

**Petit avertissement** Ce qui suit constitue la quasi-intégralité des notes de la version 2018 du cours de “MAO probabilités-statistiques” que j’ai assuré en 2017 et 2018.

Deux remarques :

- il y manque les corrigés en Python, que j’ai eu la flemme d’intégrer. . .
- ce document résulte de la fusion un peu cavalière de divers fichiers  $\text{\LaTeX}$ , d’où des problèmes de bibliographie et sans doute d’indexation.



## Chapitre 1

### Simulation de lois

Dans cette séance, nous abordons la simulation de lois sur ordinateur. L'objectif final est pratique : il s'agit de savoir produire des réalisations de lois, soit pour servir de données dans des simulations (que ce soit pour explorer des modèles, ou pour tester des hypothèses), soit pour alimenter des algorithmes stochastiques. La recherche de méthodes efficaces mène à des problèmes théoriques ; il va sans dire que nous ne sommes pas les premiers à nous y intéresser, que des solutions existent déjà et qu'elles sont implémentées dans Python. Par conséquent, si tout ce que l'on veut, c'est simuler des lois classiques, on peut se rendre directement à la section 1.8, où ces méthodes sont résumées. Elles ne couvrent cependant pas toutes les lois : pour simuler des lois un peu exotiques, les sections 1.3, 1.4 (qui présentent des méthodes de simulation universelle) et 1.5 (qui permet de simuler des mélanges de lois) peuvent être utiles. Une bonne partie du contenu peut être considérée comme une suite du cours de probabilités, et les applications comme des exercices classiques.

Pour ce chapitre on commencera par exécuter les commandes suivantes, qui importent les paquets dont nous nous servons.

```
# -*- coding: utf-8 -*-
import numpy as np
import scipy as sp
import matplotlib.pyplot as plt
import numpy.random as rnd
import scipy.stats as sts
```

#### 1.1 Simulation de lois : principes généraux

Soit  $\mu$  une loi de probabilité sur un espace mesurable quelconque. On peut d'abord se demander ce que l'on entend par "simuler la loi  $\mu$ ". Le cahier des charges naturel impose de pouvoir produire une valeur suivant la loi  $\mu$ , non pas une fois mais autant de fois que l'on veut, et de manière indépendante.

Une réaction naturelle pour les mathématiciens que vous êtes est de réduire la simulation de  $\mu$  quelconque à la simulation de  $\mu$  particuliers, et vos souvenirs de théorie de l'intégration doivent vous rappeler le théorème de Kuratowski : tout espace topologique séparable métrique  $X$  complet (dit "polonais") muni de sa tribu borélienne  $\mathcal{B}(X)$  et d'une mesure  $\mu$  de probabilité est isomorphe à  $[0, 1]$  muni de sa tribu borélienne

$\mathcal{B}([0, 1])$  et de la mesure de Lebesgue<sup>1</sup>. Donc, en gros, “ça doit suffire de savoir simuler la loi uniforme sur  $[0, 1]$ ”. En se limitant à la théorie présente dans ces notes, on peut montrer le résultat suivant (dont la formulation est empruntée à [2]) :

**Théorème 1.1.1** *Pour toute loi  $\mu$  sur  $\mathbb{R}^d$ , il existe une fonction borélienne  $f_\mu : \mathbb{R}^d \rightarrow \mathbb{R}^d$  dont l'ensemble des points de discontinuité est Lebesgue-négligeable et telle que la variable aléatoire  $f_\mu(U_1, \dots, U_d)$  soit de loi  $\mu$  si  $U_1, \dots, U_d$  sont des variables aléatoires indépendantes et de loi uniforme sur l'intervalle  $[0, 1]$ .*

**Preuve.** La preuve utilise des résultats énoncés plus loin, elle est l'objet de l'exercice 14.  $\square$

On donne en exercice une preuve constructive de ce théorème, qui fait que l'on pourrait se servir de la fonction  $f$  ainsi construite pour simuler  $\mu$  à partir de simulations de la loi uniforme sur  $[0, 1]$ . Même si, en pratique, on fera en général autrement, ceci justifie que la simulation de la loi uniforme soit notre point de départ. Nous ne discuterons pas la méthode de simulation de la loi uniforme par Numpy<sup>2</sup>.

Acceptons à partir de maintenant que la commande `rnd.rand()` retourne bien des réalisations de variables indépendantes de loi uniforme sur  $[0, 1]$ , et passons à d'autres lois.

**Remarque 1.1.2** *Nous donnons ci-dessous des méthodes de simulations de différentes lois qui ont principalement un intérêt pédagogique.*

## 1.2 Lois discrètes : méthode canonique

Un premier résultat simple va montrer le principe suivant :

*Si l'on sait simuler la loi  $\mathcal{U}([0, 1])$ , on sait simuler toute loi discrète.*

On peut évidemment se ramener au cas où la loi discrète est à support dans  $\mathbb{R}$ . On utilisera le résultat suivant :

**Proposition 1.2.1** *Soit  $(x_i)_{i \in I}$  une famille dénombrable de points de  $\mathbb{R}^d$ , et  $(p_i)_{i \in I}$  une famille d'éléments de  $[0, 1]$  telle que  $\sum_i p_i = 1$ . On suppose  $I = \mathbb{N}^*$  ou  $I = \{1, \dots, \ell\}$  et on note  $s_i = p_1 + \dots + p_i$  (et  $s_0 = 0$ ). Si  $U$  une variable de loi  $\mathcal{U}([0, 1])$ , et que l'on définit la fonction*

$$g(u) = \sum_{i \in I} x_i \mathbb{1}_{]s_{i-1}, s_i]}(u),$$

*alors  $g(U)$  suit une loi déterminée par  $\mathbb{P}(g(U) = x_i) = p_i$ .*

La preuve est simple et laissée en exercice :

**Exercice 1** *Prouvez la Proposition 1.2.1.*

1. au sens où il existe  $E_1 \subset X$ ,  $E_2 \subset [0, 1]$ , tous deux mesurables et de mesures nulles, et  $f : X \setminus E_1 \rightarrow [0, 1] \setminus E_2$  telle que  $f$  et  $f^{-1}$  sont mesurables et préservent la mesure.

2. Sophie Lemaire a dû vous dire que le générateur de nombres pseudo-aléatoires de Numpy utilisait l'algorithme “Mersenne Twister”, je n'en dirai pas plus

Cette méthode est directement applicable. L'exercice suivant propose des applications immédiates et peut être réalisé sans utiliser de boucle.

### Exercice 2

1. Créez un programme tel que `simul(probas)` retourne une réalisation d'une variable aléatoire à valeurs dans  $(0, \dots, n)$  et de loi donnée par le vecteur de probabilités `probas`. Avant de se lancer, on pourra tester ce que retourne `u < probas` si `u` est un réel.
2. Transformez le programme ci-dessus pour que `simul(probas, valeurs)` retourne une réalisation d'une variable aléatoire à valeurs `valeurs` et de loi donnée par `probas`. On fera attention à l'indexation des listes sous Python.
3. Transformez encore ce programme pour que `simul(probas, valeurs, dim)` retourne par exemple un tableau  $m \times n$  de réalisations indépendantes de la loi spécifiée ci-dessus. On pourra utiliser la commande `np.vectorize`.

Le programme ci-dessus, tel qu'il a été implémenté (sans boucle) n'est pas directement applicable au cas d'une loi à support infini. C'est en pratique suffisant pour nous puisque pour les lois discrètes à support infini usuelles (loi de Poisson, loi géométrique) il existe des méthodes *ad hoc* que l'on décrit ci-dessous. On pourra cependant, en utilisant une boucle, coder un programme qui prend en argument un  $p$  qui pourra être un vecteur ligne (couvrant donc le cas fini) ou une fonction, voir exercice 19.

**Exercice 3** Pour les deux loi suivantes :

1. loi uniforme  $\mathcal{U}(\{1, \dots, N\})$ ,
2. loi géométrique  $\mathcal{G}(p)$ ,

appliquez la Proposition 1.3.1 pour écrire des fonctions (mathématiques)  $\varphi$ , complètement explicites (donc que l'on peut implémenter sans boucle ni test), telles que  $\varphi(U)$  a la loi désirée si  $U$  suit une loi  $\mathcal{U}([0, 1])$ . Déduisez-en des fonctions Python prenant en variable d'entrée  $N$  ou  $p$  respectivement et simulant les lois désirées.

**Exercice 4** On testera la qualité des simulations en utilisant la commande `np.bincount`.

## 1.3 Méthode par inversion

Dans cette section, on suppose que  $d = 1$  : la loi  $\mu$  que l'on veut simuler vit sur  $\mathbb{R}$ . La morale de cette section sera la suivante :

*Si l'on sait simuler la loi  $\mathcal{U}([0, 1])$ , on sait simuler toute loi sur  $\mathbb{R}$ .*

La méthode pourra s'étendre au cas  $d > 1$  avec les méthodes discutées en section 1.5.

L'objet fondamental de cette section est la fonction de répartition, définie par  $F(x) = \mu([-\infty, x])$ . On rappelle, au cas où, que  $F$  est une fonction croissante, en tout point continue à droite avec une limite à gauche, a une quantité au plus dénombrable de points de discontinuité, et que  $\lim_{x \rightarrow -\infty} F(x) = 0$  et  $\lim_{x \rightarrow +\infty} F(x) = 1$ . On définit son pseudo-inverse  $F^{(-1)}$  comme la fonction

$$F^{(-1)} : \begin{array}{ll} ]0, 1[ & \rightarrow \mathbb{R} \\ q & \mapsto \inf\{x \in \mathbb{R} \mid F(x) \geq q\}. \end{array}$$

Le résultat que nous exploiterons sera le suivant :

**Proposition 1.3.1** Soit  $\mu$  une mesure de probabilité sur  $\mathbb{R}$ ,  $F$  sa fonction de répartition et  $F^{(-1)}$  son pseudo-inverse. Alors si  $U$  est une variable de loi uniforme sur  $[0, 1]$ , la variable  $F^{(-1)}(U)$  a pour loi  $\mu$ .

La preuve de cette proposition est détaillée dans l'exercice suivant.

**Exercice 5** Soit  $F^{(-1)}$  le pseudo-inverse d'une fonction de répartition  $F$ .

1. Montrez que  $q \leq F(x) \Leftrightarrow F^{(-1)}(q) \leq x$ , et que  $F \circ F^{(-1)}(q) \geq q$  avec égalité si et seulement si  $q \in \text{Im } F$ .
2. Soit  $U \sim \mathcal{U}([0, 1])$ . Montrez que  $F^{(-1)}(U)$  est une variable aléatoire de fonction de répartition  $F$ .

Le point 2. de l'exercice 5 fournit une méthode très générale et directement applicable de simulation de la loi  $\mu$  sur  $\mathbb{R}$  : on calcule (à la main)  $F^{(-1)}$ , on simule une uniforme  $U$  et on choisit  $X = F^{(-1)}(U)$ . L'exercice suivant fait le lien avec la méthode décrite en section 1.2 :

**Exercice 6** Montrez que si  $\mu$  est une loi discrète avec  $\mu(\{x_i\}) = p_i$ , alors  $X = F^{(-1)}(U)$  a exactement l'expression donnée en Proposition 1.2.1.

L'exercice suivant propose des applications pratiques.

**Exercice 7** Appliquez la méthode ci-dessus pour écrire des fonctions prenant en variable d'entrée les paramètres de la loi, et  $m, n$ , et retournant un tableau  $m \times n$  de réalisations indépendantes de la loi. Traitez le cas des lois :

1. exponentielle  $\mathcal{E}(\lambda)$ ,
2. de Weibull<sup>3</sup>  $\mathcal{W}(a, \lambda)$ .

On testera la qualité de ces simulations en utilisant la commande `plt.hist` (on aura avantage à lui donner les options `density=True, bins='auto'`).

Cette méthode très générale a cependant plusieurs défauts pratiques. D'abord, elle impose de calculer  $F^{(-1)}$ , ce qui, quand on sort des cas d'école, n'est pas forcément facile : par exemple, on ne sait pas la calculer dans le cas d'une loi normale. De plus, il faut se rappeler que les simulations de la loi  $\mathcal{U}([0, 1])$  sont des approximations qui prennent des valeurs dans un ensemble discret ; or, là où  $F$  est très plate,  $F^{(-1)}$  change de manière abrupte et peut envoyer l'ensemble "bien dense" des valeurs discrétisées de  $U$  sur un ensemble discret et très étalé.

3. qui pour  $a > 0$  et  $\lambda > 0$  a pour densité  $f(x) = a\lambda x^{a-1} \exp(-\lambda x^a) \mathbb{1}_{\mathbb{R}_+}(x)$ .

## 1.4 Méthode de rejet

Une autre méthode générale est la méthode de rejet, qui permet de simuler une loi  $\mu$ , mais conditionnée à prendre des valeurs dans une sous-partie  $B$  de  $\mathbb{R}^d$ , dès que l'on sait simuler la loi  $\mu$  non conditionnée.

Comme on va le voir, une application directe mais particulièrement intéressante est la simulation d'une loi  $\mu$  à densité  $f$ , dès que l'on sait simuler une loi  $\nu$  à densité  $g$ , avec  $f \leq \lambda g$  pour un certain  $\lambda > 0$ .

**Proposition 1.4.1** Soit  $(X_n)_n$  une suite de variables aléatoires indépendantes de loi  $\mu$  et  $B$  un borélien tel que  $\mu(B) > 0$ . Soit  $T$  le plus petit entier  $n \geq 1$  tel que  $X_n \in B$ . Alors

1.  $T$  est une variable aléatoire de loi géométrique de paramètre  $\mu(B)$ ,
2.  $X_T$  est une variable aléatoire ayant pour loi la loi conditionnelle  $\mu(\cdot|B)$ .

La preuve, encore une fois, est simple et laissée en exercice.

**Exercice 8** Prouvez la Proposition 1.4.1.

Une application directe de la Proposition 1.4.1 est le tirage uniforme suivant la loi uniforme sur un Borélien  $B$  de  $\mathbb{R}^d$ , quand on sait effectuer un tirage uniforme sur un autre Borélien  $C \supset B$ . Donnons un exemple qui nous servira plus tard :

**Exercice 9** Savez-vous tirer un point aléatoirement suivant la loi uniforme dans le carré  $[-1, +1]^2$ ? si oui, utilisez la méthode ci-dessus pour construire une fonction tirant un point uniformément dans le disque unité  $D$ . On pourra utiliser une instruction conditionnelle `if`. En moyenne, combien de tirages de l'uniforme sur  $[-1, +1]^2$  faut-il pour obtenir un tirage de l'uniforme sur  $D$ ? Transformez la fonction pour produire  $n$  points dans le disque, où  $n$  sera un argument optionnel; on pourra dans ce cas utiliser soit la méthode `append` pour une liste, soit la fonction `np.append` pour un array Numpy. Tracez avec la fonction `plt.scatter` un nuage de point obtenu par cette fonction.

Cette méthode permet également de simuler des lois à densité. On utilisera pour cela la proposition suivante :

**Proposition 1.4.2** Soit  $\nu$  une mesure de probabilité sur  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  admettant une densité  $g$  par rapport à la mesure de Lebesgue. Soit  $X_\nu$  une variable aléatoire de loi  $\nu$  et  $U$  une variable aléatoire indépendante de  $X_\nu$ , de loi  $\mathcal{U}([0, 1])$ . Alors la variable  $(X_\nu, Ug(X_\nu))$  suit une loi uniforme sur

$$E_g = \{(x, y) \in \mathbb{R}^d \times \mathbb{R}_+ \mid 0 < y < g(x)\}.$$

On utilisera conjointement avec cette proposition le lemme suivant :

**Lemme 1.4.3** Soit  $(X, Y)$  une variable aléatoire de loi uniforme sur  $E_g$  comme ci-dessus. Alors  $X$  suit la loi de densité  $g$  par rapport à la mesure de Lebesgue.

**Exercice 10** Prouvez la Proposition 1.4.2 et le Lemme 1.4.3 (dans les deux cas, appliquez le théorème du transfert).

Une conséquence immédiate des Propositions 1.4.1 et 1.4.2 est la suivante :

**Proposition 1.4.4** Soient  $\mu$  et  $\nu$  deux mesures de probabilité sur  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  de densités  $f$ ,  $g$  respectivement par rapport à la mesure de Lebesgue, et telles que  $f \leq \lambda g$  Lebesgue-presque partout, pour un certain  $\lambda > 0$ . Soit  $(X_n)_n$  une suite de variables aléatoires indépendantes de loi  $\nu$ ,  $(U_n)_n$  une suite de variables aléatoires indépendantes de loi uniforme  $\mathcal{U}([0, 1])$ , et  $T$  le plus petit entier  $n \geq 1$  tel que  $\lambda U_n g(X_n) \leq f(X_n)$ . Alors

1.  $T$  est une variable aléatoire de loi géométrique de paramètre  $1/\lambda$ ,
2.  $X_T$  est une variable aléatoire de loi  $\mu$ .

**Preuve.** La Proposition 1.4.2 montre que  $(X_n, U_n \lambda g(X_n))$  est une suite i.i.d. de loi uniforme sur  $E_{\lambda g}$ . La Proposition 1.4.1 montre à son tour que  $T$  suit une loi géométrique de paramètre  $\frac{\int f(x) dx}{\int \lambda g(x) dx} = 1/\lambda$ , et que  $(X_T, U_T \lambda g(X_T))$  suit une loi uniforme sur  $E_f$ . Le Lemme 1.4.3 montre que  $X_T$  suit une loi de densité  $f$  par rapport à la mesure de Lebesgue.  $\square$

La Proposition 1.4.4 donne à nouveau une méthode générale et facilement applicable. Le cas le plus simple est celui d'une densité  $f$  sur un intervalle borné de  $\mathbb{R}$ , acceptant une borne uniforme. Remarquez que l'on a intérêt à choisir le  $\lambda$  le plus petit possible (mais respectant, bien sûr, la contrainte  $f \leq \lambda g$ ), comme l'illustre la preuve ci-dessus.

**Exercice 11** Utilisez la méthode du rejet liée à la Proposition 1.4.4 pour simuler :

1. la loi Beta<sup>4</sup> de paramètres  $a > 1$  et  $b > 1$ . On pourra retenir que beaucoup de fonctions spéciales sont disponibles dans `scipy.special`.
2. La loi du demi-cercle<sup>5</sup>. Avant de vous lancer, réfléchissez un instant : le code est déjà écrit. Pourquoi ?
3. La loi normale de paramètres  $\mu$  et  $\sigma^2$ . On pourra utiliser pour  $\nu$  la loi de Laplace (mais on n'en donne une simulation que dans l'exercice 12).

**Remarque 1.4.5** Une manière de choisir une loi  $\nu$  facilement simulable et donnant un  $\lambda$  petit est de polygonaliser le domaine  $E_f$  et d'utiliser l'algorithme décrit dans l'exercice 21.

---

4. La loi Beta  $\beta(a, b)$  avec  $a > 0$  et  $b > 0$  a pour densité  $\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} \mathbb{1}_{[0,1]}(x)$  par rapport à la mesure de Lebesgue.

5. La loi du demi-cercle a pour densité  $\frac{2}{\pi} \sqrt{1-x^2} \mathbb{1}_{[-1,+1]}(x)$  par rapport à la mesure de Lebesgue



## 1.5 Mélanges et conditionnement

Nous allons formuler la notion de mélange dans un cadre relativement abstrait, qui couvrira de nombreuses situations, et nous apprendra le principe suivant :

*Si l'on sait simuler la loi de  $Y$  et toutes les lois conditionnelles de  $X$  sachant  $Y = y$ , alors on sait simuler la loi de  $(X, Y)$ .*

Soit  $(\mu_y)_{y \in \mathbb{R}^d}$  une collection de mesures de probabilités sur  $\mathbb{R}^d$ , et  $\nu$  une autre mesure de probabilité sur  $\mathbb{R}^d$ . On suppose que chaque  $\sigma$ -algèbre associée est soit la tribu borélienne, soit la tribu engendrée par une famille dénombrable de points (notez que l'on ne se restreint pas beaucoup à supposer que les différentes dimensions  $d$  sont les mêmes). On suppose aussi que pour tout événement  $A$ ,  $y \mapsto \mu_y(A)$  est mesurable. On dit alors qu'une mesure de probabilité  $\mu$  est un mélange de  $(\mu_y)_y$  d'intensité  $\nu$  si pour tout événement  $A$ ,

$$\mu(A) = \int \mu_y(A) d\nu(y). \quad (1.1)$$

Ce formalisme correspond au cas où une variable aléatoire  $Y$  a pour loi  $\nu$ , et qu'une autre variable  $X$  a pour "loi conditionnelle à  $Y = y$ "<sup>6</sup> la loi  $\mu_y$ . La loi de  $X$  est alors  $\mu$ .

Par conséquent, si l'on sait simuler une famille de lois  $(\mu_y)_y$  sur  $\mathbb{R}^d$  comme ci-dessus, et que l'on sait simuler une autre loi  $\nu$  sur  $\mathbb{R}^d$ , alors on sait simuler la loi  $\mu$  définie par (1.1) : il suffit de simuler  $Y$  de loi  $\nu$  puis, si l'on a obtenu un résultat  $y$ , de simuler la loi  $\mu_y$ .

**Exercice 12** *Peut-on exprimer la loi de Laplace comme un mélange de lois au sens ci-dessus ? Écrire une fonction Python qui simule une variable aléatoire de loi de Laplace.*

**Exercice 13** *Cet exercice est lié à un texte d'agrégation écrit par Florent Malrieu. On suppose qu'une falaise est habitée par quatre espèces de mouettes. Chaque espèce fabrique un nid dont le diamètre suit une loi  $\mathcal{N}(m_i, \sigma_i^2)$ , et les espèces sont en proportions  $p_i$ ,  $i = 1, \dots, 4$ . Écrire une fonction qui simule le résultat d'une observation aléatoire de nid. Dans le texte cité, on s'intéresse à cette question car il est facile d'observer des nids (plus facile en tout cas que d'observer des mouettes) et facile d'estimer les  $m_i$  et  $\sigma_i$  (en observant les espèces dans des milieux où elles vivent séparées). La difficulté est, à partir de là, d'estimer les  $p_i$  à partir d'observations aléatoires de nids.*

**Exercice 14** *Comment utiliser le "résultat" de cette section pour étendre par exemple la méthode par inversion au cas d'une loi sur  $\mathbb{R}^d$  ? Plus précisément, utilisez la Proposition 1.3.1 et la section 1.5 pour prouver le Théorème 1.1.1.*

## 1.6 Lois normales

On donne ici la méthode classique de simulation des lois normales. Remarquez que l'on a déjà donné, dans l'exercice 11, une méthode par rejet. La première méthode est l'algorithme de Box-Müller. Il est basé sur le résultat suivant :

6. Au sens où pour tout  $\varphi$  mesurable, la variable  $\mathbb{E}(\varphi(X)|\sigma(Y))$  est la fonction  $y \mapsto \int \varphi(x) d\mu_y(x)$  évaluée en  $Y$ . Voir par exemple l'excellent polycopié de J.-F. Le Gall, "intégration, probabilités et processus aléatoires".

**Lemme 1.6.1** Soient  $U_1$  et  $U_2$  deux variables aléatoires indépendantes de loi  $\mathcal{U}([0, 1])$ . Les variables  $X_1$  et  $X_2$  définies par

$$X_1 = \sqrt{-2 \log U_1} \cos(2\pi U_2) \quad X_2 = \sqrt{-2 \log U_1} \sin(2\pi U_2)$$

sont indépendantes et de même loi  $\mathcal{N}(0, 1)$ .

**Exercice 15** Prouvez le Lemme 1.6.1. Déduisez-en une fonction permettant de simuler  $X_1, \dots, X_n$  variables aléatoires i.i.d. de loi  $\mathcal{N}(0, 1)$  (pour  $n$  pair, c'est facile, et pour  $n$  impair ?).

### 1.7 Lois discrètes : méthodes particulières

On donne ici des méthodes *ad hoc* de simulation de deux lois : les lois binomiales et les lois de Poisson.

Pour les lois binomiale, la base théorique est simple : si  $B_1, \dots, B_n$  sont des variables de Bernoulli indépendantes de loi  $\mathcal{B}(p)$ , la variable  $B_1 + \dots + B_p$  suit une loi binomiale  $\mathcal{B}(n, p)$ .

**Exercice 16** Utilisez ce résultat pour donner une fonction qui simule une loi binomiale de paramètres  $n$  et  $p$ . Modifiez votre fonction pour qu'elle puisse vous donner des réalisations indépendantes de loi  $\mathcal{B}(n, p)$ .

Pour les lois de Poisson, on peut utiliser le résultat suivant :

**Lemme 1.7.1** Si  $(U_n)_n$  est une suite de variables aléatoires indépendantes de loi uniforme sur  $[0, 1]$ , la variable

$$X = \inf\{n \geq 0 \mid U_1 \times \dots \times U_{n+1} < e^{-\lambda}\}$$

suit une loi de Poisson de paramètre  $\lambda$ .

**Exercice 17** Prouvez le Lemme 1.7.1. On pourra commencer par se demander quelle loi suivent les  $-\frac{1}{\lambda} \log U_i$ . A partir de cet outil, écrivez une fonction simulant une loi de Poisson de paramètre  $\lambda$ . Comme d'habitude, modifiez cette fonction pour qu'elle puisse vous donner  $n$  réalisations indépendantes, puis un tableau  $m \times n$  de réalisations indépendantes.

### 1.8 Commandes Python

Soyons à présent un peu plus pragmatiques dans l'optique de simuler des phénomènes décrits par des lois classiques. La plupart de ces lois sont déjà implémentées dans Python, ce qui nous épargne en pratique à peu près tout ce qui précède.

**Exercice 18** Allez voir la page en ligne `numpy.random` et l'aide pour le même paquet. Trouvez les commandes permettant de simuler une loi normale, une loi binomiale, une loi exponentielle, une loi discrète dont on spécifie les valeurs et les probabilités.

## 1.9 Exercices supplémentaires

**Exercice 19** *Ecrivez une fonction qui, avec  $p$  en entrée qui est soit un vecteur ligne, soit une fonction sur  $\mathbb{N}^*$ , mais définissant dans les deux cas une probabilité, donne une réalisation de la loi  $\mathbb{P}(N = i) = p_i$  pour tout  $i$ . On généralise donc la fonction définie dans l'exercice 2, en s'autorisant une boucle.*

**Exercice 20** *Montrez la réciproque partielle suivante de l'exercice 5 : si  $X$  une variable aléatoire de fonction de répartition  $F$  continue, montrez que  $F(X) \sim U([0, 1])$  si  $F$  est continue, et que  $F$  continue équivaut à  $F(\mathbb{R}) \supset ]0, 1[$ .*

**Exercice 21** *Comment peut-on simuler une variable de loi uniforme dans un pavé  $[0, r_1] \times \dots \times [0, r_d]$  de  $\mathbb{R}^d$  ? On poursuit avec le cas  $d = 2$ . Dans ce cas, comment simuler une variable de loi uniforme dans un carré quelconque, de sommets  $A = (a_1, a_2)$ ,  $B = (b_1, b_2)$ ,  $C = (c_1, c_2)$ ,  $D = (d_1, d_2)$  ? Et dans un triangle rectangle isocèle ? Comment simuler une variable de loi uniforme dans un triangle rectangle de sommets  $A = (a_1, a_2)$ ,  $B = (b_1, b_2)$ ,  $C = (c_1, c_2)$  ? et dans un triangle quelconque de sommets  $A = (a_1, a_2)$ ,  $B = (b_1, b_2)$ ,  $C = (c_1, c_2)$  ? Programmez une fonction qui, en fonction des variables d'entrée  $A, B, C$ , effectue ce tirage. À partir de cette méthode et d'un algorithme de découpage d'un polygone en triangles, on peut construire une fonction permettant de tirer des lois uniformes sur un polygone quelconque. Ceci peut être très utile pour accélérer des méthodes par rejet.*

**Exercice 22** *Soit  $X = (X_1, \dots, X_d)$  une famille de variables i.i.d. de loi  $\mathcal{N}(0, 1)$ . Montrez que*

$$\|X\|^2 = X_1^2 + \dots + X_d^2 \quad \text{et} \quad \frac{X}{\|X\|}$$

*sont indépendantes avec  $\|X\|^2$  de loi  $\chi^2(d)$  et  $\frac{X}{\|X\|}$  de loi uniforme sur la sphère unité de  $\mathbb{R}^d$ . Utilisez ce résultat pour estimer la distance moyenne à l'équateur sur la Terre.*



## Bibliographie

- [1] M. Benaïm, N. El Karoui, *Promenade aléatoire, chaînes de Markov et simulations; martingales et stratégies.*
- [2] B. Bercu, D. Chafaï, *Modélisation stochastique et simulation.*
- [3] J.-F. Delmas, B. Jourdain, *Modèles aléatoires, applications aux sciences de l'ingénieur et du vivant.*



## Chapitre 2

# Convergence de variables aléatoires et illustrations

Le but de cette séance est de rappeler les différents modes de convergence des variables aléatoires, leurs propriétés et relations, et de voir comment les illustrer. Répétons que l'illustration dont on parle ici ne joue pas forcément le rôle d'exemple qui suit un théorème : c'est aussi un outil pour étudier un modèle et établir des conjectures sur son comportement. On va ici traiter assez peu d'exemples, et on verra plus d'illustrations dans le chapitre prochain sur les "grands résultats de convergence".

### 2.1 Rappels : définitions et propriétés générales

On commence par rappeler les définitions et propriétés élémentaires des modes de convergence pour des suites de variables aléatoires. On pourra consulter à ce sujet les classiques [2, 3, 4] (ma préférence va au troisième). Dans toute cette section,  $(X_n)_n$ ,  $(I_n)_n \dots$  représenteront une suite de variables aléatoires. Suivant les cas, les variables  $X_n$  correspondant à différentes valeurs de  $n$  vivront sur différents espaces de probabilité  $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ , ou au contraire sur un même espace  $(\Omega, \mathcal{F}, \mathbb{P})$ . Toutes les variables considérées sont à valeurs dans un espace  $\mathbb{R}^d$  muni d'une norme  $\|\cdot\|$  (dont le choix importe peu : toutes ces normes sont équivalentes).

On rappelle d'abord les définitions générales :

**Définition 2.1.1** *On dit que :*

- une suite  $(X_n)_n$  de variables aléatoires sur  $(\Omega, \mathcal{F}, \mathbb{P})$  converge presque-sûrement vers une variable  $X$ , également sur  $(\Omega, \mathcal{F}, \mathbb{P})$ , si

$$\mathbb{P}(\lim_{n \rightarrow \infty} X_n = X) = 1.$$

- une suite  $(X_n)_n$  de variables aléatoires sur  $(\Omega, \mathcal{F}, \mathbb{P})$  converge en norme  $L^p$  (où  $p \geq 1$ ) vers une variable  $X$ , également sur  $(\Omega, \mathcal{F}, \mathbb{P})$ , si

$$\lim_{n \rightarrow \infty} \mathbb{E}(\|X_n - X\|^p) = 0.$$

- une suite  $(X_n)_n$  de variables aléatoires sur  $(\Omega, \mathcal{F}, \mathbb{P})$  converge en probabilité vers une variable  $X$ , également sur  $(\Omega, \mathcal{F}, \mathbb{P})$ , si pour tout  $\epsilon > 0$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}(\|X_n - X\| > \epsilon) = 0.$$

16 CHAPITRE 2. CONVERGENCE DE VARIABLES ALÉATOIRES ET ILLUSTRATIONS

- une suite de variables aléatoires  $(X_n)_n$  sur  $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$  converge en loi vers une variable  $X$  sur  $(\Omega, \mathcal{F}, \mathbb{P})$ , si pour toute fonction continue bornée  $\varphi$  sur  $\mathbb{R}^p$ ,

$$\lim_{n \rightarrow \infty} \mathbb{E}(\varphi(X_n)) = \mathbb{E}(\varphi(X)).$$

On note  $\xrightarrow{\text{p.s.}}$ ,  $\xrightarrow{\mathbb{P}}$ ,  $\xrightarrow{L^p}$ ,  $\xrightarrow{\mathcal{L}}$  respectivement ces quatre modes de convergence.

**Remarque 2.1.2** La convergence en loi peut donc s'énoncer pour des variables vivant sur des espaces de probabilité différents puisque la convergence concerne la loi des variables et non les variables elles-mêmes.

On rappelle maintenant les relations générales entre ces différents modes de convergence :

**Proposition 2.1.3**

1. La convergence presque-sûre implique la convergence en probabilité,
2. la convergence  $L^q$  implique la convergence  $L^p$  si  $q \geq p$ , et la convergence en probabilité,
3. la convergence en probabilité implique la convergence en loi,
4. la convergence en probabilité implique la convergence presque-sûre d'une suite extraite,
5. la convergence en probabilité d'une suite  $(X_n)_n$  telle que  $(X_n^p)_n$  est uniformément intégrable implique la convergence  $L^p$ ,
6. la convergence en loi d'une suite  $(X_n)_n$  vers une constante  $c$  implique la convergence en probabilité vers  $c$ .

Rappelons qu'une suite  $(X_n)_n$  est uniformément intégrable (ou u.i.) si et seulement si elle vérifie l'une des deux conditions équivalentes suivantes :

$$\limsup_{c \rightarrow \infty} \mathbb{E}(|X_n| \mathbb{1}_{|X_n| > c}) = 0$$

$$\forall \epsilon > 0 \exists \delta > 0 \text{ tel que } \mathbb{P}(A) < \delta \Rightarrow \sup_n \mathbb{E}(|X_n| \mathbb{1}_A) < \epsilon.$$

Il est immédiat que s'il existe  $Y$  intégrable telle que pour tout  $n$ , on a  $|X_n| \leq Y$  p.s. alors  $(X_n)_n$  est u.i. Le point 5 de la Proposition 2.1.3 étend donc le théorème de convergence dominée. La Proposition 2.1.3 est résumée par la Figure 2.1.

On rappelle la propriété utile suivante (la démonstration à l'aide du lemme de Borel-Cantelli est un classique et est laissée en exercice) :

**Lemme 2.1.4** Soit une suite  $(X_n)_n$  de variables aléatoires sur  $(\Omega, \mathcal{F}, \mathbb{P})$ . Si l'on a  $\sum_n \mathbb{P}(\|X_n\| > \epsilon) < \infty$ , pour tout  $\epsilon > 0$ , alors  $X_n \xrightarrow{\text{p.s.}} 0$ .

Ce résultat s'utilise naturellement : considérant une suite  $(X_n)_n$  et sa limite éventuelle  $X$ , si l'on veut étudier le mode de convergence, on commence par se demander si  $\mathbb{P}(\|X_n - X\| > \epsilon) \rightarrow 0$  en estimant  $\mathbb{P}(\|X_n - X\| > \epsilon)$ , puis on exploite l'estimation pour se demander si la série associée converge.

On donne également les résultats de stabilité suivants :



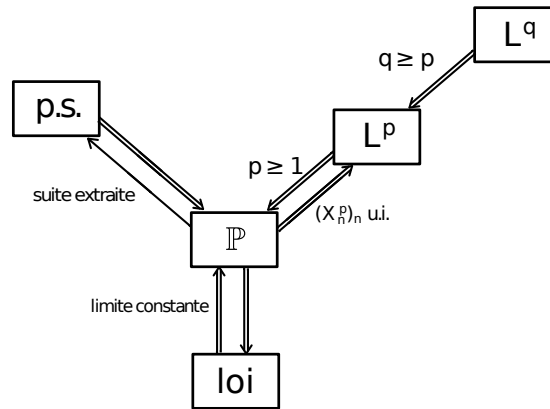


FIGURE 2.1 – La Proposition 2.1.3 en image

**Proposition 2.1.5** *Si  $f$  est une fonction continue, alors :*

- Si  $X_n \xrightarrow{p.s.} X$ , on a  $f(X_n) \xrightarrow{p.s.} f(X)$ ,
- si  $X_n \xrightarrow{\mathbb{P}} X$ , on a  $f(X_n) \xrightarrow{\mathbb{P}} f(X)$ ,
- Si  $X_n \xrightarrow{\mathcal{L}} X$ , on a  $f(X_n) \xrightarrow{\mathcal{L}} f(X)$ .

