.5.1 Introducing the intrinsic heuristic

4.5.1.1 Link with Bayesian model selection

First we show how the additive constant c_j , can be regarded as the multiplicative constant of a Bayes factor. Formally, we must have

$$\exp\{KL(\pi^J(\cdot|\mathbf{y}_n) \parallel \pi) - KL(\pi^J(\cdot|\mathbf{y}_n) \parallel \pi^J)\} \leq 1. \quad (4.4)$$

Denote the marginal measures $m^J(x) = \int_{\Theta} f(x|\theta)\pi^J(\theta) d\theta$ and $m(x) = \int_{\Theta} f(x|\theta)\pi(\theta) d\theta$. The left member of inequation (4.4) takes the form

$$m^J(\mathbf{y}_n) \exp\left\{\int_{\Theta} \pi^J(\theta|\mathbf{y}_n) \log \frac{\pi^J(\theta|\mathbf{y}_n)}{\pi(\theta)f(\mathbf{y}_n|\theta)} d\theta\right\} \quad (4.5)$$

which does not need the computation of the posterior $\pi(\cdot|\mathbf{y}_n)$ density. Then we can write the criterion under the alternative form

$$DAC_2^J(\pi|\mathbf{y}_n) = B_{J,\pi}(\mathbf{y}_n) \exp(KL\{\pi^J(\cdot|\mathbf{y}_n) \parallel \pi(\cdot|\mathbf{y}_n)\}) \quad (4.6)$$

where $B_{J,\pi}(\mathbf{y}_n)$ is the Bayes factor

$$B_{J,\pi}(\mathbf{y}_n) = \frac{m(\mathbf{y}_n)}{m^J(\mathbf{y}_n)}.$$

Thus, the rule consists of accepting π as in agreement with the data \mathbf{y}_n if $DAC_2^J(\pi|\mathbf{y}_n) \leq 1$.

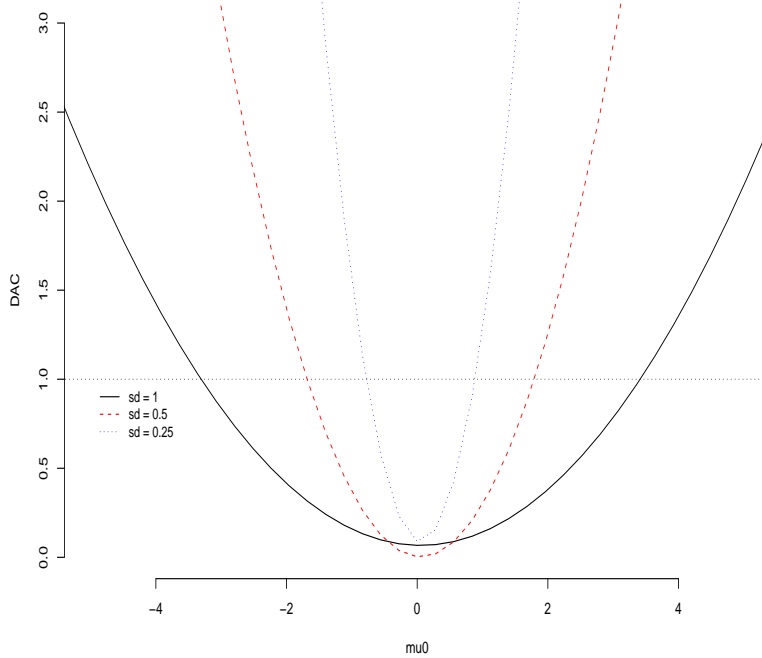


Figure 4.3: Mean evolution of DAC^J with respect to parameters μ_0 and σ_0 of a prior gaussian distribution on the mean of a bounded gaussian model (averaged on 30 samples).

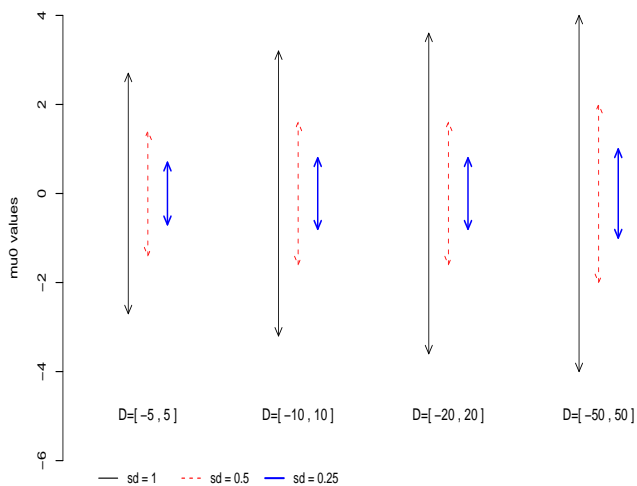


Figure 4.4: Agreement domains for the prior mean μ_0 with respect to the prior standard deviation σ_0 and some modifications of the bounded domain $D = [T_l, T_m]$ of the gaussian sampling model.

There is some interest to interpret c_j as the multiplicative unknown constant of a Bayes factor. Indeed, it is a typical issue in objective model selection about which a huge literature is dedicated. See Kass & Raftery (1995) and Andrieu *et al.* (2001) for a review. Numerous approaches are proposed to obtain default Bayes factors, and the most famous is the definition of *intrinsic Bayes factors* (IBF, Berger & Perrichi 1996).

This methodology is deeply linked to the notion of *minimal training samples* (MTS). The main idea is to use parts of the data as training samples. Remind that $\pi^J(\theta)$ is an improper prior. Let $x(l) = (x_1, \dots, x_q) \in \chi^q$ denote a part of the data \mathbf{y}_n such that $\pi^J(\theta|x(l))$ is a proper pdf and no subset of $x(l)$ leads to a proper posterior. A MTS is thus the minimal quantity of data for which all parameters in the model are identifiable. Thus conditioning π^J to a MTS gives a proper posterior, called a *posterior prior*.

Then the IBF is defined as the Bayes factor conditional to a MTS $x(l)$ by

$$B_{J,\pi}^{IBF} = B_{J,\pi}(\mathbf{y}_n)B_{\pi,J}(x(l))$$

where

$$B_{\pi,J}(x(l)) = \frac{m_J(x(l))}{m(x(l))}.$$

By construction, $B_{J,\pi}^{IBF}$ removes the arbitrariness in the choice of c_j . To reduce the dependence on MTS, using the *arithmetic* and *expected arithmetic* IBF

$$\begin{aligned} B_{J,\pi}^{AI}(\mathbf{y}_n) &= B_{J,\pi}(\mathbf{y}_n) \frac{1}{L} \sum_{l=1}^L B_{\pi,J}(x(l)), \\ B_{J,\pi}^{EAI}(\mathbf{y}_n) &= B_{J,\pi}(\mathbf{y}_n) \frac{1}{L} \sum_{l=1}^L E_{\hat{\theta}_n} [B_{\pi,J}(x(l))] \end{aligned}$$

makes sense (with $\hat{\theta}_n$ the MLE and $E_{\theta}[\cdot]$ the expectation under $\mathcal{M}(\theta)$). Other averages may be used (as the *geometric* IBF). The *Median* IBF (Berger & Perrichi 1998) appeared recently as one of the most stable version of the IBF. Finally, a nice property of the arithmetic IBF is its asymptotic equivalence with a “proper” Bayes factor arising from neutral *intrinsic priors* (appearing by a Schwarz approximation). This is a strong justification of the heuristic. For more precisions see Dass (2001). Some examples with lifetime models are given in Kim & Sun (2000).

REMARK 9. *Using such priors can reduce the computational cost of the IBF if they are proper. Note that in our case, these intrinsic priors are always proper. Denote (π^{IJ}, π^I) the intrinsic priors asymptotically equivalent to (π^J, π) for the computation of $B_{\pi,J}(\mathbf{y}_n)$. Since π^J and π are defined on the same parameter space, the intrinsic equations (48-49) in Berger & Perrichi (1996) reduce to*

$$\begin{cases} \pi^I(\theta) &= \pi(\theta), \\ \pi^{IJ}(\theta) &= \pi^J(\theta) E_{\theta} [B_{\pi,J}(x(l))] \end{cases}$$

under a condition of exchangeability for the MTS $x(l)$ (see for example Dmochowski 1996). From Theorem 1 in Berger & Perrichi (1996), π^{IJ} is always proper (since π is proper).

4.5.1.2 Main conclusions

The intrinsic heuristic leans on the use of small quantity of training data, which are chosen among the observed data (which explains the term *intrinsic*), to redefine a statistic which is formally valid but remains, in fact, difficult to assess. In the following, we are more interested in an intrinsic adaptation of DAC^J than in an adaptation of DAC_2^J . Firstly because our context is not model selection (the sampling model is the same in the marginal densities of $B_{\pi,J}$), secondly because we desire to preserve the intuitive sense of the criterion and its useful features listed in § 4.3.2.

4.5.2 Adapting DAC

4.5.2.1 The intrinsic DAC criterion

Following the intrinsic heuristic, our idea is to take a MTS $x(l)$ among the data \mathbf{y}_n to obtain a proper diffuse prior $\pi^J(\theta|x(l))$ and to use the remaining data $\mathbf{y}_n(l) = \mathbf{y}_n/x(l)$ to check the proximity between the prior information and the data information. The divergence

$$KL \{ \pi^J(\cdot|\mathbf{y}_n(l)) \parallel \pi^J(\cdot|x(l)) \}$$

is now the maximal value for the divergence between $\pi^J(\cdot|\mathbf{y}_n(l))$ and the assessed prior π conditioned to the learning information $x(l)$. This conditioning must be done in the numerator as in the denominator of DAC to attenuate the impact of the disturbing information yielded by the MTS. By a cross-validation argument, to reduce the dependence on $x(l)$, we define our adaptation as an *arithmetic* average over L available MTS. Define the *intrinsic (arithmetic) DAC criterion* by

$$\text{DAC}^{AIJ}(\pi|\mathbf{y}_n) = \frac{1}{L} \sum_{i=1}^L \frac{KL \{ \pi^J(\cdot|\mathbf{y}_n(l)) \parallel \pi(\cdot|x(l)) \}}{KL \{ \pi^J(\cdot|\mathbf{y}_n(l)) \parallel \pi^J(\cdot|x(l)) \}}. \quad (4.7)$$

Then, if the size of the MTS remains small with respect to n , DAC^{AIJ} remains small when $\pi(\theta)$ is close to $\pi^J(\theta|\mathbf{y}_n)$. When π is very few informative and is arbitrary chosen close to π^J , $\text{DAC}^{AIJ}(\pi|\mathbf{y}_n) \simeq 1$, which is similar to the behavior of DAC^J . It is easy to show that the features described in § 4.3.2 are preserved. Clearly, there is no restriction to use this adaptation.

Issues. In an industrial context where n can be small, the number L of available MTS can be problematic since the intrinsic adaptation can remain strongly dependent from some MTS containing outliers. Hence it is desirable to increase the number of MTS. Berger & Perrichi (2002) propose several ideas in this sense (especially when sufficient statistics can summarize the data information) and introduce the notion of *sequential* MTS (SMTS), including censored data. Thus the number of data combinations can increase dramatically. But the size of the SMTS can become very large and the information yielded by $\mathbf{y}_n(l)$ is likely to be far from the (ideal) information given by the whole sample. Using the special “censored” Jeffreys prior π_c^J defined by De Santis *et al.* (2001), instead of the standard π^J , is a practical alternative for simple models as the exponential distribution: a censored time can become a MTS.

However, the size q of a MTS can remain high with respect to n , especially when $\dim \Theta$ increases. It is clearly a limitation of the criterion (4.7) and incidentally, it raises the problem of the minimal number of data for which the notion of conflict with a prior is making sense. A natural idea could be to consider the minimal number of data such that the posterior distribution yields a minimal precision

on a function of interest to estimate. In § 4.7, a possible avenue for improving the adaptation is described.

4.5.2.2 Simulated examples

EXAMPLE 15. *Bernoulli model.*

Consider again the case of Bernoulli data $\mathbf{y}_n \sim \mathcal{B}_r(\theta)$ where we assume on θ a prior Beta distribution $\mathcal{B}_e(\alpha, \beta)$ on $[0, 1]$ (cf. Example 12). We choose $\theta_0 = 0.7$ as the true value of θ . The Jeffreys prior is the proper $\mathcal{B}_e(1/2, 1/2)$ distribution (moreover, these priors are conjugate). Thus, we can compare DAC^J and DAC^{AIJ} through the choice of the prior mean $\mu_0 = \alpha/(\alpha + \beta)$, fixing the prior standard deviation at 0.2. Figures 4.5 and 4.6 show the evolution of both DAC criteria, averaged on 10 simulated samples of sizes $n = 20$ (Figure 4.5) and $n = 10$ (Figure 4.6). They illustrate that DAC^{AIJ} and DAC^J can produce close agreement domains even if n is not high.

As expected, criteria are minimized when the prior mean μ_0 is close to the true value θ_0 . Moreover, domains of agreement become smaller when the prior becomes more informative with respect to the data. This feature has been noticed in Example 13. ■

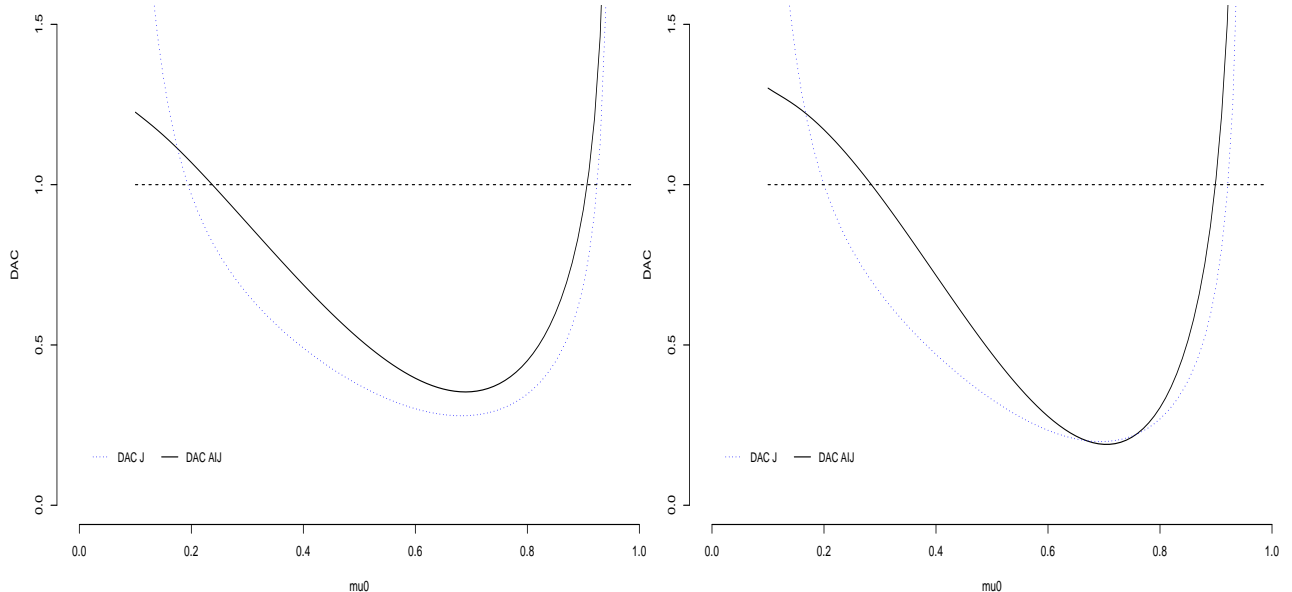


Figure 4.5: Mean evolution of DAC^J and DAC^{AIJ} in function of the mean μ_0 of the prior Beta distribution for a Bernoulli model (with $n = 20$ Bernoulli data).

Figure 4.6: Mean evolution of DAC^J and DAC^{AIJ} in function of the mean μ_0 of the prior Beta distribution for a Bernoulli model (with $n = 10$ Bernoulli data).

EXAMPLE 16. *Exponential model.*

We sampled a 10-sized *i.i.d.* data-set \mathbf{y}_n from an exponential distribution with parameter $\lambda_0 = 150^{-1}$: $\mathbf{y}_n = (142.76, 142.99, 470.3, 419.09, 185.20, 84.41, 8.13, 27.15, 573.17, 17.12)$. The MLE $\bar{y} = \hat{\lambda}_n^{-1} = 207$ is a minimal sufficient statistic and overestimates the real value. Here $\pi^J(\lambda) \propto \lambda^{-1}$ is the standard Jeffreys prior and a MTS is a single value. We choose the conjugate prior

$$\lambda \sim \mathcal{G}(a, ax_e)$$

where a embodies the size of a virtual sample of mean x_e . See Chapter 3 for more explanations. Note that π is perfectly in agreement with \mathbf{y}_n if $a = n = 10$ and $x_e = \bar{y} = \hat{\lambda}_n^{-1} = 207$. We give in Tab. 4.7 the values of DAC^{AIJ} for several prior scenarii. The adaptation rejects only few informative priors for which x_e is very far from the data (summarized by the minimal sufficient statistic \bar{y}). Evolutions of DAC^{AIJ} are displayed on Figures 4.9 and 4.10. We notice the same fact as in the ideal case of Example 13: even a good expert value can be finally rejected if π is too informative. In other words, DAC^{AIJ} as DAC^J discards unbiased priors with small variances and very biased priors with large variances.

Moreover, we provide in Tab. 4.8 the agreement domains for x_e computed by DAC^{AIJ} and the P -value procedure proposed by Evans & Moshonov (2005a). This procedure is described in a general censored case in Example 11. The agreement domain is the (5%, 95%) percentile interval of the marginal distribution of the usual minimal sufficient statistic. We observe, similarly to Example 12, that it accepts priors strongly rejected by the DAC procedure. ■

a	x_e				
	10	150	200	300	500
5	4.92	0.33	0.16	0.43	2.16
1	1.22	0.57	0.52	0.51	0.73
0.5	1.02	0.70	0.66	0.63	0.68
0.25	0.98	0.80	0.78	0.74	0.72

Tab 4.7: Values of DAC^{AIJ} for several prior scenarii.

a	DAC^{AIJ}	P -value proc. (5% - 95%)
2	55 - 500	37 - 625
5	90 - 383	41 - 510
7	102 - 355	53 - 480
10	118 - 320	65 - 460
15	137 - 290	77 - 440

Tab 4.8: Agreement domains for x_e .

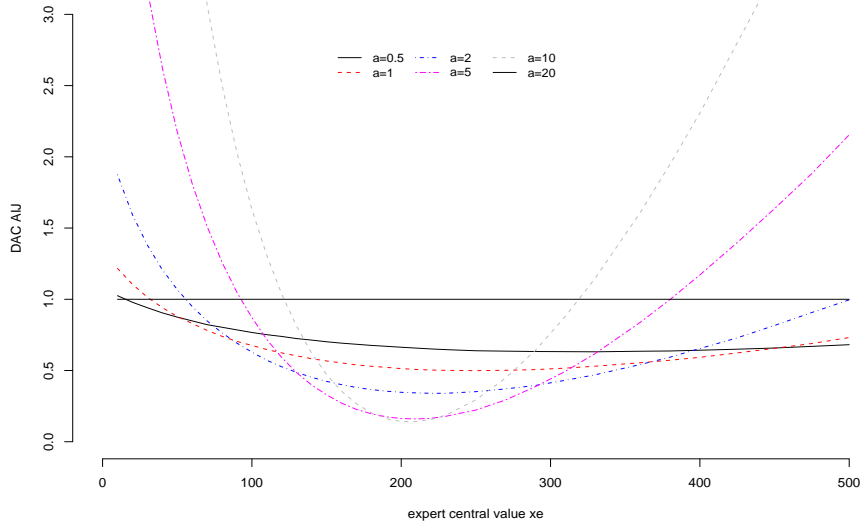


Figure 4.9: Evolutions of DAC^{AIJ} with respect to hyperparameters (x_e, a) of the conjugate prior distribution of the exponential model.

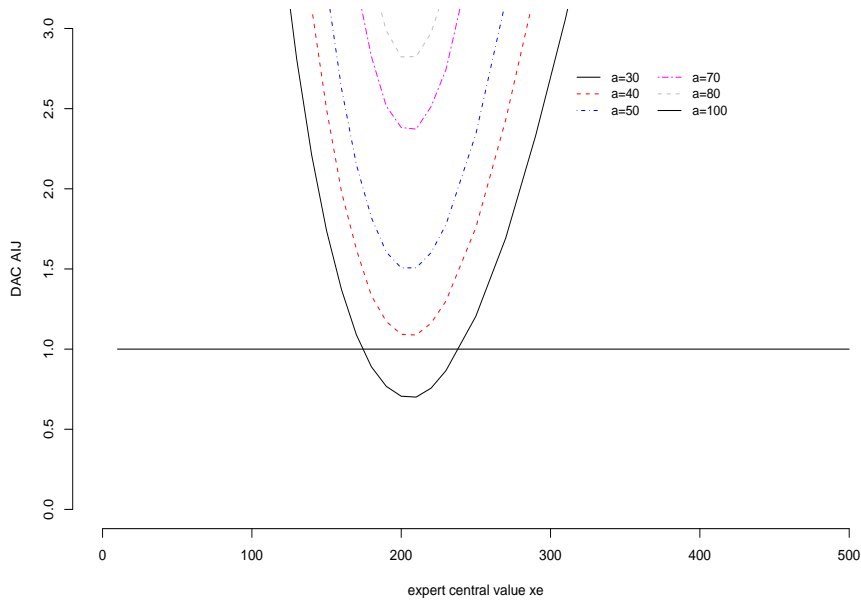


Figure 4.10: Evolutions of DAC^{AIJ} with respect to (x_e, a) .

4.5.3 A real example with the Weibull distribution

We consider the right-censored real lifetime data \mathbf{y}_n ($n = 18$) from Table C.1 (Appendix § C.1). They correspond to failure times or stopping times collected on some similar devices Σ belonging to the secondary water circuit of nuclear plants. Lifetimes are given in months. For physical reasons and according to a large consensus, those data are assumed to arise from a Weibull distribution $\mathcal{W}(\eta, \beta)$ with density

$$f(x|\eta, \beta) = \frac{\beta}{\eta} \left(\frac{x}{\eta}\right)^{\beta-1} \exp\left\{-\left(\frac{x}{\eta}\right)^\beta\right\} \mathbb{1}_{\{x \geq 0\}}.$$

The MLE is $(\hat{\eta}_n, \hat{\beta}_n) = (140.8, 4.51)$ with estimated standard deviations $\hat{\sigma}_n = (7.3, 1.8)$. The high value of $\hat{\beta}_n$ is unusual and invites (like the large estimated standard deviation) for a Bayesian estimation.

Two prior opinions on the lifetime are available on device Σ , given by independent experts \mathcal{E}_1 and \mathcal{E}_2 . They are summarized in Table C.2 (Appendix § C.1). The \mathcal{E}_1 opinion is much more informative than \mathcal{E}_2 and both are right-shifted with respect to the data. Moreover the experts are not asked at the same level of precision. \mathcal{E}_1 is a nuclear operator and speaks for a particular component while \mathcal{E}_2 can be seen as a component producer whose opinion takes into account a variety of running conditions.

Note that since the Weibull distribution does not admit conjugate prior (Soland 1966), the posterior computation is often heavy and needs numeric integration (Singpurwalla & Sun 1986) or sampling algorithms as Monte Carlo Markov Chains (MCMC). In our applications, we used importance sampling (cf Chap. 5).

We propose to consider the priors

$$\begin{cases} \eta & \sim \mathcal{G}(a, b), \\ \beta & \sim \mathcal{G}(c, d). \end{cases}$$

Since the device Σ is submitted to aging, an usual domain of main variations for the values of β is $D_\beta = [1, 5]$ (cf. Chapter 3). Since η is the 63th order percentile of the Weibull distribution, it is more tractable from the expert opinion than β . We translate approximatively the percentiles on X into the percentiles on η using the Weibull pdf, fixing $\beta = 3$. This translated knowledge and the corresponding values of a and b (assessed by least square regression) are given in Table 4.2.

	expert knowledge (on X)		translated knowledge (on η)		a	b
	(5%,95%) interval	median value	(5%,95%) interval	median value		
Expert \mathcal{E}_1	(200,300)	250	(224,336)	280	66.3	0.23
Expert \mathcal{E}_2	(100,500)	250	(112,560)	280	4.6	0.015

Table 4.2: Prior domains on X and η and hyperparameters values for $\pi_1(\eta)$.

Sun (1997) recommended to use the reference prior π^J whether one or both parameters are of interest (especially when small samples are used). In particular, when both parameters are of interest, this prior is the unique second-order matching prior. Here we place no hierarchy between the parameters, namely we do not consider the reference prior π^J in presence of a nuisance* parameter. Since π^J is improper, we have to compute DAC^{AJJ} . An uncensored MTS $x(l)$ is a couple of values (x_i, x_j)

such that $x_i \neq x_j$ and $x_i, x_j > 1$ (Sun 1997). Fortunately, $\pi^J(\theta|x_i, x_j)$ is explicit, which simplifies considerably the computation. See Appendix § 4.9.3.2 for details. When the MTS contain censored data, the special Jeffreys prior π_c^J elicited in Chapter 3 (§ 3.2.2.1) is used.

Denote

$$\begin{aligned}\tilde{\pi}(\eta, \beta) &= \pi(\eta)\pi^J(\beta) \propto \frac{1}{\beta}\pi(\eta), \\ \tilde{\tilde{\pi}}(\eta, \beta) &= \pi^J(\eta)\pi(\beta) \propto \frac{1}{\eta}\pi(\beta).\end{aligned}$$

From (4.3) we get

$$\text{DAC}^{AIJ}(\pi|\mathbf{y}_n) = \text{DAC}^{AIJ}(a, b|\mathbf{y}_n) + \text{DAC}^{AIJ}(c, d|\mathbf{y}_n) - 1. \quad (4.8)$$

where

$$\begin{aligned}\text{DAC}^{AIJ}(a, b|\mathbf{y}_n) &= \frac{1}{L} \sum_{l=1}^L \frac{KL \{ \pi^J(\cdot|\mathbf{y}_n(l)) \parallel \tilde{\pi}(\cdot|x(l)) \}}{KL \{ \pi^J(\cdot|\mathbf{y}_n(l)) \parallel \pi^J(\cdot|x(l)) \}}, \\ \text{DAC}^{AIJ}(c, d|\mathbf{y}_n) &= \frac{1}{L} \sum_{l=1}^L \frac{KL \{ \pi^J(\cdot|\mathbf{y}_n(l)) \parallel \tilde{\tilde{\pi}}(\cdot|x(l)) \}}{KL \{ \pi^J(\cdot|\mathbf{y}_n(l)) \parallel \pi^J(\cdot|x(l)) \}}.\end{aligned}$$

Thus, we can check $\pi(\eta)$ separately from any possible $\pi(\beta)$. We chosed $L = 30$ (among $n(n-1)/2$ possible uncensored and censored MTS). For expert \mathcal{E}_1 , we obtained $\text{DAC}^{AIJ}(a, b|\mathbf{y}_n) = 3.41$. For expert \mathcal{E}_2 , we obtained $\text{DAC}^{AIJ}(a, b|\mathbf{y}_n) = 1.76$. Thus, we detect the conflict between the data and the experts on the lifetime scale.

Note that the Gamma prior on η , for the first expert opinion, is very peaked and can be well approximated with a normal distribution (since $a > 30$). From equation (4.8), it is visible that no choice of $\pi(\beta)$, even a flat prior, allows to the complete prior π to be not conflicting. We think that in an industrial context, such a situation must be noticed before the Bayesian inference ; this discrepancy reflects a deep disparity between the two sources of information (data and expert).

The second expert opinion is not in this case. The scale parameter is affected by a similar conflict, but it remains possible to check that the complete prior is finally not conflicting: we must have $\pi(\beta)$ such that

$$\text{DAC}^{AIJ}(c, d|\mathbf{y}_n) \leq 0.24.$$

From the analyst viewpoint, the experts are optimistic with respect to the data (themselves do not know the data). So they seem to favor a soft aging of the device \sum (a simple reason is that they integrate some knowledge about the technical evolution in their opinions). For this reason, the Bayesian analyst should choose the expectation c/d of $\pi(\beta)$ in $[1, 2]$. For instance, the analyst selects $\text{E}[\beta] = 1.5$. Then the modelling of the second expert opinion is not conflicting for values c such that $\text{Var}[\beta] \geq 0.683$, so for $c \leq 0.65$.

This example is useful in two aspects. It shows that DAC can be used for calibration, when a first part of the prior is assessed and the second part is partially assessed. Using DAC as a prior calibration tool is the subject of the next section.

4.6 Prior calibration with DAC

In previous examples we generally considered that π was entirely assessed. The prior variance was fixed and the prior pointwise estimation θ_e of θ (a central value as the prior mean) was checked with respect to the data location. Thus, we obtained an agreement domain for θ_e . However, especially when the prior is subjective and is elicited from an expert opinion, θ_e reflects a personal viewpoint and is usually easy to assess. But common prior uncertainty measures, as the variance, are often more difficult to fix (in particular when the expert opinion is about X and not θ).

In particular, a prior modelling elicited from the expert opinion, without critical work from the analyst, can be strongly informative. This is the case with the expert opinion \mathcal{E}_1 in § 4.5.3. More generally, it happens often that *no credible information* is available on the expert. In those cases, a default rule for fixing the prior variance in a proper way is desirable.

Using DAC can help to answer this question. Denoting ω the prior hyperparameter, default values for ω can be found such that

$$\text{DAC}(\omega|\mathbf{y}_n) = 1$$

under the constraint $g(\omega) = \theta_e$ where, typically, $g(\omega) = \text{E}[\theta]$. Acting in such a way, we choose the most informative prior in accordance with the expert opinion and not conflicting with the data. In absence of a solution, θ_e should be modified with an external knowledge. Throughout examples using the exponential and the Weibull distributions, we illustrate some this strategy of calibration.

REMARK 10. *Note that DAC^{AIJ} needs to compute the posterior density $\pi(\theta|x(l))$ where $x(l)$ is a MTS. This is a limitation to the computational handiness of the adaptation, when a prior hyperparameter is unknown. This is really convenient only in the case of conjugate models and priors.*

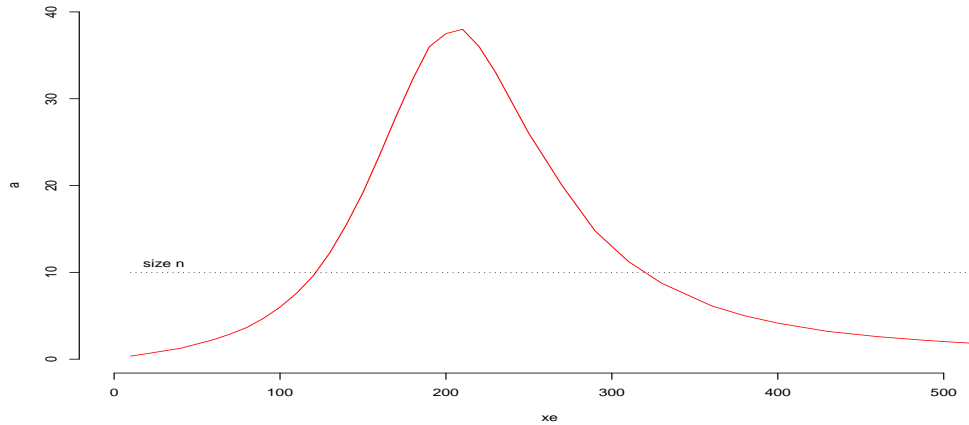
4.6.1 Application to the exponential distribution

In the exponential case, we have to use DAC^{AIJ} . Consider again the prior proposed in Example 16

$$\lambda \sim \mathcal{G}(a, ax_e)$$

where a is the size of a virtual i.i.d. sample with mean x_e and variance ax_e^2 . Thus a embodies the expert uncertainty: calibrating π is choosing a . In the following, denote \hat{a} the strictly positive value such that $\text{DAC}^{AIJ}(\hat{a}|\mathbf{y}_n) = 1$. The existence and unicity of \hat{a} can be basically proved using the following arguments. Since $\text{DAC}^{AIJ}(a|\mathbf{y}_n)$ is the sum of positive convex functions of a , it is convex in its argument a . There exists a unique minimizer $a^* = \arg \min \text{DAC}^{AIJ}(a|\mathbf{y}_n)$. Since $\text{DAC}^{AIJ}(a|\mathbf{y}_n) \rightarrow \infty$ when $a \rightarrow \infty$, for a fixed $n > 0$, we have $a^* = 0$ or $0 < a^* < \infty$. The case $a^* = 0$ corresponds to $\pi(\lambda) \propto \pi^J(\lambda)$ and $\text{DAC}^{AIJ}(a^*|\mathbf{y}_n) = 1$. This is contradictory with the fact that some priors can be not conflicting. Thus $a^* > 0$. By convexity, there exists a unique value $\hat{a} > 0$ such that $\text{DAC}^{AIJ}(\hat{a}|\mathbf{y}_n) = 1$.

An uncensored example. In Figure 4.11, the evolution of \hat{a} is displayed in function of x_e , using the data of Example 16. In the same figure, we place a boundary line at $a = n = 10$, since n is a natural limit value for a .



Tab 4.11: Evolutions of \hat{a} in function of x_e in a neighborhood of $\bar{y} = 207$.

A censored example. Assume now that \mathbf{y}_n contains $n - r$ right-censored fixed (type-I) values $\mathbf{c}_{n-r} = (c_1, \dots, c_{n-r})$. The first difference with the previous case is that n is no more a reference limit value for a . A conservative choice is clearly the number r of uncensored data. The second difference is the choice of π^J . According to De Santis *et al.* (2001), the censored Jeffreys prior

$$\pi_c^J(\lambda) \propto \lambda^{-1} \sqrt{n - \sum_{i=1}^{n-r} \exp(-\lambda c_i)}$$

(see § 3.2.2) yields a better frequentist coverage than the standard Jeffreys prior π^J for the posterior Bayes intervals.

Consider the lifetime censored datasets from Table 4.3. The MLE $\hat{\eta}_n = 1/\hat{\lambda}_n = 13,115$. On Table 4.4 we present estimations of \hat{a} computed using the priors π^J and π_c^J . We note that using π^J for the convenience of its conjugation properties can dramatically modify \hat{a} and shift the agreement domain for x_e to the largest values of X . Such results highlight the need of a fine choice of the benchmark noninformative prior when the data are poorly informative, and the need of using the full data information even it is incomplete.

4.6.2 Application to the Weibull distribution

We consider again the Weibull censored data from § 4.5.3. We change our prior, choosing a hierarchical prior defined in Chapter 3

$$\begin{aligned} \mu|\beta &\sim \mathcal{G}(a, b(a, \beta)), \\ \beta &\sim \mathcal{B}_e(p, q) \end{aligned}$$

where $\mu = \eta^{-\beta}$, where the Beta \mathcal{B}_e distribution is defined on $[\beta_l, \beta_r]$ and $b(a, \beta)$ is defined by

$$b(a, \beta) = (2^{1/a} - 1)^{-1} x_e^\beta,$$

x_e being also a prior estimation of the median lifetime. Still there, a gets a sense of a virtual sample size. Thus the complete prior is hyperparametrized by $(a, x_e, p, q, \beta_l, \beta_r)$. We do not consider here

<i>lifetime data</i>	<i>censored data</i>
671, 802, 1517	4380, 4380, 4818, 5256, 5256, 6132, 6132

Table 4.3: EDF exponential lifetime censored data *abis*.

x_e	π^J	π_c^J
3,000	0.2	2.5
5,000	1.75	5.5
10,000	7.8	6.2
15,000	12.5	1.5

Table 4.4: Estimations of \hat{a} when the standard and the censored Jeffreys priors are used.

the complete expert opinions given in Table 4.2, assuming simply that $x_e = 250$. Moreover, we fix $(\beta_l, \beta_r) = (1, 5)$ and some values of $E[\beta] = \beta_l + (\beta_r - \beta_l)p/(p + q)$. In the same idea that in the exponential case, we are also looking for default values for (a, p, q) . Note that $\pi^J(\beta)$ is proper on $[\beta_l, \beta_r]$. So using MTS in DAC^{AIJ} is required only to adjust $\pi^J(\theta'|\beta)$ where θ' is any parametrization among $\{\mu, \eta, \lambda\}$. Thus the size of a MTS $x(l)$ is 1. Similarly to § 4.5.3, we can study $\pi(\beta)$ and $\pi(\theta'|\beta)$ separately, using equation

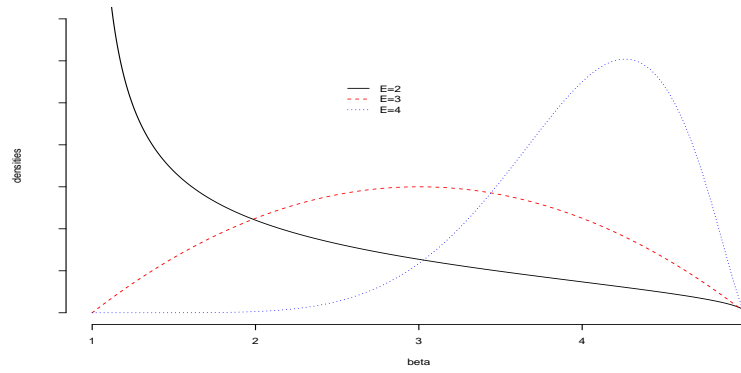
$$\text{DAC}^{AIJ}(\pi|\mathbf{y}_n) = \text{DAC}^{AIJ}(a|\mathbf{y}_n) + \text{DAC}^{AIJ}(p, q|\mathbf{y}_n) - 1. \quad (4.9)$$

Checking and calibrating $\pi(\theta'|\beta)$. The expert common central value $x_e = 250$ is very optimistic with respect to the data. We find $\hat{a} = 0.25$; thus, a non-conflicting prior on any scale parametrization cannot yield more information than 1/4 of the information yielded by a single data. This prior is very flat.

Checking and calibrating $\pi(\beta)$. We note $\hat{\sigma}$ the prior standard deviation on β such that $\text{DAC}^{AIJ}(p, q|\mathbf{y}_n) = 1$ and (\hat{p}, \hat{q}) the associate hyperparameter values. In Table 4.5 these values are computed for some classical choices of $E[\beta]$ in reliability. The elicited densities are displayed on Figure 4.12. Note that when $E[\beta] \rightarrow \hat{\beta}_n$ then σ decreases, which means that only priors modelling a strong aging are not conflicting with the data. Notice that a natural choice for (p, q) , in absence of any critical information about the expert opinion, is $(1, 1)$. Indeed, we obtain a uniform distribution on $[\beta_l, \beta_r] = [1, 5]$. Then the default value of σ is 0.53. Thus, assuming $E[\beta] = 2$, there is no other choice than eliciting (p, q) very close to $(0.5, 1.5)$ to obtain a non-conflicting prior.

$E[\beta]$	$\hat{\sigma}$	(\hat{p}, \hat{q})
2	0.5	(0.5, 1.5)
3	0.44	(2.0, 2.0)
4	0.27	(7.2, 2.4)

Table 4.5: Default values for the prior standard deviation and hyperparameters of $\pi(\beta)$.



Tab 4.12: Beta densities $\pi(\beta|\hat{p}, \hat{q})$ with respect to $E = E[\beta]$.

Calibrating the complete prior. Suppose that the Bayesian analyst recognizes that both experts have perceived at least one failure time in the past. Setting $a = 1$ for the common expert median, we obtain $\text{DAC}^{AIJ}(a|\mathbf{y}_n) = 1.16$. Suppose that the Bayesian analyst chooses $E[\beta] = 4$ (which means a strongly accelerated aging). Assuming that the complete prior remains not conflicting, we obtain a lower bound σ_{\min} on σ if we want to have

$$\text{DAC}^{AIJ}(\pi|\mathbf{y}_n) \leq 1.$$

From (4.9), this lower bound can be computed such that

$$\text{DAC}^{AIJ}(p, q|\mathbf{y}_n) = 0.84.$$

We obtain finally $\sigma_{\min} = 0.286$ in $(p, q) = (6.1, 2.03)$.

4.7 A possible improvement

In § 4.5.2, the issues raised by the definition of DAC^{AIJ} have been presented. They are especially due to the size q of the MTS, whose disturbing information can be too influential on the adaptation. In our industrial examples, this is fortunately not the case for the exponential and Weibull models, since q is respectively 1 and 2 and n remains close to 10 in the worst cases. Thus, it is desirable to decrease the amount of disturbing information yielded by the MTS into DAC^{AIJ} . In this aim, we suggest to *weight* the inferences which lead to compute $\pi^J(\theta|x(l))$ and $\pi(\theta|x(l))$ into the intrinsic adaptation. These two posteriors can be modified into *pseudoposteriors*¹ as follows.

1. Let $x(l) = x_1, \dots, x_q$ be an *i.i.d.* MTS.

2. Define the *p-weighted likelihood* by $\mathcal{L}_p(x(l); \theta) = \left(\prod_{i=1}^q f(x_i|\theta) \right)^{p/q}$.

3. Define the *weighted posteriors* by $\begin{cases} \pi_p^J(\theta|x(l)) \propto \mathcal{L}_p(x(l); \theta) \pi^J(\theta), \\ \pi_p(\theta|x(l)) \propto \mathcal{L}_p(x(l); \theta) \pi(\theta). \end{cases}$

¹This term was introduced by Walker & Hjort (2001) and Ghoshal (2004) in the quite different context of the study of posterior consistency.

4. Fix p as the minimal value in \mathbb{N} such that $\pi_p^J(\theta|x(l))$ and $\pi_p(\theta|x(l))$ are proper.

Necessarily, $p \leq q$. Thus, the mass of the likelihood of the MTS $x(l)$ is enlarged and the pseudoposteriors should yield a lower information than the real posteriors. Note that the value of p is generally limited by the choice of π^J since a proper prior π can arbitrarily be chosen regular enough to assume that $\pi_p(\theta|x(l))$ is proper for any $p \geq 0$. This scheme is easily adaptable to censored MTS, modifying the weighted likelihood and possibly π^J . Note that when π^J is proper, we have clearly $p = 0$. For NEF models, we obtain $\pi_1^J(\theta|x(l)) = \pi^J(\theta|t)$ where t is the usual minimal sufficient statistic (the mean in an *i.i.d.* case) of $x(l)$. A more interesting fact is if $p \leq q$ for multidimensional or nonconjugate models. For instance, $p = 1$ for the Weibull model, choosing π^J as the Jeffreys prior, although $q = 2$. See next proposition, which is proved in Appendix § 4.9.1.

PROPOSITION 15. Let $(x_i \neq x_j) > 1$ be a MTS for the Weibull model. Consider the parametrization (μ, β) where $\mu = \eta^{-\beta}$. Choose $\pi^J(\mu, \beta) \propto (\mu\beta)^{-1}$. Then $\pi_p^J(\mu, \beta|x_i, x_j)$ is proper for all $p > 0$, in particular for $p = 1$, and

$$\begin{cases} \mu|\beta, x_i, x_j & \sim \mathcal{G}\left(p, (p/2) \left\{x_i^\beta + x_j^\beta\right\}\right), \\ \pi_p^J(\beta|x_i, x_j) & \propto \beta^{p-1}(x_i x_j)^{p\beta/2} \left(x_i^\beta + x_j^\beta\right)^{-p}. \end{cases}$$

Some other rule could be proposed to choose p as a real value instead of an integer value. Especially, it is of interest to fix $0 < p < 1$ to obtain a vague benchmark prior $\pi_p^J(\theta|x(l))$ when it is possible. Thus we should consider this prior as yielding the same information that a fraction of one single data. See Example 17 for an illustration. But such a rule remains difficult to assess objectively.

EXAMPLE 17. *Weibull model.*

We follow the Weibull example presented at § 4.6.2. Using the same prior distribution, we however truncate $\pi(\mu|\beta)$ on $[0, 500]$. Thus π^J is proper and DAC^J can be computed for the comparison with DAC^{BIJ} . Fixing $a = 2$ and $\text{Var}[\beta] = 0.025$, we are looking for the domains on the prior median x_e and the prior mean β_0 such that $\pi(\mu|\beta)$ and $\pi(\beta)$ are not conflicting with \mathbf{y}_n , using development (4.9). Table 4.6 indicates these domains computed with DAC^J or DAC^{AIJ} . The intrinsic adaptation is computed using pseudoposterior $\pi_p^J(\cdot|x(l))$ defined in Proposition 15, for several values of p . In particular we choose $p = q = 2$, since it is the usual case where $\pi_p^J(\cdot|x(l)) = \pi^J(\cdot|x(l))$. We observe that the decreasing choice of p makes the agreement domains estimated by DAC^{AIJ} closer to the reference domain estimated by DAC^J . Thus, establishing a rule to weight $\pi^J(\theta|x(l))$ appears to be a subject of real interest.

	DAC ^J	DAC ^{AIJ}			
		p = 2	p = 1	p = 1/2	p = 1/4
η	37 - 242	45 - 236	43 - 238	40 - 240	39 - 241
β	1.42 - 4.92	1.49 - 4.78	1.45 - 4.82	1.43 - 4.85	1.42 - 4.88

Table 4.6: Agreement domains computed by DAC^J and DAC^{AIJ} (for several values of p) ■

4.8 Conclusion

Using expert opinions is frequent in industrial issues where data are collected with difficulty. These subjective perceptions can sometimes be far from the objective information yielded by the data, for instance for technical reasons. Thus checking a discrepancy or an agreement gives an information which can be used by the Bayesian analyst. The DAC criterion can be used for this detection. It is an alternative approach to the P -value procedure proposed by Evans & Moshonov (2005a, 2005b). Its handiness (here especially highlighted in the context of hierarchical priors) is, in our sense, a first argument in favor of DAC. But there are three main differences between the rules.

A first difference is that DAC indicates threshold values for the hyperparameters, such that π and \mathbf{y}_n are not conflicting, but such values remain undecidable for the P -value procedure. Thus, DAC can reject very biased priors that the P -value procedure accept, fixing usually the threshold P -values to (5%, 95%) percentiles.

