JOURNAL DE LA SOCIÉTÉ FRANÇAISE DE STATISTIQUE

MICHAEL NUSSBAUM

Équivalence asymptotique des expériences statistiques

Journal de la société française de statistique, tome 145, n° 1 (2004), p. 31-45

<http://www.numdam.org/item?id=JSFS_2004__145_1_31_0>

© Société française de statistique, 2004, tous droits réservés.

L'accès aux archives de la revue « Journal de la société française de statistique » (http://publications-sfds.math.cnrs.fr/index.php/J-SFdS) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

ÉQUIVALENCE ASYMPTOTIQUE DES EXPÉRIENCES STATISTIQUES

CONFÉRENCE LUCIEN LE CAM 2003

Michael NUSSBAUM *

RÉSUMÉ

L'idée d'approcher une suite d'expériences statistiques par une famille gaussienne remonte à Wald (1943), mais doit son développement à Lucien Le Cam, qui introduisit le terme « normalité asymptotique locale ». Ce cadre théorique est devenu un standard pour la résolution du problème de l'efficacité des tests et des estimateurs (dont l'estimateur de maximum de vraisemblance) au sens asymptotique. Il suffit d'observer que le modèle de départ est approximativement normal et hérite donc, toujours approximativement, de la structure simple des modèles normaux. Ce passage à la limite au sens du modèle est plus fort et plus riche en conséquences que les résultats portant sur les lois limites des diverses fonctions individuelles de l'échantillon, et qui découlent de l'application du théorème de la limite centrale.

ABSTRACT

The idea of approximating a sequence of statistical experiments by a gaussian family goes back to Wald (1943), but has been fully developed by Lucien Le Cam, who introduced the term "local asymptotic normality". This theoretical framework has become a standard tool for proving efficiency of tests and estimators (in particular of the maximum likelihood estimator) in an asymptotic sense. It suffices to note that the initial model is approximately normal and thus inherits, in an asymptotic sense, the simple structure of normal models. The passage to the limit in the sense of the whole model is stronger and richer in consequences than the results on limit laws of various individual functions of the sample which are consequences of the central limit theorem.

1. Définitions de base et premiers exemples

Dans cet article nous essaierons de donner une introduction élémentaire à la théorie asymptotique des expériences statistiques, une théorie intimement associée au nom de Lucien Le Cam (voir les ouvrages [15], [16]). De plus, on discutera les développements liés en statistique non paramétrique qui ont permis récemment d'élargir le champ des applications de cette théorie.

La structure fondamentale en statistique est une famille de lois $(P_\vartheta,\ \vartheta \in \Theta)$ sur un espace mesurable $(\Omega, \mathcal{A})$. Cette structure peut être appelée modèle

* Research supported by the National Science Foundation under grant DMS0306497.
Department of Mathematics, Malott Hall, Cornell University, Ithaca, NY 14853-4201, USA.
E-mail : nussbaum@math.cornell.edu

statistique; mais la terminologie *expérience statistique* a été adoptée. Une expérience $\mathcal{P}$ est définie comme un ensemble

$$\mathcal{P} = (\Omega, \mathcal{A}, (P_\vartheta, \vartheta \in \Theta)).$$

On observe une variable aléatoire X à valeurs dans Ω ayant loi P_ϑ ; la valeur ϑ du paramètre est inconnue. Puisque l'espace mesurable est déjà sous-jacent à la définition d'une loi P_ϑ, on écrit aussi $\mathcal{P} = (P_\vartheta, \vartheta \in \Theta)$.

Voici une citation directe de L. Le Cam qui décrit assez bien le point de départ (voir [13]) : *En général, la famille $\mathcal{P}$ est compliquée. On voudrait alors l'approcher par une famille plus simple.*

La méthode de choix pour obtenir une famille «plus simple» est fournie par les statistiques exhaustives. Ce sont elles qui permettent en général de réduire la dimension des données par une transformation préalable, sans perte d'information. Ainsi commençons avec un des exemples les plus élémentaires dans lesquels existe une statistique exhaustive. Dans la suite, un n-échantillon est un vecteur d'observations indépendantes et équidistribuées.

EXEMPLE 1 **n-échantillon normal, problème de position.** — *Soient X_i, $i = 1, \ldots, n$ indépendantes $N(\vartheta, 1)$. La moyenne empirique $\bar{X}_n$ est une statistique exhaustive. La loi de $\bar{X}_n$ est $\mathcal{L}(\bar{X}_n) = N(\vartheta, n^{-1})$.*

Dans ce cas, l'expérience initiale est donnée par la famille des lois jointes du n-échantillon, ϑ étant inconnu ($\vartheta \in \Theta$ disons, avec $\Theta \subset \mathbb{R}$). Si son observation se limite à la valeur de la statistique exhaustive $\bar{X}_n$, le statisticien se trouve devant la famille des lois $N(\vartheta, n^{-1})$ de celle-ci, qui, de même que la famille des lois de l'échantillon tout entier, dépend du paramètre inconnu ϑ. Ainsi on obtient deux expériences indexées par le même ϑ. Si l'on admet qu'une statistique exhaustive contient toute l'information disponible sur ϑ, alors on est amené à la notion intuitive d'expériences équivalentes.

PROPOSITION 1. — *Supposons que $\Theta \subset \mathbb{R}$. Alors les expériences données par les observations*

$$X_i = \vartheta + \xi_i,\ i = 1, \ldots, n,\ \xi_i \sim N(0,1),\ \text{indépendantes},\ \vartheta \in \Theta$$
$$Y = \vartheta + n^{-1/2}\xi,\ \ \xi \sim N(0,1),\ \vartheta \in \Theta$$

sont équivalentes.

La définition exacte de l'équivalence sera donnée plus tard. Pour le moment, nous nous contentons de l'idée imprécise que les deux expériences contiennent la même information sur ϑ.