Ces résultats seront utiles si on les combine avec la proposition suivante, puisqu'ils permettront de discuter par exemple la convergence de  $(X_n + Y_n)_n$  si l'on suppose la convergence de  $(X_n)_n$  et de  $(Y_n)_n$  :

**Lemme 2.1.6**

- Si  $X_n \xrightarrow{p.s.} X$  et  $Y_n \xrightarrow{p.s.} Y$  alors  $(X_n, Y_n) \xrightarrow{p.s.} (X, Y)$ ,
- si  $X_n \xrightarrow{\mathbb{P}} X$  et  $Y_n \xrightarrow{\mathbb{P}} Y$  alors  $(X_n, Y_n) \xrightarrow{\mathbb{P}} (X, Y)$ .

Pour la convergence en loi, il n'y a pas de résultat aussi simple, et en particulier  $X_n \xrightarrow{\mathcal{L}} X$  et  $Y_n \xrightarrow{\mathcal{L}} Y$  n'implique pas  $X_n + Y_n \xrightarrow{\mathcal{L}} X + Y$ , ni  $X_n \times Y_n \xrightarrow{\mathcal{L}} X \times Y$ . La raison en est que la loi de  $X + Y$ , par exemple, ne dépend pas que des lois de  $X$  et de  $Y$  mais de la loi du couple  $(X, Y)$ . Considérez par exemple  $X_n \sim \frac{1}{2} \delta_{-1} + \frac{1}{2} \delta_{+1}$  et  $Y_n = -X_n$ , auquel cas  $X_n \xrightarrow{\mathcal{L}} X_1$ ,  $Y_n \xrightarrow{\mathcal{L}} X_1$  mais  $X_n + Y_n = 0$  ne converge pas en loi vers  $2X_1$ .

Un résultat particulièrement utile dans cette direction<sup>1</sup> est le lemme de Slutsky :

**Lemme 2.1.7 (lemme de Slutsky)** *Si  $X_n \xrightarrow{\mathcal{L}} X$  et  $Y_n \xrightarrow{\mathcal{L}} c$  où  $c$  est une constante, alors  $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, c)$ .*

La preuve de ce lemme est donnée dans [1].

1. et dont la connaissance est un instrument efficace de discrimination entre candidats à l'agrégation

Dans toute la suite, on s'intéresse à une suite  $(X_n)_n$  de variables aléatoires, dont on suppose que l'on sait les simuler. On va voir que les deux modes de convergence les plus simples à illustrer sont la convergence presque-sûre et la convergence en loi. La littérature passe généralement sous silence la raison pour laquelle on n'illustre pas les convergences  $L^p$  et  $\mathbb{P}$ ; ici j'ai voulu expliquer pourquoi, mais comme d'habitude il est pénible d'expliquer pourquoi "ça ne marche pas".

## 2.2 Illustration de la convergence presque-sûre

La convergence presque-sûre est la plus simple à illustrer. Pour cela, on répète plusieurs fois l'opération suivante :

- on simule une réalisation de la suite  $(X_n)_n$ ,
- on trace la suite des valeurs obtenues.

Si la convergence a lieu, les tracés doivent tous avoir une asymptote. Cependant :

- Il faut faire attention à ce que, pour chaque réalisation, chaque  $X_k(\omega)$  corresponde au même  $\omega$ , et que l'on ne change pas l'aléa pour chaque  $k$ . L'exercice 23 illustre la différence entre les deux situations.
- On a l'habitude d'illustrer la convergence presque-sûre dans des cas du type "loi des grands nombres", auquel cas la limite est une constante et les trajectoires ont toutes la même asymptote. Il faut faire attention au fait que si  $X_n \xrightarrow{p.s.} X$  mais que la variable limite  $X$  n'est pas presque-sûrement constante, la limite peut dépendre du  $\omega$ , donc l'asymptote mentionnée ci-dessus n'est pas forcément la même pour toutes les réalisations de suites. Voir les exercices 23 et 30.

**Exercice 23** Soit  $(R_n)_n$  une suite i.i.d. de variables de Rademacher  $\mathbb{P}(R_1 = \pm 1) = 1/2$ . Pour un  $a \in ]-1, +1[$  fixé on pose

$$X_n = a^0 R_0 + \dots + a^{n-1} R_{n-1}.$$

Montrez que la suite  $X_n$  converge presque-sûrement, puis illustrez cette convergence en simulant plusieurs réalisations de la suite  $(X_n)_n$ . Supposons maintenant que pour la simulation de chaque  $X_n$ , on tire à nouveau les  $R_1, \dots, R_n$ , de sorte que ce qu'on calcule est

$$X'_n = a^0 R'_{n,0} + \dots + a^{n-1} R'_{n,n-1}$$

où les  $(R'_{n,k})_{n,k}$  sont indépendants. A-t-on à nouveau convergence presque-sûre ? et convergence en loi ? Enfin, on se pose les mêmes questions à propos de

$$X''_n = a^0 R_{n-1} + \dots + a^{n-1} R_0.$$

## 2.3 Illustration des convergences $L^p$ et $\mathbb{P}$

La convergence  $L^1$  ou  $L^p$  et la convergence en probabilité ont ceci de commun qu'elles sont caractérisées par la décroissance d'une espérance ou d'une famille de probabilités. Comme ces espérance et probabilité sont inconnues, on ne peut que les estimer.

On peut donc illustrer la convergence  $L^p$  via les moyennes empiriques, qui servent à estimer la quantité  $\mathbb{E}(|X_n - X|^p)$  :

- on simule, pour différents  $n$  et  $N$  grand,  $N$  réalisations  $(X_n^{(k)})_{k=1}^N$  et  $(X^{(k)})_{k=1}^N$  de  $X_n$  et de  $X$ ,
- on calcule la moyenne empirique  $\frac{1}{N} \sum_{k=1}^N |X_n^{(k)} - X^{(k)}|^p$ .

De même, on peut illustrer la convergence en probabilité  $X_n \xrightarrow{\mathbb{P}} X$  via les probabilités empiriques, qui vont servir à estimer la quantité  $\mathbb{P}(|X_n - X| > \epsilon)$  :

- on simule, pour différents  $n$  et  $N$  grand,  $N$  réalisations  $(X_n^{(k)})_{k=1}^N$  et  $(X^{(k)})_{k=1}^N$  de  $X_n$  et de  $X$ ,
- pour un  $\epsilon$  fixé, on calcule la probabilité empirique  $\frac{1}{N} \sum_{k=1}^N \mathbb{1}_{|X_n^{(k)} - X^{(k)}| > \epsilon}$ .

Pour les deux cas ci-dessus, on exploite une propriété asymptotique : les quantités calculées sont des estimateurs consistants de  $\mathbb{E}(|X_n - X|^p)$  ou  $\mathbb{P}(|X_n - X| > \epsilon)$ . La question naturelle est alors de savoir ce qu'est “ $N$  grand” ; ce choix ne peut se faire que compte tenu d'un seuil de confiance choisi par ailleurs et on se retrouve à donner des intervalles de confiance pour ces quantités. On peut déjà faire deux remarques qui soulignent les problèmes que l'on va avoir :

### Remarque 2.3.1

- *On a supposé ci-dessus que l'on savait également simuler  $X$ , or ce n'est en général pas le cas ! En particulier, il n'y a pas en général de moyen d'observer le  $X$  à partir du comportement des  $(X_n)_n$  seuls.*
- *On veut ici des intervalles de confiance pour des quantités dont on cherche à montrer qu'elles tendent vers zéro : il va donc falloir pouvoir faire tendre le diamètre de l'intervalle de confiance vers zéro aussi, et cela sans que le  $N$  croisse trop vite.*
- *Puisque l'on veut illustrer la convergence des quantités estimées, on voudrait idéalement pouvoir choisir  $N$  tel que les intervalles de confiance donnés sont valables pour tout  $n$  “assez grand”.*

On traitera plus tard dans le cours la question de l'estimation. Ici, on va principalement expliquer qu'il n'y a pas de bonne méthode pour estimer les quantités ci-dessus.

Pour  $\mathbb{E}(|X_n - X|^p)$ , et en l'absence d'information particulière sur les  $(X_n)$  et  $X$ , il n'y a pas de méthode générale et efficace. Supposons par exemple que l'on utilise l'inégalité de Bienaymé-Tchebychev : qui dit que si la suite i.i.d.  $(Y^{(k)})_{k=1}^N$  est  $L^2$ , alors avec  $\bar{Y}_N$  la moyenne empirique, on a

$$\mathbb{P}(|\bar{Y}_N - \mathbb{E}(Y^{(1)})| > \epsilon) = O\left(\frac{1}{N}\right).$$

Il est alors nécessaire d'avoir des informations supplémentaires, par exemple des bornes sur les moments des  $X^{(k)}$  et de  $X$ , pour avoir des intervalles de confiance explicites. De plus, on voudrait pouvoir appliquer ceci à  $Y^{(1)} = X_n - X$ , et cela pour une infinité de  $n$ , or les majorations ci-dessus ne sont en général pas assez robustes pour fournir des résultats valables avec probabilité grande pour tout  $n$ .

Dans le cas de la convergence en probabilité on a un résultat beaucoup plus puissant : l'inégalité de Hoeffding (démontrée dans [2]).

**Proposition 2.3.2 (inégalité de Hoeffding)** Soit  $(Y^{(k)})_{k=1}^N$  une suite i.i.d. de variables à valeurs dans un intervalle  $[a, b]$ . Alors si  $\bar{Y} = \frac{1}{N} \sum_{k=1}^N Y^{(k)}$  :

$$\mathbb{P}(|\bar{Y}_N - \mathbb{E}(Y^{(1)})| \geq t) \leq 2 \exp - \frac{2Nt^2}{(b-a)^2}$$

Celle-ci va nous permettre d'obtenir des résultats pertinents pour illustrer la convergence en probabilité :

**Exercice 24** Dans le cadre précédent, on note pour un  $\epsilon$  fixé

$$p_n = \mathbb{P}(|X_n - X| > \epsilon), \quad p_{N,n} = \frac{1}{N} \sum_{k=1}^N \mathbb{1}_{|X_n^{(k)} - X^{(k)}| > \epsilon}.$$

Montrez alors que pour  $\lambda > 1/2$  et  $n_{\max}$  on a

$$\mathbb{P}(\forall n \geq N_0, |p_{n,n} - p_n| \leq \sqrt{\frac{\lambda \log n}{n}}) \geq 1 - \frac{2}{2\lambda - 1} (N_0 - 1)^{1-2\lambda}$$

et

$$\mathbb{P}(\forall n \in \{N_0, \dots, n_{\max}\}, |p_{n_{\max},n} - p_n| \leq \sqrt{\frac{\lambda \log n}{n}}) \geq 1 - \frac{2}{2\lambda - 1} (N_0 - 1)^{1-2\lambda}.$$

On peut donc, pour illustrer la convergence en probabilité, choisir par exemple  $\lambda = 1$  et  $N_0$  tel que le dernier majorant ci-dessus soit supérieur à un niveau  $1 - \alpha$  fixé au préalable. On peut alors :

- simuler  $n_{\max} > N_0$  réalisations de  $X$  et  $X_1, \dots, X_{n_{\max}}$ ,
- calculer les  $p_{n_{\max},n}$  pour  $N_0 \leq n \leq n_{\max}$ .

On a alors avec probabilité supérieure à  $1 - \alpha$  que  $p_n$  est inférieur à  $p_{n,n} + \sqrt{\frac{\log n}{n}}$  pour tout  $n \geq N_0$ . Si l'on observe  $p_{n,n} \rightarrow 0$ , on sait alors avec probabilité supérieure à  $1 - \alpha$  que  $p_n \rightarrow 0$ , donc que l'on a convergence en probabilité. Le problème est que l'on doit effectuer  $n_{\max}^2$  simulations.

Une fois que le choix du  $N_0$  motivé, la convergence en probabilité peut être illustrée car  $p_{n_{\max},n}$  représente la proportion des  $X_n^{(k)}$  qui n'est pas dans  $] - \epsilon, +\epsilon[$ . On devrait donc observer, s'il y a convergence en probabilité, que la proportion de trajectoires de  $X_n^{(k)}$  qui se trouvent dans  $] - \epsilon, +\epsilon[$  tend vers zéro quand  $n$  croît, mais à la différence du cas d'une convergence presque-sûre, ce ne sera pas toujours les mêmes trajectoires.

**Exercice 25** On définit une suite de fonctions  $(f_n)_n$  sur  $[0, 1]$  par le fait que  $f_{2^k+\ell}$  est l'indicatrice de  $[\ell/2^k, (\ell+1)/2^k[$  pour  $\ell \in \{0, \dots, 2^k - 1\}$ . Soit  $(U_n)_n$  une suite de v.a. i.i.d. de loi uniforme sur  $[0, 1]$  et  $X_n = f_n(U_0) + \frac{1}{n} \sum_{m=1}^n U_m$ . Montrez qu'on a  $X_n \xrightarrow{\mathbb{P}} 0$  mais pas  $X_n \xrightarrow{\text{P.s.}} 0$  et illustrez ce fait (on pourra prendre  $n_{\max} = 100$  par exemple).

La conclusion de cette discussion est qu'illustrer les convergences  $L^p$  et en probabilité est assez pénible !

## 2.4 Illustration de la convergence en loi

Il existe plusieurs méthodes pour illustrer la convergence en loi, dont les plus classiques sont :

- le tracé d’histogrammes (que nous avons utilisées au TD 1),
- dans le cas de variables aléatoires discrètes, le tracé de diagrammes en bâton,
- le tracé de fonctions de répartitions empiriques.

Dans ce cours, nous ne parlerons que de la troisième méthode, qui a deux avantages notables : elle fonctionne dans tous les cas et elle est simple à coder.

Détaillons : on considère une suite  $(X_n)_n$  de variables aléatoires, et  $X$  une autre variable, et l’on note  $(F_n)_n, F$  les fonctions de répartition associées. On rappelle que  $X_n \xrightarrow{\mathcal{L}} X$  si et seulement si  $F_n(t) \rightarrow F(t)$  pour tout  $t$  point de continuité de  $F$ . Si l’on connaît de manière théorique les fonctions  $F_n$  et  $F$ , alors on peut tenter d’étudier la convergence de  $F_n(t)$  de manière théorique, et si l’on n’y arrive pas, on peut tracer directement ces fonctions. Si en revanche l’une des deux est inconnue, il faut l’obtenir par approximation.

On utilise pour cela le théorème de Glivenko-Cantelli, qui dit que si  $(Y^{(k)})_{k=1}^N$  est un  $N$ -échantillon de même loi que  $Y$ , alors sa fonction de répartition empirique, définie par

$$\widehat{F}_Y^{(N)}(t) = \frac{1}{N} \sum_{k=1}^N \mathbb{1}_{Y^{(k)} \leq t}$$

converge presque-sûrement uniformément en  $t$ , quand  $N \rightarrow \infty$ , vers la fonction de répartition  $F_Y$  de  $Y$ . Formulée en une ligne, la conclusion est

$$\mathbb{P}\left(\sup_{t \in \mathbb{R}} |\widehat{F}_Y^{(N)}(t) - F_Y(t)| \xrightarrow{N \rightarrow \infty} 0\right) = 1.$$

Rappelons que ce théorème, s’il est un peu pénible à démontrer, est essentiellement une application de la loi des grands nombres.

On va donc estimer les fonctions de répartition, par exemple celle  $F_Y$  de  $Y$ , par la fonction de répartition empirique  $\widehat{F}_Y^{(N)}$  associée à un  $N$ -échantillon  $Y^{(1)}, \dots, Y^{(N)}$  de  $Y$ , et cela en fixant un  $N$  assez grand. Grand comment ? une estimation peut être donnée par le théorème de Kolmogorov-Smirnov, qui dit que

$$\sqrt{N} \sup_{t \in [0,1]} |\widehat{F}_Y^{(N)}(t) - F_Y(t)| \tag{2.1}$$

converge en loi quand  $N \rightarrow \infty$  vers une loi donnée, appelée “loi de Kolmogorov-Smirnov” quelle que soit la loi de  $Y$ . La variable en question ayant des valeurs de l’ordre de l’unité, on admettra donc que le  $N$  doit être approximativement égal à l’inverse du carré de la précision voulue dans l’estimation de  $F_Y$ .

On se demande à présent comment tracer de manière la fonction de répartition empirique  $F_Y^{(N)}$  d’un  $N$ -échantillon  $Y^{(1)}, \dots, Y^{(N)}$ .

**Exercice 26** Soit  $Y^{(1)}, \dots, Y^{(N)}$  un  $N$ -échantillon d'une variable  $Y$ , et  $\hat{F}$  la fonction de répartition empirique associée. On note  $Y_{(1)}, \dots, Y_{(N)}$  la statistique d'ordre associée à  $Y^{(1)}, \dots, Y^{(N)}$  (autrement dit, les  $Y_{(1)}, \dots, Y_{(N)}$  sont les  $Y^{(1)}, \dots, Y^{(N)}$  ordonnés par ordre croissant). Si  $Y$  est un vecteur contenant les valeurs  $Y^{(1)}, \dots, Y^{(N)}$ , comment obtenir un vecteur  $Y_{\text{ord}}$  contenant les  $Y_{(i)}$  (dans le bon ordre)? Combien vaut  $\hat{F}$  sur l'intervalle  $[Y_{(i)}, Y_{(i+1)})$ ? Comment tracer une fonction en escalier?

La conclusion est que si  $Y$  est un vecteur contenant  $Y^{(1)}, \dots, Y^{(N)}$ , alors

```
Yord=np.sort(Y)
valeurs=np.arange(0,N)/N
plt.step(Yord,valeurs)
```

permet de tracer la fonction de répartition empirique associée à  $Y^{(1)}, \dots, Y^{(N)}$ .

Par Glivenko-Cantelli, pour  $N$  assez grand, le tracé obtenu à partir d'un  $N$ -échantillon de  $Y$  doit être proche de la fonction de répartition théorique de  $Y$ . On cherche à la comparer à la fonction de répartition de  $Y$ ; encore une fois, si on ne la connaît pas de manière théorique, on utilise la même méthode d'approximation décrite ci-dessus.

**Exercice 27** Soient  $Z$  une variable aléatoire de loi  $\mathcal{N}(0, 1)$  et  $E$  une variable aléatoire de loi  $\mathcal{E}(1)$ , indépendante de  $Z$ . À quoi ressemble la fonction de répartition de  $Y = E + Z$ ?

On va maintenant dire un mot de la manière de tracer une fonction de répartition connue théoriquement. On peut bien sûr, si l'on connaît leur expression, le faire comme pour toute fonction. Si l'on ne connaît pas cette expression mais que la fonction de répartition est celle d'une loi classique (ou encore s'exprime à partir de fonctions de répartition de lois classiques) alors on peut utiliser le paquet `scipy.stats`, dont nous allons détailler l'utilisation (rappelons que nous abrégons `scipy.stats` en `sts`).

Si l'on définit par exemple `Y=sts.norm` alors  $Y$  est une variable aléatoire de loi normale centrée réduite, et non une réalisation de cette variable aléatoire : on pourra par exemple voir ce que donne `type(Y)`. À partir de là, un certain nombre d'objets (pour être précis, je devrais ici parler de *méthodes* et non d'objets) associés à cette variable aléatoire sont accessibles : par exemple,

- `Y.pdf` est la densité d'une normale centrée réduite,
- `Y.cdf` sa fonction de répartition,
- `Y.ppf` l'inverse de la fonction de répartition (et donne donc les quantiles).

On peut accéder par `dir(Y)` à l'ensemble des méthodes associées à  $Y$ .

**Exercice 28** Vérifier de deux manières que le quantile à 97,5% est bien 1,96 approximativement, puis tracer l'allure des courbes de la densité et de la fonction de répartition d'une gaussienne centrée réduite.

#### Remarque 2.4.1

- Vous avez intérêt à retenir les syntaxes `.pdf`, `.cdf` et `.ppf` qui servent pour toutes les lois classiques.
- On a très envie, pour calculer la densité d'une variable  $X$  de la forme  $E + Z$  avec  $E$  de loi exponentielle  $\mathcal{E}(1)$  et  $Z$  de loi normale  $\mathcal{N}(0,1)$ , de définir `E=stats.expon` puis `Y=E+Z` et de faire appel à `Y.pdf`. Hélas, ça coince car même la somme  $E+Z$  n'est pas définie sous Python. On est donc obligé de faire à la main le calcul de la convolée des densités de  $E$  et  $Z$ .
- Si vous explorez tous les résultats de `dir(Y)`, vous constaterez que `Y.rvs()` donne des réalisations de la loi  $Y$ , donc d'une loi normale centrée réduite. Cette commande fait en interne appel à `rnd.normal()` (de même pour toutes les simulations d'autres lois) donc cela ne sert à rien de retenir les syntaxes `.rvs()`.

On peut à présent résumer l'illustration de la convergence en loi par le tracé de fonctions de répartitions empiriques. On procède habituellement comme suit. Si l'on connaît la fonction de répartition théorique  $F$  de la loi limite,

- on se donne une famille finie de valeurs de  $n$  (par exemple  $\{10, 100, 1000\}$ ),
- on se donne un  $\epsilon > 0$  et un niveau de confiance  $1 - \alpha > 0$  et on choisit  $N$  de telle manière que la probabilité que pour tout  $n$  de la suite, chaque  $\hat{F}_n^{(N)}$  estime  $F_n$  à  $\epsilon$  près,
- on trace les graphes de ces fonctions  $\hat{F}_n^{(N)}$  et de  $F$  : les  $\hat{F}_n^{(N)}$  doivent approcher  $F$ .

Si l'on ne connaît pas  $F$  mais que l'on sait simuler  $X$ , alors on estime  $F$  et il faut ajouter les conditions correspondantes au deuxième point ci-dessus.

**Exercice 29** *Illustrez le théorème central limite pour des sommes de variables aléatoires indépendantes  $L^2$ .*

Remarquons par ailleurs que les techniques ci-dessus peuvent également servir à identifier la loi d'une variable aléatoire. Dans l'exercice suivant, nous avons une suite de variable aléatoires qui converge presque-sûrement et cherchons à déterminer la loi de la limite. Une fois que nous avons intuité que cette loi est une loi  $\beta$ , il suffit de le vérifier.

**Exercice 30 (inspiré du texte d'agrégation public 2015-A7)** *On définit deux variables aléatoires  $A_n$  par  $A_0 = a$ ,  $B_0 = b$  et*

$$\mathbb{P}((A_{n+1}, B_{n+1} = (A_n + 1, B_n) | (A_n, B_n)) = \frac{A_n}{n + a + b}$$

$$\mathbb{P}((A_{n+1}, B_{n+1} = (A_n, B_n + 1) | (A_n, B_n)) = \frac{B_n}{n + a + b},$$

(le processus  $(A_n, B_n)_n$  est donc une urne de Pólya).

- Simulez une réalisation de la suite  $(A_n)_n$ . Semble-t-il y avoir convergence presque-sûre de  $A_n/n$ ? (si vous voulez répondre de manière théorique à la question, montrez par récurrence que  $\frac{A_n+a}{n+a+b}$  est une martingale).

24 CHAPITRE 2. CONVERGENCE DE VARIABLES ALÉATOIRES ET ILLUSTRATIONS

- *On se place dans le cas  $a = b = 1$ . Quelle loi semble suivre la limite de  $A_n/n$ ? (si vous voulez répondre de manière théorique à la question, montrez par récurrence que  $A_n + 1$  suit une loi uniforme sur  $\{1, \dots, n + 1\}$ ).*
- *Dans le cas général, montrez que la limite de  $A_n/n$  suit une loi  $\beta(a, b)$ .*



## Bibliographie

- [1] J.-F. Delmas, *Introduction au calcul des probabilités et de la statistique*.
- [2] R. Durrett, *Probability : theory and examples*.
- [3] O. Garet, A. Kurtzann, *De l'intégration aux probabilités*
- [4] J.-F. Le Gall, *Intégration, probabilités et processus aléatoires*, disponible à la page <https://www.math.u-psud.fr/~jflgall/IPPA2.pdf>
- [5] V. Rivoirard, G. Stoltz, *Statistique mathématique en action*.



## Chapitre 3

# Grands théorèmes de convergence

Le but de cette séance est de rappeler les grands théorèmes de convergence de variables aléatoires, de les discuter et illustrer. Les différentes familles de résultats de convergence sont (à chaque fois  $(a_n)_n$  désigne une suite déterministe) :

- les lois des grands nombres ou théorèmes ergodiques, qui sont du type “ $\frac{1}{a_n} \sum_{k=1}^n X_k$  converge vers une quantité déterministe”,
- les résultats du type “central limite” qui disent que “ $\frac{1}{b_n} \sum_{k=1}^n (X_k - \mathbb{E}(X_k))$  converge en loi”,
- les résultats concernant les valeurs extrêmes,
- les résultats de grande déviation,

ainsi que les résultats spécifiques aux chaînes de Markov ou aux martingales, que nous évoquerons rapidement ici et détaillerons dans les chapitres correspondants.

Nous allons à chaque fois donner des résultats sous diverses hypothèses, et discuter autant que possible les vitesses de convergence. Les références sur ces points sont, de manière générale, n’importe lequel des livres [1, 2, 4, 5]. Une lecture très intéressante sur ces points se trouve sur le blog de Terence Tao (cherchez “Terence Tao” et “275A” sous Google), et vous verrez que je lui ai beaucoup emprunté.

### 3.1 Lois des grands nombres

Dans toute la suite, on considère une suite de variables aléatoires  $(X_n)_n$ . On notera  $S_n = X_1 + \dots + X_n$ ; la loi des grands nombres concerne de manière générale le premier ordre de  $S_n$ . Les résultats classiques donnent des conditions sous lesquelles  $S_n/n$  converge vers une quantité  $\mu$  déterministe, auquel cas (si  $\mu \neq 0$ ) on a dit que  $S_n$  se comportait au premier ordre comme  $\mu n$ . On verra cependant des exemples (voir l’exercice 34) où ce premier ordre n’est pas  $O(n)$  mais par exemple  $O(n \log n)$ . Les lois des grands nombres concernent en général la convergence presque-sûre (si l’on veut être précis, on parle alors de loi forte des grands nombres) mais certains résultats ne donnent que la convergence en probabilité (on parle alors de loi faible des grands nombres).

Le résultat le plus classique est le suivant, tellement classique qu’on l’appelle en général *loi forte des grands nombres*, tout simplement :

**Théorème 3.1.1 (Kolmogorov)** Soit  $(X_n)_n$  une suite i.i.d. de variables aléatoires. Alors la suite  $(\frac{1}{n}(S_n - n\mathbb{E}(X_1)))_n$  converge presque-sûrement si et seulement si  $X_1 \in L^1$ , et alors la limite est  $\mathbb{E}(X_1)$ .

Notez que ce résultat est un “si et seulement si”. Autrement dit, aucun espoir d’avoir une convergence presque-sûre de  $S_n/n$  si  $X_1$  n’est pas intégrable. On va voir dans le Théorème 3.1.5 que l’on peut avoir une convergence en probabilité.

Il existe des variantes plus ou moins classiques à ce résultat. Le premier affaiblit l’hypothèse d’indépendance :

**Théorème 3.1.2 (Etemadi)** Soit  $(X_n)_n$  une suite de variables aléatoires deux à deux indépendantes. Alors la suite  $(\frac{1}{n}(S_n - n\mathbb{E}(X_1)))_n$  converge presque-sûrement si et seulement si  $X_1 \in L^1$ , et alors la limite est  $\mathbb{E}(X_1)$ .

Ces deux résultats peuvent être trouvés dans [1, 2, 5]. L’hypothèse d’égalité des lois peut aussi être affaiblie, par exemple en utilisant le théorème des deux séries (ou celui des trois séries) de Kolmogorov. Ce résultat étant une conséquence de la loi des grands nombres pour les martingales, que nous donnons plus loin, nous n’en parlerons pas. Rappelons par ailleurs que le théorème de Glivenko-Cantelli, que nous avons énoncé (et utilisé) dans le TP2, est une conséquence simple de cette loi des grands nombres.

Le résultat classique ce type pour les chaînes de Markov est le théorème ergodique. Nous ne le citons que pour mémoire et reviendrons dessus dans le paragraphe concernant les chaînes de Markov.

**Théorème 3.1.3** Soit  $(X_n)_n$  une chaîne de Markov sur un espace d’état dénombrable  $D$ . Si cette chaîne est irréductible et admet une probabilité invariante  $\pi$ , alors pour toute fonction  $f$  sur  $D$  intégrable par rapport à  $\pi$ , on a

$$\frac{1}{n} \sum_{k=1}^n f(X_k) \xrightarrow{\text{p.s.}} \int f d\pi.$$

Ce résultat est démontré par exemple dans [1, 2].

Enfin, on peut donner des résultats de ce type pour les martingales. Rappelons que ceci généralise le cas des sommes de variables indépendantes, puisque si  $(X_n)_n$  est une suite indépendante de variables intégrables, alors

$$M_n = \sum_{k=1}^n (X_k - \mathbb{E}(X_k))$$

est une martingale. Ces résultats dépendent de la quantité

$$\langle M \rangle_n = \sum_{k=1}^n \mathbb{E}((M_k - M_{k-1})^2 | \mathcal{F}_{k-1}).$$

On a alors le résultat suivant (voir encore une fois [1, 2]) :

**Théorème 3.1.4** Soit  $(M_n)_n$  une martingale telle que chaque  $M_n$  est de carré intégrable.

Alors :

- sur l'événement  $[\lim_n \langle M \rangle_n = \infty]$ , on a  $\frac{M_n}{\langle M \rangle_n} \xrightarrow{\text{p.s.}} 0$ ,
- sur l'événement  $[\lim_n \langle M \rangle_n < \infty]$ , la suite  $M_n$  converge presque-sûrement.

Encore une fois, nous reviendrons plus longuement dessus dans le paragraphe sur les martingales.

Dans la suite, nous nous concentrons sur les lois des grands nombres pour des sommes de variables aléatoires, et discutons des variations de ce résultat

**Exercice 31** Comment se comporte (au sens presque-sûr)  $S_n/n$  quand les  $(X_n)_n$  suivent des lois de Cauchy ? et au sens de la convergence en loi ? de la convergence en probabilité ? conjecturez le résultat, puis prouvez le (indication : la fonction caractéristique de  $X_1$  est  $t \mapsto e^{-|t|}$ ).

Le résultat ci-dessus illustre le “si et seulement si” du Théorème 3.1.1. Notons qu’il existe un analogue faible de ce théorème :

**Théorème 3.1.5** Soit  $(X_n)_n$  une suite i.i.d. de variables aléatoires. Alors la suite  $(\frac{1}{n}(S_n - n\mathbb{E}(X_1 \mathbb{1}_{|X_1| \leq n}))_n$  converge en probabilité si et seulement si  $\lim_{t \rightarrow \infty} \mathbb{P}(|X_1| \geq t) = 0$ .

Ce théorème est démontré dans [5]. Hélas, des  $X_1$  vérifiant les conditions du Théorème 3.1.5 mais pas du Théorème 3.1.1 ne sont pas facilement programmables.

Il est par ailleurs instructif de se rappeler comment prouver facilement des versions non optimales des lois des grands nombres. En effet, confronté à un problème de modélisation vous aurez souvent besoin de comprendre de manière qualitative comment se comporte votre modèle, quitte à faire des hypothèses trop fortes. L’exercice suivant vous fait retravailler des preuves utilisant la *méthode des moments*. Cela va nous permettre de montrer une loi des grands nombres pour une somme  $S_n = X_{1,n} + \dots + X_{n,n}$  où les  $X_{i,n}$  sont indépendants pour  $n$  fixé mais que les termes apparaissant dans  $S_n$  et ceux qui apparaissent dans  $S_{n'}$  ne le sont pas.

**Exercice 32** Vous rappelez-vous comment montrer la loi faible des grands nombres pour une famille i.i.d.  $L^2$  ? et la loi forte des grands nombres pour une famille i.i.d.  $L^4$  ? En déduire que si  $(X_{i,n})_{i \leq n}$  est une famille de variables aléatoires qui ont toutes la même espérance  $\mathbb{E}(X_{1,1})$  et telle que pour tout  $n$

$$X_{1,n}, \dots, X_{n,n} \text{ est une famille indépendante et } \sup_{i,n} \mathbb{E}(X_{i,n}^p) < \infty$$

alors  $S_n/n = \frac{1}{n} \sum_{k=1}^n X_{k,n}$  converge vers  $\mathbb{E}(X_{1,1})$  en probabilité si  $p = 2$  et presque-sûrement si  $p = 4$ .

**Exercice 33** On fabrique une suite  $(V_n, E_n)_{n \in \mathbb{N}^*}$  de graphes aléatoires de la manière suivante :

- l'ensemble des sommets  $V_n$  est  $\{1, \dots, n\}$  pour tout  $n$ ,
- l'ensemble d'arêtes  $E_{n+1}$  est donné comme la réunion de  $E_n$  avec des arêtes  $\{i, n+1\}$ , chacune d'entre elles étant présente avec probabilité  $1/2$  et de manière indépendante.

Écrire une fonction Python simulant cette suite. Utilisez cette fonction pour montrer que  $\#E_n / \binom{n}{2}$  converge presque-sûrement vers  $1/2$ . On peut montrer également que le nombre  $T_n$  de triangles (i.e. le nombre de triplets  $\{i, j, k\}$  de sommets tels que les arêtes  $\{i, j\}$ ,  $\{i, k\}$ ,  $\{j, k\}$  sont toutes dans  $V$ ) vérifie que  $T_n / \binom{n}{3}$  converge presque-sûrement vers  $1/8$ .

La méthode des moments permet également de traiter le cas où  $S_n$  n'évolue pas au premier ordre en  $n$ . L'exercice suivant est un classique appelé le “collectionneur de vignettes<sup>1</sup>”.

**Exercice 34** Soit  $(Y_k)_k$  une famille i.i.d. de variables de loi uniforme sur  $\{1, \dots, n\}$ . On note  $T_n$  le premier instant où les  $Y$  ont pris toutes les valeurs possibles :

$$T_n = \inf \{k \text{ tel que } \#\{Y_1, \dots, Y_k\} = k\}.$$

Écrire une fonction Python simulant  $T_n$ ; on aura intérêt à utiliser la structure de données appelée “ensemble”. Explorez le comportement en loi de  $T_n$  pour  $n \rightarrow \infty$ . On pourra en particulier expérimenter pour voir si  $T_n / (n^\alpha \log^\beta n)$  semble converger, pour des valeurs simples de  $\alpha, \beta$ . Conjecturez un résultat, puis prouvez le. Pour cela on pourra montrer que  $T_n = X_{n,1} + \dots + X_{n,n}$  pour des variables aléatoires indépendantes de loi géométrique, calculer l'espérance de  $T_n$  et majorer sa variance.

**Exercice 35** On choisit une permutation  $\sigma_n$  aléatoirement et de manière uniforme dans  $\mathfrak{S}_n$  et on s'intéresse au nombre  $C_n$  de cycles dans cette permutation. On codera une permutation de la manière suivante :

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 3 & 9 & 6 & 8 & 2 & 1 & 5 & 4 & 7 \end{pmatrix}$$

est représenté par `sigma=[3, 9, 6, 8, 2, 1, 5, 4, 7]`. On vous fournit la fonction `to_cycles` qui prend en entrée `sigma` et rend une décomposition en cycles. Observez empiriquement que l'espérance et la variance de  $C_n$  sont de l'ordre de  $\log n$  (on peut prouver en fait que  $C_n$  a même loi que la somme de  $X_{n,k}$  pour  $k = 1, \dots, n$  où  $X_{n,k} \sim \mathcal{B}(\frac{1}{n-k+1})$ ). En acceptant le résultat de cette observation, montrez que  $C_n / \log n$  tend vers 1 en probabilité. L'illustrer est assez difficile à cause de la lenteur de la convergence.

La méthode des moments permet souvent d'obtenir des résultats sous des hypothèses renforcées de moments. Une fois les résultats identifiés, on peut espérer assouplir les hypothèses, typiquement par un argument de troncation.

1. le terme anglais étant “coupon collector”, on a tendance à parler de “collectionneur de coupons”

Un point que nous n'avons pas encore abordé est celui de la vitesse de convergence dans les lois des grands nombres. Celle-ci est essentiellement donnée par un résultat du type "théorème central de la limite<sup>2</sup>" que nous étudions maintenant.

### 3.2 Théorèmes centraux de la limite

Nous avons vu que la loi des grands nombres donne le premier ordre pour  $S_n = X_1 + \dots + X_n$  :  $S_n/n \rightarrow \mu$  donne  $S_n \simeq \mu n$ . Le théorème central limite donnera le deuxième ordre, au sens où il donne des informations sur  $S_n - \mu n$  (qui ne peut être que d'un ordre inférieur à  $n$ ). Le résultat standard que nous rappelons ci-dessous donne par exemple  $S_n \simeq \mu n + Z\sqrt{n}$  où  $Z \sim \mathcal{N}(0, 1)$ .