A second difference is that, contrary to the P -value procedure, DAC accepts only prior distributions with compromising variance and bias. However, when the prior uncertainty is reasonable, both approaches lead to the same conclusions.

A final feature of DAC is that it can be used as a tool of calibration, indicating a default value for the variance of prior informative modellings.

Thus we think that DAC is an interesting alternative to the P -value procedure and a helpful method to place in the toolkit of the Bayesian analyst. In function of the available information about the conditions of the experiment and the expert credibility, the Bayesian analyst could correct the subjective beliefs or ask for other experiments to understand the discrepancy. An open issue could be detecting some outliers or the too influential data in the sample by sequential computations of DAC, increasing or randomizing the dataset.

However, there remain some important difficulties to use DAC. The choice of the noninformative benchmark prior π^J must be done carefully, being the prior with best frequentist coverage of the posterior Bayes intervals. Our intrinsic adaptation can suffer of the small size n and the high dimension of the model. A possible improvement has been proposed for this last point, as an avenue for future work.

4.9 Appendix

4.9.1 Propositions and proofs

In this subsection, the main propositions introduced or used in the text are recalled and proofs are given.

PROPOSITION 13. Geometric combination of multiple priors. *Let π_1, \dots, π_m be m priors on Θ , and let $\alpha_1, \dots, \alpha_m$ be m weights such that $0 < \alpha_i < 1$, $\sum_{i=1}^m \alpha_i = 1$. Denote $\pi(\theta) \propto \prod_{i=1}^m \pi(\theta)_i^{\alpha_i}$.*

Then

$$DAC^J(\pi|\mathbf{y}_n) \leq \sum_{i=1}^m \alpha_i DAC^J(\pi_i|\mathbf{y}_n).$$

Proof. Denoting $E[\cdot]$ the mean with respect to $\pi^J(\theta|\mathbf{y}_n)$, it is easy to see that

$$\begin{aligned} KL\{\pi^J(\cdot|\mathbf{y}_n) \parallel \pi\} &= E\left[\log \pi^J(\cdot|\mathbf{y}_n) - \sum_{i=1}^m \alpha_i \log \pi(\theta)_i\right] + \log \int_{\Theta} \prod_{i=1}^m \pi(\theta)_i^{\alpha_i} d\theta, \\ &= \sum_{i=1}^m \alpha_i KL\{\pi^J(\cdot|\mathbf{y}_n) \parallel \pi_i\} + \log \Delta \end{aligned}$$

where $\Delta = \int_{\Theta} \prod_{i=1}^m \pi_i(\theta)^{\alpha_i} d\theta$. Using the generalized Hölder inequality, we obtain

$$\Delta \leq \prod_{i=1}^m \left(\int_{\Theta} \pi_i(\theta) d\theta \right)^{\alpha_i}.$$

Thus $\Delta \leq 1$ since the π_i are densities and

$$KL\{\pi^J(\cdot|\mathbf{y}_n) \parallel \pi\} \leq \sum_{i=1}^m \alpha_i KL\{\pi^J(\cdot|\mathbf{y}_n) \parallel \pi_i\}$$

which leads straightforwardly to the statement of the proposition. \square

PROPOSITION 14. Bernoulli model. Let \mathbf{X}_n be an i.i.d. sample from a Bernoulli $\mathcal{B}_r(\theta)$ distribution. We assume on θ a prior Beta distribution $\mathcal{B}_e(\alpha, \beta)$ on $[0, 1]$. The Jeffreys prior π^J is the $\mathcal{B}_e(1/2, 1/2)$ density. Note $\delta_n = \sum_{i=1}^n X_i$. Denoting $0 < \theta_0 < 1$ the true value of the parameter. Then, $\forall(\alpha, \beta) > 0$, for any $0 < q < 1$,

$$(\log n)^q E[DAC^J(\alpha, \beta|\mathbf{X}_n) - 1] \rightarrow 0.$$

Proof. The Moivre-Laplace (or CLT) theorem indicates that under $\mathcal{B}_r(\theta_0)$,

$$\frac{\delta_n - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Thus denote

$$\delta_n = n\theta_0 + U_n\sqrt{n}$$

where $U_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \theta_0(1-\theta_0))$. Note Ψ the digamma function (the log-derivative of the gamma function). After some heavy calculations using asymptotic following developments

$$\begin{aligned} \Psi(n+1) &= \log n + \frac{1}{2n} - \frac{1}{12n^2} + o(n^{-3}), \\ \log \Gamma(n+1) &= \frac{1}{2} \log 2\pi + \left(n + \frac{1}{2}\right) \log n - n + \frac{\alpha}{12n}, \quad \text{where } 0 < \alpha < 1 \end{aligned}$$

which can be derived from Abramowitz and Stegun (1972, p. 258-260) and the exact Stirling formula given in Artin (1964, p.24), respectively, we find for $n > \max(\frac{1}{2\theta_0}, \frac{1}{2(1-\theta_0)})$ that

$$\begin{aligned} KL \{ \pi^J(\cdot | \mathbf{X}_n) | \pi^J \} &= (n + 1/2) \log n - n\Psi(1/2) \\ &\quad + U_n \sqrt{n} \{ 4 - 2\Psi(1/2) \} + K_{\theta_0}(1/2, 1/2) + o(1), \\ KL \{ \pi^J(\cdot | \mathbf{X}_n) | \pi \} &= (n + 3/2 - \alpha - \beta) \log n - n(\Psi(\beta) + \theta_0 \{ \Psi(\alpha) - \Psi(\beta) \}) \\ &\quad + U_n \sqrt{n} \{ 4 - \Psi(\alpha) - \Psi(\beta) \} + K_{\theta_0}(\alpha, \beta) + o(1). \end{aligned}$$

where

$$K_{\theta_0}(\alpha, \beta) = \log \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} - \log \sqrt{2\pi} + \left(\alpha - \frac{1}{2} \right) \log \frac{\Psi(\alpha)}{\theta_0} + \left(\beta - \frac{1}{2} \right) \log \frac{\Psi(\beta)}{1 - \theta_0}.$$

Then the asymptotic development of DAC^J gives

$$\begin{aligned} DAC^J(\alpha, \beta | \mathbf{X}_n) &= 1 + \frac{A_{\theta_0}(\alpha, \beta)}{\log n} \left\{ 1 - \frac{\Psi(1/2)}{\log n} \right\} + U_n \frac{B(\alpha, \beta)}{\sqrt{n} \log n} \left\{ 1 - \frac{\Psi(1/2)}{\log n} \right\} \\ &\quad - U_n \frac{C_{\theta_0}(\alpha, \beta)}{\sqrt{n} (\log n)^2} + \frac{D(\alpha, \beta)}{n} + o(n^{-1}) \end{aligned} \quad (4.10)$$

where

$$\begin{aligned} A_{\theta_0}(\alpha, \beta) &= \Psi(1/2) - \Psi(\beta) + \theta_0 \{ \Psi(\beta) - \Psi(\alpha) \}, \\ B(\alpha, \beta) &= 2\Psi(1/2) - \Psi(\alpha) - \Psi(\beta), \\ C_{\theta_0}(\alpha, \beta) &= A_{\theta_0}(\alpha, \beta) \{ 4 - 2\Psi(1/2) \}, \\ D(\alpha, \beta) &= 1 - \alpha - \beta. \end{aligned}$$

Note that at least one term in the development (4.10) is nonzero, except when $\pi \equiv \pi^J$ ($\Leftrightarrow \alpha = \beta = 1/2 \Leftrightarrow DAC^J = 1$). Indeed,

$$\begin{cases} A_{\theta_0}(\alpha, \beta) = 0, \\ B(\alpha, \beta) = 0, \\ D(\alpha, \beta) = 0 \end{cases} \Leftrightarrow \begin{cases} \Psi(\alpha) = \Psi(\beta), \\ \alpha + \beta = 1 \end{cases} \Leftrightarrow \alpha = \beta = 1/2.$$

To prove that

$$(\log n)^q \mathbf{E} \left[DAC^J(\alpha, \beta | \mathbf{X}_n) - 1 \right] \rightarrow 0$$

for any $0 < q < 1$, it is enough to control $\mathbf{E}[V_n]$ where $V_n = U_n [n(\log n)^q]^{-1}$. A sufficient condition is to show that $\mathbf{E}[V_n] \rightarrow 0$ when n increases. This can be done as follows. Denote $Z_n = \delta_n/n$. With

$$V_n = \frac{(Z_n - \theta_0)}{(\log n)^q},$$

we obtain by Markov's inequality, for any $M > 0$,

$$\begin{aligned} \mathbf{E} [|V_n| \mathbb{1}_{\{|V_n| \geq M\}}] &\leq M^{-1} \mathbf{E} [|V_n|^2], \\ &\leq M^{-1} \frac{\mathbf{E}[Z_n^2] + 2\theta_0 \mathbf{E}[Z_n] + \theta_0^2}{(\log n)^{2q}}, \\ &\leq M^{-1} \frac{\theta_0 (1 - \theta_0 + 2n\theta_0)}{n (\log n)^{2q}} \end{aligned}$$

which obviously tends to 0 when $n \rightarrow \infty$ followed by $M \rightarrow \infty$. This result ensures that V_n is *asymptotically uniformly integrable*. Then, from van der Vaart (1998, Theorem 2.20), we have

$$\lim_{n \rightarrow \infty} \mathbb{E}[V_n] = \lim_{n \rightarrow \infty} \frac{\mathbb{E}[U]}{n(\log n)^q}$$

where $U \sim \mathcal{N}(0, \theta_0(1 - \theta_0))$. Then $\mathbb{E}[V_n] \rightarrow 0$ and the statement of the proposition follows.

The criterion DAC^J comes to 1 in mean by upper or lower values. We thus study asymptotically the probability $P_{\theta_0}(\text{DAC}_1^J(\alpha, \beta | \mathbf{X}_n) - 1 \leq 0)$ to come to 1 by lower values. Denote Φ the cumulative function of the reduced normal distribution.

A) Suppose firstly that $A_{\theta_0}(\alpha, \beta) \neq 0$. When $B(\alpha, \beta) > 0$, from (4.10),

$$P_{\theta_0} \left\{ \text{DAC}^J(\alpha, \beta | \mathbf{X}_n) - 1 \leq 0 \right\} \stackrel{n \rightarrow \infty}{\simeq} \Phi \left(-\frac{A_{\theta_0}(\alpha, \beta)\sqrt{n}}{B(\alpha, \beta)\sqrt{\theta_0(1 - \theta_0)}} \right).$$

So

$$\begin{aligned} \lim_{n \rightarrow \infty} P_{\theta_0} \left\{ \text{DAC}^J(\alpha, \beta | \mathbf{y}_n) - 1 \leq 0 \right\} &= 0 \Leftrightarrow A_{\theta_0}(\alpha, \beta) > 0, \\ \lim_{n \rightarrow \infty} P_{\theta_0} \left\{ \text{DAC}^J(\alpha, \beta | \mathbf{y}_n) - 1 \leq 0 \right\} &= 1 \Leftrightarrow A_{\theta_0}(\alpha, \beta) < 0. \end{aligned}$$

Conversely, when $B(\alpha, \beta) < 0$,

$$P_{\theta_0} \left\{ \text{DAC}^J(\alpha, \beta | \mathbf{X}_n) - 1 \leq 0 \right\} \stackrel{n \rightarrow \infty}{\simeq} 1 - \Phi \left(\frac{A_{\theta_0}(\alpha, \beta)\sqrt{n}}{|B(\alpha, \beta)|\sqrt{\theta_0(1 - \theta_0)}} \right).$$

which gives the same convergences on the same conditions on $A_{\theta_0}(\alpha, \beta)$. Finally, suppose $B(a, \beta) = 0$. Then, if $A_{\theta_0}(\alpha, \beta) > 0$, then

$$P_{\theta_0} \left\{ \text{DAC}^J(\alpha, \beta | \mathbf{X}_n) - 1 \leq 0 \right\} \stackrel{n \rightarrow \infty}{\simeq} 1 - \Phi \left(\frac{\sqrt{n} \log n}{4 - 2\Psi(1/2)} \right) \xrightarrow{n \rightarrow \infty} 0.$$

Similarly, when $A_{\theta_0}(\alpha, \beta) < 0$, then

$$P_{\theta_0} \left\{ \text{DAC}^J(\alpha, \beta | \mathbf{X}_n) - 1 \leq 0 \right\} \stackrel{n \rightarrow \infty}{\simeq} \Phi \left(\frac{\sqrt{n} \log n}{4 - 2\Psi(1/2)} \right) \xrightarrow{n \rightarrow \infty} 1.$$

B) Secondly, suppose that $A_{\theta_0}(\alpha, \beta) = 0$ (so $C_{\theta_0}(\alpha, \beta) = 0$). Then, if $B(\alpha, \beta) > 0$,

$$P_{\theta_0} \left\{ \text{DAC}^J(\alpha, \beta | \mathbf{X}_n) - 1 \leq 0 \right\} \stackrel{n \rightarrow \infty}{\simeq} \Phi \left(\frac{(1 - \alpha - \beta) \log n}{B(\alpha, \beta)\sqrt{n}} \right) \xrightarrow{n \rightarrow \infty} 1/2.$$

We obtain the same result when $B(\alpha, \beta) < 0$. The last case is when $B(\alpha, \beta) = 0$. It appears when $\alpha = \beta$. Then, if $\alpha < 1/2$, the expression (4.10) says that the limit is attained by upper values. Conversely, if $\alpha > 1/2$, the limit is attained by lower values. \square

PROPOSITION 15. **Weibull model.** Let $(x_i \neq x_j) > 1$ be a i.i.d. minimal training sample (MTS) for the Weibull model parametrized by (μ, β) . Choose $\pi^J(\mu, \beta) \propto (\mu\beta)^{-1}$. Denote

$$\pi_p^J(\mu, \beta | x_i, x_j) \propto \{\mathcal{L}(x_i, x_j; \mu, \beta)\}^{p/2} \pi^J(\mu, \beta)$$

where $\mathcal{L}(\mathbf{x}; \mu, \beta)$ is the Weibull likelihood. Then $\pi_p^J(\mu, \beta | x_i, x_j)$ is proper for all $p > 0$, in particular for $p = 1$, and

$$\begin{cases} \mu | \beta, x_i, x_j & \sim \mathcal{G}\left(p, (p/2) \{x_i^\beta + x_j^\beta\}\right), \\ \pi_p^J(\beta | x_i, x_j) & \propto \beta^{p-1} (x_i x_j)^{p\beta/2} (x_i^\beta + x_j^\beta)^{-p}. \end{cases}$$

Proof. We have

$$\pi_p^J(\mu, \beta | x_i, x_j) \propto \mu^{p-1} \beta^{p-1} (x_i x_j)^{p\beta/2} \exp\left\{-\frac{p\mu}{2} (x_i^\beta + x_j^\beta)\right\} \mathbb{1}_{\{(\mu, \beta) \geq 0\}}.$$

Then

$$\begin{aligned} \mu | \beta, x_i, x_j & \sim \mathcal{G}\left(p, (p/2) \{x_i^\beta + x_j^\beta\}\right), \\ \pi_p^J(\beta | x_i, x_j) & \propto \beta^{p-1} \frac{(x_i x_j)^{p\beta/2}}{(x_i^\beta + x_j^\beta)^p}. \end{aligned}$$

If $(x_i, x_j) > 1$ and $x_i \neq x_j$ then $\alpha = \max(x_i, x_j)(\sqrt{x_i x_j})^{-1} > 1$. Thus

$$\begin{aligned} \beta^{p-1} \frac{(x_i x_j)^{p\beta/2}}{(x_i^\beta + x_j^\beta)^p} & \leq \beta^{p-1} \frac{(x_i x_j)^{p\beta/2}}{\max(x_i, x_j)^{p\beta}}, \\ & \leq \beta^{p-1} \exp\left(-p\beta \log \frac{\max(x_i, x_j)}{\sqrt{x_i x_j}}\right), \\ & \leq \beta^{p-1} \exp(-p\beta \log \alpha). \end{aligned}$$

Since $\alpha > 1$, $\log \alpha > 0$; then we recognize that the right hand side of the last inequality is proportional to a Gamma density with shape parameter p and scale parameter $p \log \alpha$. Thus we obtain that $p > 0 \Rightarrow \pi_p^J(\beta | x_i, x_j)$ is proper. \square

4.9.2 Kullback-Leibler divergences

Some of the expressions hereafter can be derived from Penny (2001). Ψ is the digamma function (the log-derivative of the gamma function).

Beta distributions. Let $B(a, b)$ and $B(c, d)$ be two Beta distributions. Then

$$KL\{B(a, b) \parallel B(c, d)\} = \log \frac{\Gamma(a+b)\Gamma(c)\Gamma(d)}{\Gamma(c+d)\Gamma(a)\Gamma(b)} + (a-c)\{\Psi(a) - \Psi(c)\} + (b-d)\{\Psi(b) - \Psi(d)\}.$$

Dirichlet distributions. Let $D_1 = \text{Dir}(a_1, \dots, a_d)$ and $D_2 = \text{Dir}(b_1, \dots, b_d)$ be two Dirichlet distributions. Note $a = \sum_{i=1}^d a_i$ and $b = \sum_{i=1}^d b_i$. Then

$$KL\{D_1 \parallel D_2\} = \log \frac{\Gamma(a)}{\Gamma(b)} + \sum_{i=1}^d \log \frac{\Gamma(b_i)}{\Gamma(a_i)} + \sum_{i=1}^d (a_i - b_i)(\Gamma(a_i) - \Gamma(b_i)).$$

Univariate normal bounded distributions. Let $N_1 = \mathcal{N}(\mu_1, \sigma_1^2)$ and $N_2 = \mathcal{N}(\mu_2, \sigma_2^2)$ be two univariate normal distributions defined on the bounded domain $[T_l, T_m]$. Note

$$\begin{aligned} \alpha_i &= \alpha_i(T_l, T_m, \mu_i, \sigma_i) = \left(\int_{T_l}^{T_m} \exp \left\{ -\frac{1}{2\sigma_i^2} (x - \mu_i)^2 \right\} dx \right)^{-1}, \\ N_i(x) &= \alpha_i \exp \left(-\frac{(x - \mu_i)^2}{2\sigma_i^2} \right). \end{aligned}$$

Then (denoting $E[\cdot]$ the mean with respect to N_1),

$$KL\{N_1 \parallel N_2\} = \log \frac{\alpha_1}{\alpha_2} + \left(\frac{1}{2\sigma_2^2} - \frac{1}{2\sigma_1^2} \right) E[(X - \mu_1)^2] + \frac{\mu_2^2 - \mu_1^2}{2\sigma_2^2} + \frac{\mu_1 - \mu_2}{\sigma_2^2} E[X]$$

where

$$\begin{aligned} E[(X - \mu_1)^2] &= \sigma_1^2 \{1 + (T_l - \mu_1)N_1(T_l) + (\mu_1 - T_m)N_1(T_m)\}, \\ E[X] &= \mu_1 + \sigma_1^2 \{N_1(T_l) - N_1(T_m)\}. \end{aligned}$$

Univariate normal versus uniform distributions. Let $N_1 = \mathcal{N}(\mu_1, \sigma_1^2)$ be a univariate distribution on $[T_l, T_m]$. Let U be the uniform distribution on $[T_l, T_m]$. With the notations previously defined,

$$KL\{N_1 \parallel U\} = \log \alpha_1 + \log(T_m - T_l) - \frac{1}{2\sigma_1^2} E[(X - \mu_1)^2].$$

Gamma versus univariate normal distributions. Let $G = \mathcal{G}(\gamma, \beta)$ be a Gamma distribution with mean γ/β and variance γ/β^2 . Let $N = \mathcal{N}(\mu, \sigma^2)$ be a univariate distribution. Then

$$KL(G \parallel N) = \log \frac{\beta\sqrt{2\pi\sigma^2}}{\Gamma(\gamma)} + (\gamma - 1)\Psi(\gamma) - \frac{1}{\sigma^2} \left[\frac{\gamma}{\beta} \left(\frac{1}{\beta} - \mu \right) + \frac{\mu^2}{2} \right].$$

Gamma distributions. For $i = 1, 2$, let $G_i = \mathcal{G}(\alpha_i, \beta_i)$ be two Gamma distributions with mean α_i/β_i and variance α_i/β_i^2 . Then

$$KL\{G_1 \parallel G_2\} = (\alpha_1 - \alpha_2)\Psi(\alpha_1) + \alpha_2 \log \frac{\beta_1}{\beta_2} + \log \frac{\Gamma(\alpha_2)}{\Gamma(\alpha_1)} + \alpha_1 \left(\frac{\beta_2}{\beta_1} - 1 \right).$$

4.9.3 Posterior priors

In this appendix, we give the formal expressions of the diffuse priors, proposed for defining the intrinsic adaptation of the noninformative prior π^J in the exponential and Weibull cases.

4.9.3.1 The exponential distribution

First, suppose that the MTS $x(l) = x_i$ where x_i is a single *i.i.d.* data from an uncensored sample \mathbf{y}_n . Then π^J is the Jeffreys prior and

$$\begin{aligned}\lambda|x(l) &\sim \mathcal{E}(x_i), \\ \lambda|\mathbf{y}_n(l) &\sim \mathcal{G}\left(n-1, \sum_{j=1}^n y_j - x_i\right).\end{aligned}$$

Secondly, suppose that $x(l) = c_i$ where c_i is a right-censored fixed data. Denote c_1, \dots, c_{n-r} the $n-r$ censored data into \mathbf{y}_n . From De Santis *et al.* (2001), we use the special Jeffreys prior

$$\pi_c^J(\lambda) \propto \frac{1}{\lambda} \sqrt{n - \sum_{j=1}^{n-r} \exp(-\lambda c_j)}.$$

Then

$$\pi_c^J(\lambda|c_i) \propto \sqrt{n - \sum_{j=1}^{n-r} \exp(-\lambda c_j)} \exp(-\lambda c_i).$$

To prove that $\pi_c^J(\lambda|c_i)$ is proper, we consider the marginal density

$$m_{J_c}(c_i) = \int_{\mathbb{R}^+} \pi_c^J(\lambda|c_i) S(c_i|\lambda) d\lambda$$

where S is the survival function. It is enough to show that $m_{J_c}(c_i) < \infty$. Since

$$\frac{1}{\lambda} \sqrt{n - \sum_{j=1}^{n-r} \exp(-\lambda c_j)} < \lambda^{-1/2} \sqrt{n \sum_{j=1}^{n-r} c_j} \quad \forall \lambda > 0,$$

there exists $\alpha > 0$ such that

$$m_{J_c}(c_i) < \alpha \int_{\mathbb{R}^+} \lambda^{-1/2} \sqrt{n \sum_{j=1}^{n-r} c_j} \exp(-\lambda c_i) d\lambda = \alpha \sqrt{n\pi} \sqrt{\sum_{j=1}^{n-r} \frac{c_j}{c_i}} < \infty.$$

4.9.3.2 The Weibull distribution

We suppose to have a sample $\mathbf{y}_n = (\mathbf{x}_r, \mathbf{c}_{n-r})$ where $\mathbf{x}_r = (x_1, \dots, x_r)$ are uncensored *i.i.d.* data with density function $f(x|\eta, \beta) = \beta \eta^{-\beta} x^{\beta-1} \exp(-\eta^{-\beta} x^\beta)$ for $x \geq 0$ and \mathbf{c}_{n-r} is a set of $n-r$ fixed right-censored data. Following Sun (1997), we consider the second-order matching reference prior $\pi^J(\eta, \beta) \propto (\eta\beta)^{-1}$. A uncensored MTS $x(l)$ for the Weibull model is a couple $(x_i, x_j) > 1$, $(x_i \neq x_j)$. Note that we define no hierarchy between the two parameters, namely we do not consider the reference prior in the case of a nuisance parameter.

Note that for any MTS (x_i, x_j) the posterior of the reference prior is explicit, thanks to a beautiful result of Berger *et al.* (1998):

$$\pi^{ij}(\eta, \beta) = (2(x_i x_j) |\log x_i/x_j|)^{-1} (x_i x_j)^{\beta-1} \beta \eta^{-2\beta-1} \exp\left(-\eta^{-\beta} (x_i^\beta + x_j^\beta)\right)$$

Then consider the new parametrization $\eta \rightarrow \mu = \eta^{-\beta}$, $\beta \rightarrow \beta$ with Jacobian $J(\mu, \beta) = \beta\mu^{1+1/\beta}$. The corresponding noninformative prior is $\pi^J(\mu, \beta) \propto (\mu\beta^2)^{-1}$. Thus $\pi^{ij}(\mu, \beta) = \pi^{ij}(\mu|\beta) \pi^{ij}(\beta)$ with

$$\begin{aligned}\mu|\beta &\sim \mathcal{G}\left(2, x_i^\beta + x_j^\beta\right), \\ \pi^{ij}(\beta) &= \frac{(x_i x_j)^{\beta-2}}{2|\log x_i/x_j| \left(x_i^\beta + x_j^\beta\right)^2}.\end{aligned}$$

The computation of $\text{DAC}^{AIJ}(\pi|\mathbf{y}_n)$ needs to compute the posterior density of π^J conditionally to $\mathbf{y}_{n(ij)} = (y_{(ij)1}, \dots, y_{(ij)n})$ the sample \mathbf{y}_n whose components x_i and x_j have been removed. Denote similarly $\mathbf{x}_{n(ij)}$. The posterior densities are

$$\pi^{ij}(\mu, \beta|\mathbf{y}_{n(ij)}) = \pi^{ij}(\mu|\beta, \mathbf{y}_{n(ij)}) \pi^{ij}(\beta|\mathbf{y}_{n(ij)})$$

with

$$\begin{aligned}\mu|\beta, \mathbf{y}_{n(ij)} &\sim \mathcal{G}\left(r, \sum_{k=1}^n y_{(ij)k}^\beta\right), \\ \pi(\beta|\mathbf{y}_{n(ij)}) &\propto \beta^r \frac{\left(\prod_{k=1}^r x_{(ij)k}\right)^\beta}{\left(\sum_{k=1}^n y_{(ij)k}^\beta\right)^r}.\end{aligned}$$

Chapitre 5

Méthodes de calcul bayésien

Dans les chapitres précédents, nous nous sommes focalisés sur l'étude de la connaissance experte et la calibration d'une mesure de probabilité *a priori*, en passant volontairement sous silence les méthodes de calcul bayésien, vouées à estimer *a posteriori* une fonction d'intérêt. Celles-ci interviennent par ailleurs dans le calcul du critère DAC, proposé au chapitre 4. L'objet du présent chapitre est donc de présenter et analyser les méthodes numériques nous permettant de mener les calculs nécessaires à une analyse bayésienne complète.

Table locale

5.1	Introduction	140
5.2	Algorithmes d'échantillonnage	142
5.3	L'algorithme BRM : une étude critique	147
5.4	L'algorithme Population Monte-Carlo	152
5.5	Conclusions	164

5.1 Introduction

Considérons un échantillon de données observées \mathbf{y}_n , qui sont supposées être des représentations d'une variable aléatoire de densité $f(x|\theta)$, de paramètre $\theta \in \Theta$. Dans le cadre bayésien, θ est lui aussi supposé aléatoire et muni d'une densité $\pi(\theta)$ dominée par une mesure de référence ν sur l'espace mesurable Θ^1 . L'estimation bayésienne est fondée sur la loi *a posteriori* de densité

$$\pi(\theta|\mathbf{y}_n) \propto \pi(\theta)\mathcal{L}(\mathbf{y}_n;\theta)$$

où $\mathcal{L}(\mathbf{y}_n;\theta)$ désigne la vraisemblance de l'échantillon \mathbf{y}_n .

Ce chapitre est consacré aux méthodes numériques permettant d'estimer *a posteriori* une fonction d'intérêt $\Phi(\theta)$. En général, dans un cadre d'étude fiabiliste, $\Phi(\theta) = \theta$, $\Phi(\theta) = \mathbb{E}[X|\theta]$ (durée de vie moyenne) ou $\Phi(\theta) = S(t|\theta)$ où t est une valeur de temps fixée (survie au temps t). L'estimation ϕ la plus traditionnelle est réalisée sous le *coût quadratique* (Robert 2001). Il s'agit de l'espérance *a posteriori*

$$\begin{aligned} \phi &= \int_{\Theta} \Phi(\theta)\pi(\theta|\mathbf{y}_n) d\theta, \\ &= \frac{\int_{\Theta} \Phi(\theta)\mathcal{L}(\mathbf{y}_n;\theta)\pi(\theta) d\theta}{\int_{\Theta} \mathcal{L}(\mathbf{y}_n;\theta)\pi(\theta) d\theta}. \end{aligned}$$

Le plus souvent, ces intégrales ne sont pas explicites et il est indispensable de recourir à des méthodes d'approximation. Trois types de méthode sont théoriquement susceptibles de fournir une bonne estimation de ces intégrales.

1. Les méthodes d'approximation de Lindley (1980) et de Laplace (Tierney *et al.* 1986), fondée sur le développement asymptotique des quantités à intégrer.
2. Les méthodes d'intégration numérique de type Newton-Cotes, Runge-Kutta, etc. qui offrent plutôt de bons résultats lorsque $d = \dim \Theta$ est petite, mais présentent des instabilités numériques (Bacha 1996).
3. Les algorithmes d'échantillonnage.

Selon Bacha (1996), les deux premières méthodes restent peu adaptées à notre cadre d'étude, où les données \mathbf{y}_n sont en faible nombre et censurées, au contraire des méthodes d'échantillonnage. Dans ce chapitre, nous considérons exclusivement ces dernières.

Le livre de Robert & Casella (2004) offre un panorama complet des méthodes d'échantillonnage et constitue la référence majeure de notre chapitre. Signalons l'article en français de Guillin *et al.* (2005) qui offre une introduction très claire à la construction de ces méthodes, en évitant au maximum les notations mathématiques, ainsi qu'un aperçu historique de leur développement au sein de la communauté bayésienne.

¹Cette mesure sera induite et non exprimée dans les intégrales de ce chapitre, pour faciliter la lecture.

Les méthodes d'échantillonnage peuvent elles-mêmes se diviser en deux catégories :

- les *méthodes de Monte-Carlo par chaînes de Markov* (MCMC),
- les méthodes d'*échantillonnage préférentiel* (ou d'*importance sampling*).

Dans les deux cas, elles visent à simuler un échantillon $\theta_M = (\theta_1, \dots, \theta_M)$ de la distribution de densité $\pi(\theta|\mathbf{y}_n)$. Idéalement, obtenir un tel échantillon permet l'estimation

$$\phi_M = \frac{1}{M} \sum_{i=1}^M \Phi(X|\theta_i)$$

qui converge vers ϕ presque sûrement quand $M \rightarrow \infty$, de par la loi des grands nombres ; estimation dont la précision peut être estimée par

$$\sigma_{\phi_M}^2 = \frac{1}{M-1} \sum_{i=1}^M \{\Phi(\theta_i) - \phi_M\}^2.$$

Cependant, si l'objectif est le même, précisons que ces deux méthodes sont par construction très différentes. Les méthodes MCMC sont itératives et produisent à l'itération t un échantillonnage $\theta_M^{(t)}$ qui sont des réalisations d'une chaîne de Markov ayant pour distribution stationnaire $\pi(\theta|\mathbf{y}_n)$. Les algorithmes d'échantillonnage préférentiel produisent un échantillon simulé selon une distribution *instrumentale*, de densité idéalement proche de $\pi(\theta|\mathbf{y}_n)$; certaines parties de l'échantillon peuvent ensuite être conservées, après rééchantillonnage, de telle façon que l'échantillon final suive approximativement $\pi(\theta|\mathbf{y}_n)$.

Ce chapitre se divise en trois parties. Premièrement, nous faisons quelques brefs rappels sur les méthodes MCMC et les algorithmes d'échantillonnage préférentiel. Par la suite, nous concentrons notre étude sur ces derniers. En effet, nous proposons dans une seconde partie une analyse critique de l'algorithme BRM proposé par Bacha & Celeux (1996). Avantage par sa facilité d'utilisation, cet algorithme est ainsi employé dans le récent logiciel REXPERT (Procaccia & Procaccia 2005), dédié à la fiabilité industrielle. Nous montrons que la méthode présente cependant des défauts théoriques et peut provoquer des erreurs d'estimation importantes. Il est impératif, avant toute utilisation de cette technique, d'obtenir des indications sur la proximité de π et de \mathbf{y}_n . Le lien est ainsi fait avec le chapitre 4, où est introduit et étudié le critère DAC qui fournit de telles indications. Par ailleurs, nous faisons un certain nombre de propositions pour appliquer la méthode aux modélisations bayésiennes de Weibull développées au chapitre 3.

La troisième partie du chapitre est consacrée à l'utilisation de l'algorithme *Population Monte Carlo* (PMC), introduit par Cappé *et al.* (2004) puis amélioré par Douc *et al.* (2005) et Celeux *et al.* (2006). Après une description de la méthode (ainsi qu'une comparaison de construction avec BRM) nous étudions son application dans le cadre des modélisations bayésiennes de Weibull développées au chapitre 3. Une comparaison est faite, via le logiciel WinBUGS, avec les méthodes MCMC.

Rappelons que dans notre cadre d'étude, l'échantillon \mathbf{y}_n est souvent constitué de données manquantes, comprenant éventuellement des données incomplètes (voir chapitre 1). Notons \mathbf{z} les données manquantes du problème et $k(\mathbf{z}|\mathbf{y}_n, \theta)$ leur densité conditionnellement à l'échantillon observé. Dans un

contexte de données de durée de vie censurées à droite, les données manquantes \mathbf{z} sont les $n - r$ temps de défaillance survenant après les temps de censure observés c_1, \dots, c_{n-r} , supposés généralement fixés indépendamment de l'expérience. Alors

$$k(\mathbf{z}|\mathbf{y}_n, \theta) = \prod_{i=1}^{n-r} \frac{f(z_i|\theta)}{S(c_i|\theta)} \mathbb{1}_{]c_i, +\infty[}(z_i)$$

où S est la fonction de survie du modèle. D'une façon générale, la densité des données observées peut s'écrire sous la forme marginale

$$f_y(\mathbf{y}_n|\theta) = \int g(\mathbf{y}_n, \mathbf{z}|\theta) d\mathbf{z}$$

où g est donc la densité des données complètes. Tout au long de ce chapitre, nous accordons une place importante à la prise en compte de ces données manquantes dans les méthodes d'estimation.

5.2 Algorithmes d'échantillonnage

5.2.1 Méthodes MCMC

Outre notre référence principale au livre de Robert & Casella (2004), on trouve dans Bacha (1996) un résumé clair des principales méthodes MCMC ainsi que des remarques d'utilisation. Nous n'en faisons ici qu'un bref rappel.

Le principe général d'une méthode MCMC est le suivant. Partant d'une distribution de densité $\pi_0(\theta|\mathbf{y}_n)$, la méthode produit une chaîne de Markov de réalisations $\theta^{(1)}, \dots, \theta^{(t)}, \dots$, qui a pour distribution stationnaire $\pi(\theta|\mathbf{y}_n)$. Ainsi, la convergence de l'estimation

$$\phi_T = \frac{1}{T} \sum_{t=1}^T \Phi(\theta^{(t)})$$

vers ϕ est assurée par l'ergodicité de cette chaîne de Markov. Plus formellement, en notant t_0 le "temps de chauffe" de la chaîne - soit le nombre d'itérations t à partir duquel la chaîne de Markov peut être considérée en régime stationnaire - on a

$$\frac{1}{(T - t_0 + 1)} \sum_{t=t_0}^T \Phi(\theta^{(t)}) \xrightarrow{p.s.} \phi.$$

Une méthode MCMC est entièrement déterminée par le choix d'un *noyau markovien*, que l'on peut comprendre comme la généralisation au cadre continu de la matrice de transition d'une chaîne de Markov à états discrets. On peut le définir, au temps t de la méthode, par

$$\begin{aligned} K(\theta, A) &= P(\theta^{(t)} \in A | \theta^{(t-1)} = \theta), \\ &= \int_A \kappa(\theta, \tilde{\theta}) d\tilde{\theta} \end{aligned}$$

où $\tilde{\theta} \mapsto \kappa(\theta, \tilde{\theta})$ est une densité sur Θ (dite *densité de transition*). Parmi les méthodes MCMC, l'algorithme de Hastings-Metropolis (Hastings 1970) et l'algorithme de Gibbs (Geman & Geman 1984) sont les plus utilisées et ont donné lieu de nombreux algorithmes dérivés. Le logiciel BUGS/WinBUGS (Spiegelhalter *et al.* 1999) permet de simuler des chaînes MCMC et d'estimer les paramètres d'un modèle, par échantillonnage de Gibbs.

5.2.1.1 Algorithme d'Hastings-Metropolis

Cet algorithme produit une chaîne de Markov de noyau K_H en faisant appel, à l'itération $t \rightarrow t+1$, à deux étapes :

1. on simule des réalisations de θ en utilisant un noyau markovien *auxiliaire*, de densité $\kappa(\theta^t, \cdot)$ arbitrairement fixée ;
2. on procède à une étape d'acceptation-rejet sur ces réalisations selon une probabilité $p(\theta^t, \cdot)$ dépendante de $\kappa(\theta^t, \cdot)$.

À la fin de chaque itération, une partie des valeurs simulées à l'itération précédente est conservée et une autre partie est remise à jour *via* cette seconde étape. Au final, les échantillons produits ne sont pas *i.i.d.* La probabilité $p(\theta, \tilde{\theta})$ est construite de façon à ce que la loi cible $\pi(\theta|\mathbf{y}_n)$ soit une loi invariante pour la chaîne de noyau K_H , quel que soit le choix de $\kappa(\theta, \tilde{\theta})$. Il est cependant nécessaire que le rapport

$$\pi(\theta|\mathbf{y}_n)/\kappa(\theta|\tilde{\theta})$$

soit connu à une constante multiplicative près indépendante de $\tilde{\theta}$. Voir le chapitre 7 de Robert & Casella (2004) pour plus de précisions. L'itération $t \rightarrow t+1$ de l'algorithme s'écrit formellement comme suit. On rappelle la notation

$$\bar{\pi}(\theta|\mathbf{y}_n) = \mathcal{L}(\mathbf{y}_n; \theta) \pi(\theta).$$

Algorithme 2. Hastings-Metropolis.

Itération $t \rightarrow t+1$

1. simuler $\theta \sim \kappa(\theta^t, \cdot)$;
2. simuler $u \sim \mathcal{U}_{[0,1]}$;
3. calculer la probabilité d'acceptation

$$p(\theta^t, \theta) = \begin{cases} \min \left\{ 1, \frac{\bar{\pi}(\theta|\mathbf{y}_n)/\kappa(\theta^t, \theta)}{\bar{\pi}(\theta^t|\mathbf{y}_n)/\kappa(\theta, \theta^t)} \right\} & \text{si } \bar{\pi}(\theta^t|\mathbf{y}_n)\kappa(\theta^t, \theta) \neq 0, \\ 1 & \text{sinon.} \end{cases}$$

4. poser $\theta^{(t+1)} = \begin{cases} \theta & \text{si } u \leq p(\theta^t, \theta), \\ \theta^{(t)} & \text{sinon.} \end{cases}$.

Très usuellement, on choisit $\kappa(\theta, \tilde{\theta}) = \kappa(\tilde{\theta})$, ce qui permet de simplifier l'écriture de la probabilité d'acceptation. La densité $\kappa(\theta)$ doit alors mimer correctement le comportement de $\pi(\theta|\mathbf{y}_n)$ pour assurer une bonne convergence de l'algorithme.

5.2.1.2 Algorithme de Gibbs

Au chapitre 3, nous avons remarqué que la structure de la distribution *a posteriori* des paramètres (μ, β) du modèle de Weibull

$$\begin{aligned} \mu|\mathbf{y}_n, \beta &\sim \mathcal{G}(a, b(\beta)), \\ \beta|\mathbf{y}_n &\sim \pi(\beta|\mathbf{y}_n) \end{aligned}$$

permettait de simplifier la simulation des couples (μ, β) : sachant des réalisations de β , il est aisé de simuler des réalisations de μ de par la connaissance explicite de $\pi(\mu|\mathbf{y}_n, \beta)$. C'est sur de telles propriétés de dépendance conditionnelle entre paramètres que l'on construit l'algorithme de Gibbs.

Plus généralement, si l'on peut écrire $\theta_1, \dots, \theta_p$ une partition du vecteur des paramètres θ pour le modèle étudié, et si l'on peut spécifier complètement les lois de comportement *a posteriori* conditionnelles

$$\begin{aligned} & \pi(\theta_1|\theta_2, \dots, \theta_p, \mathbf{y}_n), \\ & \pi(\theta_2|\theta_1, \theta_3, \dots, \theta_p, \mathbf{y}_n), \\ & \dots \\ & \pi(\theta_{p-1}|\theta_1, \dots, \theta_{p-2}, \theta_p, \mathbf{y}_n), \\ & \pi(\theta_p|\theta_1, \dots, \theta_{p-1}, \mathbf{y}_n) \end{aligned}$$

alors on peut facilement simuler, pas par pas, des réalisations conditionnelles de $\theta_1, \dots, \theta_p$. En itérant le procédé un grand nombre de fois, la chaîne de Markov produite par ces simulations répétées a pour distribution stationnaire $\pi(\theta|\mathbf{y}_n)$. Formellement, l'algorithme s'écrit comme suit.

Algorithme 3. Gibbs.

Itération $t \rightarrow t+1$:

1. **simuler** $\theta_1^{(t+1)} \sim \pi(\theta_1|\theta_2^{(t)}, \dots, \theta_p^{(t)}, \mathbf{y}_n)$;
2. **pour** $j = 2, \dots, p-1$, **simuler** $\theta_j^{(t+1)} \sim \pi(\theta_j|\theta_1^{(t+1)}, \dots, \theta_{j-1}^{(t+1)}, \theta_{j+1}^{(t+1)}, \dots, \theta_p^{(t)}, \mathbf{y}_n)$;
3. **simuler** $\theta_p^{(t+1)} \sim \pi(\theta_p|\theta_1^{(t+1)}, \dots, \theta_{p-1}^{(t+1)}, \mathbf{y}_n)$.

Cependant, il arrive rarement que l'on puisse spécifier complètement toutes les lois conditionnelles. Si le conditionnement peut être optimisé de façon à obtenir le plus grand nombre possible de lois explicites, il est souvent nécessaire de coupler cet algorithme à des étapes d'Hastings-Metropolis pour obtenir la simulation d'une loi conditionnelle connue à une constante multiplicative près. Le tout forme donc un algorithme hybride qui, parfois, peut être lourd d'exécution.

5.2.1.3 Difficultés des méthodes MCMC

En contrepartie d'une relative facilité d'implémentation, les méthodes MCMC peuvent présenter des difficultés importantes.

La convergence - c'est-à-dire l'atteinte de l'équilibre de la chaîne de Markov - peut être très lente, surtout lorsque $\pi(\theta|\mathbf{y}_n)$ est difficile à approximer par une densité de transition κ facilement simulable. ; par ailleurs, une faible maniabilité de l'expression de $\pi(\theta|\mathbf{y}_n)$ rend cette convergence parfois difficile à prouver.

Par ailleurs, même si convergence il y a, détecter le "temps de chauffe" est loin d'être évident. Une importante littérature est ainsi consacrée à la recherche de critères d'arrêt des méthodes MCMC. Voir ainsi les livres de Robert (1998) et Robert & Casella (2004, chap. 12) et l'article de Chauveau & Diebolt (1999).

5.2.2 Méthodes d'échantillonnage préférentiel

Les algorithmes d'échantillonnage préférentiel, ou *importance sampling*, sont des méthodes de simulation qui, à partir d'un échantillon *i.i.d.* $\theta_M = \theta_1, \dots, \theta_M$ provenant d'une densité *instrumentale* ou d'*importance* $\rho(\theta)$, vont proposer une estimation de ϕ . En effet, puisque

$$\phi = \int_{\Theta} \Phi(\theta) \frac{\pi(\theta|\mathbf{y}_n)}{\rho(\theta)} \rho(\theta) d\theta,$$

une estimation naturelle de ϕ est

$$\phi_M = \sum_{i=1}^M \bar{\omega}_i \Phi(\theta_i) \quad (5.1)$$

où

$$\begin{aligned} \bar{\omega}_i &= \frac{\omega_i}{\sum_{j=1}^M \omega_j}, \\ \omega_i &= \omega(\theta_i|\mathbf{y}_n) = \frac{\bar{\pi}(\theta_i|\mathbf{y}_n)}{\rho(\theta_i)} = \frac{\mathcal{L}(\mathbf{y}_n; \theta_i) \pi(\theta_i)}{\rho(\theta_i)}. \end{aligned} \quad (5.2)$$

En application directe de la loi des grands nombres, sous l'hypothèse que $\text{Supp}(\pi(\theta|\mathbf{y}_n)) \subset \text{Supp}(\rho(\theta))$ et que $\int_{\Theta} \Phi(\theta) \pi(\theta|\mathbf{y}_n) d\theta < \infty$, on a

$$\phi_M \xrightarrow{p.s.} \phi.$$

Si de plus $\int_{\Theta} \Phi^2(\theta) \frac{\pi(\theta|\mathbf{y}_n)}{\rho(\theta)} \pi(\theta) d\theta < \infty$, alors

$$\sqrt{M} \{\phi_M - \phi\} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_{\phi}^2)$$

$$\text{où } \sigma_{\phi}^2 = \left\{ \int_{\Theta} \bar{\pi}(\theta|\mathbf{y}_n) d\theta \right\}^{-2} \left\{ \int_{\Theta} (\Phi(\theta) - \phi)^2 \omega(\theta|\mathbf{y}_n) \bar{\pi}(\theta|\mathbf{y}_n) d\theta \right\}.$$

Ce type de méthode peut apparaître ingénieux mais fragile dans des cas pratiques, l'estimation ϕ_M étant rendue fortement dépendante des poids ω_i . On peut résumer informellement un bon choix de ρ :

- $\rho(\theta)$ doit rester proche de $\pi(\theta|\mathbf{y}_n)$ tout en étant facile à simuler ; dans ce cas, les poids $\bar{\omega}_i$ restent relativement équilibrés et réduisent la variabilité de (5.1) ;
- la densité $\rho(\theta)$ doit posséder des queues au moins aussi lourdes que celles de $\pi(\theta|\mathbf{y}_n)$.