EXEMPLE 2 **n-échantillon normal, problème d'échelle.** — *Soient X_i, $i = 1, \ldots, n$ indépendantes $N(0, \sigma^2)$. La variance empirique $S_n^2 = n^{-1}\Sigma_{i=1}^n X_i^2$ est une statistique exhaustive. La loi de S_n^2 est celle de $n^{-1}\sigma^2\chi_n^2$, où χ_n^2 est une variable aléatoire de loi chi-carré à n degrés de liberté.*

Supposons que le paramètre inconnu σ^2 satisfasse à $\sigma^2 \in \Theta$, où $\Theta \subset (0, \infty)$. Ici la réduction de la dimension du modèle a eu lieu, mais on pourrait essayer de pousser plus loin la simplification, en se demandant si la famille des lois $\mathcal{L}(n^{-1}\sigma^2\chi_n^2)$, $\sigma^2 \in \Theta$ ne peut pas faire l'objet d'une simplication supplémentaire grâce au théorème de la limite centrale. En effet, on a

$$\sqrt{n}(S_n^2 - \sigma^2) \stackrel{\mathcal{L}}{\Longrightarrow} N(0, 2\sigma^4). \tag{1}$$

Si la loi d'une statistique exhaustive a une loi limite normale, comme c'est le cas ici, on serait tenté de conclure qu'aussi la famille des lois (l'expérience initiale) converge vers une expérience gaussienne. On voudrait réécrire (1) sous la forme

$$S_n^2 \approx N\left(\sigma^2, 2\sigma^4 n^{-1}\right) \tag{2}$$

sans pour autant préciser la relation $\approx$, et l'expérience limite serait la famille

$$\left(N\left(\sigma^2, 2\sigma^4\right), \sigma^2 \in \Theta\right). \tag{3}$$

Toutefois, étant donné que la convergence en loi est une convergence faible, on a besoin d'une version plus forte de (1). Rappelons que la distance de la variation totale $\|\cdot\|_{TV}$ des lois P, Q est définie comme

$$\|P - Q\|_{TV} = 2 \sup_{A \text{ mesurable}} |P(A) - Q(A)| = \int |p - q| \, d\mu$$

où p, q sont les densités par rapport à μ de P et Q. Il est bien connu que, sous certaines conditions de régularité, le théorème de la limite centrale s'applique au sens plus fort de la variation totale (voir van der Vaart [22], 2.31). Essayons d'en expliquer la raison. Dans notre cas la loi χ_n^2 est continue et très régulière, donc la loi de $\sqrt{n}(S_n^2 - \sigma^2)$ a une densité. La loi elle-même tendant vers la loi $N(0, 2\sigma^4)$, il serait naturel que la densité converge aussi. Pour les densités, le théorème de la limite centrale est connu sous le nom de théorème de la limite locale. Si ce théorème s'applique point par point pour une suite de densités, il s'ensuit par le Lemme de Scheffé que pour les deux densités ($p_{\sigma,n}$ et q_σ, disons) on a

$$\int |p_{\sigma,n} - q_\sigma| \to 0, \text{ si } n \to \infty,$$

et donc, pour la distance de la variation totale,

$$\left\|\mathcal{L}(\sqrt{n}(S_n^2 - \sigma^2)) - N(0, 2\sigma^4)\right\|_{TV} \to 0, \text{ si } n \to \infty.$$

L'application $x \mapsto n^{-1/2}x + \sigma^2$ est mesurable et bijective. Par conséquent,

$$\left\|\mathcal{L}(S_n^2) - N(\sigma^2, 2n^{-1}\sigma^4)\right\|_{TV} \to 0, \text{ si } n \to \infty.$$

Par un argument supplémentaire, on déduit que cette convergence est uniforme sur tout ensemble de paramètres de la forme $\sigma^2 \in \Theta \subset (a, b)$, $a > 0$.

La loi de la statistique exhaustive converge donc en variation totale, uniformément. Puisque l'approximation normale peut être faite pour les probabilités de tous les événements, uniformément en le paramètre inconnu, il est légitime de qualifier d'asymptotiquement normale l'expérience initiale.

Formulons donc un constat d'équivalence asymptotique : il s'agit d'une comparaison d'expériences censées contenir asymptotiquement, lorsque $n \to \infty$, le même montant d'information sur le paramètre inconnu σ^2. Evidemment, si on veut clarifier le sens précis de ce qui précède, ce sera en liaison étroite avec un concept à définir d'exhaustivité asymptotique.

PROPOSITION 2. — *Supposons que* $\sigma^2 \in \Theta \subset (a, b)$, $a > 0$. *Alors les expériences données par les observations*

$$X_i = \sigma\xi_i,\ i = 1, \ldots, n,\ \xi_i \sim N(0,1),\ \text{indépendantes},\ \sigma^2 \in \Theta$$
$$Y = \sigma^2 + n^{-1/2}\sqrt{2}\sigma^2\xi,\quad \xi \sim N(0,1),\ \sigma^2 \in \Theta$$

sont asymptotiquement équivalentes.

Étant donné que l'on désire l'approximation la plus simple de l'expérience initiale, le deuxième modèle ci-dessous n'est pas encore entièrement satisfaisant. Bien qu'elle soit normale d'espérance σ^2, la deuxième expérience est hétéroscédastique, c'est-à-dire que la variance aussi y dépend de σ^2. Pour tous les problèmes d'inférence statistique, un modèle gaussien homoscédastique serait préférable. Pour obtenir l'homoscédasticité, on se sert d'une procédure de stabilisation de la variance. Rappelons le principe de cette idée bien connue en statistique. Revenons à la relation (1) et notons que pour une fonction régulière g (deux fois dérivable, disons), celle-ci entraîne

$$\sqrt{n}(g(S_n^2) - g(\sigma^2)) \overset{\mathcal{L}}{\Longrightarrow} N(0, 2\sigma^4(g'(\sigma^2))^2). \tag{4}$$

La fonction $g(x) = \log x$ a pour dérivée $g'(x) = x^{-1}$; on obtient donc

$$\sqrt{n}(\log S_n^2 - \log\sigma^2)) \overset{\mathcal{L}}{\Longrightarrow} N(0,2),$$

et on pourrait réécrire (2) sous la forme

$$\log S_n^2 \approx N\left(\log\sigma^2, 2n^{-1}\right). \tag{5}$$

Dans ce cas, la transformation $S_n^2 \mapsto \log(S_n^2)$ s'avère stabilisante pour la variance. Ce raisonnement simple est valable pour la convergence en loi ; afin de le justifier au niveau des expériences, on doit utiliser la convergence en variation totale.

PROPOSITION 3. — *Supposons que* $\sigma^2 \in \Theta \subset (a, b)$, $a > 0$. *Alors les expériences données par les observations*

$$X_i = \sigma\xi_i,\ i = 1, \ldots, n,\ \xi_i \sim N(0,1),\ \text{indépendantes},\ \sigma^2 \in \Theta$$
$$Y = \log\sigma^2 + n^{-1/2}\sqrt{2}\xi,\quad \xi \sim N(0,1),\ \sigma^2 \in \Theta$$

sont asymptotiquement équivalentes.