Ici aussi, nous allons discuter l'énoncé standard et les extensions que l'on peut obtenir en améliorant ses différentes méthodes de preuve. Nous allons aussi discuter la vitesse de convergence.

Le résultat standard est le suivant. Nous allons rappeler plusieurs méthodes de preuve.

**Théorème 3.2.1** *Si  $(X_n)_n$  est une suite i.i.d. de carré intégrable, alors*

$$\frac{S_n - n\mathbb{E}(X_1)}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \text{var}(X_1))$$

Nous allons voir une première illustration élémentaire, qui servira plus tard :

**Exercice 36** *Soit  $(X_n)_n$  une suite de variables i.i.d. de loi de Bernoulli  $\mathcal{B}(p)$ . Illustrez le théorème ci-dessus.*

La preuve la plus courante et la plus courte utilise **les fonctions caractéristiques** et le théorème de Lévy-Cramér. On rappelle que si  $Z$  est une variable aléatoire (réelle ici, mais ce qui suit s'étend au cas de variables à valeurs dans  $\mathbb{R}^d$  par exemple), sa fonction caractéristique est

$$\Phi_Z : \mathbb{R} \rightarrow \mathbb{C} \\ t \mapsto \mathbb{E}(e^{itZ}) .$$

On écrit alors

$$\Phi_{\frac{S_n - n\mathbb{E}(X_1)}{\sqrt{n}}}(t) = e^{-it\sqrt{n}\mathbb{E}(X_1)} \left( \Phi_{X_1}\left(\frac{t}{\sqrt{n}}\right) \right)^n$$

et

$$\Phi_{X_1}\left(\frac{t}{\sqrt{n}}\right) = \Phi_{X_1}(0) + \frac{t}{\sqrt{n}} \Phi'_{X_1}(0) + \frac{t^2}{2n} \Phi''_{X_1}(0) + o\left(\frac{t^2}{2n}\right)$$

or  $\phi_{X_1}(0) = 1$ ,  $\Phi'_{X_1}(0) = i\mathbb{E}(X_1)$ ,  $\Phi''_{X_1}(0) = -\mathbb{E}(X_1^2)$  et on obtient immédiatement

$$\Phi_{\frac{S_n - n\mathbb{E}(X_1)}{\sqrt{n}}}(t) \rightarrow e^{-\frac{t^2}{2\text{var}(X_1)}} .$$

Cette preuve est la plus courte mais elle n'est pas celle qui s'étend le plus facilement.

---

2. oui, c'est le terme correct

Une autre preuve utilise la **méthode de Lindeberg**. On l'écrit ci-dessous dans le cas où les  $X_k$  sont centrés et de variance 1 (ce qui ne restreint pas la portée de la preuve car on peut s'y ramener en translatant et normalisant les  $X_k$ ), et admettent un troisième moment (on peut ensuite par troncation qu'il suffit d'avoir un deuxième moment). Elle consiste à remarquer pour commencer que si  $(N_n)_n$  est une famille i.i.d. de loi  $\mathcal{N}(0, 1)$  alors  $\frac{N_1 + \dots + N_n}{\sqrt{n}}$  suit exactement la loi  $\mathcal{N}(0, 1)$ . Ensuite, pour  $\varphi$  une fonction infiniment dérivable sur  $\mathbb{R}$ , on écrit :

$$\begin{aligned} & \mathbb{E}\left(\varphi\left(\frac{X_1 + \dots + X_n}{\sqrt{n}}\right)\right) - \mathbb{E}\left(\varphi\left(\frac{N_1 + \dots + N_n}{\sqrt{n}}\right)\right) \\ &= \sum_{k=1}^n \mathbb{E}\left(\varphi\left(\frac{X_1 + \dots + X_k + N_{k+1} + \dots + N_n}{\sqrt{n}}\right) - \varphi\left(\frac{X_1 + \dots + X_{k-1} + N_k + \dots + N_n}{\sqrt{n}}\right)\right) \end{aligned}$$

puis on donne pour chaque terme de la somme un développement de Taylor au troisième ordre. L'avantage de cette méthode est qu'elle permet plus facilement d'arriver à des variations du théorème, dans lesquelles on s'intéresse à une somme de  $(\tilde{X}_{n,k})_k$  indépendants pour  $k$  fixé mais pas pour différents  $n$ . On arrive alors au raffinement qui est le sens direct du théorème suivant :

**Théorème 3.2.2 (Lindeberg-Feller)** *Soit  $K_n$  une suite d'entiers avec  $K_n \rightarrow \infty$  quand  $n \rightarrow \infty$ , et soit pour tout  $n$  une famille  $\tilde{X}_{n,1}, \dots, \tilde{X}_{n,K_n}$  de variables aléatoires indépendantes. On note  $\tilde{\sigma}_n^2 = \sum_{k=1}^{K_n} \text{var}(\tilde{X}_{n,k})$ ; alors la condition*

$$\forall \epsilon > 0 \quad \frac{1}{\tilde{\sigma}_n^2} \sum_{k=1}^{K_n} \mathbb{E}\left(\left(\tilde{X}_{n,k} - \mathbb{E}(\tilde{X}_{n,k})\right)^2 \mathbb{1}_{|\tilde{X}_{n,k} - \mathbb{E}(\tilde{X}_{n,k})| > \epsilon \tilde{\sigma}_n}\right) \xrightarrow{n \rightarrow \infty} 0 \quad (3.1)$$

est vérifiée si et seulement si on a

$$\frac{1}{\tilde{\sigma}_n^2} \max_{1 \leq k \leq K_n} \text{var}(\tilde{X}_{n,k}) \xrightarrow{n \rightarrow \infty} 0 \quad (3.2)$$

et

$$\frac{1}{\tilde{\sigma}_n} \sum_{k=1}^{K_n} (\tilde{X}_{n,k} - \mathbb{E}(\tilde{X}_{n,k})) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

### Remarque 3.2.3

- On retrouve le résultat standard donné par le Théorème 4.1.2 en prenant  $\tilde{X}_{n,k} = X_k/\sqrt{n}$  (c'est pour souligner le fait que les  $\tilde{X}_{n,k}$  contiennent une normalisation que nous utilisons la notation  $\tilde{X}$ ).
- La condition (3.1) est appelée "condition de Lindeberg". Elle signifie qu'aucun  $X_{n,k}$  ne joue un rôle prédominant dans la somme  $\sum_{k=1}^{K_n} (\tilde{X}_{n,k} - \mathbb{E}(\tilde{X}_{n,k}))$ .
- On peut bien sûr démontrer ce résultat en utilisant les fonctions caractéristiques, mais c'est moins naturel qu'avec la méthode de Lindeberg.



**Exercice 37** On utilise les notations de l'exercice 35. On pose  $\tilde{X}_{n,k} = \frac{X_{n,n-k} - \frac{1}{k}}{\sqrt{\log n}}$ ; en utilisant le Théorème 3.2.2, montrez que  $\frac{C_n - \log n}{\sqrt{\log n}}$  converge en loi vers une variable centrée réduite puis illustrez ce résultat.

**Remarque 3.2.4** Une condition suffisante pour (3.1) est qu'il existe  $\delta > 0$  tel que

$$\frac{1}{\tilde{\sigma}_n^{2+\delta}} \sum_{k=1}^n \mathbb{E}(|\tilde{X}_{n,k} - \mathbb{E}(\tilde{X}_{n,k})|^{2+\delta}) = 0. \quad (3.3)$$

Cette condition (3.3) est appelée "condition de Lyapounov".

**Exercice 38** On reprend l'exercice 34 mais cette fois-ci le collectionneur ne cherche qu'à obtenir une proportion  $\rho \in ]0, 1[$  des vignettes disponibles. Montrez que le temps  $T_n^{(\rho)}$  peut s'écrire comme  $X_{n,1} + \dots + X_{n,r_n}$  avec  $r_n \sim \rho n$ . Utilisez le Théorème 3.2.2 pour montrer que pour tout  $\rho \in ]0, 1[$  il existe deux constantes  $m_\rho$  et  $\sigma_\rho^2$  pour lesquelles  $\frac{T_n^{(\rho)} - m_\rho n}{\sqrt{\sigma_\rho^2 n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$ . On pourra prouver que la condition (3.3) est vérifiée avec  $\delta = 2$  en utilisant l'inégalité<sup>3</sup>

$$\mathbb{E}\left(|G_p - \frac{1}{p}|^4\right) \leq \frac{K}{p^4}$$

pour  $G_p$  suivant une loi géométrique de paramètre  $p$ , et  $K$  une certaine constante (indépendante de  $p$ ). Et pour  $\rho = 1$ ? On rappelle que  $E(T_n) \sim n \log n$  et que  $\text{var}(T_n) \sim \frac{\pi^2}{6} n^2$ ; a-t-on convergence en loi de  $\frac{T_n - n \log n}{\pi n / \sqrt{6}}$ ?

La troisième méthode de preuve du théorème central limite passe encore par **les moments**. Elle utilise le résultat suivant : si une suite de variables aléatoires  $(Y_n)_n$  vérifie  $\mathbb{E}(Y_n^p) \rightarrow \mathbb{E}(Z^p)$  pour tout  $p \in \mathbb{N}$ , où  $Z \sim \mathcal{N}(0, 1)$  alors  $Y_n \xrightarrow{\mathcal{L}} Z$  (et on peut remplacer ici  $\mathcal{N}(0, 1)$  par toute loi "déterminée par ses moments", i.e. qui est la seule avec les moments donnés — et c'est le cas de la loi normale). *Mea culpa* : je n'ai pas trouvé d'application raisonnablement simple et intéressante.

On verra dans les chapitres concernant martingales et chaînes de Markov qu'il existe des théorèmes centraux de la limite pour ce type de processus.

Parlons maintenant de vitesse de convergence dans le théorème central limite. Le résultat principal est le théorème suivant :

**Théorème 3.2.5 (Berry-Esséen)** Soit  $(X_n)_n$  une suite de variables aléatoires i.i.d. admettant des moments d'ordre 3. On note  $F_n$  la fonction de répartition de  $\frac{S_n - n\mathbb{E}(X_1)}{\sqrt{n \text{var}(X_1)}}$  et  $F$  la fonction de répartition de  $\mathcal{N}(0, 1)$ . Alors

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq \frac{C\rho}{\sigma^3 \sqrt{n}}$$

où  $C \simeq 0,4748$  est une constante universelle et  $\rho = \mathbb{E}(|X_1 - \mathbb{E}(X_1)|^3)$ .

3. facile à montrer.

Ce résultat montre donc essentiellement que la vitesse de convergence dans le théorème central limite, Théorème 3.1.1, est en  $O(1/\sqrt{n})$ .

**Exercice 39** On reprend le contexte et le code de l'exercice 36. On note  $F_n^{(N)}$  la fonction de répartition empirique approchant  $F_n$  associée au  $N$ -échantillon. En utilisant le fait que  $\sup_{x \in \mathbb{R}} |F_n^{(N)}(x) - F(x)|$  est atteint en l'un des points de l'échantillon (faites un dessin!), donner une estimation de cette quantité pour différentes valeurs de  $n$  et la comparer au majorant donné dans le Théorème 3.2.5.

Avant de passer à un autre sujet, mentionnons le fait que dans le Théorème 4.1.2, on ne peut renforcer le résultat pour avoir une convergence en probabilité ou presque-sûre. Cela peut se prouver de plusieurs manières mais nous utiliserons pour cela un résultat appelé *loi du log itéré*.

**Théorème 3.2.6** Soit  $(X_n)_n$  une suite i.i.d. de carré intégrable, et

$$V_n = \frac{1}{\sqrt{2n \log \log n}} (S_n - n\mathbb{E}(X_1)).$$

Alors presque-sûrement

$$\limsup_{n \rightarrow \infty} V_n = +\sqrt{\text{var}(X_1)} \quad \liminf_{n \rightarrow \infty} V_n = -\sqrt{\text{var}(X_1)}.$$

**Exercice 40**

1. En admettant le Théorème 3.2.6, montrez que l'on ne peut avoir convergence en probabilité ou presque-sûre dans le Théorème 4.1.2, à moins que les variables  $X$  ne soient déterministes.
2. On voudrait illustrer le Théorème 3.2.6; c'est impossible car les échelles de temps sont trop lentes. On peut en revanche observer que  $V_n$  prend des valeurs "éloignées de zéro" pour des temps arbitrairement grand. Simulez une trajectoire de  $V_n$  pour des  $X_n$  de loi de Rademacher.

### 3.3 Valeurs extrêmes

Soit encore  $(X_n)_n$  une famille i.i.d. de variables réelles. Les résultats du type "central limite" portaient sur la convergence en loi de  $(S_n - a_n)/b_n$  pour des suites  $(a_n)_n$  et  $(b_n)_n$ . Nous nous intéressons ici à la variable donnant les maxima de la suite :

$$M_n = \max_{k=1, \dots, n} X_k$$

et à l'éventuelle convergence en loi de  $(M_n - a_n)/b_n$  quand  $n \rightarrow \infty$ .

On peut commencer par se demander quel est le comportement de  $M_n$  lui-même.

**Exercice 41** Soit  $x_F = \sup\{x \in \mathbb{R} \mid \mathbb{P}(X_1 \leq x) < 1\} \in ]-\infty, +\infty]$ . On a  $M_n \xrightarrow{\mathbb{P}} x_F$ .

La question de la convergence de  $(M_n - a_n)/b_n$  consiste alors à étudier le comportement de  $x_F - M_n$  (si  $x_F < \infty$ ) ou la vitesse de divergence de  $M_n$  (si  $x_F = \infty$ ). Le résultat standard en la matière est le suivant :

**Théorème 3.3.1 (Fisher-Fréchet-Gnedenko-Tippett)** *Soit  $(X_n)_n$  une famille i.i.d. de variables réelles. S'il existe deux suites,  $(a_n)_n$  de réels quelconques et  $(b_n)_n$  de réels strictement positifs, telles que  $(M_n - a_n)/b_n$  converge en loi quand  $n \rightarrow \infty$ , alors cette loi limite est nécessairement de l'un des quatre types suivants :*

- une masse de Dirac,
- une loi de Gumbel, i.e. de fonction de répartition  $x \mapsto e^{-e^{-x}}$ ,
- une loi de Weibull<sup>4</sup>, i.e. de fonction de répartition  $x \mapsto e^{-(-x)^a} \mathbf{1}_{\mathbb{R}_-}(x) + \mathbf{1}_{\mathbb{R}_+}(x)$  avec  $a > 0$ ,
- une loi de Fréchet, i.e. de fonction de répartition  $x \mapsto e^{-x^{-a}} \mathbf{1}_{\mathbb{R}_+}(x)$  avec  $a > 0$ .

Nous n'allons pas démontrer ce résultat mais l'illustrer dans des cas correspondant à différentes situations.

**Exercice 42** *Dans les cas suivants, prouvez puis illustrez le résultat.*

1. Si  $X_1 \sim \mathcal{B}(p)$  avec  $p \in ]0, 1[$ , on a  $M_n \xrightarrow{\mathcal{L}} \delta_1$ .
2. Si  $X_1 \sim \mathcal{U}([0, \theta])$  pour  $\theta > 0$ , on a  $(M_n - \theta)/\frac{\theta}{n} \xrightarrow{\mathcal{L}}$  une loi de Weibull avec  $a = 1$  (qui est la même chose que la loi de  $-E$  pour  $E \sim \mathcal{E}(1)$ ).
3. Si  $X_1 \sim \mathcal{E}(\lambda)$  avec  $\lambda > 0$ , on a  $(M_n - \frac{\log n}{\lambda})/\frac{1}{\lambda} \xrightarrow{\mathcal{L}}$  une loi de Gumbel.
4. Si  $X_1 \sim \mathcal{C}(1)$  alors  $M_n/\frac{n}{\pi} \xrightarrow{\mathcal{L}}$  une loi de Fréchet avec  $a = 1$ .

---

4. c'est la loi de  $-X$  pour  $X$  une loi de Weibull définie précédemment ; les deux conventions de notation existent.



## Bibliographie

- [1] Ph. Barbe, M. Ledoux, *Probabilité*.
- [2] M. Benaïm, N. El Karoui, *Promenade aléatoire, chaînes de Markov et simulations; martingales et stratégies*.
- [3] B. Bercu, D. Chafaï, *Modélisation stochastique et simulation*.
- [4] O. Garet, A. Kurtzmann, *De l'intégration aux probabilités*.
- [5] P. Toulouse, *Thèmes de probabilités et statistiques*.



## Chapitre 4

# Utilisation des chaînes de Markov pour la simulation

Dans cette séance, nous abordons l'utilisation des chaînes de Markov pour la résolution de problèmes numériques. Nous commencerons par nous intéresser à la méthode de Monte-Carlo qui permet de calculer des intégrales  $\int f d\pi$  lorsque  $f$  est une fonction sur un ensemble dénombrable  $E$  et que l'on connaît une chaîne de Markov sur  $E$  de probabilité invariante  $\pi$ . Ce problème va nous mener à étudier deux questions théoriques, concernant :

- la construction d'une chaîne de Markov ayant une probabilité invariante donnée *a priori*,
- la vitesse de convergence d'une chaîne de Markov vers sa mesure d'équilibre.

Les réponses à ces deux questions nous permettront de répondre à un autre problème numérique, qui est la recherche des minima d'une fonction  $V$  sur  $E$  par un algorithme stochastique.

Les notations utilisées dans l'ensemble du texte seront les suivantes :  $(X_n)_n$  sera toujours une chaîne de Markov de matrice de transition  $(P(x, y))_{x, y \in E}$  sur un ensemble  $E$  de cardinal (fini)  $d$ , avec la convention  $P(x, y) = \mathbb{P}(X_{n+1} = y | X_n = x)$ , de sorte que si  $\mu_0 = (\mu_0(x))_{x \in E}$  (vu comme un vecteur ligne) est la loi de  $X_0$  alors  $\mu_n = \mu_0 P^n$  est la loi de  $X_n$ . De même, si  $f$  est une fonction sur  $E$  alors  $Pf$  sera la fonction

$$Pf(x) = \sum_{y \in E} P(x, y)f(y) = \mathbb{E}_x(f(X_1)).$$

On notera  $\mathbb{P}_x$  (resp.  $\mathbb{P}_\mu$ ) les probabilités conditionnelles à  $X_0 = x$  p.s. (resp. sous l'hypothèse que  $X_0$  suit la loi  $\mu$ ), de même pour  $\mathbb{E}_x$  (resp.  $\mathbb{E}_\mu$ ).

Nous commençons avec un paragraphe rapide sur la simulation des chaînes de Markov et sur des théorèmes sur le comportement asymptotique de ces chaînes.

### 4.1 Simulation et résultats classiques pour les chaînes de Markov

On cherche ici à simuler une chaîne de Markov décrite par sa matrice de transition  $T$ , qui est définie par le fait que  $T_{i,j} = \mathbb{P}(X_{n+1} = i | X_n = j)$ .

Rappelons qu'une chaîne de Markov peut également être donnée sous la forme  $X_{n+1} = f_n(X_n, U_{n+1})$  où  $(U_n)_n$  est une suite i.i.d. de loi  $\mathcal{U}([0, 1])$  indépendante de

$X_0$ , cette dépendance étant en général exprimée sous la forme d'un algorithme impliquant le tirage de  $U_{n+1}$ , comme c'est le cas dans l'algorithme de Metropolis que nous verrons plus loin.

Supposons que la matrice de transition est donnée sous la forme d'une liste de listes, par exemple  $T = [[1/6, 1/3, 1/2], [1/2, 1/4, 1/4], [1/2, 1/2, 0]]$  pour la matrice  $T = \begin{pmatrix} 1/6 & 1/3 & 1/2 \\ 1/2 & 1/4 & 1/4 \\ 1/2 & 1/2 & 0 \end{pmatrix}$ . Dans ce cas, et si l'on indexe (suivant l'habitude de Python) les trois sommets comme 0, 1 et 2 alors conditionnellement à  $X_n = i$  la variable  $X_{n+1}$  vaut 0, 1 ou 2 avec des probabilités données par  $T[i]$ . On peut alors utiliser la commande `rnd.choice(d, T[i])` où  $d$  est le cardinal de l'espace d'état (ici  $d = 3$ ).

**Exercice 43** *Écrivez une fonction Python `Markov` qui prend en entrée  $X_0$ ,  $n$  et  $T$  et simule une trajectoire  $(X_0, \dots, X_n)$ .*

On rappelle qu'il existe un résultat de type "loi des grands nombres" pour les chaînes de Markov : c'est le théorème ergodique.

**Théorème 4.1.1** *Soit  $(X_n)_n$  une chaîne de Markov sur un espace d'état dénombrable  $E$ . Si cette chaîne est irréductible et admet une probabilité invariante  $\pi$ , alors pour toute fonction  $f$  sur  $D$  intégrable par rapport à  $\pi$ , on a*

$$\frac{1}{n} \sum_{k=1}^n f(X_k) \xrightarrow{\text{p.s.}} \int f d\pi.$$

Notons que l'irréductibilité assure que la probabilité invariante est unique si elle existe. Dans le cas où  $D$  est fini, il suffit de supposer que la chaîne est irréductible, l'existence de  $\pi$  étant alors automatique. Ce résultat est démontré par exemple dans [1, 2].

Il existe également un résultat de type "théorème central de la limite" :

**Théorème 4.1.2** *Soit  $(X_n)_n$  une chaîne de Markov sur un espace d'état fini  $D$ . Si cette chaîne est irréductible et qu'on note  $\pi$  sa probabilité invariante, alors pour toute fonction  $f$  sur  $D$ , il existe  $\sigma_f^2$  tel que*

$$\sqrt{n} \left( \frac{1}{n} \sum_{k=1}^n f(X_k) - \int f d\pi \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_f^2).$$

Le gros défaut pratique de cet énoncé est que  $\sigma_f$  n'est défini que de manière implicite, de sorte que ce résultat ne peut servir pour donner des intervalles de confiance. Pour des énoncés généraux (dont celui-ci) voir [2].

**Exercice 44** *On considère la même chaîne que dans l'exercice précédent. Comment déterminer la probabilité invariante  $\pi$  de cette chaîne ? On utilisera `np.linalg.eig`, dont on pourra consulter le fichier d'aide ; ne pas oublier que la probabilité invariante est vecteur propre de la transposée de  $T$  et pas de  $T$  même. Illustrer ensuite les deux théorèmes ci-dessus dans le cas de la chaîne de matrice de transition définie ci-dessus et pour  $f = \mathbb{1}_{\{0\}}$ . La variance  $\sigma_f$  étant inconnue, on l'estimera par la variance empirique d'un  $N$ -échantillon de  $X_n$  pour  $n$  assez grand.*



## 4.2 Méthodes de Monte-Carlo

Le principe général de la méthode de Monte-Carlo est le suivant : si  $(X_n)_n$  est une chaîne de Markov irréductible sur un ensemble fini  $E$ , qui admet une probabilité invariante  $\pi$  sur  $E$  (ce qui est toujours le cas si  $E$  est fini), alors pour toute fonction  $f$  sur  $E$ , le théorème ergodique donne :

$$\frac{1}{n} \sum_{k=1}^n f(X_k) \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \int f \, d\pi. \quad (4.1)$$

On peut donc espérer calculer  $\int f \, d\pi$  en simulant une trajectoire d'une chaîne de Markov de probabilité invariante  $\pi$ . Deux questions se posent naturellement :

- puisque l'objet de départ est  $\pi$ , sait-on construire une chaîne de Markov (irréductible) de probabilité invariante  $\pi$  ?
- quelle est la vitesse de convergence dans (4.1) ?

Remarquons au sujet du premier point que la méthode ci-dessus ne sera intéressante que s'il est moins coûteux numériquement de simuler la chaîne que de calculer tous les  $\pi(x)$  et de sommer les  $f(x)\pi(x)$ . Nous allons décrire dans la section 4.3 une méthode, dite de Metropolis–Hastings, de simulation d'une chaîne de Markov de probabilité invariante  $\pi$ , qui aura cette qualité.

Au sujet du deuxième point le théorème 4.1.2 devrait nous permettre de donner au moins un intervalle de confiance asymptotique pour  $\int f \, d\pi$ , malheureusement ce résultat est en pratique difficile à exploiter. On va donc renforcer nos hypothèses pour avoir une estimation de  $|P^n f(x) - \int f \, d\pi|$ . Si cette quantité est  $\leq \epsilon_n$ , alors si l'on se donne  $N$  réalisations indépendantes  $X_n^{(1)}, \dots, X_n^{(N)}$  de  $X_n$ , le théorème central limite appliqué à  $f(X_n^{(k)})_k$  implique

$$\sqrt{N} \left( \frac{1}{N} \sum_{k=1}^N f(X_n^{(k)}) - \int f \, d\mu_n \right) \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, \sigma_{\text{emp}}^2) \quad (4.2)$$

où cette fois-ci

$$\sigma_{\text{emp}}^2 = \int (f(x) - \int f \, d\mu_n)^2 \, d\mu_n \leq 4\|f\|_{\infty}^2$$

et on peut donc obtenir un intervalle de confiance asymptotique pour  $\int f \, d\mu_n$ , qui est  $\epsilon_n$ -proche de  $\int f \, d\pi$  (mais au prix de  $N$  simulations de  $X_n$ , au lieu de la seule simulation de  $X_1, \dots, X_n$ ).

Nous donnons dans la section 4.3 l'algorithme de Metropolis–Hastings, méthode de construction d'une chaîne de Markov de mesure invariante  $\pi$  ; on aura même pour  $\pi$  une propriété plus forte appelée réversibilité. Dans la section 4.4, nous définissons une famille de probabilités  $\pi$ , les mesures de Gibbs, qui a un intérêt tout particulier pour les applications, nous explicitons la méthode de Metropolis–Hastings dans le cas de cette famille, et l'utilisons pour obtenir notre première vraie simulation. Nous donnons dans la section 4.5 des estimations de  $|P^n f(x) - \int f \, d\pi|$ , sous l'hypothèse que la chaîne est (ça tombe bien) réversible pour la probabilité  $\pi$ . Dans la section 4.6, nous explicitons ces estimations dans le cas des mesures de Gibbs et en tirons la méthode du recuit simulé, qui permet de trouver les minima d'une fonction.

### 4.3 Algorithme de Metropolis–Hastings

Dans cette section, nous allons donner une méthode permettant de construire une matrice de transition  $P$  qui sera réversible pour une matrice donnée  $\pi$ . Commençons par rappeler la définition de la réversibilité :

**Définition 4.3.1** *On dit que la matrice de transition  $(P(x, y))_{x, y}$  est réversible par rapport à  $\pi$  si pour tous  $x, y$  de  $E$  on a*

$$P(x, y) \pi(x) = P(y, x) \pi(y). \quad (4.3)$$

Il est immédiat que si  $P$  est réversible par rapport à  $\pi$ , alors  $\pi$  est invariante pour  $P$ . Plus précisément, l'invariance de  $\pi$  est caractérisée par l'égalité

$$\sum_{y \in E} P(x, y) \pi(x) = \sum_{y \in E} P(y, x) \pi(y),$$

donc l'invariance de  $\pi$  par  $P$  correspond à l'égalité "en moyenne" de  $P_{x, y} \pi(x)$  alors que la réversibilité correspond à une égalité terme à terme. Ceci explique que (4.3) soit aussi appelée la condition de bilan détaillé.

On part d'une probabilité  $\pi$  dont on suppose qu'elle charge tous les points :  $\pi(x) > 0$  pour tout  $x$ , et on suppose donnée une matrice de transition  $Q$ , dite matrice de sélection, qui a la propriété  $Q(x, y) = 0 \Rightarrow Q(y, x) = 0$ . Pour cette matrice, on considère une fonction  $h : \mathbb{R}_+^* \rightarrow ]0, 1]$  telle que  $h(u) = uh(1/u)$  (remarquons que  $h(u) = \min(1, u)$  ou  $h(u) = \frac{u}{1+u}$  conviennent) et on pose

$$\alpha(x, y) = h\left(\frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)}\right)$$

(avec la convention que  $\alpha(x, y) = 0$  si  $Q(x, y) = 0$ ).

**Lemme 4.3.2** *On définit*

$$P(x, y) = \begin{cases} \alpha(x, y)Q(x, y) & \text{si } x \neq y, \\ 1 - \sum_{y \neq x} P(x, y) & \text{sinon.} \end{cases} \quad (4.4)$$

*Alors  $P$  est un noyau de transition qui est réversible pour  $\pi$ , irréductible si  $Q$  l'est, et aperiodique si  $h(u) < 1$  pour tout  $u$ .*

La preuve est laissée en exercice :

**Exercice 45** *Prouvez le Lemme 4.3.2.*

La matrice de transition donnée ci-dessus correspond à un algorithme d'évolution très simple : si l'on a  $X_n = x$ , alors le choix de la valeur de  $X_{n+1}$  se fait de la manière suivante :

1. on choisit  $y$  suivant la loi  $(Q(x, y))_{y \in E}$ ,
2. on calcule  $A(x, y) = h\left(\frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)}\right)$ ;

3. on tire un  $U$  de loi  $\mathcal{U}([0, 1])$ ;
  - si  $U \leq A(x, y)$  alors on accepte la sélection et on choisit  $X_{n+1} = y$ ;
  - sinon on refuse la sélection et on choisit  $X_{n+1} = x$ .

Cet algorithme simple permet donc de simuler une chaîne de Markov qui va être réversible par rapport à  $\pi$ . On verra justement dans la section 4.5 que l'estimation des vitesses de convergence est plus simple pour les chaînes réversibles.

**Remarque 4.3.3** *On n'a pas besoin d'explicitier la matrice  $Q$  : il suffit de savoir définir un algorithme qui simule la chaîne de transitions données par  $Q$ , et de connaître les rapports  $Q(x, y)/Q(y, x)$ . On n'a pas besoin non plus de connaître explicitement  $\pi(x)$  : il suffit de connaître les rapports  $\pi(x)/\pi(y)$ , donc les constantes de normalisation ne jouent aucun rôle.*

**Exercice 46** Soit  $V = (\mathbb{Z}/r\mathbb{Z})^d$  un réseau  $d$ -dimensionnel (on considère  $(\mathbb{Z}/r\mathbb{Z})^d$  plutôt que  $\{1, \dots, r\}^d$  pour se simplifier un peu la vie, vous comprendrez bientôt pourquoi). On appelle configuration de sphères dures sur  $S$  une application  $x : V \rightarrow \{0, 1\}$  telle que  $x(v) \neq x(v')$  si  $v$  et  $v'$  sont voisins. La situation  $x(v) = 1$  décrit la présence en  $v$  d'une "sphère" qui est assez grosse pour empêcher la présence d'une autre sphère sur les sites voisins (mais pas sur les sites plus lointains). On note  $M$  l'ensemble des configurations de sphères dures sur  $V$ , et  $\pi$  la probabilité uniforme sur  $M$ . On considère l'algorithme suivant, définissant une configuration  $x_n$  aléatoire : si l'on a  $X_n = x$ , alors

1. on choisit  $v$  uniformément dans  $V$ ,
2. si les sites voisins de  $v$  ne sont pas tous libres, on ne fait rien,
3. si tous les sites voisins de  $v$  sont libres, on donne à  $x(v)$  la valeur d'une variable de Bernoulli  $\mathcal{B}(1/2)$ .

Montrez que ceci est un algorithme de Metropolis pour  $h(u) = \frac{u}{1+u}$  et définit donc une chaîne de Markov irréductible apériodique de mesure invariante  $\pi$ . En déduire une estimation numérique du nombre moyen de sphères  $\mathbb{E}_\pi(\sum_{v \in V} x(v))$  (en utilisant des  $N = 100$ -échantillons avec  $n = 10\,000$  pour  $r = 10$ ,  $d = 2$ ). Pour cela on aura intérêt à définir une fonction  $\epsilon$  qui à  $k$  associe le  $k$ -ième vecteur de la base canonique.

## 4.4 Mesures de Gibbs

Soit  $V$  une fonction de  $E$  dans  $\mathbb{R}$ , et  $T > 0$  un paramètre appelé la température. On appelle mesure de Gibbs associée à la fonction  $V$ , à la température  $T$ , la probabilité définie par

$$\pi_{V,T}(x) = \frac{1}{Z} e^{-V(x)/T}, \quad (4.5)$$

où  $Z$  est une constante de normalisation :

$$Z = \sum_{x \in E} e^{-V(x)/T}.$$

L'une des difficultés de la simulation directe des lois de Gibbs réside dans la constante  $Z$ , *a priori* inconnue et dont le calcul est en pratique impossible – ou en tout cas aussi compliqué que le calcul de  $\int f d\pi_{V,T}$ . C'est là que l'on se réjouit de la Remarque 4.3.3. L'algorithme est de plus simplifié par le fait que pour tous  $x, y$  de  $E$ ,

$$\frac{\pi_{V,T}(y)}{\pi_{V,T}(x)} = e^{-(V(y)-V(x))/T}$$

qui ne dépend que des différences  $V(y) - V(x)$ . Explicitons dans ce cadre l'algorithme de Metropolis : partant de  $x \in E$ , l'évolution est donnée simplement par

1. choisir  $y \in E$  suivant la loi  $(Q(x, y))_{y \in E}$  ;
2. calculer  $A(x, y) = h \left( \frac{Q(y, x)}{Q(x, y)} e^{-\Delta V(x, y)/T} \right)$  où  $\Delta V(x, y) = V(y) - V(x)$  ;
3. on tire un  $U$  de loi  $\mathcal{U}([0, 1])$  ;
  - si  $U \leq A(x, y)$  alors on accepte la sélection et on choisit  $X_{n+1} = y$  ;
  - sinon on refuse la sélection et on choisit  $X_{n+1} = x$ .

Les mesures de Gibbs sont fondamentales en physique, et plus particulièrement en physique statistique où  $E$  représente l'espace d'état d'un système physique,  $x \in E$  une configuration donnée et  $V(x)$  l'énergie de cette configuration. Les mesures de Gibbs sont alors les lois d'équilibre macroscopique du système<sup>1</sup>. Notre premier exercice de vraie simulation concerne l'utilisation de la méthode de Metropolis–Hastings pour simuler la loi de Gibbs dans un modèle simple d'aimant.

**Exercice 47** On considère un réseau fini  $R = \{1, \dots, r\}^2$ , qui représente les positions des atomes dans un bloc de métal. On note  $a \sim b$  si deux sites  $a$  et  $b$  sont voisins. Chaque atome  $a$  a un moment magnétique (une "micro-aimantation") qui est orienté soit vers le haut, soit vers le bas : on note  $\sigma(a) \in \{-1, +1\}$  le moment magnétique (que l'on appelle habituellement "spin") en  $a = (i, j) \in R$ . La configuration du bloc est donc décrite par  $\sigma = (\sigma(a))_{a \in R}$  et on a donc  $E = \{-1, +1\}^R$ . On suppose que les moments magnétiques différents ont tendance à se repousser, de sorte que si deux atomes voisins ont des spins différents, l'énergie du système est plus élevée que s'ils sont alignés. On suppose en revanche que deux atomes qui ne sont pas immédiatement voisins n'interagissent pas directement l'un avec l'autre. On modélise ceci en posant

$$V(\sigma) = - \sum_{(a,b) | a \sim b} \sigma(a) \cdot \sigma(b) \quad (4.6)$$

où la somme porte sur l'ensemble des couples  $a, b$  de  $R$  qui sont voisins.

1. On considère la chaîne de Markov dont l'évolution est donnée comme suit : partant de  $\sigma$ , on choisit  $a \in R$  uniformément et on renverse le spin en ce site, sans toucher au reste : on passe en  $\sigma'$  vérifiant  $\sigma'(a) = -\sigma(a)$  et  $\sigma'(b) = \sigma(b)$  pour  $b \neq a$ . Si l'on note  $Q$  la matrice de transition associée (à ne surtout pas expliciter), a-t-elle la propriété  $Q(\sigma, \sigma') > 0 \Rightarrow Q(\sigma', \sigma) > 0$  ? A-t-on une relation plus précise entre  $Q(\sigma, \sigma')$  et  $Q(\sigma', \sigma)$  ?

1. Au sens où les mesures  $\pi_{V,T}$  sont les mesures  $\pi$  qui maximisent l'entropie  $S(\pi) = - \int \log \pi(x) d\pi$  sous la contrainte que l'énergie totale  $\int V d\pi$  est fixée.

