Likelihood ratio inequalities with applications to various mixtures.

E. Gassiat

Laboratoire de Mathématiques, Equipe de Probabilités, Statistique et Modélisation, Bâtiment 425, Université de Paris-Sud, 91 405 Orsay Cedex, France.

Abstract: We give two simple inequalities on likelihood ratios. A first application is the consistency of the maximum-penalized marginal-likelihood estimator of the number of populations in a mixture with Markov regime. The second application is the derivation of the asymptotic power of the likelihood ratio test under loss of identifiability for contiguous alternatives. Finally, we propose self-normalized score tests that have exponentially decreasing level and asymptotic power 1.

Résumé : Nous donnons deux inégalités pour les rapports de vraisemblance. Une première application est la consistance de l'estimateur de vraisemblance marginale pénalisée du nombre de composants dans un mélange de populations à régime markovien. Une deuxième application est le calcul de la puissance asymptotique du test de rapport de vraisemblance dans les cas de non identifiabilité des paramètres sous l'hypothèse nulle, et pour des alternatives contigües. Enfin, nous proposons un test du score auto-normalisé dont l' erreur de première espèce décroit exponentiellement vite et la puissance asymptotique est égale à 1.

1 Introduction

Let $(X_i)_{i\in\mathbb{N}}$ be a strictly stationary sequence of random variables with distribution $\mathbb{P}$ on a polish space $\mathbb{X}$. Let $\mathcal{G}$ be a set of densities g with respect to some positive measure ν, such that the marginal density f of all X_i's is an element of $\mathcal{G}$; f will thus be fixed throughout the paper. For any g in $\mathcal{G}$, set

$$\ell_n(g) = \sum_{i=1}^{n} \log g(X_i). \tag{1}$$

When the variables (X_i) are independently distributed, ℓ_n is the likelihood of the observations. Otherwise, call it marginal-likelihood. In this paper, we mainly investigate the use of ℓ_n as an inference tool for the number of populations in various mixtures of populations.

The testing problem of the number of populations in a mixture is a typical example of testing problems exhibiting a lack of identifiability of the alternative under the null hypothesis. This lack of identifiability leads to the degeneracy of the Fisher information of the model, so that the classical chi-square theory does not apply. In two recent papers, D. Dacunha-Castelle and E. Gassiat [2], [3] proposed a theory of reparametrization to solve such problems, that they called "locally conic parametrization". Roughly speaking, the idea is to approach the null hypothesis using directional submodels in which the Fisher information is normalized to be uniformly equal to one. This theory provides a guideline to solve general testing problems in which non identifiability occurs, even using other statistics than likelihood ratios, or handling non identically distributed or dependent observations (see [3] for an application to ARMA models). The theory was developed in [2] and [3] to obtain the asymptotic level of the likelihood ratio test (LRT) for the number of populations in a mixture with independent observations. Recently, Liu and Shao [7] proposed an alternative proof to derive this asymptotic distribution. The idea is to give an expansion of the likelihood ratio using a "generalized differentiable in quadratic mean condition", and to handle simultaneously Hellinger distances and Pearson distances.

In this paper, we prove two general simple inequalities on likelihood ratios (or marginal-likelihood ratios) that allow to derive, in a very simple way, upper bounds for the LRT statistic. We then apply it to several mixtures of populations. For any $g \in \mathcal{G}$, define

$$s_g = \frac{\frac{g-f}{f}}{\| \frac{g-f}{f} \|_2} = \frac{\frac{g}{f} - 1}{\| \frac{g}{f} - 1 \|_2}, \tag{2}$$

where $\| \cdot \|_2$ is the norm in $L^2(f d\nu)$. Let us now state the first inequality.

Inequality 1.1

$$\sup_{g \in \mathcal{G} : \ell_n(g) - \ell_n(f) \geq 0} \| \frac{g-f}{f} \|_2 \leq 2 \sup_{g \in \mathcal{G}} \frac{\sum_{i=1}^n s_g(X_i)}{\sum_{i=1}^n (s_g)_-^2 (X_i)},$$

with $(s_g)_-(x) = \min\{0, s_g(x)\}$

The second inequality is the following.