Tout l'art de l'échantillonnage préférentiel consiste donc à choisir finement une densité ρ qui, en plus de vérifier les conditions ci-dessus, doit présenter des caractéristiques délicates à vérifier (il faut par exemple établir des critères de proximité avec $\pi(\theta|\mathbf{y}_n)$, densité que l'on ne connaît pas explicitement). Ce qui explique que l'échantillonnage préférentiel ait été quelque peu délaissé, par le passé, au profit des méthodes MCMC. Cependant, différentes méthodes de construction de $\rho(\theta)$ ont été proposées. On peut les séparer en deux catégories.

1. Les méthodes *statiques*, où $\rho(\theta)$ est déterminée de façon unique en préalable à la simulation. Parmi elles, la méthode WLB (*Weighted Likelihood Bootstrap*), proposée par Newton (1991) et Newton & Raftery (1994) construit $\rho(\theta)$ de façon non paramétrique, par apprentissage à partir de l'information apportée par \mathbf{y}_n . Bacha & Celeux (1996) ont pointé les difficultés théoriques et concrètes rencontrées par l'utilisation de cette méthode. Ils ont proposé à la place l'algorithme BRM (*Bayesian Restoration Maximization*), qui fait l'objet d'une étude critique dans la section suivante. Celui-ci est simple d'utilisation, réservé aux modèles à données manquantes et n'admet pas d'*a priori* impropre. Il semble donc adapté aux problèmes bayésiens industriels mais rencontre des problèmes théoriques et appliqués.
2. Les méthodes *adaptatives*, où $\rho(\theta)$ est construite par itérations successives, en fonction de l'adaptation des simulations courantes $\theta_1, \dots, \theta_M \sim \rho(\theta)$ à la densité cible $\pi(\theta|\mathbf{y}_n)$. Une étude de l'algorithme PMC (*Population Monte Carlo*) est ainsi proposée au § 5.4.

Traditionnellement, les algorithmes d'échantillonnage préférentiel sont associés à une étape de ré-échantillonnage par densité de préférence, notée SIR (*Sampling Importance Resampling*, Rubin 1987 & 1988), qui permet d'obtenir à partir de $\theta_M \sim \rho(\theta)$ un échantillon $\tilde{\theta}_P$, non *i.i.d.*, de taille $P < M$ qui suit approximativement $\pi(\theta|\mathbf{y}_n)$. On procède de la manière suivante.

Algorithme 4. SIR.

1. Normaliser les poids ω_i en $\bar{\omega}_i = \omega_i / \sum_{j=1}^M \omega_j$.
2. Simuler les indicateurs $(i_1, \dots, i_M) \in \mathcal{N}^M$ selon la loi multinomiale $\mathcal{M}_M(\cdot | \bar{\omega}_1, \dots, \bar{\omega}_M)$ tels que $\sum_{k=1}^M i_k = P$.
3. Sélectionner $(\tilde{\theta}_1, \dots, \tilde{\theta}_P) = (\underbrace{\theta_1, \dots, \theta_1}_{i_1 \text{ fois}}, \underbrace{\theta_2, \dots, \theta_2}_{i_2 \text{ fois}}, \dots, \underbrace{\theta_M, \dots, \theta_M}_{i_M \text{ fois}})$.

Ce faisant, par l'application de la loi des grands nombres, $\tilde{\theta}_P$ suit approximativement $\pi(\theta|\mathbf{y}_n)$ lorsque $M/P \rightarrow \infty$. Rubin (1987) considère (sans justification théorique) qu'un rapport $M/P = 20$ est généralement suffisant. Dans les applications, on choisira $M/P = 30$.

REMARQUE 11. *Vis-à-vis des MCMC, les méthodes d'échantillonnage préférentiel apportent le gain suivant : un échantillon d'importance est réutilisable lorsqu'on décide de modifier $\pi(\theta)$ en $\pi^*(\theta)$ (par exemple pour une étude de sensibilité). Puisque*

$$\pi^*(\theta|\mathbf{y}_n) \propto \frac{\pi^*(\theta)}{\pi(\theta)} \pi(\theta|\mathbf{y}_n),$$

il est aisé d'obtenir un nouvel estimateur a posteriori en modifiant les poids ω_i en

$$\omega_i^* = \frac{\pi^*(\theta_i)}{\pi(\theta_i)} \omega_i.$$

Ainsi, on peut se ramener à l'utilisation d'a priori conjugués ou pratiques pour estimer a posteriori une fonction d'intérêt.

5.3 L'algorithme BRM : une étude critique

5.3.1 Définition de la méthode

L'algorithme BRM propose la construction de la densité instrumentale $\rho(\theta)$ à partir de l'information apportée par les données \mathbf{y}_n mais en utilisant aussi l'information *a priori* modélisée par $\pi(\theta)$. Il n'est applicable que dans le cadre des modèles à données manquantes. Ces données manquantes sont reconstituées à partir d'une simulation marginale *a priori*, ce qui implique que π doit être propre. Le principe de l'algorithme est le suivant : en maximisant la vraisemblance de M échantillons reconstitués, on obtient alors un échantillon $\hat{\theta}_M$ d'estimations de θ . En choisissant $\rho(\theta)$ comme la densité de $\hat{\theta}_M$ (ou comme un estimateur non paramétrique de cette densité), $\rho(\theta)$ est intuitivement susceptible d'être proche de $\pi(\theta|\mathbf{y}_n)$. Selon le but à atteindre, on peut coupler cette méthode avec des méthodes de rééchantillonnage. Le couplage BRM-SIR est ainsi conseillé pour obtenir des réalisations de la loi *a posteriori*. Voir Bacha (1996, chapitre 2) pour plus de précisions sur le choix à effectuer.

Soit \mathbf{z} les données manquantes. L'algorithme s'écrit alors, avec les notations précisées dans l'introduction de ce chapitre,

1. Pour $i = 1, \dots, M$
 - (B) simuler $\theta_i \sim \pi(\theta)$;
 - (R) simuler $\mathbf{z}_i \sim k(\mathbf{z}|\mathbf{y}_n, \theta_i)$;
 - (M) calculer $\hat{\theta}_i = \arg \max_{\Theta} g(\mathbf{y}_n, \mathbf{z}_i|\theta)$.
2. On choisit ρ comme la densité des tirages $\hat{\theta}_M = (\hat{\theta}_1, \dots, \hat{\theta}_M)$.

En général, la densité ρ n'est pas explicite et une estimation non paramétrique $\hat{\rho}$ est nécessaire. Ce faisant, on "lisse" la répartition discrète des valeurs $\hat{\theta}_1, \dots, \hat{\theta}_M$ par le biais de fonctions continues centrées sur les θ_i . La technique utilisée est la méthode des noyaux gaussiens qui estime $\rho(\theta)$ par

$$\hat{\rho}(\theta) = \frac{1}{M|F|} \sum_{i=1}^M N(F^{-1}(\theta_i - \theta)) \quad (5.3)$$

où F est une matrice carrée non singulière d'ordre $d = \dim \Theta$ et $N(x)$ est la densité de la loi $\mathcal{N}(0, I_d)$ avec I_d l'identité en dimension d . La matrice F est choisie selon le principe du lissage maximum (Terrel 1990) qui cherche à minimiser

$$\mathbb{E} \left[\left(\int_{\Theta} \hat{\rho}(\theta) d\theta - \rho(\theta) \right)^2 \right].$$

Le cas exponentiel est une exception, ρ étant explicite, et fait l'objet de la sous-partie suivante. Il permet de montrer que le caractère "intuitif" de l'algorithme peut être trompeur et que ρ n'est malheureusement pas toujours un bon choix de densité d'importance. De fait, le gros défaut de BRM est de rendre les poids ω_i de la formule (5.2) parfois très dépendants des valeurs $\hat{\theta}_M$. Supposons par exemple qu'il y ait peu de données manquantes. Dans ce cas, les MLE $\hat{\theta}_i$ sont très proches les uns des autres, et les queues de $\rho(\theta)$ risquent d'être plus légères que celles de $\pi(\theta|\mathbf{y}_n)$. Ce phénomène est amplifié par le choix de noyaux gaussiens. En conséquence, la variance empirique de l'estimation risque d'être abusivement faible.

5.3.2 Difficultés et précautions

Afin donc de mesurer les difficultés théoriques de BRM, nous étudions son application au modèle exponentiel. Il s'agit d'un cas très simple, pour lequel la densité $\rho(\theta)$ des $(\hat{\theta}_j)$ est explicite et ne nécessite pas de construire une estimation non paramétrique lorsque π est conjugué. Bien entendu, nous n'avons nul besoin d'un algorithme d'échantillonnage puisque $\pi(\theta|\mathbf{y}_n)$ est alors entièrement connue et aisément simulable. Nous montrons sur ce cas simple que ρ s'avère un candidat dangereux car borné sur Θ .

Soit $f(x|\lambda)$ la densité du modèle exponentiel $\mathcal{E}(\lambda)$. On place sur λ un *a priori* conjugué $\mathcal{G}(a, b)$ de densité

$$\pi(\lambda) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} \exp(-b\lambda) \mathbb{1}_{\{\lambda>0\}}.$$

On suppose que l'échantillon observé \mathbf{y}_n contient $0 \leq r < n$ valeurs (x_1, \dots, x_r) non censurées et $\mathbf{c}_{n-r} = (c_1, \dots, c_{n-r})$ valeurs censurées fixées (type-I). Alors, en notant $t(\mathbf{y}_n) = \sum_{i=1}^n y_i$, on a

$$\hat{\lambda}_i = \frac{n}{t(\mathbf{y}_n) + t_i}$$

où $t_i = \sum_{k=1}^{n-r} z_k^i$ et, de par la propriété d'absence de mémoire,

$$\begin{aligned} z_{|\lambda}^k &\sim \mathcal{E}(\lambda), \\ t_{i|\lambda} &\sim \mathcal{G}(n-r, \lambda). \end{aligned}$$

La densité inconditionnelle de t_i est alors

$$\begin{aligned} f_T(t) &= \int_{\mathbb{R}^+} f_T(t|\lambda) \pi(\lambda) d\lambda, \\ &= \frac{\Gamma(n-r+a)b^a}{\Gamma(n-r)\Gamma(a)(b+t)^{a+n-r}} t^{n-r-1}. \end{aligned}$$

Par un simple changement de variable, on trouve

$$\rho(\lambda) = \frac{\Gamma(n-r+a)}{\Gamma(n-r)\Gamma(a)} \left(\frac{(n-t(\mathbf{y}_n)\lambda)^{-1}}{(n-(t(\mathbf{y}_n)-b)\lambda)^a} \right)^{n-r} n b^a \lambda^{a-1} \mathbb{1}_{\{0 \leq \lambda < n/t(\mathbf{y}_n)\}}.$$

La densité des MLE reconstitués est donc bornée sur $[0, n/t(\mathbf{y}_n)[$, ce qui induit que la densité-cible réelle de l'échantillonnage est la restriction de $\pi(\lambda|\mathbf{y}_n)$ sur $[0, n/t(\mathbf{y}_n)[$. Ainsi, ρ représente théoriquement un mauvais choix de densité d'importance.

Cependant, ce défaut résulte du fait que l'on privilégie la connaissance de la densité $\rho(\theta)$ à toute reconstruction non paramétrique. Cette lecture "myope" de l'algorithme permet certes de mettre en valeur la fragilité de BRM, mais on peut imaginer un bon choix de noyaux non bornés, avec des queues acceptables, qui permettrait d'évacuer cette difficulté théorique. L'utilisation de noyaux gaussiens dans (5.3) ayant tendance à exagérer la concentration des MLE reconstitués, des noyaux de Cauchy de densités

$$\rho_C(\theta|\alpha, \theta_0) = \left(\frac{\alpha}{\pi}\right) \frac{1}{(\theta - \theta_0)^2 + \alpha^2}$$

paraissent susceptibles d'améliorer l'estimation non paramétrique de ρ . En effet, les lois de Cauchy sont connues pour n'avoir ni espérance ni variance, et les variables de Cauchy n'obéissent pas au théorème central limite ; la probabilité d'obtenir des valeurs éloignées de θ_0 est en fait trop élevée pour

permettre à la moyenne empirique de converger. Ainsi, cette loi se caractérise par un fort étalement et des queues très lourdes.

REMARQUE 12. *Par ailleurs, pour des raisons déjà évoquées au chapitre 4, un a priori industriel π reste dans les faits proche des données. Ainsi, comme le note Celeux (1996), en dépit du fait que le support de ρ est borné, la procédure BRM, dans le cas exponentiel, permet d'estimer convenablement la majeure partie de la masse de $\pi(\cdot|\mathbf{y}_n)$.*

D'une façon générale, l'utilisation de BRM est simple (et possiblement améliorable), mais son gros défaut est de fournir potentiellement des poids ω_i trop déséquilibrés en fonction de l'éloignement entre π et \mathbf{y}_n . On propose donc de précéder l'emploi de BRM du calcul du critère DAC défini au chapitre 4, qui semble bien adapté au problème. Si jamais $\text{DAC} > 1$ ou si le pourcentage de données censurées est faible, on proscrie l'utilisation de BRM et un autre choix de densité d'importance doit être fait. Un examen "visuel" de la proximité entre information *a priori* et objective peut suffire lorsque la dimension de θ est restreinte à 1 ou 2. Dans des cas de dimension supérieure, elle devient moins évidente à appréhender, d'où l'utilité de DAC.

5.3.3 Une densité instrumentale de référence

Dans le cadre du chapitre 3, on a défini $\pi(\beta)$ sur un compact $[\beta_l, \beta_r]$ pour le modèle de Weibull. Dans ce même chapitre, nous faisons un choix *a priori* nous permettant d'obtenir explicitement $\pi(\mu|\beta, \mathbf{y}_n)$:

$$\begin{aligned}\mu|\beta &\sim \mathcal{G}(a, b(a, \beta)), \\ \beta &\sim \mathcal{B}_e(p, q) \text{ définie sur } [\beta_l, \beta_r].\end{aligned}$$

Ainsi, on s'assure que la densité instrumentale

$$\begin{aligned}\rho_R(\mu, \beta) &= \pi(\theta'|\beta, \mathbf{y}_n)\rho_R(\beta), \\ \beta &\sim \mathcal{U}_{[\beta_l, \beta_r]}\end{aligned}$$

respecte les conditions souhaitées par Geweke (1989) et conserve des queues toujours plus lourdes que $\pi(\theta|\mathbf{y}_n)$. On a alors $\omega(\theta', \beta|\mathbf{y}_n) = \omega(\beta|\mathbf{y}_n)$ et

$$\omega(\beta|\mathbf{y}_n) = \beta^r \left(\prod_{i=1}^r x_i \right)^\beta (\beta - \beta_l)^{p-1} (\beta_r - \beta)^{q-1} \frac{b^a(a, \beta)}{\left(b(a, \beta) + \sum_{i=1}^n y_i^\beta \right)^{a+r}} \mathbb{1}_{\{\beta_l \leq \beta \leq \beta_r\}}.$$

Ce choix de référence nous permet donc, au travers d'applications numériques, d'apprécier le comportement de la densité instrumentale proposée par BRM.

5.3.4 Comparaisons

On simule 30 échantillons de taille $n = 20$ selon le modèle de Weibull de paramètres $(\eta_0, \beta_0) = (112, 3)$ (soit d'espérance 100 et de variance 36^2). On remplace $n - r$ valeurs de cet échantillon par une censure simulée selon la loi uniforme sur $[110, 130]$. Quatre pourcentages de censure seront considérés : 5%, 25%, 50%, 75%. On fixe $[\beta_l, \beta_r] = [1, 5]$ et $p = q$. Ainsi, $\pi(\beta)$ est toujours centrée sur β_0 . On

choisit deux valeurs $p = 1.5$ et $p = 15$, ce qui entraîne $\sigma_\beta = 1$ et $\sigma_\beta = 0.36$. D'autre part, on suppose connaître la médiane marginale *a priori* x_e . On utilise donc, d'après le chapitre 3, la formule

$$b(a, \beta) = (2^{1/a} - 1)^{-1} x_e^\beta.$$

On choisit $x_e = 100$, ce qui est parfaitement en accord avec les données, et $x_e = 50$, ce qui témoigne d'une expertise très pessimiste. Enfin, a correspondant à une taille d'échantillon fictif, on sélectionne $a = 3$ et $a = 10$. Dans le deuxième cas, s'il y a forte censure, l'*a priori* apporte une information sur θ' plus importante que les données.

Dans les tableaux 5.1 et 5.2, on présente pour chaque choix de densité instrumentale, selon plusieurs caractéristiques de π , un triplet de valeurs correspondant aux estimations de Monte-Carlo de la durée de vie moyenne *a posteriori* (l'espérance prédictive sur X), notée D , de son écart-type σ_D et le pourcentage de survie prédictif à $t_0 = 130$ (soit la valeur maximale de la censure), noté S_{130} . Ces estimations sont calculées après une étape SIR sur les simulations des différentes densités instrumentales, à partir de $M = 10^4$ particules. À titre de comparaison, on indique également l'estimation moyenne de ces quantités par maximisation de la vraisemblance (EMV).

De façon quasi-systématique, la méthode BRM sous-estime σ_D . Comme prévu, cette estimation peut être fortement erronée lorsque la censure est importante et que l'*a priori* est éloigné des données (cf. tableau 5.2, 3ème ligne). L'estimation de D souffre elle aussi de ce conflit potentiel entre données et *a priori*. Lorsque l'*a priori* est faiblement informatif vis-à-vis des données ($a = 3$, $\sigma_\beta = 1$), l'estimation de D par BRM reste très dépendante de $x_e = 50$. Par ailleurs, ces défauts d'estimation et de proximité avec $\pi(\theta|\mathbf{y}_n)$ induisent dans cet exemple des taux de survie surévalués.

a	σ_β	% cens.	EMV	BRM	référence (ρ_R)
3	0.36	25%	112 39 31%	110 40 30%	110 40 31%
		50%	129 39 48%	128 40 48%	126 46 45%
10	1	5%	100 35 21%	101 33 21%	100 37 21%
		25%	115 39 35%	112 35 31%	111 42 31%
		50%	133 43 52%	122 38 42%	123 49 51%

TAB. 5.1 – Triplets (D, σ_D, S_{130}) pour un *a priori* en accord avec les données ($x_e = 100$). BRM offre une estimation raisonnable des quantités *a posteriori*, vis-à-vis de l'estimation de référence (dernière colonne).

a	σ_β	% cens.	EMV	BRM	référence (ρ_R)
3	1	25%	113 39 33%	112 37 32%	108 48 30%
		50%	127 40 48%	124 34 44%	128 61 43%
		75%	162 53 70%	138 35 60%	133 114 62%
	0.36	25%	111 34 30%	112 37 32%	105 40 27%
		50%	130 42 49%	127 38 47%	123 48 42%
		75%	158 47 71%	150 47 64%	154 61 61%
10	1	5%	102 35 22%	104 39 25%	86 45 16%
		25%	114 37 33%	115 39 34%	97 55 24%
		50%	129 41 48%	131 50 48%	119 81 36%

TAB. 5.2 – Triplets (D, σ_D, S_{130}) pour un *a priori* décalé avec les données ($x_e = 50$). BRM sous-estime fortement σ_D dans de nombreux cas, et l'estimation de la durée de vie D est parfois erronée (3 dernières lignes), vis-à-vis de l'estimation de référence (dernière colonne).

5.4 L'algorithme Population Monte-Carlo

5.4.1 Définition

La recherche d'une bonne densité instrumentale $\rho(\theta)$ est l'enjeu de cette section. Nous l'avons vu précédemment, le choix de $\rho(\theta)$ est loin d'être évident, malgré la simplicité des conditions de Geweke. Cappé *et al.* (2004), sous le nom d'algorithme PMC (*Population Monte Carlo*), ont proposé d'introduire une dimension itérative dans la production d'échantillons préférentiels dans le but d'adapter $\rho(\theta)$ à la densité cible $\pi(\theta|\mathbf{y}_n)$. Cette adaptation est implémentée de façon séquentielle : au pas t de l'algorithme, M réalisations de θ sont simulées à partir d'une densité instrumentale courante dépendant des M réalisations du pas $t-1$. À l'étape finale $t=T$, les M dernières réalisations sont utilisées pour estimer ϕ_M selon l'expression (5.1).

Au fur et à mesure des itérations, le bon ou le mauvais comportement de la densité instrumentale courante est jugé à travers la distribution des poids d'importance, et idéalement, la forme de $\rho(\theta)$ se rapproche de celle de $\pi(\theta|\mathbf{y}_n)$. L'avantage de cette méthode est de permettre une grande généralité de choix instrumentaux à chaque pas de l'algorithme, et d'aboutir potentiellement à une meilleure efficacité que les algorithmes classiques d'échantillonnage préférentiel. Par ailleurs, un gain important de la méthode est le coût plutôt faible d'itérations en t avant de parvenir à la stabilité de la distribution instrumentale.

On trouve dans l'article de Guillin *et al.* (2005) un résumé très clair de la construction de PMC et une comparaison de l'algorithme avec les méthodes MCMC ; le bon comportement de PMC vis-à-vis des résultats parfois insatisfaisants des méthodes MCMC est ainsi mis en lumière. Au cours de cette section, nous produirons quelques comparaisons numériques, dans notre contexte d'étude industrielle, qui illustreront là aussi le meilleur comportement de PMC.

Différences avec BRM. Alors que BRM est un algorithme d'échantillonnage statique, PMC est *dynamique* : il agit de manière adaptative. Nous verrons qu'un tirage *a priori* est nécessaire pour PMC ; cependant celui-ci ne sert qu'à l'initialisation de l'algorithme. On peut donc utiliser PMC pour estimer une grandeur *a posteriori*, pour laquelle l'*a priori* est non informatif, à la différence de BRM ; il suffit de simuler initialement une loi faiblement informative. Par ailleurs, BRM n'est utilisable qu'en présence de données manquantes ; l'utilisation de PMC n'y est pas restreinte.

5.4.1.1 Forme générique

L'implémentation de l'algorithme s'articule autour de l'itération des étapes suivantes :

1. on simule M variables aléatoires (ou *particules*) suivant une densité choisie *a priori* ou courante ;
2. on calcule le poids de chaque particule suivant (5.2) et on rééchantillonne parmi les M particules, proportionnellement à ces poids, pour obtenir un nouvel échantillon.

On note $\mathcal{F}_{M,t}$ l'ensemble des particules $\theta_{1,t}, \dots, \theta_{M,t}$ et des poids $\omega_{1,t}, \dots, \omega_{M,t}$ calculés à l'itération t de l'algorithme. Ces M particules sont ainsi simulées selon M distributions instrumentales $\rho_{i,t}(\theta)$ ($i = 1, \dots, M$), densités qui sont donc construites sachant $\mathcal{F}_{M,t-1}$. En effet, toutes les informations d'échantillonnage et de rééchantillonnage à l'itération $t-1$ de l'algorithme sont potentiellement

utilisables pour adapter $\rho(\theta)$ à l'itération t . La construction proprement dite des densités $\rho_{i,t}(\theta)$ et l'utilisation de la connaissance passée symbolisée par $\mathcal{F}_{M,t}$ fera l'objet du § 5.4.1.4.

Algorithme 5. PMC générique.

1. Au temps $t = 0$: faire un choix *a priori* de $\theta_{1,0}, \dots, \theta_{M,0}$.
2. Au temps $t = 1, \dots, T$, pour $i = 1, \dots, M$:
 - (a) simuler $\theta_{i,t} \sim \rho_{i,t}(\theta)$ et calculer $\omega_{i,t} = \frac{\mathcal{L}(\mathbf{y}_n; \theta_{i,t}) \pi(\theta_{i,t})}{\rho_{i,t}(\theta_{i,t})}$;
 - (b) calculer les poids normalisés $\bar{\omega}_i$ et rééchantillonner les particules courantes en utilisant les poids $\bar{\omega}_i$;
 - (c) construire $\rho_{i,t+1}$ sachant $\mathcal{F}_{M,t}$.

Un estimateur asymptotiquement sans biais de ϕ est alors donné par

$$\phi_{M,T} = \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^M \bar{\omega}_{i,t} \phi(\theta_{i,t}).$$

La variance de cet estimateur décroît en T et M . Par ailleurs, pour tout $t = 1, \dots, T$, sous des conditions peu restrictives, $\sum_{i=1}^M \bar{\omega}_{i,t} \Phi(\theta_{i,t})$ converge en probabilité vers ϕ et un théorème central limite s'applique quand $M \rightarrow \infty$. Voir Douc *et al.* (2005).

5.4.1.2 Données manquantes

L'algorithme est modifié en ajoutant une étape où les données manquantes \mathbf{z} sont reconstituées via la simulation courante des paramètres. Celeux *et al.* (2006) proposent une version de PMC qui mime un échantillonnage de Gibbs en produisant les \mathbf{z} et les θ selon leurs distributions conditionnelles respectives. On fait en effet le choix d'une nouvelle densité instrumentale

$$\rho_{i,t}^J(\theta, \mathbf{z}) = k(\mathbf{z} | \mathbf{y}_n, \theta_{i,t-1}) \rho_{i,t}(\theta | \mathbf{z}) \quad (5.4)$$

où $\rho_{i,t}(\theta | \mathbf{z})$ doit être proche de $\pi(\theta | \mathbf{y}_n, \mathbf{z})$. Le poids non normalisé s'écrit alors (voir l'introduction du chapitre pour les notations)

$$\omega_{i,t} = \frac{g(\mathbf{y}_n, \mathbf{z}_{t,i} | \theta_{i,t}) \pi(\theta_{i,t})}{k(\mathbf{z}_{t,i} | \mathbf{y}_n, \theta_{i,t-1}) \rho_{i,t}(\theta_{i,t} | \mathbf{z}_{t,i})}. \quad (5.5)$$

et l'algorithme prend la forme suivante.

Algorithme 6. PMC à données manquantes.

- Au temps $t = 1, \dots, T$:
1. pour $i = 1, \dots, M$:
 - (a) simuler de façon indépendante $\mathbf{z}_{i,t} \sim k(\mathbf{z} | \mathbf{y}_n, \tilde{\theta}_{i,t-1})$ et $\theta_{i,t} \sim \rho_{i,t}(\theta | \mathbf{z}_{i,t})$,
 - (b) calculer le poids (5.5) ;

2. rééchantillonner les particules courantes selon les poids normalisés ;
3. construire $\rho_{i,t+1}$ sachant $\mathcal{F}_{M,t}$, pour $i = 1, \dots, M$ (cf. § 5.4.1.4).

Pour certains problèmes, la reconstitution des données manquantes n'est pas rédhibitoire pour une bonne estimation bayésienne, par exemple *via* des algorithmes d'Hastings-Metropolis (cf. Celeux *et al.* 2000). Mais généralement, cette estimation est difficile, voire impossible, et se priver des possibilités de complétion d'un modèle à données manquantes du type

$$\mathbf{z}|\mathbf{y}_n, \theta \sim k(\mathbf{z}|\mathbf{y}_n, \theta) \propto g(\mathbf{y}_n, \mathbf{z}|\theta),$$

qui facilite le calcul des poids, est dommageable.

5.4.1.3 Réduction de variance par rao-blackwellisation

La grande variabilité des poids (5.5), accentuée par la présence de données manquantes simulées, peut être diminuée en considérant une approximation de la densité instrumentale marginale de $\theta_{i,t}$

$$\rho_{i,t}^J(\theta) = \int \rho_{i,t}^J(\theta, \mathbf{z}) d\mathbf{z}. \quad (5.6)$$

En effet, la variance de la distribution $\rho_{i,t}^J(\theta)$ sur l'ensemble des $\theta_{i,t}$ est inférieure à celle de la distribution jointe $\rho_{i,t}^J(\theta, \mathbf{z})$ sur l'ensemble des couples $(\theta_{i,t}, \mathbf{z}_{i,t})$, de par le théorème de Rao-Blackwell (Gelfand & Smith 1990, Robert & Casella, 2004, p. 130). Connue pour accroître la stabilité des schémas d'augmentation de données, cette stratégie porte le nom de rao-blackwellisation dans la terminologie des MCMC. Si cette stratégie ne modifie pas les propriétés de convergence de PMC, elle requiert cependant la possibilité de séparer la simulation des données manquantes et des valeurs courantes du paramètre θ .

Une estimation de (5.6) par *importance sampling* permet de réutiliser l'ensemble des données manquantes qui ont été simulées. On obtient alors

$$\rho_{i,t}^J(\theta_{i,t}) \simeq d_{i,t} = \frac{1}{M} \sum_{l=1}^M \frac{k(\mathbf{z}_{l,t}|\mathbf{y}_n, \theta_{i,t-1}) \rho_{i,t}(\theta_{i,t}|\mathbf{z}_{l,t})}{k(\mathbf{z}_{l,t}|\mathbf{y}_n, \theta_{l,t-1})}.$$

Les poids $\omega_{i,t}$ profitent de cette stratégie en approximant à la fois les distributions *a posteriori* et instrumentales marginales :

$$\omega_{i,t} = \frac{n_{i,t}}{d_{i,t}}$$

où

$$n_{i,t} = \frac{1}{M} \sum_{l=1}^M \frac{g(\mathbf{y}_n, \mathbf{z}_{l,t}|\theta_{i,t}) \pi(\theta_{i,t})}{k(\mathbf{z}_{l,t}|\mathbf{y}_n, \theta_{l,t-1})}.$$

Notons cependant que le calcul de l'estimateur rao-blackwellisé exige un coût quadratique $\mathcal{O}(M^2)$, qui a tendance à ralentir l'algorithme.

5.4.1.4 Construction itérative de densités instrumentales

Nous nous intéressons au choix de la densité instrumentale courante $\rho_{i,t}(\theta|\mathbf{z})$ dans la formulation générique (5.4). Rappelons que la construction de cette densité est conditionnée à la connaissance des

simulations et poids d'importance produits à l'itération $t - 1$ de l'algorithme. Cette connaissance est, dans les raisonnements précédents, symbolisée par $\mathcal{F}_{M,t-1}$.

Cappé *et al.* (2004) proposent de choisir un mélange de D noyaux de transition Q_k ($1 \leq k \leq D$) différents que l'on définit par

$$Q_k(\theta', A) = \int_A q_k(\theta', \theta) d\theta \quad (5.7)$$

où $A \subset \Theta$ et $\theta \mapsto q_k(\theta', \theta)$ est une densité fixée dès le départ de l'algorithme². Le vecteur θ' est donc un paramètre permettant de mettre à jour les caractéristiques de q_k ; un exemple simple est celui d'une densité gaussienne dont on actualise l'espérance en utilisant l'estimation $\theta_{i,t-1}$:

$$q_k(\theta_{i,t-1}, \theta) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} (\theta - \theta_{i,t-1})^2 \right\}.$$

Ainsi,

$$\rho_{i,t}(\theta | \mathbf{z}_{i,t}) = \sum_{k=1}^D \alpha_k^t q_k(\theta_{i,t-1}, \theta) \quad (5.8)$$

où $0 \leq \alpha_k^t \leq 1$ et $\sum_{k=1}^D \alpha_k^t = 1$. L'évolution des poids $\alpha_1^t, \dots, \alpha_k^t$ dirige seule l'adaptation séquentielle de ρ à $\pi(\theta | \mathbf{y}_n)$. L'approximation souhaitée est d'autant meilleure que D et T sont grands. Ces poids sont logiquement initialisés par $1/D$. En l'état actuel de l'art, deux grandes méthodes ont été proposées pour établir une règle de modification des α_k^t au cours des itérations t .

1. D-kernel PMC. Cappé *et al.* (2004) utilisent un indicateur $K_{i,t}$ de taille M qui indique l'origine de la simulation $\theta_{i,t}$

$$\theta_{i,t} \sim Q_{K_{i,t}}(\theta_{i,t-1}, \cdot).$$

Le nouveau poids est alors calculé comme une proportion du nombre de $\theta_{i,t}$ qui "survivent" à l'étape SIR du pas t de l'algorithme :

$$\alpha_k^{t+1} = \sum_{i=1}^M \bar{\omega}_{i,t} \mathbb{1}_k(K_{i,t})$$

où $\bar{\omega}_{i,t}$ est la normalisation de $\omega_{i,t}$, avec

$$\omega_{i,t} = \frac{g(\mathbf{y}_n, \mathbf{z}_{t,i} | \theta_{i,t}) \pi(\theta_{i,t})}{k(\mathbf{z}_{t,i} | \mathbf{y}_n, \theta_{i,t-1}) q_{K_{i,t}}(\theta_{i,t-1}, \theta_{i,t})}.$$

Cependant, Douc *et al.* (2005) prouvent que cette méthode présente un défaut majeur : lorsque M augmente, les poids de tous les noyaux convergent vers $1/D$, ce qui démontre que l'algorithme n'a rien d'adaptatif. Le défaut de la méthode est de ne pas établir de règle d'adaptation à partir de laquelle dériver la construction du poids $\omega_{i,t}$.

²en considérant que $\theta' \mapsto Q_k(\theta', \cdot)$ est dominé par la mesure de référence ν indiquée en introduction du chapitre.

2. D-kernel PMC rao-blackwellisé. Douc *et al.* (2005) proposent de rao-blackwelliser l'utilisation des noyaux dans les poids de l'algorithme, en choisissant

$$\omega_{i,t} = \frac{g(\mathbf{y}_n, \mathbf{z}_{t,i} | \theta_{i,t}) \pi(\theta_{i,t})}{k(\mathbf{z}_{t,i} | \mathbf{y}_n, \theta_{i,t-1}) \sum_{k=1}^D \alpha_k^t q_{K_{i,t}}(\theta_{i,t-1}, \theta_{i,t})}. \quad (5.9)$$

En effet les auteurs prouvent que cette version de PMC s'adapte à la densité cible $\varphi(\theta) = \pi(\theta | \mathbf{y}_n)$ en convergeant vers un optimum au sens de la divergence de Kullback-Leibler. En d'autres termes, la quantité

$$\int \int \log \left(\frac{\varphi(\theta) \varphi(\theta')}{\varphi(\theta) \sum_{k=1}^D \alpha_k q_k(\theta, \theta')} \right) (\varphi \otimes \varphi)(d\theta, d\theta')$$

décroit à chaque pas de l'algorithme. Par ailleurs, une loi faible des grands nombres et un théorème de limite centrale assurent la convergence en probabilité et en loi du mélange vers $\varphi(\theta)$ sous les conditions suivantes, peu restrictives.

A1 : $\forall k \in \{1, \dots, D\}, \varphi \otimes \varphi \{q_k(\theta, \theta') = 0\} = 0$; cette condition indique que la distribution des poids d'importance est finie presque sûrement.

A2 : $\forall k \in \{1, \dots, D\}, \int \int |\log q_k(\theta, \theta')| (\varphi \otimes \varphi)(d\theta, d\theta') < \infty$; cette condition est automatiquement remplie si tous les q_k dominent φ , au sens où φ/q_k est bornée.

Le choix d'un mélange fait de PMC une méthode flexible, permettant d'obtenir en général un bon approximateur de $\pi(\theta | \mathbf{y}_n)$, explicite et donc réutilisable. En effet, la convergence de l'algorithme est liée à la stabilité des α_k^t au cours des itérations t . Si ces poids évoluent peu, l'algorithme a sélectionné le mélange le plus proche de $\pi(\theta | \mathbf{y}_n)$ au sens de Kullback-Leibler. On dispose ainsi d'une règle intuitive pour stopper l'algorithme au temps T . Comme nous l'observerons dans les exemples, le nombre d'itérations T est souvent très faible (pas plus d'une dizaine généralement), ce qui est très avantageux vis-à-vis des méthodes MCMC, habituellement coûteuses. Cependant, un cas où tous les α_k^T sont nuls sauf un (qui vaut nécessairement 1) indique que le mélange proposé est initialement un mauvais candidat. L'application concrète de la méthode est évidemment facilitée par le choix d'un mélange de densités aisément simulables. Selon la variance de la densité de probabilité finale 1, on peut juger si le mélange initial est trop ou pas assez informatif, et le modifier en conséquence.

5.4.1.5 Forme générique finale de PMC

De par la présence des données manquantes dans nos échantillons, on peut enfin appliquer la rao-blackwellisation complète proposée au § 5.4.1.2. Remarquons que dans (5.8), la complétion des données manquantes intervient de façon sous-jacente par l'utilisation de $\theta_{i,t-1}$. Autrement dit, les données manquantes simulées à l'itération t de l'algorithme apportent une information qui n'est prise en compte par le mélange qu'à l'itération $t + 1$. L'algorithme final prend alors la forme suivante.

Algorithme 7. D-kernel PMC rao-blackwellisé, à données manquantes.

- Au temps $t = 0$:
 1. faire un choix *a priori* de $\theta_{1,0}, \dots, \theta_{M,0}$;
 2. fixer $\alpha_k^1 = 1/D$ pour $k = 1, \dots, D$.
- Au temps $t = 1, \dots, T$:
 1. Pour $i = 1, \dots, M$,
 - (a) simuler par la loi multinomiale les origines $K_{i,t} \sim \mathcal{M}_M(\alpha_1^t, \dots, \alpha_D^t)$;
 - (b) simuler de façon indépendante $\mathbf{z}_{i,t} \sim k(\mathbf{z}|\mathbf{y}_n, \theta_{i,t-1})$ et $\theta_{i,t} \sim Q_{K_{i,t}}(\theta_{i,t-1}, \cdot)$;
 - (c) calculer $n_{i,t} = \frac{1}{M} \sum_{l=1}^M \frac{g(\mathbf{y}_n, z_{i,t}|\theta_{i,t})\pi(\theta_{i,t})}{k(\mathbf{z}_{l,t}|\mathbf{y}_n, \theta_{l,t-1})}$;
 - (d) calculer $d_{i,t} = \frac{1}{M} \sum_{l=1}^M \frac{k(\mathbf{z}_{l,t}|\mathbf{y}_n, \theta_{i,t-1})}{k(\mathbf{z}_{l,t}|\mathbf{y}_n, \theta_{l,t-1})} \sum_{k=1}^D \alpha_k^t q_{K_{i,t}}(\theta_{i,t-1}, \theta_{i,t})$;
 - (e) calculer $\omega_{i,t} = n_{i,t}/d_{i,t}$.
 2. rééchantillonner les particules courantes selon les poids normalisés $\bar{\omega}_{i,t}$;
 3. calculer les nouveaux poids du mélange $\alpha_k^{t+1} = \sum_{i=1}^M \bar{\omega}_{i,t} \mathbb{1}_k(K_{i,t})$.

5.4.2 Application formelle au modèle $\mathcal{B}(\eta_0, \eta_1, \beta)$

Nous formalisons ici l'application de l'algorithme PMC dans sa version D-kernel rao-blackwellisé au modèle à risques concurrents $\mathcal{B}(\eta_0, \eta_1, \beta)$. En découle évidemment l'application de PMC aux modèles exponentiel et de Weibull sous-jacents, dont les densités sont notées f_E et f_W . Pour faciliter l'écriture, nous modifions la paramétrisation du modèle. On note

$$\lambda = \eta_0^{-1}, \quad \mu = \eta_1^{-\beta}.$$

Notons $\mathbf{y}_n = (\mathbf{x}_r, \mathbf{c}_{n-r})$ les données observées de taille n , avec $\mathbf{x}_r = (x_1, \dots, x_r)$ des données non censurées de taille r et $\mathbf{c}_{n-r} = (c_1, \dots, c_{n-r})$ des données fixées de censure à droite. Les données manquantes \mathbf{z} du modèle sont les origines des données observées. Elles sont alors de taille $r + 2(n - r)$ et

$$k(\mathbf{z}|\mathbf{y}_n, \lambda, \mu, \beta) = k(\mathbf{z}^{\text{ew}}|\mathbf{y}_n, \lambda, \mu, \beta) k_E(\mathbf{z}^{\text{e}}|\mathbf{y}_n, \lambda) k_W(\mathbf{z}^{\text{w}}|\mathbf{y}_n, \mu, \beta)$$

$$\text{avec } \begin{cases} k_E(\mathbf{z}^{\text{e}}|\mathbf{y}_n, \lambda) &= \prod_{i=1}^{n-r} \frac{f_E(z_i|\lambda) \mathbb{1}_{\{z_i > c_i\}}}{\exp(-\lambda c_i)}, \\ k_W(\mathbf{z}^{\text{w}}|\mathbf{y}_n, \mu, \beta) &= \prod_{i=1}^{n-r} \frac{f_W(z_i|\mu, \beta) \mathbb{1}_{\{z_i > c_i\}}}{\exp(-\mu c_i^\beta)}, \\ k(\mathbf{z}^{\text{ew}}|\mathbf{y}_n, \lambda, \mu, \beta) &= \prod_{i=1}^r \left\{ \frac{f_E(z_i|\lambda) \mathbb{1}_{\{z_i > x_i\}}}{\exp(-\lambda x_i)} \right\}^{1-\delta(x_i)} \left\{ \frac{f_W(z_i|\mu, \beta) \mathbb{1}_{\{z_i > x_i\}}}{\exp(-\mu x_i^\beta)} \right\}^{\delta(x_i)} \end{cases}$$

et $\delta(X) = \delta(X|\lambda, \mu, \beta)$ la probabilité que X soit d'origine exponentielle. On a

$$\delta(X|\lambda, \mu, \beta) = \frac{h_E(X|\lambda)}{h_E(X|\lambda) + h_W(X|\mu, \beta)}$$

où h_E et h_W représentent les taux de défaillance respectifs des modèles exponentiel et de Weibull.

Enfin, la vraisemblance des données complètes $g(\mathbf{y}_n, \mathbf{z}|\lambda, \mu, \beta)$ est le produit des densités exponentielles ou de Weibull correspondant à l'origine de chaque donnée observée ou reconstituée, de par l'indépendance des deux modèles. Il nous faut donc obtenir une partition courante $(\mathbf{e}_{i,t}, \mathbf{w}_{i,t})$ des données reconstituées en données d'origine exponentielle ou de Weibull. Chacune des partitions est de taille n .

Choix de mélange instrumental. Nous choisissons ici d'utiliser la modélisation *a priori* définie au chapitre 3. Soit

$$\begin{aligned}\lambda &\sim \mathcal{G}(a_0, b_0), \\ \mu|\beta &\sim \mathcal{G}(a_1, b_1(a, \beta)), \\ \beta &\sim \mathcal{B}_e(p, q) \text{ définie sur } [\beta_l, \beta_r].\end{aligned}$$

En notant $\mathbf{e}_{i,t} = (e_{i,t}[1], \dots, e_{i,t}[n])$, à l'itération (i, t) de l'algorithme, on a alors

$$\begin{aligned}\lambda|\mathbf{e}_{i,t} &\sim \mathcal{G}\left(a_0 + n, b_0 + \sum_{l=1}^n e_{i,t}[l]\right), \\ \mu|\beta, \mathbf{w}_{i,t} &\sim \mathcal{G}\left(a_1 + n, b_1(a, \beta) + \sum_{l=1}^n w_{i,t}^\beta[l]\right)\end{aligned}$$

et $\pi(\beta|\mathbf{w}_{i,t})$ est définie à une constante près. On propose donc de choisir pour densité instrumentale courante

$$\begin{cases} \rho_{i,t}(\theta|\beta) &= \pi(\lambda|\mathbf{e}_{i,t})\pi(\mu|\beta, \mathbf{w}_{i,t}) \\ \beta &\sim \sum_{k=1}^D \alpha_k^t \mathcal{B}_e(p_k, q_k). \end{cases}$$

Le choix d'un mélange de lois Bêta sur $[\beta_l, \beta_r]$ fait sens vu le choix de $\pi(\beta)$. D est choisi suffisamment grand pour permettre le choix de plusieurs espérances, et pour chaque espérance le choix de plusieurs variances décroissantes. Notons $I_D = \{1, \dots, D\}$.

1. Soit I_1, \dots, I_d une partition régulière de I_D telle que chaque I_k soit de taille fixe L . On doit avoir $d \geq 2$ (d'où un choix $D = 4$ au minimum).
2. Soit d valeurs $(\beta_1, \dots, \beta_d)$ de β régulièrement réparties sur $[\beta_l, \beta_r]$.
3. Soit L variances $\sigma_{\beta,1}^2, \dots, \sigma_{\beta,L}^2$ classées par valeurs décroissantes. D'après notre expérience, pour $D \geq 10$ et $L \geq 3$, nous proposons une règle du type

$$\begin{aligned}\sigma_{\beta,1}^2 &= \text{Var}_\pi[\beta], \\ \sigma_{\beta,i+1} &= \sigma_{\beta,i}(1 - 1/L) \text{ pour } i = 1, \dots, L.\end{aligned}$$

4. $\forall k \in \{1, \dots, p\}$, on choisit L lois $\mathcal{B}_e(p_{k,i}, q_{k,i})$ avec $i = 1, \dots, L$ de même espérance β_k et de variances $\sigma_{\beta,1}^2, \dots, \sigma_{\beta,L}^2$.

5.4.3 Applications numériques

Nous utilisons pour ces applications numériques le modèle de Weibull. Soit $\theta = (\eta, \beta)$. On considère les données \mathbf{y}_n de taille 10, censurée à 60%, provenant du jeu $B215$ du tableau 2.3 (chapitre 2 p. 46).

défaillances réelles	censures
4380, 1791, 1611, 1291	6132, 5694, 5296, 4818, 4818, 4380

TAB. 5.3 – Données $B215$.

Le MLE (calculé par une méthode de Newton-Raphson) est $\hat{\theta} = (8752, 1.21)$, d'écart-types estimés $(2500, 0.5)$. La vraie valeur du paramètre est $\theta_0 = (8760, 2.15)$.