2. Supposons que dans l'évolution  $\sigma \rightarrow \sigma'$  ci-dessus, on ait choisi le site  $a$ . Montrez que

$$V(\sigma') - V(\sigma) = 2\sigma(a) \cdot \sum_{b|b\sim a} \sigma(b). \quad (4.7)$$

Définissez une fonction `deltaV` qui calcule la quantité donnée en (4.7) si on lui donne  $\sigma$  (comme un array Numpy) et  $a$  comme un tuple. On pourra définir avant cela une fonction `voisins` qui prend en entrée  $a$  et retourne l'ensemble des voisins de  $a$ .

3. Reprenez l'algorithme de Metropolis–Hastings avec  $h(u) = \min(1, u)$  pour  $Q$  et  $\pi_{V,T}$ . Observez que l'évolution de matrice de transition  $P$  est définie simplement comme suit : partant de  $\sigma$ ,
- on choisit un site  $a \in R$  uniformément,
  - on calcule  $\Delta V = 2\sigma(a) \cdot \sum_{b|b\sim a} \sigma(b)$  et on simule une variable de loi uniforme  $U \sim \mathcal{U}([0, 1])$ ,
  - si  $U < e^{-\Delta V/T}$ , on reverse le spin en  $a$ .
4. Montrer que la chaîne de Markov définie ci-dessus est irréductible apériodique. Par conséquent, la loi après  $n$  itérations de l'algorithme converge vers la probabilité invariante quand  $n \rightarrow \infty$ . Ecrivez une fonction qui exploite l'algorithme de Metropolis pour tirer une configuration suivant l'état de Gibbs du modèle d'Ising. Affichez-la en utilisant la commande `matshow` de Matplotlib, et comparez-la à une configuration obtenue en choisissant en chaque site un spin  $\pm 1$  de manière équiprobable. De manière authentiquement pifométrique, on vous recommande les valeurs  $r = 20$ ,  $n = 10000$  et  $T$  variable. Pour un meilleur effet, commencez avec  $T = 10$  et diminuez la valeur de 1 en 1 jusqu'à 1.
5. Pour une configuration  $\sigma$  donnée, on peut définir l'aimantation macroscopique  $A(\sigma) = \frac{1}{r^2} \sum_{a \in R} \sigma(a)$ . Pour différents tirages à différentes valeurs de  $T$ , calculez  $|A(\sigma)|$ . La valeur observée indique-t-elle que les spins ont tendance à s'aligner, ou bien peut-il être simplement l'effet du hasard ? pour répondre à cette question, donnez un intervalle de fluctuation à 95% pour  $|A(\sigma)|$  dans le cas où  $\sigma$  est obtenu en tirant uniformément, en chaque site, un spin  $\pm 1$ .

## 4.5 Chaînes réversibles et trou spectral

L'un des intérêts de la notion de matrice réversible est la suivante : si l'on définit un produit scalaire sur l'ensemble des fonctions de  $E$  dans  $\mathbb{R}$  par

$$\langle f, g \rangle_\pi = \int fg \, d\pi, \quad (4.8)$$

alors on a

**Lemme 4.5.1** *La matrice  $P$  est réversible par rapport à  $\pi$  si et seulement si elle est autoadjointe pour  $\langle \cdot, \cdot \rangle_\pi$  en tant qu'application  $f \mapsto Pf$ .*

La conséquence immédiate est que  $P$  est diagonalisable dans le groupe orthogonal pour le produit scalaire  $\langle \cdot, \cdot \rangle_\pi$ , de valeurs propres réelles. On a donc :

**Lemme 4.5.2** Si  $P$  est réversible par rapport à une probabilité  $\pi$ , alors ses valeurs propres peuvent s'écrire

$$1 = v_1 \geq v_2 \geq \dots \geq v_d \geq -1.$$

Si de plus  $P$  est irréductible, alors  $1 = v_1 > v_2$ . Si de plus  $P$  est irréductible et apériodique alors  $v_d > -1$ .

La preuve est laissée en exercice :

**Exercice 48** Prouvez le lemme 4.5.2. Pour les deuxième et troisième assertions, utilisez le théorème de convergence ergodique (qui dit que  $\frac{1}{n} \sum_{k=1}^n P^k f$  converge ponctuellement vers  $\int f d\pi$ ) et le théorème de convergence en loi (qui dit que sous certaines conditions  $P^n f \rightarrow \int f d\pi$  converge ponctuellement vers  $\int f d\pi$ ).

L'intérêt de cette notion est donné par le lemme suivant, où l'on note  $\|f\|_\pi^2 = \langle f, f \rangle_\pi$ .

**Lemme 4.5.3** Soit  $P$  réversible par rapport à  $\pi$ . Si l'on note  $\rho = \max(|v_2|, |v_d|)$  alors pour toute fonction  $f$  sur  $E$  on a pour tout  $n$  :

$$\|P^n f - \int f d\pi\|_\pi \leq \rho^n \|f - \int f d\pi\|_\pi. \quad (4.9)$$

**Preuve.** On note  $\Psi_1, \dots, \Psi_d$  une base orthonormée de diagonalisation, où chaque  $\Psi_i$  est vecteur propre pour la valeur propre  $v_i$ . On a

$$f = \sum_{i=1}^d \langle \Psi_i, f \rangle_\pi \Psi_i$$

et comme  $\Psi_1 = \mathbb{1}$ , la fonction constante égale à 1, on a

$$f - \left( \int f d\pi \right) \mathbb{1} = \sum_{i=2}^d \langle \Psi_i, f \rangle_\pi \Psi_i$$

d'où en appliquant  $P^n$

$$P^n f - \left( \int f d\pi \right) \mathbb{1} = \sum_{i=2}^d \langle \Psi_i, f \rangle_\pi v_i^n \Psi_i$$

et par Pythagore :

$$\|P^n f - \int f d\pi\|_\pi = \sum_{i=2}^d |\langle \Psi_i, f \rangle_\pi|^2 v_i^{2n} \leq \rho^{2n} \sum_{i=2}^d |\langle \Psi_i, f \rangle_\pi|^2 = \rho^{2n} \|f - \int f d\pi\|_\pi^2. \quad \square$$

**Corollaire 4.5.4** *Sous les mêmes hypothèses, en supposant de plus que  $\pi(x) > 0$  pour tout  $x \in E$ , alors pour toute fonction  $f$  sur  $E$  et tout  $x \in E$  on a quel que soit  $n$  :*

$$\left| \mathbb{E}_x(f(X_n)) - \int f d\pi \right| \leq c(x) \rho^{n-1} \|f - \int f d\pi\|_\pi \quad (4.10)$$

où  $c(x)^2 = P^2(x, x)/\pi(x)$  (attention,  $P^2(x, x)$  et pas  $P(x, x)^2$ ).

**Preuve.** Supposons  $\int f d\pi = 0$  (ce qui revient à remplacer  $f$  par  $f - \int f d\pi$ ). On pose  $\Phi = P^{n-1}f$ . On a

$$P\Phi(x) = \sum_y P(x, y)\Phi(y) = \sum_y \frac{P(x, y)}{\pi(y)} \Phi(y)\pi(y) = \langle g_x, \Phi \rangle_\pi$$

où  $g_x(y) = P(x, y)/\pi(y)$ . Donc par Cauchy-Schwarz

$$|P^n f(x)|^2 = |P\Phi(x)|^2 \leq \|g_x\|_\pi^2 \|\Phi\|_\pi^2 \leq \|g_x\|_\pi^2 \rho^{2n-2} \|f\|_\pi^2$$

où  $\|g_x\|_\pi^2 = \sum_y P(x, y)^2/\pi(y)$ ; en utilisant la réversibilité on a  $P(x, y)^2/\pi(y) = P(x, y)P(y, x)/\pi(x)$ , donc  $\|g_x\|_\pi^2 = c(x)^2$ .  $\square$

Le Corollaire 4.5.4 montre en particulier que, partant de la mesure  $\mu_0 = \delta_x$ , on a convergence exponentielle avec taux  $\rho$  de  $\mu_n = \mu_0 P^n$  vers  $\pi$ . La suite va consister à donner une majoration de  $\rho$  (que l'on veut aussi petit que possible).

On va maintenant définir la forme de Dirichlet associée à  $P$  et  $\pi$  : c'est la forme quadratique sur l'ensemble des fonctions de  $E$  dans  $\mathbb{R}$ , donnée par

$$\mathcal{E}(f, f) = \langle f, (\text{id} - P)f \rangle_\pi. \quad (4.11)$$

**Exercice 49** *Montrer que, si  $P$  est réversible par rapport à une probabilité  $\pi$ , alors*

$$\mathcal{E}(f, f) = \frac{1}{2} \sum_{x, y \in E} |f(x) - f(y)|^2 P(x, y)\pi(x).$$

On définit encore le trou spectral

$$\lambda(P) = \inf \left\{ \frac{\mathcal{E}(f, f)}{\|f - \int f d\pi\|_\pi^2} \mid f \notin \mathbb{R}\mathbb{1} \right\}. \quad (4.12)$$

Cette quantité peut encore se relier aux valeurs propres  $v_i$  :

**Lemme 4.5.5** *Si  $P$  est irréductible et réversible par rapport à une probabilité  $\pi$ , alors  $\lambda(P) = 1 - v_2$ .*

**Preuve.** Remarquons que  $\mathcal{E}(f, f) = \mathcal{E}(f - \int f d\pi, f - \int f d\pi)$ . Par conséquent,

$$\lambda(P) = \inf \left\{ \frac{\langle f, (\text{id} - P)f \rangle_\pi}{\|f\|_\pi^2} \mid f \in (\mathbb{R}\mathbb{1})^\perp \right\} = \inf \text{sp} (\text{id} - P)|_{(\mathbb{R}\mathbb{1})^\perp} = 1 - v_2. \quad \square$$

Ces résultats vont être utilisés de la manière suivante : on dit qu'on a une inégalité de Poincaré s'il existe  $C > 0$  tel que pour tout  $f$

$$\|f - \int f d\pi\|_{\pi}^2 \leq C \mathcal{E}(f, f).$$

Une telle inégalité s'obtient habituellement en estimant directement la forme de  $\mathcal{E}(f, f)$  donnée dans l'exercice 49. Si elle est vérifiée, on a par définition  $\lambda(P) \geq C^{-1}$  donc  $v_2 \leq 1 - C^{-1}$ . Si l'on sait par ailleurs que  $|v_d| \leq |v_2|$  (par exemple si  $v_d \geq 0$ ), alors on a  $\rho \leq 1 - C^{-1}$ , inégalité que l'on peut injecter dans (4.10). On verra dans la section 4.6 que l'on sait obtenir des inégalités de Poincaré pour les chaînes de Markov obtenues par l'algorithme de Metropolis pour des mesures de Gibbs.

#### 4.6 Mesures de Gibbs et recuit simulé

Des calculs plus pénibles que réellement compliqués (voir Théorème 3.3.11 et section 3.3.4 de [1]) permettent de montrer, pour une chaîne de Markov obtenue par l'algorithme de Metropolis à partir d'une mesure de Gibbs associée à un potentiel dont on suppose pour simplifier qu'il vérifie

$$Q(x, y) > 0 \Rightarrow V(x) \neq V(y) \quad (4.13)$$

(hypothèse qui n'est pas vraie pour le modèle d'Ising), que l'on a une inégalité de Poincaré avec  $C = d^3 e^{C(V)/T}$  où  $d = \text{card } E$ ,  $C(V)$  est une constante qui ne dépend que de  $V$  (et en particulier pas de  $T$ ), qui est hélas en pratique difficile à calculer. Par ailleurs, si l'on note  $P_T$  la matrice de transition obtenue par l'algorithme de Metropolis,  $\rho_T, v_{T,2}$  les quantités  $\rho, v_2$  associées, alors on a l'observation simple suivante :

**Lemme 4.6.1** *Il existe  $T_0$  tel que pour tout  $T < T_0$ ,  $\rho_T = v_{T,2}$ .*

**Preuve.** On note  $P_0 = \lim_{T \rightarrow 0} P_T$ . On voit facilement que cette limite existe, et à partir de (4.4) et de l'hypothèse que  $Q(x, y) > 0 \Rightarrow V(x) \neq V(y)$  que

$$P_0(x, y) = \begin{cases} (\#\{y \in E \mid Q(x, y) > 0\})^{-1} & \text{si } Q(x, y) > 0 \text{ et } V(y) < V(x) \\ 0 & \text{sinon.} \end{cases}$$

Par conséquent, si l'on ordonne les points  $x$  de  $E$  par valeurs de  $V$  croissantes, la matrice  $P_0$  est triangulaire. Les éléments diagonaux étant positifs, ses valeurs propres sont positives ou nulles. Par conséquent  $v_{0,d} \geq 0$  et la valeur propre de plus grand module après 1 est  $v_{0,2}$ , donc les quantités associées  $\rho_0$  et  $v_{0,2}$  sont égales. Ces quantités étant continues en les paramètres de la matrice, donc en  $T$ , elles restent vraies dans un voisinage de 0.  $\square$

Les deux observations ci-dessus impliquent (voir la section 4.5) que pour  $T$  assez petit :

$$\rho_T \leq 1 - d^{-3} e^{-C(V)/T}. \quad (4.14)$$

Par ailleurs, si l'on note

$$V_{\min} = \{x \in E \mid V(x) = \inf_E V\}$$

alors on a immédiatement le lemme suivant :



**Lemme 4.6.2** *On a la convergence*

$$\pi_{V,0} \stackrel{\text{def.}}{=} \lim_{T \rightarrow 0} \pi_{V,T} = \frac{1}{\text{card } V_{\min}} \sum_{x \in V_{\min}} \delta_x. \quad (4.15)$$

**Preuve.** Soient  $x, y$  deux points de  $E$ . On a

$$\frac{\pi_{V,T}(y)}{\pi_{V,T}(x)} = e^{-(V(y)-V(x))/T}.$$

Par conséquent, si  $V(y) > V(x)$  alors  $\lim_{T \rightarrow 0} \frac{\pi_{V,T}(y)}{\pi_{V,T}(x)} = 0$  donc  $\lim_{T \rightarrow 0} \pi_{V,T}(y) = 0$ . Cela est nécessairement vrai pour tout  $y \notin V_{\min}$ . Pour  $x, y \in V_{\min}$  on a  $\pi_{V,T}(x) = \pi_{V,T}(y)$  pour tout  $T$ .  $\square$

On a donc montré les deux points suivants :

1. par le Corollaire 4.5.4 et (4.14), partant de n'importe quelle distribution  $\mu$ , la probabilité  $\mu_n$  converge exponentiellement vite vers  $\pi_{V,T}$  – mais avec un taux

$$-\log\left(1 - \frac{1}{d^3} e^{-C(V)/T}\right) \underset{T \rightarrow 0}{\sim} d^{-3} e^{-C(V)/T}$$

donc qui s'approche rapidement de zéro lorsque  $T \rightarrow 0$ ;

2. on a  $\lim_{T \rightarrow 0} \pi_{V,T} = \frac{1}{\text{card } V_{\min}} \sum_{x \in V_{\min}} \delta_x$ .

On peut espérer que, si l'on applique l'évolution  $P_T$  donnée par l'algorithme de Metropolis en faisant changer la température  $T$  par paliers, et cela assez lentement pour qu'entre deux changements on ait eu le temps de s'approcher de  $\pi_{V,T}$ , alors en fin de compte on devrait s'approcher de l'état  $\pi_{V,0}$ . Le résultat suivant prouve que cette intuition est juste :

**Théorème 4.6.3** *Si  $V$  a la propriété (4.13), que  $\inf_E V > 0$  et que l'on choisit une suite de températures  $(T(n))_n$  données par*

$$T(n) = \frac{1}{k} \quad \text{pour} \quad e^{C(k-1)} \leq n < e^{Ck},$$

avec  $C > C(V)$ , alors on a

$$\lim_n P_{T(n)} \dots P_{T(1)} f = \frac{1}{\text{card } V_{\min}} \sum_{x \in V_{\min}} f(x).$$

**Preuve.** On note  $n_k = e^{Ck}$ , et  $\pi_T$  au lieu de  $\pi_{V,T}$ . On se contente de montrer la convergence aux temps  $n_k$ . Par le Corollaire (4.5.4), on a

$$\begin{aligned} & \left| \mathbb{E}(f(X_{n_k} | X_{n_{k-1}} = x)) - \int f d\pi_{1/k} \right|^2 \\ & \leq \frac{P^2(x,x)}{\pi_{1/k}(x)} \rho_{1/k}^{2(n_k - n_{k-1} - 1)} 4 \|f\|_{\infty}^2. \end{aligned}$$

On a  $P^2(x, x) \leq 4$  et les majorations grossières

$$\pi_{1/k}(x) = \frac{1}{Z_{1/k}} e^{-kV(x)} \geq (d e^{-k \inf V})^{-1} e^{-k \sup V} = d^{-1} e^{-k\beta}$$

où l'on note  $\beta = \frac{1}{2}(\sup V - \inf V) > 0$  et

$$\rho_{1/k}^{2(n_k - n_{k-1} - 1)} = \left(1 - \frac{1}{d^3} e^{-kC(V)}\right)^{2(n_k - n_{k-1} - 1)} \leq \exp -\alpha e^{k(C-C(V))}$$

pour une constante  $\alpha$ . On a donc

$$\begin{aligned} & \left| \mathbb{E}(f(X_{n_k} | X_{n_{k-1}} = x)) - \int f d\pi_{1/k} \right| \\ & \leq 16d \|f\|_\infty^2 \exp(\beta k - \alpha e^{k(C-C(V))}) \end{aligned}$$

et le majorant tend vers zéro. On conclut donc par le Lemme 4.6.2.  $\square$

On obtient donc sous les hypothèses du Théorème, que partant de n'importe quel  $x \in E$  on atteindra presque-sûrement un minimum de  $V$  en appliquant un algorithme de Metropolis tel que décrit en section 4.4, avec une température variable déterminée comme ci-dessus. Cet algorithme s'appelle le "recuit simulé", par analogie avec la technique métallurgique où l'on obtient un métal durci en le chauffant avant de le laisser refroidir lentement, et ce plusieurs fois.

L'idée qui sous-tend cet algorithme est la suivante : on teste des changements de configuration en les sélectionnant au hasard suivant  $Q$ . Si le changement fait baisser  $V$  (cas  $\Delta V < 0$ ) alors on accepte la nouvelle configuration (car la condition  $U \leq e^{-\Delta V/T}$  est forcément vraie). Si le changement fait augmenter  $T$  (cas  $\Delta V > 0$ ), on peut l'accepter ou le refuser, suivant que  $U \leq e^{-\Delta V/T}$  ou non. Le premier mécanisme tend à faire diminuer  $V$  ; le deuxième évite que l'on se retrouve coincé en un minimum local. Le paramètre modifiant la tendance à accepter un changement défavorable est  $T$  (on l'accepte d'autant plus que  $T$  est grand) ; choisir  $T$  décroissant vers 0 nous assure que l'on finira par se fixer en un minimum, la décroissance lente doit nous assurer que l'on aura pris assez de risque pour explorer "toutes" les possibilités avant de se fixer.

Les obstructions techniques sont de deux types :

- On doit choisir un schéma de décroissance par palier avec  $C > C(V)$ , mais  $C(V)$  est inconnu,
- on a la convergence presque-sûre, mais on n'a pas explicité la vitesse de convergence (on aurait pu faire mieux dans les démonstrations ci-dessus, mais on n'aurait rien eu de très instructif).

En pratique, il sera facile de voir si l'algorithme converge vers un minimum de  $V$ . En faisant tourner des simulations, on verra si l'algorithme se comporte bien, et en particulier s'il semble avoir l'une des pathologies suivantes :

- une convergence trop lente, due à un algorithme qui continue trop longtemps à accepter les sélections obtenues par  $Q$  (ce qui se produit quand la température décroît trop lentement) ;

- une convergence trop rapide, en général vers un minimum local, due à un algorithme qui se met trop rapidement à refuser les sélections obtenues par  $Q$  (ce qui se produit quand la température décroît trop rapidement).

On aura donc intérêt à jouer à varier la valeur de  $C \dots$  ou même à changer de schéma de décroissance par palier. Nous allons appliquer ce résultat à un problème classique, où nous observerons ce phénomène.

**Exercice 50** *Un voyageur doit visiter  $r$  villes, que l'on représente par  $r$  points  $M_1, \dots, M_r$  du plan. Sa ville de départ doit être la même que sa ville d'arrivée, mais on suppose (cela simplifie les notations) qu'il peut choisir cette ville aussi. Puisqu'il passera dans chaque ville une et une seule fois, son parcours est déterminé par une permutation  $\sigma$  de  $\{1, \dots, r\}$ , donc  $E = \mathfrak{S}_r$ . Les villes visitées sont alors, dans l'ordre,  $M_{\sigma(1)}, M_{\sigma(2)} \dots$ . La distance entre les villes  $i$  et  $j$  sont données par  $d(i, j)$ ; le voyageur parcourra donc une distance*

$$V(\sigma) = \sum_{i=1}^r d(M_{\sigma(i)}, M_{\sigma(i+1)}) \quad (4.16)$$

(où l'on considère que  $r+1 = 1$  : on prend bien en compte la distance parcourue pour retourner au point de départ après avoir visité toutes les villes). On souhaite trouver un itinéraire  $\sigma$  pour lequel la distance parcourue totale est minimale. Il y a  $r!$  itinéraires; tous les tester serait beaucoup trop long. On va donc utiliser le recuit simulé.

1. *Ecrivez une fonction `distancetotale` qui, pour des variables d'entrée `M` et `sigma` qui sont respectivement une liste de paires de points représentant les coordonnées  $(x_1, y_1), \dots, (x_r, y_r)$  de  $r$  points  $M_1, \dots, M_r$  du plan et une permutation de  $\{1, \dots, n\}$ , calcule la distance totale  $V(\sigma)$  telle que définie par la relation (4.16) (on pourra utiliser un `%`, qui calcule le "modulo").*
2. *On considère la transition dont l'évolution est la suivante : partant de  $\sigma$ , on choisit au hasard et uniformément deux points distincts, qui s'écrivent donc  $\sigma(i)$  et  $\sigma(j)$ . Si par exemple on visitait  $\sigma(i)$  avant  $\sigma(j)$  (c'est-à-dire si  $i < j$ ) alors on échange  $\sigma(i)$  et  $\sigma(j)$  dans l'itinéraire (par exemple si le couple choisi est 3, 6 alors  $[2, 3, 1, 4, 6, 5]$  devient  $[2, 6, 1, 4, 3, 5]$ ). Montrez que l'évolution en question a la propriété  $Q(\sigma, \sigma') = Q(\sigma', \sigma)$ , puis écrivez une fonction `Qpas` qui pour une variable d'entrée `sigma` représentant une permutation  $\sigma$ , retourne `sigmap` représentant  $\sigma'$  définie comme ci-dessus.*
3. *Pour un  $T > 0$ , reprenez l'algorithme de Metropolis–Hastings pour  $Q$  et  $\pi_{V,T}$ . Observez que l'évolution de matrice de transition  $P_T$  est définie simplement comme suit : partant de  $\sigma$ ,*
  - (a) *on modifie l'itinéraire  $\sigma$  en  $\sigma'$  comme ci-dessus,*
  - (b) *on calcule  $\Delta V = V(\sigma') - V(\sigma)$  et on simule une uniforme  $U \sim \mathcal{U}([0, 1])$ ,*
  - (c) *si  $U < e^{-\Delta V/T}$ , on sélectionne  $\sigma'$ , sinon on conserve  $\sigma$ .*
4. *Pour les points  $M_1, \dots, M_r$ , on tire  $r$  points au hasard<sup>2</sup> dans  $[0, 1]^2$ . Appliquer l'évolution ci-dessus avec une température variable suivant le schéma donné*

2. Si l'on est un peu plus curieux de vraies données, on peut aller récupérer les positions des villes françaises sur [data.gouv.fr](http://data.gouv.fr)

dans le Théorème 4.6.3, en faisant varier la valeur de  $C$ . On pourra essayer aussi avec une variation plus rapide  $T(n) = C/n$  et comparer les résultats. On pensera en particulier aux éléments de discussion donnés avant l'énoncé de l'exercice.

#### 4.7 Exercices supplémentaires

**Exercice 51** On considère un sous-graphe  $(G, A)$  du réseau  $\{1, \dots, 10\}^3$ , donné par un ensemble de points  $G$  et une matrice d'adjacence symétrique  $A$ . On veut "ranger" dans cette boîte des boules de diamètre  $R$  vérifiant  $1 < R < \sqrt{2}$ , de sorte que l'on ne peut mettre deux boules sur des sites voisins mais qu'il n'y a pas d'autre contrainte. On veut utiliser le recuit simulé pour trouver une configuration permettant de ranger le plus grand nombre possible de boules dans la boîte. On décrit par  $x : \Lambda \rightarrow \{0, 1\}$  une configuration.

1. Montrez que l'algorithme suivant : partant de  $x$ ,
  - (a) on choisit un site  $s$  au hasard dans  $G$ ,
  - (b) si tous les sites voisins (au sens de l'adjacence dans  $S$ ) sont libres, on ajoute une boule en  $s$ ,
 définit une matrice de transition  $Q$  vérifiant  $Q(x, y) = 0 \Leftrightarrow Q(y, x) = 0$ .
2. Montrez que la fonction  $V$  définie par

$$V(x) = 1000 - \text{card}\{s \in \Lambda \mid x(s) = 1\}$$

(donc le nombre de sites libres dans la configuration  $x$ ) vérifie bien pour  $x \neq y$  la relation  $Q(x, y) > 0 \Rightarrow V(x) \neq V(y)$ .

3. Appliquez l'algorithme du recuit simulé pour trouver une configuration avec un nombre maximal de boules. On pourra faire varier le schéma de décroissance de la température en fonction du comportement observé de l'algorithme. Comme exemples de  $(G, A)$  on pourra partir du réseau cubique  $\{1, \dots, 10\}^3$  et lui enlever aléatoirement des arêtes et/ou sommets.

## Bibliographie

- [1] M. Benaïm, N. El Karoui, *Promenade aléatoire, chaînes de Markov et simulations; martingales et stratégies.*
- [2] B. Bercu, D. Chafaï, *Modélisation stochastique et simulation.*
- [3] J.-F. Delmas, B. Jourdain, *Modèles aléatoires, applications aux sciences de l'ingénieur et du vivant.*



## Chapitre 5

# Estimateurs paramétriques, intervalles de confiance et tests d'hypothèses

Dans cette séance, nous abordons les estimateurs de paramètres de lois. Comme son nom l'indique, un estimateur donne une estimation d'un paramètre ; or, une estimation n'est rien sans une indication de l'erreur associée. Les intervalles de confiance ont justement pour utilité de préciser cette erreur. La seule difficulté conceptuelle est que, l'estimation étant basée sur des résultats aléatoires, la précision elle-même est aléatoire. Nous allons discuter la théorie et utiliser la simulation pour illustrer la qualité des estimateurs et intervalles de confiance obtenus par l'approche théorique.

### 5.1 Estimateurs : définitions

Nous commençons par rappeler les définitions générales.

**Définition 5.1.1** *Un modèle paramétrique est une famille de probabilités indexées par un paramètre  $\theta$  qui vit dans un espace de dimension finie :*

$$\mathcal{P} = \{\mathbb{P}_\theta, \theta \in \Theta\},$$

où  $\Theta \subset \mathbb{R}^d$  pour un certain  $d$ . On dira que le modèle est identifiable si  $\Theta \ni \theta \mapsto \mathbb{P}_\theta$  est injective.

Cette définition abstraite appelle déjà plusieurs remarques. Tout d'abord, l'identifiabilité est une hypothèse naturelle : puisque ce que l'on observe, ce sont des variables aléatoires de loi  $\mathbb{P}_\theta$ , on n'a aucun moyen de distinguer  $\theta$  de  $\theta'$  si  $\mathbb{P}_\theta = \mathbb{P}_{\theta'}$ . Dans la suite, on supposera par défaut que les modèles sont identifiables. Par ailleurs, en pratique, on travaillera le plus souvent avec un  $n$ -échantillon  $X_1, \dots, X_n$  d'une loi  $\mathbb{P}_\theta$ , de sorte que le modèle considéré sera de la forme

$$\mathcal{P}^{\otimes n} = \{\mathbb{P}_\theta^{\otimes n} \mid \theta \in \Theta\}.$$

#### Exemple 1

- Si  $X_1, \dots, X_n$  est un  $n$ -échantillon de loi normale  $\mathcal{N}(m, \sigma^2)$  avec  $m, \sigma^2$  inconnus, alors le modèle naturel associé est

$$\mathcal{P}^{\otimes n} = \{\mathbb{P}_{m, \sigma^2}^{\otimes n} \mid (m, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+\}$$

où  $\mathbb{P}_{m, \sigma^2}$  est la loi normale  $\mathcal{N}(m, \sigma^2)$  sur  $\mathbb{R}$ .

- On considère les  $n + 1$  premiers pas  $(X_k)_{k=0}^n$  d'une chaîne de Markov sur un espace d'état fini  $E$ , de loi initiale  $\mu$  et de matrice de transition  $P$ . Le modèle associé est l'ensemble des lois  $\mathbb{P}_{\mu, P}^{(n)}$  :

$$\mathbb{P}_{\mu, P}^{(n)}(x_0, \dots, x_n) = \mu(x_0)P_{x_0, x_1} \dots P_{x_{n-1}, x_n},$$

qui n'est pas de la forme  $\mathcal{P}^{\otimes n}$ . Il apparaît naturellement lorsque l'on veut estimer la matrice de transition d'une chaîne de Markov à partir de l'observation des  $n$  premiers pas.

- On considère l'ensemble des lois à support dans  $[0, 1]$  qui sont absolument continues par rapport à la mesure de Lebesgue. Elles ne constituent pas un modèle paramétrique, donc n'entrent pas dans le cadre de cette séance.

**Définition 5.1.2** Si  $X$  est une variable aléatoire dont la loi appartient à un modèle paramétrique  $\mathcal{P} = \{\mathbb{P}_\theta, \theta \in \Theta\}$ , une statistique  $Z$  est une variable  $X$ -mesurable, autrement dit une fonction  $\varphi(X)$  de  $X$  ; cette statistique est un estimateur d'un paramètre  $g(\theta)$  si presque-sûrement  $Z \in g(\Theta)$ .

On formule la notion d'estimateur de  $g(\theta)$  pour couvrir le cas où l'on n'estime pas tous les paramètres à la fois : par exemple, dans l'exemple ci-dessus où  $\theta = (m, \sigma^2)$  on peut vouloir parler d'un estimateur de  $m$ .

Il faut remarquer qu'être un estimateur n'est pas une propriété bien intéressante : par exemple, dans le cas d'un  $n$ -échantillon gaussien comme ci-dessus, n'importe quelle constante positive est un estimateur de  $\sigma^2$ . On va donc définir plusieurs qualités possibles pour un estimateur. On commence par définir le biais et le risque quadratique :

**Définition 5.1.3** Soit  $\mathcal{P}$  un modèle comme ci-dessus, et  $Z$  un estimateur de  $g(\theta)$ . Son biais est la quantité

$$b(\theta) = \mathbb{E}_\theta(Z) - g(\theta)$$

et son risque quadratique est

$$r(\theta) = \mathbb{E}_\theta((Z - g(\theta))^2)$$

où, dans les deux cas,  $\mathbb{E}_\theta$  est l'espérance par rapport à  $\mathbb{P}_\theta$ .

Différentes qualités éventuelles d'un estimateur  $Z$  de  $g(\theta)$  seront les suivantes :

- être sans biais, c'est-à-dire avoir un biais  $b(\theta)$  nul pour tout  $\theta$ ,
- avoir un risque quadratique faible.



Notez que si l'on veut comparer deux estimateurs  $Z_1$  et  $Z_2$  par leur risque quadratique, une inégalité n'a d'intérêt que si elle a lieu pour tout  $\theta$  : on dira par exemple que  $Z_1$  a un risque quadratique plus faible que  $Z_2$  si  $r_{Z_1}(\theta) \leq r_{Z_2}(\theta)$  pour tout  $\theta \in \Theta$ . On parlera plus loin d'estimateur UMVB pour "uniformément de variance minimale, sans biais", donc pour un estimateur  $Z$  de  $g(\theta)$  vérifiant  $b_Z(\theta) = 0$  et  $r_Z(\theta) \leq r_{Z'}(\theta)$  pour tout autre  $Z'$  estimateur sans biais de  $\theta$ .

Lorsque l'on a un modèle  $\mathcal{P}_n$ , l'estimateur  $Z_n$  de  $g(\theta)$  dépend *a priori* de la taille  $n$  de l'échantillon, et l'on note  $b_n(\theta)$  et  $r_n(\theta)$  ses biais et risque quadratique. D'autres qualités possibles de l'estimateur sont alors :

- être asymptotiquement sans biais, c'est-à-dire vérifier,  $\lim_{n \rightarrow \infty} b_n(\theta) = 0$ ,
- être fortement consistant c'est-à-dire vérifier que pour tout  $\theta$  on a  $Z_n \xrightarrow{\text{P.S.}} g(\theta)$ ,
- être faiblement consistant, c'est-à-dire vérifier que pour tout  $\theta$  on a  $Z_n \xrightarrow{\text{P}} g(\theta)$ .

Toutes ces qualités concernant le biais et la consistance proviennent en général, soit de théorèmes généraux (loi des grands nombres par exemple), soit de calculs explicites.

### Exercice 52

1. Montrez que l'on a la décomposition risque-variance suivante pour le risque quadratique d'un estimateur  $Z$  :

$$r(\theta) = b(\theta)^2 + \text{var } Z.$$

2. Montrez l'inégalité

$$\mathbb{P}(|Z - g(\theta)| > \varepsilon) \leq \varepsilon^{-2} r(\theta).$$

Déduisez-en que si l'on a une suite  $(Z_n)_n$  d'estimateurs de  $g(\theta)$  avec la propriété  $\lim_{n \rightarrow \infty} r_n(\theta) = 0$  pour tout  $\theta$ , alors l'estimateur est faiblement consistant. Quelle condition suffisante sur  $r_n(\theta)$  peut-on énoncer pour avoir la (forte) consistance ?

Le point 2. montre que la vitesse de décroissance en  $n$  du risque quadratique donne une indication de la vitesse de convergence de l'estimateur.

Une dernière qualité recherchée d'une suite d'estimateurs est l'existence d'une loi asymptotique, c'est-à-dire le fait qu'il existe une suite  $(a_n)_n$  avec  $a_n \rightarrow \infty$ , telle que  $a_n(Z_n - g(\theta))$  converge en loi, vers une loi non triviale. On dit dans ce cas que la suite d'estimateurs  $(Z_n)_n$  est de vitesse  $(a_n)_n$ . Lorsque  $a_n = \sqrt{n}$  et que la loi limite est une normale centrée, on parle de normalité asymptotique. Nous verrons l'intérêt d'une telle propriété lorsque nous étudierons les intervalles de confiance asymptotiques. Nous indiquons dans la section 5.3 un résultat général donnant une telle propriété, mais pour les cas qui ne sont pas couverts par ce résultat, l'existence d'une loi asymptotique se prouve au cas par cas.

**Exercice 53** Soit  $U_1, \dots, U_n$  un  $n$ -échantillon de loi  $\mathcal{U}([0, \theta])$ .

1. Montrez que  $Z_1 = 2\bar{U}_n$  est un estimateur sans biais, consistant et asymptotiquement normal de  $\theta$ .

2. Montrez que  $Z_2 = \max(U_1, \dots, U_n)$  est un estimateur consistant de  $\theta$  et que  $n(Z_2 - \theta)$  converge en loi vers une loi que l'on identifiera.
3. Pour un  $\theta$  quelconque, simulez un 100-échantillon  $U_1, \dots, U_n$  et définissez les estimateurs  $Z_1$  et  $Z_2$  correspondant aux 100 valeurs de  $n$ . Tracez les trajectoires de  $Z_1$  et  $Z_2$ . Lequel des deux estimateurs semble converger le plus vite vers  $\theta$  ?

## 5.2 Estimateurs : méthodes d'obtention

Nous rappelons rapidement les méthodes les plus classiques d'obtention d'estimateurs. Notons qu'en général, ces méthodes n'assurent aucune bonne propriété aux estimateurs proposés et qu'il faut, une fois l'estimateur obtenu, l'étudier au cas par cas. Pour certains types de modèles, la méthode du maximum de vraisemblance assure de bonnes propriétés à l'estimateur : nous reviendrons là-dessus à la section 5.3.