Avec la deuxième expérience ci-dessus on a poussé au plus loin la simplification, en obtenant un simple modèle gaussien de translation. Mais le désavantage maintenant est que le paramètre inconnu (la moyenne) n'est plus le σ^2 originel mais la transformée $\log \sigma^2$, ce qui rend plus compliquées les procédures statistiques telles que l'estimation de σ^2. Toutefois, cette réduction des modèles n'est pas dépourvue d'intérêt, et elle est le sujet principal de cet exposé.

EXEMPLE 3 ***n*-échantillon de Poisson.** — *Soient* X_i, $i = 1, \ldots, n$ *indépendantes, de loi de Poisson* $\mathrm{Po}(\vartheta)$. *Ici encore, la moyenne empirique* $\bar{X}_n$ *est une statistique exhaustive.*

Par le théorème de la limite centrale on a

$$\sqrt{n}(\bar{X}_n - \vartheta) \stackrel{\mathcal{L}}{\Longrightarrow} N(0, \vartheta). \tag{6}$$

Ici toutefois la convergence en variation totale n'a pas lieu puisque la loi $\mathcal{L}(\bar{X}_n)$ est discrète (on a $\mathcal{L}(n\bar{X}_n) = \mathrm{Po}(n\vartheta)$). Mais d'après ce qu'on sait sur les théorèmes limites, la convergence en variation totale pourrait bien avoir lieu après un lissage adéquat. Cette idée raisonnable nous conduit à réexaminer le concept de l'équivalence statistique.

Rappelons l'équivalence de l'Exemple 1 (n-échantillon gaussien, par statistique exhaustive) : posons $P_{n,\vartheta} = \mathcal{L}(X_1, \ldots, X_n)$ et $Q_{n,\vartheta} = \mathcal{L}(\bar{X}_n) = N(0, n^{-1}\vartheta)$. De plus, soit $P_{n,\vartheta}(\cdot|\bar{X}_n = x)$ la loi conditionnelle de $(X_1, \ldots, X_n)$ étant donné $\bar{X}_n$. D'après la définition d'une statistique exhaustive, $P_{n,\vartheta}(\cdot|\bar{X}_n = x)$ ne dépend pas du paramètre ϑ, c'est-à-dire

$$P_{n,\vartheta}(A|\bar{X}_n = x) = P_{n,\cdot}(A|\bar{X}_n = x) \tag{7}$$

pour tout Borélien A de $\mathbb{R}^n$. Mais indépendamment de l'exhaustivité, dans cette situation très régulière la loi conditionnelle peut être choisie comme le noyau Markovien (fonction de transition)

$$K_\vartheta(A, x) = P_{n,\vartheta}(A|\bar{X}_n = x)$$

permettant de reconstituer la loi $P_{n,\vartheta}$ à partir de la loi $Q_{n,\vartheta}$: si l'opération $K_\vartheta : Q_{n,\vartheta} \mapsto K_\vartheta Q_{n,\vartheta}$ est définie par

$$K_\vartheta Q_{n,\vartheta}(A) = \int K_\vartheta(A, x) Q_{n,\vartheta}(dx) = \int P_{n,\vartheta}(A|\bar{X}_n = x) Q_{n,\vartheta}(dx), \quad A \in \mathcal{A},$$

alors on a $K_\vartheta Q_{n,\vartheta} = P_{n,\vartheta}$. L'exhaustivité (7) implique que le noyau K_ϑ peut être choisi indépendamment de ϑ ; donc il existe un noyau Markovien K tel que

$$KQ_{n,\vartheta} = P_{n,\vartheta} \text{ pour tout } \vartheta \in \Theta. \tag{8}$$

Pour ce qui est de la relation inverse, il est évident qu'il existe un autre noyau Markovien K' qui permet d'obtenir $Q_{n,\vartheta}$ à partir de $P_{n,\vartheta}$. Posons

$X = (X_1, \ldots, X_n)$ et considérons l'application non-aléatoire $t(X) = \bar{X}_n$; le noyau Markovien trivial défini par

$$K'(B, x) = \mathbf{1}_B(t(x))$$

pour tout Borélien B de $\mathbb{R}$ est celui qui satisfait à

$$K' P_{n,\vartheta}(B) = \int \mathbf{1}_B(t(x)) P_{n,\vartheta}(dx) = \int_B Q_{n,\vartheta}(dx) = Q_{n,\vartheta}(B).$$

On a donc

$$K' P_{n,\vartheta} = Q_{n,\vartheta} \text{ pour tout } \vartheta \in \Theta. \qquad (9)$$

Les deux relations (8), (9) qui sont satisfaites dans le cadre d'une statistique exhaustive inspirent la définition suivante.

DÉFINITION 1 (**Δ-écart de Le Cam**). — *Soient* $\mathcal{P} = (P_\vartheta, \vartheta \in \Theta)$ *et* $\mathcal{Q} = (Q_\vartheta, \vartheta \in \Theta)$ *deux expériences, indexées par le même ensemble d'indices* Θ, *mais dont les espaces d'observations peuvent être distincts. La déficience (le défaut) de* $\mathcal{P}$ *par rapport à* $\mathcal{Q}$ *est*

$$\delta(\mathcal{P}, \mathcal{Q}) = \inf_K \sup_{\vartheta \in \Theta} \| Q_\vartheta - K P_\vartheta \|_{TV}$$

(inf *sur tous les noyaux Markoviens), et le* Δ*-écart est*

$$\Delta(\mathcal{P}, \mathcal{Q}) = \max(\delta(\mathcal{P}, \mathcal{Q}), \delta(\mathcal{Q}, \mathcal{P})).$$

Il s'agit ici d'une définition simplifiée, qui est valable sous les conditions de régularité suivantes. Pour chacune des deux expériences, $\mathcal{P} = (\Omega_\mathcal{P}, \mathcal{A}_\mathcal{P}, (P_\vartheta, \vartheta \in \Theta))$ (resp., $\mathcal{Q} = (\Omega_\mathcal{Q}, \mathcal{A}_\mathcal{Q}, (Q_\vartheta, \vartheta \in \Theta))$), l'espace des observations $\Omega_\mathcal{P}$ (resp., $\Omega_\mathcal{Q}$) est un espace polonais (espace métrique séparable et complet) et $\mathcal{A}_\mathcal{P}$ (resp., $\mathcal{A}_\mathcal{Q}$) est la tribu Borélienne correspondante. En outre, la famille $(P_\vartheta, \vartheta \in \Theta)$ (resp., $(Q_\vartheta, \vartheta \in \Theta)$) est dominée par une mesure sigma-finie. Si ces conditions ne sont pas remplies, la définition fait intervenir des objets plus abstraits comme des noyaux Markoviens généralisés (voir Le Cam [15], Chap. 2.3 ou van der Vaart [23], Chap. 8).