Nous proposons d'utiliser la modélisation *a priori* définie au chapitre 3 et reprise dans la sous-partie précédente, pour la paramétrisation $\theta = (\mu, \beta)$. On cherche à vérifier le bon comportement de $\rho(\theta)$ lorsque l'*a priori* est en accord avec les vraies valeurs du paramètre mais reste relativement peu informatif. On choisit donc $\pi(\beta)$ comme la loi uniforme sur $[0.5, 3.8]$, donc centrée en 2.15. On considère que la médiane marginale sur X correspond à la médiane simulée

$$x_e = \eta_0(\log 2)^{1/\beta_0} \simeq 7380$$

et l'on accorde *a priori* $a = 3$ données équivalentes à cet avis d'expert, ce qui représente légèrement moins de la moitié de l'information apportée par les données censurées (équivalente environ à l'apport de 6.5 données selon l'indicateur \tilde{n} défini au chapitre 3, § 3.5.2). Notons que la médiane *a priori* sur η est alors estimée à 8760.

5.4.3.1 Utilisation de PMC

Dans un premier temps, nous utilisons $D = 12$ densités $\mathcal{B}_e(p_k, q_k)$ sur $[0.5, 3.8]$, construites de la même façon qu'au § 5.4.2. Cependant, on modifie à chaque itération l'espérance de ces lois, la rendant peu éloignée de l'estimation PMC de β courante. Au bout de quelques itérations (pour $M = 2.10^5$ particules), l'évolution des estimations PMC de θ est largement stabilisée (cf. figure 5.2). L'évolution des poids α_k^t est fournie sur la figure 5.1, mais reste peu lisible, puisqu'ils s'écartent peu de $1/D^3$. Au bout d'un certain nombre d'itérations, l'utilisation de PMC rajoute plus de bruit qu'elle n'apporte d'information pour stabiliser les poids. Numériquement, ces poids privilégient un mélange de même espérance $\tilde{\beta} = 1.904$ et d'écart-types décroissants $(1.32, 0.99, 0.33)$ (le poids le plus fort désignant 0.33). L'estimation finale de θ est $\tilde{\theta} = (8505, 1.93)$ avec les écart-types $(1340, 0.29)$. Ces résultats proches des vraies valeurs illustrent le bon comportement mélangeant de PMC.

³Ce comportement est probablement dû à la modification des paramètres des composantes du mélange (les résultats de Douc *et al.* (2004) ne s'appliquent pas ici)

Dans un second temps, nous ne modifions pas les densités du mélange à chaque itération. Parmi les différentes espérances initiales (1.16, 1.82, 2.48, 3.14), seules 1.16 et 1.82 sont privilégiées par les poids. La densité du mélange de probabilité finale la plus forte est d'espérance 1.82 et d'écart-type 0.16. Nous obtenons une estimation finale $\tilde{\theta} = (8488, 1.92)$ d'écart-types (1336, 0.32).

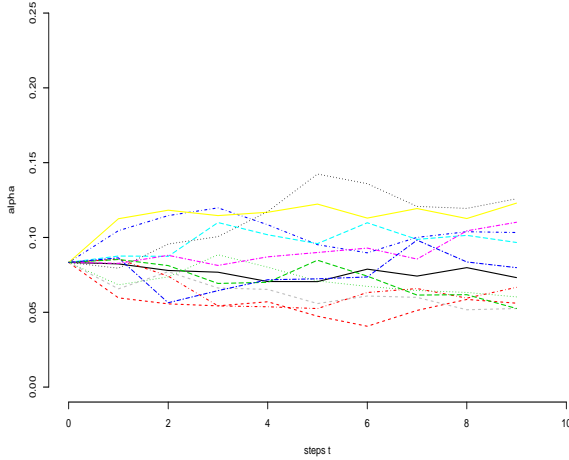


FIG. 5.1 – Évolution des α_k^t , $k = 1, \dots, 12$.

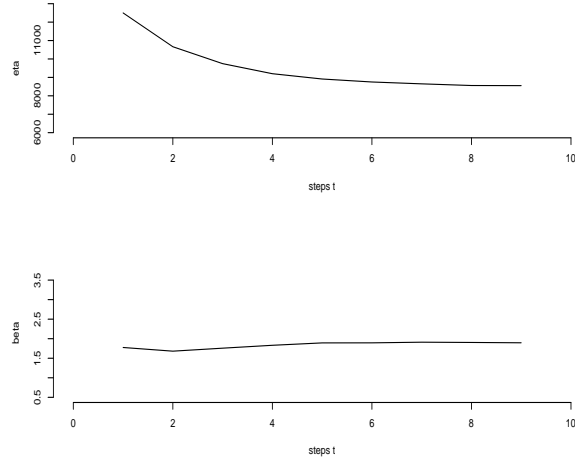


FIG. 5.2 – Évolution des estimations PMC de η (en haut) et β .

Troisièmement, on juge que $D = 12$ est un nombre trop élevé de densités à manier. On propose alors un mélange de 4 densités $\mathcal{B}_{e,i}$, $i = 1, \dots, 4$, telles que leurs espérances soient respectivement $e_1 = e_2 = 1.49$ et $e_3 = e_4 = 2.48$ (soit les valeurs au 1/3 et au 2/3 de $[\beta_l, \beta_r]$). On fixe tout d'abord les écarts-types $\sigma_1 = \sigma_3 = 0.33$ et $\sigma_2 = \sigma_4 = 0.165$. On trouve alors

$$\beta \sim 0.065 \mathcal{B}_{e,3} + 0.935 \mathcal{B}_{e,4}$$

ce qui revient à dire que l'*a posteriori* est au mieux approximé par $\mathcal{B}_{e,4}$. Clairement, l'espérance est mauvaise au vu de β_0 mais surtout l'écart-type paraît trop faible pour être réaliste. Le choix de mélange initial est donc mal adapté, ce qu'indique la différence de valeur importante entre les α_k^T (cf. figure 5.3). On augmente alors les écarts-types, en choisissant $\sigma_1 = \sigma_3 = 0.45$ et $\sigma_2 = \sigma_4 = 0.25$. On trouve alors (cf. figure 5.4)

$$\beta \sim 6.10^{-3} \mathcal{B}_{e_1} + 0.03 \mathcal{B}_{e_2} + 0.48 \mathcal{B}_{e_3} + 0.48 \mathcal{B}_{e_4}$$

ce qui entraîne une espérance $E[\beta] = 2.4$ et un écart-type de 0.34. Ce mélange n'est visiblement pas optimal, puisque les poids restent peu équilibrés. Cependant, cette densité instrumentale paraît beaucoup plus raisonnable que la précédente.

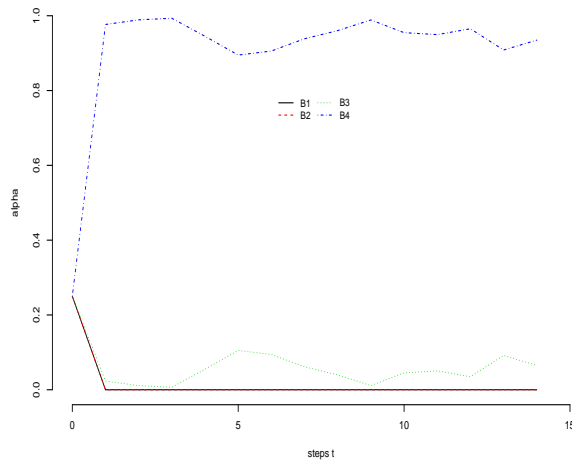


FIG. 5.3 – Évolution des α_k^t , $k = 1, \dots, 4$ pour un mauvais choix initial.

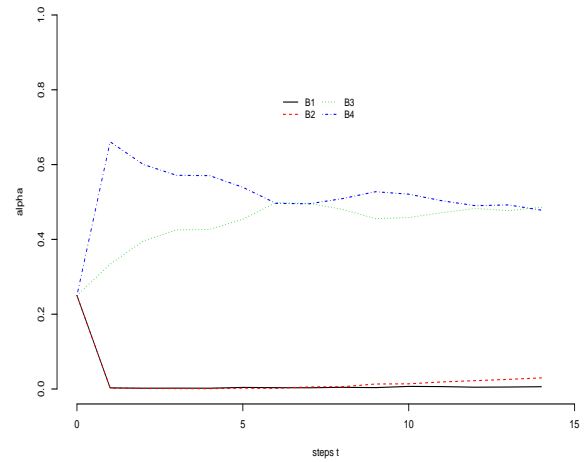


FIG. 5.4 – Évolution des α_k^t , $k = 1, \dots, 4$ pour ce même choix, après modification des écarts-types.

Quatrièmement, sachant $\beta_0 = 2.15$, nous cherchons le meilleur mélange de densités Bêta d'espérance β_0 et de variances décroissantes qui approxime $\pi(\beta|\mathbf{y}_n)$. Nous formons donc un mélange initial de densités $\mathcal{B}_{e,i}$ d'écarts-types respectifs $(0.66, 0.33, 0.165, 0.1)$, pour $i = 1, \dots, 4$. Après une dizaine d'itérations, les poids et les estimations PMC se sont stabilisés (cf. figure 5.5) et l'on obtient le mélange

$$\beta \sim 0.18 \mathcal{B}_{e_1} + 0.26 \mathcal{B}_{e_2} + 0.37 \mathcal{B}_{e_3} + 0.19 \mathcal{B}_{e_4}$$

ce qui donne un écart-type final de 0.284.

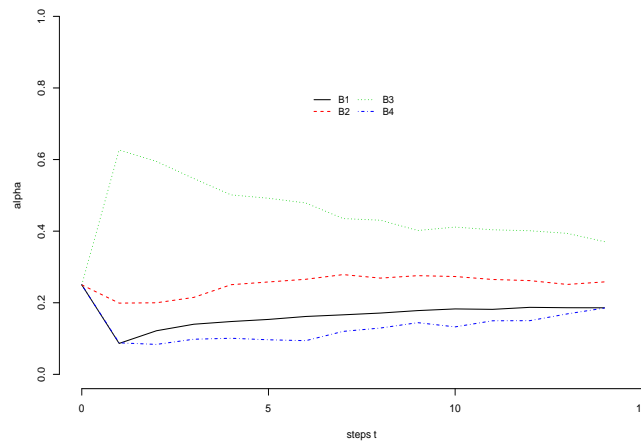


FIG. 5.5 – Évolution des α_k^t , $k = 1, \dots, 4$.

5.4.3.2 Utilisation de MCMC

Enfin, afin de comparer l'algorithme PMC avec les MCMC classiques (en sus des études déjà effectués dans Douc *et al.* (2004), Celeux *et al.* (2005) et Guillin *et al.* (2005) sur d'autres modèles), nous implémentons sur WinBUGS 1.4 le modèle bayésien selon le code suivant (voir Goldstein 2005 pour une revue des principales étapes d'utilisation du logiciel). L'initialisation des chaînes MCMC proposée ici correspond à $\beta = 2.15$ et $\eta = 8000$ et reste donc proche des vraies valeurs. 10^5 itérations sont utilisées et les résultats d'estimation de la table 5.4 sont moyennées sur 10 chaînes indépendantes. Par ailleurs, nous récapitulons dans le tableau 5.5 quelques valeurs de temps de calcul des algorithmes PMC (§ 5.4.3.1 et MCMC).

model

```

data

list(Y = c(4380,1791,1611,1291,6132,5694,5296,4818,4818,4380), D =
c(0,0,0,0,6132,5694,5296,4818,4818,4380), N = 10, N1 = 4, a =3 ,
xe=7380, p=1, q=1, B1=0.5,B2=3.8)

inits

list(beta0 = 0.5,mu=1.5625E-9)

{
  for( i in 1 : N1 ) {
    Y[i] ~ dweib(beta,lambda)
  }
  for( i in (N1+1) : N ) {
    Y[i] ~ dweib(beta,lambda)I(, D[i])
  }

  beta0 ~ dbeta(p,q) # a priori des effets aléatoires
  beta <- beta0*(B2-B1) + B1
  b <- pow(xe,beta)/( pow(2,(1/a)) - 1.0)
  mu ~ dgamma(a,b)
  lambda <- pow(mu,(1/beta))
  eta <- 1.0/lambda
}

```

La surestimation de η (ainsi que la valeur disproportionnée de l'écart-type sur ce paramètre) est corrélée à une sous-estimation importante de β . La convergence des chaînes MCMC sur η , sur cet exemple, n'est visiblement pas obtenue après 100 000 itérations. La figure 5.6 témoigne, sur les 5000 dernières particules, de l'instabilité d'une chaîne sur η , bien que l'estimation de β atteigne rapidement une certaine stabilité, après quelques centaines d'itérations. Quelques essais avec 10^6 itérations ne modifient pas substantiellement les résultats.

Vis-à-vis de PMC, les méthodes MCMC sont bien moins efficaces sur cet exemple d'échantillon. Plusieurs essais sur d'autres données présentant des caractéristiques proches (taille, taux de censure)

T_0		moyenne	médiane	écart-type	intervalle 5%-95%
50 000	η	15070	12420	9998	(4735, 40500)
	β	1.140	1.139	0.0722	(1.007, 1.287)
30 000	η	14930	12380	10460	(4673, 40300)
	β	1.139	1.138	0.0729	(1.005, 1.286)
10 000	η	15320	12410	10340	(4670, 39950)
	β	1.139	1.138	0.0727	(1.021, 1.285)
5 000	η	16500	12360	10220	(4640, 39750)
	β	1.143	1.140	0.0720	(1.009, 1.296)

TAB. 5.4 – Estimations MCMC moyennées sur les T_0 dernières itérations de 10 chaînes MCMC de $M = 10^5$ particules produites par WinBUGS.

Nombre de particules M	Tps PMC	Tps MCMC
5.10^4	32 s	25 s
10^5	46 s	50 s
2.10^5	1 min 12 s	1 min 40 s
10^6	3 min 37 s	8 min 20 s

TAB. 5.5 – Temps de calcul des algorithmes PMC et MCMC (moyennés sur 10 exécutions sur Pentium III. On considère ici 10 itérations d'adaptation pour PMC.

ont là encore montré la plus grande fiabilité de PMC. Par ailleurs, malgré le coût $\mathcal{O}(M^2)$ des étapes de rao-blackwellisation dans PMC, cet algorithme se révèle très concurrentiel en temps de calcul effectif vis-à-vis de l'exécution d'une MCMC dans un logiciel public.

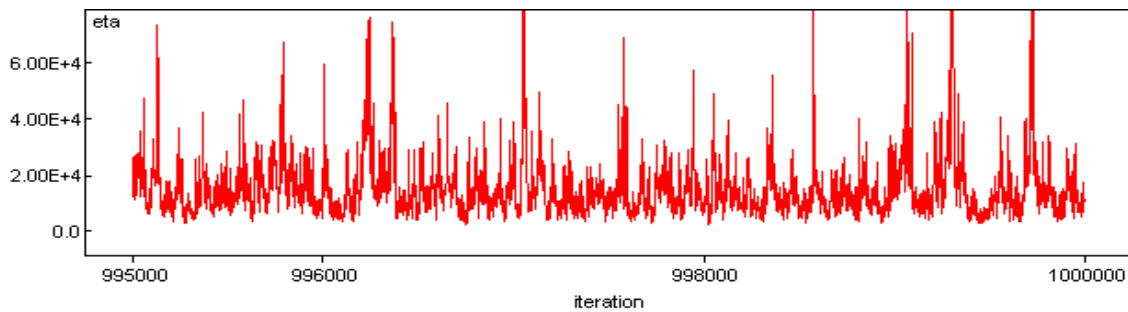


FIG. 5.6 – 5000 dernières itérations d'une chaîne MCMC sur η .

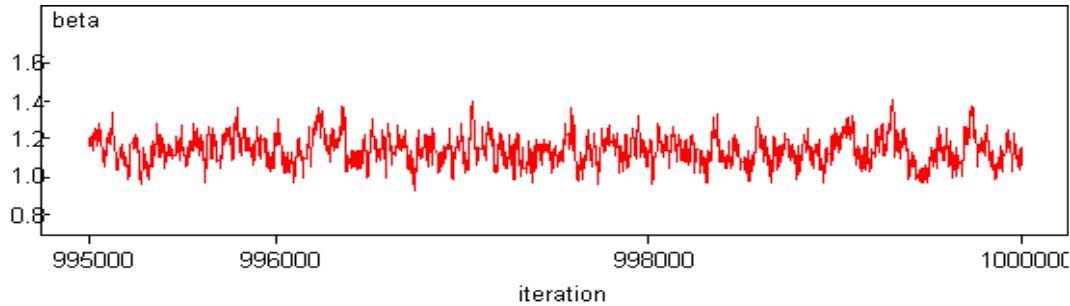


FIG. 5.7 – 5000 dernières itérations d’une chaîne MCMC sur β .

5.5 Conclusions

L’objectif d’un fiabiliste ou d’un durabiliste est à préciser clairement. S’il s’agit d’obtenir simplement des estimations ponctuelles du paramètre θ , dans le cadre restreint des *a priori* modélisés au chapitre 3, et non des simulations *a posteriori* de θ , l’échantillonnage préférentiel statique est parfaitement adapté : une proposition de densité instrumentale $\rho(\theta)$ est faite au § 5.3.3, respectant toutes les conditions requises pour une bonne estimation. Elle permet en outre de montrer le comportement parfois problématique de la densité instrumentale construite par Bacha (1996) *via* la méthode BRM. Dans un cadre plus général, nous conseillons d’utiliser le critère DAC construit et étudié au chapitre 4 pour fournir une indication du bien-fondé du choix de la densité instrumentale ρ .

Si dans un cadre plus général, on cherche à obtenir une densité maniable et réutilisable, à partir de laquelle on peut simuler sans effort, les méthodes d’échantillonnage préférentiel adaptatif, et plus particulièrement la version D-kernel rao-blackwellisée de l’algorithme PMC, semblent appropriées. Notons que les défauts de cette version de PMC - le choix d’un mélange de noyaux et surtout le choix de leurs variances - astreignent cependant son usage à des modèles de faible dimension, et, en l’état actuel de l’art, obligent à calibrer l’initialisation de l’algorithme “à la main”.

Lorsque l’échantillon \mathbf{y}_n est fortement censuré, nous avons connu certaines difficultés à sélectionner les variances des densités du mélange. Remarquons cependant que l’usage de cet algorithme est très général : il n’est pas restreint aux modèles à données manquantes et il permet d’approximer la distribution *a posteriori* lorsque l’*a priori* n’est pas informatif. On fait dans ce cas un choix initial d’un mélange de densités de variances relativement large, en utilisant par exemple comme référence l’estimation de la variance d’un estimateur fréquentiste.

Conclusions et perspectives

6.1 Principaux résultats, difficultés et motivations

La modélisation bayésienne de la loi de Weibull et sa calibration ont constitué le cœur de ce travail de thèse, axé sur certains problèmes rencontrés par les ingénieurs durabilistes et fiabilistes. L'article *An alternative competing risk model to the Weibull distribution for modelling aging in lifetime data analysis*, présenté en Annexe A, étudie une modélisation permettant d'améliorer la représentation classique par la loi de Weibull de la durée de vie d'un composant industriel. De façon plus générale, le chapitre 4 aborde un problème fréquemment rencontré par les analystes bayésiens quel que soit le contexte de l'étude ; il importe seulement d'utiliser une connaissance essentiellement subjective pour modéliser un *a priori* π . Outre ceux présentés dans l'article Bertholon *et al.* (2006), nos principaux résultats sont les suivants.

1. Nous sommes capables de prendre en compte une connaissance fractionnaire et subjective de la durée de vie d'un composant pour construire une modélisation *a priori* facilement calibrable par un analyste industriel au travers d'un paramètre homogène à une taille d'échantillon virtuel ; il est en effet accessible à la compréhension de l'expert et permet une discussion aboutissant à un consensus. Il permet aussi à un analyste industriel d'exercer sa critique de l'expertise.
2. Nous avons défini et utilisé le critère DAC qui permet de déterminer de façon objective si les informations provenant de l'expertise, de données observées, sont abusivement éloignées et menacent le bien-fondé du résultat d'inférence. Il apparaît comme une alternative à une procédure très récente, aux propriétés et à la facilité d'emploi moindres. Le cadre d'utilisation de cet outil est typiquement celui d'une étude appliquée, et son emploi est pertinent dans les problèmes de durabilité et de fiabilité industrielles.

Le critère nécessite cependant une adaptation dans de nombreux cas d'application ; celle que nous proposons peut parfois se révéler médiocre si les échantillons de données disponibles se révèlent de taille très faible, de censure forte ou si le modèle paramétrique choisi est de dimension élevée. Deux objectifs conjoints de recherche seraient donc de tenter d'améliorer cette adaptation comme de proposer une méthodologie permettant de juger dans quelle mesure la comparaison

entre une connaissance *a priori* forte et un faible nombre de données industrielle fait sens.

En sus de ces résultats, nous avons consacré un chapitre aux méthodes d'estimation bayésienne, pour lesquelles nous donnons également quelques conseils méthodologiques.

D'autres problèmes rencontrés dans l'industrie peuvent cependant s'opposer à une application directe des outils et méthodes présentés dans cette thèse. Ainsi, élargir notre méthodologie à des données industrielles plus générales, par exemple censurées à gauche comme à droite, nous paraît être un enjeu important et le sujet de futures recherches. Par ailleurs, certains objectifs de recherche nous sont apparus comme des compléments ou des généralisations de notre travail. Brièvement, nous en présentons trois dans la suite de ce chapitre.

L'un des premiers objectifs est de définir et d'appliquer une méthodologie similaire à un nouveau modèle de défaillance tenant compte des possibles défaillances d'un composant \sum à la sollicitation, à la suite des travaux préliminaires de Celeux & Rodionov (2002). Ce type de modèle intéresse au moins autant les industriels que les modèles de Weibull sur lesquels nous nous sommes focalisés dans ce travail de thèse. Nous donnons au § 6.2 une définition formelle de ce modèle ainsi qu'une liste des principaux objectifs de recherche.

Par ailleurs, au cours de divers entretiens avec les industriels qui ont motivé ce travail de thèse et approuvé la méthodologie proposée, il nous est souvent apparu que si les connaissances expertes sont nombreuses, elles ont peine à être exprimées en termes compréhensibles à la fois pour les ingénieurs et les statisticiens. C'est pourquoi, définir une modélisation *a priori* aux hyperparamètres facilement calibrables *via* des indicateurs comme une taille de données virtuelles nous est apparu comme un objectif intéressant. Au § 6.3, nous faisons quelques propositions de recherche en ce sens ; une généralisation des *pseudoposterior priors* introduits au chapitre 4 (§ 4.7) nous semble en particulier être un outil intéressant.

Enfin, dans la continuité du chapitre 5, nous faisons au § 6.4 une proposition de construction de lois instrumentales visant à améliorer l'échantillonnage préférentiel non adaptatif (statique). Nous avons mis en évidence les défauts de l'algorithme BRM. En conservant l'idée d'utiliser à la fois la connaissance des données et la connaissance *a priori* lorsque les données sont incomplètes, nous proposons de définir une loi instrumentale par une moyenne de *posterior priors* sur des données non censurées et reconstituées. Nous donnons ainsi un premier exemple qui montre le bon comportement du tirage d'importance.

6.2 Un modèle de défaillance à la sollicitation

Ce modèle a originellement été proposé par Celeux & Rodionov (2002) dans une communication à l'ESReDA mais n'a pas fait l'objet de travaux ultérieurs. Il vise à prendre en compte la situation suivante : un composant \sum peut être soumis à deux types de défaillances qui entrent en compétition ; l'une est accidentelle et l'autre est due à son vieillissement, suscité par des chocs à la sollicitation. Il s'agit donc d'un modèle construit sur le même principe que $\mathcal{B}(\eta_0, \eta_1, \beta)$.

6.2.1 Définition formelle

Formellement, on considère que le composant Σ peut subir une défaillance au bout de N chocs, et l'on suppose qu'un choc est la réalisation d'un processus de Poisson d'intensité ω . Ce paramètre est ainsi le taux d'apparition d'un choc. Si l'on nomme S le nombre de sollicitations du matériel sur la période $[0, T]$, on a

$$\omega = S/T.$$

Dans une première approche, on peut considérer que S et T (et donc ω) sont connus. Si l'on suppose également que les chocs sont indépendants entre eux et que l'apparition d'un choc ne dépend pas de l'âge du matériel, l'occurrence de N chocs est une somme de N lois exponentielles indépendantes et suit donc une loi Gamma $\mathcal{G}(N, \omega)$ de densité $f_C(t|N, \omega)$. Le taux de défaillance de ce processus de vieillissement s'écrit

$$h_C(t|N, \omega) = \frac{(\omega t)^{N-1} \exp(-\omega t)}{\int_t^\infty (\omega u)^{N-1} \exp(-\omega u) du}.$$

Par ailleurs, le composant Σ est potentiellement soumis à des défaillances accidentelles avec un taux $h_A(t|\lambda) = \lambda$. On le supposera indépendant de l'âge du composant et des N chocs qu'il peut subir. De la même façon que pour le modèle $\mathcal{B}(\eta_0, \eta_1, \beta)$, on modélise ce comportement par une loi exponentielle de paramètre λ , de densité notée $f_A(t|\lambda)$.

Ainsi, si l'on nomme X le temps de défaillance du composant Σ , on aura

$$X = \min(A, C)$$

où

$$\begin{aligned} A &\sim \mathcal{E}(\lambda), \\ C &\sim \mathcal{G}(N, \omega) \end{aligned}$$

et A et C sont indépendantes. La fiabilité globale du composant s'écrit

$$P(X > t) = S_A(t|\lambda)S_C(t|N, \omega) = \exp(-\lambda t) \int_t^\infty \frac{\omega^N}{\Gamma(N)} u^{N-1} \exp(-\omega u) du$$

où S_A et S_C sont les fonctions de survie des modèles en compétition.

6.2.2 Nature et vraisemblance des données

Les données industrielles disponibles sont le plus souvent censurées à droite (on les supposera fixées indépendamment de l'expérience). De même que pour le modèle $\mathcal{B}(\eta_0, \eta_1, \beta)$, il existe un autre type de données manquantes : on peut ne pas connaître l'origine d'une défaillance effectivement relevée. En l'affectant préférentiellement à l'un des modèles en compétition, on censure une donnée de durée de vie provenant de l'autre modèle.

Ces données se présentent donc sous la forme $\mathbf{D} = \{(y_1, \mathbf{e}_1), \dots, (y_n, \mathbf{e}_n)\}$ où y_i est un temps de défaillance ou un temps de censure à droite observé, et chaque vecteur $\mathbf{e}_i \in \{0, 1\}^2$ désigne l'origine de la défaillance. La première composante de ce vecteur désigne une cause accidentelle et la deuxième

un vieillissement. Ainsi, le vecteur $\mathbf{e}_i = (0, 1)$ indique une défaillance par vieillissement certifiée. Si l'origine de y_i est inconnue mais que y_i n'est pas une censure, on note $\mathbf{e}_i = (1, 1)$. Enfin, si y_i est une censure (à droite), on note $\mathbf{e}_i = (0, 0)$. Listons alors les densités des différentes données rencontrées :

1. pour les n_1 observations telles que $\mathbf{e}_i = (1, 0)$, $f(y_i, \mathbf{e}_i) = f_A(y_i|\lambda)S_C(y_i|N, \omega)$,
2. pour les n_2 observations telles que $\mathbf{e}_i = (0, 1)$, $f(y_i, \mathbf{e}_i) = f_C(y_i|N, \omega)S_A(y_i|\lambda)$,
3. pour les n_{12} observations telles que $\mathbf{e}_i = (1, 1)$, $f(y_i, \mathbf{e}_i) = f_A(y_i|\lambda)S_C(y_i|N, \omega) + f_C(y_i|N, \omega)S_A(y_i|\lambda)$,
4. pour les n_0 observations telles que $\mathbf{e}_i = (0, 0)$, $f(y_i, \mathbf{e}_i) = S_A(y_i|\lambda)S_C(y_i|N, \omega)$.

Alors la vraisemblance des données observées s'écrit

$$\mathcal{L}(\mathbf{D}; \lambda, N, \omega) = \lambda^{n_1} \prod_{i/\mathbf{e}_i=(0,1)} h_C(y_i|N, \omega) \prod_{j/\mathbf{e}_j=(1,1)} (\lambda + h_C(y_j|N, \omega)) \prod_{k=1}^n S_A(y_k|\lambda)S_C(y_k|N, \omega).$$

6.2.3 Objectifs de recherche

Dans un cadre d'étude fréquentiste, Celeux & Rodionov (2002) ont proposé et testé les algorithmes EM et SEM pour estimer les paramètres (λ, N) . Les équations du maximum de vraisemblance sont assez pénibles à résoudre et une résolution par moments a été proposée pour estimer plus simplement les paramètres. Si numériquement les estimations semblent donner de bons résultats, aucun résultat théorique n'est encore disponible, à notre connaissance, sur l'unicité du maximum de vraisemblance, la consistance et la proximité de cet estimateur. Voici un premier objectif de recherche.

A priori, les problèmes d'estimation sont fondamentalement les mêmes que ceux que nous avons rencontrés pour estimer les paramètres du modèle $\mathcal{B}(\eta_0, \eta_1, \beta)$. La faible taille et la censure des données industrielles ajoutées à la nature de ce modèle à risques masqués font que dans la pratique, l'information manquante est prépondérante sur l'information disponible. Il est raisonnable de considérer plutôt une estimation bayésienne des paramètres.

L'inférence bayésienne est ici considérablement simplifiée vis-à-vis des lois de Weibull auxquelles nous nous sommes intéressées au chapitre 3. En effet, les modèles exponentiel et Gamma sont conjugués. Si l'on peut, au sein d'une procédure d'estimation similaire à celle développée par Bertholon *et al.* (2006), reconstituer l'origine de chaque donnée, nous pouvons manier des lois *a posteriori* conditionnelles explicites.

Un second objectif de recherche est donc d'adapter la démarche entreprise au chapitre 3 pour délimiter et modéliser la connaissance *a priori* disponible, d'une façon à la fois pratique et compréhensible ; nécessité sera de fournir de nouveaux indicateurs compréhensibles par l'expert et l'analyste bayésien.

Enfin, Celeux & Rodionov (2002) ont proposé une amélioration possible du modèle, qui consiste à prendre en compte les *remplacement préventifs*. Ce choix revient à trier certaines données de censure à droite, en localisant un temps réel de défaillance, masqué, entre deux valeurs de censure successives. La vraisemblance des données est alors modifiée, ce qui modifie évidemment les algorithmes d'estimation. Un travail complet d'estimation reste à faire sur ce nouveau modèle, apprécié par les industriels.

6.3 Une méthodologie de calibration

Une autre piste de recherche, entrevue pendant la thèse, concerne l'usage des *posterior priors* et *pseudoposterior priors* introduits au chapitre 4. Ils nous paraissent des constructions formelles susceptibles d'être généralisées et utilisées pour la calibration d'un a priori arbitraire $\pi(\theta|\omega)$ où ω est le vecteur des hyperparamètres que l'on cherche à fixer.

6.3.1 Une modélisation d'un *a priori* subjectif

Nous faisons tout d'abord l'hypothèse que le modèle paramétrique $\mathcal{M}(\theta)$ est justifié vis-à-vis des données. Des méthodes permettant de valider cette hypothèse sont notamment proposées par Bayarri & Berger (2000).

Supposons maintenant que la connaissance subjective disponible permette au moins de définir une représentation *idéale* de la connaissance *a priori*, au sens où elle respecte au maximum les spécifications de l'expertise. Celle-ci est supposée s'exprimer sur la variable observable X . Sa densité s'écrit $f(x|\theta_e)$ où θ_e est une estimation de θ , agréée par l'expert. Oakley & O'Hagan (2006) préconisent des méthodes non paramétriques, comme des jeux d'histogrammes, pour obtenir de l'expert un accord indirect sur une estimation du paramètre.

Supposons alors que l'analyste bayésien puisse juger l'expert comme un fournisseur de données fictives $\tilde{\mathbf{x}}_a = \tilde{x}_1, \dots, \tilde{x}_a \in \chi^a$. Il fixe une taille $a \in \mathbb{N}^*$, et propose une première représentation *a priori* perçue comme un *a posteriori* moyenné sur les réalisations de $\tilde{\mathbf{x}}_a$, soit comme l'*expected posterior prior*

$$\pi_a(\theta|\theta_e) = \int_{\chi^a} \pi^J(\theta|\tilde{x}_1, \dots, \tilde{x}_a) \prod_{i=1}^a f(\tilde{x}_i|\theta_e) d\tilde{x}_1 \dots d\tilde{x}_a,$$

où π^J est un *a priori* non informatif choisi pour modéliser l'ignorance. Si la taille a est en-dessous de la taille q d'un *échantillon minimal d'entraînement* (MTS), qui est la quantité minimale de donnée $x(l)$ telle que $\pi^J(\theta|x(l))$ est propre, nous suggérons l'emploi d'un *expected pseudoposterior prior*

$$\pi_a(\theta|\theta_e) = \int_{\chi^q} \pi_a^J(\theta|\tilde{x}_1, \dots, \tilde{x}_q) \prod_{i=1}^q f(\tilde{x}_i|\theta_e) d\tilde{x}_1 \dots d\tilde{x}_q \quad (6.1)$$

où

$$\pi_a^J(\theta|\tilde{x}_1, \dots, \tilde{x}_q) \propto \left(\prod_{i=1}^q f(\tilde{x}_i|\theta) \right)^{a/q} \pi^J(\theta).$$

Le choix d'un $a \in \mathbb{R}^+$ (et non plus dans \mathbb{N}^*) est alors limité par la condition que $\pi_a^J(\theta|\tilde{x}_1, \dots, \tilde{x}_q)$ soit propre. Mais pour certains modèles, comme la famille exponentielle naturelle ou le modèle de Weibull, on peut choisir a aussi petit que l'on veut (en particulier inférieur à 1 si l'on désire un *a priori* très plat). Voir le § 4.7 pour plus de précision. Un tel résultat, sur un modèle réputé délicat à utiliser comme Weibull, laisse espérer que cette méthode permette de modéliser de larges gammes d'*a priori* "idéaux". Ceux-ci sont parfois entièrement explicites, comme dans l'exemple suivant.

EXEMPLE 18. *Modèle exponentiel.*

Supposons obtenir d'un expert une estimation x_e de la durée de vie médiane d'un composant de taux

de défaillance λ . Alors l'estimation *a priori* de λ est $\lambda_e = x_e^{-1} \log 2$. Fixons $a < 1$. Alors

$$\pi_a(\lambda|\lambda_e) = a^{a+1} \lambda_e \frac{\lambda^{a-1}}{(\lambda_e + a\lambda)^{a+1}}. \blacksquare$$

6.3.2 Objectifs de recherche

Nous proposons plusieurs objectifs de recherche liés à l'utilisation des *expected pseudoposterior priors* en analyse bayésienne subjective. Il nous semble en effet que cette approche peut se révéler plus compréhensible pour un analyste industriel. Cependant, le plus souvent, $\pi_a(\theta|\theta_e)$ est une densité définie *via* une intégrale et à un facteur multiplicatif près, et donc peu pratique d'utilisation. On désire alors fixer le vecteur ω des hyperparamètres de $\pi(\theta) = \pi(\theta|\omega)$, une densité *a priori* de forme arbitraire et de maniement plus simple. Il est alors naturel de sélectionner

$$\omega^* = \arg \min_{\omega \in \Omega} \mathcal{D}(\omega) \quad (6.2)$$

où

$$\mathcal{D}(\omega) = KL \{ \pi_a(\cdot|\theta_e) \parallel \pi(\cdot|\omega) \}.$$

Cette entropie relative est convexe en son argument $\pi(\cdot|\omega)$. Ainsi, le minimum existe et est unique sous les conditions que π est identifiable, soit que l'application $\Omega : \omega \mapsto \pi(\cdot|\omega)$ est injective, et que $\mathcal{D}(\omega)$ soit deux fois différentiable en ω . Notons que si $\pi_a(\cdot|\theta_e)$ est définie à un facteur multiplicatif près, celui-ci n'a aucun rôle dans le calcul de ω^* . Il est seulement nécessaire de pouvoir simuler des tirages $(\theta_1, \dots, \theta_M)$ de densité $\pi_a(\theta|\theta_e)$ afin d'estimer $\mathcal{D}(\omega)$ et ses dérivées en une valeur ω , par des méthodes de Monte-Carlo. Ces méthodes sont alors typiquement couplées à des méthodes de gradient, ce qui peut les alourdir considérablement.

La première difficulté réside donc dans le traitement de la minimisation (6.2). On peut par exemple s'épargner de lourds calculs en décidant de choisir π dans une classe de densités à la fois adaptées à ce type de résolution et cependant suffisamment flexibles pour être facilement simulables. Par exemple, si nos *a priori* sont bornés, Barron & Sheu (1991) proposent d'utiliser des suites de familles exponentielles pour approximer $\pi_a(\cdot|\theta_e)$ au sens de Kullback.

Ce choix n'est cependant pas toujours judicieux. Il nous semble intéressant de choisir un certain nombre de familles d'*a priori* traditionnellement utilisées et de fournir des algorithmes de minimisation adaptés. Ainsi, nous avons déjà obtenu quelques résultats sur des *a priori* à la structure hiérarchisée (l'un d'entre eux est fourni en Annexes § 6.5.1).

L'autre difficulté que nous percevons réside dans le *satisfecit* de l'approximation. Une fois ω^* estimé, une question légitime est de savoir si la densité $\pi(\theta|\omega^*)$ approxime bien $\pi_a(\omega|\theta_e)$. Une erreur *relative* peut être calculée, indiquant la pertinence du choix de π . Ainsi, l'erreur relative ρ en norme L_1 (en variation totale) peut être majorée ainsi (cf. Pollard 2006, chapitre 3)

$$\rho \leq \frac{\sqrt{2\mathcal{D}(\omega^*)}}{\|\pi_a(\cdot|\theta_e)\|_1}$$

où

$$\|\pi_a(\cdot|\theta_e)\|_1 = \int_{\Theta} |\pi_a(\theta|\theta_e)| d\theta.$$

Fournir un certain nombre de résultats dans la même veine, faisant intervenir des notions d'erreur compréhensibles par un analyste industriel, nous paraît être un objectif de recherche intéressant lorsque l'on cherche à simplifier la modélisation.

Notons qu'un autre problème relié à la projection d'information par Kullback est de savoir si $\pi(\theta|\omega^*)$ ne transmet pas d'information supplémentaire *a priori* dans l'inférence à celle qui est désirée par l'analyste bayésien ; ceci peut se traduire par une réduction drastique de la variance de θ selon $\pi(\theta|\omega^*)$ vis-à-vis de la variance de θ selon $\pi(\theta|\theta_e)$. Il nous semble donc important d'obtenir une procédure de comparaison de ces deux *a priori* vis-à-vis de l'information apportée par les données.

6.4 Une proposition d'échantillonnage préférentiel non adaptatif

Nous avons fourni au § 5.3 une critique de l'algorithme BRM qui propose comme densité instrumentale $\rho(\theta)$ la densité des estimations du maximum de vraisemblance $\hat{\theta}_1, \dots, \hat{\theta}_M$ d'échantillons à données manquantes reconstituées. Sur l'exemple de la loi de Weibull de paramètre $\theta = (\mu, \beta)$, *via* une comparaison numérique avec une densité $\rho_R(\theta)$ de référence construite au § 5.3.3, les défauts de BRM ont été mis en évidence pour l'estimation bayésienne de fonctions de θ .

On construit cette densité $\rho_R(\mu, \beta)$ dans un certain cadre *a priori*, où $\pi(\beta)$ est bornée et $\pi(\mu|\beta, \mathbf{y}_n)$ est explicite. Dans un cadre plus général, et notamment sans supposer que l'espace des paramètres Θ est borné, nous désirons proposer un autre candidat $\rho^J(\theta)$.

Pour le moment, notre démarche reste purement empiriste et s'appuie sur le maniement des *posterior priors* introduits au chapitre 4 pour l'adaptation du critère DAC. Soit π^J un *a priori* non informatif et notons $x(1), \dots, x(L)$ des *échantillons minimaux d'entraînement* (MTS) choisis parmi des données \mathbf{y}_n . Nous avons remarqué que lorsque la taille des MTS est faible par rapport n , les sommes arithmétiques de *posterior priors*

$$\sum_{i=1}^L \pi^J(\theta|x(i)), \tag{6.3}$$

ont des queues lourdes et privilégient des valeurs de θ situées dans les valeurs hautes de la vraisemblance. De la même façon que BRM, nous proposons de remplacer les données manquantes dans \mathbf{y}_n par des données simulées marginalement selon la loi *a priori* afin de prendre en compte toute l'information disponible. On espère ainsi que la combinaison (6.3) privilégie les valeurs de θ ayant une fréquence élevée dans la distribution *a posteriori*. On construit donc une densité instrumentale candidate $\rho^J(\theta)$ de la façon suivante.

1. On simule $\theta_1, \dots, \theta_M \sim \pi(\theta)$.

2. Pour $j = 1, \dots, M$, on simule $\mathbf{z}_j \sim k(\mathbf{z}|\mathbf{y}_n, \theta_j)$ et on reconstitue M échantillons de taille n $\mathbf{x}_n^{(1)}, \dots, \mathbf{x}_n^{(M)}$
3. Soit $x^{(i)}(l)$ un MTS dans l'échantillon $\mathbf{x}_n^{(i)}$.
Soit L le nombre de MTS par échantillon.
4. On choisit alors comme densité instrumentale

$$\rho^J(\theta) = \frac{1}{ML} \sum_{i=1}^M \sum_{l=1}^L \pi^J(\theta|x^{(i)}(l)). \quad (6.4)$$

$$(6.5)$$

Dans l'exemple suivant, nous nous plaçons dans les mêmes conditions expérimentales qu'au § 5.3.4, en utilisant le modèle de Weibull. Nous comparons alors $\rho_R(\theta)$ et $\rho^J(\theta)$, que nous remplaçons en perspective avec BRM. Sur cet exemple, le choix de $\rho^J(\theta)$ est bien meilleur que BRM et donne des résultats très proches du choix de référence $\rho_R(\theta)$.

EXEMPLE 19. *Modèle de Weibull.*

Nous reprenons les échantillons de Weibull censurés, simulés au § 5.3.4. Dans les tableaux 6.1 et 6.2, pour chaque choix de densité instrumentale et selon plusieurs scénarii *a priori*, nous indiquons un triplet de valeurs correspondant aux estimations de Monte-Carlo de la durée de vie moyenne *a posteriori* (l'espérance prédictive sur X), notée D , de son écart-type σ_D et le pourcentage de survie prédictif à $t_0 = 130$ (soit la valeur maximale de la censure), noté S_{130} . Nous donnons en Annexes § 6.5.2 le descriptif de ρ^J (qui est entièrement explicite) et la méthode de simulation d'importance. ■

a	σ_β	% censure	BRM	ρ^J	référence ρ_R
3	0.36	25%	110 40 30%	110 40 31%	110 40 31%
		50%	128 40 48%	126 46 45%	126 46 45%
10	1	5%	101 33 21%	101 38 21%	100 37 21%
		25%	112 35 31%	112 42 31%	111 42 31%
		50%	122 38 42%	123 50 41%	123 49 51%

TAB. 6.1 – Triplets (D, σ_D, S_{130}) pour un *a priori* en accord avec les données ($x_e = 100$). BRM et le choix de densité d'importance ρ^J selon (6.4) fournissent de bonnes estimations vis-à-vis de l'estimation de référence (dernière colonne).

a	σ_β	% censure	BRM	ρ^J	référence ρ_R
3	1	25%	106 48 29%	107 48 30%	108 48 30%
		50%	124 34 44%	129 62 43%	128 61 43%
		75%	138 35 60%	135 115 62%	133 114 62%
	0.36	25%	112 37 32%	104 39 27%	105 40 27%
		50%	127 38 47%	122 47 42%	123 48 42%
		75%	150 47 64%	153 61 61%	154 61 61%
10	1	5%	104 39 25%	85 45 16%	86 45 16%
		25%	115 39 34%	96 56 23%	97 55 24%
		50%	131 50 48%	118 80 35%	119 81 36%

TAB. 6.2 – Triplets (D, σ_D, S_{130}) pour un *a priori* décalé avec les données ($x_e = 50$). Alors que BRM souffre de défauts déjà perçus au § 5.3.4, le choix ρ^J de densité d'importance (6.4) reste satisfaisant, vis-à-vis du choix de référence (dernière colonne).

6.5 Annexes

6.5.1 Minimisation hiérarchique

Quand π est construit de façon hiérarchique, la proposition 16 permet de séparer par niveau hiérarchique les étapes de la minimisation

$$\omega^* = \arg \min_{\omega \in \Omega} \mathcal{D}(\omega)$$

où

$$\mathcal{D}(\omega) = KL \{ \pi_a(\cdot | \theta_e) \parallel \pi(\cdot | \omega) \}.$$

PROPOSITION 16. *Soit $(\theta_1, \theta_2) \in \Theta_1 \times \Theta_2 \subset \Theta$ et $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 \subset \Omega$. Supposons pouvoir écrire $\pi(\theta | \omega) = \pi(\theta_1 | \theta_2, \omega_1) \pi(\theta_2 | \omega_2)$. Alors $\omega^* = (\omega_1^*, \omega_2^*)$ où*

$$\begin{cases} \omega_2^* = \arg \min_{\omega_2} \mathcal{D}_2(\omega_2), \\ \omega_1^* = \arg \min_{\omega_1} \mathcal{D}_1(\omega_1) \end{cases}$$

où

$$\begin{aligned} \mathcal{D}_1(\omega_1) &= KL \{ \pi_a(\theta | \theta_e) \parallel \pi(\theta_1 | \theta_2, \omega_1) \}, \\ \mathcal{D}_2(\omega_2) &= KL \{ \pi_a(\theta_2 | \theta_e) \parallel \pi(\theta_2 | \omega_2) \}, \end{aligned}$$

Preuve. Nous donnons la preuve dans le cas où $\pi_a(\theta | \theta_e)$ est définie par (6.1). Avec $\pi^J(\theta) = \pi^J(\theta_1 | \theta_2) \pi^J(\theta_2)$, et en notant $\tilde{\mathbf{x}}_{\mathbf{q}} = \tilde{x}_1, \dots, \tilde{x}_q$, on obtient immédiatement que

$$\pi_a^J(\theta | \tilde{x}_1, \dots, \tilde{x}_q) = \pi_a^J(\theta_1 | \tilde{\mathbf{x}}_{\mathbf{q}}) \pi_a^J(\theta_2 | \tilde{\mathbf{x}}_{\mathbf{q}})$$

Notons $H = - \int_{\Theta} \pi_a(\theta | \theta_e) \log \pi_a(\theta | \theta_e) d\theta$ l'entropie de $\pi_a(\cdot | \theta_e)$. Notons également $\nu_e(\tilde{\mathbf{x}}_{\mathbf{q}}) = \prod_{i=1}^q f(\tilde{x}_i | \theta_e)$.