La méthode la plus simple est la **méthode des moments**. Elle consiste, si l'on se place dans le cadre d'un modèle  $\mathcal{P}^{\otimes n}$ , à identifier des fonctions  $\psi$  et  $\varphi_1, \dots, \varphi_p$ , telles que

$$g(\theta) = \psi(\varphi_1(\mathbb{E}(X)), \dots, \varphi_p(\mathbb{E}(X^p))).$$

On estime alors  $g(\theta)$  par

$$\hat{g}(\theta) = \psi(\varphi_1(\bar{X}_n), \dots, \varphi_p(\bar{X}_n^p)).$$

En gros : si  $g(\theta)$  est une fonction des moments, on estime  $g(\theta)$  par la même fonction des moments empiriques. Tout ceci est bien plus compréhensible avec un exemple.

**Exemple 2** Soit  $X_1, \dots, X_n$  un  $n$ -échantillon de loi  $\beta(a, b)$ . On a

$$\mathbb{E}(X) = \frac{a}{a+b} \quad \text{var } X = \frac{ab}{(a+b)^2(a+b+1)}.$$

On en tire

$$a = \mathbb{E}(X) \left( \frac{\mathbb{E}(X)(1 - \mathbb{E}(X))}{\mathbb{E}(X^2) - \mathbb{E}(X)^2} - 1 \right) \quad b = (1 - \mathbb{E}(X)) \left( \frac{\mathbb{E}(X)(1 - \mathbb{E}(X))}{\mathbb{E}(X^2) - \mathbb{E}(X)^2} - 1 \right).$$

La méthode des moments suggère donc comme estimateurs

$$\hat{a} = \bar{X}_n \left( \frac{\bar{X}_n(1 - \bar{X}_n)}{\bar{X}_n^2 - \bar{X}_n} - 1 \right) \quad \hat{b} = (1 - \bar{X}_n) \left( \frac{\bar{X}_n(1 - \bar{X}_n)}{\bar{X}_n^2 - \bar{X}_n} - 1 \right).$$

Si les fonctions  $\psi, \varphi_1, \dots$  sont affines, alors par linéarité, l'estimateur proposé est sans biais. Si les fonctions  $\psi, \varphi_1, \dots$  sont continues et que les lois  $\mathbb{P}_\theta$  admettent des moments d'ordre  $p$  alors la loi des grands nombres assure la consistance de l'estimateur. Il faut cependant le rappeler au cas par cas.

### Exercice 54

- Montrez que  $Z_1$  dans l'exercice 53 ci-dessus aurait pu être trouvé par la méthode des moments.

- Soit  $P_1, \dots, P_n$  un  $n$ -échantillon de loi de Poisson de paramètre  $\lambda$ . Proposez deux estimateurs de  $\lambda$  différents, trouvés par la méthode des moments. Ces estimateurs sont-ils biaisés ? sont-ils consistants ? Le calcul théorique des risques quadratiques est pénible ; utilisez une simulation pour les comparer (on pourra se restreindre à  $\lambda \in \Lambda = [1, 3]$ ).

La **méthode par insertion** est similaire : il s'agit, si  $g(\theta)$  s'écrit comme une fonction d'un autre paramètre  $h(\theta)$ , par exemple  $g(\theta) = \psi(h(\theta))$ , et que l'on connaît un estimateur  $Z_h$  de  $h(\theta)$ , de proposer  $\psi(Z_h)$  pour estimateur de  $g(\theta)$ . Si l'estimateur  $Z_h$  est consistant (respectivement faiblement consistant) comme estimateur de  $h(\theta)$  et que  $\psi$  est continue (respectivement uniformément continue) alors  $\psi(Z_h)$  est consistant (respectivement faiblement consistant) comme estimateur de  $g(\theta)$ . Encore une fois, rien ne vaut un exemple :

**Exercice 55** Soit  $Z_1, \dots, Z_n$  un  $n$ -échantillon de loi  $\mathcal{N}(0, \sigma^2)$ . Calculez  $\mathbb{E}(|Z_1|)$  et déduisez-en un estimateur de  $\sigma^2$ . Cet estimateur est-il consistant ? Est-il sans biais ? si vous n'arrivez pas à répondre de manière théorique à l'une de ces questions, vous pouvez utiliser la simulation.

Enfin, la **méthode du maximum de vraisemblance** est la plus complexe mathématiquement mais aussi la plus universelle. On suppose que toutes les lois  $\mathbb{P}_\theta$  sont absolument continues par rapport à une mesure commune  $\mu$ . Cette mesure n'a pas besoin d'être une probabilité, et les applications les plus courantes sont les cas où  $\mu$  est la mesure de Lebesgue (l'hypothèse est alors que toutes les lois  $\mathbb{P}_\theta$  sont à densité) ou la mesure de comptage sur  $\mathbb{N}$  (l'hypothèse est alors que toutes les lois  $\mathbb{P}_\theta$  sont à support dans  $\mathbb{N}$ ). On note alors  $f_\theta$  la densité de  $\mathbb{P}_\theta$  par rapport à  $\mu$ , c'est-à-dire que  $f_\theta = \frac{d\mathbb{P}_\theta}{d\mu}$ . Une réalisation  $X$  de loi  $\mathbb{P}_\theta$  étant donnée, on propose comme estimateur de  $\theta$  la valeur (si elle est unique) de  $\theta$  qui maximise

$$V : \theta \mapsto f_\theta(X).$$

Encore une fois, en général on travaillera avec un  $n$ -échantillon  $X_1, \dots, X_n$ , de sorte que la densité à considérer sera  $f_\theta^{\otimes n}$  et que l'on maximisera

$$V_n : \theta \mapsto f_\theta(X_1) \dots f_\theta(X_n),$$

ou plutôt son logarithme.

**Exemple 3** On considère le modèle  $\{\mathbb{P}_{m, \sigma^2}^{\otimes n} \mid (m, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+\}$ . La fonction à maximiser est

$$\theta \mapsto \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X_i - m)^2}{2\sigma^2}}.$$

On passe donc à la log-vraisemblance

$$\begin{aligned} \log V_n(\theta) &= -\frac{\sum_{i=1}^n (X_i - m)^2}{2\sigma^2} - n \log \sigma + \text{constante} \\ &= -n \left( \frac{\overline{X^2} - 2m\overline{X} + m^2}{2\sigma^2} + \log \sigma \right) + \text{constante}. \end{aligned}$$

Quel que soit  $\sigma^2$ , le  $m$  maximiseur est  $\bar{X}$ . En réinjectant ce résultat dans l'expression on trouve que le  $\sigma^2$  maximiseur est  $\bar{X}^2 - \overline{X^2}$ .

On verra à la section 5.3 que l'estimateur du maximum de vraisemblance possède de bonnes propriétés pour certaines familles de modèles.

**Exercice 56** Reprenez les cas d'un  $n$ -échantillon  $U_1, \dots, U_n$  de loi  $\mathcal{U}([0, \theta])$ , d'un  $n$ -échantillon  $E_1, \dots, E_n$  de loi  $\mathcal{E}(\lambda)$ , d'un  $n$ -échantillon  $P_1, \dots, P_n$  de loi  $\mathcal{P}(\lambda)$ , et donnez leurs estimateurs obtenus par la méthode du maximum de vraisemblance.

**Exercice 57** On considère un  $n$ -échantillon de loi  $\mathbb{P}_\alpha$ , de densité  $\frac{1+\alpha x}{2} \mathbb{1}_{[-1, +1]}(x)$  avec  $\alpha \in [-1, +1]$ . Quelle équation doit vérifier l'estimateur du maximum de vraisemblance de  $\alpha$ ? on verra que l'on ne peut donner son expression explicite (donc pas prouver ses propriétés à la main). Le fichier `xx.alpha.inconnu` donne un 1 000 000-échantillon de loi  $\mathbb{P}_\alpha$ , pour un certain  $\alpha$ . Comment a été simulé cet échantillon, d'après vous? En utilisant la fonction `scipy.optimize.fsolve` pour trouver le  $\alpha$  maximisateur, illustrez le fait que l'estimateur du maximum de vraisemblance est consistant.

### 5.3 Borne de Cramér-Rao et modèles exponentiels

On revient ici sur le risque quadratique associé à un estimateur. On a vu dans l'exercice 52 que la vitesse de décroissance en  $n$  du risque quadratique donne une indication de la vitesse de convergence de l'estimateur. Il est donc intéressant d'avoir un risque quadratique minimal. La borne de Cramér-Rao donne une borne inférieure pour ce risque quadratique, montrant que l'on ne peut espérer faire mieux qu'une certaine quantité.

On se place maintenant dans le cadre d'un modèle régulier, c'est à dire que l'on suppose :

- que  $\Theta$  est un ouvert,
- que toutes les lois  $\mathbb{P}_\theta$  ont même support  $S$  et sont absolument continues par rapport à une mesure commune  $\mu$ , et on note encore  $f_\theta = \frac{d\mathbb{P}_\theta}{d\mu}$ ,
- que  $\theta \mapsto \log f_\theta$  est deux fois continûment dérivable en  $\mu$ -presque tout point de  $S$ , et de carré intégrable.

On définit alors

**Définition 5.3.1** On suppose que le modèle  $\{\mathbb{P}_\theta \mid \theta \in \Theta \subset \mathbb{R}\}$  est régulier. On appelle information de Fisher du modèle la fonction

$$I(\theta) = \mathbb{E}_\theta \left( \left( \frac{\partial}{\partial \theta} \log f_\theta(X) \right)^2 \right) = -\mathbb{E}_\theta \left( \frac{\partial^2}{\partial \theta^2} \log f_\theta(X) \right). \quad (5.1)$$

Cette définition valable pour  $\theta$  réel s'étend immédiatement au cas  $\Theta \subset \mathbb{R}^d$ . Il faut alors remplacer les dérivées par des gradients et les carrés par des produits scalaires (de sorte que le résultat  $I(\theta)$  est toujours un réel).

On a la borne inférieure suivante sur le risque quadratique :

**Théorème 5.3.2 (borne de Cramér-Rao)** Pour  $\{\mathbb{P}_\theta \mid \theta \in \Theta\}$  un modèle régulier, on a (sous quelques conditions de régularité supplémentaires) que tout estimateur sans biais de  $g(\theta)$  a un risque quadratique supérieur à  $(g'(\theta))^2 / I(\theta)$ .

La preuve est donnée dans [2]. Un estimateur sans biais dont le risque quadratique est, pour tout  $\theta$ , égal à la borne inférieure ci-dessus sera appelé UMVB, pour “de variance uniformément minimale, sans biais”.

**Exercice 58** Reprenez les trois exemples de l'exercice 56. Peut-on leur appliquer le théorème précédent ? Calculez l'information de Fisher du modèle, puis comparez le risque quadratique des estimateurs du maximum de vraisemblance à la borne de Cramér-Rao.

Si l'on se restreint encore quant au type de modèle considéré, on peut obtenir un résultat général qui montre que ces modèles atteignent la borne de Cramér-Rao, au moins asymptotiquement. Commençons par définir ces modèles : on dit qu'un modèle est exponentiel s'il est régulier avec des densités  $f_\theta$  de la forme

$$f_\theta(x) = \phi(x) \exp(a(\theta)h(x) - b(\theta)).$$

**Exemple 4** Tout modèle  $\mathcal{P}^{\otimes n}$  associé à des  $n$ -échantillons de loi  $\mathcal{P}(\lambda)$ ,  $\mathcal{N}(\mu, \sigma^2)$ ,  $\mathcal{E}(\lambda)$  est un modèle exponentiel.

On a les deux théorèmes suivants, démontrés dans [2] :

**Théorème 5.3.3** Pour  $\{\mathbb{P}_\theta \mid \theta \in \Theta\}$  un modèle exponentiel (avec quelques conditions de régularité), alors  $h(X)$  est l'estimateur du maximum de vraisemblance de  $b'(\theta)/a'(\theta)$ , et il est UMVU.

**Exercice 59** Expliquez les résultats de l'exercice 58 à la lumière de ce théorème.

Le théorème ci-dessus montre que la méthode du maximum de vraisemblance fournit un estimateur UMVU... mais pas forcément pour  $\theta$  lui-même. Le résultat suivant montre qu'elle fournit aussi un estimateur pour  $\theta$  qui atteint asymptotiquement la borne de Cramér-Rao :

**Théorème 5.3.4** Soit  $\mathcal{P} = \{\mathbb{P}_\theta \mid \theta \in \Theta\}$  un modèle exponentiel (avec quelques conditions de régularité supplémentaires). On considère  $\mathcal{P} = \mathcal{P}^{\otimes n}$  le modèle associé à un  $n$ -échantillon. L'estimateur du maximum de vraisemblance  $\hat{\theta}_n$  de  $\theta$  vérifie

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I_1(\theta)^{-1}) \quad (5.2)$$

où  $I_1(\theta)$  est l'information de Fisher associée à  $\mathcal{P}$  (donc à une seule variable  $X_1$ ).

La preuve utilise la méthode delta décrite dans le Lemme 5.4.3.

L'exercice 68 énoncé plus loin illustre le fait que la normalité asymptotique est un phénomène général qui ne se limite pas aux modèles exponentiels.

### 5.4 Intervalles de confiance : définitions

On commence par donner une définition générale d'un intervalle de confiance associé à une réalisation  $X$  d'un modèle  $\mathcal{P}$ .

**Définition 5.4.1** *Un intervalle de confiance pour  $g(\theta)$  est un intervalle aléatoire  $I(\omega)$ , dont les bornes sont des fonctions mesurables de  $X$ . On dit que l'intervalle de confiance est de niveau (de confiance)  $1 - \alpha$  pour  $\alpha \in ]0, 1[$ , si  $\mathbb{P}(g(\theta) \in I) \geq 1 - \alpha$ .*

Notez bien que dans l'expression  $\mathbb{P}(\theta \in I)$ , c'est  $I$  qui est aléatoire !

Un intervalle peut être bilatère, c'est-à-dire de la forme  $I(\omega) = [A(\omega), B(\omega)]$ , ou bien unilatère, c'est-à-dire de l'une des formes  $I(\omega) = [A(\omega), +\infty[$  ou  $I(\omega) = ] - \infty, B(\omega)]$ . Les intervalles de confiance unilatères prendront un sens quand nous ferons le lien entre intervalles de confiances et tests paramétriques. Pour l'instant, nous allons nous concentrer sur le cas bilatère.

Un intervalle de confiance de niveau  $1 - \alpha$  pour le paramètre  $\theta$  donne une indication supplémentaire par rapport à un estimateur  $\hat{\theta}$  : il dit qu'avec probabilité supérieure à  $1 - \alpha$ , on connaît  $\theta$  à avec une certaine précision. Il est évident que plus on choisit  $\alpha$  petit (donc plus on veut de certitude sur notre estimation), plus l'intervalle devra être grand (et donc moins on aura d'information – mais avec plus de certitude).

Le lien avec les estimateurs vient du fait que l'on obtient en général les intervalles de confiance à partir des estimateurs. Si par exemple on a un estimateur  $\hat{\theta}$  de  $\theta$ , on cherchera  $\varepsilon$  tel que  $\mathbb{P}(|\hat{\theta} - \theta| > \varepsilon) \leq \alpha$ . Dans ce cas, l'intervalle  $[\hat{\theta} - \varepsilon, \hat{\theta} + \varepsilon]$  sera un intervalle de confiance de niveau  $1 - \alpha$ . Un tel intervalle peut s'obtenir de différentes façons : soit l'on connaît exactement la loi de  $\hat{\theta} - \theta$ , soit l'on utilise des inégalités de type Markov, Bienaymé-Tchebychev, ou Hoeffding (voir les séances précédentes).

Il est important de remarquer que, lorsque l'on recherche un intervalle de confiance ou intervalle de confiance asymptotique de niveau de confiance  $1 - \alpha$ , on impose une probabilité  $\geq 1 - \alpha$ . Si cette inégalité a été obtenue à partir de majorations peu efficaces, on va avoir un intervalle de confiance inutilement grand. Autrement dit on n'obtient pas, pour un niveau de confiance donné, l'information optimale sur la précision de l'estimation. Dans la suite, nous allons proposer un certain nombre d'intervalles de confiance de niveau  $1 - \alpha$ , puis utiliserons dans l'exercice 63 la simulation pour estimer le niveau réel des intervalles de confiance, c'est-à-dire la valeur de la probabilité  $\mathbb{P}(g(\theta) \in I)$ .

#### Exercice 60

1. Soit  $X_1, \dots, X_n$  un  $n$ -échantillon de loi normale de variance connue  $\mathcal{N}(m, \sigma_0^2)$ . On propose l'estimateur  $\bar{X}_n$  de  $m$ . Donnez un intervalle de confiance pour  $m$  en utilisant le fait que  $\bar{X}_n - m$  suit une loi connue.
2. Soit  $X_1, \dots, X_n$  un  $n$ -échantillon de loi normale de variance inconnue  $\mathcal{N}(m, \sigma_0^2)$ . Quel estimateur peut-on proposer pour  $m$  ? et pour  $\sigma^2$  ? Donnez des intervalles de confiance pour  $m$  et  $\sigma^2$ .
3. Soit  $B_1, \dots, B_n$  un  $n$ -échantillon de variables de Bernoulli de paramètre  $p$ . On propose l'estimateur  $\bar{B}_n$  de  $p$ . Donnez trois intervalles de confiance pour  $p$  :

- l'un en utilisant l'inégalité de Bienaymé-Tchebychev et l'inégalité  $p(1-p) \leq 1/4$ ,
- l'autre en utilisant l'inégalité de Bienaymé-Tchebychev et en résolvant une équation du second degré,
- la dernière en utilisant l'inégalité de Hoeffding.

On va aussi s'intéresser à la notion d'intervalle de confiance asymptotique, associé à une réalisation  $X_1, \dots, X_n$  d'un modèle  $\mathcal{P}_n$  (qui n'est pas forcément de la forme  $\mathcal{P}^{\otimes n}$ , autrement dit les  $X_1, \dots, X_n$  ne sont pas forcément i.i.d).

**Définition 5.4.2** *Un intervalle de confiance asymptotique pour le paramètre  $\theta$  est la donnée pour tout  $n$  d'un intervalle de confiance  $I_n(\omega)$ ; on dit qu'il est niveau (de confiance) asymptotique  $1 - \alpha$  pour  $\alpha \in ]0, 1[$ , si  $\liminf_{n \rightarrow \infty} \mathbb{P}(\theta \in I_n) \geq 1 - \alpha$ .*

C'est pour obtenir des intervalles de confiance asymptotiques que la notion de vitesse prend tout son intérêt. Supposons par exemple qu'une suite d'estimateurs  $(Z_n)_n$  est de vitesse  $(a_n)_n$ , c'est à dire que

$$a_n(Z_n - \theta) \xrightarrow{\mathcal{L}} \text{loi limite.}$$

Alors si l'on note  $q_{\alpha/2}, q_{1-\alpha/2}$  les quantiles de la loi limite, et que l'on suppose que ce ne sont pas des atomes<sup>1</sup>, on a immédiatement un intervalle de confiance asymptotique :

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\theta \in \left[Z_n - \frac{q_{1-\alpha/2}}{a_n}, Z_n - \frac{q_{\alpha/2}}{a_n}\right]\right)$$

**Exercice 61** *On reprend  $B_1, \dots, B_n$  un  $n$ -échantillon de variables de Bernoulli de paramètre  $p$  et l'estimateur  $\bar{B}_n$  de  $p$ . Quelle est la vitesse de  $\bar{B}_n$  comme estimateur de  $p$ ? Donnez un intervalle de confiance asymptotique pour  $p$ .*

Un résultat utile pour discuter de la normalité asymptotique d'estimateurs est ce que l'on appelle la méthode delta :

**Lemme 5.4.3** *Soit  $(\hat{\theta}_n)_n$  une suite de variables aléatoires telle que pour tout  $\theta \in \Theta$ ,*

$$a_n(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} Z_\theta$$

*où  $(a_n)_n$  est une suite qui croît vers  $+\infty$  et  $Z_\theta$  une variable aléatoire dont la loi dépend de  $\theta$ . Soit  $g$  une fonction à valeurs dans  $\mathbb{R}^q$ , différentiable sur un ouvert contenant  $\Theta$ . On a*

$$a_n(g(\hat{\theta}_n) - g(\theta)) \xrightarrow{\mathcal{L}} Dg(\theta)(Z_\theta).$$

Autrement dit, en général, si  $(\hat{\theta}_n)_n$  est de vitesse  $(a_n)_n$ , alors  $(g(\hat{\theta}_n))_n$  aussi.

---

1. rappelons que si  $X_n \xrightarrow{\mathcal{L}} X$  alors  $\mathbb{P}(X_n \leq x) \rightarrow \mathbb{P}(X \leq x)$  n'est vrai *a priori* que si  $x$  n'est pas un atome de  $X$ .

**Exercice 62** On reprend  $B_1, \dots, B_n$  un  $n$ -échantillon de variables de Bernoulli de paramètre  $p$  et l'estimateur  $\bar{B}_n$  de  $p$ . En utilisant la méthode delta, proposez un estimateur qui vérifie une convergence vers une loi qui ne dépend pas de  $p$ . En déduire un nouvel intervalle de confiance asymptotique pour  $p$ .

Les exercices suivants utilisent la simulation pour estimer le niveau réel d'un intervalle de confiance, c'est-à-dire la valeur de la probabilité  $\mathbb{P}(g(\theta) \in I)$ .



**Exercice 63**

- Reprenez les quatre intervalles de confiance ou intervalles de confiances asymptotiques proposés pour le paramètre  $p$  d'une loi binomiale  $(n, p)$ . Pour  $n = 10, 50, 100$  et différentes valeurs de  $p$ , répétez  $N = 10000$  fois l'opération suivante : simulez  $B_1, \dots, B_n$ , calculez les intervalles de confiance associés et la proportion, sur les  $N$  réalisations, de fois où  $p$  est bien dans l'intervalle.
- Dans l'événement ci-dessus, on est en train d'estimer la probabilité  $\mathbb{P}(p \in I)$  à partir d'un  $N$ -échantillon de loi binomiale. Quelle précision donne l'expérience ci-dessus ?

**Exercice 64** Reprenez les différents intervalles de confiance proposés dans l'exercice 60 et estimez leur niveau réel.

**5.5 Tests d'hypothèses : définitions générales**

Nous partons d'un modèle paramétrique  $\mathcal{P} = \{\mathbb{P}_\theta, \theta \in \Theta\}$ , et nous donnons deux sous-ensembles disjoints  $\Theta_0$  et  $\Theta_1$  de  $\Theta$ . Nous ne supposons pas que  $\Theta_0 \cup \Theta_1 = \Theta$ ; la raison en sera plus claire sur des exemples. Nous allons considérer les deux hypothèses suivantes :

$$H_0 : \theta \in \Theta_0 \quad H_1 : \theta \in \Theta_1.$$

On verra toujours  $H_0$  comme l'hypothèse a priori, et  $H_1$  comme l'hypothèse alternative. Autrement dit, le test vise "à ne rejeter  $H_0$  que si l'on a de bonnes raisons de le faire", et le choix des critères de rejet va dépendre de la forme de  $H_1$ .

**Exemple 5** Dupond et Dupont jouent à pile ou face : pile fait gagner Dupond, face fait gagner Dupont. A priori, la pile est équilibrée (la probabilité  $p$  de faire pile vaut  $1/2$ ) mais les deux policiers se soupçonnent de tricher. Si Dupond veut faire le test, il va naturellement considérer  $H_0^d : p = 1/2$  contre  $H_1^d : p < 1/2$  (puisque si Dupont triche, c'est pour gagner). Inversement, si Dupont veut faire le test, il va considérer  $H_0^t : p = 1/2$  contre  $H_1^t : p > 1/2$ . On voit bien que les ensembles  $\Theta_0$  et  $\Theta_1$  ne sont pas complémentaires, ce qui correspond à des hypothèses faites sur le modèle, hypothèses qu'il va s'agir d'exploiter.

**Définition 5.5.1** Un test de  $H_0$  contre  $H_1$  est une fonction  $\phi(X)$  à valeurs dans  $\{0, 1\}$ , à laquelle on associe la règle de décision : si  $\phi(X) = 0$ , on conserve  $H_0$  et si  $\phi(X) = 1$ , on rejette  $H_0$ . On définit les erreurs de première espèce et de seconde espèce associées :

$$\begin{array}{ll} \underline{\alpha} : \Theta_0 & \rightarrow [0, 1] \\ \theta & \mapsto \mathbb{P}_\theta(\phi(X) = 1) \end{array} \quad \begin{array}{ll} \underline{\beta} : \Theta_1 & \rightarrow [0, 1] \\ \theta & \mapsto \mathbb{P}_\theta(\phi(X) = 0). \end{array}$$

On dit qu'un test est de niveau  $\alpha$  si  $\sup_{\theta \in \Theta_0} \underline{\alpha} \leq \alpha$ .

Les erreurs de première et seconde espèce caractérisent les probabilités des deux manières de se tromper : respectivement, rejeter  $H_0$  à tort (donc observer  $\phi(X) = 1$  alors que  $\theta \in \Theta_0$ ), ou conserver  $H_0$  à tort (donc observer  $\phi(X) = 0$  alors que  $\theta \in \Theta_1$ ). On parle aussi de la puissance du test, qui est simplement la fonction  $1 - \underline{\beta}$ .

### 5.6 Tests d'ajustement à un paramètre

Dans cette section, nous considérons uniquement des tests d'ajustement à un paramètre, autrement dit : nous testons si la valeur d'un sous-paramètre  $g(\theta)$  est bien une valeur proposée  $g(\theta_0)$ . La construction du test va dépendre de l'hypothèse alternative, mais nous commençons par discuter la situation d'un test bilatère, et explicitons le lien avec les intervalles de confiance. Etudions un exemple :

**Exemple 6** On considère le modèle associé à un  $n$ -échantillon  $X_1, \dots, X_n$  de loi  $\mathcal{N}(m, \sigma_0^2)$ , où  $\sigma_0^2$  est connu et  $m$  inconnu. On considère les hypothèses  $H_0 : m = m_0$  contre  $H_1 : m \neq m_0$ . Un tel test est appelé bilatère. D'après la séance précédente, si l'on note  $\bar{X}_n$  la moyenne empirique de l'échantillon, alors l'intervalle  $[\bar{X}_n - u_\alpha \frac{\sigma_0}{\sqrt{n}}, \bar{X}_n + u_\alpha \frac{\sigma_0}{\sqrt{n}}]$  (où  $u_\alpha = q_{1-\alpha/2}$ , quantile associé à la loi normale centrée réduite), est un intervalle de confiance de niveau de confiance  $\alpha$ . Il est par conséquent évident que

$$\phi(X_1, \dots, X_n) = \mathbb{1}_{m_0 \notin [\bar{X}_n - u_\alpha \frac{\sigma_0}{\sqrt{n}}, \bar{X}_n + u_\alpha \frac{\sigma_0}{\sqrt{n}}]}$$

est un test de niveau  $\alpha$ , puisque sa probabilité dans le cas  $m = m_0$  (correspondant ici à l'hypothèse  $H_0$ ) est inférieure à  $\alpha$ .

Observons maintenant que

$$\begin{aligned} m_0 \notin [\bar{X}_n - u_\alpha \sigma_0, \bar{X}_n + u_\alpha \sigma_0] &\Leftrightarrow \bar{X}_n > m_0 + u_\alpha \sigma_0 \quad \text{ou} \quad \bar{X}_n < m_0 - u_\alpha \sigma_0 \\ &\Leftrightarrow \frac{\bar{X}_n - m_0}{\sigma_0} > u_\alpha \quad \text{ou} \quad \frac{\bar{X}_n - m_0}{\sigma_0} < -u_\alpha. \end{aligned} \tag{5.3}$$

On peut donc définir le test comme basé sur l'observation suivante :

- sous l'hypothèse  $H_0$ , on sait, grâce à l'estimateur  $\bar{X}_n$ , définir une fonction de  $X_1, \dots, X_n$  et de paramètres connus, dont on connaît la loi sous  $H_0$  : ici, c'est  $Z = \frac{\bar{X}_n - m_0}{\sigma_0}$ , qui suit une loi normale  $\mathcal{N}(m_0, \sigma_0^2)$  ;
- sous l'hypothèse  $H_1$ , on sait que  $\bar{X}_n$  sera plutôt plus grande (si  $m > m_0$ ) que sous  $H_0$ , ou plutôt plus petite (si  $m < m_0$ ) que sous  $H_0$ , et la même chose est vraie pour  $Z$ .

On va donc rejeter l'hypothèse  $H_0$  au profit de l'hypothèse  $H_1$  si les valeurs observées de  $\bar{X}_n$  sont "trop grandes pour être crédibles sous  $H_0$ " ou "trop petites pour être crédibles sous  $H_1$ ". On se sert alors de la loi explicite sous  $H_0$  pour préciser ce qu'on entend par "trop grandes/petites pour être crédibles sous  $H_0$ ". C'est l'une des formulations de droite dans (5.3).

La correspondance ci-dessus entre intervalles de confiance et tests bilatères fonctionne encore avec les tests unilatères, c'est-à-dire du type  $H'_0 : m = m_0$  contre  $H'_1 : m > m_0$  ou bien  $H''_0 : m = m_0$  contre  $H''_1 : m < m_0$ , à condition de considérer des intervalles de confiance unilatère. Dans ce cas-là, il est cependant plus intuitif de raisonner sur le comportement des variables aléatoires, qui sont en général un estimateur du paramètre que l'on veut tester.

En résumé, on construit en général un test d'ajustement à un paramètre de la manière suivante :

- on propose un estimateur du paramètre testé,
- on trouve une fonction  $Z$  de cet estimateur, qui implique pour seule quantité inconnue le paramètre testé, et dont on connaît exactement la loi sous l'hypothèse  $H_0$ ,
- on étudie comment il se comporte sous  $H_1$  par opposition à  $H_0$  (plutôt plus grand, plutôt plus petit...)
- on définit la règle de rejet de  $H_0$  en conséquence.

On va illustrer ce principe avec des exercices :

**Exercice 65** On considère un  $n$ -échantillon  $X_1, \dots, X_n$  de loi  $\mathcal{N}(m, \sigma^2)$ . Rappelez les estimateurs usuels de  $m$  et  $\sigma^2$  dans les différents cas ( $m$  connu ou inconnu,  $\sigma$  connu ou inconnu). Construisez par exemple les tests  $m = m_0$  contre  $m > m_0$  pour  $\sigma$  connu,  $m = m_0$  contre  $m \neq m_0$  pour  $\sigma$  inconnu,  $\sigma = \sigma_0$  contre  $\sigma < \sigma_0$  pour  $m$  inconnu.

#### Exercice 66

1. Soit  $X_1, \dots, X_n$  un  $n$ -échantillon de loi  $\mathcal{N}(m, \sigma^2)$  avec  $m$  et  $\sigma$  inconnus. Donnez un estimateur de  $m$ , et déterminez sa loi. Définissez un test de  $H_0 : m = m_0$  contre  $H_1 : m > m_0$ .
2. Définissez une fonction qui, si on lui donne une variable `xech` représentant le  $n$ -échantillon  $X_1, \dots, X_n$  en entrée, donne le résultat du test. Codez cette fonction de manière à ce qu'elle puisse accepter en entrée un tableau  $N \times n$  représentant  $N$  différents  $n$ -échantillons.
3. On se place dans le cas  $m = m_0 = 2$ ,  $\alpha = 5\%$ , et, pour conserver notre ignorance de  $\sigma$ , on tirera au hasard une valeur dans  $[1, 2]$ ; on veut estimer le niveau réel  $\underline{\alpha}(m_0)$ . On va donc faire  $N$  fois l'expérience qui consiste à simuler  $n$  variables de loi  $\mathcal{N}(m, \sigma^2)$ , à appliquer le test à ce  $n$ -échantillon et à compter combien de fois, sur les  $N$ , on rejette (à tort, donc) l'hypothèse  $H_0$ . Sachant que la vraie valeur de  $p = \underline{\alpha}(m_0)$  est de l'ordre de 0,05, quelle valeur de  $N$  choisir pour avoir une estimation (de niveau de confiance 95%) de  $p$  à 0,01 près ?
4. Implémentez l'expérience ci-dessus, et estimez le niveau réel du test pour différentes valeurs de la taille  $n$  par exemple (5,10,50,100).
5. On veut maintenant estimer la puissance du test. Estimez la puissance du test pour  $n = 5, 10, 50, 100$  et  $m$  variant de 2,1 à 3 par pas de 0,1. Représentez les résultats.

## 5.7 Exercices

**Exercice 67** Soit  $X_1, \dots, X_n$  un  $n$ -échantillon de loi  $\mathcal{N}(\theta, \theta^2)$ , où  $\theta$  est un réel strictement positif. La densité correspondante est notée  $f_\theta(x)$ .

1. A quels estimateurs pouvez-vous penser pour estimer le paramètre  $\theta$  ? Quelle est leur variance ?

2. Calculez la log-vraisemblance associée aux observations pour le paramètre  $\theta$  et en déduire que le maximum de vraisemblance  $\hat{\theta}_n$  est défini par :

$$\hat{\theta}_n = -\frac{\bar{X}}{2} + \sqrt{X^2 + \frac{1}{4}\bar{X}^2}.$$

3. Ecrivez un programme simulant  $N$  réalisations de l'estimateur  $\hat{\theta}_n$  et illustrez sa consistance, que vous vérifierez également par le calcul. Comparez les convergences vers  $\theta$  de  $\hat{\theta}_n$  et des estimateurs proposés en 1.
4. Montrez que  $I_1(\theta)$  vaut  $3/\theta^2$  puis voyez si le  $N$ -échantillon construit en 2. a une variance proche de  $(nI_1(\theta))^{-1}$ .

**Exercice 68** Reprenez l'énoncé de l'exercice 57. Tracez la fonction de répartition empirique d'un 1000-échantillon de la variable  $\sqrt{n}(\hat{\alpha}_n - \alpha)$  où  $\hat{\alpha}_n$  est l'estimateur de  $\alpha$  obtenu pour  $n = 1000$ , à partir des données `xx_alpha_inconnu`. Qu'illustre ce tracé ?

## Bibliographie

- [1] J.-F. Delmas, *Introduction au calcul des probabilités et de la statistique*.
- [2] V. Rivoirard, G. Stoltz, *Statistique mathématique en action*.



## Chapitre 6

# Martingales et applications

### 6.1 Quelques résultats de convergence

Les résultats classiques de convergence sont donnés dans [1], [2].

**Théorème 6.1.1** Soit  $(M_n)_n$  une martingale. Si cette martingale est bornée dans  $L^p$  au sens où  $\sup_n \mathbb{E}(|M_n|^p) < \infty$ , alors :

- pour  $p = 1$ , le processus  $M_n$  converge presque-sûrement vers une variable aléatoire intégrable,
- pour  $p > 1$ , le processus  $M_n$  converge presque-sûrement et dans  $L^p$  vers une variable  $L^p$ .

Attention à ne pas affirmer la convergence  $L^1$  dans le cas  $p = 1$ , qui n'est vraie que sous l'hypothèse supplémentaire que  $(X_n)_n$  est uniformément intégrable. Un contre-exemple standard est  $M_n = 2^{X_1 + \dots + X_n}$  où  $(X_n)_n$  est i.i.d. de loi  $\frac{1}{3} \delta_{-1} + \frac{1}{3} \delta_{+1}$ . On a souvent l'usage des résultats concernant les sous- ou sur-martingales ; on rappelle qu'une sur-martingale est un processus  $(A_n)_n$  tel que  $\mathbb{E}(A_{n+1} | \mathcal{F}_n) \leq A_n$  et une sous-martingale un processus  $(B_n)_n$  tel que  $\mathbb{E}(B_{n+1} | \mathcal{F}_n) \geq B_n$ .

**Théorème 6.1.2** Soit  $(A_n)_n$  une sur-martingale positive, alors elle converge presque-sûrement vers une variable aléatoire  $A_\infty$  intégrable.

Soit  $(B_n)_n$  une sous-martingale telle que  $\sum_n \mathbb{E}(B_n^+) < \infty$ , alors elle converge presque-sûrement vers une variable aléatoire  $B_\infty$  intégrable.