Dans les exemples 1-3, si $\mathcal{P} = (P_\vartheta, \vartheta \in \Theta)$ est la famille d'origine, T est une statistique exhaustive et $\mathcal{Q} = (Q_\vartheta, \vartheta \in \Theta)$ est la famille $Q_\vartheta = \mathcal{L}(T|P_\vartheta)$ alors on a $\Delta(\mathcal{P}, \mathcal{Q}) = 0$. La dernière relation est interprétée comme l'équivalence (au sens strict) des expériences.

DÉFINITION 2. — *Supposons que* $\mathcal{P}_n = (P_{n,\vartheta}, \vartheta \in \Theta)$ *et* $\mathcal{Q}_n = (Q_{n,\vartheta}, \vartheta \in \Theta)$ *sont deux suites d'expériences, indexées par* n *et telles que l'espace des observations peut aussi dépendre de* n. *Les suites* $\mathcal{P}_n, \mathcal{Q}_n$ *sont asymptotiquement équivalentes si* $\Delta(\mathcal{P}_n, \mathcal{Q}_n) \to 0$.

Pour éclaircir la signification statistique de la déficience, rappelons le cadre classique de la théorie de la décision. Pour un espace de paramètres Θ et un

espace mesurable de décisions $(E, \mathcal{E})$, soit $W : E \times \Theta \mapsto [0, \infty)$ une fonction de perte telle que pour tout $\vartheta \in \Theta$, $W(\vartheta, \cdot)$ soit mesurable par rapport à $(E, \mathcal{E})$. Une règle de décision randomisée dans l'expérience $\mathcal{P} = (\Omega, \mathcal{A}, (P_\vartheta, \vartheta \in \Theta))$ est un noyau Markovien $t(\cdot, \omega)$ qui à chaque $\omega \in \Omega$ associe une mesure de probabilité sur $(E, \mathcal{E})$. Le risque de t en ϑ est défini par

$$r_t(\vartheta) = \int W(e, \vartheta) t(de, \omega) P_\vartheta(d\omega).$$

PROPOSITION 4 ***(Caractérisation de la déficience).*** — *Deux expériences $\mathcal{P}, \mathcal{Q}$ vérifient la relation $\delta(\mathcal{P}, \mathcal{Q}) \leqslant \varepsilon$ si et seulement si : pour tout $\varepsilon > 0$, pour tout problème de décision avec fonction de perte W telle que $0 \leqslant W \leqslant 1$, pour toute fonction de décision t disponible dans $\mathcal{Q}$, il existe une fonction de décision t^* disponible dans $\mathcal{P}$ telle que*

$$r_{t^*}(\vartheta) \leqslant r_t(\vartheta) + \varepsilon, \quad \vartheta \in \Theta$$

(la règle de décision t^* est aussi bonne que t, à ε près).

Soulignons que cette caractérisation concerne la déficience $\delta(\mathcal{P}, \mathcal{Q})$ de $\mathcal{P}$ par rapport à $\mathcal{Q}$, et que δ n'est pas symétrique. Par conséquent, si pour l'expression symétrisée (le Δ-écart) on a $\Delta(\mathcal{P}, \mathcal{Q}) \leqslant \varepsilon$, alors les risques disponibles dans $\mathcal{Q}$ sont aussi disponibles dans $\mathcal{P}$, à ε près, et vice versa.

Revenons maintenant à l'Exemple 3, où la loi de la statistique exhaustive $\tilde{X}_n = n\bar{X}_n = \sum_{i=1}^n X_i$ est une loi de Poisson $\mathrm{Po}(n\vartheta)$. D'abord, de la convergence (6), on voudrait déduire une approximation par l'expérience gaussienne

$$\bar{X}_n \approx N\left(\vartheta, n^{-1}\vartheta\right). \tag{10}$$

Mais cette expérience est hétéroscédastique. Ici la fonction $g(x) = x^{1/2}$ s'avère stabilisante pour la variance, puisque, par analogie avec (4), on obtient $g'(x) = (4x)^{-1/2}$ et par conséquent

$$\sqrt{n}(g(\bar{X}_n) - g(\vartheta)) \xrightarrow{\mathcal{L}} N(0, \vartheta(g'(\vartheta))^2) = N(0, 1/4).$$

De là on pourrait réécrire (10) sous la forme

$$\bar{X}_n^{1/2} \approx N\left(\vartheta^{1/2}, (4n)^{-1}\right), \tag{11}$$

et obtenir ainsi une approximation gaussienne simple, qui reste à établir au sens fort moyennant des noyaux Markoviens. La méthode la plus élégante pour cela a été récemment obtenue dans [2], Théorème 4. Soit U une variable aléatoire uniforme sur $[-1/2, 1/2)$, indépendante de $\tilde{X}_n$. On pose

$$Z_n = \operatorname{sgn}\left(\tilde{X}_n + U\right)\sqrt{\left|\tilde{X}_n + U\right|} \tag{12}$$

et on montre que

$$\left\|\mathcal{L}(Z_n) - N((n\vartheta)^{1/2}, 1/4)\right\|_{TV} \leqslant C(n\vartheta)^{-1/2}$$

où C ne dépend pas de ϑ et de n. Ici, le noyau Markovien est représenté par l'opération de « lissage » $\tilde{X}_n \mapsto \tilde{X}_n + U$ qui est aussi inversible : la valeur de $\tilde{X}_n + U$ identifie celle de $\tilde{X}_n$ qui prend des valeurs entières. Par conséquent, on a obtenu à la fois les deux noyaux K, K' qui font converger vers 0 les deux déficiences $\delta(\mathcal{P}, \mathcal{Q})$ et $\delta(\mathcal{Q}, \mathcal{P})$).

PROPOSITION 5. — *Supposons que $\vartheta \in \Theta \subset (a, b)$, $a > 0$. Alors les expériences données par les observations*

$$X_i,\ i = 1, \ldots, n,\ X_i \sim \mathrm{Po}(\vartheta),\ \textit{indépendantes},\ \vartheta \in \Theta$$
$$Y = \vartheta^{1/2} + \frac{1}{2} n^{-1/2} \xi, \quad \xi \sim N(0,1),\ \vartheta \in \Theta$$

sont asymptotiquement équivalentes.