Inequality 1.2

$$\sup_{g \in \mathcal{G}} (\ell_n(g) - \ell_n(f)) \leq \frac{1}{2} \sup_{g \in \mathcal{G}} \frac{(\sum_{i=1}^n s_g(X_i))^2}{\sum_{i=1}^n (s_g)_-^2 (X_i)}.$$

Inequality 1.1 and inequality 1.2 are proved in section 5. Inequality 1.2 allows to prove the tightness of the LRT (or marginal LRT) statistic under simple assumptions. It is used in section 2 to prove the consistency of the estimator of the number of populations with Markov regime using penalized marginal-likelihood. This estimator has been used for instance by Leroux and Puterman [6], but was only known not to underestimate the number of populations with asymptotic probability 1, see [5], [9]. Inequality 1.1 allows to prove that the Pearson distance of the maximum likelihood estimator of the marginal density to the true density is $\sqrt{n}$- consistent under simple assumptions. We use it to derive the asymptotic expansion of the LRT statistic under contiguous alternatives for testing problems exhibiting a lack of identifiability of the alternative under the null hypothesis. The result applies in particular to the LRT for the number of populations with independent observations, so that it is possible to find the asymptotic power of the test under contiguous alternatives. We also propose self-normalized score tests which have asymptotic level 0 and power 1 by a use of recent results on self-normalized empirical processes obtained by Bercu, Gassiat and Rio [1].

2 Maximum-penalized marginal-likelihood estimation of the number of populations in a mixture with Markov regime.

Let $\mathcal{P} = \{p_\theta, \theta \in \Theta\}$ be a set of densities with respect to ν, Θ being a compact finite dimensional set. Let also $(Z_i)_{i \in \mathbb{N}}$ be a sequence of stationary random variables taking values in a finite subset of Θ such that, conditionally to (Z_i), the variables X_i are independent, with densities p_{Z_i}. Then, the marginal density of X_1 is a finite mixture of populations, that is

$$f = \sum_{i=1}^p \pi_i^0 p_{\theta_i^0}$$

where $\theta_1^0, \ldots, \theta_p^0$ are the possible values for the Z_i's, and $\pi_i^0 = \mathbb{P}(Z_1 = \theta_i^0)$. If $(Z_i)_{i \in \mathbb{N}}$ is a Markov chain, the sequence is said to be a finite mixture of populations with Markov regime, and the number of populations is the number of different states of the Markov chain having positive probability. Define for any integer p

$$\mathcal{G}_p = \{\sum_{i=1}^{p} \pi_i p_{\theta_i} \ : \ (\theta_1, \dots, \theta_p) \in \Theta^p, \ (\pi_1, \dots, \pi_p) \in [0,1]^p, \sum_{i=1}^{p} \pi_i = 1\}.$$

In this section, the set $\mathcal{G}$ of possible densities for the marginal distribution of the X_i's will be set to

$$\mathcal{G} = \bigcup_{p=1}^{P} \mathcal{G}_p = \mathcal{G}_P.$$

For any g in $\mathcal{G}$, define the number of populations as $p(g) = \min\{p \in \{1, \dots, P\} \ : \ g \in \mathcal{G}_p\}$, and let $p_0 = p(f)$. Define $\mathcal{S}$ as the set of functions s_g (see (2)) for g in $\mathcal{G}_P$.

We now define the maximum-penalized marginal-likelihood estimator of p_0, the number of populations, as a maximizer $\hat{p}$ of

$$T_n(p) = \max\{\ell_n(g) \ : \ g \in \mathcal{G}_p\} - a_n(p)$$

over $\{1, \dots, P\}$.
Let us introduce some assumptions.
(A1): $a_n(\cdot)$ is increasing, $a_n(p_1) - a_n(p_2)$ tends to infinity as n tends to infinity for any $p_1 > p_2$, and $a_n(p)/n$ tends to 0 as n tends to infinity for any p.
(A2): the parametrization $\theta \to p_\theta(x)$ is continuous for all x, and one can find a function m in $L^1(f d\nu)$ such that for any $g \in \mathcal{G}_P$, $| \log g | \leq m$.
(A3): (Z_i) is an irreducible Markov chain such that, if β is its mixing rate function, and $H_\beta(u)$ is the entropy with bracketing of $\mathcal{S}$ with respect to the norm $\| \cdot \|_{2,\beta}$ (see [4] for definitions), then

$$\int_0^1 \sqrt{H_\beta(u)} du < +\infty.$$

In [4] it is proved that $\| \cdot \|_{2,\beta} \geq \| \cdot \|_2$, so that under **(A3)**, $\mathcal{S}$ admits an envelope function F such that $\int F^2 f d\nu < +\infty$. Assumption **(A3)** also ensures the uniform tightness of $(\sum_{i=1}^n s(X_i))/\sqrt{n})$ for $s \in \mathcal{S}$. Indeed, it is proved by Doukhan, Massart and Rio [4] that under this assumption, $(\sum_{i=1}^n s(X_i))/\sqrt{n})_{s \in \mathcal{S}}$ satisfies a uniform central limit theorem.
We now have:

Theorem 2.1 *Under (A1), (A2) and (A3), $\hat{p}$ converges in probability to the number of populations p_0.*

Proof.

$$\begin{aligned}
\mathbb{P}(\hat{p} > p_0) &\leq \sum_{p=p_0+1}^{P} \mathbb{P}\left(T_n(p) \geq T_n(p_0)\right) \\
&= \sum_{p=p_0+1}^{P} \mathbb{P}\left(\sup_{g \in \mathcal{G}_p} (\ell_n(g) - \ell_n(f)) - \sup_{g \in \mathcal{G}_{p_0}} (\ell_n(g) - \ell_n(f)) \geq a_n(p) - a_n(p_0)\right) \\
&\leq \sum_{p=p_0+1}^{P} \mathbb{P}\left(\sup_{s \in \mathcal{S}} \frac{(\sum_{i=1}^n s(X_i))^2}{\sum_{i=1}^n s_-^2(X_i)} \geq a_n(p) - a_n(p_0)\right)
\end{aligned}$$

by applying inequality 1.2 and since $\sup_{g \in \mathcal{G}_{p_0}} (\ell_n(g) - \ell_n(f)) \geq 0$. Now, using Theorem 1 of [4], under (A3)

$$\sup_{s \in \mathcal{S}} \frac{1}{n} (\sum_{i=1}^{n} s(X_i))^2 = O_{\mathbb{P}}(1), \tag{3}$$

and since the usual norm in $L^2(fd\nu)$ is upper bounded by $\| \cdot \|_{2,\beta}$, $\mathcal{S}^2$ is $\mathbb{P}$-Glivenko-Cantelli (see [8]), so that in probability

$$\lim_{n\to+\infty} \inf_{s\in\mathcal{S}} \frac{1}{n} \sum_{i=1}^{n} s_-^2(X_i) = \inf_{s\in\mathcal{S}} \| s_- \|_2^2 . \tag{4}$$

But under the assumptions,

$$\inf_{s\in\mathcal{S}} \| s_- \|_2 > 0. \tag{5}$$

Indeed, if (5) does not hold, there exists a sequence s^n of functions in $\mathcal{S}$ such that $\| (s^n)_- \|_2$ converges to 0. This implies that $(s^n)_-$ converges to 0 in $L^1(fd\nu)$ and ν-a.s. for a subsequence. But $\int (s^n)_- d\nu = -\int (s^n)_+ d\nu$, with $s_+(x) = \max\{0, s(x)\}$, since all functions s in $\mathcal{S}$ verify $\int s f d\nu = 0$. Thus $(s^n)_+$ converges also to 0 in $L^1(fd\nu)$, and ν-a.s. for a subsequence. But under (A3), $\mathcal{S}$ admits a square integrable envelope function F, so that along a subsequence s^n converges ν-a.s. to 0, and in $L_2(fd\nu)$. This contradicts the fact that all functions s in $\mathcal{S}$ verify $\int s^2 f d\nu = 1$. Applying (3), (4), (5),

$$\sup_{s\in\mathcal{S}} \frac{(\sum_{i=1}^n s(X_i))^2}{\sum_{i=1}^n s_-^2(X_i)} = O_{\mathbb{P}}(1)$$

and $\mathbb{P}(\hat{p} > p_0)$ tends to 0 as n tends to infinity using (A1). Also,

$$\mathbb{P}\left(\hat{p} < p_0\right) \leq \sum_{p=1}^{p_0-1} \mathbb{P}\left(\sup_{g\in\mathcal{G}_p} \frac{\ell_n(g) - \ell_n(f)}{n} \geq \frac{a_n(p) - a_n(p_0)}{n} \right).$$