Il existe également des résultats de type "loi des grands nombres" et "central limite" pour les martingales. Ces résultats dépendent de la quantité

$$\langle M \rangle_n = \sum_{k=1}^n \mathbb{E}((M_k - M_{k-1})^2 | \mathcal{F}_{k-1}).$$

On a alors le résultat suivant (voir encore une fois [1] ou [2]) :

**Théorème 6.1.3** Soit  $(M_n)_n$  une martingale de carré intégrable. Alors :

- sur l'événement  $[\lim_n \langle M \rangle_n = \infty]$ , on a  $\frac{M_n}{\langle M \rangle_n} \xrightarrow{\text{p.s.}} 0$ ,

- sur l'événement  $[\lim_n \langle M \rangle_n < \infty]$ , la suite  $M_n$  converge presque-sûrement.

Donner une vitesse de convergence en général est un peu pénible (voir [2]). On se contentera de donner un théorème central limite, qui donne une condition pour avoir une convergence en loi  $\frac{1}{\sqrt{a_n}} M_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$ . On va y retrouver des conditions dites “de Lindeberg” et “de Lyapounov” analogues à celles que nous avons rencontrées pour les sommes de variables indépendantes dans le TP3.

**Théorème 6.1.4** Soit  $(M_n)_n$  une martingale de carré intégrable, et soit  $(\langle M \rangle_n)_n$  son processus croissant. On suppose que  $(a_n)_n$  est une suite de réels positifs croissant vers l'infini telle que

$$\frac{\langle M \rangle_n}{a_n} \xrightarrow{\mathbb{P}} \ell$$

pour un  $\ell \geq 0$  déterministe, et que pour tout  $\epsilon > 0$

$$\frac{1}{a_n} \sum_{k=1}^{a_n} \mathbb{E}(|\Delta M_k|^2 \mathbf{1}_{|\Delta M_k| \geq \epsilon \sqrt{a_n}} | \mathcal{F}_{k-1}) \xrightarrow{\mathbb{P}} 0 \quad (6.1)$$

(où  $\Delta M_k = M_k - M_{k-1}$ ). Alors on a

$$\frac{1}{\sqrt{a_n}} M_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \ell)$$

et si  $\ell > 0$

$$\sqrt{a_n} \frac{M_n}{\langle M \rangle_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \ell^{-1}).$$

La condition (6.1) est appelée *condition de Lindeberg* (sur les accroissements  $\Delta M_k$  de la martingale). Elle est vraie s'il existe  $\delta > 0$  tel que

$$\frac{1}{a_n} \sum_{k=1}^n \mathbb{E}(|\Delta M_k|^{2+\delta} | \mathcal{F}_{k-1}) \xrightarrow{\mathbb{P}} 0, \quad (6.2)$$

condition dite *de Lyapounov*.

Un dernier résultat qui nous sera utile est le théorème de Robbins-Siegmund :

**Théorème 6.1.5** Soient  $(V_n)_n$ ,  $(A_n)_n$ ,  $(B_n)_n$  trois suites positives  $(\mathcal{F}_n)_n$ -adaptées telles que presque-sûrement  $V_0$  est finie et

$$\mathbb{E}(V_{n+1} | \mathcal{F}_n) \leq V_n + A_n - B_n.$$

Alors sur l'événement  $[\sum_n A_n < \infty]$ , on a  $V_n \xrightarrow{\text{p.s.}} V_\infty$  où  $V_\infty$  est une variable aléatoire finie presque-sûrement, et de plus

$$\sum_n B_n < \infty.$$



## 6.2 Processus de Galton-Watson

Les processus de Galton-Watson ont été introduits par Galton et Watson en 1874, pour étudier la question de la disparition éventuelle d'un nom de famille en fonction de la distribution du nombre d'enfants mâles qu'aura un homme de la famille<sup>1</sup>.

Une formulation générale est la suivante : si  $Z$  est une variable aléatoire à valeurs dans  $\mathbb{N}$ , on dit que  $(X_n)_n$  est un processus de Galton-Watson de loi de reproduction  $Z$  si  $X_0 = 1$  et

$$X_{n+1} = \sum_{k=1}^{X_n} Z_{k,n+1} \quad (6.3)$$

où  $(Z_{k,n+1})_k$  est une suite de variables aléatoires i.i.d., indépendante de  $X_0, \dots, X_n$  et de même loi que  $Z$ . On notera dans tout le texte  $\mu$  et  $\sigma^2$  les espérance et variance de  $Z$ . Notons qu'en particulier on utilise la convention qu'une somme  $\sum_{k=1}^0$  vaut zéro, de sorte que si  $X_n = 0$  alors  $X_{n+1} = 0$ . On notera également

$$Y_n = \sum_{k=0}^n X_k = 1 + \sum_{k=1}^n X_k \quad (6.4)$$

la population totale jusqu'à la génération  $n$ . Les processus de Galton-Watson sont utiles pour étudier la question de l'évolution d'une population, que ce soit celle des mâles de nom de famille donné, de malades contaminés par un virus, ou de neutrons émis dans une masse d'atome d'uranium 235. Dans un tel cadre, on s'intéresse naturellement aux questions suivantes : quelle est la probabilité  $\mathbb{P}(X_n \rightarrow 0)$  que la population finisse par s'éteindre ? si ce n'est pas le cas, la population  $X_n$  va-t-elle rester bornée ou bien exploser ? et pour la population totale  $Y_n$  ? Nous allons, bien sûr, répondre à ces questions. Mais d'abord, un peu de simulation.

**Exercice 69** *Ecrivez une fonction `simulGW` qui, pour une variable d'entrée `n`, donne en sortie une variable `x` qui est une réalisation de la suite  $(X_0, \dots, X_n)$ . On aura le droit d'utiliser une boucle. On commencera par l'écrire en supposant qu'on a déjà défini une fonction `simulZ` telle que `simulZ(k)` donne `k` réalisations indépendantes de la variable  $Z$  si  $k \neq 0$  (l'intérêt est que, si l'on veut changer de loi pour  $Z$ , il suffira de faire la modification dans `simulZ` pour qu'elle se répercute partout où l'on se sert de  $Z$ ).*

*Simulez ensuite des trajectoires du processus  $(X_n)_n$  dans les cas suivants :*

- dans le cas où  $Z$  prend les valeurs 0, 1, 2 avec des probabilités  $p_0, p_1, p_2$  ;
- dans le cas où  $Z$  suit une loi géométrique de paramètre  $p$  à valeurs dans  $\mathbb{N}$ , c'est-à-dire  $\mathbb{P}(Z = k) = p(1-p)^k$ .

*Testez vos programmes avec différents choix de  $p_0, p_1, p_2$  dans le premier cas, de  $p$  dans le deuxième. Notez les valeurs pour lesquelles on semble avoir  $X_n \rightarrow 0$  à chaque fois, et celles pour lesquelles ce n'est pas le cas.*

1. plus précisément, Galton a posé la question dans le *Educational Times* en 1873, Watson y a répondu et ils ont écrit ensemble un article, *On the probability of extinction of families* en 1874.

On s'intéresse d'abord à la probabilité d'extinction

$$q = \mathbb{P}(\lim_{n \rightarrow \infty} X_n = 0) = \mathbb{P}(\exists n \text{ tel que } X_n = 0). \quad (6.5)$$

Cette probabilité s'exprime simplement :

**Lemme 6.2.1** Si l'on note  $q_n = \mathbb{P}(X_n = 0)$ , alors la suite  $(q_n)_n$  converge et  $q = \lim_n q_n$ .

**Preuve.** Puisque par définition,  $X_n = 0$  implique  $X_{n+1} = 0$ , la suite d'événements  $(X_n = 0)_n$  est croissante, donc la suite  $(q_n)_n$  est croissante et majorée par 1, donc elle converge. De plus, les valeurs de  $(X_n)_n$  étant entières,  $X_n \rightarrow 0$  si et seulement si  $X_n = 0$  pour  $n$  assez grand. On a donc

$$q = \mathbb{P}(\lim_{n \rightarrow \infty} X_n = 0) = \mathbb{P}\left(\bigcup_n (X_n = 0)\right) = \lim_{n \rightarrow \infty} \mathbb{P}(X_n = 0) = \lim_n q_n. \quad \square$$

On va commencer par étudier une situation où la réponse est simple, en donnant au passage une identité qui nous servira plus tard.

### Exercice 70

1. Montrez l'identité  $\mathbb{E}(X_{n+1}|X_n) = \mathbb{E}(Z) \times X_n$ .
2. Déduisez-en que  $\mathbb{E}(X_n) = \mu^n$ , où  $\mu = \mathbb{E}(Z)$ .
3. Prouvez que  $\mathbb{P}(X_n \neq 0) \leq \mu^n$ . Déduisez-en que  $q = 1$  si  $\mu < 1$ .

Le cas  $\mu < 1$  est appelé le cas *sous-critique* ; on sait déjà que  $q = 1$  si  $\mu < 1$ . Dans le cas  $\mu > 1$ , la théorie des martingales nous donne déjà quelques informations. En effet, le point 1 nous dit que le processus

$$M_n = X_n / \mu^n$$

est une martingale.

### Exercice 71

1. Montrez que  $M_n$  converge presque-sûrement vers une variable aléatoire intégrable  $M_\infty$ .
2. Montrez que dans le cas sous-critique, on a  $M_\infty = 0$  presque-sûrement. Est-ce contradictoire avec le fait que  $\mathbb{E}(M_n) = 1$  pour tout  $n$  ?
3. Montrez que  $\mathbb{E}(M_{n+1}^2) = \mathbb{E}(M_n^2) + \frac{\sigma^2}{\mu^{n+2}}$ . Montrez que  $\sup_n \mathbb{E}(M_n^2) < \infty$  si et seulement si  $\mu > 1$ .
4. Montrez alors que  $M_n$  converge aussi vers  $M_\infty$  au sens  $L^2$ , et que  $M_\infty$  a pour espérance 1 et pour variance  $\sigma^2 / (\mu(\mu - 1)) - 1$ .
5. Pourquoi ne simule-t-on pas un échantillon de  $M_n$  pour  $n$  grand, afin de vérifier la convergence ci-dessus ?

On a donc montré que dans le cas  $\mu > 1$  on avait  $X_n \sim \mu^n Z_\infty$  mais on ne sait pas si  $Z_\infty$  peut s'annuler, et on n'a pas d'information dans le cas  $\mu = 1$ . Pour régler ces problèmes, on va utiliser la fonction génératrice  $G_n$  de  $X_n$ . Rappelons ce qu'est la fonction génératrice :

**Définition 6.2.2** Soit  $A$  une variable aléatoire à valeurs dans  $\mathbb{N}$ . Sa fonction génératrice est la fonction notée  $G_A$  suivante :

$$\begin{aligned} G_A : [0, 1] &\rightarrow [0, 1] \\ s &\mapsto \mathbb{E}(s^A) = \sum_k \mathbb{P}(A = k) s^k. \end{aligned}$$

On rappelle quelques propriétés des fonctions génératrices :

**Lemme 6.2.3** La fonction génératrice  $G_A$  d'une variable aléatoire à valeurs dans  $\mathbb{N}$  est une fonction convexe. Elle est strictement convexe à moins que  $\mathbb{P}(A \leq 1) = 1$ . Elle vérifie les relations

$$G_A(1) = 1, \quad G'_A(1) = \mathbb{E}(A), \quad G''_A(1) = \mathbb{E}(A^2) - \mathbb{E}(A), \quad G_A^{(k)}(0) = k! \mathbb{P}(A = k).$$

On suppose à partir de maintenant que  $\mathbb{P}(Z \leq 1) \neq 1$ .

**Exercice 72** Calculez  $G$  dans les deux cas mentionnés dans l'exercice 69 (variable à valeurs dans  $\{0, 1, 2\}$  et loi géométrique dans  $\mathbb{N}$ ). Tracez dans une même fenêtre la courbe de  $G$  dans trois situations correspondant à  $\mu < 1$ ,  $\mu = 1$  et  $\mu > 1$ , et la courbe de la fonction  $s \mapsto s$ .

On note à partir de maintenant

$$G_n = G_{X_n} \quad \text{et} \quad G = G_Z.$$

On a immédiatement, par le Lemme 6.2.3, que  $q_n = G_n(0)$ . Le résultat central concernant  $q$  est le suivant :

**Lemme 6.2.4** Si  $\mu \leq 1$ , alors  $q = 1$ . Si  $\mu > 1$  alors  $q$  est l'unique solution dans  $[0, 1[$  de  $q = G(q)$ .

Remarquez que 1 est toujours solution de  $s = G(s)$ . D'après [3], Watson en avait conclu de manière incorrecte que  $q = 1$  dans tous les cas.

**Preuve.** La première étape est prouvée dans l'exercice suivant :

**Exercice 73** Montrez que  $\mathbb{E}(s^{X_{n+1}} | X_n) = G(s)^{X_n}$ . Dédisez-en que  $G_n = G^{\circ n}$ , où  $G^{\circ n} = G \circ \dots \circ G$  ( $n$  fois), puis que  $q_{n+1} = G(q_n)$ .

On peut alors utiliser un résultat d'analyse simple :

**Lemme 6.2.5** Soit  $f$  une fonction convexe, dérivable, de  $[0, 1]$  dans  $[0, 1]$  avec  $f(1) = 1$ . Si  $f$  n'est pas la fonction identité, alors pour tout  $s \in [0, 1[$ , la suite  $(f^{\circ n}(s))_n$  :

- converge vers 1 si  $f'(1) \leq 1$
- converge vers l'unique solution de  $s = f(s)$  sur  $[0, 1[$  si  $f'(1) > 1$ .

Puisque  $G'(1) = \mu$  par le Lemme 6.2.3, la preuve est terminée.  $\square$

Ce lemme justifie la nomenclature suivante :

- si  $\mu = 1$ , on dit qu'on est dans le cas critique,
- si  $\mu < 1$ , on dit qu'on est dans le cas sous-critique,
- si  $\mu > 1$ , on dit qu'on est dans le cas sur-critique.

**Exercice 74** Reprenez vos simulations de l'exercice 69. Pour  $Z$  de loi  $\mathcal{G}_{\mathbb{N}}(p)$ , dans trois situations illustrant les trois possibilités sous-critique, critique et sur-critique, simulez des 500-échantillons de  $X_{20}$  et calculez la probabilité empirique d'avoir  $X_{20} = 0$ . Quel intervalle de confiance de niveau 95% peut-on obtenir ? Comparez aux valeurs de  $q$  données par le Lemme 6.2.4. Pour déterminer les solutions de  $G(q) = q$ , on pourra soit procéder de manière théorique, soit utiliser la commande `scipy.optimize.fsolve`.

Une question naturelle est de savoir, dans les cas critique et sur-critique, quel est le comportement de la population *conditionnellement à la non-extinction*. Pour répondre à cette question dans le cas critique, on commence par donner une asymptotique sur la probabilité de non-extinction :

**Proposition 6.2.6** Dans le cas critique, on a

$$\lim_{n \rightarrow \infty} n \mathbb{P}(X_n \neq 0) = \frac{2}{\sigma^2}.$$

**Preuve.** On donne une preuve sommaire. D'abord, on a pour  $s \geq 0$  au voisinage de zéro :

$$G(1-s) = 1-s + \frac{\sigma^2}{2} s^2 + o(s^2).$$

Par conséquent, pour  $s > 0$ ,

$$\frac{1}{1-G(1-s)} - \frac{1}{s} = \frac{\sigma^2}{2} + r(s)$$

où  $r$  est une fonction vérifiant  $\lim_{s \rightarrow 0^+} r(s) = 0$ . On a donc

$$\begin{aligned} \frac{1}{n} \left( \frac{1}{1-G_n(1-s)} - \frac{1}{s} \right) &= \frac{1}{n} \sum_{k=0}^{n-1} \left( \frac{1}{1-G(G_k(1-s))} - \frac{1}{1-G_k(1-s)} \right) \\ &= \frac{\sigma^2}{2} + \frac{1}{n} \sum_{k=0}^{n-1} r(1-G_k(1-s)) \end{aligned}$$

ce qui montre que pour tout  $s \in ]0, 1]$  ceci tend vers  $\frac{\sigma^2}{2}$ . On utilise maintenant le fait que

$$n \mathbb{P}(X_n \neq 0) = n(1-G_n(0)) = \left( \frac{1}{n} \left( \frac{1}{1-G_n(0)} - 1 \right) + \frac{1}{n} \right)^{-1}.$$

Ceci conclut la preuve.  $\square$

**Exercice 75** (facultatif) détaillez la preuve ci-dessus.

On peut maintenant vérifier empiriquement ce résultat :

**Exercice 76** Reprenez les simulations de l'exercice 69. Pour des choix de paramètres correspondant au cas critique, simulez des 500-échantillons et estimez empiriquement la probabilité d'avoir  $X_n \neq 0$  pour  $n = 10, 20, 30, 40, 50$ . Comparez-la à  $\frac{2}{\sigma^2 n}$ . La comparaison (en termes d'ordre de grandeur des erreurs d'estimation) est-elle pertinente ?

Une conséquence simple du résultat ci-dessus est le suivant :

**Corollaire 6.2.7** Dans le cas critique, on a

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}(X_n | X_n \neq 0) = \frac{\sigma^2}{2}.$$

La preuve est laissée en exercice :

**Exercice 77** Prouvez le lemme ci-dessus. On reviendra à la définition :

$$\mathbb{E}(X_n | X_n \neq 0) = \sum_k k \mathbb{P}(X_n = k | X_n > 0).$$

**Exercice 78** Reprenez les simulations de l'exercice 69. Pour des choix de paramètres correspondant au cas critique, simulez des 100-échantillons de  $X_n$  conditionné à vérifier  $X_n > 0$ , pour  $n = 10, 20, 30, 40, 50$ . On utilisera pour cela une méthode du rejet. A partir de ces simulations, estimez empiriquement l'espérance  $\mathbb{E}(X_n | X_n \neq 0)$  pour  $n = 10, 20, 30, 40, 50$ , puis comparez-la à  $\frac{\sigma^2 n}{2}$ .

On passe maintenant au cas sur-critique. Les informations obtenues sur le comportement de  $Z_n = X_n/\mu^n$  montrent immédiatement le résultat suivant :

**Proposition 6.2.8** Dans le cas sur-critique, conditionnellement à la non-extinction, on a explosion : plus précisément,

$$\mathbb{P}(X_n \rightarrow \infty | X_n \neq 0) = 1.$$

### 6.3 Le bandit à deux bras

### 6.4 Processus autorégressifs



## Bibliographie

- [1] M. Benaïm, N. El Karoui, *Promenade aléatoire, chaînes de Markov et simulations; martingales et stratégies.*
- [2] B. Bercu, D. Chafaï, *Modélisation stochastique et simulation.*
- [3] J.-F. Delmas, B. Jourdain, *Modèles aléatoires, applications aux sciences de l'ingénieur et du vivant.*





## Chapitre 7

### Énoncés

On trouvera dans les pages suivantes les énoncés, suivis des corrigés correspondants, pour l'année 2018 :

- un “devoir maison”,
- le partiel,
- l'examen final,
- l'examen de deuxième session (l'unique candidat m'ayant posé un lapin, il n'y a pas de corrigé pour cet énoncé).

# MAO Probabilités-Statistiques

## DM 1 : le battage de cartes

On va essayer de répondre empiriquement à la question “combien de fois faut-il battre un jeu de carte pour qu’il soit bien mélangé” si la technique de battage des cartes est “on coupe et on bat”. On considère un jeu (standard) de 52 cartes qui est initialement dans un ordre donné (par exemple les pique de 1 à 10 puis roi, dame, valet, puis les trèfle, etc.) et on représente ses états possibles par les permutations  $\sigma$  de  $\{1, \dots, 52\}$ . On définit l’indice d’une permutation comme le nombre de sous-suites maximales (pour l’inclusion) avec la propriété d’être croissantes et faites de nombres consécutifs. Par exemple, la permutation  $(7, 3, 1, 8, 4, 2, 5, 6)$  a une décomposition  $(7, 8), (3, 4, 5, 6), (1, 2)$  donc pour indice 3.

1. Soit  $\sigma$  une permutation de  $\{1, \dots, n\}$ . Quelles valeurs peut prendre son indice ? montrer que celui-ci vaut

$$1 + \text{card}\{i \in \{1, \dots, n-1\} \mid \sigma^{-1}(i) > \sigma^{-1}(i+1)\}.$$

2. On représente dans Python la permutation comme une liste d’entiers, par exemple `sigma=[7, 3, 1, 8, 4, 2, 5, 6]`. Écrivez une fonction qui calcule l’indice d’une permutation `sigma` en utilisant `numpy.sort` et `numpy.diff`. Vous ne manquez pas de tester votre fonction sur des exemples simples.
3. Simulez un nombre  $N = 10000$  de permutations aléatoires de  $\{1, \dots, 52\}$  et calculez leurs indices (il existe une commande `numpy.random` dédiée). Comptez ensuite, pour chacune des valeurs possibles, la fréquence empirique de cette valeur. Ceci nous donne une idée des valeurs “typiques” d’une permutation aléatoire.

Étudions maintenant l’effet d’un battage : on le modélise en supposant que :

- on coupe le paquet en deux tas, mettant dans le premier tas les  $N$  premières cartes où  $N$  suit une loi  $\mathcal{B}(52, 1/2)$
- on mélange les deux paquets en choisissant à chaque instant la carte du dessus d’un tas ou de l’autre, aléatoirement avec des probabilités proportionnelles au nombre de cartes restant dans le paquet.

Cette méthode s’appelle en français le “mélange américain” et en anglais le “riffle shuffle” ou “dovetail shuffle”.

Comme la question du mélange est cruciale, je préfère détailler en prenant l’exemple hypothétique où on n’aurait que cinq cartes  $\{C_1, C_2, C_3, C_4, C_5\}$ .

- le “tirage” de  $N$  donne  $N = 2$ . Les deux paquets sont donc  $\{C_1, C_2\}$  et  $\{C_3, C_4, C_5\}$ .

- on choisit une première carte : elle viendra du premier paquet avec probabilité  $2/5$  (auquel cas ce sera  $C_2$ ) et du deuxième paquet avec probabilité  $3/5$  (auquel cas ce sera  $C_5$ ). Si c'est le deuxième cas qui se produit, alors  $\{C_5\}$  devient le tas "mélangé", et les deux paquets restants sont  $\{C_1, C_2\}$  et  $\{C_3, C_4\}$ . On choisit alors la deuxième carte, qui viendra du premier paquet avec probabilité  $2/4$  (auquel cas ce sera  $C_2$ ) et du deuxième paquet avec probabilité  $2/4$  (auquel cas ce sera  $C_4$ ). Si c'est le premier cas qui se produit, le tas mélangé devient  $\{C_2, C_5\}$  et les deux paquets restants sont  $\{C_1\}$  et  $\{C_3, C_4\}$ . On continue ainsi jusqu'à épuisement des deux paquets.
4. Écrivez une fonction Python qui prend en entrée une permutation `sigma` et rend la permutation correspondant au résultat d'un battage de `sigma`. Remarquez que le battage est déterminé par la donnée de  $N \sim \mathcal{B}(52, 1/2)$  puis d'un tirage sans remise de 52 boules parmi 52 dans lesquelles  $N$  sont de type "1" et  $52 - N$  de type "2". On pourra utiliser la commande `numpy.random.choice`.
  5. Supposons que l'on parte d'un paquet neuf, et qu'on le batte de manière répétée. On note  $\sigma_n$  la permutation après  $n$  battages, et  $i_n$  son indice; ces deux quantités sont des variables aléatoires. En utilisant votre fonction précédente, effectuez une simulation de  $\sigma_n$  pour  $n = 0, \dots, 12$  et tracez la courbe donnant les  $i_n$  correspondants.
  6. Simulez un nombre  $N = 10000$  de trajectoires et calculez la loi empirique de  $i_n$  pour ces différentes valeurs.
  7. On rappelle que l'écart en variation totale de deux lois à support fini  $X$  est

$$\|\mu - \nu\|_{\text{VT}} = \frac{1}{2} \sum_{x \in X} |\mu(x) - \nu(x)|.$$

En admettant que les simulations ci-dessus ont donné une bonne estimation de la loi  $\mu$  de l'indice d'une permutation aléatoire et de la loi  $\nu_n$  de  $i_n$ , tracez la courbe  $n \mapsto \|\mu - \nu_n\|_{\text{VT}}$ .

8. Combien de fois, en fin de compte, pensez-vous qu'il faut battre un jeu de cartes ?

Remarquons que l'on a étudié ici la loi de l'indice  $i_n$  plutôt que celle de la permutation  $\sigma_n$  car la première est plus simple à simuler (son support est de cardinal 52 au lieu d'être de cardinal  $52! \simeq 8 \times 10^{67}$ ) mais on observe le même phénomène de cutoff (décroissance brutale de la variation totale) sur les lois des permutations (voir à ce sujet l'article [1], dont le contenu dépasse largement le cadre de ce cours).

## Références

- [1] D. Bayer, P. Diaconis. *Trailing the dovetail shuffle to its lair*, Annals of Applied Probability 2.

## MAO Probabilités-Statistiques

### Corrigé du DM 1 : le battage de cartes

1. Il y a au moins une suite, donc l'indice vaut au moins 1, et la permutation  $(1, \dots, n)$  a justement pour indice 1. Il y a au plus  $n$  suites, donc l'indice vaut au plus  $n$ , et la permutation  $(n, \dots, 1)$  a justement pour indice  $n$ . Il n'est pas difficile de fabriquer un exemple de suite ayant pour indice  $k \in \{2, \dots, n-1\}$  donc l'indice peut prendre les valeurs  $\{1, \dots, n\}$ .

Pour prouver la formule générale, on peut remarquer qu'une suite telle que définie par l'énoncé est exactement une suite maximale  $(i, i+1, \dots, i+\ell)$  telle que  $\sigma^{-1}(i) < \sigma^{-1}(i+1) < \dots < \sigma^{-1}(i+\ell)$  (cette condition signifie que  $i, i+1, \dots, i+\ell$  apparaissent dans cet ordre. Par conséquent, on peut définir la famille de sous-suites de la manière suivante : on prend celle qui contient 1, elle s'arrête au  $i$  minimal tel que  $\sigma^{-1}(i) > \sigma^{-1}(i+1)$ . Si  $i \leq n-1$  alors on commence une nouvelle sous-suite, etc. On voit qu'on épuise  $\{1, \dots, n\}$  en ayant une première sous-suite, plus une supplémentaire à chaque fois qu'on a un  $i \in \{1, \dots, n-1\}$  tel que  $\sigma^{-1}(i) > \sigma^{-1}(i+1)$ . Cela prouve la formule.

Si on trouve l'argument ci-dessus fumeux, on peut prouver la relation par récurrence : pour  $n = 1$  elle est triviale ; si maintenant on considère une permutation de  $n+1$  on considère d'abord la liste des images  $(\sigma(1), \dots, \sigma(n+1))$  dont on supprime  $n+1$ . Le nombre de sous-suites maximales dans cette liste est donné par la formule  $1 + \text{card}\{i \in \{1, \dots, n-1\} \mid \sigma^{-1}(i) > \sigma^{-1}(i+1)\}$ . On note temporairement  $c_n$  le cardinal défini ci-dessus. Quand on ajoute le terme  $n+1$  : si  $\sigma^{-1}(n) < \sigma^{-1}(n+1)$  alors on peut ajouter  $n+1$  à la suite qui contient  $n$ , donc le nombre de suites ne change pas, mais de même on a  $c_{n+1} = c_n$ , donc la formule reste vraie. De même, si  $\sigma^{-1}(n) > \sigma^{-1}(n+1)$  alors il faut créer une nouvelle sous-suite qui est composée de  $n+1$  seul, donc le nombre de suites augmente de 1 mais  $c_{n+1} = c_n + 1$ . On a donc la formule générale.

2. Voir la fonction `indice`. On ne manque pas de le tester sur la permutation de l'énoncé, ainsi que sur des exemples simples comme  $(1, \dots, 10)$ ,  $(10, \dots, 1)$ , et d'autres pour lesquels on peut calculer l'indice à la main. Remarquez que toute la suite dépend de la fonction qui calcule l'indice, il est donc rentable de passer trente secondes à vérifier qu'elle fonctionne.
3. Voir le code. On observe le résultat donné ici en Figure 1. Il semble que les valeurs typiques soient centrées autour de  $27/28$ , et dispersées de quelques unités.
4. Supposons qu'on a tiré le  $N$ . On peut voir facilement que (conditionnellement

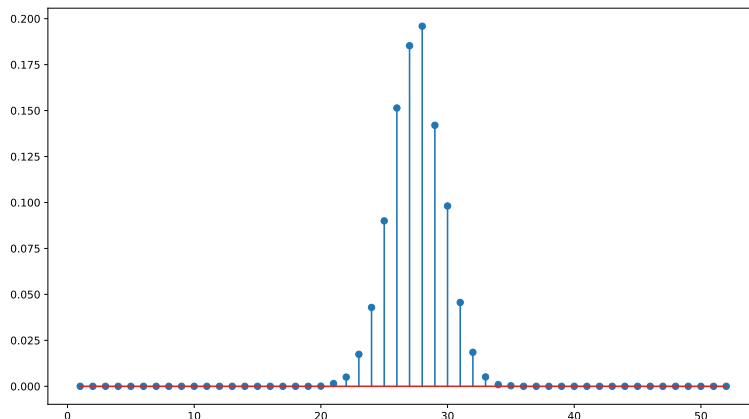


FIGURE 1 – loi empirique de l’indice d’une permutation uniforme aléatoire

au  $N$ ), toute séquence de  $N$  valeurs “1” et de  $n - N$  valeurs “2” (indiquant de quel paquet vient quelle carte) a la même probabilité. Par conséquent, il suffit de se donner les positions des “1” en ordonnant un tirage sans remise de  $N$  valeurs dans  $\{1, \dots, 52\}$ . Voir la fonction `battage`.

5. Voir la Figure 2. On observe qu’après environ huit battages, l’indice varie d’une manière qui semble aléatoire, et que l’amplitude des fluctuations est de l’ordre des fluctuations de l’indice d’une permutation uniforme aléatoire (ceci d’après la question 3). On peut donc penser qu’après environ huit battages, les permutations obtenues en continuant à battre sont à peu près uniformément distribuées, et peut-être même (approximativement) indépendantes les unes des autres.
6. Voir le code. On obtient la Figure 3.
7. Après un plateau (bien compréhensible : l’indice vaut typiquement 2 après un battage, 4 après deux battages...), on observe une décroissance rapide de la variation totale entre 6 et 8 battages. Il semble qu’il faille battre les cartes huit fois au moins.

Un commentaire au sujet de ce résultat, extrait de “Shuffling cards and stopping times” de David Aldous et Persi Diaconis (ce dernier, avant d’être probabiliste, a été magicien professionnel et sait donc de quoi il parle) :

*Magicians and card cheats have long taken advantage of such patterns. Suppose a deck of 52 cards in known order is shuffled 3 times and cut arbitrarily in between these shuffles. Then a card is taken out, noted and replaced in a different position. The noted card can be determined with near certainty! Gardner (1977) describes card tricks based on the inefficiency of too few riffle shuffles.*

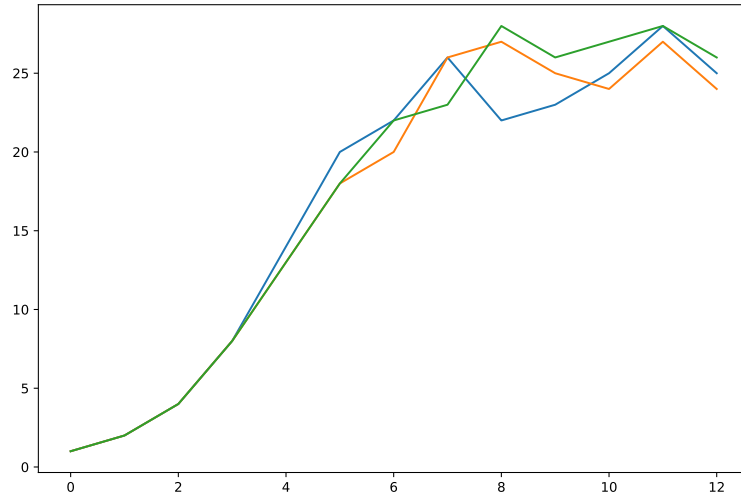


FIGURE 2 – trois simulations de l'évolution de l'indice en fonction du nombre de bat-tages, en partant d'un paquet ordonné

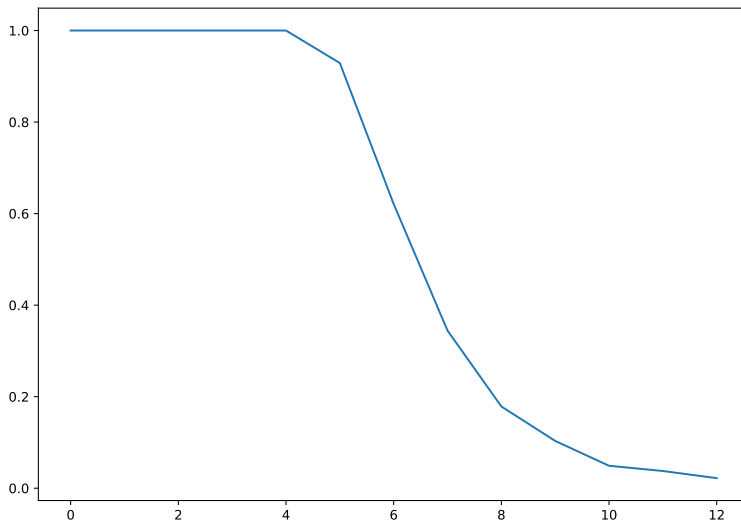


FIGURE 3 – évolution de la variation totale entre (les estimations de) la loi de  $i_n$  et la loi de l'indice d'une permutation uniforme aléatoire

# MAO Probabilités-Statistiques

## Partiel

Ce sujet est (très) largement inspiré d'un texte d'agrégation que vous pourrez trouver à l'adresse <http://agreg.org/Textes/public2015-A9.pdf> (mais cela ne vous servira à rien d'y aller le voir dans l'immédiat).

Je vous demande de faire un effort de lisibilité et de penser votre copie comme un rapport, plus que comme une suite de réponses. Cela signifie qu'il faut, au minimum, expliquer à tout instant ce que vous faites, même si l'explication était déjà présente dans la question : votre copie doit être lisible seule. En particulier, je ne veux surtout pas qu'elle ressemble à

1.  $\alpha = 0,48$ .
2.  $\hat{\alpha}_n = \frac{1}{2} \bar{X}_n^2$ .

mais plutôt à

1. On s'intéresse à la probabilité théorique de la fin du monde. On trouve en utilisant la méthode de Philippulus que  $\alpha = 0,48$ .
2. On propose d'utiliser l'estimateur  $\hat{\alpha}_n = \frac{1}{2} \bar{X}_n^2$  qui, d'après le théorème de Calys, est asymptotiquement normal.

Par ailleurs, je ne veux pas avoir à deviner quelle fonction Python sert à quoi, ni devoir faire tourner moi-même vos simulations pour les exploiter. C'est à vous des les commenter pour expliquer ce que, d'après vous, elles illustrent, suggèrent, infirment, ou autre.

Dans la liste ci-dessous, les questions mentionnées comme facultatives sont à traiter "à la carte". Elles sont indépendantes les unes des autres et des questions non facultatives.

On étudie un modèle de constitution des partis politiques dans une population (supposée infinie). Dans ce modèle, les citoyens entrent en politique les uns après les autres, de telle manière qu'à l'instant  $n \in \mathbb{N}^*$ , exactement  $n$  citoyens sont membres d'un parti ; on va noter  $B_n$  le paysage politique au temps  $n$ , qui sera donné comme une partition de  $\{1, \dots, n\}$ . Au temps  $n = 1$ , il existe un seul parti  $b_1$ , dont l'unique membre est le citoyen 1, et le paysage politique est donc

$$B_1 = \{b_1\} \quad \text{avec} \quad b_1 = \{1\}.$$

Au temps  $n + 1$ , le paysage politique  $B_{n+1}$  est obtenu à partir de  $B_n$  en fonction du comportement, considéré comme aléatoire, du  $n + 1$ -ième citoyen :

- avec probabilité  $\frac{\theta}{\theta+n}$ , ce citoyen crée un nouveau parti : dans ce cas  $B_{n+1} = B_n \cup \{\{n+1\}\}$ ,
- s'il ne crée pas un nouveau parti, alors il adhère à un parti existant qu'il choisit suivant une probabilité proportionnelle à la taille du parti.