De même que dans la Proposition 3, l'approximation est obtenue par une expérience gaussienne de translation, mais le paramètre de celle-ci est $\vartheta^{1/2}$.

2. Modèle paramétrique : normalité asymptotique locale

Soit $\mathcal{P}_n = (P_{n,\vartheta},\ \vartheta \in \Theta)$ une suite de modèles paramétriques réguliers, $\Theta \subset \mathbb{R}^k$, engendrée par n variables indépendantes équidistribuées, et dans laquelle $P_{n,\vartheta}$ est donc une loi produit P_ϑ^n. Supposons que l'estimateur de maximum de vraisemblance $\hat{\vartheta}_n$ satisfasse à

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) \overset{\mathcal{L}}{\Longrightarrow} N_k(0, J_\vartheta^{-1}),$$

où J_ϑ est la matrice d'information de Fisher au point ϑ. Souvent l'estimateur de maximum de vraisemblance est aussi une statistique exhaustive, ou du moins est exhaustive au sens asymptotique. D'après ce qu'on a vu (*cf.* (2), (3), (10)), on serait tenté de chercher une approximation de l'expérience $\mathcal{P}_n$ par la famille

$$\left(N_k(\vartheta, n^{-1} J_\vartheta^{-1}), \vartheta \in \Theta\right) \tag{13}$$

qui est une expérience gaussienne hétéroscédastique, généralisant celle de la Proposition 2 (pour Y). Cette idée a été développée par Le Cam, qui s'est aussi rendu compte que la suite approchante (13) n'est pas attrayante du point de vue de la théorie de la décision. En effet, la présence du paramètre ϑ dans la matrice de covariance et la structure essentiellement arbitraire de J_ϑ^{-1}, $\vartheta \in \Theta$ ne permettent pas de considérer (13) comme une simplification.

2.1. La méthode locale

Une approximation plus prometteuse est fournie par la méthode locale. On considère une suite d'expériences restreintes où le paramètre ϑ ne varie que dans un voisinage $\Theta_n(\vartheta_0)$ d'une valeur ϑ_0 connue, et où le diamètre du voisinage est de l'ordre de $n^{-1/2}$. Une telle restriction peut être justifiée par deux arguments : premièrement, pour chaque règle de décision, le risque sur l'ensemble restreint de paramètres $\Theta_n(\vartheta_0)$ fournit des bornes inférieures pour l'expérience non restreinte. De plus, ces bornes inférieures sont souvent raisonnables car le ϑ_0, présumé connu d'abord, peut être remplacé par un estimateur préalable qui identifierait le vrai ϑ avec une précision de l'ordre de $n^{-1/2}$. Choisissons donc $\vartheta_0 \in \Theta$ (centre de localisation) et posons $\vartheta = \vartheta_0 + n^{-1/2}h$, où h est un paramètre local. Nous avons alors pour tout compact $K \subset \mathbb{R}^k$

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta_0) \stackrel{\mathcal{L}}{\Longrightarrow} N(h, J_{\vartheta_0}^{-1}),\ h \in K\ . \tag{14}$$

Ici, si ϑ_0 est présumé connu et $\hat{\vartheta}_n$ est exhaustif, on a de nouveau une suite de statistiques exhaustives qui converge en loi vers la famille

$$\left(N(h, J_{\vartheta_0}^{-1}),\ h \in \mathbb{R}^k\right). \tag{15}$$

Une telle famille gaussienne de translation offre tous les avantages de la simplicité. Pour obtenir une borne inférieure du risque de l'estimation de ϑ, on effectue un changement de paramètre parallèle : pour le risque quadratique et un estimateur $\tilde{\vartheta}_n$ quelconque de ϑ on a

$$\begin{aligned} n\,E_\vartheta\left(\tilde{\vartheta}_n - \vartheta\right)^2 &= n\,E_\vartheta\left(\tilde{\vartheta}_n - \vartheta_0 + n^{-1/2}h\right)^2 = E_\vartheta\left(n^{1/2}(\tilde{\vartheta}_n - \vartheta_0) - h\right)^2 \\ &= E_\vartheta\left(\tilde{h}_n - h\right)^2 \end{aligned} \tag{16}$$

où on a posé $\tilde{h}_n = n^{1/2}(\tilde{\vartheta}_n - \vartheta_0)$; $\tilde{h}_n$ s'interprète comme un estimateur de h. Il reste à démontrer de façon rigoureuse la convergence de l'expérience restreinte et localisée vers la limite gaussienne (15). Pour cela, Le Cam a développé une méthode directe à partir de la convergence faible (14), par des noyaux de Markov qui s'apparentent au lissage (12) (voir [15], section 11.8 , à comparer aussi Müller [17]).

2.2. Le processus de vraisemblance

Une autre méthode, beaucoup plus efficace, repose sur le lien profond entre l'équivalence des expériences et leurs processus de vraisemblance. Ce dernier est défini de la façon suivante. Soit $\mathcal{P} = (P_\vartheta, \vartheta \in \Theta)$ une famille sur $(\Omega, \mathcal{A})$, dominée par un de ses éléments P_{ϑ_0}, où $\vartheta_0 \in \Theta$. La densité $\Lambda(\vartheta)(\omega) = dP_\vartheta/dP_{\vartheta_0}(\omega)$ engendre une variable aléatoire $\Lambda(\vartheta)$ si l'argument ω suit la loi $\mathcal{L}(\omega) = P_{\vartheta_0}$. L'ensemble de variables aléatoires

$$\Lambda_\mathcal{P} = (\Lambda(\vartheta),\ \vartheta \in \Theta)$$

toutes définies sur l'espace de probabilité $(\Omega, \mathcal{A}, P_{\vartheta_0})$ forme un processus stochastique indexé par ϑ, le processus de vraisemblance de l'expérience $\mathcal{P}$. La loi $\mathcal{L}(\Lambda_{\mathcal{P}})$ de ce processus est l'ensemble des lois marginales finies. Un autre résultat clé de Le Cam est : si $\Lambda_{\mathcal{P}}$ et $\Lambda_{\mathcal{Q}}$ sont des processus de vraisemblance associés aux expériences $\mathcal{P}, \mathcal{Q}$, alors on a

$$\Delta(\mathcal{P}, \mathcal{Q}) = 0 \text{ si et seulement si } \mathcal{L}(\Lambda_{\mathcal{P}}) = \mathcal{L}(\Lambda_{\mathcal{Q}}). \tag{17}$$