Under (A2), the set $\{\log(\frac{g}{f}), g \in \mathcal{G}_p\}$ is Glivenko-Cantelli in $f\nu$-probability, so that $\sup_{g\in\mathcal{G}_p} \frac{\ell_n(g)-\ell_n(f)}{n}$ converges in probability to

$$-\inf_{g\in\mathcal{G}_p} \int f \log \frac{f}{g},$$

which is negative using (A2) and the fact that $p < p_0$. Finally, $\hat{p}$ converges to p_0 in probability

3 Asymptotic power of the LRT test under loss of identifiability.

In this section, we assume that the X_i's are independent, and $\mathcal{G}$ is the set of possible densities of the X_i's, not necessarily mixtures of populations. Define the extended set of scores $\mathcal{S}$ as the set of functions s_g (see (2)) for g in $\mathcal{G}$. Define $H_{[],2}(u)$ as the entropy with bracketing of $\mathcal{S}$ with respect to the norm $\| \cdot \|_2$ in $L^2(fd\nu)$. Let us introduce the assumption **(B):**

$$\int_0^1 \sqrt{H_{[],2}(u)} du < +\infty.$$

Under **(B)**, the set $\mathcal{S}$ is Donsker.
Define now the set of scores $\mathcal{D}$ as the set of functions d in $L^2(fd\nu)$ such that one can find a sequence g_n in $\mathcal{G}$ satisfying $\| \frac{g_n - f}{f} \|_2 \to 0$ and $\| d - s_{g_n} \|_2 \to 0$.
With such a (g_n), define, for all $t \in [0,1]$, $g_t = g_n$, where $n \leq \frac{1}{t} < n+1$. We thus have that, for any $d \in \mathcal{D}$, there exists a parametric path $(g_t)_{0\leq t\leq\alpha}$ such that for any $t \in [0,\alpha]$, $g_t \in \mathcal{G}$, $t \to \| \frac{g_t - f}{f} \|_2$ is continuous, tends to 0 as t tends to 0, and $\| d - s_{g_t} \|_2 \to 0$ as t tends to 0.

We first prove that all densities in $1/\sqrt{n}$-neighborhoods of f leading to a score d on a parametric path define contiguous probabilities $\mathbb{P}_n$ for the sequence $X_1, \dots, X_n$. We refer the reader to [10] for the definition of contiguity and general results on likelihood ratios for contiguous measures. Using the reparametrization $\| \frac{g_u - f}{f} \|_2 = u$, for any $d \in \mathcal{D}$, there exists a parametric path $(g_u)_{0 \leq u \leq \alpha}$ such that

$$\int \left(\frac{g_u - f}{f} - ud \right)^2 f d\nu = o(u^2).$$

Now,

$$\begin{aligned}
\int \left(\sqrt{\frac{g_u}{f}} - 1 - \frac{u}{2} d \right)^2 f d\nu &= \int \left(\frac{\frac{g_u}{f} - 1 - ud}{2} - \frac{(\sqrt{\frac{g_u}{f}} - 1)^2}{2} \right)^2 f d\nu \\
&\leq \frac{1}{2} \int \left(\frac{g_u - f}{f} - ud \right)^2 f d\nu + \frac{1}{2} \int \left(\sqrt{\frac{g_u}{f}} - 1 \right)^4 f d\nu \\
&\leq o(u^2) + u^2 \int \left(\frac{\frac{g_u - f}{f}}{u} \right)^2 \left(\frac{\sqrt{\frac{g_u}{f}} - 1}{\sqrt{\frac{g_u}{f}} + 1} \right)^2 f d\nu.
\end{aligned}$$

Applying the dominated convergence theorem to the second term proves that the upper bound is $o(u^2)$ so that the parametric path $(g_u)_{0 \leq u \leq \alpha}$ is differentiable in quadratic mean, with score function d. As a consequence, any $g_{c/\sqrt{n}}$ along such a path defines a $\mathbb{P}_n = (g_{c/\sqrt{n}} \nu)^{\otimes n}$ mutually contiguous to $(f\nu)^{\otimes n}$. Fix such a $g_{c/\sqrt{n}}$, and let d_0 be the associated score. We now have the Theorem:

Theorem 3.1 *Under (B),*

$$\sup_{g \in \mathcal{G}} (\ell_n(g) - \ell_n(f)) = \frac{1}{2} \sup_{d \in \mathcal{D}} \left(\max\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n d(X_i); 0\} \right)^2 + o_{\mathbb{P}_n}(1).$$