Si par exemple au temps  $n = 9$  on a  $B_9 = \{b_1, b_2, b_3, b_4\}$  avec

$$b_1 = \{1\}, \quad b_2 = \{2, 5, 6, 9\}, \quad b_3 = \{3, 4, 8\}, \quad b_4 = \{7\}$$

alors avec probabilité  $\frac{\theta}{\theta+9}$  on a  $B_{10} = \{b_1, b_2, b_3, b_4, b_5\}$  avec  $b_5 = 10$ ; et conditionnellement au fait que ce ne soit pas le cas, avec probabilité  $1/9$  on a  $B_{10} = \{b'_1, b_2, b_3, b_4\}$  où  $b'_1 = \{1, 10\}$ , avec probabilité  $4/9$  on a  $B_{10} = \{b_1, b'_2, b_3, b_4\}$  où  $b'_2 = \{2, 5, 6, 9, 10\}$ , etc.

Dans la majeure partie de ce qui suit, on ne s'intéressera pas à la composition détaillée de  $B_n$ , mais seulement à la liste  $E_n$  des effectifs de chacun de ces partis. Autrement dit, plutôt que de dire que  $B_9 = \{\{1\}, \{2, 5, 6, 9\}, \{3, 4, 8\}, \{7\}\}$  on se contentera de dire que  $E_9 = (1, 4, 3, 1)$ , autrement dit le premier parti a 1 membre, le deuxième 4, ... (attention,  $E_n$  est bien une famille ordonnée).

1. Décrivez l'évolution de  $E_n$ . Dans quel ensemble  $E_n$  prend-il ses valeurs ?
2. Écrivez une fonction Python qui prend en entrée  $\theta$  et une famille  $E$  d'effectifs et rend un  $E'$  qui est le résultat de l'évolution donnée ci-dessus. Autrement dit, si  $E_n = E$  alors la fonction retourne  $E_{n+1}$  ( $n$  devra être déterminé à partir de  $E$ ).
3. Quel est le comportement qualitatif de  $E_n$  pour  $\theta$  très petit ? pour  $\theta$  très grand ?
4. On note  $|B_n|$  le cardinal de  $B_n$ , qui est le nombre de partis politiques (dans l'exemple donné en introduction,  $|B_9| = 4$ ). Montrez que l'on peut écrire  $|B_n| = \xi_1 + \dots + \xi_n$  où les  $\xi_k$  sont des variables indépendantes de lois de Bernoulli indépendantes mais pas forcément identiquement distribuées.
5. Déduisez du point précédent que

$$\mathbb{E}(|B_n|) = \sum_{k=0}^{n-1} \frac{\theta}{\theta+k} \quad \text{var}(|B_n|) = \sum_{k=1}^{n-1} \frac{\theta k}{(\theta+k)^2}. \quad (1)$$

6. Donnez des équivalents de  $\mathbb{E}(|B_n|)$  et  $\text{var}(|B_n|)$  puis montrez que  $\frac{|B_n|}{\log n} \xrightarrow{\mathbb{P}} \theta$ . Quelle méthode utiliseriez vous pour illustrer facilement cette convergence ?



7. *Question facultative* : illustrez cette convergence. Si vous n'êtes pas satisfait, demandez-vous quelle est la vitesse de convergence donnée par la preuve détaillée au point 6.
8. Cette même preuve du point 6 vous permet-elle de démontrer la convergence presque-sûre ? Que semblent indiquer les simulations à propos d'une éventuelle convergence presque-sûre ?
9. On admet que  $\frac{|B_n| - \theta \log n}{\sqrt{\theta \log n}}$  converge en loi vers une normale centrée réduite. Comment feriez-vous pour illustrer cette convergence ? *La suite naturelle de l'exercice serait de vous demander un intervalle de confiance pour  $\theta$ , mais nous n'avons pas encore parlé d'intervalles de confiance.*
10. *Question facultative* : illustrez la convergence en question 9. Vous risquez de ne pas être satisfait.
11. *Question facultative* : on note  $U_n$  le nombre de partis qui ont un unique membre qui est donc leur créateur (dans l'exemple donné en introduction,  $U_9 = 2$ ). Conditionnellement à  $U_n = a$ , quelles sont les valeurs possibles de  $U_{n+1}$ , et avec quelles probabilités ? Déduisez-en que  $\mathbb{E}(U_n) = \frac{n\theta}{n+\theta-1}$ . A-t-on convergence de  $U_n$  vers  $\theta$  presque-sûrement ? en loi ? pour un autre mode de convergence ?
12. *Question facultative* : dans cette question on s'intéresse à la distribution des tailles de partis. On note  $C_{n,k}$  le nombre de partis de taille  $k$  (dans l'exemple donné en introduction,  $C_{9,1} = 2, C_{9,2} = 0, C_{9,3} = 1, C_{9,4} = 1$ ). On propose la loi suivante pour  $(C_{n,1}, \dots, C_{n,n})$  : pour tout  $n$ -uplet d'entiers naturels  $a_1, \dots, a_n$  vérifiant  $\sum_{k=1}^n k a_k = n$ , on a

$$\mathbb{P}(C_{n,1} = a_1, \dots, C_{n,n} = a_n) = \frac{n!}{\theta(\theta+1) \dots (\theta+n-1)} \prod_{k=1}^n \frac{1}{a_k!} \left(\frac{\theta}{k}\right)^{a_k}.$$

On pourra prouver cette formule<sup>1</sup> (par récurrence) et/ou la vérifier empiriquement dans le cas  $n = 5$  (auquel cas il n'y a que six possibilités pour la famille  $(C_{n,k})_{k=1}^n$ ), en simulant un  $N$ -échantillon de  $B_n$  avec  $N$  choisi suffisamment grand.

13. *Question facultative et ouverte* : on peut se poser de nombreuses autres questions, soit sur ce modèle, par exemple :
- pour un coefficient  $\alpha \in ]0, 1[$ , comment croît le nombre de partis ayant au moins  $\alpha n$  adhérents ?
  - à quelle vitesse croît la taille du plus grand parti ?
- soit sur des modèles légèrement différents (qui vous obligeront donc à modifier votre code) :
- on peut supposer par exemple qu'un parti qui n'a pas enregistré de nouvelle adhésion depuis un certain temps (qui peut être fixe ou proportionnel à sa taille) finit par se dissoudre, ou par se fondre dans un autre,

1. j'avoue que c'est un peu pénible.

- on peut supposer qu'un citoyen peut changer de parti (à des moments déterministes ou aléatoires),
- on peut supposer que si un citoyen ne fonde pas de nouveau parti, il choisit celui auquel il adhère avec une probabilité non pas proportionnelle à son effectif mais dépendant de sa composition (dans l'exemple donné en introduction, le dixième citoyen pourrait choisir le parti auquel il adhère avec des probas proportionnelles à  $1, 2 + 5 + 6 + 9, 3 + 4 + 8, 7$ , ou encore proportionnelles à  $1, 9, 8, 7$ ).

Utilisez des simulations pour conjecturer les réponses aux nouvelles questions sur l'ancien modèle, ou aux anciennes questions sur les nouveaux modèles.

## MAO Probabilités-Statistiques

### Correction du partiel

1. On a  $E_1 = \{1\}$  et  $E_{n+1}$  dépend de  $E_n$  de la manière suivante : si  $E_n = (e_1, \dots, e_\ell)$ , alors avec probabilité  $\frac{\theta}{\theta+n}$  on a  $E_{n+1} = (e_1, \dots, e_\ell, 1)$  et conditionnellement à ce que ne soit pas le cas on a  $E_{n+1} = (e_1, \dots, e_{k-1}, e_k + 1, e_{k+1}, \dots, e_\ell)$  avec probabilité  $\frac{e_k}{\theta+n}$ . On peut remarquer que de manière équivalente, on a conditionnellement à  $E_n = (e_1, \dots, e_\ell)$

$$E_n = (e_1, \dots, e_\ell, 1) \text{ avec probabilité } \frac{\theta}{\theta+n}$$

$$E_{n+1} = (e_1, \dots, e_{k-1}, e_k + 1, e_{k+1}, \dots, e_\ell) \text{ avec probabilité } \frac{e_k}{\theta+n}.$$

On voit que  $E_n$  est à valeurs dans l'ensemble des familles  $(e_1, \dots, e_\ell)$  d'entiers naturels non nuls tels que  $\sum_k e_k = n$ . Si l'on comprend la question de manière plus stricte et que l'on veut l'ensemble *exact* des valeurs possibles, la réponse est la même mais il faut alors justifier que toute famille de ce type a une probabilité non nulle d'être observée. C'est le cas car par exemple on a une probabilité non nulle que  $E_k = (1, \dots, 1)$  et conditionnellement à cela une probabilité non nulle que le premier parti croisse jusqu'à la valeur  $e_1$ , etc.

2. Voir la fonction `evolution` dans le fichier joint.
3. En faisant tourner la fonction `evolution` avec  $\theta = 0,1$  on voit qu'il y a peu de créations de nouveaux partis et qu'en général c'est le premier parti qui récolte les nouveaux adhérents. Ceci est facile à expliquer : comme les créations de parti sont rares, le premier parti a en général eu le temps de beaucoup grossir avant d'être confronté à la concurrence, mais avec notre modèle un gros parti a plus de chances d'attirer les adhérents. À l'inverse pour  $\theta = 100$  on a création d'un nouveau parti presque à chaque tour donc les partis ont presque tous un unique adhérent et cela restera vrai tant que  $n$  est petit devant  $\theta$ .

En anticipant un peu sur les notations, on trace quelques courbes d'évolution de  $|B_n|$  pour différentes valeurs de  $\theta$  (voir code). On obtient la figure 1. On voit que le nombre de partis augmente beaucoup plus vite quand  $\theta$  est grand que quand il est petit.

4. À chaque tour,  $|B_n|$  peut croître de 1 (avec probabilité  $\frac{\theta}{\theta+n-1}$ ) ou rester constant, et les accroissements aux différents tours sont indépendants les uns des autres. On a donc  $|B_n| = \xi_1 + \dots + \xi_n$  où les  $\xi_k$  sont des variables indépendantes de lois  $\xi_k \sim \mathcal{B}(\frac{\theta}{\theta+k-1})$ .

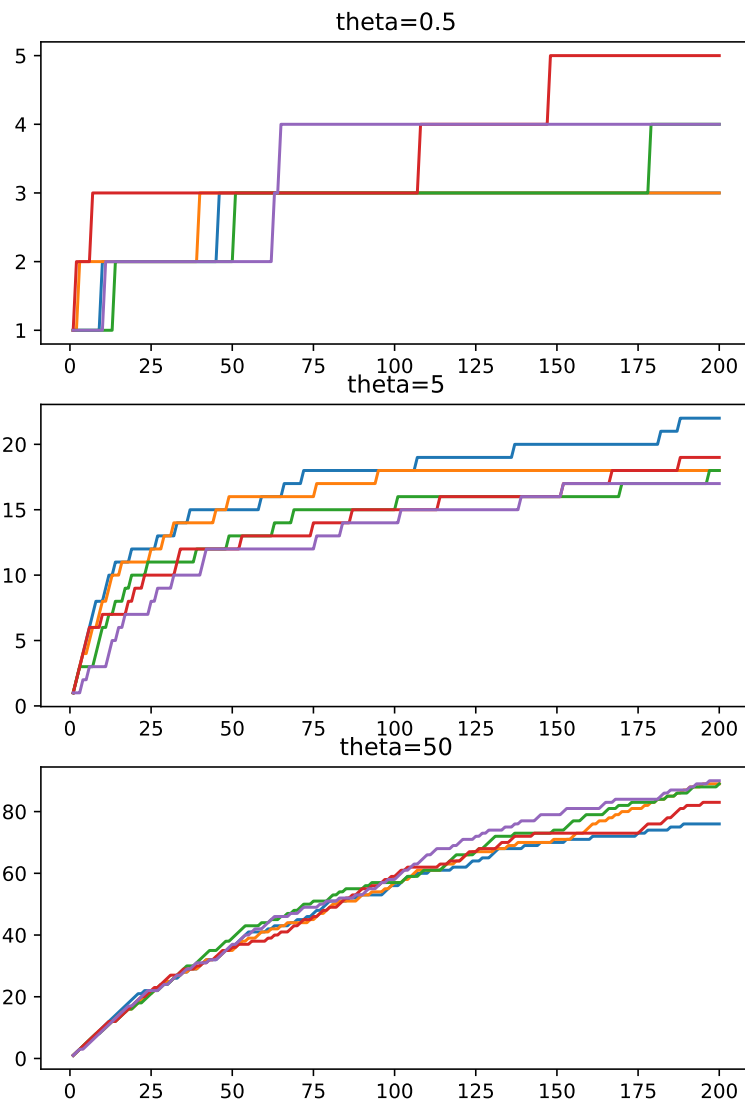


FIGURE 1 – Tracé de  $|B_n|$ , question 3

5. On a immédiatement par le point précédent que

$$\mathbb{E}(|B_n|) = \sum_{k=1}^n \mathbb{E}(\xi_k) \quad \text{var}(|B_n|) = \sum_{k=1}^n \text{var}(\xi_k) \quad (1)$$

(notez que la première identité utilise uniquement la linéarité de l'espérance mais que la deuxième utilise l'indépendance entre les  $\xi_k$ ). Comme  $\mathbb{E}(\xi_k) = \frac{\theta}{\theta+k-1}$  et  $\text{var}(\xi_k) = \frac{\theta(k-1)}{(\theta+k-1)^2}$ , on obtient immédiatement

$$\mathbb{E}(|B_n|) = \sum_{k=0}^{n-1} \frac{\theta}{\theta+k-1} \quad \text{var}(|B_n|) = \sum_{k=1}^{n-1} \frac{\theta k}{(\theta+k)^2}.$$

6. Les termes  $\frac{\theta}{\theta+k-1}$  et  $\frac{\theta k}{(\theta+k)^2}$  sont tous les deux équivalents en  $k \rightarrow \infty$  à  $\theta/k$ , qui est le terme général d'une série divergente. Par conséquent,  $\mathbb{E}(|B_n|)$  et  $\text{var}(|B_n|)$  sont équivalents à  $\sum_{k=1}^n \theta/k$ , dont on sait qu'il est lui-même équivalent à  $\theta \log n$ .

On a

$$\frac{B_n}{\log n} - \theta = \frac{|B_n| - \mathbb{E}(|B_n|)}{\log n} + \frac{\mathbb{E}(|B_n|) - \theta \log n}{\log n}.$$

Le deuxième terme tend vers zéro de manière déterministe quand  $n \rightarrow \infty$ , donc pour  $n$  assez grand

$$\begin{aligned} \mathbb{P}\left(\left|\frac{|B_n|}{\log n} - \theta\right| > \epsilon\right) &\leq \mathbb{P}\left(\left|\frac{|B_n| - \mathbb{E}(|B_n|)}{\log n}\right| > \frac{\epsilon}{2}\right) \\ &\leq \frac{\text{var}(|B_n|)}{(\epsilon/2 \log n)^2} \end{aligned}$$

(on a utilisé Bienaymé-Tchebychev). Par la discussion ci-dessus, ce majorant est un  $O\left(\frac{1}{\log n}\right)$ , donc on a prouvé la convergence en probabilité de  $\frac{|B_n|}{\log n}$  vers  $\theta$ . On a vu qu'illustrer une convergence en probabilité était pénible mais ici on parle de convergence vers une constante, donc la convergence en probabilité équivaut à la convergence en loi. On choisira donc ici d'illustrer la convergence en loi, par la méthode standard de tracé des fonctions de répartition empirique.

7. Voir le code (le faire tourner est long). On voit sur la figure 2... qu'on ne voit pas grand chose ! On voit que le "centre" de la courbe croît lentement vers la vraie valeur  $\theta = 5$ . Cela nous rappelle que la majoration obtenue au point précédent nous a donné une vitesse de convergence en  $1/\log n$ , donc une convergence très lente. Cela pourrait n'être que la conséquence de l'utilisation d'une majoration grossière (Bienaymé-Tchebychev) mais notre simulation semble indiquer que la convergence est en effet lente.

8. La majoration obtenue au point 6 donne un majorant non sommable. Elle ne permet pas d'appliquer la méthode standard (par Borel-Cantelli) de preuve de la convergence presque-sûre. De plus, les simulations (voir figure 3) ne montrent pas a priori de convergence presque-sûre (ou alors ce serait vers une variable aléatoire non constante, mais ce n'est pas possible car on a convergence en probabilité vers une constante). Encore une fois il n'est pas à exclure que la convergence soit simplement trop lente pour être observée.

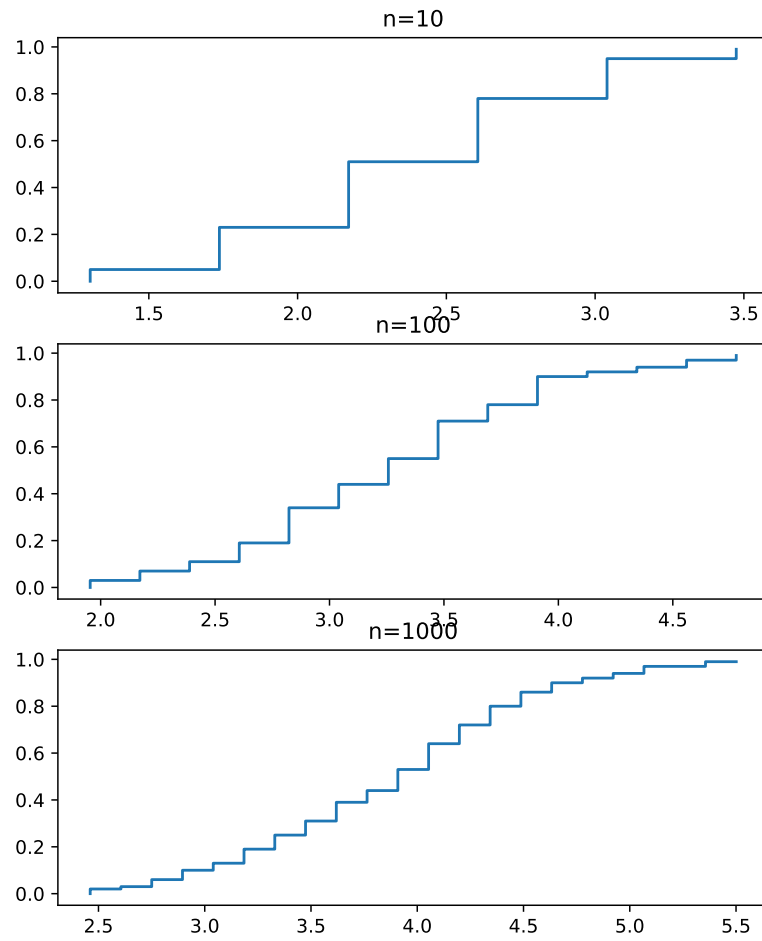
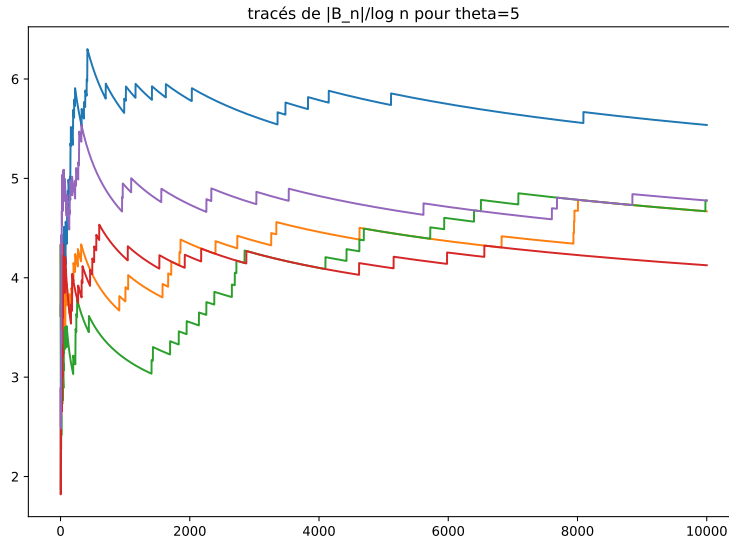


FIGURE 2 – Tracé de  $|B_n|/\log n$ , question 7

FIGURE 3 – Tracé de  $|B_n|/\log n$ , question 8

9. Pour illustrer la convergence en loi de  $\frac{|B_n| - \theta \log n}{\sqrt{\theta \log n}}$  vers une normale centrée réduite, on voudrait appliquer la méthode des fonctions de répartition empirique : pour plusieurs valeurs de  $n$ , on simule un  $N$ -échantillon et on trace la fonction de répartition empirique à partir de l'échantillon (pour  $N$  assez grand – on rappelle que l'erreur uniforme est en  $1/\sqrt{N}$ ).
10. Voir le code associé (il suffit de modifier le code la question 7). Comme on le voit en figure 4, la convergence ne saute pas aux yeux. Encore une fois, c'est que celle-ci est trop lente.
11. Supposons qu'au temps  $n$  il y a  $a$  partis monomembres. Le  $n+1$ -ième citoyen va soit fonder son propre parti, ce qui ferait un parti monomembre de plus et alors  $U_{n+1} = a + 1$ ; soit adhérer à un parti qui avait strictement plus d'un membre, et alors le nombre de partis monomembre ne change pas, i.e.  $U_{n+1} = a$ ; soit adhérer à un parti qui était jusque là monomembre, et alors le nombre de ces partis diminue d'une unité, i.e.  $U_{n+1} = a - 1$ . Si l'on précise les probabilités correspondantes, on a conditionnellement à  $U_n = a$

$$U_{n+1} = \begin{cases} a + 1 & \text{avec probabilité } \frac{\theta}{\theta+n} \\ a & \text{avec probabilité } \frac{n-a}{\theta+n} \\ a - 1 & \text{avec probabilité } \frac{a}{\theta+n} \end{cases}$$

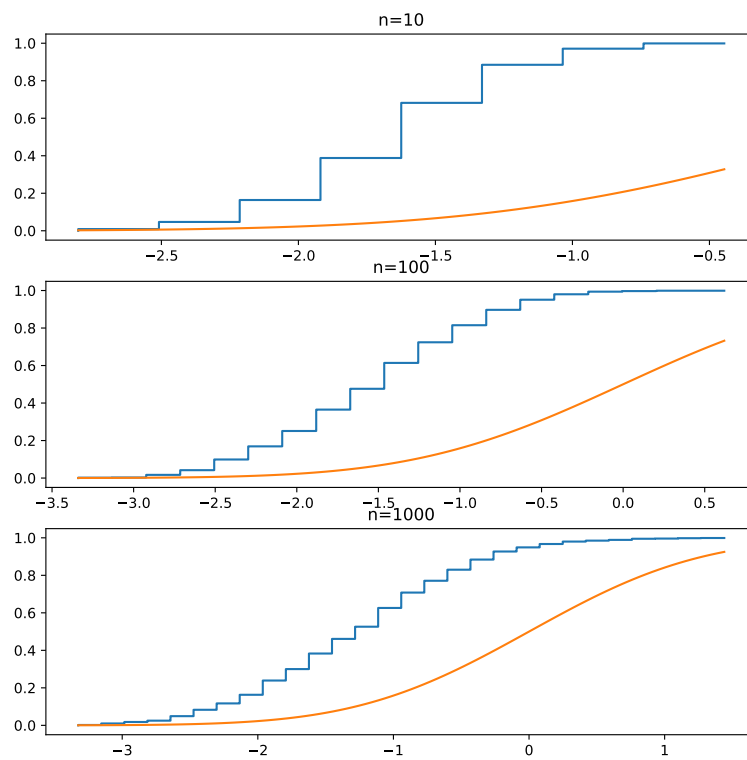


FIGURE 4 – Tracé de  $|B_n|/\log n$ , question 10



Par conséquent,

$$\begin{aligned}\mathbb{E}(U_{n+1}|U_n) &= (U_n + 1) \frac{\theta}{\theta + n} + U_n \frac{n - U_n}{\theta + n} + (U_n - 1) \frac{U_n}{\theta + n} \\ &= U_n \left(1 - \frac{1}{\theta + n}\right) + \frac{\theta}{\theta + n}\end{aligned}$$

et donc

$$\mathbb{E}(U_{n+1}) = \mathbb{E}(U_n) \times \left(1 - \frac{1}{\theta + n}\right) + \frac{\theta}{\theta + n}.$$

On trouve donc  $\mathbb{E}(U_{n+1}) - \theta = (\mathbb{E}(U_n) - \theta) \times \left(1 - \frac{1}{\theta + n}\right)$ , d'où

$$\mathbb{E}(U_n) = \frac{\theta n}{\theta + n - 1}.$$

On commence par tracer des trajectoires de  $U_n$ . On n'a pas l'air d'avoir convergence presque-sûre (cf. figure 5). On trace alors les fonctions de répartition em-

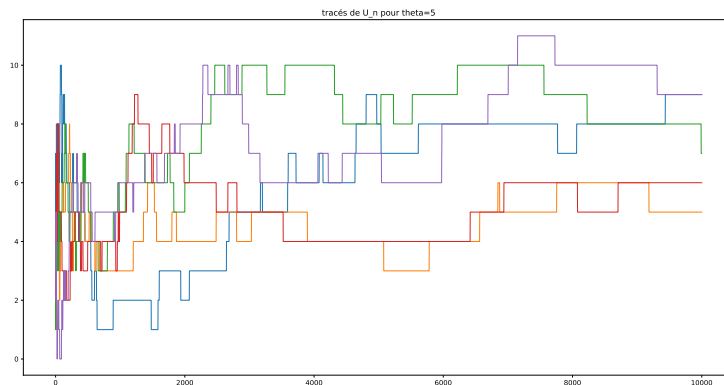


FIGURE 5 – Tracé de  $U_n$ , question 11

piriques pour différentes valeurs de  $n$ . Le résultat est la figure 6. Conclusion : on ne voit pas bien ce qui se passe ! On peut tenter d'observer le comportement de  $\mathbb{P}(U_n = u)$  pour quelques valeurs de  $u$  dans une certaine gamme de valeurs. On se limite pour des raisons de lisibilité à quelques valeurs de  $u$  dans la zone correspondante, par exemple  $\{3,4,5,6\}$ . Le résultat des simulations est visible en figure 7 : il semble aller dans le sens d'une convergence en loi.

12. Donnons une preuve théorique de la formule. On note  $C_n$  le  $n$ -uplet  $(C_{n,1}, \dots, C_{n,n})$ . Pour  $n = 1$  on a  $C_n = (1)$  avec probabilité 1, or on voit immédiatement que la formule proposée donne bien une probabilité 1 à cet événement. On va ensuite travailler par récurrence : supposons que la formule est vraie jusqu'au

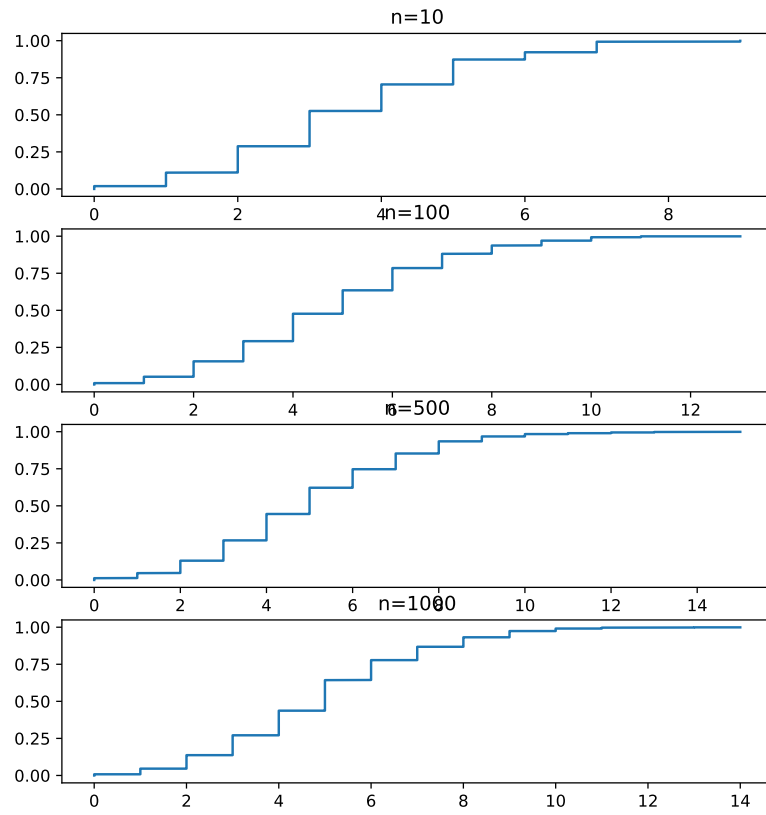
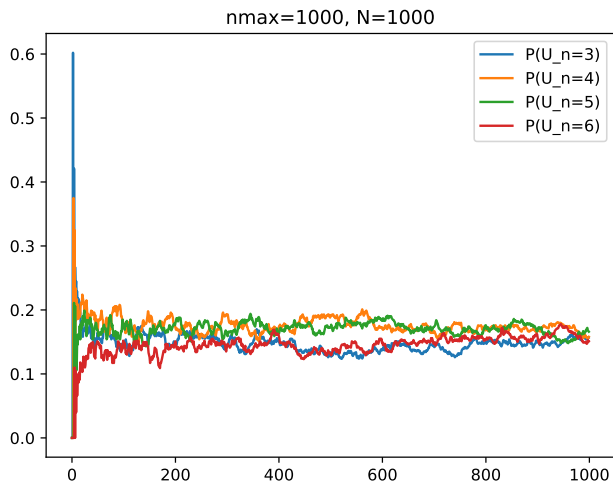


FIGURE 6 – Fonctions de répartition empiriques de  $U_n$ , question 11

FIGURE 7 – Probabilités empiriques  $\mathbb{P}(U_n = u)$ , question 11

rang  $n$ , et considérons une famille  $(a_1, \dots, a_{n+1})$  d'entiers naturels vérifiant  $\sum_{k=1}^{n+1} k a_k = n + 1$ . On va exprimer à partir de quelles valeurs des  $C_n$  on peut obtenir  $C_{n+1} = (a_1, \dots, a_{n+1})$ . Traitons à part le cas  $a_{n+1} = 1$  (et  $a_k = 0$  pour  $k \leq n$ ) : si  $C_{n+1, n+1} = 1$  c'est que  $C_{n, n} = 1$  et que le  $n + 1$ -ième citoyen a adhéré au parti unique déjà existant. Par récurrence cet événement a probabilité  $\frac{n!}{\theta(\theta+1)\dots(\theta+n-1)} \frac{\theta}{n} \times \frac{n}{\theta+n}$  et ceci est bien égal à  $\frac{(n+1)!}{\theta(\theta+1)\dots(\theta+n)} \frac{\theta}{n+1}$  qui est la valeur proposée par la formule de l'énoncé. Supposons maintenant que  $a_{n+1} = 0$ . Alors  $C_{n+1} = (a_1, \dots, a_{n+1})$  peut avoir été obtenu (si  $a_1 \geq 1$ ) à partir de  $C_n = (a_1 - 1, a_2, \dots, a_n)$  si le  $n + 1$ -ième citoyen a créé son propre parti, à partir de  $C_n = (a_1, \dots, a_{k-1}, a_k + 1, a_{k+1} - 1, a_{k+2}, \dots, a_n)$  (pour  $k = 1, \dots, n - 1$ ) si le  $n + 1$ -ième citoyen a adhéré à un parti qui avait  $k$  membres.

Le premier cas a pour probabilité  $\frac{n!}{\theta(\theta+1)\dots(\theta+n-1)}$  fois

$$\frac{1}{(a_1 - 1)!} \theta^{a_1 - 1} \prod_{k=2}^n \frac{1}{a_k!} \left(\frac{\theta}{k}\right)^{a_k} \times \frac{\theta}{\theta + n}$$

et le deuxième pour probabilité  $\frac{n!}{\theta(\theta+1)\dots(\theta+n-1)}$  fois

$$\prod_{j \neq k, k+1}^n \frac{1}{a_j!} \left(\frac{\theta}{j}\right)^{a_j} \frac{1}{(a_k + 1)!} \left(\frac{\theta}{k}\right)^{a_k + 1} \frac{1}{(a_{k+1} - 1)!} \left(\frac{\theta}{k+1}\right)^{a_{k+1} - 1} \times \frac{k(a_k + 1)}{\theta + n}$$

et la somme de ces différents termes vaut  $\frac{n!}{\theta(\theta+1)\dots(\theta+n-1)} \prod_{k=1}^n \frac{1}{a_k!} \left(\frac{\theta}{k}\right)^{a_k}$  fois

$$\begin{aligned} & \frac{a_1}{\theta} \frac{\theta}{\theta+n} + \sum_{k=1}^{n-1} \frac{1}{a_k+1} \frac{\theta}{k} a_{k+1} \frac{k+1}{\theta} \frac{k(a_k+1)}{\theta+n} \\ &= \frac{1}{\theta+n} \left( a_1 + \sum_{k=1}^{n-1} (k+1) a_{k+1} \right) \\ &= \frac{n+1}{\theta+n} \end{aligned}$$

et la formule de récurrence est vérifiée.

Pour vérifier empiriquement la formule dans le cas  $n = 5$ , remarquons que les seules valeurs possibles pour  $C_n$  dans ce cas sont

$$(0,0,0,0,1), (1,0,0,1,0), (0,1,1,0,0), (2,0,1,0,0), \\ (1,2,0,0,0), (3,1,0,0,0), (5,0,0,0,0)$$

Voir le code. Le résultat en est

```
les probas empiriques pour N=10000 sont
  pour (0, 0, 0, 0, 1) : 0.0072
  pour (1, 0, 0, 1, 0) : 0.0495
  pour (0, 1, 1, 0, 0) : 0.0340
  pour (2, 0, 1, 0, 0) : 0.1666
  pour (1, 2, 0, 0, 0) : 0.1253
  pour (3, 1, 0, 0, 0) : 0.4078
  pour (5, 0, 0, 0, 0) : 0.2096
les probas théoriques pour N=10000 sont
  pour (0, 0, 0, 0, 1) : 0.0079
  pour (1, 0, 0, 1, 0) : 0.0496
  pour (0, 1, 1, 0, 0) : 0.0331
  pour (2, 0, 1, 0, 0) : 0.1653
  pour (1, 2, 0, 0, 0) : 0.1240
  pour (3, 1, 0, 0, 0) : 0.4134
  pour (5, 0, 0, 0, 0) : 0.2067
```

ce qui compte tenu des approximations semble indiquer que la formule est bonne.

## MAO Probabilités-Statistiques

### Correction de l'examen final

#### Exercice 1 (loi de Kolmogorov-Smirnov)

Voir le code joint. Le tracé pour  $n = 50$  et  $N = 500$  dans le cas d'une loi  $\beta(2,3)$  et d'une loi  $\Gamma(3,1)$  est illustré dans la Figure 1, on voit qu'on a déjà une bonne adéquation. Rappelons que le  $n$  contrôle la convergence théorique et que le  $N$  contrôle la qualité de notre approximation de la loi de  $\sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|$ .