Pour expliquer ce résultat au niveau heuristique, notons d'abord que le processus $\Lambda_{\mathcal{P}}$ représente une statistique exhaustive. Plus précisément, si ω sont les données, on définit une statistique $T(\omega)$ à valeurs dans un espace de haute dimension comme l'ensemble

$$T(\omega) = (\Lambda(\vartheta)(\omega), \vartheta \in \Theta)$$

(ceci prend des valeurs dans l'espace $\mathbb{R}^{\Theta}$). Selon le critère de factorisation de Neyman, si g_ϑ est la projection $\mathbb{R}^{\Theta} \mapsto \mathbb{R}$ dont la valeur est la coordonnée correspondant à ϑ, on a

$$\Lambda(\vartheta)(\omega) = dP_\vartheta / dP_{\vartheta_0}(\omega) = g_\vartheta\left(T(\omega)\right),$$

et par conséquent T est exhaustive. La famille des lois

$$(\mathcal{L}(T|P_\vartheta), \vartheta \in \Theta)$$

est donc une expérience equivalente à $\mathcal{P}$. Il suffit maintenant de remarquer que cette famille est déja déterminée par un de ses éléments, c'est-à-dire $\mathcal{L}(T|P_{\vartheta_0}) = \mathcal{L}(\Lambda_{\mathcal{P}})$. Or on a pour chaque fonction mesurable bornée h qui ne dépend que d'un nombre fini de coordonnées h

$$E_\vartheta h(T) = E_{\vartheta_0} h(T) \frac{dP_\vartheta}{dP_{\vartheta_0}} = E_{\vartheta_0} h(T) g_\vartheta\left(T\right),$$

ce qui est une fonctionelle de la loi $\mathcal{L}(T|P_{\vartheta_0})$. On a donc démontré (17) sans trop de rigueur, en supposant que la famille $\mathcal{P}$ soit dominée par un de ses éléments.

Le critère (17) suggère un résultat analogue approximatif

$$\Delta(\mathcal{P}_n, \mathcal{Q}_n) \to 0 \text{ si et seulement si } \mathcal{L}(\Lambda_{\mathcal{P}_n}) \overset{\mathcal{L}}{\approx} \mathcal{L}(\Lambda_{\mathcal{Q}_n}) \text{ pour } n \to 0 \tag{18}$$

où la relation $\overset{\mathcal{L}}{\approx}$ signifie proximité en loi, une notion à préciser. Si une expérience limite est visée comme dans (15), la suite $\mathcal{Q}_n$ est constante et la proximité se réduit à la convergence en loi.

2.3. Normalité asymptotique locale (LAN) dans le cas d'un n-échantillon

Dans le cas des variables indépendantes, la convergence en loi du processus de vraisemblance se vérifie de la façon suivante. D'abord, pour le modèle gaussien

de translation (15), ce processus prend la forme

$$\Lambda_{\mathcal{Q}}(h) = \exp\left(h^\top J_{\vartheta_0}^{1/2}\xi - \frac{1}{2}h^\top J_{\vartheta_0}h\right)$$

où ξ est une variable gaussienne standard dans $\mathbb{R}^k$. Dans la famille $\mathcal{P}_n = (P_{n,\vartheta}, \vartheta \in \Theta)$, $\Theta \subset \mathbb{R}^k$, où $P_{n,\vartheta}$ est une loi produit P_ϑ^n, on a effectué la localisation $\vartheta = \vartheta_0 + n^{-1/2}h$ introduisant le nouveau paramètre local h. Soit f_h la densité de $P_{\vartheta_0+n^{-1/2}h}$ par rapport à la mesure de Lebesgue. Dans les cas réguliers où f_h est dérivable en h, on obtient $f_h/f_0 - 1 \approx n^{-1/2}$ et le logarithme de la vraisemblance peut s'écrire

$$\begin{aligned}\log \Lambda_{\mathcal{P}_n}(h) &= \sum_{i=1}^n \log \frac{f_h}{f_0}(X_i) \approx \sum_{i=1}^n \left(\frac{f_h}{f_0}(X_i) - 1\right) - \frac{1}{2}\sum_{i=1}^n \left(\frac{f_h}{f_0}(X_i) - 1\right)^2 \\ &\xrightarrow{\mathcal{L}} h^\top J_{\vartheta_0}^{1/2}\xi - \frac{1}{2}h^\top J_{\vartheta_0}h \qquad (19)\end{aligned}$$

par un théorème de type limite centrale et une loi des grands nombres. La convergence en loi de $\Lambda_{\mathcal{P}_n}$ vers $\Lambda_{\mathcal{Q}}$ est donc vérifiée, en passant par les logarithmes.

Rappelons que l'expérience limite gaussienne (15) a été établie dans un cadre local, autour de ϑ_0, en supposant implicitement que ϑ_0 peut être estimé avec une précision de l'ordre de $n^{-1/2}$. La théorie développée autour de cette idée de Le Cam peut être nommée paradigme LAN (local asymptotic normality). Le champ des applications s'est avéré beaucoup plus étendu que celui des variables indépendantes; voir Strasser [21], Genon-Catalot et Picard [7], van der Vaart [22], Shiryaev et Spokoiny [20].

L'utilité de cette approche est limitée dans le cas non paramétrique. Supposons que tous les P_ϑ sont des lois sur $[0,1]$, que la famille $\mathcal{P}_n$ des lois produits $P_{n,\vartheta} = P_\vartheta^n$ est paramétrisée par la densité de Lebesgue $\vartheta = f$ de P_ϑ et que Θ est identifié à un ensemble de densités Σ de dimension infinie. De nouveau, la localisation autour d'une densité centrale f_0 selon $f - f_0 \approx n^{-1/2}$ et la propriété LAN sont souvent possibles. Mais typiquement le centre f_0 ne peut pas être estimé à la précision $n^{-1/2}$, dû au fait que le problème est mal posé (au sens analytique). En effet pour la fonction de répartition empirique $\hat{F}_n$ on a

$$\hat{F}_n(t) - \int_0^t f(t)dt = O_p(n^{-1/2})$$

mais l'application $f \mapsto \int_0^t f(t)dt$ n'a pas d'inverse continu, donc la vitesse d'estimation $n^{-1/2}$ n'est pas possible pour f. Puisque les bornes du risque obtenues à travers la méthode LAN reposent sur la renormalisation par $n^{1/2}$ (*cf.* (16)), il s'ensuit que le paradigme LAN n'est pas adéquat pour l'estimation globale de la densité.