Proof. Following the proof of Theorem 2.1, one can see that (3) and (4) hold again. Since $f \in \mathcal{G}$, one may apply inequality 1.1 to obtain

$$\sup_{g \in \mathcal{G}: \ell_n(g) - \ell_n(f) \geq 0} \| \frac{g - f}{f} \|_2 = O_{\mathbb{P}}(n^{-1/2}). \tag{6}$$

Taylor expansion gives that $\log(1+u) = u - \frac{u^2}{2} + u^2 R(u)$, with $\lim_{u \to 0} R(u) = 0$. Thus for any g,

$$\begin{aligned}
\ell_n(g) - \ell_n(f) &= \| \frac{g - f}{f} \|_2 \sum_{i=1}^n s_g(X_i) - \frac{1}{2} \| \frac{g - f}{f} \|_2^2 \sum_{i=1}^n (s_g(X_i))^2 \\
&\quad + \| \frac{g - f}{f} \|_2^2 \sum_{i=1}^n (s_g(X_i))^2 R(\| \frac{g - f}{f} \|_2 \, s_g(X_i)).
\end{aligned}$$

By (B), $\mathcal{S}$ admits a square integrable envelope function F, and

$$\max_{i=1,\dots,n} F(X_i) = o_{\mathbb{P}}(\sqrt{n}).$$

Moreover, under (B) $\frac{1}{n} \sum_{i=1}^n (s_g(X_i))^2 = O_{\mathbb{P}}(1)$. These facts, together with (6) lead to

$$\sup_{g \in \mathcal{G}: \ell_n(g) - \ell_n(f) \geq 0} \| \frac{g - f}{f} \|_2^2 \sum_{i=1}^n (s_g(X_i))^2 R(\| \frac{g - f}{f} \|_2 \, s_g(X_i)) = o_{\mathbb{P}}(1),$$

so that

$$\sup_{g\in\mathcal{G}}\left(\ell_n(g)-\ell_n(f)\right)=\sup_{g\in\mathcal{G}}\left\{\parallel\frac{g-f}{f}\parallel_2\sum_{i=1}^n s_g(X_i)-\frac{1}{2}\parallel\frac{g-f}{f}\parallel_2^2\sum_{i=1}^n(s_g(X_i))^2\right\}+o_{\mathbb{P}}(1),$$

which implies that

$$2\sup_{g\in\mathcal{G}}\left(\ell_n(g)-\ell_n(f)\right)\leq\sup_{g:\ell_n(g)-\ell_n(f)\geq 0}\frac{\left(\max\{\frac{\sum_{i=1}^n s_g(X_i)}{\sqrt{n}};0\}\right)^2}{\frac{\sum_{i=1}^n s_g^2(X_i)}{n}}+o_{\mathbb{P}}(1).$$

But under (B), $\mathcal{S}^2$ is Glivenko-Cantelli in probability so that

$$\sup_{g\in\mathcal{G}}\mid\frac{\sum_{i=1}^n s_g^2(X_i)}{n}-1\mid=o_{\mathbb{P}}(1),$$

and

$$2\sup_{g\in\mathcal{G}}\left(\ell_n(g)-\ell_n(f)\right)\leq\sup_{g:\ell_n(g)-\ell_n(f)\geq 0}\left(\max\{\frac{1}{\sqrt{n}}\sum_{i=1}^n s_g(X_i);0\}\right)^2+o_{\mathbb{P}}(1).$$

Let $\mathcal{G}_n=\{g\in\mathcal{G}:\parallel\frac{g-f}{f}\parallel_2\leq n^{-1/4}\}$. Using (6), we obtain that

$$2\sup_{g\in\mathcal{G}}\left(\ell_n(g)-\ell_n(f)\right)\leq\sup_{g\in\mathcal{G}_n}\left(\max\{\frac{1}{\sqrt{n}}\sum_{i=1}^n s_g(X_i);0\}\right)^2+o_{\mathbb{P}}(1)$$

Now, $\sup_{g\in\mathcal{G}_n}\parallel s_g-\mathcal{D}\parallel_2$ tends to 0 as n tends to infinity so that for a sequence u_n decreasing to 0, and with

$$\Delta_n=\{s_g-d\ :\ g\in\mathcal{G}_n,\ d\in\mathcal{D},\ \parallel s_g-d\parallel_2\leq u_n\},$$

we obtain that

$$2\sup_{g\in\mathcal{G}}\left(\ell_n(g)-\ell_n(f)\right)\leq\left(\max\{\sup_{d\in\mathcal{D}}\frac{1}{\sqrt{n}}\sum_{i=1}^n d(X_i)+\sup_{\delta\in\Delta_n}\frac{1}{\sqrt{n}}\sum_{i=1}^n\delta(X_i);0\}\right)^2+o_{\mathbb{P}}(1).$$