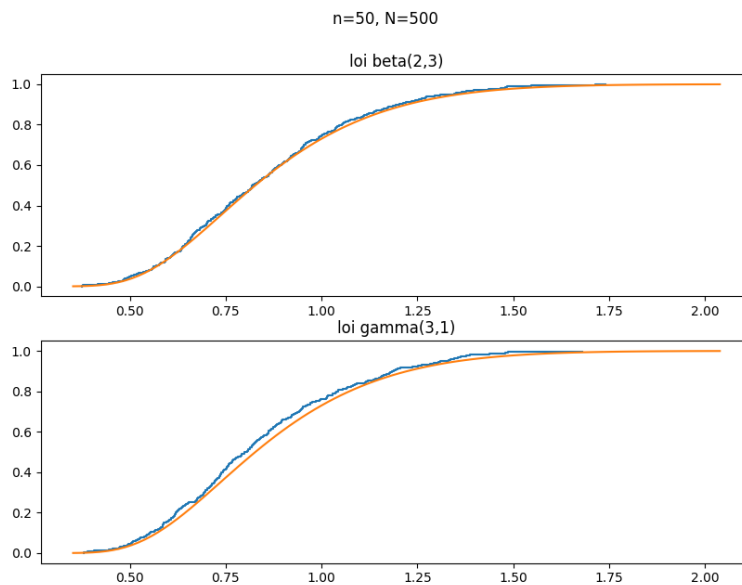


FIGURE 1 – Illustration de l'Exercice 1

#### Exercice 2 (points fixes d'une permutation aléatoire)

1. Puisque l'on travaille avec une loi uniforme, il suffit de compter : le nombre de permutations de  $\{1, \dots, n\}$  qui ont  $i$  pour point fixe est en bijection avec le nombre de permutations de  $\{1, \dots, n\} \setminus \{i\}$ , il y en a donc  $(n-1)!$ . On a

- par conséquent  $\mathbb{P}(i \in F_n) = \frac{(n-1)!}{n!} = \frac{1}{n}$ , d'où  $\mathbb{E}(f_n) = \mathbb{E}(\sum_{i=1}^n \mathbb{1}_{i \in F_n}) = \sum_{i=1}^n \frac{1}{n} = 1$ .
2. De la même manière que ci-dessus, on trouve  $\mathbb{P}(E \subset F_n) = \frac{(n-k)!}{n!}$  et donc si l'on note temporairement  $\mathcal{P}_k(n)$  l'ensemble des parties à  $k$  éléments de  $\{1, \dots, n\}$  et  $f_{n,k}$  le nombre de parties à  $k$  éléments de  $F_n$ , l'espérance cherchée vaut  $\mathbb{E}(\sum_{E \in \mathcal{P}_k(n)} \mathbb{1}_{E \in F_n}) = \sum_{E \in \mathcal{P}_k(n)} \frac{(n-k)!}{n!} = \binom{n}{k} \times \frac{(n-k)!}{n!} = \frac{1}{k!}$ .
  3. Avec les notations ci-dessus on a  $f_{n,k} = \binom{f_n}{k} = \frac{f_n(f_n-1)\dots(f_n-k+1)}{k!}$ , ce qui répond à la question. On a donc  $\mathbb{E}(f_n(f_n-1)\dots(f_n-k+1)) = 1$ .
  4. Puisque  $f_n$  est à valeurs dans  $\{0, \dots, n\}$  on a  $\mathbb{E}(s^{f_n}) = \sum_{k=0}^n \mathbb{P}(f_n = k) s^k$ , donc  $\varphi_n$  est bien un polynôme de degré  $n$  au plus. Il est immédiat que l'on peut intervertir dérivation et espérance, et alors  $\frac{d^k}{ds^k} \varphi_n(s) = \mathbb{E}(f_n(f_n-1)\dots(f_n-k+1))$  qui vaut 1 par la question précédente. L'unique polynôme vérifiant les deux contraintes ci-dessus est  $\sum_{k=0}^n \frac{(s-1)^k}{k!}$ , c'est donc bien l'expression de  $\varphi_n$ .
  5. La fonction génératrice d'une loi de Poisson est  $\sum_{k=0}^{\infty} \frac{e^{-1}}{k!} s^k = e^{(s-1)}$ . Puisque cette série est convergente pour tout  $s \in [0, 1]$ , on a immédiatement  $\varphi_n(s) \rightarrow \mathbb{E}(s^P)$  et donc convergence en loi de  $f_n$  vers  $\mathcal{P}(1)$ .
  6. Voir le code et la Figure 2 : ça a l'air de converger très vite.

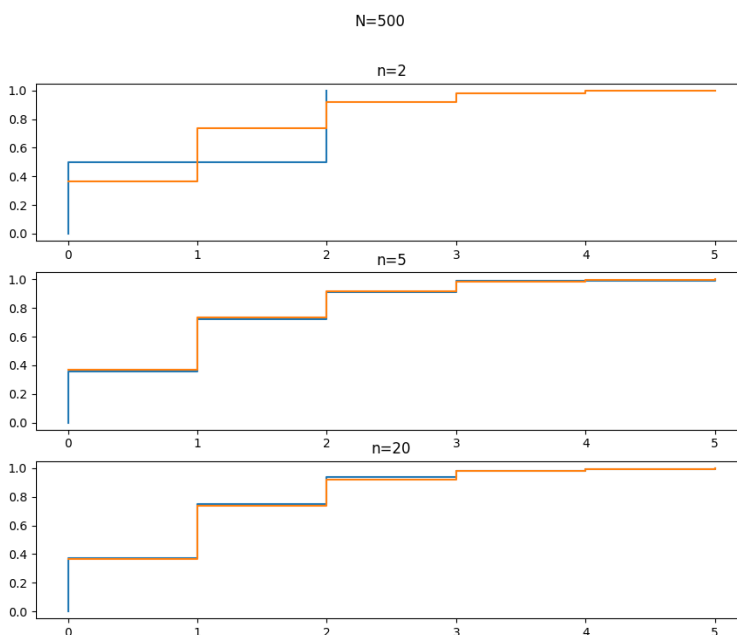


FIGURE 2 – Illustration de l'Exercice 2

**Exercice 3 (nombre de triangles dans un graphe d'Erdős-Rényi)**

1. Un triangle  $abc$  est un triangle du graphe si et seulement si  $\{a, b\}$ ,  $\{b, c\}$  et  $\{c, a\}$  sont des arêtes du graphe. Les trois événements sont indépendants et de même probabilité  $\frac{1}{8}$ , et c'est le cas pour chacun des  $\binom{n}{3}$  choix de triplets  $a, b, c$ , donc  $\mathbb{E}(t_n) = \sum_{a,b,c} \mathbb{E}(\mathbb{1}_{abc \in T_n}) = \frac{1}{8} \binom{n}{3}$ .
2. Comme  $t_n = \sum_{a,b,c} \mathbb{1}_{abc \in T_n}$  on a immédiatement  $t_n^2 = \sum_{a,b,c} \sum_{d,e,f} \mathbb{1}_{abd \in T_n} \mathbb{1}_{def \in T_n}$  et on obtient l'expression voulue. Comme pour  $k$  fixé  $\binom{n}{k} = \frac{n(n-1)\dots(n-k+1)}{k!}$  on a immédiatement  $\binom{n}{k} = \frac{n^k}{k!} + O(n^{k-1})$ . Estimons d'abord

$$\sum_{a,b,c,d,e,f} \mathbb{E}(\mathbb{1}_{abc \in T_n} \mathbb{1}_{def \in T_n}) \mathbb{1}_{\text{card}\{a,b,c,d,e,f\}=6}.$$

On a  $abc \in T_n$  et  $def \in T_n$  si et seulement si  $\{a, b\}$ ,  $\{b, c\}$ ,  $\{c, a\}$ ,  $\{d, e\}$ ,  $\{e, f\}$ ,  $\{f, d\}$  sont toutes des arêtes du graphe, et la contrainte  $\text{card}\{a, b, c, d, e, f\}$  fait que ces six événements sont indépendants. Par ailleurs, il y a  $\binom{n}{6}$  tels 6-uplets et  $\binom{6}{3}$  manières de fabriquer deux triangles avec. On a donc

$$\sum_{a,b,c,d,e,f} \mathbb{E}(\mathbb{1}_{abc \in T_n} \mathbb{1}_{def \in T_n}) \mathbb{1}_{\text{card}\{a,b,c,d,e,f\}=6} = \frac{1}{2^6} \binom{n}{6} \binom{6}{3} = \frac{1}{2^6 (3!)^2} n^6 + O(n^5).$$

Si l'on considère la somme contrainte à  $\text{card}\{a, b, c, d, e, f\} = 5$  on voit que la probabilité de l'événement " $abc \in T_n$  et  $def \in T_n$ " reste la même, mais il n'y a plus que  $\binom{n}{5}$  tels 5-uplets,  $\binom{5}{3}$  manières de faire un premier triangle avec, et 3 manières de choisir le point "double". Dans la somme contrainte à  $\text{card}\{a, b, c, d, e, f\} = 4$  on voit qu'il y a nécessairement deux points en commun donc il n'y plus que cinq arêtes, donc à  $a, b, c, d, e, f$  fixés la probabilité de l'événement est  $\frac{1}{2^5}$  et il y a  $\binom{n}{4}$  manières de choisir le 4-uplet,  $\binom{4}{3}$  manières de faire un premier triangle, et 3 manières de choisir l'arête "double". Dans la somme contrainte à  $\text{card}\{a, b, c, d, e, f\} = 4$  il n'y plus que trois arêtes, donc à  $a, b, c, d, e, f$  fixés la probabilité de l'événement est  $\frac{1}{2^3}$  et il y a  $\binom{n}{3}$  manières de choisir le 3-uplet. Dans tous les cas on voit que les termes où  $\text{card}\{a, b, c, d, e, f\} \leq 5$  sont au plus des  $O(n^5)$ , donc  $\mathbb{E}(t_n^2) = \frac{1}{2^6 (3!)^2} n^6 + O(n^5)$ .

3. Comme  $\mathbb{E}(t_n)^2 = \left(\frac{1}{2^3 3!} n^3 + O(n^2)\right)^2$  on a  $\text{var}(t_n) = O(n^5)$ . Par Bienaymé-Tchebychev on a alors pour  $\epsilon > 0$  :

$$\begin{aligned} \mathbb{P}\left(\left|\frac{t_n}{\binom{n}{3}} - \frac{1}{8}\right| > \epsilon\right) &= \mathbb{P}\left(\left|t_n - \frac{1}{8} \binom{n}{3}\right| > \epsilon \binom{n}{3}\right) \\ &\leq \epsilon^{-2} \left(\frac{1}{2^3 3!} n^3 + O(n^2)\right)^{-2} \text{var}(t_n) \end{aligned}$$

qui, d'après les résultats précédents, est un  $O(1/n)$  donc tend vers zéro. On a bien la convergence en probabilité demandée.

4. En conservant les termes en  $n^5$  ci-dessus, on s'aperçoit qu'en fait  $\text{var}(t_n) = O(n^4)$ . Dans ce cas la majoration ci-dessus donne que  $\mathbb{P}\left(\left|\frac{t_n}{\binom{n}{3}} - \frac{1}{8}\right| > \epsilon\right)$  est un  $O(1/n^2)$  donc est sommable. Par une application standard de Borel-Cantelli, on a la convergence presque-sûre vers zéro de  $\frac{t_n}{\binom{n}{3}}$  vers  $\frac{1}{8}$ .
5. On voit immédiatement que de manière générale le coefficient  $i, j$  de  $A_n^k$  est le nombre de chemins (orientés) de longueur  $k$  allant de  $i$  à  $j$  en parcourant les arêtes du graphe. En particulier, le coefficient  $i, i$  de  $A_n^3$  est le nombre de chemins de longueur 3 de  $i$  à  $i$ . Or un triangle contenant  $i$  est exactement un chemin de longueur 3 une fois que l'on a oublié l'orientation, et il y a deux sens de parcours d'un triangle. Ce coefficient  $i, i$  de  $A_n^3$  est donc bien le double du nombre de triangles du graphe qui ont  $i$  comme l'un de leurs sommets. Comme chaque triangle a trois sommets, il apparaît trois fois dans  $\sum_{i=1}^n (A_n^3)_{i, i}$  et on a bien  $t_n = \frac{1}{6} \text{tr}(A_n^3)$ .
6. Voir le code joint. En allant jusqu'à  $n = 100$ , on obtient la Figure 3. Si l'on n'est pas satisfait, on peut aller plus loin mais c'est plus long. On semble bien observer une convergence presque-sûre. Et pourquoi a-t-on illustré la conver-

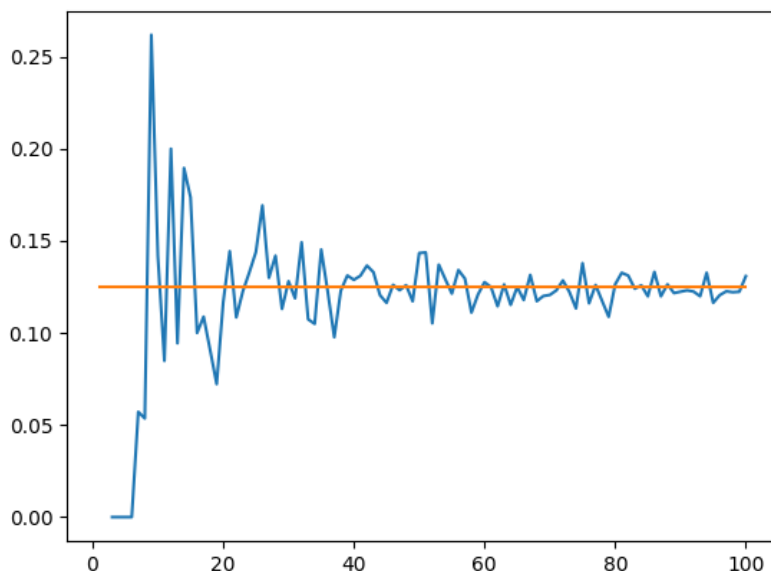


FIGURE 3 – Illustration de l'Exercice 3

gence presque-sûre plutôt que la convergence en probabilité ? il y a deux raisons à ça : d'une part, c'est plus facile, d'autre part la convergence presque-sûre implique la convergence en probabilité donc si on illustre la première alors on a



illustré la deuxième...

#### Exercice 4 (résolution du problème de Dirichlet)

1. Sur l'événement  $n < \tau_{\partial D}$  on a  $X_n \in \text{int } D$ . En conditionnant au premier pas, on a

$$\begin{aligned} \mathbb{E}(f(X_{n+1}) | X_n) &= f(X_n + (0, +1)) \times \mathbb{P}(X_{n+1} = X_n + (0, +1) | X_n) \\ &\quad + f(X_n + (0, -1)) \times \mathbb{P}(X_{n+1} = X_n + (0, -1) | X_n) \\ &\quad + f(X_n + (+1, 0)) \times \mathbb{P}(X_{n+1} = X_n + (+1, 0) | X_n) \\ &\quad + f(X_n + (-1, 0)) \times \mathbb{P}(X_{n+1} = X_n + (-1, 0) | X_n) \\ &= \frac{1}{4} \left( f(X_n + (0, +1)) + f(X_n + (0, -1)) \right. \\ &\quad \left. + f(X_n + (+1, 0)) + f(X_n + (-1, 0)) \right) \\ &= Pf(X_n) \\ &= f(X_n) \end{aligned}$$

par hypothèse que  $f - Pf = 0$  sur  $\text{int } D$ . On a montré que  $(f(X_n))_n$  est une martingale sous  $\mathbb{P}_x$  et,  $\tau_{\partial D}$  étant un temps d'arrêt presque-sûrement fini,  $(f(X_{\inf(n, \tau_{\partial D})}))_n$  en est une aussi. En particulier, en prenant  $n \rightarrow \infty$  ci-dessus on a par convergence dominée  $\mathbb{E}_x(f(X_{\tau_{\partial D}})) = \mathbb{E}_x(f(X_0))$ , or par définition de  $f$  et de  $X_0$  on a  $\mathbb{E}_x(f(X_{\tau_{\partial D}})) = \mathbb{E}_x(g(X_{\tau_{\partial D}}))$  et  $\mathbb{E}_x(g(X_0)) = f(x)$ .

2. On a en utilisant la propriété de Markov, pour tout  $x \in \text{int } D$  :

$$\begin{aligned} \mathbb{E}_x(f(X_{\partial D})) &= \mathbb{E}_x \left( \mathbb{E}_{X_1}(f(X_{\partial D})) \right) \\ &= \mathbb{E}_{x+(0,+1)}(f(X_{\partial D})) \mathbb{P}_x(X_1 = x + (0, +1)) \\ &\quad + \mathbb{E}_{x+(0,-1)}(f(X_{\partial D})) \mathbb{P}_x(X_1 = x + (0, -1)) \\ &\quad + \mathbb{E}_{x+(+1,0)}(f(X_{\partial D})) \mathbb{P}_x(X_1 = x + (+1, 0)) \\ &\quad + \mathbb{E}_{x+(-1,0)}(f(X_{\partial D})) \mathbb{P}_x(X_1 = x + (-1, 0)) \end{aligned}$$

donc  $f(x) = Pf(x)$  pour  $x \in \text{int } D$ . Par ailleurs conditionnellement à  $X_0 = x \in \partial D$  on a  $\tau_{\partial D} = 0$  d'où  $f(x) = \mathbb{E}_x(f(X_0)) = g(x)$ .

3. On va simplement estimer  $\mathbb{E}_x(g(X_{\tau_{\partial D}}))$  par la moyenne empirique sur un grand nombre de réalisations de  $g(X_{\tau_{\partial D}})$ . Comme ce processus est à valeurs dans  $[-1, +1]$ , l'inégalité de Hoeffding donne que la probabilité d'une déviation de  $t$  entre moyenne empirique et moyenne théorique est majorée par  $2 \exp -\frac{Nt^2}{2}$ .

Avec un niveau de confiance de 99% on peut donc obtenir une précision  $\sqrt{\frac{2 \log 200}{N}}$  pour une taille d'échantillon  $N$  donnée.

4. Voir le code. On obtient pour  $r = 50$  la Figure 4.

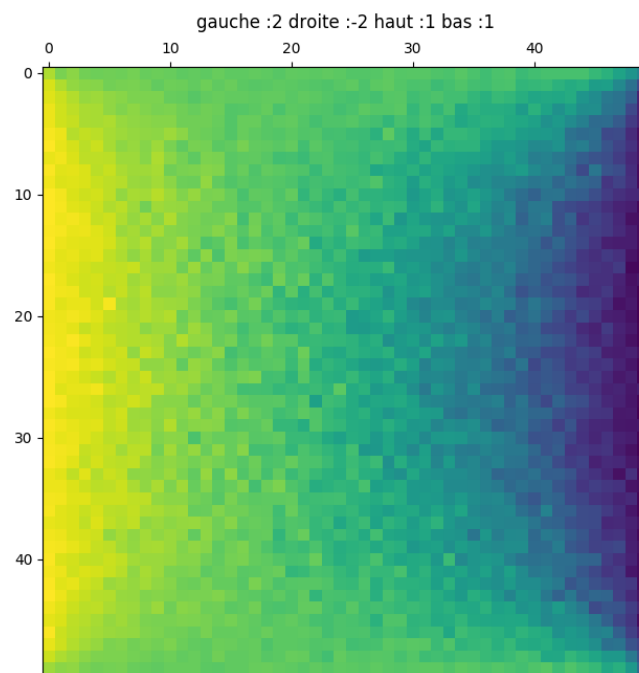


FIGURE 4 – Illustration de l'Exercice 4

# MAO Probabilités-Statistiques

## Examen de seconde session

Cet énoncé est basé sur un texte utilisé lors de la troisième épreuve orale (dite “de modélisation”) du concours de l’agrégation. Dans cette épreuve, on vous propose un texte décrivant un problème “pratique” à étudier, une modélisation de ce problème et quelques résultats mathématiques liés. Le rôle du candidat est de commenter ces résultats, ce qui veut dire en démontrer une partie mais surtout discuter la pertinence de leurs hypothèses et les modifications qu’on peut leur apporter, et les appliquer pour donner une réponse au problème pratique de départ.

L’informatique joue là-dedans un rôle prépondérant puisqu’il permet d’illustrer les résultats mais aussi de conjecturer leurs extensions en testant leurs conclusions sous des hypothèses modifiées.

De l’énoncé d’origine, nous n’avons gardé que la partie “modèle”, et réorganisé le reste pour mieux vous suggérer le type de questions que vous pourrez vous poser. Le texte d’origine complet est disponible à l’adresse

<http://agreg.org/Textes/public2015-A7.pdf>

Certains insectes comme les fourmis utilisent pour se reconnaître des signaux chimiques. Ainsi, chaque fourmi porte sur son corps et dépose partout où elle passe des quantités infinitésimales d'une substance, appelée phéromone, qui est caractéristique de la fourmilière à laquelle elle appartient. Elle peut ainsi sentir si le chemin qu'elle suit a déjà été emprunté par une de ses congénères ou non, ou s'il a été emprunté par un individu d'une autre fourmilière. Ceci a des conséquences qualitatives importantes sur la façon dont les fourmis se déplacent et nous nous proposons d'étudier quelques modèles simples de ce type de phénomènes.

### 1. Le modèle

Considérons la situation où, pour se rendre d'un endroit à un autre, les individus d'une fourmilière ont le choix entre deux trajets disjoints. Chaque fois qu'une fourmi emprunte l'un des deux chemins, elle y dépose une certaine quantité de phéromone et cette quantité peut éventuellement dépendre de la quantité déjà présente sur le chemin. Nous supposons que chaque fourmi choisit de façon aléatoire le chemin qu'elle emprunte, en affectant à chacun une probabilité qui est proportionnelle à la quantité de phéromone qui y est présente.

Notons  $a$  et  $b$  les deux chemins possibles. Nous modéliserons les trajets successifs suivis par les fourmis par une suite  $X = (X_1, X_2, \dots)$  de variables aléatoires à valeurs dans  $\{a, b\}$ . Pour chaque  $n \geq 1$ , notons  $\alpha_n$  et  $\beta_n$  les quantités de phéromones présentes respectivement sur les chemins  $a$  et  $b$  après le  $n$ -ième trajet. Ce sont deux variables aléatoires réelles que nous supposons strictement positives pour éviter d'avoir à discuter des cas particuliers. Par convention, les nombres  $\alpha_0$  et  $\beta_0$  représentent les quantités de phéromones présentes au début de l'expérience et font donc partie des données du problème. L'autre donnée est une fonction  $r : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$ , dite de renforcement, qui modélise la façon dont les fourmis déposent leur phéromone pendant le trajet. Nous ferons l'hypothèse que la quantité de phéromone présente sur un chemin après le passage d'une fourmi est l'image par  $r$  de la quantité présente avant son passage. Il semble donc raisonnable de supposer que  $r(x) \geq x$  pour tout  $x \in \mathbb{R}_+^*$ . Nous nous intéresserons à trois fonctions particulières, qui sont définies par

$$(1) \quad \forall x > 0, \quad r_0(x) = x, \quad r_1(x) = x + 1, \quad r_2(x) = \rho x,$$

où  $\rho \in ]1, +\infty[$  est un paramètre réel.

On construit les variables aléatoires  $X_n, \alpha_n, \beta_n$  par récurrence sur  $n$ . Sachant  $\alpha_n$  et  $\beta_n$ , les variables  $X_{n+1}, \alpha_{n+1}$  et  $\beta_{n+1}$  sont indépendantes de  $X_1, \dots, X_n, \alpha_1, \beta_1, \dots, \alpha_n, \beta_n$  et de loi donnée par

$$\begin{aligned} \mathbb{P}[X_{n+1} = a, \alpha_{n+1} = r(\alpha_n), \beta_{n+1} = \beta_n | \alpha_n, \beta_n] &= \frac{\alpha_n}{\alpha_n + \beta_n} \\ \mathbb{P}[X_{n+1} = b, \alpha_{n+1} = \alpha_n, \beta_{n+1} = r(\beta_n) | \alpha_n, \beta_n] &= \frac{\beta_n}{\alpha_n + \beta_n}. \end{aligned}$$

Notons que la suite de variables aléatoires  $X = (X_1, X_2, \dots)$  n'est pas en général une chaîne de Markov. En revanche, la suite des triplets  $((X_n, \alpha_n, \beta_n))_{n \geq 1}$  en est une.

Dans l'étude du modèle, nous allons nous intéresser en particulier au nombre total de passages sur chacun des deux chemins avant un temps donné. Définissons les deux quantités suivantes :

$$\forall n \geq 1, \quad A_n = \#\{k \in \{1, \dots, n\} : X_k = a\} \quad \text{et} \quad B_n = \#\{k \in \{1, \dots, n\} : X_k = b\}.$$

Elles satisfont la relation  $A_n + B_n = n$ . Nous poserons  $A_0 = B_0 = 0$ .

Dans ce sujet, nous nous intéresserons uniquement au cas des renforcements linéaire ( $r(x) = x + 1$ ) et géométrique ( $r(x) = \rho x$ ). Pour simplifier les notations et les simulations on remplacera "a" et "b" par "1" et "0" respectivement ;  $X_n$  sera donc à valeurs

dans  $\{0, 1\}$  et  $A_n, B_n$  représenteront

$$A_n = \#\{k \in \{1, \dots, n\} \mid X_k = 1\}, \quad B_n = \#\{k \in \{1, \dots, n\} \mid X_k = 0\}.$$

On va commencer par écrire deux fonctions permettant de simuler l'expérience ci-dessus.

### Exercice 1

1. *Ecrivez deux fonctions `trajectoire_lin` et `trajectoire_geo` qui prennent en entrée des variables `n`, `alpha0`, `beta0`, plus `rho` dans le cas géométrique, et donnent en sortie une variable `x` qui représente le  $n$ -uplet  $(X_1, \dots, X_n)$ , dans les cas respectifs des renforcements linéaire et géométrique.*
2. *Dans le cas linéaire, simulez plusieurs fois des trajectoires de  $A_n/n$  (c'est-à-dire la suite  $A_k/k$  pour  $k = 1, \dots, n$  pour  $n$  bien choisi), et tracez les courbes correspondantes; d'abord dans le cas  $\alpha_0 = \beta_0 = 1$ , puis dans d'autres cas, par exemple  $\alpha_0 = 1$  et  $\beta_0 = 2$ . Quel type de convergence semble-t-on avoir pour  $(A_n/n)_n$ ?*
3. *Même question dans le cas géométrique. On considérera une seule valeur de  $\rho$ , que l'on prendra relativement faible, par exemple 1,2.*

Les deux programmes de simulation ci-dessus sont utilisés dans toute la suite. Demandez à votre chargé de cours de vérifier que vos résultats sont cohérents, cela vous évitera de vous fourvoyer pour la suite.

### 1 Cas du renforcement linéaire : $r(x) = x + 1$

On se place dans cette section dans le cas où  $r(x) = x + 1$ . On va montrer la convergence de  $A_n/n$ ; on va pouvoir expliciter de manière théorique la limite dans le cas  $\alpha_0 = \beta_0 = 1$ , puis on la conjecturera dans les autres cas, et on utilisera les simulations pour vérifier cette intuition.

### Exercice 2

1. *Exprimez  $\alpha_n + \beta_n$  en fonction de  $\alpha_0 + \beta_0$ . Exprimez une relation entre  $\alpha_n$  et  $A_n$ , entre  $\beta_n$  et  $B_n$ .*
2. *Calculez  $\mathbb{E}(A_n \mid \mathcal{F}_n)$  où  $\mathcal{F}_n$  est la tribu engendrée par  $X_1, \dots, X_n$ . Montrez que les suites  $(\frac{\alpha_0 + A_n}{\alpha_0 + \beta_0 + n})_n$  et  $(\frac{\beta_0 + B_n}{\alpha_0 + \beta_0 + n})_n$  sont des martingales bornées.*
3. *En déduire que  $(\frac{A_n}{n})_n$  converge presque-sûrement vers une variable aléatoire intégrable qu'on note  $L$ .*

On va maintenant chercher à identifier la loi de  $L$ . Dans l'exercice suivant, on va le faire de manière théorique dans le cas  $\alpha_0 = \beta_0 = 1$ .

**Exercice 3** *On se place dans le cas  $\alpha_0 = \beta_0 = 1$ .*

1. *Montrez par récurrence que  $\frac{1+A_n}{2+n}$  suit une loi uniforme sur  $\{\frac{1}{2+n}, \frac{2}{2+n}, \dots, \frac{1+n}{2+n}\}$ .*
2. *Déduisez-en que  $L$  suit une loi uniforme sur  $[0, 1]$ .*

3. Illustrez le point précédent : simulez un  $N$ -échantillon de  $A_n/n$  (on aura le droit d'utiliser une boucle), où vous choisirez  $n$  assez grand pour que  $A_n/n$  semble assez proche de sa valeur limite, à partir de vos simulations de l'exercice 1. Tracez la fonction de répartition empirique du  $N$ -échantillon considéré et comparez-la à la fonction de répartition (théorique) d'une loi uniforme.

On aimerait identifier la loi de  $L$  dans les autres cas. On ne sait le faire de manière théorique, mais on a des raisons de penser que c'est une loi Bêta. On rappelle qu'une loi Bêta de paramètres  $a > 0$  et  $b > 0$ , notée  $\beta(a, b)$ , a pour densité

$$x \mapsto \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} \mathbb{1}_{[0,1]}(x).$$

Si l'on connaissait  $a$  et  $b$ , on pourrait comparer les fonctions de répartition, empirique de  $A_n/n$  et théorique de la loi  $\beta(a, b)$ , mais ici  $a$  et  $b$  sont inconnus. On va donc commencer par les estimer.

#### Exercice 4

1. On rappelle que si  $L \sim \beta(a, b)$  alors

$$\mathbb{E}(L) = \frac{a}{a+b} \quad \text{et} \quad \text{var}(B) = \frac{ab}{(a+b)^2(a+b+1)},$$

ce qui implique

$$a = \mathbb{E}(L) \left( \frac{\mathbb{E}(L)(1-\mathbb{E}(L))}{\mathbb{E}(L^2) - \mathbb{E}(L)^2} - 1 \right) \quad b = (1-\mathbb{E}(L)) \left( \frac{\mathbb{E}(L)(1-\mathbb{E}(L))}{\mathbb{E}(L^2) - \mathbb{E}(L)^2} - 1 \right).$$

Si l'on vous donne un  $N$ -échantillon  $L_1, \dots, L_N$  de loi  $\beta(a, b)$ , quels estimateurs de  $a$  et  $b$  pouvez-vous proposer? Quelle bonne propriété ont-ils de manière (presque) immédiate?

2. Ecrivez une fonction estimab qui, si on lui donne en entrée une variable  $\mathbb{1}$  représentant un  $N$ -échantillon comme ci-dessus, renvoie deux variables `achap`, `bchap` qui sont les estimateurs proposés à la question précédente.
3. Pour  $\alpha_0 = 1$ ,  $\beta_0 = 2$ , simulez un  $N$ -échantillon de  $A_n/n$  pour un  $n$  suffisamment grand (on ne fera pas varier  $n$ ; prenez par exemple  $n = N = 1000$ ). En assimilant chaque  $A_n/n$  à un  $L$  de loi  $\beta(a, b)$  (autrement dit, en faisant comme si chaque  $A_n/n$  suivait vraiment une loi  $\beta(a, b)$ ) proposez une valeur pour  $a$  et  $b$ . Recommencez avec d'autres valeurs de  $\alpha$  et  $\beta$ , puis proposez une formule (très simple) pour  $a$  et  $b$  en fonction de  $\alpha$  et  $\beta$ .
4. Vérifiez cette intuition : simulez un  $N$ -échantillon de  $A_n/n$ , où vous choisirez  $n$  assez grand pour que  $A_n/n$  semble assez proche de sa valeur limite, à partir de vos simulations de l'exercice 1. Tracez la fonction de répartition empirique du  $N$ -échantillon considéré et comparez-la à la fonction de répartition empirique d'une loi  $\beta(a, b)$  pour les  $a$  et  $b$  proposés.

## 2 Cas du renforcement géométrique : $r(x) = \rho x$

On se place dans cette section dans le cas où  $r(x) = \rho x$  pour un  $\rho > 1$ . On suppose jusqu'à la dernière question de cette section que  $\alpha_0 = \beta_0 = 1$ . On a donc

$$\alpha_n = \rho^{A_n} \quad \beta_n = \rho^{B_n},$$

de sorte que

$$\mathbb{P}(A_{n+1} = A_n + 1 \mid A_n, B_n) = \frac{\rho^{A_n}}{\rho^{A_n} + \rho^{B_n}} \quad \mathbb{P}(A_{n+1} = A_n \mid A_n, B_n) = \frac{\rho^{B_n}}{\rho^{A_n} + \rho^{B_n}}.$$

On va montrer que presque-sûrement, il existe un rang à partir duquel la suite  $(X_n)_n$  est stationnaire, et que la valeur correspondante est 0 ou 1 avec probabilités  $1/2$  et  $1/2$ . Plus formellement, on va montrer que

$$\mathbb{P}(\exists N \mid \forall n \geq N, X_n = 0) = \frac{1}{2} \quad \mathbb{P}(\exists N \mid \forall n \geq N, X_n = 1) = \frac{1}{2}. \quad (1)$$

**Exercice 5** On note  $\Delta_n = |A_n - B_n|$ .

1. Montrez que

$$\mathbb{P}(\Delta_{n+1} - \Delta_n = +1 \mid A_n, B_n) = \begin{cases} \frac{\rho^{\Delta_n}}{1 + \rho^{\Delta_n}} & \text{si } A_n > B_n \\ \frac{\rho^{\Delta_n}}{1 + \rho^{\Delta_n}} & \text{si } A_n < B_n \\ 1 & \text{si } A_n = B_n \end{cases}$$

(les deux première expressions sont bien identiques). Et si l'on remplace  $+1$  par  $-1$  ci-dessus ?

2. En déduire que

$$\begin{aligned} \mathbb{P}(\Delta_{n+1} - \Delta_n = +1 \mid \Delta_n) &= \frac{\rho^{\Delta_n}}{1 + \rho^{\Delta_n}} \mathbb{1}_{\Delta_n \neq 0} + \mathbb{1}_{\Delta_n = 0} \\ \mathbb{P}(\Delta_{n+1} - \Delta_n = -1 \mid \Delta_n) &= \frac{1}{1 + \rho^{\Delta_n}} \mathbb{1}_{\Delta_n \neq 0}. \end{aligned}$$

3. On définit une fonction  $f$  sur  $\mathbb{N}$  par  $f(0) = 0$  et  $f(n) = \sum_{k=0}^{n-1} \rho^{-\frac{k(k+1)}{2}}$ . Vérifiez l'égalité

$$\rho^n (f(n+1) - f(n)) = f(n) - f(n-1)$$

et montrez qu'elle implique l'égalité

$$\mathbb{E}(f(\Delta_{n+1}) \mid \Delta_n) = f(\Delta_n) + \mathbb{1}_{\Delta_n = 0}.$$

Peut-on avoir  $f(m) = f(n)$  pour  $m \neq n$  ?

4. Dédisez-en que le processus  $f(\Delta_n)$  est une sous-martingale positive et bornée, et converge donc presque-sûrement. Montrez que  $f(\Delta_n) \neq f(\Delta_{n+1})$ . Concluez-en que  $f(\Delta_n)$  converge presque sûrement vers  $\sum_{k=0}^{\infty} \rho^{-\frac{k(k+1)}{2}}$ , puis que  $\Delta_n$  tend presque-sûrement vers  $+\infty$ . Dédisez-en que pour tout  $D$ , la probabilité  $\mathbb{P}(\Delta_N \geq D)$  tend vers 1 quand  $N \rightarrow \infty$ .
5. Montrez que, pour  $N$  fixé,

$$\mathbb{P}(\forall n \geq N, \Delta_{n+1} > \Delta_n \mid \Delta_N = d) = \prod_{k=d}^{+\infty} \frac{1}{1 + \rho^{-k}}.$$

6. On note  $Q(d)$  le terme de droite. Montrez qu'il définit une suite croissante qui converge vers 1 en  $+\infty$ . Montrez les majorations successives, pour  $D \in \mathbb{N}$  :

$$\begin{aligned} \mathbb{P}(\exists N \text{ t.q. } \forall n \geq N, \Delta_{n+1} > \Delta_n) &\geq \sup_N \mathbb{P}(\forall n \geq N, \Delta_{n+1} > \Delta_n) \\ &\geq \sup_N \sum_{d=0}^{\infty} Q(d) \mathbb{P}(\Delta_N = d) \\ &\geq Q(D) \times \sup_N \mathbb{P}(\Delta_N \geq D). \end{aligned}$$

7. Dédisez-en que  $\mathbb{P}(\exists N \mid \forall n \geq N, \Delta_{n+1} > \Delta_n) = 1$ , puis concluez à la propriété (1). Ceci vous semble-t-il cohérent avec Les simulations faites à l'exercice 1 ?

On conjecture maintenant que la propriété de stationnarité (1) reste vraie pour tous  $\alpha_0, \beta_0$ , avec des probabilités différentes de  $1/2$  et  $1/2$ . On suppose maintenant admis qu'on a une convergence presque-sûre de  $(X_n)_n$  vers une variable aléatoire  $G$  de loi de Bernoulli, de paramètre  $p$ , avec d'après ce qui précède  $p = 1/2$  lorsque  $\alpha_0 = \beta_0 = 1$ .

### Exercice 6

Montrez que le paramètre  $p$  ne dépend que de  $\beta_0/\alpha_0$ . Ecrivez une fonction `estimp`, qui, prend en entrée des variables `N`, `alpha0`, `beta0`, et `rho`, et donne en sortie une variable `pchap` qui estime le paramètre  $p$  de la loi  $G$ . Comment choisir vos paramètres pour avoir une estimation à  $0,01$  près ? On pourra fixer  $\alpha_0 = 1$  et tracer la courbe du  $p$  estimé pour différentes valeurs de  $\beta_0$ .