3. Équivalence asymptotique non paramétrique

Dans le cas de n variables indépendantes équidistribuées sur $\mathbb{R}$, la fonction de répartition empirique $\hat{F}_n$ est toujours une statistique exhaustive. Le point de départ peut donc être le théorème de la limite centrale pour cette statistique :

$$\sqrt{n}(\hat{F}_n(t) - F(t)) \stackrel{\mathcal{L}}{\Longrightarrow} B \circ F(t), \text{ pour un pont brownien } B.$$

D'abord on voudrait en déduire une expérience gaussienne approchante analogue à (10) et (13), qui maintenant pourrait prendre la forme

$$dy(t) = f(t)dt + n^{-1/2} f^{1/2}(t) dW(t), \ t \in [0,1]$$

(« signal f observé dans un bruit blanc »). Mais ce modèle hétéroscédastique n'est pas valable au sens statistique (c'est-à-dire trivial) puisque les lois du processus $y(t)$, $t \in [0,1]$ sont orthogonales si les f sont différents. Cet effet est causé par la présence du facteur $f^{1/2}(t)$ auprès de $dW(t)$, c'est-à-dire dans le coefficient de diffusion. Pour résoudre ce problème d'hétéroscédasticité, une stabilisation de la variance, analogue à celle de (11), serait souhaitable. Dans le cas de (11) on a utilisé la racine carrée comme transformation ; maintenant on pourrait prendre en compte le rôle spécial de la racine de la densité $f^{1/2}$ lié à la distance de Hellinger. Ceci suggère comme approximation gaussienne et valable de la famille $\mathcal{P}_n$ des lois produits un modèle de signal avec bruit blanc homoscédastique dans lequel le signal serait $f^{1/2}$.

3.1. Approximation par un modèle de signal avec bruit blanc

Passons à l'énoncé du théorème correspondant (voir [18]). Considérons pour $\alpha \in (0,1)$, $M > 0$ une classe de fonctions Hölderiennes

$$\mathcal{H}^\alpha(M) = \{f : \ |f(x) - f(y)| \leqslant M \ |x-y|^\alpha\}. \tag{20}$$

Pour $\alpha, M, \varepsilon > 0$ donnés, définissons un espace de paramètres

$$\Sigma_d(\alpha, M, \varepsilon) = \mathcal{H}^\alpha(M) \cap \{\text{densités sur } [0,1] \text{ bornées inférieurement par } \varepsilon\}.$$

THÉORÈME 1. — *Soit $\Sigma = \Sigma_d(\alpha, M, \varepsilon)$ pour $\varepsilon > 0$, $M > 0$ et $\alpha > 1/2$ fixés. Alors les expériences données par les observations*

$$\begin{aligned} &X_i,\ i = 1, \ldots, n \text{ indépendantes, de densité } f \\ &dy(t) = f^{1/2}(t)dt + \frac{1}{2} n^{-1/2} dW(t), \ t \in [0,1], \end{aligned} \tag{21}$$

où $f \in \Sigma$, sont asymptotiquement équivalentes.

La preuve de ce résultat repose sur la relation (18) concernant le processus de vraisemblance, où il faut préciser la notion de proximité en loi $\stackrel{\mathcal{L}}{\approx}$ pour deux suites. Pour cela on utilise le coupling, c'est-à-dire la construction des processus sur le même espace de probabilité qui soient proches au sens

métrique. Les noyaux de Markov réalisant l'équivalence asymptotique ne sont pas donnés de façon explicite, la méthode est donc fort indirecte. Des travaux plus récents (Carter [4], Brown, Carter, Low et Zhang [2]) ont réussi à refaire la preuve en exhibant les noyaux de Markov qui effectueraient un lissage du processus empirique comme dans (12).

Il faut remarquer qu'un précurseur du Théorème 1 pour les modèles paramétriques a été déjà démontré par Le Cam dans [14]. Ceci concernait les ensembles de densités Σ qui sont de dimension finie dans la métrique de Hellinger. Ce cas se ramène essentiellement à celui d'une famille paramétrique de densités f_ϑ, $\vartheta \in \Theta \subset \mathbb{R}^k$, et le modèle de bruit blanc gaussien (21) avec signal $f_\vartheta^{1/2}$ pourrait être compris comme résultat d'une stabilisation de la variance dans (13). Il y a d'autres variantes de l'approximation globale gaussienne, voir Pfanzagl [19].

Un précurseur plus immédiat du Théorème 1 a été le résultat de Brown et Low [1] concernant la relation entre un modèle de signal avec bruit blanc et sa version discrétisée, c'est-à-dire avec la régression non paramétrique gaussienne. Soit f une fonction sur $[0,1]$ appartenant à un classe Hölderienne $\mathcal{H}^\alpha(M)$ (*cf.* (20)) où $\alpha > 1/2$. Alors l'équivalence asymptotique se produit entre les deux modèles

$$\begin{aligned} &Y_i,\ i = 1, \ldots, n \text{ indépendantes, de loi } N(f(i/n), 1) \\ &dy(t) = f(t)dt + n^{-1/2}dW(t),\ t \in [0,1]. \end{aligned}$$

La preuve élégante de ce résultat s'appuie sur l'exhaustivité dans des modèles gaussiens et est donc constructive au sens des noyaux de Markov. Enfin, Brown, Zhang [3] ont établi que, pour les classes Hölderiennes, la borne $\alpha > 1/2$ est exacte, dans le Théorème 1 ainsi que dans la régression non paramétrique gaussienne ci-dessus, en donnant des contre-exemples dans le cas $\alpha = 1/2$. Le cas de la régression non gaussienne a été discuté dans [10], [11]. De plus, il a été démontré que le résultat du Théorème 1 se reproduit dans le cas où le modèle des variables indépendantes équidistribuées serait remplacé par un certain processus de diffusion (voir [8]).