Alors

$$\begin{aligned} \mathcal{D}(\omega) &= -H - \int_{\Theta_2} \int_{\Theta_1} \int_{\chi^q} \pi_a^J(\theta_1 | \theta_2, \tilde{\mathbf{x}}_{\mathbf{q}}) \pi_a^J(\theta_2 | \tilde{\mathbf{x}}_{\mathbf{q}}) \log \{ \pi(\theta_1 | \theta_2, \omega_1) \pi(\theta_2 | \omega_2) \} \nu_e(\tilde{\mathbf{x}}_{\mathbf{q}}) d\tilde{\mathbf{x}}_{\mathbf{q}} d\theta_1 d\theta_2, \\ &= -H - \int_{\Theta_2} \int_{\chi^q} \left\{ \int_{\Theta_1} \pi_a^J(\theta_1 | \theta_2, \tilde{\mathbf{x}}_{\mathbf{q}}) \log \pi(\theta_1 | \theta_2, \omega_1) d\theta_1 \right\} \pi_a^J(\theta_2 | \tilde{\mathbf{x}}_{\mathbf{q}}) \nu_e(\tilde{\mathbf{x}}_{\mathbf{q}}) d\tilde{\mathbf{x}}_{\mathbf{q}} d\theta_2 \\ &\quad - \int_{\Theta_2} \int_{\chi^q} \pi_a^J(\theta_2 | \tilde{\mathbf{x}}_{\mathbf{q}}) \log \pi(\theta_2 | \omega_2) \nu_e(\tilde{\mathbf{x}}_{\mathbf{q}}) d\tilde{\mathbf{x}}_{\mathbf{q}} d\theta_2, \\ &= \int_{\Theta_2} \int_{\chi^q} \mathcal{D}_1(\omega_1 | \tilde{\mathbf{x}}_{\mathbf{q}}, \theta_2) \nu_e(\tilde{\mathbf{x}}_{\mathbf{q}}) \pi_a^J(\theta_2 | \tilde{\mathbf{x}}_{\mathbf{q}}) d\tilde{\mathbf{x}}_{\mathbf{q}} d\theta_2 + \mathcal{D}_2(\omega_2) + \Delta \end{aligned}$$

$$\text{où } \mathcal{D}_1(\omega_1 | \tilde{\mathbf{x}}_{\mathbf{q}}, \theta_2) = KL \{ \pi_a^J(\theta_1 | \theta_2, \tilde{\mathbf{x}}_{\mathbf{q}}) \parallel \pi(\theta_1 | \theta_2) \}$$

et

$$\begin{aligned} \Delta &= -H - \int_{\Theta_1} \int_{\Theta_2} \int_{\chi^q} \pi_a^J(\theta_1 | \theta_2, \tilde{\mathbf{x}}_{\mathbf{q}}) \pi_a^J(\theta_2 | \tilde{\mathbf{x}}_{\mathbf{q}}) \log \{ \pi_a^J(\theta_1 | \theta_2, \tilde{\mathbf{x}}_{\mathbf{q}}) \} \nu_e(\tilde{\mathbf{x}}_{\mathbf{q}}) d\tilde{\mathbf{x}}_{\mathbf{q}} d\theta_1 d\theta_2 \\ &\quad - \int_{\Theta_2} \int_{\chi^q} \pi_a^J(\theta_2 | \tilde{\mathbf{x}}_{\mathbf{q}}) \log \left\{ \int_{\chi^q} \pi_a^J(\theta_2 | \tilde{\mathbf{x}}_{\mathbf{q}}) \nu_e(\tilde{\mathbf{x}}_{\mathbf{q}}) d\tilde{\mathbf{x}}_{\mathbf{q}} \right\} \nu_e(\tilde{\mathbf{x}}_{\mathbf{q}}) d\tilde{\mathbf{x}}_{\mathbf{q}} d\theta_2. \end{aligned}$$

Puisque Δ ne dépend pas de ω , minimiser $\mathcal{D}(\omega)$ revient à minimiser

$$\int_{\Theta_2} \int_{\mathcal{X}^q} \mathcal{D}_1(\omega_1 | \tilde{\mathbf{x}}_q, \theta_2) \nu_e(\tilde{\mathbf{x}}_q) \pi_a^J(\theta_2 | \tilde{\mathbf{x}}_q) d\tilde{\mathbf{x}}_q d\theta_2 + \mathcal{D}_2(\omega_2).$$

Ces deux termes sont positifs et indépendants. Les minimiseurs sont alors

$$\begin{aligned} \omega_2^* &= \arg \min_{\omega_2 \in \Omega_2} \mathcal{D}_2(\omega_2), \\ \omega_1^* &= \arg \min_{\omega_1 \in \Omega_1} - \int_{\Theta_2} \int_{\Theta_1} \left\{ \int_{\mathcal{X}^q} \pi_a^J(\theta_1 | \theta_2, \tilde{\mathbf{x}}_q) \pi_a^J(\theta_2 | \tilde{\mathbf{x}}_q) \nu_e(\tilde{\mathbf{x}}_q) d\tilde{\mathbf{x}}_q \right\} \log \pi(\theta_1 | \theta_2) d\theta_1 d\theta_2 \end{aligned}$$

ce qui est équivalent à l'énoncé de la proposition.

6.5.2 Simulation d'importance pour la loi de Weibull

Nous reprenons ici les conditions de l'exemple 19 où le calcul des estimations *a posteriori* placées dans les tableaux 6.1 et 6.2 nécessitent de pouvoir simuler selon la densité instrumentale $\rho^J(\mu, \beta)$ définie au § 6.4.

Un MTS pour Weibull est un couple $(x_i, x_j) = x(l)$ tel que $x_i \neq x_j$ et $x_i, x_j > 1$ (Sun, 1997). Notons qu'agréablement, d'après le § 4.9.3, la densité $\rho^J(\mu, \beta)$ est explicite et s'écrit,

$$\rho^J(\mu, \beta) = \frac{1}{ML} \sum_{i=1}^M \sum_{j=1}^{L_1} \sum_{j=1}^{L_2} \rho^{ij}(\mu | \beta)$$

où $L = L_1 L_2$ et

$$\begin{aligned} \mu | \beta &\sim \mathcal{G}\left(2, x_i^\beta + x_j^\beta\right), \\ \rho^{ij}(\beta) &= \frac{(x_i x_j)^{\beta-2}}{2 |\log x_i / x_j| (x_i^\beta + x_j^\beta)^2}. \end{aligned}$$

La difficulté de la méthode d'échantillonnage est donc de simuler facilement des réalisations de $\rho^{ij}(\beta)$. Une simple méthode d'acceptation-rejet (Robert & Casella 2004, chap. 2) permet d'effectuer rapidement cette simulation en sélectionnant les MTS les plus appropriés. On peut remarquer que $\forall \beta > 0$, $x_i^\beta + x_j^\beta \geq \max^\beta(x_i, x_j)$. On en déduit que

$$\begin{aligned} \frac{(x_i x_j)^\beta}{(x_i^\beta + x_j^\beta)^2} &\leq \exp\left(-2\beta \log \frac{\max(x_i, x_j)}{\sqrt{x_i, x_j}}\right), \\ &\leq \exp(-\beta |\log x_i / x_j|). \end{aligned}$$

Alors $\pi^{ij}(\beta) \leq T(x_i, x_j) g_{ij}(\beta)$ où g_{ij} est la densité de la loi exponentielle $\mathcal{E}(|\log x_i / x_j|)$, et

$$T(x_i, x_j) = \frac{(x_i x_j)^{-2}}{2 (\log x_i / x_j)^2}.$$

Ainsi, le nombre d'appels au générateur aléatoire effectués par un algorithme de rejet suit la loi géométrique $\mathcal{G}(T^{-1}(x_i, x_j))$. L'espérance du nombre d'appels à ce générateur est $T(x_i, x_j)$. Alors,

$$\frac{\max(x_i, x_j)}{\min(x_i, x_j)} > \exp(1/\sqrt{2}) \simeq 2.028 \Rightarrow T(x_i, x_j) < 1.$$

Ce faisant, on propose de sélectionner de composer un MTS d'une valeur basse et d'une valeur haute dans les échantillons pour minimiser le temps de calcul de la méthode. Numériquement, cette méthode a toujours donné des résultats rapides, même sans appliquer cette règle pour chaque MTS.

Annexe **A**

The competing risk model $\mathcal{B}(\eta_0, \eta_1, \beta)$

L'article commençant page suivante, co-rédigé avec Henri Bertholon et Gilles Celeux, fait l'objet du rapport de recherche INRIA n°5265 (2004), et a été accepté pour publication par *Lifetime Data Analysis* sous le titre *An alternative competing risk model to the Weibull distribution for modelling aging in lifetime data analysis*. Dans cet article, on note T la variable aléatoire associée à la durée de vie (notée souvent X dans les chapitres de la thèse, dans un cadre parfois plus large que la durabilité). Par ailleurs, les données sont notées \mathbf{y} à la place de \mathbf{y}_n .

The following article, coauthored with Henri Bertholon and Gilles Celeux, has been previously published in 2004 as the INRIA research report n°5265. After some improvements it has been accepted for publication by *Lifetime Data Analysis* in 2006.

An alternative competing risk model to the Weibull distribution for modelling aging in lifetime data analysis

Abstract. A simple competing risk distribution as a possible alternative to the Weibull distribution in lifetime analysis is proposed. This distribution corresponds to the minimum between exponential and Weibull distributions. Our motivation is to take account of both accidental and aging failures in lifetime data analysis. First, the main characteristics of this distribution are presented. Then the estimation of its parameters are considered through maximum likelihood and Bayesian inference. In particular the existence of a unique consistent root of the likelihood equations is proved. Decision tests to choose between an exponential, Weibull and this competing risk distribution are presented. And this alternative model is compared to the Weibull model from numerical experiments on both real and simulated data sets, especially in an industrial context.

Keywords: Failure Time Distribution; Aging; Weibull Distribution; Accidental Failure; Competing risk Model; EM algorithm; Bayesian Inference; Importance sampling; Likelihood ratio Test.

A.1 Introduction

In a reliability context, two of the most employed lifetime distributions are the exponential and the Weibull distributions (see for instance Meeker and Escobar, 1998, pp.79 & 85-86). The exponential distribution $\mathcal{E}(\eta)$ whose reliability function is

$$S_E(t) = \exp\left(-\frac{t}{\eta}\right), \quad (\text{A.1})$$

η being the scale parameter, is modelling accidental failure times of a no aging material cleared of infant mortality defects. While the versatile Weibull $\mathcal{W}(\eta, \beta)$ distribution, with reliability function

$$S_W(t) = \exp\left\{-\left(\frac{t}{\eta}\right)^\beta\right\} \quad (\text{A.2})$$

and hazard rate

$$h_W(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1} \quad (\text{A.3})$$

can be used for modelling infant mortality defects when the shape parameter $\beta < 1$ or aging when $\beta > 1$. Note that when $\beta = 1$ the Weibull distribution reduces to an exponential distribution with scale parameter η .

Reliability feedback experience data are often modelled with the Weibull distribution and an important question is to decide if $\beta = 1$ versus $\beta < 1$ when concerned with infant mortality or $\beta = 1$ versus $\beta > 1$ when concerned with aging. This question can be solved using likelihood ratio tests (see d'Agostino and Stephens, 1986). In this article we are interested in modelling possible aging of a material cleared of infant mortality defects. If, for instance, aging is diagnosed, then further statistical inference is made assuming that the observed failure times arise from a Weibull distribution. When accidental failures and failures due to aging, with an increasing aging hazard function, work together in a system, it means that the occurrence of accidental failures can be regarded as negligible as compared to the occurrence of failures caused by material aging. This assumption could appear to be reasonable in many circumstances, but there are a lot of situations where neglecting accidental failures can introduce an important bias in statistical inference on material lifetimes. Even when aging

is the most frequent cause of failure, accidental failures can remain numerous.

Thus a more realistic way of modelling failure times is to consider a competing risk model which takes into account the fact that a failure can be caused by aging or by accident. This model is defined as follows. A failure time is the realization of a random variable $B = \min(E, W)$ where E is a random variable with an exponential distribution $\mathcal{E}(\eta_0)$ and W is a random variable with a Weibull distribution $\mathcal{W}(\eta_1, \beta)$ where $\beta > 1$, E and W being independent. Consequently, the distribution of B is characterized by three parameters η_0 , η_1 and β . It will be denoted $\mathcal{B}(\eta_0, \eta_1, \beta)$. Its failure rate function is the sum of the exponential and Weibull failure rates

$$h_B(x) = \frac{1}{\eta_0} + \frac{\beta}{\eta_1} \left(\frac{x}{\eta_1} \right)^{\beta-1}. \quad (\text{A.4})$$

The aim of this paper is to analyze the possibility of using a $\mathcal{B}(\eta_0, \eta_1, \beta)$ instead of a $\mathcal{W}(\eta, \beta)$ distribution for modelling aging. The deliberate choice of taking $\beta > 1$ is justified by the fact we consider systems cleared from infant failures but possibly submitted to aging. Such a context is current when studying nuclear plant systems lifetimes (see for instance Lannoy and Procaccia, 2001, or Bacha *et al.*, 1998). Failures due to infant mortality will not be considered in this paper.

The use of competing risk models is well known and a large family of distributions has been notably presented in Park and Padgett (2004). Theory of competing risks is now readily available at a far more sophisticated level than the \mathcal{B} distribution (see for instance Goetghebeur and Ryan, 1995 and Chan and Meeker, 1999). But the \mathcal{B} distribution deserves a special interest since it represents one of the simplest, most practical and intuitive competing risk models, with only three parameters.

The article is organized as follows. In Section 2 the main characteristics of the $\mathcal{B}(\eta_0, \eta_1, \beta)$ distribution are presented. In Section 3, the estimation of the $\mathcal{B}(\eta_0, \eta_1, \beta)$ distribution is considered. First maximum likelihood estimation of the three parameters of the \mathcal{B} distribution from possibly right censored data is presented through the EM algorithm. Then, Bayesian estimation of those parameters is presented using an importance sampling approach to approximate the posterior distribution of the parameters. Section 4 is concerned with hypothesis tests. In particular, the important problem of testing a Weibull distribution against a \mathcal{B} distribution is considered. Section 5 is devoted to the presentation of numerical experiments on both simulated, prospective and real data sets and a short discussion section ends the paper.

A.2 Characteristics of the \mathcal{B} distribution

Let a random variable (r.v.) $B = \min(E, W)$, where the r.v. E has an exponential distribution with mean value η_0 and the r.v. W has a Weibull distribution with scale parameter η_1 and shape parameter β , E and W being independent. The main characteristics of the \mathcal{B} probability distribution are as follows. Its hazard function, its reliability (or survival) function and its probability density function (pdf) are successively given by

$$h_B(t) = h_E(t) + h_W(t) = \frac{1}{\eta_0} + \frac{\beta}{\eta_1} \left(\frac{t}{\eta_1} \right)^{\beta-1}, \quad t \geq 0$$

$$S_B(t) = S_E(t) \times S_W(t) = \exp \left\{ -\frac{t}{\eta_0} - \left(\frac{t}{\eta_1} \right)^\beta \right\}, \quad t \geq 0$$

$$f_B(t) = \left[\frac{1}{\eta_0} + \frac{\beta}{\eta_1} \left(\frac{t}{\eta_1} \right)^{\beta-1} \right] \exp \left\{ -\frac{t}{\eta_0} - \left(\frac{t}{\eta_1} \right)^\beta \right\}, \quad t \geq 0. \quad (\text{A.5})$$

On Figure A.1 are displayed examples of \mathcal{B} pdf's with the corresponding exponential and Weibull pdf's.

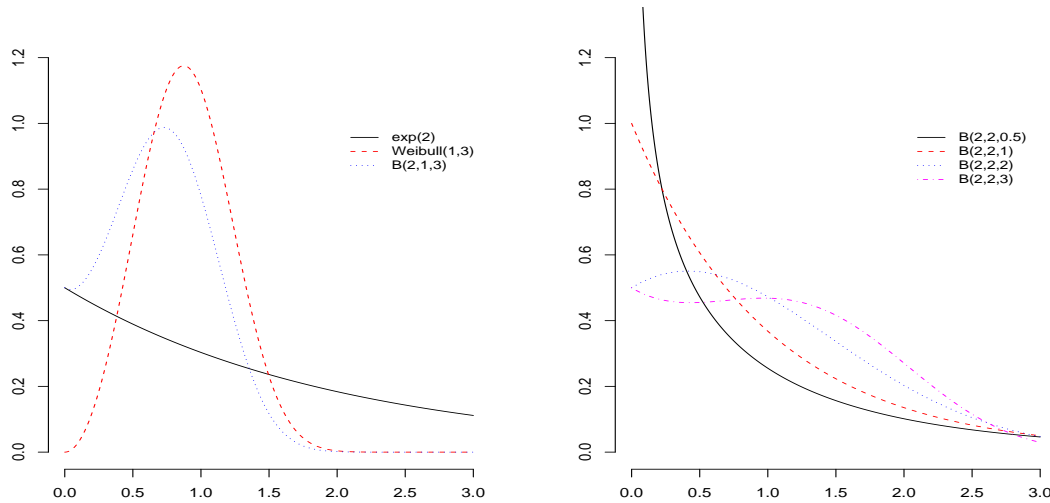


Figure A.1: Examples of \mathcal{B} pdf's.

In view of analyzing the roles of scale parameters η_0 and η_1 in failure time data analysis, it is of interest to examine the particular case of a shape parameter $\beta = 2$ where calculations are straightforward. In this setting, the probability that a failure arises from E (accidental failure) is given by

$$P(B = E) = P(E \leq W) = \frac{\eta_1}{\eta_0} \frac{\sqrt{\pi}}{2} \exp \left[\frac{\eta_1^2}{4\eta_0^2} \right] \operatorname{erfc} \left(\frac{\eta_1}{2\eta_0} \right)$$

where the function erfc is

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{+\infty} \exp(-u^2) du.$$

Table A.1 displays the evolution of this probability as a function of ratio $\frac{\eta_0}{\eta_1}$.

η_0/η_1	0.1	0.2	0.5	1	1.5	2	5	10
$P(B = E)$	0.98	0.93	0.75	0.54	0.42	0.34	0.15	0.08

Table A.1: Probability of an accidental failure as a function of ratio η_0/η_1 in a $\mathcal{B}(\eta_0, \eta_1, 2)$ distribution.

Then the moment generating function of \mathcal{B} (for $\beta = 2$) can be written as (see Bertholon 2001 for details)

$$G(u) = 1 + \exp \left[\frac{1}{2} \eta_1 \left(\frac{1}{\eta_0} - u \right)^2 \right] \left[\eta_1 u \frac{\sqrt{\pi}}{2} \operatorname{erfc} \left\{ \frac{1}{2} \eta_1 \left(\frac{1}{\eta_0} - u \right) \right\} \right]. \quad (\text{A.6})$$

From this expression, it is possible to derive the mean and the variance :

$$E[T] = \eta_1 \exp \left[\frac{\eta_1^2}{4\eta_0^2} \right] \frac{\sqrt{\pi}}{2} \operatorname{erfc} \left(\frac{\eta_1}{2\eta_0} \right)$$

and

$$\begin{aligned} \operatorname{Var}[T] &= 2\eta_1 \exp \left[-\frac{\eta_1^2}{4\eta_0^2} \right] \left(-\frac{\eta_1^2}{2\eta_0} \frac{\sqrt{\pi}}{2} \operatorname{erfc} \left(\frac{\eta_1}{2\eta_0} \right) + \frac{1}{2} \eta_1 \exp \left[-\frac{\eta_1^2}{4\eta_0^2} \right] \right) - \\ &\quad \left[\eta_1 \exp \left[\frac{\eta_1^2}{4\eta_0^2} \right] \frac{\sqrt{\pi}}{2} \operatorname{erfc} \left(\frac{\eta_1}{2\eta_0} \right) \right]^2. \end{aligned}$$

Now coming back to the general setting, not surprisingly, $\mathcal{B}(\eta_0, \eta_1, \beta) \approx \mathcal{W}(\eta_1, \beta)$ as $\eta_0 \gg \eta_1$ and $\mathcal{B}(\eta_0, \eta_1, \beta) \approx \mathcal{E}(\eta_0)$ as $\eta_1 \gg \eta_0$. Since we are presumably concerned with situations where aging can be sensitive, it is reasonable to assume that $\eta_0 \geq \eta_1$, because $\eta_1 > \eta_0$ implies a predominant frequency of accidental failure. In the following, this a priori assumption will be made in the Bayesian framework.

A.3 Estimating the parameter of the \mathcal{B} distribution

A.3.1 Maximum Likelihood Estimation

The \mathcal{B} distribution is a three-parameter distribution and estimating its parameters can be thought of as more difficult than estimating the two parameters of a Weibull distribution. When using the maximum likelihood principle to obtain estimates of the \mathcal{B} distribution parameters, two difficulties arise. The first one, underlined by Friedman and Gertsbakh (1980), is that there is a path in the multivariate parameter space along which the likelihood function tends to infinity. However we prove in the Appendix that the likelihood equations have a unique consistent root, even in a type-I censored case. This result justifies the maximum likelihood method and answers to a significant question raised for instance in Flehinger *et al.* (2002).

The second obstacle is that direct likelihood maximization often leads to unstable numerical results. To get round this difficulty it is however possible to take profit of the fact that the \mathcal{B} distribution can be regarded as an incomplete data model. This enables to use estimation algorithms as the EM algorithm for maximum likelihood (ML) estimation (Dempster, Laird and Rubin, 1977) or Data augmentation algorithms for Bayesian inference (Tanner and Wong, 1987). As it will be noticed in the next section, the advantage of this choice is that the maximization in the M step of the EM algorithm is made separately with respect to the parameters of the exponential and Weibull distributions. This explains why the numerical estimates are obtained in a more stable way by the EM algorithm. Notice that the EM algorithm has been previously used several times to estimate competing risk models with masked data (see Usher and Hogdson, 1988, Bacha *et al.* 1998, Flehinger *et al.* 2002, Craiu and Duchesne 2004, among others). We next present this algorithm in detail.

A.3.2 The EM algorithm

Let $\mathbf{y} = (y_1, \dots, y_n)$ be a sample from a \mathcal{B} distribution which can contain type I right censored data (fixed censoring time). Each y_i can be written as $y_i = (t_i, \delta_i)$, where

$$\delta_i = \begin{cases} 0 & \text{if } t_i \text{ is a censoring time,} \\ 1 & \text{if } t_i \text{ is a failure time.} \end{cases}$$

Thus the observed likelihood which is

$$\begin{aligned} L(\eta_0, \eta_1, \beta | \mathbf{y}) &= \prod_{i=1}^n [f_B(t_i)]^{\delta_i} [S_B(t_i)]^{1-\delta_i} \\ &= \prod_{i=1}^n [h_B(t_i)]^{\delta_i} S_B(t_i), \end{aligned}$$

can be written as

$$L(\eta_0, \eta_1, \beta | \mathbf{y}) = \left[\prod_{i=1}^n \left(\frac{1}{\eta_0} + \frac{\beta}{\eta_1} \left(\frac{t_i}{\eta_1} \right)^{\beta-1} \right)^{\delta_i} \right] \exp \left[-\frac{1}{\eta_0} \sum_{i=1}^n t_i - \sum_{i=1}^n \left(\frac{t_i}{\eta_1} \right)^{\beta} \right]. \quad (\text{A.7})$$

The \mathcal{B} distribution is the distribution of a competing risk model with missing data. The missing data are binary indicator values associated to the failure times. If t_i is a failure time, we define $z_i = (z_i^E, z_i^W)$ where $z_i^E = 1$ and $z_i^W = 0$ if the failure time t_i arose from the exponential distribution and $z_i^W = 1$ and $z_i^E = 0$ if t_i arose from the Weibull distribution. By convention, if t_i is a censoring time ($\delta_i = 0$), $z_i^E = 0$ and $z_i^W = 0$. Thus, the complete data set can be written as $\mathbf{x} = (x_i = (y_i, z_i), i = 1, \dots, n) = (\mathbf{y}, \mathbf{z})$. The density of a complete observation x_i is

$$\begin{aligned} f(x_i) &= [f_E(t_i)]^{z_i^E} [S_E(t_i)]^{1-z_i^E} [f_W(t_i)]^{z_i^W} [S_W(x_i)]^{1-z_i^W}, \\ f(x_i) &= [h_E(t_i)]^{z_i^E} [h_W(t_i)]^{z_i^W} S_E(t_i) S_W(t_i). \end{aligned}$$

And the complete loglikelihood can be written

$$l(\theta | \mathbf{x}) = \sum_{i=1}^n z_i^E \ln [h_E(t_i)] + z_i^W \ln [h_W(t_i)] + \ln [S_E(t_i)] + \ln [S_W(t_i)]. \quad (\text{A.8})$$

The EM algorithm consists of maximizing the conditional expectation of the complete likelihood knowing the observed data and a current value $\tilde{\theta}$ of the parameter in an iterative two-step algorithm (Dempster, Laird and Rubin 1977, McLachlan and Krishnam 1997). The E step is calculating this conditional expectation, denoted by $Q(\theta | \tilde{\theta})$, and the M step is maximizing $Q(\theta | \tilde{\theta})$ with respect to θ .

E step

It consists of calculating $Q(\theta | \tilde{\theta})$, $\tilde{\theta}$ being the current parameter value.

$$\begin{aligned} Q(\theta | \tilde{\theta}) &= E \left(l(\theta | \mathbf{x}) | \mathbf{y}, \tilde{\theta} \right) \\ &= \sum_{i=1}^n \left[E \left(z_i^E | \mathbf{y}, \tilde{\theta} \right) \ln [h_E(t_i)] + E \left(z_i^W | \mathbf{y}, \tilde{\theta} \right) \ln [h_W(t_i)] + \ln [S_E(t_i)] + \ln [S_W(t_i)] \right] \\ &= \sum_{i=1}^n \tilde{p}_E(y_i) \ln [h_E(t_i)] + \tilde{p}_W(y_i) \ln [h_W(t_i)] + \ln [S_E(t_i)] + \ln [S_W(y_i)] \end{aligned}$$

where $\tilde{p}_E(y_i)$ (resp. $\tilde{p}_W(y_i)$) is the conditional probability that the observed failure time y_i is assigned to the exponential (resp. Weibull) distribution. These probabilities are calculated as follows :

$$\tilde{p}_E(y_i) = \begin{cases} 0 & \text{if } y_i \text{ is a censored value } (\delta_i = 0), \\ P(z_i^E = 1|y_i, \tilde{\theta}) = P(E_i \leq W_i|y_i, \tilde{\theta}) & \text{else} \end{cases}$$

with

$$\begin{aligned} P(E_i \leq W_i|y_i, \tilde{\theta}) &= \frac{f_E(t_i)S_W(t_i)}{f_E(t_i)S_W(t_i) + f_W(t_i)S_E(t_i)}, \\ &= \frac{h_E(t_i)}{h_E(t_i) + h_W(t_i)}. \end{aligned}$$

Similarly,

$$\tilde{p}_W(y_i) = \begin{cases} 0 & \text{if } \delta_i = 0 \\ \frac{h_W(t_i)}{h_E(t_i) + h_W(t_i)} & \text{if } \delta_i = 1. \end{cases}$$

$Q(\theta|\tilde{\theta})$ can be written

$$Q(\theta|\tilde{\theta}) = Q_E(\eta_0|\tilde{\theta}) + Q_W(\eta_1, \beta|\tilde{\theta}) \tag{A.9}$$

where

$$Q_E(\eta_0|\tilde{\theta}) = \sum_{i=1}^n [\tilde{p}_E(x_i) \ln(h_E(x_i)) + \ln(S_E(x_i))]$$

and

$$Q_W(\eta_1, \beta|\tilde{\theta}) = \sum_{i=1}^n [\tilde{p}_W(x_i) \ln(h_W(x_i)) + \ln(S_W(x_i))].$$

This additive decomposition of $Q(\theta|\tilde{\theta})$ between the contribution of the exponential and the Weibull distributions will facilitate the M step which is now described.

M step

It consists of deriving $\hat{\theta} = \arg \max_{\theta} Q(\theta|\tilde{\theta})$. From (A.9) it leads to derive $\hat{\eta}_0 = \arg \max_{\eta_0} Q_E(\eta_0|\tilde{\theta})$, and $(\hat{\eta}_1, \hat{\beta}) = \arg \max_{(\eta_1, \beta)} Q(\eta_1, \beta|\tilde{\theta})$. Thus the following equations are straightforwardly obtained

$$\hat{\eta}_0 = \frac{\sum_{i=1}^n t_i}{\sum_{i=1}^n \tilde{p}_E(y_i)},$$

$$\frac{1}{\hat{\beta}} + \frac{\sum_{i=1}^n \tilde{p}_W(y_i) \ln(t_i)}{\sum_{i=1}^n \tilde{p}_W(y_i)} - \frac{\sum_{i=1}^n (t_i)^{\hat{\beta}} \ln(t_i)}{\sum_{i=1}^n (t_i)^{\hat{\beta}}} = 0,$$

and

$$\hat{\eta}_1 = \left(\frac{\sum_{i=1}^n (t_i)^{\hat{\beta}}}{\sum_{i=1}^n \tilde{p}_W(y_i)} \right)^{\frac{1}{\hat{\beta}}}.$$

And the resulting $\hat{\theta}$ becomes the current parameter value.

The EM algorithm increases the observed likelihood $L(\eta_0, \eta_1, \beta|\mathbf{y})$ at each iteration and is expected to converge toward the ML estimate of η_0, η_1 and β under certain conditions (Dempster, Laird and Rubin 1977, McLachlan and Krishnam 1997). In particular, the initial position of EM has to be close enough to the ML estimate of η_0, η_1 and β and the sample size large enough to ensure the convergence of the EM algorithm to the unique consistent solution of the likelihood equations. From our experiments, EM converges to the ML estimate of η_0, η_1 and β for moderately censored sample with moderate sizes.

A.3.3 Bayesian inference through importance sampling

In many cases, the estimation of reliability distribution functions has to be done from small and highly censored samples. In such a situation, ML estimation turns out to be imprecise or even unreliable (see for instance Bacha *et al.*, 1998). Bayesian inference can be expected to be useful since, in many circumstances including engineering applications, there exists some expert knowledge on the underlying failure mechanism which can be translated into good prior information on the failure distribution parameters. Bayesian inference concerning competing risk models involving Weibull distributions has been considered by several authors including Berger and Sun (1993), Bacha *et al.* (1998) and Basu *et al.* (2003).

In Bayesian inference, a prior probability distribution $\pi(\theta)$ is specified for the parameter to be estimated and leads to the posterior distribution $\pi(\theta|\mathbf{y}) \propto L(\theta|\mathbf{y})\pi(\theta)$ from which the inference is based. Approximating the posterior distribution of the parameters $\theta = (\eta_0, \eta_1, \beta)$ of a \mathcal{B} distribution or more generally of a competing risk model is a difficult task for which Monte Carlo approximation using Monte Carlo Markov Chains (MCMC) methods (a good reference on MCMC methods is Robert and Casella, 1999) as Gibbs sampling or Hasting-Metropolis algorithm can be used (see Berger and Sun, 1993). However, it has been noticed in Bacha *et al.* (1998) that MCMC methods can encounter prohibitively slow convergence situation especially when there is a small amount of observed failure times. Moreover, assessing the convergence of a Markov chain generated by a MCMC algorithm is a difficult theoretical challenge and consequently the quality of the posterior distribution approximation is dubious. Thus it can present some advantages to use importance sampling techniques instead. This is the approach considered in the present article.

Importance sampling (see Robert and Casella 1999, chapter 3) is based on the simulation of θ'_i 's ($i = 1, \dots, M$) from an instrumental distribution $\rho(\theta)$. The difference between the distribution of interest $\pi(\theta|\mathbf{y})$ and the instrumental distribution $\rho(\theta)$ is corrected using importance weights

$$\omega_i = \frac{\pi(\theta_i|\mathbf{y})/\rho(\theta_i)}{\sum_{j=1}^M \pi(\theta_j|\mathbf{y})/\rho(\theta_j)}$$

to preserve that, for any function of interest $h(\theta)$,

$$\sum_{j=1}^M \omega_j h(\theta_j) \approx \int h(\theta)\pi(\theta|\mathbf{y})d\theta. \quad (\text{A.10})$$

The art of importance sampling lies in choosing a good importance function ρ . This choice is paramount to ensure that convergence to the posterior distribution π occurs at the right rate, the minimum requirement being that the variance of the importance weights ω_i is finite (Robert and

Casella 1999, chapter 3). Importance sampling techniques received a renewed interest since a few years. In particular the population Monte Carlo approach that we chose received benefit from earlier works on MCMC algorithms to select good proposal distributions. It thus comes as a further advance that exploits dependence on previous iterations without requiring ergodicity and the theoretical apparatus of Markov chain theory that are needed when using MCMC algorithms. Following Celeux *et al.* (2006), in the spirit of data augmentation methods (Tanner and Wong 1987, Tanner 1991, Robert and Casella 1999), it can be taking profit of the missing data structure of the problem to produce a simple and feasible importance function by simulating missing data. It leads to propose two adaptative importance sampling schemes that are now presented.

The SRE scheme Using the ideas developed in Steele, Raftery and Emond (2003) a first scheme, denoted SRE in the following, is now described. Denoting $L(\theta|\mathbf{y}, \mathbf{z})$ the completed likelihood and $k(\mathbf{z}|\mathbf{y}, \theta) = L(\theta|\mathbf{y}, \mathbf{z})/L(\theta|\mathbf{y})$ being the conditional density of the missing data knowing the observed data, it can be described as follows.

- Compute $\hat{\theta} = \arg \max_{\theta} L(\theta|\mathbf{y})$;
- For $j = 1, \dots, M$
 1. Generate $\mathbf{z}^{(j)}$ from the conditional distribution of the missing data $k(\mathbf{z}|\mathbf{y}, \hat{\theta})$;
 2. Generate $\theta^{(j)}$ from $\pi(\theta|\mathbf{y}, \mathbf{z}^{(j)})$;
 3. Compute $r^{(j)} = \frac{L(\theta^{(j)}|\mathbf{y}, \mathbf{z}^{(j)})\pi(\theta^{(j)})}{k(\mathbf{z}^{(j)}|\mathbf{y}, \hat{\theta})\pi(\theta^{(j)}|\mathbf{y}, \mathbf{z}^{(j)})}$ and $\omega^{(j)} = r^{(j)} / \sum_{s=1}^M r^{(s)}$;
- Generate K realizations $(\hat{\theta}^{(1)}, \dots, \hat{\theta}^{(K)})$ from $(\theta^{(1)}, \dots, \theta^{(M)})$ using a multinomial distribution with probabilities $(\omega^{(1)}, \dots, \omega^{(M)})$.

Even if the proposal distribution is a distribution on θ, \mathbf{z} given \mathbf{y} and not a marginal distribution on θ given \mathbf{y} , the resulting sample $\hat{\theta}^{(1)}, \dots, \hat{\theta}^{(K)}$ can be approximatively regarded as a sample from the posterior distribution $\pi(\theta|\mathbf{y})$ and a standard estimator of $E_{\pi}(h(\theta))$ is $\sum_{j=1}^M \omega^{(j)} h(\theta^{(j)})$. As Tanner (1991), the authors of this scheme justify it by saying that $\pi(\mathbf{z}|\mathbf{y}, \hat{\theta})$ is a good surrogate for the predictive density $\pi(\mathbf{z}|\mathbf{y})$. A possible problem of this scheme is that, typically, Bayesian inference can be useful in small sample setting for which maximum likelihood can provide unreliable estimates (see Bacha *et al.* 1998). Thus in such cases it is doubtful that initiating the sampling scheme from $\hat{\theta}$ is a good choice. A more Bayesian criticism is that this choice does not take into account the intrinsic variability in θ due to the prior distribution, so the importance function could be too concentrated.

The Population Monte Carlo scheme The Population Monte Carlo scheme proposed by Cappé *et al.* (2004) is an iterated scheme that produces, at each iteration, a sample approximatively simulated from $\pi(\theta|\mathbf{y})$ and some approximatively unbiased estimators of integrals under that distribution. The novelty of the method is that the iterated call to importance sampling based on the current importance sampling sample allows for a progressive selection of the most relevant points of the sample. Convergence results on PMC scheme can be found in Douc *et al.* (2005). The most general version of

this procedure is as follows:

- For $j = 1, \dots, M$, choice of $\theta_0^{(j)}$;
- Step i. ($i = 1, \dots$): for $j = 1, \dots, M$
 1. Generate $\theta_i^{(j)}$ from $q_{ij}(\theta|\theta_{i-1})$ where $\theta_{i-1} = (\theta_{i-1}^{(1)}, \dots, \theta_{i-1}^{(M)})$;
 2. Compute $r^{(j)} = \frac{f(y|\theta_i^{(j)})\pi(\theta_i^{(j)})}{q_{ij}(\theta_i^{(j)}|\theta_{i-1})}$ and $\omega^{(j)} = r^{(j)} / \sum_{s=1}^M r^{(s)}$.

By an importance sampling argument (see Cappé *et al.*, 2004), the choice of q_{ij} is fairly unrestricted and this proposal distribution can depend on the previous sample or even on the whole sequence of samples simulated so far. A specific version of this procedure, denoted PMCH in the following, which takes profit of the missing data structure of the problem is as follows:

- For $j = 1, \dots, M$, choice of $\theta_0^{(j)}$;
- Step i. ($i = 1, \dots$)
 - a) For $j = 1, \dots, M$
 1. Generate $\mathbf{z}^{(j)}$ from $k(\mathbf{z}|\mathbf{y}, \theta_{i-1}^{(j)})$;
 2. Generate $\theta_i^{(j)}$ from $\pi(\theta|\mathbf{y}, \mathbf{z}^{(j)})$;
 3. Compute $r^{(j)} = \frac{L(\theta_i^{(j)}|\mathbf{y}, \mathbf{z}^{(j)})\pi(\theta_i^{(j)})}{k(\mathbf{z}^{(j)}|\mathbf{y}, \theta_{i-1}^{(j)})\pi(\theta_i^{(j)}|\mathbf{y}, \mathbf{z}^{(j)})}$ and $\omega^{(j)} = r^{(j)} / \sum_{s=1}^M r^{(s)}$;
 - b) Resample the $(\theta_i^{(j)})$ using the weights $w^{(j)}$.

Choosing the prior distribution In this article, we are concerned with Bayesian inference in an informative context where experts are expected to be able to give good prior information on the parameters (η_0, η_1, β) of the \mathcal{B} distribution. Following Erto (1982), Berger and Sun (1993) and Bacha *et al.* (1998) and owing to our own experience, it is assumed that the shape parameter β is supposed to be in an interval $[\beta_\ell, \beta_r]$. Since, here, we are interested in aging, it is assumed that $\beta_\ell = 1$ and a typical value for β_r is $\beta_r = 5$. The prior density chosen for β is a uniform distribution on this interval. Now, in order to use conjugate Gamma prior distributions for the scale parameters η_0 and η_1 , the chosen prior distribution is as follows. Putting $\lambda_0 = 1/\eta_0$ and $\lambda_1 = 1/\eta_1$, we have

$$\pi(\beta, \lambda_1, \lambda_0) = \pi(\beta)\pi(\lambda_1|\beta)\pi(\lambda_0|\beta, \lambda_1), \quad (\text{A.11})$$

where $\pi(\beta)$ is the uniform distribution on $[\beta_\ell, \beta_r]$,

$$\pi(\lambda_1|\beta) = \frac{\lambda_1^{c\beta-1} d^c \beta \exp(-d\lambda_1^\beta)}{\Gamma(c)} \mathbf{I}_{(0,\infty)}(\lambda_1)$$

and

$$\pi(\lambda_0|\beta, \lambda_1) \propto \frac{\lambda_0^{a-1} b^a \exp(-b\lambda_0)}{\Gamma(a)} \mathbf{I}_{(0,\lambda_1]}(\lambda_0).$$

It means that the prior distribution of λ_0 is a Gamma distribution $\mathcal{G}(a, b)$ truncated in λ_1 . And, it can easily be proved that the conditional distribution of λ_1 knowing β is such that the prior distribution of λ_1^β knowing β is a Gamma distribution $\mathcal{G}(c, d)$. This choice of prior distributions has been proposed by Berger and Sun (1994) for the Weibull model conditionally to the variations of β , similarly to Berger and Sun (1993) for Poly-Weibull distributions. Clearly Gamma distributions have been chosen to push the estimation through. Fortunately, the Gamma distribution is a versatile distribution and this choice does not jeopardize a good treatment of expert knowledge for the scale parameters.

Hyperparameters a, b, c and d are for instance chosen in the following way. Experts are asked to give intervals of possible values for η_0 and η_1 . Denoting $[\eta_{0\ell}, \eta_{0r}]$ respectively $[\eta_{1\ell}, \eta_{1r}]$ those intervals, it leads to

$$\begin{aligned} a &= \alpha \frac{(1/\eta_{0\ell} + 1/\eta_{0r})^2}{(1/\eta_{0\ell} - 1/\eta_{0r})^2} \\ b &= 2\alpha \frac{(1/\eta_{0\ell} + 1/\eta_{0r})}{(1/\eta_{0\ell} - 1/\eta_{0r})^2} \end{aligned}$$

where α is chosen small enough to ensure a large variance of the Gamma distribution, and

$$c = \tau \frac{(1/\eta_{1\ell}^{\beta_\ell} + 1/\eta_{1r}^{\beta_r})^2}{(1/\eta_{1\ell}^{\beta_\ell} - 1/\eta_{1r}^{\beta_r})^2} \quad (\text{A.12})$$

$$d = 2\tau \frac{(1/\eta_{1\ell}^{\beta_\ell} + 1/\eta_{1r}^{\beta_r})}{(1/\eta_{1\ell}^{\beta_\ell} - 1/\eta_{1r}^{\beta_r})^2} \quad (\text{A.13})$$

where τ can be chosen in the same manner as α . This choice is ensuring the prior distributions to be centered on the means of expert intervals and the possibility of enlarging the prior variance.

Implementation of the SRE and PMCH schemes for the \mathcal{B} distribution The implementation of the two above described importance sampling schemes does not involve difficulties and is not detailed here. However some comments are to be made. First, it is important to build an importance function ρ with heavier tails than the posterior to be approximated (Robert and Casella 1999, chapter 3). In that purpose, it is beneficial to enlarge the missing data space. Thus, the missing data we considered are not reduced to be the binary vectors z_i indicating the distribution (exponential or Weibull) from which failure times occur. They include the failure time of the alternative distribution (exponential or Weibull) not assigned to the observed failure time, and also the failure times beyond the censoring times for both the exponential and the Weibull distributions. Then, the completed likelihood takes the form

$$L(\eta_0, \eta_1, \beta|\mathbf{y}, \mathbf{z}) = L(\eta_0|e_1, \dots, e_n) \times L(\eta_1, \beta|w_1, \dots, w_n)$$

where e_i (resp. w_i) are either the observed failure times t_i assigned to the exponential (resp. Weibull) distribution or the simulated failure times according to the exponential distribution $\mathcal{E}(\eta_0)$ (resp. $\mathcal{W}(\eta_1, \beta)$).

Generating $\lambda_0^{(j)}$ and $\lambda_1^{(j)}$ is easy since by conjugate properties, the conditional distribution of $\lambda_0^{(j)}$ is a $\mathcal{G}(a+n, b + \sum_{i=1}^n e_i^{(j)})$ and the conditional distribution of $(\lambda_1^{(j)})^{\beta^{(j)}}$ is a $\mathcal{G}(c+n, d + \sum_{i=1}^n (w_i^{(j)})^{\beta^{(j)}})$. Since, no conjugate prior exists for the shape parameter of a Weibull distribution, generating $\beta^{(j)}$ is carried out with an accept-reject algorithm:

1. Generate β from $\pi(\beta)$.
2. Generate u from a uniform distribution on $[0, 1]$.
3. Put $\beta^{(j)} = \beta$ if $u \leq \frac{\pi(\beta|\mathbf{y}, \mathbf{z}^{(j)}, \eta_0^{(j-1)}, \eta_1^{(j-1)})}{\pi(\beta) \max_{\beta} L(\beta, \eta_0^{(j-1)}, \eta_1^{(j-1)}|\mathbf{y}, \mathbf{z}^{(j)})}$, otherwise goto 1.

A.4 Assessing the failure distribution

The objective of this section is to give a procedure for choosing the most relevant model among the three following ones, given in the increasing order of complexity: the exponential, Weibull and \mathcal{B} model. In a reliability context, the exponential model is often the first proposed distribution. But, if aging is suspected, the Weibull model can be expected to be more appropriate. As for the \mathcal{B} model, it has already been seen that it is even more complex than the Weibull model.

Towards that end, the standard Likelihood Ratio Test (denoted LRT) is an adequate goodness-of-fit test, in an asymptotic framework. More precisely, as these three models are embedded, the two-step following procedure is proposed:

- First the standard exponential versus Weibull test is applied (see for example Lawless 1982, pp. 173-174). As a first step, this enables to detect aging. Under the null hypothesis $H_0: \beta = 1$, the LRT statistic $2 \ln \left(\frac{L_W(\hat{\eta}, \hat{\beta})}{L_E(\hat{\eta}_0)} \right)$ converges to χ_1^2 in distribution.