But using (B), the definition of Δ_n and the maximal inequality p.286 of [10],

$$\sup_{\delta\in\Delta_n}\frac{1}{\sqrt{n}}\sum_{i=1}^n\delta(X_i)=o_{\mathbb{P}}(1),$$

so that

$$2\sup_{g\in\mathcal{G}}\left(\ell_n(g)-\ell_n(f)\right)\leq\sup_{d\in\mathcal{D}}\left(\max\{\frac{1}{\sqrt{n}}\sum_{i=1}^n d(X_i);0\}\right)^2+o_{\mathbb{P}}(1)\qquad(7)$$

Moreover, using the differentiability in quadratic mean along the parametric paths, one obtains that for a sequence of finite subsets $\mathcal{D}_k$ increasing to $\mathcal{D}$, one has for any k

$$2\sup_{g\in\mathcal{G}}\left(\ell_n(g)-\ell_n(f)\right)\geq\sup_{d\in\mathcal{D}_k}\left(\max\{\frac{1}{\sqrt{n}}\sum_{i=1}^n d(X_i);0\}\right)^2+o_{\mathbb{P}}(1).$$

Therefore, equality holds in (7). Since $(f\nu)^{\otimes n}$ and $\mathbb{P}_n$ are mutually continguous, all $o_{\mathbb{P}}(1)$ are $o_{\mathbb{P}_n}(1)$.
Define $(W(d))_{d\in\mathcal{D}}$ the centered gaussian process with covariance the scalar product in $L^2(fd\nu)$. One has:

Corollary 3.1 *Assume that (B) holds. Under* $\mathbb{P}_n$, $\sup_{g\in\mathcal{G}}(\ell_n(g)-\ell_n(f)$ *converges in distribution to*

$$\frac{1}{2}\sup_{d\in\mathcal{D}}\left(\max\{W(d)+c\int dd_0 f d\nu;0\}\right)^2.$$

The Corollary follows from Theorem 3.1 by an application of Le Cam's third Lemma in metric spaces, see [11].
The result may be applied to mixtures of populations.

4 Self-normalized score tests.

We still assume that the variables X_i's are independent with common unknown density f. Let us now consider the testing problem of H_0: "$g = f$" against H_1: "$g \neq f$, $g \in \mathcal{G}$" for some particular density f. To investigate the asymptotic level and the asymptotic power of a test, one has to know about the (large or moderate) deviations of the testing statistic.
Let $\mathcal{S}$ be as in section 3, $v(n)$ some sequence tending to $+\infty$, and define the test function ϕ_n by $\phi_n = 1$ if and only if

$$\sup_{s\in\mathcal{S}}\frac{(\sum_{i=1}^n s(X_i))^2}{\sum_{i=1}^n s^2(X_i)} \geq v(n),$$

and $\phi_n = 0$ otherwise. Let α_n be the level of the test and $1-\beta_n(g)$ its power function.
Introduce the assumption **(C):** $\mathcal{D}$ is $f\nu$-Donsker, and for any positive δ, there exists a finite covering $(B_i)_{i\in I}$ of $\mathcal{S}$ such that, for any $i \in I$, one can find functions (l_i, u_i) satisfying for any $s \in B_i$, $l_i^2 \leq s^2 \leq u_i^2$, $su_i \geq 0$, $sl_i \geq 0$, and $\int(u_i - l_i)^2 f d\nu \leq \delta$.
One has:

Theorem 4.1 *Assume (C). Let* $v(n) \to +\infty$ *with* $v(n) = o(n)$. *Then for any* $g \in \mathcal{G}$ *such that* $g \neq f$, *one has*

$$\lim_{n\to+\infty}\frac{1}{v(n)}\log\alpha_n = -\frac{1}{2} \text{ and } \lim_{n\to+\infty} 1-\beta_n(g) = 1.$$

Proof. Under (C), the self normalized ratio $\sup_{s\in\mathcal{S}}\frac{\sum_{i=1}^n s(X_i)}{\sqrt{\sum_{i=1}^n s^2(X_i)}}$ obeys a moderate deviations principle by Theorem 3.1 of [1], and the result on α_n follows.
The result on $\beta_n(g)$ is a consequence of (C) and the fact that $v(n) \to +\infty$.
Remark. Assumption (C) holds in particular when $\mathcal{G}$ is the set of mixtures of gaussian densities with different means and same variance.

5 Proof of inequalities 1.1 and 1.2.

We have

$$\begin{aligned}\ell_n(g)-\ell_n(f) &= \sum_{i=1}^n \log\left(1+\|\frac{g-f}{f}\|_2\, s_g(X_i)\right)\\ &\leq \|\frac{g-f}{f}\|_2\sum_{i=1}^n s_g(X_i) - \frac{1}{2}\|\frac{g-f}{f}\|_2^2\sum_{i=1}^n (s_g)_-^2(X_i),\end{aligned}$$

since for any real number u, $\log(1+u) \leq u - \frac{1}{2}u_-^2$. As soon as $\ell_n(g)-\ell_n(f) \geq 0$,

$$\|\frac{g-f}{f}\|_2\sum_{i=1}^n s_g(X_i) \geq \frac{1}{2}\|\frac{g-f}{f}\|_2^2\sum_{i=1}^n (s_g)_-^2(X_i)$$

and inequality 1.1 follows.
Now, for any $g \in \mathcal{G}$,

$$\begin{aligned}
\ell_n(g) - \ell_n(f) &\leq \sup_{0 \leq p \leq \|\frac{g-f}{f}\|_2} \sum_{i=1}^n \log\left(1 + p s_g(X_i)\right) \\
&\leq \sup_{0 \leq p \leq \|\frac{g-f}{f}\|_2} \left[p \sum_{i=1}^n s_g(X_i) - \frac{p^2}{2} \sum_{i=1}^n (s_g)_-^2 (X_i) \right] \\
&\leq \sup_{p \geq 0} \left[p \sum_{i=1}^n s_g(X_i) - \frac{p^2}{2} \sum_{i=1}^n (s_g)_-^2 (X_i) \right] \\
&\leq \frac{1}{2} \frac{\left(\sum_{i=1}^n s_g(X_i)\right)^2}{\sum_{i=1}^n (s_g)_-^2 (X_i)}
\end{aligned}$$

and inequality 1.2 follows.

Références

[1] Bercu, B., Gassiat, E. and Rio, E. Concentration inequality, large and moderate deviations for self-normalized empirical processes. *Annals of Probability*, to appear.

[2] Dacunha-Castelle, D. and Gassiat, E. Testing in locally conic models. *ESAIM Prob. and stat.*, 1, 1997.

[3] Dacunha-Castelle, D. and Gassiat, E. Testing the order of a model using locally conic parametrization: population mixtures and stationary ARMA processes. *Annals of Statistics*, 27, 4 1178-1209, 1999.

[4] Doukhan, P., Massart, P. and Rio, E. Invariance principles for absolutely regular empirical processes. *Ann. Inst. Henri Poincaré*, 31, 2, 393-427, 1995.

[5] Leroux, B.G. Consistent estimation of a mixing distribution. *Annals of Statistics*, 20, 1350-1360, 1992.

[6] Leroux, B.G. and Puterman, M.L. Maximum-penalized likelihood estimation for independent and Markov dependent mixture models. *Biometrics*, 48, 545-558, 1992.

[7] Liu, X. and Shao, Y. Asymptotics of the likelihood ratio under loss of identifiability with applications to finite mixture models. *Preprint*, 2001.

[8] Peskir, G. and Weber, M. Necessary and sufficient conditions for the uniform law of large numbers in the stationary case. *Functional analysis IV*, Budkovic, D. (ed) et al. Proceedings of the postgraduate school and conference held at the Inter-university Center, Dubrovnic, Croatia, 1993. Mat. Institut, Var. Publ. Ser., Aarhus Univ. 43, 165-190, 1994.

[9] Ryden, T. Estimating the order of hidden Markov models. *Statistics*, 26, 345-354, 1995.

[10] Van der Vaart, A. *Asymptotic statistics.* Cambridge University Press, 1998.

[11] Van der Vaart, A. and Wellner, J. *Weak convergence and empirical processes.* Springer, New-York, 1996.