3.2 Densité spectrale d'une suite gaussienne stationnaire

Soit $y(1), \ldots, y(n)$ une suite gaussienne stationnaire telle que $Ey(1) = 0$, de fonction d'autocovariance

$$\gamma(h) = Ey(t)y(t+h) = \int_{-\pi}^{\pi} \exp(\mathrm{i}\, h\omega)\, f(\omega)\, d\omega$$

où f est la densité spectrale définie sur $[-\pi, \pi]$. La fonction f est non négative, symétrique ($f(\omega) = f(-\omega)$) et nous supposons $f \in L_2[-\pi, \pi]$. Encore une fois, nous considérons le cas non paramétrique, c'est-à-dire nous supposons que $f \in \Sigma$, une classe de fonctions régulières de dimension infinie. Les sujets classiques comme l'estimation de f, les vitesses de convergence, l'optimalité, etc. ont été étudiés en détail. De plus, la propriété de normalité asymptotique locale (LAN) a été établie dans le cadre paramétrique, voir Davies [5] et

Dzhaparidze [6] où l'information de Fisher (l'expression J_ϑ figurant dans (19)) est déterminée de la façon suivante. Supposons une famille paramétrique régulière de densités spectrales : f_ϑ, $\vartheta \in \Theta \subset \mathbb{R}$: alors

$$J_\vartheta = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\frac{\partial}{\partial \vartheta} \log f_\vartheta(\omega) \right)^2 d\omega.$$

Ceci suggère un modèle de signal avec bruit blanc

$$dZ_\omega = \log f_\vartheta(\omega) d\omega + 2\pi^{1/2} n^{-1/2} dW_\omega,\ \omega \in [-\pi, \pi]$$

où $\vartheta \in \Theta \subset \mathbb{R}$, pour tous les familles paramétriques régulières. En fait, ce dernier modèle aurait la même information de Fisher asymptotique.

Passons à l'énoncé du théorème correspondant (voir [9]). Considérons, pour $\alpha \in (0,1)$ et $M > 0$, une classe de fonctions Hölderiennes (20). Pour α, M et $\varepsilon \in (0,1)$ donnés, définissons un espace de paramètres

$$\Sigma_s(\alpha, M, \varepsilon) = \mathcal{H}^\alpha(M) \cap \left\{ \text{fonctions } f \text{ sur } [0,1] \text{ à valeurs dans } (\varepsilon, \varepsilon^{-1}) \right\}.$$

THÉORÈME 2. — *Soit* $\Sigma = \Sigma_s(\alpha, M, \varepsilon)$ *pour* $\varepsilon \in (0,1)$, $M > 0$ *et* $\alpha > 1/2$ *donnés. Soit* ω_j, $j = 1, \ldots, n$ *un réseau de points situés uniformément dans* $[-\pi, \pi]$. *Alors les trois expériences données par les observations*

Y_i, $i = 1, \ldots, n$, *stationnaire, centré, gaussien à densité spectrale* f

Z_i, $i = 1, \ldots, n$ *indépendantes, ayant la loi* $N(0, f(\omega_i))$

$dZ_\omega = \log f(\omega) d\omega + 2\pi^{1/2} n^{-1/2} dW_\omega$, $\omega \in [-\pi, \pi]$,

où $f \in \Sigma$, *sont asymptotiquement équivalentes.*

Ce résultat a deux composantes. La première effectue la réduction de la suite stationnnaire à un modèle de variables gaussiennes indépendantes. Ce dernier est une version non paramétrique du modèle gaussien d'échelle de l'Exemple 2. Dans ce cadre, la deuxième partie du théorème s'apparente à la Proposition 3. Les démonstrations connues jusqu'à présent sont indirectes au sens de l'existence des noyaux markoviens (voir [9]).

Références

[1] BROWN L. D. and LOW M. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24** 2384-2398.

[2] BROWN L. D., CARTER A. V., LOW M. G. and ZHANG C.-H. (2004). Equivalence theory for density estimation, Poisson processes and Gaussian white noise with drift. To appear, *Ann. Statist.* **32** (5).

[3] BROWN L. D. and ZHANG C.-H. (1998). Asymptotic nonequivalence of nonparametric experiments when the smoothness index is 1/2. *Ann. Statist.* **26**, 279-287.

[4] CARTER A. (2002). Deficiency distance between multinomial and multivariate normal experiments. *Ann. Statist.* **30** 708-730.

[5] DAVIES R.B. (1973). Asymptotic inference in stationary Gaussian time-series, *Adv. Appl. Probab.* **5**, 469-497.

[6] DZHAPARIDZE K. (1986). *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series.* Springer-Verlag, New York Inc.

[7] GENON-CATALOT V. et PICARD D. (1993). *Eléments de Statistique Asymptotique.* Mathématiques et Applications 11, Springer Verlag, Paris.

[8] GENON-CATALOT V., LARÉDO C., NUSSBAUM M. (2002). Asymptotic equivalence of estimating a Poisson intensity and a positive diffusion drift. *Ann. Statist.* **30** 731-753.

[9] GOLUBEV G., NUSSBAUM M. and ZHOU H. (2004). Asymptotic equivalence of spectral density estimation and Gaussian white noise. En préparation.

[10] GRAMA I. and NUSSBAUM M. (1998). Asymptotic equivalence for nonparametric generalized linear models. *Prob. Theor. Rel. Fields*, **111**, 167-214.

[11] GRAMA I. and NUSSBAUM M. (2002). Asymptotic equivalence for nonparametric regression. *Math. Meth. Statist.* **11** (1) 1-36.

[12] BROWN L. D. and LOW M. (1996). Asymptotic equivalence of nonparametric regression and white noise, *Ann. Statist.* **24** 2384-2398 (1996).

[13] LE CAM L. (1969). *Théorie Asymptotique de la Décision Statistique.* Les Presses de l'Université de Montréal.

[14] LE CAM L. (1985). Sur l'approximation de familles de mesures par des familles gaussiennes. *Ann. Inst. Henri Poincaré* **21** (3) 225-287.

[15] LE CAM L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag, New York.

[16] LE CAM L. and YANG G. (2000). *Asymptotics in Statistics, 2nd ed..* Springer-Verlag, New-York.

[17] MÜLLER D. W. (1981). The increase of risk due to inaccurate models. *Symposia Mathematica.* Instituto Nazionale di Alta Mathematica, Vol. 25.

[18] NUSSBAUM M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24**, 2399-2430.

[19] PFANZAGL J. (1995). On local and global asymptotic normality. *Math. Meth. Statist.* **4** 115-136

[20] SHIRYAEV A. N. and SPOKOINY V. (2000). *Statistical Experiments and Decisions : Asymptotic Theory.* World Scientific, Singapore.

[21] STRASSER H. (1985). *Mathematical Theory of Statistics.* de Gruyter, Berlin.

[22] VAN DER VAART A. W. (1998). *Asymptotic Statistics.* Cambridge University Press.

[23] VAN DER VAART A. W. (2002). The statistical work of Lucien Le Cam. *Ann. Statist.* **30** 631-682.

[24] WALD A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* **54** 426-482.