- Secondly, to go further in the case where the previous test leads to choose a Weibull model, the LRT is now applied to discriminate between Weibull and \mathcal{B} models. In other words, here, the question is to know whether it is worth taking into account a possible accidental cause of failure. To be more precise, for this second test, denoting $\lambda_0 = 1/\eta_0$, the null hypothesis (corresponding to the Weibull model) is defined as $H_0: \lambda_0 = 0$ ($\eta_0 = +\infty$) (which means that the accidental component vanishes). Since the parameter λ_0 is constrained to be ≥ 0 , the distribution under H_0 of the LRT statistic (with notations of § A.3.2)

$$\begin{aligned} 2 \ln \left(\frac{L_{\mathcal{B}}(\hat{\lambda}_0, \hat{\lambda}_1, \hat{\beta})}{L_W(\hat{\lambda}, \hat{\beta})} \right) &= 2 \left[\sum_{i=1}^n \delta_i \ln \left(\hat{\lambda}_0 + \hat{\lambda}_1^{\hat{\beta}} x_i^{\hat{\beta}-1} \right) - \sum_{i=1}^n \left(\hat{\lambda}_0 x_i + \hat{\lambda}_1^{\hat{\beta}} x_i^{\hat{\beta}} \right) \right] - \\ &2 \left[\sum_{i=1}^n \delta_i \left(\tilde{\beta} \ln \tilde{\lambda} + \ln \tilde{\beta} \right) + (\tilde{\beta} - 1) \sum_{i=1}^n \delta_i \ln x_i - \sum_{i=1}^n \left(\tilde{\lambda}^{\tilde{\beta}} x_i^{\tilde{\beta}} \right) \right] \\ &\sim \frac{1}{2} \chi_1^2 + \frac{1}{2} \delta_0 \end{aligned}$$

where δ_0 denotes the Dirac distribution in 0. A general presentation of this result can be found in Gourieroux and Monfort (1996, chapter 21). Beyond standard regularity conditions that are fulfilled

by the \mathcal{B} distribution, their line of proof needs that the Taylor expansion of $\ln L_{\mathcal{B}}$ about $\lambda_0 = 0$ is possible, which is true since $\ln L_{\mathcal{B}}$ is well defined in a neighborhood of $\lambda_0 = 0$.

Thus the application of the LRT is straightforward. Proceeding in this manner, it is possible to examine gradually the complexity of models. Notice that it is also possible to make use of the BIC criterion to select one of the three models. The BIC criterion associated to a model \mathcal{X} with a parameter of dimension k and the maximum likelihood estimate $\hat{\theta}$ (or a Bayesian estimate) can be written as $\ln L_{\mathcal{X}}(\hat{\theta}) - \frac{k}{2} \ln n > 0$ where n is the size of the sample. The selected model is the one which presents higher BIC criterion.

A.5 Numerical experiments

A.5.1 Estimations and tests on simulated data

In this section, estimations of exponential, Weibull and \mathcal{B} distributions on simulated data from a $\mathcal{B}(200, 100, 2)$ distribution are considered. Note that for such a distribution, the probability for a failure to be accidental is 0.34. Samples of decreasing size have been simulated (30 samples for each size) and have been censored with the fixed value 100, which gives approximately 20% censored data. Results of estimation on two censored samples (of size 500 and 50) are summarized in Tables A.2 and A.3. Estimation results are summarized with the median and the [25% – 75%] percentile interval over the 30 samples. The median value is preferred to the mean value to summarize those Monte-Carlo numerical experiments because it can happen that the estimate of η_0 in the \mathcal{B} distribution tends to high values.

In next tables “ML” denomination embodies the use of EM algorithm for the \mathcal{B} distribution, a Newton-Raphson descent algorithm (with adaptative decreasing step) applied to the Weibull distribution and the direct likelihood maximisation for the exponential distribution. In the Bayesian setting, prior distributions have been designed from the equations given in Section A.3.3 from the following prior informations. The shape parameter β is in $[\beta_l, \beta_r] = [1, 5]$, the scale parameter η_0 of the exponential component is concentrated on $(\eta_{0l}, \eta_{0r}) = (1, 300)$ and the scale parameter η_1 of the Weibull component is concentrated on $(\eta_{1l}, \eta_{1r}) = (1, 200)$.

In those tables the estimation of the mean residual survival time

$$MRST(t_0) = \frac{\int_t^\infty S(u) du}{S(t)}$$

has been added with its empirical standard deviation into parentheses, S denoting the survival function and t_0 a fixed value of time. In the aim to predict the behavior of the model beyond the observed distribution, we choose $t_0 = (\eta_0 + \eta_1)/2 = 150$ which theoretically implies $S_{\mathcal{B}}(t_0) = 4.97\%$ and $MRST_{\mathcal{B}}(t_0) = 25.25$. Compared with the MRST obtained with the Weibull distribution, the improvement of the prediction is especially highlighted throughout the treatment of 50-sized data.

It is to be noticed that the estimation of parameter η_0 seems to be more sensitive to censoring time and sample size than the two other ones. However, the LRT test and BIC criterion choose the correct distribution in most cases. In the small sample size and censored cases, the model choice favoring the Weibull distribution is not surprising since the estimation of η_0 in \mathcal{B} modelling tends to high values,

so that the \mathcal{B} distribution is well approximated with a Weibull distribution.

	<i>exponential</i>		<i>Weibull</i>			<i>B model</i>			
	η	MRST(150)	η	β	MRST(150)	η_0	η_1	β	MRST(150)
ML	82.3	82.3	78.2	1.47	35.75	170.5	100.4	2.21	25.15
	(75.1-86.5)	(8.48)	(70.5-84.8)	(1.33-1.69)	(14.27)	(153.4-195.7)	(84.2-119.7)	(1.91-2.29)	(11.25)
PMCH	82.5	82.5	83.4	1.60	32.90	211.2	95.5	1.95	26.63
	(75.1-87.1)	(9.0)	(73.2-89.9)	(1.38-1.72)	(14.38)	(185.3-226.9)	(86.5-120.1)	(1.90-2.22)	(10.88)

Table A.2: ML and Bayesian estimations of parameters (with interquartile intervals) and MRST (with standard deviations) of 500-sized $\mathcal{B}(200, 100, 2)$ samples (20% of censored data).

	<i>exponential</i>		<i>Weibull</i>			<i>B model</i>			
	η	MRST(150)	η	β	MRST(150)	η_0	η_1	β	MRST(150)
ML	82.4	82.4	80.6	1.82	30.0	345.1	83.6	1.84	28.84
	(64.1-95.4)	(23.7)	(58.5-95.8)	(1.37-2.14)	(22.4)	(179.4-567.8)	(58.9-114.7)	(1.42-2.27)	(22.6)
PMCH	80.8	80.8	81.3	1.74	26.7	235.1	85.5	1.85	23.12
	(69.5-90.7)	(15.7)	(62.5-91.4)	(1.53-2.10)	(17.6)	(181.5-262.7)	(63.5-98.7)	(1.67-2.09)	(12.5)

Table A.3: ML and Bayesian estimations of parameters (with interquartile intervals) and MRST (with standard deviations) of 50-sized $\mathcal{B}(200, 100, 2)$ samples (20% of censored data).

<i>size</i>	<i>censoring %</i>	<i>exponential</i>		<i>Weibull</i>		<i>B model</i>	
		LRT	BIC	LRT	BIC	LRT	BIC
500	0%	0%	0%	6.6%	3.3%	93.4%	96.7%
	20%	0%	0%	3.3%	3.3%	96.7%	96.7%
200	0%	0%	0%	6.7%	3.3%	93.3%	96.7%
	20%	0%	0%	10%	3.3%	90%	96.7%
50	0%	3.3%	3.3%	23.3%	20%	73.4%	76.7%
	20%	3.3%	6.6%	86.7%	30%	10%	63.4%

Table A.4: Model choice for samples (from the $H_1 = \mathcal{B}$ distribution) of sizes 500, 200 and 50 (5%-level tests) by LRT/BIC, tested on 30 samples.

The performance of both approaches under the null hypothesis of a Weibull distribution is studied in Table A.5. It presents the proportion of samples (among 100 items) from a $\mathcal{W}(100, 2)$ distribution for which the LRT (at 5%-level) and BIC statistics accept the null hypothesis. The BIC method appears to be more appropriate when the proportion of censored data increases, because the asymptotic approximation of the LRT is deteriorating.

A.5.2 Estimation and tests on prospective data

The following data are sampled from a *prospective* mortality table given by INSEE (INSEE 2001) for 100,000 people born in 1993. It means that the individual survivals are projections in the future. This mortality table was presented in Bertholon (2001). In this work the EM estimation of the \mathcal{B} distribution parameters were $\hat{\eta}_0 = 1662$ years, $\hat{\eta}_1 = 97.4$ years, and $\hat{\beta} = 12.4$. The value of shape

<i>censoring %</i>	LRT	BIC	<i>censoring %</i>	LRT	BIC
0%	0.96	0.98	0%	0.95	0.98
20%	0.90	0.96	20%	0.91	0.95
50%	0.57	0.74	50%	0.85	0.93
80%	0.51	0.68	80%	0.73	0.83

Table A.5: Proportions of well-recognized samples of size 50 (left) and 200 (right) the $H_0 = \mathcal{W}(100, 2)$ distribution.

parameter β is unusual. It indicates that aging is accelerating after some time, which is not really surprising concerning idealised human life. The \mathcal{B} estimated mean is numerically computed from

$$E(X) = \int_0^\infty \exp \left\{ -\frac{x}{\hat{\eta}_0} - \left(\frac{x}{\hat{\eta}_1} \right)^{\hat{\beta}} \right\} dx \simeq 90.8 \text{ years.}$$

Note that according to these results the mean lifetime should be 1662 years if humans would not be submitted to aging! If aging is only taken into account (as it is generally the case) with a Weibull distribution the mean lifetime remains close to 97 years for people born in 1993. On Figure A.3 survival functions for the Weibull and \mathcal{B} estimated distributions are displayed.

From these data 30 sets of size 1,000 and 100 respectively are randomly sampled, for estimating the parameters of the three considered lifetime models and analyzing the results of hypothesis tests. All samples are considered uncensored. Because of relatively large sample sizes ($n \geq 100$), maximum likelihood was used to estimate the parameters of the models. The ML mean estimates and the MTS mean estimates at $t_0 = 97$ years are presented in Table A.6, LRT test results in Table A.7 and Weibull plots for the 100-sized case in Figure A.2.

<i>size</i>	<i>exponential</i>	<i>Weibull</i>		<i>B model</i>	
	$\hat{\eta}$	$(\hat{\eta}, \hat{\beta})$	<i>MRTS(97)</i>	$(\hat{\eta}_0, \hat{\eta}_1, \hat{\beta})$	<i>MRTS(97)</i>
1,000	83.1	(87.5,7.5)	4.5	(1285.5,90.2,13.8)	2.1
100	84.2	(88.8,10.2)	3.0	(1510.5,88.8,14.4)	1.5

Table A.6: ML mean estimates and MTS mean estimates at $t_0 = 97$ years from 30 samples of sizes 1,000 and 100 coming from the prospective mortality table.

<i>size</i>	<i>exponential</i>	<i>Weibull</i>	<i>B model</i>	<i>Model choice</i>
1,000	-5425.45	-4167.23	-3472.17	<i>B model</i>
100	-543.09	-392.24	-351.95	<i>B model</i>

Table A.7: Loglikelihood mean values and model choice for samples of sizes 1,000 and 100 from the prospective mortality table (5%-level tests)

From Weibull plots in Figure A.2, it clearly appears that the \mathcal{B} distribution provides a better fit to the data than the Weibull distribution. Neglecting the accidental cause of dying produces an unreliable analysis of prospective human lifetime. Extensive differences of estimated survival functions for the Weibull and the \mathcal{B} distributions appear in Figure A.3.

An interesting characteristic of the \mathcal{B} distribution can be observed on Weibull plots, in Figure A.2. The empirical distribution is divided into two parts (in the proximity of $\log t = 4.1$), corresponding respectively to the underlying exponential distribution (from $\log t \in [0, 4.1]$) and the Weibull distribution. From the formal expression of \mathcal{B} distribution, the Weibull plot $\log(-\log S_W(t)) = \log(t/\eta_0 + (t/\eta_1)^\beta)$ displays a knee in the neighbourhood of $\log t = 4.1$. Thus, aging becomes the prominent cause of failure for $t > t_o \simeq \exp(4.1) = 60.34$ years.

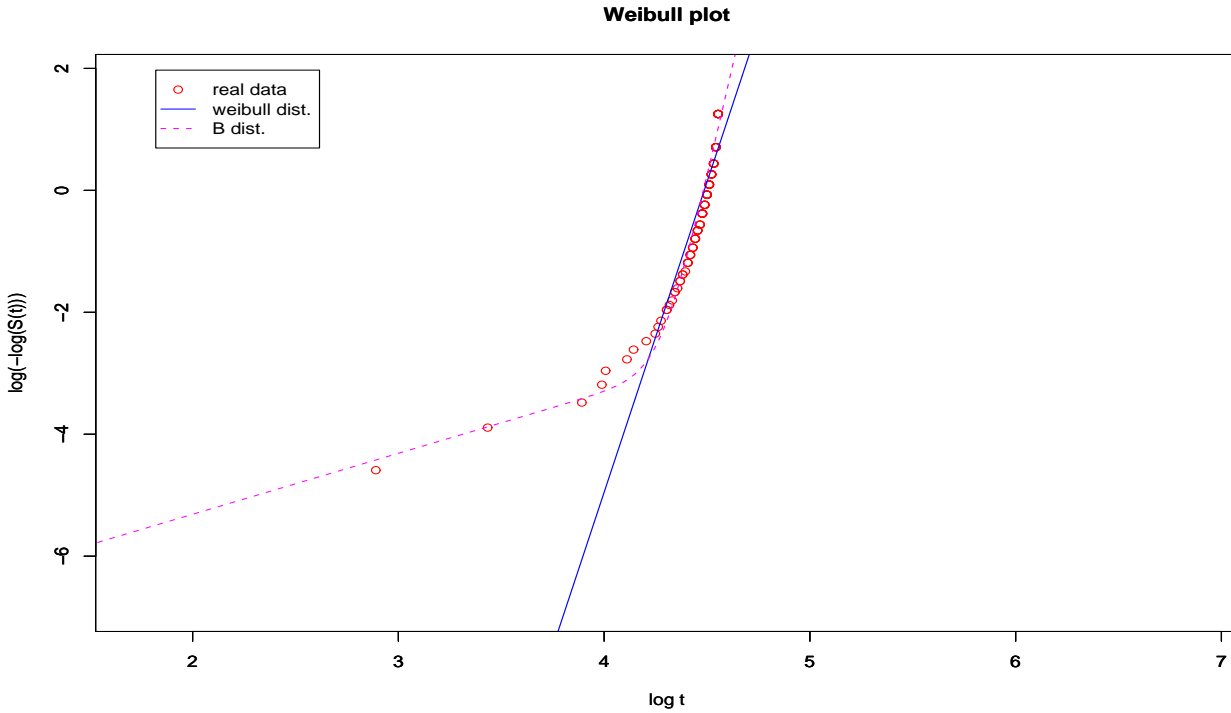


Figure A.2: Weibull plots for a 100-sized randomized sample from the prospective mortality data set and the mean estimation from Table A.6.

A.5.3 Analysis of vehicles failures

Twelve failure intervals (trimesters) $[0 - 1)$ to $[11 - 12)$ are given for 50 fleet vehicles, owing to the LTM society of transportation from Nice (France). In this application it is considered that the failures occur at the end of each interval of time. So the considered failures times are $(\underbrace{1, \dots, \dots}_{n_1}, \dots, \underbrace{12, \dots}_{n_{12}})$ where $(n_1, \dots, n_{12}) = (2, 3, 3, 2, 2, 4, 5, 4, 5, 6, 5, 9)$. Estimations are given in Table A.8 and corresponding Weibull plots are displayed on Figure A.4.

	<i>Weibull</i>		<i>B model</i>		
ML	$\hat{\eta} = 8.83$	$\hat{\beta} = 2.52$	$\hat{\eta}_0 = 16.86$	$\hat{\eta}_1 = 11.01$	$\hat{\beta} = 8.10$
PMCH	$\tilde{\eta} = 10.35$	$\tilde{\beta} = 2.41$	$\tilde{\eta}_0 = 20.75$	$\tilde{\eta}_1 = 10.50$	$\tilde{\beta} = 5.10$

Table A.8: ML and Bayesian estimations from the vehicles failures sample.

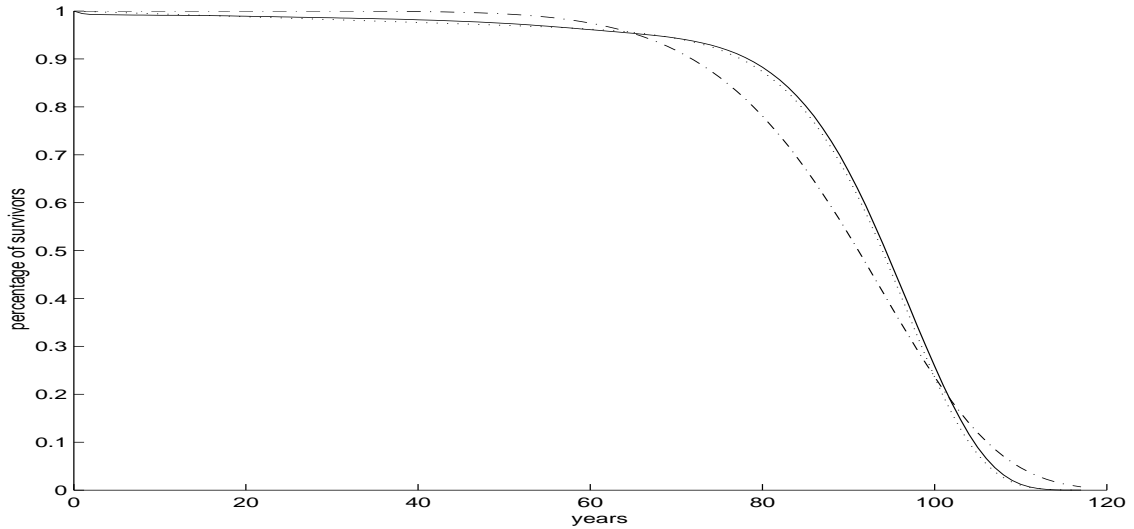


Figure A.3: Survival functions for the 100,000-sized life table [— : empirical survival function;: \mathcal{B} estimated survival function; - - -: Weibull estimated survival function].

First it can be noticed a good accordance between ML and Bayesian inference. The \mathcal{B} distribution is clearly more adapted than a simple Weibull distribution, and Bayesian estimation provides a better fit than ML estimation. Here prior domains were chosen as $[1, 6]$ for β and $[1, 30]$ for both η_0 and η_1 . The \mathcal{B} estimation shows a strong aging behavior (which seems not surprising for very used vehicles) approximatively for 60% of cases, competing with a high rate of accidents. Notice that the use of the \mathcal{B} distribution could be expected to be well-adapted to the case of vehicles failures that may be generated by technical causes (related to aging) or accidental causes.

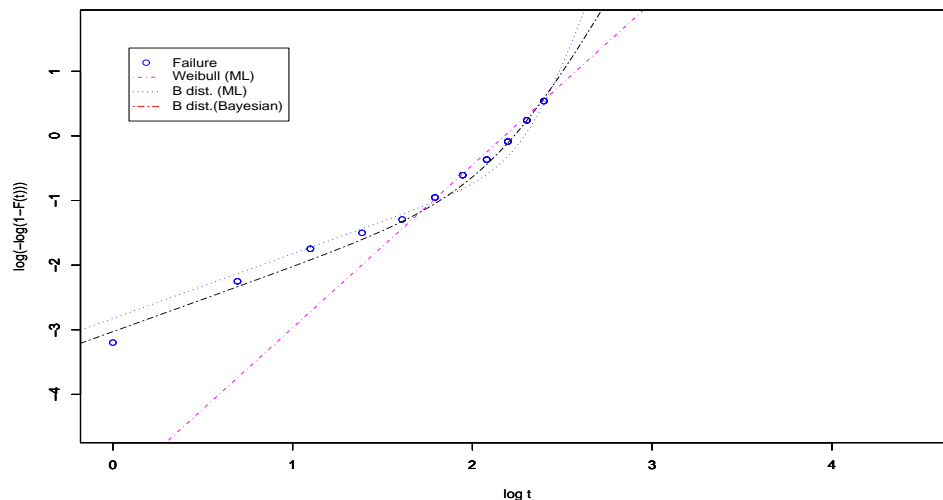


Figure A.4: Weibull probability plots of vehicle failures data.

A.6 Discussion

The \mathcal{B} distribution is a natural and simple model to analyse various competing risks of death or failure. It can be mentioned here that other models have already been proposed. One of them is the Failure-Time Model of Chan and Meeker (1999) which includes at least one more parameter. It requires thus large sample sizes to get estimations in a stable way. In the present paper the \mathcal{B} distribution has been considered as an alternative to the Weibull distribution for modelling aging. In some circumstances it can be thought of as a more realistic model to describe aging than the Weibull distribution. And a clear strategy of testing has been designed to choose between exponential, Weibull and \mathcal{B} distributions.

The versatile Weibull distribution notably appears to have difficulty in taking into account all aspects of lifetime data. An illustration is highlighted by the life table data. Figure A.3 shows that the best estimated Weibull distribution is unable to fit the empirical survival function given by the entire table. On the contrary, the \mathcal{B} distribution gives a perfect adjustment to this curve.

Nevertheless a possible drawback of this model is that it involves three parameters. Indeed it can be difficult to estimate those parameters for small-sized and/or censored samples. For this very reason, Bayesian inference is desirable to get reliable estimations. We have proposed an efficient importance sampling method to approximate the posterior estimation of the \mathcal{B} distribution parameters. One of our research perspectives is to design simple and reliable elicitation procedures for taking into account prior information regarding the \mathcal{B} distribution in a proper way.

Note that covariates could be taken into account in a simple way with the \mathcal{B} distribution. It would consist of partitioning the failures into two classes (one for exponential failures and one for Weibull failures) and to explain this classification with the covariates through a linear logistic regression.

The \mathcal{B} distribution can finally reach a level of information good enough for industrial applications and may be considered by its practical use as a reference distribution for reliability studies.

A.7 Appendix: existence of a consistent root of the likelihood equations

The aim of this section is to prove that the likelihood equations for the \mathcal{B} distribution have a root which is consistent and asymptotically normally distributed. The proof uses a Chanda's theorem that is recalled first.

THEOREM 4.(Chanda 1954). *Let $f(x|\theta)$ be a probability density function, $\theta = (\theta_1, \dots, \theta_k)$ being a vector parameter belonging to the parameter space Θ , and x_1, \dots, x_n be independent observations of a random variable X with density $f(x|\theta)$. The likelihood equations are given by $\frac{\partial \ln L}{\partial \theta} = 0$, where $\ln L = \sum_{i=1}^n \ln f(x_i|\theta)$. Let θ_0 denote the true value of θ . It is assumed that θ_0 lies at some point in Θ . Then, if Conditions 1-3 below hold, there exists a unique consistent estimator θ_n , solution of the likelihood equations. Furthermore, $\sqrt{n}(\theta_n - \theta_0)$ is asymptotically normally distributed with mean zero and covariance matrix $I_{\theta_0}^{-1}$, where I_{θ_0} is the Fisher information matrix.*

- **CONDITION 1** For almost all x and for all $\theta \in \bar{\Theta}$ $\frac{\partial \ln f}{\partial \theta_r}$, $\frac{\partial^2 \ln f}{\partial \theta_r \partial \theta_s}$ et $\frac{\partial^3 \ln f}{\partial \theta_r \partial \theta_s \partial \theta_t}$ exist for all $r, s, t = 1, \dots, k$.
- **CONDITION 2** For almost all x and for all $\theta \in \bar{\Theta}$ $\left| \frac{\partial f}{\partial \theta_r} \right| < F_r(x)$, $\left| \frac{\partial^2 f}{\partial \theta_r \partial \theta_s} \right| < F_{rs}(x)$ and $\left| \frac{\partial^3 f}{\partial \theta_r \partial \theta_s \partial \theta_t} \right| < H_{rst}(x)$, where H_{rst} is such that $\int_{-\infty}^{+\infty} H_{rst}(x) f(x) dx \leq M < \infty$, and $F_r(x)$ and $F_{rs}(x)$ are bounded for all $r, s, t = 1, \dots, k$.
- **CONDITION 3** For all $\theta \in \bar{\Theta}$ the matrix $I(\theta) = \int_{-\infty}^{+\infty} \left(\frac{\partial \ln f}{\partial \theta} \right) \left(\frac{\partial \ln f}{\partial \theta} \right)' f dx$ is positive definite.

The three conditions of the Chanda theorem are now checked for the likelihood equations of the \mathcal{B} distribution. It is to be remarked that the proof below includes the possibility of censored observations. In such a case, denoting c such a censored observation, the expression of the pdf is then

$$f(c) = \exp \left\{ -\frac{c}{\eta_0} - \left(\frac{c}{\eta_1} \right)^\beta \right\}$$

instead of (A.5). Henceforth, x may denote a censored observation in what follows.

First, the partial derivatives of the pdf f are of the following form

$$\frac{\partial \ln f}{\partial \theta_r} = \frac{\partial f}{\partial \theta_r} \frac{1}{f}$$

$$\frac{\partial^2 \ln f}{\partial \theta_r \partial \theta_s} = \frac{1}{f} \frac{\partial^2 f}{\partial \theta_r \partial \theta_s} - \frac{\partial f}{\partial \theta_r} \frac{\partial f}{\partial \theta_s} \frac{1}{f^2}$$

$$\frac{\partial^3 \ln f}{\partial \theta_r \partial \theta_s \partial \theta_t} = 2 \frac{\partial f}{\partial \theta_r} \frac{\partial f}{\partial \theta_s} \frac{\partial f}{\partial \theta_t} \frac{1}{f^3} - \frac{\partial^2 f}{\partial \theta_r \partial \theta_t} \frac{\partial f}{\partial \theta_s} \frac{1}{f^2} - \frac{\partial f}{\partial \theta_r} \frac{\partial^2 f}{\partial \theta_s \partial \theta_t} \frac{1}{f^2} - \frac{\partial f}{\partial \theta_t} \frac{\partial^2 f}{\partial \theta_r \partial \theta_s} \frac{1}{f^2} + \frac{\partial^3 f}{\partial \theta_r \partial \theta_s \partial \theta_t} \frac{1}{f^3},$$

where θ is the vector parameter $\theta = (\eta_0, \eta_1, \beta)$. By induction, it can be proved that f and its partial derivatives of any order (denoted below by g) can be written in the following way

$$\exp \left\{ -\frac{x}{\eta_0} - \left(\frac{x}{\eta_1} \right)^\beta \right\} \left[P\left(\frac{1}{\eta_0}, \frac{1}{\eta_1}, \beta\right) + \sum_{k_1=0}^{M_1} \sum_{k_2=1}^{M_2} \sum_{k_3=0}^1 Q_{k_1 k_2 k_3}\left(\frac{1}{\eta_0}, \frac{1}{\eta_1}, \beta\right) \left(\ln \left(\frac{x}{\eta_1} \right) \right)^{k_1} \left(\frac{x}{\eta_1} \right)^{k_2 \beta - k_3} \right]$$

where $P\left(\frac{1}{\eta_0}, \frac{1}{\eta_1}, \beta\right)$ and $Q_{k_1 k_2 k_3}\left(\frac{1}{\eta_0}, \frac{1}{\eta_1}, \beta\right)$ are polynomials in $\frac{1}{\eta_0}, \frac{1}{\eta_1}$ and β . Consequently, Condition 1 is satisfied.

Secondly, any partial derivative g is a continuous function in x and θ . Thus g is bounded for $\theta \in \bar{\Omega}$ and x in any closed interval. Therefore to check Condition 2, it suffices to consider its behavior for large values of x . It is easily seen that there exist positive numbers A and B such that g is inferior to $e^{-Bx} \times x^A$ for sufficiently large x and $\theta \in \bar{\Omega}$. Since $e^{-Bx} \times x^A$ is bounded, Condition 2 is satisfied. As for Condition 3, $I(\theta)$, which is a covariance matrix, is positive definite unless it exists a, b, c not all equal to zero such that $a \frac{\partial \ln f}{\partial \eta_0} + b \frac{\partial \ln f}{\partial \eta_1} + c \frac{\partial \ln f}{\partial \beta} = 0$. A simple examination of the derivatives shows straightforwardly that they are not collinear. Thus the three conditions of Chanda theorem are verified.

Appendix **B**

An industrial Bayesian analysis with Weibull distributions

This article summarizes the main points of Chapter 3. Moreover it proposes a default calibration method for the shape parameter β of the Weibull distribution. Finally, a real example is treated, considering the methodological aspects developed in Chapter 3. In the text, “Bousquet (2006)” refers to the complete thesis document.

Abstract. The context of our study is industrial reliability, where lifetime data are usually censored and in small number. Background information is available from experts. Our prior subjective knowledge is only about the lifetime of an industrial component and not about the parameters of a Weibull distribution which represents this lifetime. We propose to focus the discussion between the experts and the industrial analyst about the size of *virtual* data representing the variability of the expert opinion. Indeed, this size is one of the scarce indicators that both can understand. The prior calibration is made easy, and some methods and indicators including a default calibration method are proposed to help the Bayesian analyst (they can be extended to inferences on other distributions than Weibull). Besides, the posterior computation by importance sampling is simple and satisfying. Finally, through a real example, the flexibility of the elicitation is illustrated.

Keywords: reliability, durability, Bayesian analysis, Weibull distribution, expert opinion, subjective prior, virtual data, censored data, Kullback-Leibler.

B.1 Introduction

In a reliability context, one of the most employed lifetime distributions is the Weibull distribution (Lawless 1982, chap. 4). This versatile $\mathcal{W}(\eta, \beta)$ distribution, with density function

$$f_W(t|\eta, \beta) = \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1} \exp\left\{-\left(\frac{t}{\eta}\right)^\beta\right\}$$

and hazard rate $h_W(t|\eta, \beta) = \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1}$ where $\eta, \beta > 0$, can be used for modelling infant mortality defects when the shape parameter $\beta < 1$, aging when $\beta > 1$ or accidental failure (with a constant rate) when $\beta = 1$. In this case, it reduces to an exponential distribution with scale parameter η .

In numerous industrial studies, reliability *feedback experience lifetime* (FEL) data $\mathbf{t}_n = t_1, \dots, t_n$ are often modelled with the Weibull distribution and the parameters are estimated by maximum likelihood inference using Newton-Raphson (NR) or EM algorithms. However, the good behavior of likelihood maximization is ensured only when the sample size n is large enough and when the lifetimes are little censored (Bacha 1996). Otherwise, using Bayesian inference techniques is relevant when prior knowledge is available (Robert 2001). This is the context of our study, where some experts provided informations about the lifetime T of an industrial component \sum . The main difficulty is to translate this expert opinion into information on (η, β) , through the choice of a prior distribution with density $\pi(\eta, \beta)$. Then the posterior density

$$\pi(\eta, \beta|\mathbf{t}_n) = \frac{\mathcal{L}(\mathbf{t}_n; \eta, \beta) \pi(\eta, \beta)}{\int_{\mathbb{R}} \int_{\mathbb{R}} \mathcal{L}(\mathbf{t}_n; \eta, \beta) \pi(\eta, \beta) d\eta d\beta}$$

allows to obtain estimates of reliability functions $h(\eta, \beta)$ which are of interest for the industrial analyst, as the survival function. Here $\mathcal{L}(\mathbf{t}_n; \eta, \beta)$ denotes the likelihood of the data.

This article is focused on the choice of $\pi(\eta, \beta)$. We consider that the expert(s) (considered as male(s) for simplicity) can express his (their) beliefs only about the marginal distribution \mathcal{M} of T with density

$$m(t) = \int_{\mathbb{R}} \int_{\mathbb{R}} f_W(t|\eta, \beta) \pi(\eta, \beta) d\eta d\beta$$

and not directly about (η, β) . We take profit of this feature to build a prior for the scale parametrization of the Weibull model (see Prop. 17). Typically, he (they) can give prior estimates of the mean, the median, the mode or percentiles of \mathcal{M} . However, he (they) can have difficulties to estimate his self uncertainty; indeed, an expert is often not a statistician and notions as standard deviation remain fuzzy. To obtain good estimations of this prior uncertainty, discussions between the expert and the Bayesian analyst (who gathers the prior information, possibly weight it and then infer both the data and the prior) are desirable. Hence the need of indicators which are understandable for both of them.

After some recalls and precisions about the meaning of the Weibull parameters in Section ??, we propose in Section B.3 an improvement of the approach proposed by Berger and Sun (1993) and Bacha *et al.* (1998) to elicit prior distributions. The main point of this improvement is to consider the prior distribution as an approximation of the posterior distribution coming from a noninformative prior and *virtual data* whose size a is a relevant indicator of uncertainty.

According to the specifications of the prior knowledge, we obtain a flexible hierarchical prior family for which we propose some strategies of calibration in Section B.4. Especially, a default calibration method is suggested. Independently from the expert, the Bayesian analyst must often proceed to a final recalibration step. In this aim, an indicator measuring the effective size of the censored dataset \mathbf{t}_n is given to locate the strength of the subjective information with respect to the strength of the objective data information. Besides, bibliographical results are recalled and can be proceeded in a reliable and simple way. Then the consensus between several experts is briefly studied.

Finally, we show in Section B.5 that the computation of the posterior distribution is simple using an importance sampling algorithm. In Section B.6, a full Bayesian analysis is led for real data and two expert opinions to illustrate the main points of our elicitation method.

B.2 Data and parameters

B.2.1 The statistical context

For $n \in \mathbb{N}^*$, let $\mathbf{T}_n = T_1, \dots, T_n \sim \mathcal{W}(\eta, \beta)$ be independently and identically distributed real- or vector-valued random variables in the sample space \mathbb{R}_+^n with probability density function (pdf) $f_W(t|\eta, \beta)$ and survival (reliability) function $S_W(t|\eta, \beta)$.

The available data are as follows: let $\mathbf{t}_n = (t_1, \dots, t_n)$ be an observed sample of n data. Usually \mathbf{t}_n contains r uncensored *i.i.d.* data $\mathbf{x}_r = (x_1, \dots, x_r)$ following $\mathcal{W}(\eta, \beta)$ and $n - r$ fixed (progressive type-I) right-censored values, denoted $\mathbf{c}_{n-r} = (c_1, \dots, c_{n-r})$. Thus, the observed likelihood can be written as

$$\mathcal{L}(\mathbf{t}_n; \eta, \beta) = \frac{\beta^r}{\eta^r} \prod_{i=1}^r \left(\frac{x_i}{\eta} \right)^{\beta-1} \exp \left\{ - \sum_{j=1}^n \left(\frac{t_j}{\eta} \right)^\beta \right\}.$$

B.2.2 Meaning of Weibull parameters

The usual Weibull parameters (η, β) have different senses. The scale parameter η is the 63rd percentile of the distribution and is homogeneous to T . But the shape parameter β has no dimension.

An expert viewpoint about β appears to be *qualitative*: the value of β reflects a knowledge about the *prospective* (or future) behavior of the component Σ . This knowledge is intrinsic to the component, independently of the *quantitative* knowledge of the failure time which is translated by a knowledge on functions of (η, β) as the mean, the median, etc. This qualitative meaning of β can be highlighted as follows: when the component Σ is submitted to aging, aging rate is given by

$$\frac{\partial h_W(t|\eta, \beta)}{\partial t} = \frac{\beta(\beta-1)}{\eta^2} \left(\frac{t}{\eta} \right)^{\beta-2}$$

and aging acceleration can be measured by

$$\frac{\partial^2 h_W(t|\eta, \beta)}{\partial t^2} = \frac{\beta(\beta-1)(\beta-2)}{\eta^3} \left(\frac{t}{\eta} \right)^{\beta-3}.$$

Thus, as we said before, the exponential distribution ($\beta = 1$) can be used for modelling a component which is only submitted to accidental failure, without aging rate. Face to aging, engineers usually try

to slow down aging by preventive care (such that the estimated β remains between 1 and 2).

In our study, the available expert opinions are mainly quantitative and give little information about β . This is a significant difference with the conditions required for instance by Singpurwalla and Song (1986) and Singpurwalla (1988), who proposed hierarchical prior buildings with numerous hyperparameters. Note that Berger and Sun (1993) assumed the same hypothesis of a prominent prior knowledge on β . Here, we shall simply consider that $\text{Supp}(\pi(\beta))$ is bounded for objective reasons.

In the following, we use for convenience the parametrization $\mu = \eta^{-\beta}$. More generally, we denote θ the parameters of the model, especially when we define or use techniques which can be applied in more general settings than Weibull inference.

B.3 The prior modelling

B.3.1 The prior distribution of the shape parameter

We assume that an objective knowledge of β is available such that the prior domain can be bounded on $[\beta_l, \beta_r] \subset (0, \infty)$. Following arguments from Erto (1982), Berger and Sun (1993) and especially Bacha *et al.* (1998), this is a reasonable assumption since β directs the kinetics of aging. On mechanical systems, for physical reasons, $\beta > 5$ is never estimated on real Weibull data (except when they come from a 3-parameter Weibull distribution and must be reduced from a burn-in time). Usually, β stays in $[1, 2.5]$. See Lannoy and Procaccia (2001) for an engineering viewpoint. Moreover, when aging is assumed, we have clearly $\beta_l = 1$.

Following Bacha (1996), we propose to use a Beta $\mathcal{B}_e(p, q)$ distribution for the prior on β ,

$$\pi(\beta) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \frac{(\beta - \beta_l)^{p-1} (\beta_r - \beta)^{q-1}}{(\beta_r - \beta_l)^{p+q-1}} \mathbb{1}_{[\beta_l, \beta_r]}(\beta)$$

because of its flexibility. It can be calibrated in terms of variance and central value independently of (β_l, β_r) . Moreover, $E[\beta]$ depends only of p/q and can be chosen independently of $\text{Var}[\beta]$ (and similarly for the mode which depends only of $(p-1)/(q-1)$). Bacha *et al.* (1998) provided a methodology to calibrate $\pi(\beta)$ which lays on the specification of an observation β_m . Jenkinson (2005) gives a review of the calibration methods of a Beta prior distribution if some values of β can be built from past experiments or expert questioning. However, such methodologies remain *ad hoc*. Moreover, since experts are not statisticians, the questioning has to be indirect.

We propose the following method to obtain prior estimations of β . About the component Σ , ask to the experts the probabilities p_0 and p_1 to fall down before the times t_0 and $t_1 > t_0$, respectively. Then

$$\frac{1-p_1}{1-p_0} = \exp \left\{ -\mu x_0^\beta \left(\left\{ \frac{x_1}{x_0} \right\}^\beta - 1 \right) \right\} = (1-p_0) \left(\frac{x_1}{x_0} \right)^{\beta-1}$$

and deduce the prior estimation

$$\beta_m = \log \left\{ \frac{\log(1-p_1)}{\log(1-p_0)} \right\} \log^{-1}(x_1/x_0). \quad (\text{B.1})$$

In next subsection, we focus our study on the prior distribution of the scale parameter η . A hyperparameter a which can be interpreted as a virtual sample size is defined. When fixed, this virtual size will be used afterwards (in § B.4.1.2) to propose a calibration method of $\pi(\beta)$ when no *credible* information about the uncertainty of β is available.

B.3.2 The prior distribution of the scale parameter

B.3.2.1 Extending the approach of Berger and Sun (1993)

We extend here the approach of Berger and Sun (1993). Conditionally to β , they choose for the scale parameter η the generalized inverse gamma distribution

$$\eta|\beta \sim \mathcal{GIG}(a, b, \beta)$$

with density

$$f(\eta|a, b, \beta) = \frac{b^a \beta}{\Gamma(a)} \frac{1}{\eta^{a\beta+1}} \exp\left(-\frac{b}{\eta^\beta}\right) 1_{[0;+\infty[}(\eta) \quad (\text{B.2})$$

where $(a, b, \beta) > 0$, with moment

$$\mathbb{E}[\eta^k] = \frac{b^{k/\beta} \Gamma(a - k/\beta)}{\Gamma(a)} \quad \forall a\beta > k > 0,$$

and mode $M_a[\eta] = (b\beta/a\beta + 1)^{1/\beta}$. This family is closed by scale transformation (i.e. $X \sim \mathcal{GIG}(a, b, \beta) \Rightarrow \forall c > 0, cX \sim \mathcal{GIG}(a, bc^\beta, \beta)$) which makes it interesting to represent prior informations on a scale parameter. Moreover, if we consider the reparametrization $\mu = \eta^{-\beta}$, we obtain prior and posterior

$$\begin{aligned} \mu|\beta &\sim \mathcal{G}(a, b), \\ \mu|\beta, \mathbf{t}_n &\sim \mathcal{G}\left(a + r, b + \sum_{i=1}^n t_i\right). \end{aligned}$$

Thus the posterior computation (see § B.5) is made easier. However, Berger and Sun (1993) do not provide a meaning to the hyperparameters a and b . This choice remains controversial since $\pi(\mu|\beta) = \pi(\mu)$. Then we propose an alternative choice.

Suppose that the quantitative expert opinion can be represented by a *virtual* Weibull sample $\tilde{\mathbf{x}}_m = (\tilde{x}_1, \dots, \tilde{x}_m)$. Let $\pi^J(\eta|\beta) \propto \eta^{-1}$ be the conditional Jeffreys prior. If our conditional prior on η can be represented by the posterior $\pi^J(\eta|\beta, \tilde{\mathbf{x}}_m)$, we obtain

$$\begin{cases} a &= m, \\ b &= b(a, \beta) = \sum_{i=1}^m \tilde{x}_i^\beta. \end{cases}$$

Thus the choice of b as a function of β appears to be natural. Besides, a takes the meaning of the size of a virtual sample yielding the same information as the expert opinion. Hence a is now a *calibration* hyperparameter since the Bayesian analyst can modulate the strength of the prior *quantitative* opinion through this simple hyperparameter. This sample size is an easy interpretable parameter and a good focus point for a discussion between the analyst and the expert. We propose some default calibration values for a and questioning suggestions in § B.4.

Ideally, $\pi(\mu|\beta)$ has a natural hierarchical structure. Assuming $\tilde{\mathbf{x}}_{\mathbf{m}} \stackrel{i.i.d.}{\sim} \mathcal{W}(\eta_e, \beta_e)$ where (η_e, β_e) are prior estimations, we get

$$\begin{aligned} b|\beta &\sim \mathcal{G}(a, \eta_e^{-\beta}), \\ \mu|\beta, b &\sim \mathcal{G}(a, b). \end{aligned}$$

However, for simplicity reasons we prefer to choose a determinist expression of b in function of the available prior specifications. This is done in the next paragraph.

B.3.2.2 Translating the quantitative expert opinion

The expert opinion is *quantitative* when he talks in terms of lifetime. Many authors in Bayesian litterature use expert opinions on the values of the parameter vector (see Bacha 1996, Lijoi 2003 or Wisse *et al.* 2005). It can be understandable when it concerns the exponential model, since the mean lifetime is the parameter. But usually the expert does not know the statistical model and the prior information on parameters corresponds more to a “transformation” of the expert knowledge by the Bayesian analyst.

Thus, Sinpurwalla and Song (1986) then Sinpugpurwalla (1988) considered that an expert was able to speak about the median lifetime $m = \eta(\log 2)^{1/\beta}$: he is supposed to give an estimation of the prior mean $E[m]$. Fixing $\pi(\beta)$, they obtain a complete prior on both parameters. In our industrial context, to be more general, we consider that an expert can give an estimation of a quantity characterizing the *marginal distribution* \mathcal{M} of lifetime T with density

$$m_W(t) = \int_{\mathbb{R}} \int_{\mathbb{R}} f_W(t|\mu, \beta) \pi(\mu|\beta) \pi(\beta) d\mu d\beta. \quad (\text{B.3})$$

An intensive litterature deals with the questioning of experts (see Daneskhah 2004 for a review). Some discussion techniques as the *bisection* method (Garthwaite *et al.* 2005) have been proposed with success to obtain from experts quantitative information about the behavior of a studied system. In our case, two questions are essential : “Can you give a representative value t_e of the lifetime T of component Σ ?” and “What is the probability α for the component Σ to fail before time t_e ?” (the context of both questions differ : the first one concerns *reliability* and the second one deals with *durability* ; see for instance Lawless 2000). Thus, we assume that the expert can be solicited to give a lifetime value t_e and the *specification* of t_e with respect to \mathcal{M} . This value can be perceived as the estimation

1. of the α order percentile, namely $P(T < t_e) = \alpha \in]0, 1[$;
2. of expectation $E[T]$;
3. of mode $M_d[T]$ (however, several private discussions with reliability specialists conclude to the very weak probability that an industrial expert can really specify a mode).

Through discussion, specifying precisely the nature of an expert opinion is essential. For instance, some authors have noticed that the experts tend to give a median value although they understand t_e as a mean (Schieren 1993, Lannoy and Procaccia 2001). The specification can be reinforced by some open questions as

1. When the age of a group of components reaches t_e , which proportion $1 - \alpha$ is still in use ?

2. Is the value t_e a mean of past events, taking into account extreme values or not ?
3. Is the value t_e the failure time which has been observed the most frequently in the past ?

Then, conditionally to the choice of a , the choice of b is given in the following proposition from the knowledge of the marginal distribution \mathcal{M} .

PROPOSITION 17. *Let $t_e \in \mathbb{R}^{+*}$, $a > 0$ and $\alpha \in]0, 1]$. Denote*

$$\begin{aligned} b_1(a, \beta) &= \left((1 - \alpha)^{-1/a} - 1 \right)^{-1} t_e^\beta, \\ b_2(a, \beta) &= \left(\frac{\Gamma(a)}{\Gamma(1 + 1/\beta)\Gamma(a - 1/\beta)} \right)^\beta t_e^\beta, \\ b_3(a, \beta) &= \frac{a\beta + 1}{\beta - 1} t_e^\beta. \end{aligned}$$

Denote B_i , $i = 1, 2, 3$ the induced prior modellings. Then, for all choice of $\pi(\beta)$ such that $\text{Supp}(\pi(\beta)) \subset [\beta_l, \beta_r]$,

- (i) $P(T \leq t_e | B_1) = \alpha$,
- (ii) $E[T | B_2] = t_e$ if $a > \beta_l^{-1}$,
- (iii) $M_d[T | B_3] = t_e$ if $\beta_l > 1$.

where the expectation is taken with respect to the marginal distribution \mathcal{M} with density (B.3).

Proof. See Appendix .1.

Notice that an expert can give more than one specification. For instance, he can give a *credibility interval* $[t_{e1}, t_{e2}]$ and probabilities (α_1, α_2) such that $P(T \leq t_{ei}) = \alpha_i$ for $i = 1, 2$. We propose to reduce any expert opinion to “discrete” specifications $i = 1, \dots, P$ then to select couples (a_i, b_i) . Finally, we sum all specifications through the prior

$$\mu | \beta \sim \mathcal{G} \left(\sum_{i=1}^P P \gamma_i a_i, \sum_{i=1}^P b_i (P \gamma_i a_i, \beta) \right) \tag{B.4}$$

which is the posterior coming from successive Bayesian inferences on virtual samples. Weights $\gamma_i > 0$ (such that $\sum_{i=1}^P \gamma_i = 1$) are fixed in function of the relative trust in the specifications. Later in the article, an application will exemplify (and clarify) this choice. It is reinforced by the convexity properties of applications $\beta \mapsto b(a, \beta)$. Especially, it appears obviously that $b_1(a, \beta)$ (thus any convex sum of several $b_1(a, \beta)$) “mimics” the geometrical behavior of $\beta \mapsto \sum_{i=1}^a \tilde{x}_i^\beta$. It needs some weighty calculations (not given here) but can easily been numerically checked.

In next figures we display some densities $m(t)$ of the prior marginal distribution \mathcal{M} , where $t_e = 100$, $[\beta_l, \beta_r] = [1.1, 5]$, $(p, q) = (1.5, 1.5)$ and $a = 2$. Figure B.1 shows $m(t)$ when t_e is successively considered as the prior median/mean/mode. Figure B.2 shows the evolution of $m(t)$ when t_e is the α order percentile, for several values of α . The correlation between parameters induced by the prior modelling is revealed by the *convex hull* of the prior sampling ; in Figure B.3, 95% of prior simulated (η, β) values are used for delimiting the convex hull. The form of the correlation remains near the form of the theoretical joint confidence area showed by Wu (2002) ; the inverse distribution of parameter values which is usually perceived is coherent with the classical behavior of frequentist estimations of the parameters (using unknown data): if β is overestimated then η is underestimated, and conversely.

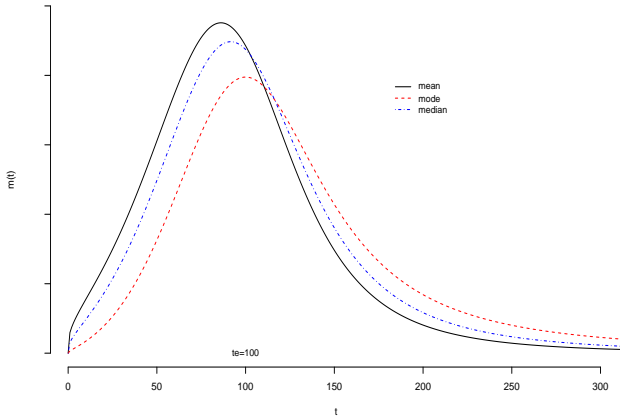


Figure B.1: Densities $m(t)$ of the prior marginal distribution \mathcal{M} (indexed by the specification of $t_e = 100$).

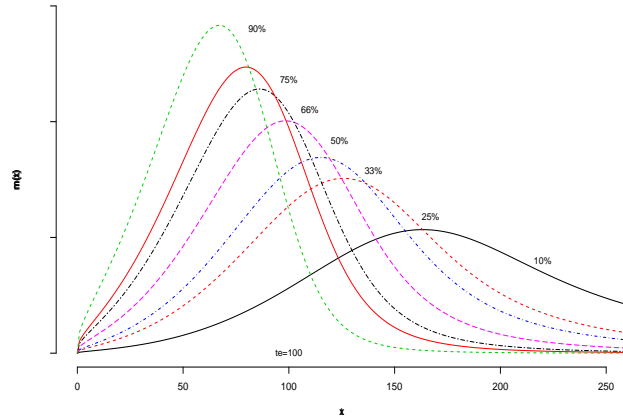


Figure B.2: Densities $m(t)$ of the prior marginal distribution \mathcal{M} (indexed by the specification of α).

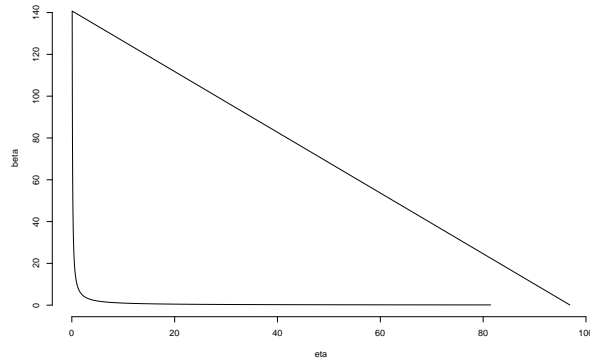


Figure B.3: Convex hull of a prior sampling of parameters (η, β) .

B.4 Prior calibration

Some authors as Singpurwalla and Song (1986) build prior modellings with numerous hyperparameters, which allow to modify the expert information in location and uncertainty. As Lindley and Singpurwalla (1986) noticed, it has been observed that experts tend to produce “location and scale bias”. In this work, our opinion is that the full expert knowledge is given through t_e (and α if t_e is a percentile value). No objective criterion allows the Bayesian analyst to modify this expert knowledge. The only possibility is to modulate the expert uncertainty through the choice of a . In our sense, three steps of calibration should be considered.

1. A *default calibration* step; according to the number and the nature of specifications, we propose some default values for a . This is done in § B.4.1. Besides, when a has been chosen, we propose in § B.4.1.2 a default method for calibrating $\pi(\beta)$ when no information is available about the uncertainty of β .

2. A *combined calibration* step; because a can be understood as a virtual sample size, it is a good parameter for a discussion between the expert and the Bayesian analyst. First values of a proposed by the default calibration step can be high when the expert is very precise ; the intent is to make the expert react to his self recklessness and correct a .
3. A *final recalibration* step; the Bayesian analyst can have some prior knowledge about the reliability of the expert opinion. A huge litterature is dedicated to this analysis (see Cooke and Goossens 2001 for a review). This analysis can be based on other subjective beliefs (which are not explored here) and the comparison with objective indicators. Here, if $a < n$ and if the data \mathbf{t}_n are uncensored, the posterior distribution is ensured to be more dependent from the objective data knowledge. In § B.4.2, a similar upper bound for a is defined when the data \mathbf{t}_n are censored. Besides, some bibliographical results are given to moderate the expert opinions and the consensus of several expert opinions is studied.

B.4.1 Default prior calibration

B.4.1.1 Default calibration of the scale parameter

This calibration can be separated according to the kind of specification. The value of a that we proposed is the minimal number a_{min} of past virtual data that seem necessary to assess this specification.

Mean, median or mode ? A confusion between those three specification is easy (especially the two first ones). Therefore a default choice must be done. We suggest to choose $a_{min} = 1$ for the mean and $a_{min} = 2$ for the median. Specifying the mode requires intuitively more past virtual data (and the certain knowledge of aging). We suggest $a_{min} = 3$ since it is the minimal number of data to define the modal class of a continuous distribution. In Table B.1, we give estimations of the standard deviation σ and the *skewness* γ of \mathcal{M} , using the prior choices done for displaying Figures B.1 and B.2, using several values of a . The mode specification appears as giving heavier tails and favorizing higher values of X . We recommend to use the median specification by default since it ensures a larger uncertainty than the mean specification.

a	mean		mode		median	
	σ	γ	σ	γ	σ	γ
2	67.4	4.5	360.7	31.3	94.2	5.5
3	53.8	2.4	185.0	12.5	66.0	2.7
4	50.0	1.91	151.6	9.8	58.2	1.94
5	48.1	1.45	136.7	7.5	55.8	1.46
10	44.0	1.26	110.7	6.3	50.3	1.13

Table B.1: Estimations of standard deviation σ and skewness γ of distribution \mathcal{M} .

Quantile specification. Our working hypothesis is that the expert has perceived as much past virtual data as it is necessary to obtain the precision α when he specifies t_e as a percentile. For instance, the Bayesian analyst can propose $a = 10$ when $\alpha = 10\%$. This is similar to the numerous *histogram* methods that have been developed for the asking of experts. Thus, Van Noortwijk *et al.* (1992) propose to segment \mathbb{R}_+ in separate intervals, choosing boundaries near to real observed data (or censored data). Since those lifetimes can be representative for the expert (even he is supposed not

to know the data), he can be asked about his probability of failure. Next example illustrates our way of thinking.

EXAMPLE 20. *Let $(50, 80, 90, 170)$ be some (possibly censored) observed failure times. Divide the lifetime scale in $D_1 = [0, 90)$ and $D_2 = [90, \infty)$. Let (n_1, n_2) be the numbers of virtual data in D_1 and D_2 , respectively. Suppose that the expert gives 66% chance of breaking down in D_1 . Then $n_1 = 2n_2$. Divide now D_1 in $D_{1,a} = [0, 50)$ and $D_{1,b} = [50, 90)$. Indicate to the expert that the effective lifetime is contained in D_1 . Suppose that he indicates 25% chance of breaking down in $D_{1,b}$; to a similar question on a partition of D_2 , $D_{1,a}$ or $D_{1,b}$, suppose that he is unable to answer. He is supposed to have “perceived” at least $n_1 = 4$ virtual data. Finally, he can be given $n_1 + n_2 = 6$ virtual data. ■*

Credibility interval specification. Suppose to obtain from an expert the interval $[t_{e1}, t_{e2}]$ and probabilities (α_1, α_2) such that $P(T \leq t_{ei}) = \alpha_i$ for $i = 1, 2$. Adding two percentile specifications needs to normalize a_1, a_2 such that $a = \gamma_1 a_1 + \gamma_2 a_2$ is the maximal size attainable by the separate specifications. Thus, specifications (50%, 90%) will give the same default size $a = 10$ than specifications (10%, 90%). Moreover, if the prior domain for β has been elicited, the *coherence* of the credibility interval with the Weibull model must be checked using formula (B.1). If

$$\beta_e = \log \left\{ \frac{\log(1 - \alpha_2)}{\log(1 - \alpha_1)} \right\} \log^{-1}(t_{e2}/t_{e1}) \quad (\text{B.5})$$

is not in $[\beta_l, \beta_r]$ the orders should be modified such that $\beta_e = \beta_0$ (for instance the middle or a bound of $[\beta_l, \beta_r]$). Solving $(?) = \beta_0$ with a Newton-Raphson algorithm is simple (see Bousquet 2006 for details). Orders are weighted with the constant convergence rate $\gamma_1/(1 - \gamma_1)$. Thus, if $\gamma_1 \rightarrow 1$, α_1 remains stable.

Algorithm 8. Weighting the credibility α .

1. Let $0 < \alpha_1^0 < \alpha_2^0 < 1$ and $\alpha_0 = (\alpha_1^0, \alpha_2^0)$. Denote $\ell_0 = (x_{e,2}/x_{e,1})^{\beta_0}$, $\ell(\alpha) = \frac{\log(1 - \alpha_2)}{\log(1 - \alpha_1)}$ and choose a precision ε . Fix $0 < \rho \ll 1$.

2. Step $k = 0, \dots, K$:

- compute the vector $\delta_k = \begin{pmatrix} (1 - \alpha_1^k) \log(1 - \alpha_1^k) (\ell(\alpha_k) - \ell_0) / \ell(\alpha_k) \\ -(1 - \alpha_2^k) \log(1 - \alpha_1^k) (\ell(\alpha_k) - \ell_0) \end{pmatrix}$;
- compute $h_k = -\frac{(1 - \alpha_1^k)}{(1 - \alpha_2^k) \ell(\alpha_k)}$.
- compute $\alpha^{k+1} = \begin{cases} \alpha_1^k - \rho h_k (\alpha_2^k - \alpha_1^k), \\ \alpha_2^k - \rho h_k (\alpha_2^k - \alpha_1^k) \frac{\gamma_1}{1 - \gamma_1} \end{cases}$
- stop when $\|\delta_k\| \leq \varepsilon$

EXAMPLE 21. *Choose $\beta_0 = 3$ and $(t_{e,1}, t_{e,2}) = (200, 300)$. Fixing $(\alpha_1^0, \alpha_2^0) = (0.05, 0.95)$, we obtain $\beta_e \simeq 10.03$. The expert opinion induces an unrealistic parameter shape. By default, fix $\gamma = (1/2, 1/2)$. Then we obtain $\alpha = (0.3, 0.7)$. If we choose now $\alpha_1^0 = 0.25$, we obtain $\beta_e \simeq 5.78$. The induced aging remains unreasonable. With equal weights, we find $\alpha = (0.4, 0.8)$. Assuming much more credibility in*

the lower bound, we fix $\gamma = (0.95, 0.05)$. Then $\alpha = (0.26, 0.64)$. Conversely, with $\gamma = (0.05, 0.95)$, we obtain $\alpha = (0.55, 0.93)$. ■

B.4.1.2 Default calibration of the shape parameter

Once a has been chosen, it can be used to calibrate $\pi(\beta)$ if no real information about the uncertainty of β is available. Ideally, since we see $\pi(\eta|\beta)$ as a reference posterior density with respect to virtual data $\tilde{\mathbf{x}}_{\mathbf{a}}$, $\pi(\beta)$ should be chosen as

$$\pi^J(\beta|\tilde{\mathbf{x}}_{\mathbf{a}}) \propto \beta^{a-1} \frac{\left(\prod_{i=1}^a \tilde{x}_i\right)^\beta}{\left(\sum_{j=1}^a \tilde{x}_j^\beta\right)^a}, \tag{B.6}$$

with $\pi^J(\beta)$ the noninformative reference prior. Because this prior distribution is not tractable (it is well known that the Weibull distribution does not admit any continuous conjugate, cf. Soland 1969), we made an arbitrary choice of $\pi(\beta)$. But the Beta prior $\pi(\beta|p, q)$ can be elicited as the minimizer of the relative entropy

$$\int_{\beta_l}^{\beta_r} \pi^J(\beta|\tilde{\mathbf{x}}_{\mathbf{a}}) \log \frac{\pi^J(\beta|\tilde{\mathbf{x}}_{\mathbf{a}})}{\pi(\beta|p, q)} d\beta \tag{B.7}$$

under a constraint on the mean or the mode, which is tractable using (B.1). However, the values of the virtual sample $\tilde{\mathbf{x}}_{\mathbf{a}}$ remain unknown. Therefore we propose to sample virtual samples from the $\mathcal{W}(\eta_e, \beta_e)$ distribution, where η_e and β_e are prior estimations. Thus we replace $\pi^J(\beta|\tilde{\mathbf{x}}_{\mathbf{a}})$ in (B.7) by the *expected posterior prior*

$$\pi_a^J(\beta) = \int_{\mathbb{R}_+^a} \pi^J(\beta|\tilde{\mathbf{x}}_{\mathbf{a}}) \prod_{i=1}^a f_W(\tilde{x}_i|\eta_e, \beta_e) d\tilde{x}_1 \dots d\tilde{x}_a.$$

Then minimizing (B.7) in (p, q) is similar to minimize

$$\log \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} + (p+q) \log(\beta_r - \beta_l) - (p+q) \mathbb{E} \left[\log \frac{\beta_r - \beta}{\beta - \beta_l} \right] \tag{B.8}$$

where the mean $\mathbb{E}[\cdot]$ is with respect to $\pi_a^J(\beta)$. When the prior mean (or mode) of $\pi(\beta)$ is chosen, the solution is unique. This minimization can be done using Monte Carlo estimations of the extreme right term of (B.8). In our applications, this minimization gave promising results. But it needs more work to ensure that the Kullback-Leibler projection does not lead to an overestimation of the prior information.

B.4.2 Recalibration and consensus

B.4.2.1 Comparing subjective and objective knowledge

In this subsection we give the definition of an indicator \tilde{n} which measures the effective size of the available censored data $\mathbf{t}_{\mathbf{n}}$. The virtual size a should be compared to \tilde{n} to locate the subjective information with respect to the objective data information. A similar indicator can be easily defined for other models than the Weibull distribution.

Denote \mathbf{x}_m an *i.i.d.* Weibull sample of size m yielding the same data information of \mathbf{t}_n . Formally, denoting

$$\mathcal{D}^{J_c, J}(\mathbf{t}_n, \mathbf{x}_m) = \int_{\Theta} \pi_c^J(\theta | \mathbf{t}_n) \log \frac{\pi_c^J(\theta | \mathbf{t}_n)}{\pi_c^J(\theta | \mathbf{x}_m)} d\theta \quad (\text{B.9})$$

the Kullback-Leibler divergence between two reference posterior distributions, we define our indicator of effective size by

$$\tilde{n} = \arg \min_m E_{X_m} \mathcal{D}^{J_c, J}(\mathbf{t}_n, \mathbf{x}_m). \quad (\text{B.10})$$

From Lin *et al.* (2006, Prop. 1), the existence and unicity of \tilde{n} is ensured. In (B.9), π^J and π_c^J are noninformative priors with good frequentist coverage of the posterior Bayes intervals for \mathbf{x}_m and \mathbf{t}_n , respectively. Indicator \tilde{n} is of interest since it improves the conservative choice r (the number of uncensored data) which underestimates the effective data information. Especially, when all available data are censored ($r = 0$) we have $\tilde{n} > 0$. Of course, when \mathbf{t}_n is uncensored, we have $\tilde{n} = n$.

Usually π^J and π_c^J are said *coverage matching priors*; see Datta (1996) and Ghosal (1999) for reviews and precisions. The definition follows: denoting $\theta_n(\alpha)$ the posterior α -quantile of θ based on observations \mathbf{t}_n (i.e. $P_c^J(\theta \leq \theta_n(\alpha) | \mathbf{t}_n) = \alpha$), it means that

$$P_{\theta}(\theta \leq \theta_n(\alpha)) = P_c^J(\theta \leq \theta_n(\alpha) | \mathbf{t}_n) + \mathcal{O}(n^{-i/2})$$

where the left-hand side is the frequentist probability, where θ is fixed in Θ and $\theta_n(\alpha)$ is random, and i is (ideally) the highest attainable value (the coverage order). In our applications, π^J was chosen as the *reference prior* (Berger and Bernardo 1992) which is at least of order two according to Sun (1993). Since \mathbf{t}_n is censored, we chose π_c^J as the special Jeffreys prior (including censored times) which can be derived from De Santis *et al.* (2001). The authors show it has better posterior coverage than the standard Jeffreys prior. From Bousquet (2006, Chapter 3), it is defined as follows, when \mathbf{t}_n contains $n - r$ right-censored fixed values c_1, \dots, c_{n-r} .

Let γ be the Euler constant ($\gamma \simeq 0.57722$). Let $\gamma_1 = \pi^2/6 + \gamma^2 - 2\gamma > 0$ and $\gamma_2 = -2(1 - \gamma)$. Denote

$$\begin{aligned} \tilde{\delta}(\mu, \beta) &= \delta^2(\mu, \beta) + [\delta(\mu, \beta) - 1] (\gamma_1 + \gamma_2 \log \mu + \log^2 \mu) + \pi^2/6 - 1, \\ \delta(\mu, \beta) &= n - \sum_{i=1}^{n-r} \exp(-\mu c_i^{\beta}). \end{aligned}$$

Then the special Jeffreys prior for the Weibull parametrization (μ, β) is

$$\pi_c^J(\mu, \beta) \propto (\mu\beta)^{-1} \sqrt{\tilde{\delta}(\mu, \beta)}.$$

B.4.2.2 Percentile orders correction

Suppose that the discussion between the expert and the analyst came to an agreement about the order α of a percentile value t_e . Then, independently from the expert, the analyst can correct this order using the results averaged by numerous authors about the real sense of the percentile prior estimations. Face to the results of various experiments in reliability, those estimations are given credibility

orders that usually are very overestimated. Illustrated by Meyer and Booker (1987) and Lannoy and Procaccia (2001), a tacit rule is summarized in next table, modifying the order of percentile then giving us upper bounds a^* for a .

expert opinion	reality	a^*
5%	25%	4
20%	33%	3
25%	40%	2
75%	60%	2
80%	66%	3
95%	75%	4

Table B.2: Tacit reduction of percentile order between expert opinion and reality.

B.4.2.3 Consensus between experts

When priors $\pi_1(\theta), \dots, \pi_M(\theta)$ are modelling several available (independent) expert opinions, the *convex weighted combination*

$$\pi(\theta) = \frac{\prod_{i=1}^M \pi_i^{\varpi_i}(\theta)}{\int_{\Theta} \prod_{i=1}^M \pi_i^{\varphi_i}(\theta) d\theta},$$

where $\sum_{i=1}^M \varphi_i = 1$, minimizes the Kullback-Leibler information loss

$$KL(\pi; \pi_1, \dots, \pi_M | \beta) = \sum_{i=1}^M \varpi_i \int_{\mathbb{R}} \pi(\theta) \log \frac{\pi(\theta)}{\pi_i(\theta)} d\theta$$

and carries out an optimal *consensus* of the opinions. See Liisberg (1991) ou Alturazza *et al.* (2004) for more precisions about this elicitation. The relative importance of the experts is judged through the choices of weights ϖ_i . Cooke *et al.* (1988) or Budescu and Rantilla (2000) among others propose several criteria, like the past error rate, to fix the weights through methodologies whose most famous is probably the Delphi method (see Linstone and Turoff 2002 for a review). Fortunately we obtain

$$\begin{aligned} \mu | \beta &\sim \mathcal{G} \left(\sum_{i=1}^M \varpi_i a_i, \sum_{i=1}^M \varpi_i b_i(a_i, \beta) \right), \\ \beta &\sim \mathcal{B}_e \left(\sum_{i=1}^M \varpi_i p_i, \sum_{i=1}^M \varpi_i q_i \right) \end{aligned}$$

when $\pi_i(\beta)$ is a Beta density $\mathcal{B}_e(p_i, q_i)$ defined on a common domain $[\beta_l, \beta_r]$. The consensus distribution on μ appears as the posterior distribution coming from a consensus virtual sample whose size is the weighted sum of all sizes. If experts can not be considered as independent (usually when $M > 2$), correlations must be add to the modelling. Face to this issue, O'Hagan (2003, 2005) gives numerous arguments to define a consensus (by discussion means) before the modelling. But we obtain the same kind of prior modelling.

B.5 Posterior computation

Berger and Sun (1993) provided steps of Gibbs sampling when $\pi(\beta)$ was log-concave. Our choice could be similar. Alternatively, since $\text{Supp}(\pi(\beta))$ is bounded, it is easy to use *importance sampling* to estimate the posterior mean of a function of interest $h(\theta)$

$$\begin{aligned} I &= \int_{\Theta} h(\theta) \pi(\theta | \mathbf{t}_n) d\theta, \\ &= \int_{\Theta} h(\theta) \frac{\pi(\theta | \mathbf{t}_n)}{\rho(\theta)} \rho(\theta) d\theta \end{aligned}$$

where $\rho(\theta)$ is any function such that $\text{Supp}(\pi(\theta | \mathbf{t}_n)) \subset \text{Supp}(\rho(\theta))$ (Robert and Casella 2004). Choosing ρ as a density, it is easy to estimate I by

$$\hat{I}_M = \sum_{i=1}^M \omega_i h(\theta_i)$$

where the θ_i are simulated from $\rho(\theta)$ and the weights $\omega_i = \omega'_i / \sum_{j=1}^M \omega'_j$ with

$$\omega'_i = \frac{\pi(\theta_i) \mathcal{L}(\mathbf{t}_n; \theta_i)}{\rho(\theta_i)}.$$

Under mild conditions, a limit central theorem ensures the convergence of \hat{I}_M to I when $M \rightarrow \infty$. The main difficulty is that the tails of π must be heavier than the tails of $\pi(\theta | \mathbf{t}_n)$ (and ideally, $\rho(\theta)$ should be close to $\pi(\theta | \mathbf{t}_n)$). In our case, with $\theta = (\mu, \beta)$, we obtain

$$\begin{aligned} \mu | \beta, B, \mathbf{t}_n &\sim \mathcal{G} \left(a + r, b(a, \beta) + \sum_{i=1}^n t_i^\beta \right), \\ \pi(\beta | B, \mathbf{t}_n) &\propto \beta^r \left(\prod_{j=1}^r x_j \right)^\beta (\beta - \beta_l)^{p-1} (\beta_r - \beta)^{q-1} b^a(a, \beta) \left\{ b(a, \beta) + \sum_{i=1}^n t_i^\beta \right\}^{-(a+r)} \mathbb{1}_{[\beta_l, \beta_r]}(\beta). \end{aligned}$$

Thus, choosing $\rho(\mu, \beta) = \pi(\mu | B, \beta, \mathbf{t}_n) \mathbb{1}_{\{\beta_l \leq \beta \leq \beta_r\}} / (\beta_r - \beta_l)$, we respect the conditions of a satisfying importance sampling and obtain the theoretical (unnormalized) weight

$$\omega'(\beta | B, \mathbf{t}_n) = \beta^r \left(\prod_{i=1}^r x_i \right)^\beta (\beta - \beta_l)^{p-1} (\beta_r - \beta)^{q-1} \frac{b^a(a, \beta)}{\left(b(a, \beta) + \sum_{i=1}^n t_i^\beta \right)^{a+r}} \mathbb{1}_{\{\beta_l \leq \beta \leq \beta_r\}}.$$

Especially, an industrial reliabilist is interested in the computation of the following *predictive* quantities:

1. the *mean lifetime*

$$\mathbb{E}[T | B, \mathbf{t}_n] = \int_{\mathcal{R}} \Gamma(1 + 1/\beta) \left\{ b(a, \beta) + \sum_{i=1}^n t_i^\beta \right\}^{1/\beta} \frac{\Gamma(a + r - 1/\beta)}{\Gamma(a + r)} \pi(\beta | B, \mathbf{t}_n) d\beta;$$

2. the *survival at time t_0*

$$\begin{aligned} S(t_0 | B, \mathbf{t}_n) &= P(T > t_0 | B, \mathbf{t}_n), \\ &= \int_{\mathcal{R}} \left(1 + \frac{t_0^\beta}{\left\{ b(a, \beta) + \sum_{i=1}^n t_i^\beta \right\}} \right)^{-(a+r)} \pi(\beta | B, \mathbf{t}_n) d\beta; \end{aligned}$$

3. the *residual lifetime after time* t_0 (see for instance Finkelstein 2006)

$$\begin{aligned} \text{MRTF}(t_0|B, \mathbf{t}_n) &= E[T - t_0|T > t_0, B, \mathbf{t}_n], \\ &= \frac{1}{S(t_0|B, \mathbf{t}_n)} \int_{t_0}^{\infty} S(x|B, \mathbf{t}_n) dx. \end{aligned}$$

Thus Monte Carlo estimations of these quantities need simply posterior sampling of β , or uniform importance sampling on (β_l, β_r) .

B.6 A numerical example

We consider the right-censored real lifetime data \mathbf{t}_n ($n = 18$) from Table B.3. They correspond to failure times or stopping times collected on some similar devices \sum belonging to the secondary water circuit of nuclear plants. Lifetimes are given in months. For physical reasons and according to a large consensus, those data are assumed to arise from a Weibull distribution $\mathcal{W}(\eta, \beta)$. The maximum likelihood estimator (MLE) is $(\hat{\eta}_n, \hat{\beta}_n) = (140.8, 4.51)$ with estimated standard deviations $\hat{\sigma}_n = (7.3, 1.8)$. The high value of $\hat{\beta}_n$ is unexpected and suggests a Bayesian estimation.

real failure times:	134.9, 152.1, 133.7, 114.8, 110.0, 129.0, 78.7, 72.8, 132.2, 91.8
right-censored times :	70.0, 159.5, 98.5, 167.2, 66.8, 95.3, 80.9, 83.2

Table B.3: Lifetimes (months) of nuclear components (from secondary water circuits).

Two prior opinions on the lifetime are available on device \sum , given by independent experts \mathcal{E}_1 and \mathcal{E}_2 . They are summarized in Table B.4. The \mathcal{E}_1 opinion is much more informative than \mathcal{E}_2 and both are right-shifted with respect to the data. Moreover the experts are not asked at the same level of precision. \mathcal{E}_1 is a nuclear operator and speaks for a particular component while \mathcal{E}_2 can be seen as a component producer whose opinion takes into account a variety of running conditions. Thus the expert opinions can be considered as independent.

	credibility intervals (5%,95%)	median value
expert \mathcal{E}_1	[200,300]	250
expert \mathcal{E}_2	[100,500]	250

Table B.4: Expert opinions about the lifetime T .

Aging is assumed: we choose $\beta_l = 1$. For technical reasons we choose $\beta_r = 5$. Using (B.1), the underlying prior estimates of β proposed by the experts take values in $\{8.02, 10.0, 11.7\}$ (expert \mathcal{E}_1) and $\{2.11, 2.53, 2.84\}$ (expert \mathcal{E}_2). Thus, the first expert opinion seems dubious since it induces an unreasonable aging. Note that when we replace orders (5%, 95%) by (33%, 66%) (found using Algorithm 1) we obtain prior values of β near to 2.5. For this reason, we prefer these corrected orders. Then, for each prior density $\pi(\beta|(p, q)_{\mathcal{E}_i})$ ($i = 1, 2$), we specify for the *mode* the same value $\beta^* = 2.5$. Indeed, we necessarily have $(p, q)_{\mathcal{E}_i} > 1$ which allows not to obtain flat priors. This is the starting point of the elicitation method in Bacha *et al.* (1998).

Now let (a_1, a_2) be the virtual size for each expert. A default prior for the combination of both expert opinions is

$$\begin{aligned}\mu|\beta &\sim \mathcal{G}(a, b(\beta)), \\ \beta &\sim \mathcal{B}_e(p, q)\end{aligned}$$

with

$$\begin{cases} a &= \varpi_1 a_{\mathcal{E}_1} + (1 - \varpi_1) a_{\mathcal{E}_2}, \\ b(\beta) &= \varpi_1 b_{\mathcal{E}_1}(\beta) + (1 - \varpi_1) b_{\mathcal{E}_2}(\beta), \\ p &= \varpi_1 p_{\mathcal{E}_1} + (1 - \varpi_1) p_{\mathcal{E}_2}, \\ q &= \varpi_1 q_{\mathcal{E}_1} + (1 - \varpi_1) q_{\mathcal{E}_2}, \\ \\ a_{\mathcal{E}_1} &= 3(\gamma_1 a_1 + \gamma_2 a_2 + \gamma_3 a_3), \\ a_{\mathcal{E}_2} &= 3(\gamma'_1 a'_1 + \gamma'_2 a'_2 + \gamma'_3 a'_3), \\ q_{\mathcal{E}_i} &= 1 + (\delta - 1)(p_{\mathcal{E}_i} - 1) \text{ for } i = 1, 2, \\ \delta &= (\beta_r - \beta_l)/(\beta^* - \beta_l), \end{cases}$$

and

$$\begin{aligned}b_{\mathcal{E}_1}(\beta) &= \left(\frac{1}{0.66^{1/3\gamma_1 a_1}} - 1\right)^{-1} 200^\beta + \left(\frac{1}{0.5^{1/3\gamma_2 a_2}} - 1\right)^{-1} 250^\beta + \left(\frac{1}{0.33^{1/3\gamma_3 a_3}} - 1\right)^{-1} 300^\beta, \\ b_{\mathcal{E}_2}(\beta) &= \left(\frac{1}{0.95^{1/3\gamma'_1 a'_1}} - 1\right)^{-1} 100^\beta + \left(\frac{1}{0.5^{1/3\gamma'_2 a'_2}} - 1\right)^{-1} 250^\beta + \left(\frac{1}{0.05^{1/3\gamma'_3 a'_3}} - 1\right)^{-1} 500^\beta.\end{aligned}$$

We have no precision about the relative legitimacy of the percentiles. Thus we fix $\gamma_i = \gamma'_i = 1/3$. Similarly, we have no objective criterion allowing us to favor any expert. Thus we fix $\varpi_1 = 1/2$.

The minimal number of virtual data to obtain the percentiles (25%, 50%, 75%) is 4. To specify percentiles (25%, 75%), we need $a_1 = a_3 = 2a_2$. By normalization, we obtain the default values $a_1 = a_3 = 1.6$ and $a_2 = 0.8$. For the second expert opinion, by the same method, we obtain $a'_1 = a'_3 = 10a'_2$ and $a'_1 + a'_2 + a'_3 = 20$. Then $a'_2 = 20/21 \simeq 0.95$ and $a'_1 = a'_3 \simeq 9.5$. Note that the censored dataset \mathbf{t}_n yields approximatively as much information as $\tilde{n} \simeq 11$ *i.i.d.* data (see § B.4.2.1). Thus, the second posterior distribution will be more dependent from the prior than the data (in an approximate ratio of 66%) if a_2 is not modified (for instance using Table B.2).

Then, since a_1 and a_2 are fixed, the prior uncertainty on β is elicited as explained in § B.4.1.2. We use $(\eta_e, \beta_e) = (290, 2.5)$ to sample virtual data, since the median of $\mathcal{W}(\eta_e, \beta_e)$ is 250. We obtained prior variances $\sigma_1^2 = 1.07$ and $\sigma_2^2 = 0.67$, respectively. We obtained finally $(p, q)_{\mathcal{E}_1} = (1.19, 1.31)$ for the first expert, and $(p, q)_{\mathcal{E}_2} = (3.13, 4.56)$ for the second expert.

Finally, the complete prior corresponds to a virtual sample of size $a = 12$, which implies that the posterior distribution is approximatively as much dependent from subjective knowledge than frequentist knowledge. Separate prior densities on parameters (η, β) and the marginal density $m(t)$ are displayed on Figures B.4, B.5 and B.6. Marginal densities $m(t)$ get good compromises between the specifications. For the expert \mathcal{E}_1 , empirical percentiles of order (33%, 50%, 66%) are (197, 250, 298). For the expert \mathcal{E}_2 , empirical percentiles of order (5%, 50%, 95%) are (98, 255, 494) (these results are computed on 10^5 sampled particles). Posterior survival functions $S(t)$ are displayed in Figure B.7.

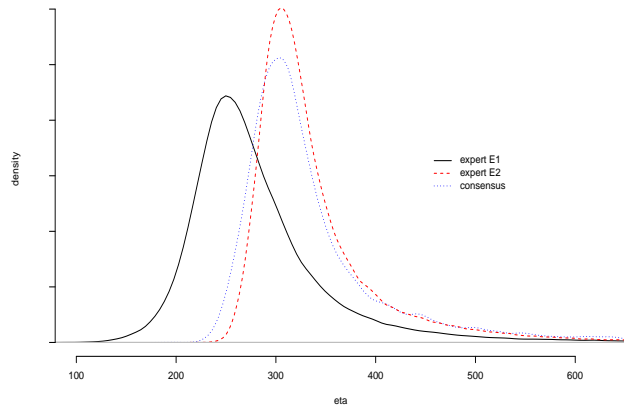


Figure B.4: Default prior densities $\pi(\eta)$ for separate and consensus expert opinions.

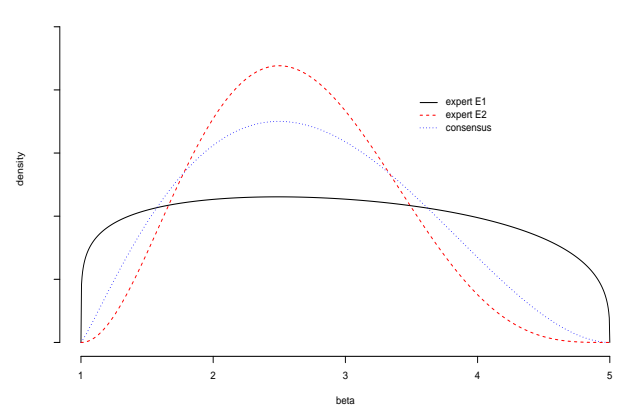


Figure B.5: Default prior densities $\pi(\beta)$ for separate and consensus expert opinions.

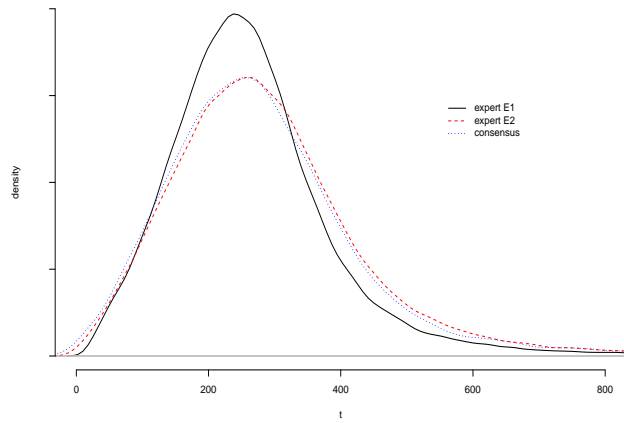


Figure B.6: Default marginal prior densities $m(t)$ for separate and consensus expert opinions.

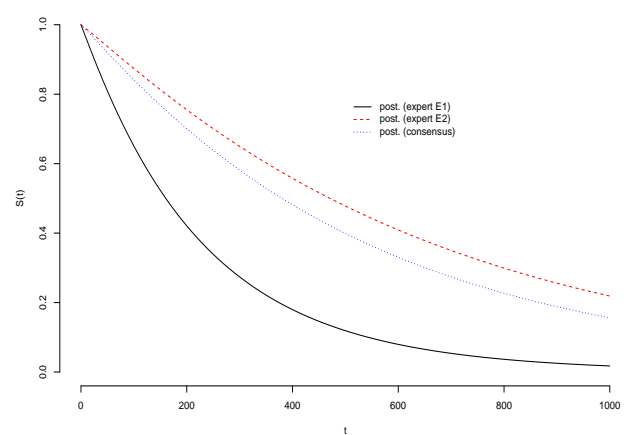


Figure B.7: Posterior survival functions $S(t)$ for separate and consensus expert opinions.

B.7 Conclusion

In this article, we have provided a prior modelling of the Weibull parameters which is practical to use. Some strategies of calibration have been proposed when the prior uncertainty is difficult to assess. Considering the size of a virtual sample, yielding an information which is approximatively the same than the expert information, is practical and subjective and objective data informations can be easily compared. An indicator \tilde{n} is defined to replace the size n when data \mathbf{t}_n are censored, in a more general setting than the Weibull analysis. Finally, the posterior computation remains simple.

.1 Appendix : proof of proposition 17

Denote $F_W(t)$ the Weibull distribution function. We have

$$\begin{aligned} P(T \leq t_e | B_1) &= \int_{\mathbb{R}} \int_{\mathbb{R}} F_W(t_e | \mu, \beta) \pi(\mu | B_1, \beta) \pi(\beta) d\mu d\beta, \\ &= 1 - \int_{\mathbb{R}} \left(\frac{1}{1 + \frac{x_e^\beta}{b_1(\beta)}} \right)^a \pi(\beta) d\beta, \\ &= 1 - \int_{\mathbb{R}} (1 - \alpha) \pi(\beta) d\beta = \alpha. \end{aligned}$$

This results holds even if $\text{Supp}(\pi(\beta))$ is not compact. Assuming $a > \beta_l^{-1}$,

$$\begin{aligned} E[T | B_2] &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mu^{-1/\beta} \Gamma(1 + 1/\beta) \pi(\mu | B_2, \beta) \pi(\beta) d\mu d\beta, \\ &= \int_{\mathbb{R}} \Gamma(1 + 1/\beta) b_2^{1/\beta}(\beta) \frac{\Gamma(a - 1/\beta)}{\Gamma(a)} \pi(\beta) d\beta = t_e. \end{aligned}$$

This result holds if $\text{Supp}(\pi(\beta)) \subset [\beta_l, \infty[$ for any $\beta_l > 0$. Finally, a mode of distribution \mathcal{M} is located in $t_e \neq 0$ such that the derivative $m'_W(t)$ be zero. We have

$$m_W(t) = \int_{\mathbb{R}} \frac{ab^a}{(b + t^\beta)^{a+1}} \beta t^{\beta-1} \pi(\beta) d\beta.$$

With $\text{Supp}(\pi(\beta))$ compact, the derivative is defined and

$$m'_W(t) = - \int_{\mathbb{R}} \frac{ab^a t^{\beta-2}}{(b + t^\beta)^{a+2}} [t^\beta (a\beta + 1) - b(\beta - 1)] \pi(\beta) d\beta. \quad (11)$$

When $a > \beta_l^{-1}$, choosing $b = b_3$ visibly allows to obtain $m'(t_e) = 0$. Besides, $\forall x > 0$, we have

$$m'_W(t | B_3) = \int_{\mathbb{R}} \frac{ab_3^a t^{\beta-2}}{(b_3 + t^\beta)^{a+2}} (a\beta + 1) \pi(\beta) \{t_e^\beta - t^\beta\} d\beta$$

whose sign is the same than $t_e - t$. Then the unicity of the mode is ensured.

Annexe **C**

Tables de données (retour d'expérience)

C.1 Durées de vie de composants nucléaires (circuit d'eau secondaire)

défaillances réelles :	134.9, 152.1, 133.7, 114.8, 110.0, 129.0, 78.7, 72.8, 132.2, 91.8
censures à droite :	70.0, 159.5, 98.5, 167.2, 66.8, 95.3, 80.9, 83.2

TAB. C.1 – Durées de vie X (mois) de composants du circuit secondaire.

	Intervalle de confiance (5%,95%)	Valeur médiane
Expert \mathcal{E}_1	(200,300)	250
Expert \mathcal{E}_2	(100,500)	250

TAB. C.2 – Opinions d'expert sur la durée de vie X .

Glossaire

D.1 Glossaire des termes statistiques

Plan d'échantillonnage : ensemble des conditions qui président à la production d'un échantillon de données ; par exemple : taille, type de données, validité ou censure. Voir par exemple Cadieu *et al.* (2004) pour plus de précisions.

Statistique d'ordre : soit (x_1, \dots, x_n) un échantillon de loi quelconque. La statistique d'ordre i de l'échantillon, notée x_i^* , est la $i^{\text{ième}}$ valeur de l'échantillon trié par ordre croissant de valeur :

$$x_1^* \leq x_2^* \leq \dots \leq x_n^*.$$

Voir par exemple Gaudoin (2002) pour une revue très claire des principales caractéristiques de la distribution des statistiques d'ordre.

Convergence presque sûre : supposons que $\{X_n\}$ soit une suite de variables aléatoires. On dit que $\{X_n\}$ converge vers X *presque sûrement* quand $n \rightarrow \infty$, et on écrit $\{X_n\} \xrightarrow{p.s.} X$ si

$$P(\{\omega : X_n(\omega) \rightarrow X(\omega)\}) = 1.$$

Il s'agit de la notion la plus naturelle de convergence d'une suite de variable aléatoires, qui peut se rapprocher de la notion de convergence simple des suites d'applications (avec un point de vue probabiliste).

Identifiabilité : un modèle paramétrique $\mathcal{M}(\theta)$ où $\theta \in \Theta$ est identifiable si $\theta_1 \neq \theta_2 \Leftrightarrow \mathcal{M}(\theta_1) \neq \mathcal{M}(\theta_2)$; autrement dit, 2 jeux de paramètres différents donnent des modèles différents.

Inférence bayésienne : démarche logique permettant de calculer ou réviser la probabilité d'une hypothèse. Cette démarche est régie par l'utilisation de règles strictes de combinaison des probabilités, desquelles dérive le théorème de Bayes. Dans la perspective bayésienne, une probabilité n'est pas interprétée comme le passage à la limite d'une fréquence, mais plutôt comme la traduction numérique d'un état de connaissance, par exemple le degré de confiance accordé à une hypothèse.

Intervalle de confiance : soit (X_1, \dots, X_n) un échantillon de la loi de probabilité P_θ . On appelle intervalle de confiance de niveau (ou seuil) α sur une fonction $h(\theta)$ un intervalle aléatoire $[T_1, T_2]$, où $T_1 \leq T_2$ sont deux statistiques, fonction de l'échantillon, telles que

$$P_\theta(h(\theta) \in [T_1, T_2]) = 1 - \alpha.$$

Si on réalise M simulations indépendantes de la loi P_θ , les variables aléatoires (T_1, T_2) prendront des valeurs particulières (t_1, t_2) (dites *réalisations*). L'encadrement sera alors vrai ou faux. Pour $\alpha = 0.05$, si on répète 100 fois la série de M expériences pour obtenir 100 intervalles, on peut s'attendre à ce que 5 d'entre eux ne contiennent pas la vraie valeur de $h(\theta)$.

Fonction pivotale : une fonction pivotale pour le paramètre θ est une variable aléatoire fonction des observations (X_1, \dots, X_n) et de θ dont la loi ne dépend pas de θ . L'intérêt majeur d'une fonction pivotale est de permettre d'exprimer un intervalle de confiance sur θ .

Paramètre de nuisance : un paramètre de nuisance θ_2 apparaît dans un modèle statistique $\mathcal{M}(\theta)$, où $\theta = (\theta_1, \theta_2)$ comme un paramètre de valeur inconnue, sans rôle pour l'estimation d'une fonction d'intérêt dépendante uniquement de θ_1 (dit paramètre d'intérêt). Par exemple, si l'on s'intéresse uniquement à l'espérance d'une loi normale, un écart-type inconnu est un paramètre de nuisance. Un problème majeur de la statistique est donc de réduire voire de supprimer l'influence de ce paramètre sur l'inférence. Voir par exemple Basu (1977) pour plus de précisions.

Tribu : une tribu (ou σ -algèbre) \mathcal{A} d'un espace fondamental Ω est un ensemble non vide de parties de Ω si

1. $\forall A \in \mathcal{A}, \bar{A} \in \mathcal{A}$;
2. Pour toute famille dénombrable $\{A_i; i \geq 1\}$ d'éléments de \mathcal{A} ,

$$\bigcup_{i \geq 1} A_i \in \mathcal{A}.$$

Borélien : la tribu borélienne de χ , noté $\mathcal{B}(\chi)$, est la tribu engendrée par la famille \mathcal{F} des ouverts de χ , c'est-à-dire la plus petite tribu $\sigma(\mathcal{F})$ contenant \mathcal{F} . Lorsque $\chi \in \mathbb{R}^{*+}$ (par exemple χ étant un ensemble de temps de défaillance), on a

$$\mathcal{B}(\chi) = \sigma([x_1, x_2]; 0 < x_1 < x_2).$$

D.2 Glossaire des termes d'ingénierie

Ce glossaire regroupe les termes d'ingénierie les plus fréquemment employés dans ce document. Pour une liste plus complète, voir par exemple Bacha *et al.* (1998, pages 107-114).

Durabilité : aptitude d'une entité à accomplir une fonction requise, dans des conditions données d'utilisation et de maintenance, jusqu'à ce qu'un état limite soit atteint (norme EN 13-306, 2001).

Fiabilité : aptitude d'une entité à accomplir une fonction requise, dans des conditions données d'utilisation, pendant un temps donné (normes NF X 60-500 et NF X 50-120).

Maintenabilité : aptitude d'une entité à être maintenue ou rétablie, sur un intervalle de temps donné, dans des conditions d'utilisation et d'accomplissement de la maintenance données (normes NF X 60-500 et 60-010).

Maintenance : ensemble des actions destinées à maintenir ou rétablir une entité dans un état lui permettant d'accomplir une fonction requise (normes NF X 60-500 et 60-010). On distingue en général deux types de maintenance :

1. **corrective** : effectuée après la détection d'une panne;

2. **préventive** : ayant pour objectif de réduire la probabilité de défaillance ou de dégradation d'une entité ou d'un service ; elle est déclenchée selon un échéancier établi à partir d'un nombre prédéterminé d'unités d'usage (maintenance *systématique*) et/ou des critères prédéterminés, significatifs de l'état de dégradation de l'entité ou du service (maintenance *conditionnelle*).

Mécanisme de dégradation : processus physique, chimique ou autre ayant entraîné une défaillance (norme NF X 60-150).

Mode de défaillance : effet par lequel une défaillance est observée (norme NF X 60-010). Exemples : fissure, rupture.

Sûreté de fonctionnement : ensemble des propriétés qui décrivent la disponibilité d'une entité et les facteurs qui la conditionnent : fiabilité, maintenabilité et logistique de maintenance (norme ISO 8402). En France, cet ensemble inclut aussi la sécurité.

Taux de défaillance : désigne la proportion d'entités qui ont survécu à un instant arbitraire t et ne sont plus en vie à l'instant $t+1$, proportion ramenée à l'unité de temps (norme NF X 06-501).

Temps de fonctionnement : la durée de fonction effective d'une entité, synonyme de "temps de service", s'exprime le plus souvent en temps cumulé d'heures de fonctionnement ou en nombre cumulé de sollicitations.

Bibliographie

- ABEL, P.S. & SINGPURWALLA, N.D. (1994). To survive or to fail : that is the question, *Am. Statist.*, **48**, pp. 18-21.
- ABRAMOWITZ, M. & STEGUN, I. A. (EDS.) (1972). "Psi (Digamma) Function.", "Bernoulli and Euler Polynomials and the Euler-Maclaurin Formula". In *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th printing. New York : Dover.
- D'AGOSTINO, R.M. & STEPHENS, M.A. (1986). *Goodness-of-Fit Techniques*, Marcel Dekker : New York.
- ALI, S.M. AND SILVEY, D. (1966). A general class of coefficients of divergence of one distribution from another, *J.Roy.Stat.Soc.*, **28**, pp. 131-142.
- ALTUZARRA, A., MORENO-JIMENEZ, J.M. & SALVADOR, M. (2004). Searching for consensus in AHP-Group decision making. A Bayesian perspective, *Rapport de recherche*, Université de Saragosse.
- ANDRADE, J.A.A. & O'HAGAN, A. (2006). Bayesian robustness modeling using regularly varying distributions, *Bayesian Analysis*, **1**, pp. 169-188.
- ANDRIEU, C. & ROBERT, C.P. (2001). Controlled MCMC for optimal sampling, *Rapport technique 2001-25, Cahiers du CEREMADE*, Université Paris Dauphine.
- ANDRIEU, C., DOUCET, A., FITZGERALD, W.J. AND PÉREZ, J.M. (2001). Bayesian Computational Approaches to Model Selection, *Nonlinear and Non Gaussian Signal Processing*, Smith, R.L., Young, P.C. and Walkden. A. (Eds), Cambridge University Press.
- ANGERS, J.F. (2000). Credence and Robustness Behavior, *Metron*, **58**, pp. 81-108.
- ARTIN, E. (1964). *The Gamma function*, Holt Rinehart Winston, New York.
- ATWOOD, C.L., SHAH, V.N. & GALYEAN, W.J. (1999). Analysis of pressurized water reactor primary coolant leak events caused by thermal fatigue, *ESREL Proceedings*.
- BACHA, M. (1996). "Inférence statistique pour des modèles de durées de vie et applications", *Thèse de doctorat*, Université de Rouen.
- BACHA, M. & CELEUX, G. (1996). BRM-IS : un algorithme d'estimation bayésienne pour modèles à données incomplètes, *Proceedings des XXVIII journées de statistique*, Québec.
- BACHA, M., CELEUX, G., IDÉE, E., LANNOY, A. & VASSEUR, D. (1998). *Estimation de modèles de durées de vie fortement censurées*, Eyrolles.

- BAIN, L. & ENGELHARDT, M. (1991). *Statistical Analysis of Reliability and Life Testing Models*, Marcel Dekker.
- BASU, D. (1977). On the elimination of nuisance parameters, *JASA*, **72**, pp. 355-366.
- BASU, S., SEN, A. & M. BANERJEE (2003). Bayesian analysis of competing risks with partially masked cause of failure, *Applied Statistics*, **52**, pp. 77-93.
- BAYARRI, M.J. AND BERGER, J.O. (2000). P-values for composite null models, *JASA*, **95**, pp. 1127-1142.
- BERGER, J.O. (1985) *Statistical Decision Theory and Bayesian Analysis* (2nd edition), Springer-Verlag.
- BERGER, J.O & BERNARDO, J.M. (1992). On the development of reference priors (with discussion). In : J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, Eds., *Bayesian Statistics 4*, Oxford University Press, pp. 35-60.
- BERGER, J.O. & SUN, D. (1993). Bayesian analysis for the Poly-Weibull Distribution, *JASA*¹, **88**, pp. 1412-1418.
- BERGER, J.O. AND SUN, D. (1994). Bayesian sequential reliability for Weibull and related distributions, *Ann. Inst. Statist. Math.*, **46**, pp. 221-249.
- BERGER, J.O (1994). An overview of robust Bayesian analysis, *Test*, **3**, pp. 5-124.
- BERGER, J.O & BERNARDO, J.M. (1992). On the development of reference priors (with discussion). In : J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, Eds., *Bayesian Statistics 4*, Oxford University Press, pp. 35-60.
- BERGER, J.O. AND PERRICHI, L.R. (1996). The Intrinsic Bayes Factor for Model Selection and Prediction, *JASA*, **91**, pp. 109-122.
- BERGER, J.O. AND PERRICHI, L.R. (1998). Accurate and stable Bayesian Model Selection : the Median Intrinsic Bayes Factor, *Sankhyā : the Indian Journal of Statistics*, **60**, pp. 1-18.
- BERGER, J.O., PERRICHI, L.R. & VARSHAVSKY, J.A. (1998). Bayes Factors and Marginal Distributions in invariant situations, *Sankhya : The Indian Journal of Statistics*, **60**, pp. 307-321.
- BERGER, J.O., AND PERICCHI, L.R. (2002). Training Samples in Objective Bayesian Model Selection, ISDS Discussion Paper 02-14.
- BERNARDO, J.M. (1979). Reference Posterior Distributions for Bayesian Inference, *J. R. Statist. Soc.*, **41**, pp. 113-147.
- BERNARDO, J.M., BERGER, J.O., DAWID, A.P., BAYARRI, M.J., HECKERMAN, D., DEGROOT M.H., LINDLEY, D.V., SMITH A.F.M. & WEST, M. (EDS) (1980-2003). *Bayesian Statistics*, Vol.1 à 7.
- BERNARDO, J.M. (1997). Noninformative Priors Do Not Exist : A Discussion (*with discussion*), *J. Statist. Planning and Inference*, **65**, pp. 159-189.
- BERTHOLON, H. (2001). "Une modélisation du vieillissement", *Thèse de doctorat*, Université Joseph Fourier, Grenoble.
- BERTHOLON, H., BOUSQUET, N. & CELEUX, G. (2006). An alternative competing risk model to the Weibull distribution for modelling aging in lifetime data analysis, *Lifetime Data Analysis*, bientôt publié.

¹Journal of the American Statistical Association

- BIERNACKI C., CELEUX G., VILLAIN B. & VÉRITÉ B. (1998) 'Utilisation des opinions d'experts pour l'analyse et la dégradation des structures passives', *Rapport de Recherche INRIA*.
- BILLY, F., CLAROTTI, C.A. & LANNOY, A. (2004). Inférence des paramètres de lois exponentielles et de Weibull : comparaison des approches classiques et bayésiennes, *Textes des Conférences "Risques & Opportunités"*, Congrès $\lambda\mu$ 14, Bourges.
- BILLY, F., BOUSQUET, N. & CELEUX, G. (2005). Modelling and eliciting expert knowledge with fictitious data, *Proceedings of the WEJ Workshop*, Cadarache (France)
- BILLY, F., BOUSQUET, N., CELEUX, G. & REMY, E. (2006). Notions et mesures de cohérence bayésienne entre connaissance *a priori* et données observées, *Textes des Conférences "Risques & Opportunités"*, Congrès Lambda-Mu 15, Lille.
- BITOUZÉ, D., LAURENT, B. & MASSART, P. (1999). A Dvoretzky-Kiefer-Wolfowitz type inequality for the Kaplan-Meier estimator, *Annales de l'Institut Henri Poincaré*, **35**, pp. 735-764.
- BONNEVIALLE, A.-M. & BILLY, F. (2006). Réactualisation de données de fiabilité issues du REX : faisabilité d'une méthode bayésienne subjective, *Actes du congrès Lambda-Mu*, Lille.
- BOUSQUET, N. (2003). "Estimation de modèles de durée de vie à risques concurrents", *Mémoire de DEA*, Université Joseph Fourier, Grenoble.
- BOUSQUET, N. (2005). Introducing a new way of eliciting prior distributions for Weibull inference in an industrial context, *Communications in Dependability and Quality Management*, **8**, pp. 51-57.
- BOUSQUET, N. (2006). A Bayesian analysis of industrial lifetime data with Weibull distributions, rapport de recherche INRIA RR-6025 (*soumis*).
- BOUSQUET, N., CELEUX, G. & REMY, E. (2005). A protocol for integrating FED and expert data in a study of durability, *Proceedings of the WEJ Workshop*, Cadarache (France)
- BOX, G.E.P. AND TIAO, G.C. (1973). *Bayesian Inference in Statistical Analysis*, Reading, MA :Addison-Wesley.
- BUDESCU, D.V. & RANTILLA, A. K. (2000). Confidence in aggregation of expert opinions. *Acta Psychologica*, **104**, pp. 371-398.
- CADIEUX, J., COLIN, B. & MONGA, E. (2004). Efficacité asymptotique des plans d'échantillonnage, XXXVIèmes Journées de Statistiques de Montpellier.
- CANFIELD, R.V. & BORGMAN, L.E. (1975). Some Distributions of Time to Failure for Reliability Applications, *Technometrics*, **17**, pp. 263-268.
- CAPPÉ, O., GUILLIN, A., MARIN, J.-M. & ROBERT, C.P. (2004). Population Monte Carlo, *Journal of Computational and Graphical Statistics*, **13**, pp. 907-929.
- CASTANIER, B. (1997). "Estimation des variances pour des modèles de durées de vie censurées", *Mémoire de DEA*, INRIA Rhône-Alpes.
- CELEUX, G. & DIEBOLT, J. (1984). "Reconnaissance de mélange de densité et classification - un algorithme d'apprentissage probabiliste : l'algorithme SEM", *Rapport de recherche 349*, INRIA.
- CELEUX, G. & DIEBOLT, J. (1992). A Stochastic Approximation type EM Algorithm for the Mixture Model, *Stochastics and Stochastics Reports*, **41**, pp. 119-134.
- CELEUX, G. & DIEBOLT, J. (1994). L'algorithme SEM comme approximation de l'algorithme d'augmentation de données. In : *XXVI^{es} Journées de Statistiques*, Neuchâtel, pp. 192-194.
- CELEUX, G. (1996). Estimation of failure times involving Weibull distributions via stochastic algorithms, In : *XVIIèmes Rencontres Franco-Belges de Statisticiens*, Marne-la-Vallée.

- CELEUX, G., HURN, M. & ROBERT, C.P. (2000). Computational and inferential difficulties with mixture posterior distributions, *JASA*, **95**, pp. 957-970.
- CELEUX, G. & RODIONOV, A. (2002). A shock model for assessing component aging reliability, *Proceedings of 22rd ESREDA Seminar*, Madrid.
- CELEUX, G., MARIN, J.-M. & ROBERT, C.P. (2006) Iterated importance sampling in missing data problems, *Computational Statistics and Data Analysis*, **12**, pp. 3386-3404.
- CELEUX, G., BILLY, F., BOUSQUET, N. & REMY, E. (2006a). Notions et mesures de cohérence bayésienne entre connaissance *a priori* et données observées, Actes du congrès $\lambda\mu$ 15, Lille.
- CELEUX, G., BILLY, F., BOUSQUET, N. & REMY, E. (2006b). Modélisation de la durée de vie par une loi de Weibull : comparaison de trois approches bayésiennes, Actes du congrès $\lambda\mu$ 15, Lille.
- CHAN, V. & MEEKER, W.Q. (1999). A Failure-Time Model for Infant-Mortality and Wearout Failure Modes, *IEEE Transactions on Reliability*, **48**, pp. 377-387.
- CHANDA, K.C. (1954). A note on the Consistency and Maxima of the Roots of Likelihood Equations, *Biometrika*, **41**, pp. 56-61.
- CHAUVEAU, D., & DIEBOLT, J. (1999). An Automated Stopping Rule for MCMC Convergence Assessment, *Computational Statistics*, **14**, pp. 419-442.
- CHEN, Y., HOLLANDER, M. & LANDBERG, N.A. (1982). Small-sample results for the Kaplan-Meier estimator, *JASA*, **77**, pp. 141-144.
- CLARKE B.S. ET BARRON A.R. (1990). Information-Theoretic Asymptotics of Bayes Methods, *IEEE Trans. Infor. Theory*, **36**, pp. 453-471.
- CLARKE B.S. (1996). Implications of reference priors for prior information and for sample size, *JASA*, **91**, pp. 173-184.
- CLARKE B.S. (1999). Asymptotic Normality of the Posterior in Relative Entropy, *IEEE Trans. Infor. Theory*, **45**, pp. 165-176.
- CLAROTTI C.A. (1998). 'Les techniques fréquentielles et bayésiennes au service de l'ingénieur de sûreté de fonctionnement'. *Rapport final du projet ISdF 8/96*.
- CONSONNI, G., VERONESE, P. & GUTIERREZ-PENA, E. (2004a) Reference priors for natural exponential families having a simple quadratic variance function, *J. Multivariate Analysis*, **88**, pp. 335-364.
- CONSONNI, G., VERONESE, P. & GUTIERREZ-PENA, E. (2004b) Compatible Priors for Bayesian Model Comparison with an Application to the Hardy-Weinberg Equilibrium Model, technical report, University of Pavia.
- COOK, S. & RUBIN, D.B. (2006). Constructing Vague but Proper Prior Distributions in Complex Bayesian Models, *soumis*.
- COOKE, R.M., MENDEL, M. & THIJS, W. (1988). Calibration and Information in Expert Resolution, *Automatica*, **24**, pp. 87-94.
- COOKE, R.M. & GOOSSENS, L.H.J. (2001). *Expert judgement elicitation in risk assessment*. Nederland : Kluwer Academic Publishers.
- COPEN, E.C. (1975). *The difficulty of Assessing Uncertainty*, Society of Petroleum Engineers and American Institute of Mining, Metallurgy and Petroleum Engineers, 50th Fall Conference.
- COVER, T.M. AND THOMAS, J.A. (1991). *Elements of Information Theory*. New York : Wiley.

- COX, D.R. (1953). Some simple approximate tests for Poisson variates, *Biometrika*, **40**, pp. 354-360.
- CRAIU, R.V. & DUCHESNE, T. (2004). Inference based on the EM algorithm for the competing risks model with masked causes of failure, *Biometrika*, **91**, pp. 543-558.
- D'AGOSTINO, R.B. & STEPHENS, M.A. (1986). *Goodness-of-fit techniques*, New York and Basel : Dekker.
- DANESHKHAH, A.R. (2004). Psychological Aspects Influencing Elicitation of Subjective Probability, *research report*, University of Sheffield.
- DANESHKHAH, A.R., OAKLEY, J.E. & O'HAGAN, A. (2006). Nonparametric Prior Elicitation with Imprecisely assessed Probabilities, *research report*, University of Sheffield.
- DASS, S.C. (2001). Propriety of Intrinsic Priors in Invariant Testing Situations, *Journal of Statistical Planning and Inference*, **92**, pp. 147-162.
- DAWID, A.P. (1982). The Well-Calibrated Bayesian (with discussion), *JASA*, **77**, pp. 605-613.
- DAWID, A.P. & LAURITZEN, S. (2000). Compatible prior distribution, in *Bayesian Methods with Application to Science Policy and Official Statistics, ISBA proceedings*, pp. 109-118.
- DATTA, G.S. & GHOSH, J.K. (1995). On priors providing frequentist validity for Bayesian inference, *Biometrika*, **82**, pp.37-45.
- DATTA, G.S. (1996). On priors providing frequentist validity for Bayesian inference for multiple parametric functions, *Biometrika*, **83**, pp. 287-298.
- DEMPSTER, A.P., LAIRD, N.M. & RUBIN, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion), *Journal. Roy. Statist. Soc.*, **39**, pp. 1-38.
- DE FINETTI, B. (1961). "The Bayesian Approach to the Rejection of Outliers", in *Proceedings of the Fourth Berkeley Symposium on Probability and Statistics*, Berkeley : University of California Press, pp. 199-210.
- DE FINETTI, B. (1974). *Theory of Probability*, New York : Wiley.
- DE SANTIS, F., MORTERA, J. & NARDI, A. (2001). Jeffreys priors for survival models with censored data, *J. Statist. Planning and Inference*, **99**, pp. 193-209.
- DIEBOLT, J. & IP, E.H.S. (1996). Stochastic EM; method and application, in : *Markov Chain Monte Carlo in Practice*, W.R. Gilks, S. Richardson and D.J. Spiegelhalter Eds, Chapman & Hall, pp. 259-273.
- DMOCHOWSKI, J. (1996). Intrinsic Priors via Kullback-Leibler Geometry, in *Bayesian Statistics 5*, Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M. Eds, pp. 543-549, Oxford University Press.
- DODSON, B. (2006). *The Weibull Analysis Handbook, second edition*, ASQ Quality Press, Milwaukee.
- DOSS, H. & CHIANG, Y.-C. (1994). Choosing the resampling scheme when bootstrapping : a case study in reliability, *JASA*, **89**, pp. 298-307.
- DOUC, R., GUILLIN, A., MARIN, J.-M. & ROBERT, C.P. (2007). Convergence of adaptive sampling schemes, *Annals of Statistics*, **35**, 1.
- EFRON, B. (1982). The Jackknife, the Bootstrap and Other Resampling Plans, *CBMS*, **38**, SIAM-NSF.
- EFRON, B. & TIBSHIRANI, R.J. (1998). *An Introduction to the Bootstrap*, Chapman & Hall.

- ERTO, P. (1982). New practical Bayes estimators for the 2-parameter Weibull distribution, *IEEE Transactions on Reliability*, **31**, pp. 194-197.
- EVANS, M. & MOSHONOV, H. (2005a). Checking for Prior-Data conflict, Dept. of Statistics, University of Toronto, Technical Report 0413, accepté par *Bayesian Analysis*.
- EVANS, M. & MOSHONOV, H. (2005b). Checking for Prior-Data conflict with Hierarchically Specified Priors, Dept. of Statistics, University of Toronto, Technical Report 0503.
- FINKELSTEIN, M. (2006). Aging : damage accumulation versus increasing mortality rate, *ALT 2006*, Angers.
- FLEHINGER, B.J., REISER, B. AND YASHCHIN, E. (2002). Parametric Modeling for Survival with Competing Risks and Masked Failure Causes, *Lifetime Data Analysis*, **8**, pp. 177-203.
- FRIEDMAN, L. & GERTSBAKH, I.B. (1980). Maximum Likelihood Estimation in a Minimum-Type Model with Exponential and Weibull failure modes, *JASA*, **75**, pp. 460-465.
- GARTHWAITE, P.H., KADANE, J.B. & O'HAGAN, A. (2005). Statistical methods for eliciting probability distributions, *JASA*, **100**, pp. 680-701.
- GAUDOIN, O. (2002). Statistique Non Paramétrique, *Notes de cours ENSIMAG*.
- GEMAN, S. & GEMAN, D. (1984). Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, pp. 721-741.
- GEWEKE, J. (1989). Bayesian Inference in Econometric Models using Monte Carlo Integration, *Econometrica*, **57**, pp. 1317-1339.
- GELMAN, A., MENG, X. AND STERN, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies, *Statistica Sinica*, **6**, pp. 733-808.
- GIESBRECHT, F. & KEMPTHORNE, O. (1976). Maximum Likelihood Estimation in the Three-Parameter Lognormal Distribution, *Journal. Roy. Statist. Soc.*, **38**, pp. 257-264.
- GHOSH, J.K. & MUKERJEE, R. (1993). On priors that match posterior and frequentist distribution functions, *Can. J. Statis.*, **21**, pp. 89-96.
- GHOSH, J.K. (1994). *Higher order asymptotics*, Institute of Mathematical Statistics, Hayward.
- GHOSH, M., REID, N. & FRASER, D.A.S. (2005) Ancillary statistics : a review, *soumis*.
- GHOSHAL, S. (1999a). A review of consistency and convergence of posterior distribution, *Proceedings of Varanashi Symposium in Bayesian Inference*, Banaras Hindu University.
- GHOSHAL, S. (1999b). Probability matching priors for non-regular cases, *Biometrika*, **86**, pp. 956-964.
- GHOSHAL, S. (2004). Bridging maximum likelihood and Bayes, and beyond, slides, Duke University.
- GOETGHEBEUR, E. & RYAN, L. (1995). Competing risks survival analysis, *Biometrika*, **42**, pp. 821-833.
- GOLDSTEIN, H. (2005). Review of WinBUGS, *technical report*, Institute of Education, University of London.
- GOURIEROUX, C. & MONFORT, A. (1996). *Statistique et modèles économétriques*, Economica : Paris.
- GULLIN, A., MARIN, J.-M. & ROBERT & C.P. (2005). Estimation bayésienne approximative par échantillonnage préférentiel, *Revue de Statistique Appliquée*, **54**, pp. 79-95.
- HARTIGAN, J.A. (1983). *Bayes' Theory*, New York : Springer-Verlag.

- HARTIGAN, J.A. (1998). The Maximum Likelihood Prior, *Annals of Statistics*, **26**, pp. 2083-2103.
- HARDER, M.H. (1990). Effects of censoring on the validity of confidence intervals, *JASA*, **85**, pp. 90-96.
- HARTIGAN, J.A. (1964). Invariant prior densities, *Ann. Math. Statist.*, **35**, pp. 836-845.
- HARTIGAN, J.A. & HARTIGAN, P.M. (1985a). The dip test of unimodality, *Annals of Statistics*, **13**, pp. 70-84.
- HARTIGAN, P.M. (1985b). Computation of the Dip Statistic to Test for Unimodality, *Applied Statistics* (JRSS C), **34**, pp. 320-325.
- HASTINGS, W. (1970). Monte Carlo Sampling Methods using Markov Chains and their Applications, *Biometrika*, **57**, pp. 97-109.
- HEALY, M.J.R. & WESTMACOTT, M. (1956). Missing values in experiments analyzed on automatic computers, *Applied Statistics*, **5**, pp. 203-206.
- HILL, B.M. (1974). "On Coherence, Inadmissibility and Inference About Many Parameters in the Theory of Least Squares", in *Studies in Bayesian Econometrics and Statistics*, eds. S.E. Fienberg and A. Zellner, Amsterdam : North-Holland, pp. 555-584.
- HOADLEY, B. (1971). Asymptotic Properties of Maximum Likelihood Estimators for the Independent not Identically Distributed Case, *Ann. Math. Statist.*, **42**, pp. 1977-1991.
- HOLLANDER, M. & PEÑA, E.A. (1992). A chi-squared goodness-of-fit test for randomly censored data, *JASA*, **87**, pp. 458-463.
- IDÉE, E., LANNOY, A. AND MESLIN, T. (2001). *Estimation of a lifetime law for equipment on the basis of a highly right multicensored sample and expert assessments*, rapport de recherche 01-10b de l'équipe LAMA, Université de Savoie. Également Séminaire ESReDA *Lifetime Management*, Erlangen.
- INSEE (2001) : <http://www.insee.fr>.
- IP, E.H.S. (1994). Using the stochastic EM algorithm in multivariate hierarchical models, *technical report*, Stanford University.
- ISDF (2000). "Méthodes de collecte et de traitement d'avis d'experts et guide de mise en oeuvre en sûreté de fonctionnement". *Rapport de projet IsdF n° 6/98*.
- JEFFREYS, H. (1946). An invariant form for the prior probability in estimation problems, *Proceedings of the Royal Society of London*, **186**, pp. 453-461.
- JEFFREYS, H. (1961). *Theory of Probability* (3rd ed.), London : Oxford University Press.
- JENKINSON, D. (2005). The elicitation of probabilities - A review of the statistical literature, *research report*, Open University and University of Sheffield.
- JENNRICH, R.I. & JAMSHIDIAN, M. (1996). Standard errors for EM estimation, *Computing Science and Statistics*, **29**(2), pp. 463-470
- KADANE, J.B. (1980). Predictive and structural methods for eliciting prior distributions, in *Bayesian Analysis in Econometrics and Statistics* (ed. A. Zellner), Amsterdam : North-Holland.
- KADANE, J.B. & WOLFSON, J.A. (1998). Experiences in elicitation, *The Statistician*, **47**, pp. 3-19.
- KAMINSKIY, M.P. & KRIVTSOV, V.V. (2005). A Simple Procedure for Bayesian Estimation of Weibull Distribution, *IEEE Transactions on Reliability*, ??

- KASS, R.E. & VAIDYANATHAN, S. (1992). Approximate Bayes factors and Orthogonal Parameters, with Applications to testing equality of two binomial proportions, *Journal of the Royal Statistical Society*, **54**, pp. 129-144.
- KASS, R.E. & RAFTERY, A.E. (1995). Bayes factors, *J. Amer. Stat. Ass.*, **90**, pp. 773-795.
- KASS, R.E. & WASSERMAN, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *JASA*, **90**, pp. 928-934.
- KASS, R.E. & WASSERMAN, L. (1996). The selection of prior distributions by formal rules, *JASA*, **91**, pp. 1343-1370.
- KIM, S.W. & SUN, D. (2000). Intrinsic Priors for Model Selection Using an encompassing model with applications to censored failure time data, *Lifetime Data Analysis*, **6**, pp. 251-269.
- LAIRD, N.M., DEMPSTER, A.P. & RUBIN, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion), *J. Roy. Statist. Soc.*, **39**, pp. 1-38.
- LANNOY, A. (1995). Analyse quantitative et utilité du REX pour la maintenance et la sécurité des matériels, *Collectif EDF R&D n° 93*, Eyrolles.
- LANNOY, A. & PROCACCIA, H. (1994). *Méthodes avancées d'analyse des bases de données du retour d'expérience industriel*, Eyrolles.
- LANNOY, A. & PROCACCIA, H. (2001). *L'utilisation du jugement d'expert en sûreté de fonctionnement*, Tec & Doc.
- LANNOY, A. & PROCACCIA, H. (2005). *Évaluation et maîtrise du vieillissement industriel*, Tec & Doc.
- LANNOY, A., PROCACCIA, H. & PROCACCIA, M. (2005). What is aging?, *Proceedings of the International Workshop on Practical Applications of Age-Dependent Reliability Models and Analysis of Operational Data*, Institut de Radioprotection et de Sûreté Nucléaire, Fontenay-aux-Roses.
- LAWLESS, J.F. (1982). *Statistical Models and Methods for Lifetime Data*, Wiley.
- LEHMANN, E.L. (1986). *Testing Statistical Hypotheses* (second edition), Wiley : New York, 1986.
- LEI, S. & SMITH, M.R. (2003). Evaluation of several nonparametric bootstrap methods to estimate confidence intervals for software metrics, *IEEE Transactions on Software Engineering*, **29**, pp. 996-1004.
- LEUCARI, V. & CONSONNI, G. (2003). Compatible priors for causal Bayesian networks, in *Bayesian Statistics 7*, Oxford University Press, pp. 597-606.
- LIISBERG, C. (1991). Possible Low-Priced, robust expert systems using neural networks and minimal entropy coding, *Expert Systems with Applications*, **3**, pp. 249-257.
- LILJOI, A. (2003). Approximating priors by finite mixtures of conjugate distributions for an exponential family, *J. Statist. Planning and Inference*, **113**, pp. 419-435.
- LIN, X., PITTMAN, J. & CLARKE, B. (2006). Bayesian Effective Samples and Parameter Size, *soumis*.
- LINDLEY, D.V. (1980). Approximate Bayesian Methods, in J.M. Bernardo, M.H. DeGroot, D.V. Lindley & A.F.M. Smith (eds), *Bayesian Statistics*, University Press, pp. 223-245.
- LO, S.H., MACK, Y.P. & WANG, J.L. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator, *Proba. Theory Related Fields*, **80**, pp. 461-473.

- LOUIS, T. (1982). Finding the observed information matrix when using the EM algorithm, *J. Roy. Statist. Society*, **44**, pp. 226-233.
- LUCAS, W. (1993). When is Conflict Normal?, *JASA*, **88**, pp. 1433-1437.
- MARIN, J.-M. (2006). Conjugate compatible prior distributions between two nested models from the exponential distribution, soumis.
- MCLACHLAN, G.J. & KRISHNAN, T. (1997). *The EM algorithm and Extensions*, Wiley : New York.
- MEEKER, W.Q. & ESCOBAR L.A. (1998). *Statistical Methods for Reliability Data*, Wiley.
- MEYER, M. & BOOKER, J.M. (1987). *Source of correlation between experts : empirical results from two extremes*, NRC report, NUREG/CR-4814.
- MENG, X.L. & RUBIN, D.B. (1991). Using EM to obtain asymptotic variance-covariance matrices : the SEM algorithm, *JASA*, **86**, pp. 899-909.
- NEAL, R. (1999). Regression and classification using Gaussian process priors, *Bayesian Statistics 6*, Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M. Eds, 69-95, Oxford University Press.
- NEAL, R.M. (2001). Transferring prior information between models using imaginary data, *Technical Report No. 0511*, Dept. of Statistics, University of Toronto.
- NEWTON, M. (1991). The Weighted Likelihood Bootstrap and an Algorithm for Prepivoting, *thèse de doctorat*, University of Washington.
- NEWTON, M. & RAFTERY, A.E. (1994). Approximate Bayesian Inference with the Weighted Likelihood Bootstrap, *J. Roy. Statist. Soc.(Ser.B)*, **56**, pp. 3-48.
- NIELSEN, S.F. (2000). The stochastic EM algorithm : estimation and asymptotic results, *Bernoulli*, **6**, pp. 457-489.
- PARK, C. & PADGETT, W.J. (2004). Analysis of Strength Distributions of Multi-Modal Failures Using the EM Algorithm, *technical report 220*, Department of Statistics, University of South Carolina.
- PERROUD, G. & MOUREAU, O. (2004). Aide à la décision à partir de signaux faibles, *Textes des Conférences "Risques & Opportunités"*, Congrès $\lambda\mu$ 14, Bourges.
- PETERSON, A.V. (1977). Expressing the Kaplan-Meier Estimator as a function of empirical subsurvival functions, *JASA*, **72**, pp. 854-858.
- PROCACCIA, M. & PROCACCIA, H. (2005), *Manuel utilisateur du logiciel REXPERT*, société SIAD-COM.
- OAKLEY, J.E. & O'HAGAN, A. (2006). Uncertainty in prior elicitation : a nonparametric approach. Research Report No. 521/02, Department of Probability and Statistics, University of Sheffield.
- O'HAGAN, A. (1979). On outlier rejection phenomena in Bayes inference, *J. Roy. Statist. Soc. B*, **41**, 358-367.
- O'HAGAN, A. (1988). Modelling with heavy tails. , In *Bayesian Statistics 3*, J. M. Bernardo et al (Eds.), Oxford University Press, pp. 345-359.
- O'HAGAN, A. (1990). On outliers and credence for location parameter inference. *JASA*, **85**, pp. 172-176.
- O'HAGAN, A. (1998). Eliciting expert beliefs in substantial practical applications, *The Statistician*, **47**, Part.1, 21-35.

- O'HAGAN, A. (2003). HSSS model criticism (with discussion). In : *Highly Structured Stochastic Systems*, P. J. Green, N. L. Hjort and S. T. Richardson (eds), Oxford University Press, pp. 423-453.
- O'HAGAN, A. (2005). Elicitation, *Significance*, June, pp. 84-86.
- PEERS, H.W. (1965). On confidence sets and Bayesian probability points in the case of several parameters, *J. Roy. Statist. Soc.*, **27**, pp. 9-16.
- PENNY, W.D. (2001). *KL-Divergences of normal, gamma, Dirichlet and Wishart densities*, Technical Report, Wellcome Dpt of Cognitive Neurology, University College London.
- PÉREZ, J.M. (1998). *Development of Conventional Prior Distributions for Model Comparisons*, Ph.D. Thesis, Purdue University.
- PÉREZ, J.M. AND BERGER, J. (2002). Expected posterior prior distributions for model selection, *Biometrika*, **89**, pp. 491-512.
- POLLARD, D. (2006). *Asymptotia* (ouvrage en progression).
- PORTER, P.S., RAO, S.T., KU, J.-Y., POIROT, R.L & DAKINS, M. (1997). Small Sample Properties of Nonparametric Bootstrap t Confidence Intervals, *J. Air Waste Manage. Assoc.*, **47**, pp. 1197-1203.
- PRESS, S.J. (2003). *Subjective and Objective Bayesian Statistics* (second edition), New York : Wiley.
- RAMAMOORTHY, R.V. & GHOSH, J.K. (2003). *Bayesian nonparametrics*, Springer.
- ROBERT, C.P. (1998). *Discretization and MCMC Convergence Assessment*, Lecture Notes 135, Springer-Verlag : New York.
- ROBERT, C.P. (2001). *The Bayesian Choice. A Decision-Theoretic Motivation* (second edition), Springer. En français : *Le choix bayésien. Principes et pratique* (2006), Springer.
- ROBERT, C.P. & CASELLA, G. (2004). *Monte Carlo Statistical Methods (second edition)*, Springer.
- RODIONOV, A. (2005). A Weibull model parameters estimation using strong censored data, *Proceedings of the International Workshop on Practical Applications of Age-Dependent Reliability Models and Analysis of Operational Data*, Institut de Radioprotection et de Sûreté Nucléaire, Fontenay-aux-Roses.
- ROELEN, A.L.C., COOKE, R.M & GOOSSENS, L.H.J. (2004). *Assessment of the validity of expert judgement techniques and their application at Air Traffic Control the Netherlands*. Amsterdam :LVNL.
- RUBIN, D. (1987). A Noniterative Sampling/Importance Resampling alternative to the Data Augmentation Algorithm for creating a few imputations when fractions of missing information are modest : The SIR Algorithm, *JASA*, **82**, pp. 543-546.
- RUBIN, D. (1988). Using the SIR Algorithm to Simulate Posterior Distributions, in *Bayesian Statistics 3*, Bernardo J., DeGroot M., Lindley D. & Smith A. (eds), Oxford University Press, pp. 395-402.
- ROVERATO, A. & CONSONNI, G. (2004). Compatible prior distributions for DAG models, *J. Roy. Statist. Soc.*, **66**, pp. 47-61.
- SCHAFFER, R.E. (1969). *Bayesian reliability demonstration, phase I - data for the a priori distribution*, RADC-TR-69-389, Rome Air Development Center.
- SCHAFFER, R.E. & SHEFFIELD, T.S. (1971). *Bayesian reliability demonstration, phase II - data for the a priori distribution*, RADC-TR-71-139, Rome Air Development Center.

- SCHIEREN, G.A. (1993). Median Worklife, Mean Age at Final Separation, or Transition Probabilities to calculate Expected Lost Earnings?, *Journal of Forensic Economics*, **1**, pp. 103-109.
- SINANOVIĆ, S. & JOHNSON, D.H. (2003). Towards a Theory of Information Processing, J. Franklin Institute, soumis à *IEEE Trans. Signal Processing*.
- SINGPURWALLA, N.D. & SONG, M.S. (1986). An analysis of Weibull lifetime data incorporating expert opinion, in *Probability and Bayesian Statistics* (R.Viertl ed.), Plenum Pub.Corp., pp. 431-442.
- SINGPURWALLA, N.D. (1988). An interactive PC-Based procedure for reliability assessment incorporating expert opinion and survival data, *JASA*, **83**, pp. 43-51.
- SINGPURWALLA, N.D. (2006). *Reliability and Risk : a Bayesian perspective*, Wiley.
- SINGPURWALLA, N.D., SCHAFER, R.E. & MANN, N.R. (1999). *Methods for Statistical Analysis of Reliability and Life Data*, Springer-Verlag.
- SOLAND, R. (1969). Bayesian analysis of the Weibull process with unknown scale and shape parameters, *IEEE Transactions on Reliability*, **18**, pp. 181-184.
- SPIEGELHALTER, D.J. & SMITH, A.F.M. (1982), Bayes factor for linear and for log-linear models with vague prior information, *J. Roy. Stat Soc.*, **44**, pp. 377-387.
- SPIEGELHALTER, D.J., THOMAS, A. & BEST, N.G. (1999). *WinBUGS User Manual*, MRC Biostatistics Unit.
- STEELE, R., RAFTERY, A.E. & EMOND, M. (2003). Computing Normalizing Constants for Finite Mixture Models via Incremental Mixture Importance Sampling (IMIS), *technical Report 436*, Department of Statistics, University of Washington.
- SUN, D. (1997). A note on noninformative priors for Weibull distributions, *J. Statist. Planning and Inference*, **61**, pp. 319-338.
- SUNDBERG, R. (2001). Comparison of confidence procedures for type I censored exponential lifetimes, *Lifetime Data Analysis*, **7**, pp. 393-413.
- SWEETING, T.J., DATTA, G.S. AND GHOSH M. (2005). *Nonsubjective priors via predictive relative entropy regret*, Research Report No.253, Dpt of Statistical Science, University College London.
- SYVERSVEEN, A.R. (1998). Noninformative Bayesian priors. Interpretation and problems with construction and applications, Preprint Statistics No. 3, 1998, Department of Mathematical Sciences, NTNU, Trondheim.
- TANNER, M. & WONG, W. (1987). The Calculation of Posterior Distributions by Data Augmentation, *JASA*, **82**, pp. 528-550.
- TANNER, M. (1991). *Tools for Statistical Inference, Observed Data and Data Augmentation Methods*, Lecture Notes in Statistics, Springer-Verlag : New York.
- TERREL, G. (1990). The Maximal Smoothing Principle in Density Estimation, *JASA*, **85**, pp. 470-477.
- TIERNEY, L. & KADANE, J.B. (1986). Accurate Approximations for Posterior Moments and Marginal Densities, *JASA*, **81**, pp. 82-86.
- TINGLEY, M. & FIELD, C. (1990). Small-sample confidence intervals, *JASA*, **85**, pp. 427-434.
- USHER, J.S. & HODGSON, T.J. (1988). Maximum Likelihood analysis of component reliability using masked system life data, *IEEE Trans. on Reliability*, **37**, pp.550-555.

- USUREAU, E. (2001). *Application des méthodes bayésiennes pour l'optimisation des coûts de développement des produits nouveaux*, thèse de doctorat, Institut des Sciences et Techniques d'Angers.
- VAN DER VAART, A.W. (1998). *Asymptotic Statistics*, Cambridge University Press.
- VAN NOORTWIJK, J.M., DEKKER, R., COOKE, R.M. & MAZZUCHI, T.A. (1992). Expert judgment in Maintenance Optimization, *IEEE Transactions on Reliability*, **41**, pp. 427-431.
- WALD, A. (1949). Note on the consistency of the maximum likelihood estimate, *Ann. Math. Statist.*, **20**, pp.595-601.
- WALKER, S. & Hjort, N.L. (2001). On Bayesian consistency, *Journal of the Royal Statistical Society B*, **63**, 811-821.
- WANG, C.-P. & Ghosh, M. (2000). Bayesian analysis of bivariate competing risks models, *Sankhyā : The Indian Journal of Statistics*, **62**, pp. 388-401.
- WASSERMAN, L. (2000). Asymptotic inference for mixture models using data dependent priors, *J. Roy. Statist. Soc.*, pp. 159- ?
- WELCH, B.N. & PEERS, B. (1963). On formulae for confidence points based on integrals of weighted likelihoods, *J. Roy. Statist. Soc.*, **35**, pp. 318-329.
- WISSE, B., BEDFORD, T. & QUIGLEY, J. (2005). Combining Expert Judgements in the Bayes Linear Methodology, *Proceedings of the Workshop on the Use of Expert Judgement in Decision-Making*, CEA & European Commission, Aix-en-Provence.
- WOLFSON, L.J. (1995). Elicitation of priors and utilities for Bayesian analysis, **thèse de doctorat**, Carnegie Mellon University, Pittsburgh.
- WU, C.F.J. (1983). On the convergence of the EM algorithm, *Annals of Statistics*, **11**, pp. 95-103.
- WU, S.J. (2002). Estimations of the the parameters of the Weibull distribution with progressively censored data, *J. Japan Statist. Soc.*, **32**, pp. 155-163.
- YANG, R. & BERGER, J.O. (1997). A Catalog of Noninformative Priors, *ISDS Discussion Paper 97-42*.
- ZELLNER, A. (1977). Maximal data information prior distributions, in A. Aykæ and C. Brumat. eds., *New developments in the applications of Bayesian methods*, Amsterdam.
- ZELLNER, A. (1996). Models, prior information and Bayesian analysis, *Journal of Econometrics*, **75**, pp. 51-68.

Une méthodologie d'analyse bayésienne pour la prévision de la durée de vie de composants industriels

Résumé. Le contexte de ce travail de recherche est celui de l'utilisation des outils de la statistique bayésienne pour estimer les grandeurs caractérisantes de la durée de vie d'un composant industriel, modélisée par un modèle statistique paramétrique. En effet, le cadre d'étude est celui d'un faible nombre de données de défaillance, par ailleurs censurées, et les méthodes fréquentistes (maximisation de la vraisemblance) proposent des estimations souvent erronées de ces grandeurs. Plus spécifiquement, le modèle de Weibull est étudié, le composant étant supposé soumis au vieillissement. Un modèle mettant en concurrence défaillances par accident et vieillissement est développé, permettant d'améliorer la représentation de la durée de vie. Une connaissance experte est disponible; nous proposons alors une famille de modélisations *a priori* apte à représenter cette connaissance tout en étant facilement calibrable par un analyste industriel. L'un des enjeux importants de cette étude est de positionner cette information subjective à l'information objective, apportée par les données, et ainsi de contrôler la calibration. L'apport majeur de cette thèse est la définition et l'utilisation du critère DAC, dans un cadre plus général, qui détecte une éventuelle incohérence entre les données et la modélisation *a priori*. Celui-ci offre une amélioration par rapport à la seule approche existante. Il apparaît comme un outil pratique, préalable à l'inférence. Des propriétés de calibration par défaut achèvent d'en faire un instrument intéressant de l'analyste industriel. Enfin, un récapitulatif didactique de certaines méthodes de calcul bayésien achève ce travail, dont l'objectif est de fournir un ensemble d'outils adéquats et maniables à un ingénieur.

Mots-clés : fiabilité, durabilité, durée de vie, REX, expert industriel, modèles exponentiel, Weibull, à risques compétitifs, inférence bayésienne, conflit entre données et *a priori*, algorithmes EM, SEM, BRM, PMC.

A methodology of Bayesian analysis for the lifetime prediction of industrial components

Abstract. This thesis takes place in a context when Bayesian techniques are used to estimate magnitudes in industrial reliability and durability, when lifetime is represented by a parametric statistical model. We firstly focus our study around the Weibull distribution, modelling components submitted to aging. A competing risk model between aging and accidental failure is developed. Bayesian technics are used because lifetime data are typically in a few number and contain censored values. Besides, some expert knowledge is provided. Thus, a prior family is elicited such that its calibration is simple for an industrial analyst. Especially, the location of the subjective information with respect to the objective data information is a point of interest. The main contribution of the thesis is the definition and the study of the DAC criterion, which measures a possible discrepancy between a prior and available data, in a larger setting than reliability. DAC improves the existing techniques, has some default calibration properties and constitutes a helpful technique in the toolkit of the Bayesian industrial analyst. Finally, a review of importance sampling approaches - which appear to be especially adapted to Bayesian computation in our industrial settings - ends this thesis work, whose first aim is to propose methodological avenues for engineers.

Keywords: reliability, durability, lifetime, feedback experience data, expert opinion, exponential, Weibull and competing risk model, Bayesian inference, prior-data conflict, algorithms EM, SEM, BRM, PMC.

AMS Classification: