Available online at www.sciencedirect.com
ScienceDirect
C. R. Acad. Sci. Paris, Ser. I 344 (2007) 645–650
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Statistique/Probabilités

Un principe fonctionnel de grandes déviations en estimation non-paramétrique

Djamal Louani [a,b], Sidi Mohamed Ould Maouloud [b]

[a] *L.S.T.A, Université de Paris 6, 75252 Paris cedex 05, France*
[b] *Université de Reims, moulin de la Housse, BP 1039, 51687 Reims cedex 2, France*

Reçu le 15 décembre 2005 ; accepté après revision le 13 mars 2007

Disponible sur Internet le 10 mai 2007

Présenté par Paul Deheuvels

Résumé

L'objet de cette Note est d'établir un principe fonctionnel de grandes déviations pour le processus $U_n(f)$ indexé par une famille de fonctions $\mathcal{F}$ et de déduire de façon unifiée des résultats de grandes déviations pour les estimateurs de la fonction de régression et de la fonctions de densité. Nos résultats sont alors appliqués au problème de selection de modèles. ***Pour citer cet article : D. Louani, S.M. Ould Maouloud, C. R. Acad. Sci. Paris, Ser. I 344 (2007).***

Abstract

A functional large deviation principle in nonparametric estimation. The subject of this Note is to establish a functional large deviations principle for the process $U_n(f)$ indexed by a family of functions $\mathcal{F}$ and to deduce large deviations results in an unified way for the density and regression functions in the nonparametric estimation problem. Our results are then applied to model selection. ***To cite this article: D. Louani, S.M. Ould Maouloud, C. R. Acad. Sci. Paris, Ser. I 344 (2007).***

Abridged English version

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a sequence of i.i.d. $\mathbb{R} \times [a, b]$-valued random vectors with a joint density function g with respect to Lebesgue measure. Denote by g_X and g_Y the marginal density functions of X_1 and Y_1 respectively. Taking $\mathcal{F}$ as a compact subset of the space $\mathcal{C}([a, b])$ of continuous real functions over the interval $[a, b]$ and $z_0 \in \mathbb{R}$, define for $f \in \mathcal{F}$, the process

$$W_n(f) := W_n(f)(z_0) = \frac{1}{nh_n} \sum_{i=1}^{n} f(Y_i) K\left(\frac{X_i - z_0}{h_n}\right),$$

where h_n is a sequence of bandwidths and K is a positive kernel integrating to 1.

Adresses e-mail : louani@ccr.jussieu.fr (D. Louani), sidi-mohamed.ould-maouloud@etudiant.univ-reims.fr (S.M. Ould Maouloud).

1631-073X/$ – see front matter
doi:10.1016/j.crma.2007.03.017

Consider now the vector process U_n, indexed by the family $\mathcal{F}$ and defined for any $f \in \mathcal{F}$ by

$$U_n(f) = \big(W_n(f), W_n(\mathbf{1})\big)$$

where **1** indicates the function identically equal to 1.

The main result of this Note gives a functional large deviation principle for the process U_n in the space $\mathcal{C}(\mathcal{F}) \times \mathbb{R}$. This result allows us to deduce in an unified way large deviation theorems for both the kernel density estimator and the Nadaraya–Watson estimator type of the regression function. Some implications in model selection with respect to the large deviations criterion are also deduced. Our results are stated under some regularity conditions over the kernel K, the densities g, g_X, and the class of functions $\mathcal{F}$ which is assumed to be compact with respect to the uniform topology. The main result of this Note may be summarized by the fact that the process U_n indexed by $\mathcal{F}$ satisfies a functional large deviation principle with convergence rate nh_n and a good rate function.

Usually, results of this type are proved by stating a large deviation principle for the finite-dimension version of the process before investigating its exponential tightness. Our method of proof is quite different. In first place, we establish a large deviation principle for the vector process $(A_n, B_n, W_n(\mathbf{1}))$ defined in (1) below. Using this result and the contraction principle, we state a large deviations principle for an exponential approximating process U_n^m of the process U_n. Finally, we prove that the process U_n^m is a good exponential approximation of the process U_n, i.e., the statement (2) holds true.

1. Introduction

Soit $(X_1, Y_1), \ldots, (X_n, Y_n), \ldots$ une suite de vecteurs aléatoires i.i.d. à valeurs dans $\mathbb{R} \times [a, b]$ ayant pour densité jointe la fonction g et pour densités marginales les fonctions g_X et g_Y. Soit $(h_n)_{n \geqslant 1}$ une suite de paramètres de lissage (i.e., une suite de réels positifs convergeant vers zéro). Dans toute la suite de ce travail K désigne une fonction réelle positive telle que $\int K(u)\,\mathrm{d}u = 1$. Soit $\mathcal{F}$ une partie compacte de l'espace $\mathcal{C}([a, b])$ des fonctions réelles continue sur l'intervalle $[a, b] \subset \mathbb{R}$. Pour $z_0 \in \mathbb{R}$ et $f \in \mathcal{F}$, on considère le processus suivant indexé par la famille de fonctions $\mathcal{F}$

$$W_n(f) := W_n(f)(z_0) = \frac{1}{nh_n} \sum_{i=1}^{n} f(Y_i) K\left(\frac{X_i - z_0}{h_n}\right).$$

Ce processus a été introduit par Einmahl et Mason (2000) dans [2] sous une forme plus complexe. Il leur a permis d'unifier l'étude de la consistance uniforme, par rapport à une famille de fonctions $\mathcal{F}$, de l'estimateur du type Nadaraya–Watson de la fonction de régression et de l'estimateur à noyau de la fonction de densité. Ils ont aussi obtenu des vitesses de convergence pour ces estimateurs.

Le processus $W_n(f)$ nous permet dans le cadre de ce travail d'établir de façon unifiée des principes de grandes déviations pour l'estimateur de type Nadaraya–Watson de la fonction de régression et de l'estimateur à noyau de la fonction de densité. Pour cela, on considère le processus vectoriel indexé par la famille de fonctions $\mathcal{F}$ défini par

$$U_n(f) = \big(W_n(f), W_n(\mathbf{1})\big)$$

où **1** désigne la fonction identiquement égale à 1. Notons par $\mathcal{C}(\mathcal{F})$ l'espace des fonctions réelles continues sur $(\mathcal{F}, \|\ \|_\infty)$, où $\|\ \|_\infty$ désigne la norme infinie, que l'on munit de la norme

$$\|Z\|_{\mathcal{C}(\mathcal{F})} = \sup_{f \in \mathcal{F}} \big|Z(f)\big|.$$

L'espace $(\mathcal{C}(\mathcal{F}), \|\ \|_{\mathcal{C}(\mathcal{F})})$ ainsi formé est un espace vectoriel normé. Considérons maintenant l'espace $\mathcal{C}(\mathcal{F}) \times \mathbb{R}$ muni de la norme

$$\big\|(Z, t)\big\|_{\mathcal{C}(\mathcal{F}) \times \mathbb{R}} = \max\big\{\|Z\|_{\mathcal{C}(\mathcal{F})}, |t|\big\}.$$

Pour $(Z, t) \in \mathcal{C}(\mathcal{F}) \times \mathbb{R}$ et δ un réel positif, la boule ouverte $B_{(Z,t),\delta}$ dans $\mathcal{C}(\mathcal{F}) \times \mathbb{R}$ de centre (Z, t) et de rayon δ est définie par

$$B_{(Z,t),\delta} = \big\{(Z', t') \in \mathcal{C}(\mathcal{F}) \times \mathbb{R}\colon\ \big\|(Z, t) - (Z', t')\big\|_{\mathcal{C}(\mathcal{F}) \times \mathbb{R}} < \delta\big\}.$$

Établir un principe de grandes déviations (PGD) pour le processus U_n indexé par la famille de fonctions $\mathcal{F}$ consiste à trouver une suite ϵ_n de réels positifs convergeant vers zéro et une fonction I semi-continue inférieurement sur $\mathcal{C}(\mathcal{F}) \times \mathbb{R}$ telles que

$$\begin{aligned}
&\text{Pour tout fermé } F \text{ de } \mathcal{C}(\mathcal{F}) \times \mathbb{R}\\
&\limsup_{n\to\infty} \epsilon_n \log P(U_n \in F) \leqslant -\inf_{(Z,t)\in F} I(Z,t).\\
&\text{Pour tout ouvert } G \text{ de } \mathcal{C}(\mathcal{F}) \times \mathbb{R}\\
&\liminf_{n\to\infty} \epsilon_n \log P(U_n \in G) \geqslant -\inf_{(Z,t)\in G} I(Z,t).
\end{aligned}$$

On dira alors que le processus U_n satisfait le PGD dans $\mathcal{C}(\mathcal{F}) \times \mathbb{R}$ avec la vitesse ϵ_n^{-1} et la fonction de taux I.

Le résultat obtenu pour le processus U_n permet alors de déduire des résultats de grandes déviations à la fois pour l'estimateur à noyau de la fonction de densité et pour l'estimateur de type Nadaraya–Watson de la fonction de régression. Ces résultats ont été obtenu séparément par D. Louani dans [3] et [4]. On rappelle que la fonction de densité g_X est estimée par $W_n(\mathbf{1})$ et que l'estimateur de la fonction de régression $m(f, z_0) := E(f(Y)|X = z_0)$ est

$$m_n(f, z_0) := \frac{W_n(f)}{W_n(\mathbf{1})}.$$

Les implications du résultat principale relatifs à ces deux estimateurs sont donnés dans les Corollaire 1 et 2 respectivement.

La méthode de démonstration classique de résultats de grandes déviations fonctionnels consiste à établir en premier lieu des résultats pour la version fini-dimensionnelle du processus pour ensuite réaliser sa tension exponentielle. La démarche entreprise dans cette Note est différente. Elle consiste à établir un principe des grandes déviations dans $\mathbb{R}^{2m+1}$, pour le processus vectoriel $(A_n, B_n, W_n(\mathbf{1}))$ où les vecteurs $A_n = (A_n^1, \ldots, A_n^m)$ et $B_n = (B_n^1, \ldots, B_n^m)$ sont définis par

$$A_n^j = \frac{1}{nh_n}\sum_{i=1}^n Y_i \mathbf{1}_{[t_{j-1},t_j]}(Y_i) K\left(\frac{X_i - z_0}{h_n}\right) \quad \text{et} \quad B_n^j = \frac{1}{nh_n}\sum_{i=1}^n \mathbf{1}_{[t_{j-1},t_j]}(Y_i) K\left(\frac{X_i - z_0}{h_n}\right) \tag{1}$$

pour $j = 1, 2, \ldots, m$. Les t_j formant ici une partition uniforme de l'intervalle $[a, b]$, i.e., $a = t_0 < t_1 < \cdots < t_m = b$ et $t_j - t_{j-1} = (b-a)/m$, $1 \leqslant j \leqslant m$ et $\mathbf{1}_A$ désigne la fonction indicatrice de l'ensemble A. Nous construisons ensuite une approximation exponentielle U_n^m du processus U_n définie par

$$U_n^m = \left(W_n^m, W_n(\mathbf{1})\right)$$

où,

$$W_n^m(f) = \frac{1}{nh_n}\sum_{i=1}^n \left(\sum_{j=1}^m \left(\alpha_j(f) Y_i + \beta_j(f)\right)\mathbf{1}_{[t_{j-1},t_j]}(Y_i)\right) K\left(\frac{X_i - z_0}{h_n}\right)$$

avec $\alpha_j(f) = \frac{f(t_{j-1}) - f(t_j)}{t_{j-1} - t_j}$ et $\beta_j(f) = \frac{t_{j-1} f(t_j) - t_j f(t_{j-1})}{t_{j-1} - t_j}$; $j = 1, 2, \ldots, m$.

Un PGD pour U_n^m est alors établi à l'aide du résultat obtenu pour le processus $(A_n, B_n, W_n(\mathbf{1}))$ et du principe de contraction. En utilisant le Théorème 4.2.16 dans [1], on conclut notre preuve en montrant que le processus U_n^m est une bonne approximation exponentielle de U_n, i.e, pour tout $\tau > 0$

$$\lim_{m\to\infty} \limsup_{n\to\infty} \frac{1}{nh_n} \log P\left(\left\| U_n - U_n^m \right\|_{\mathcal{C}(\mathcal{F})\times R} > \tau\right) = -\infty. \tag{2}$$

2. Résultats

Nous donnons maintenant l'ensemble des hypothèses nécessaires à l'établissement de nos résultats :

(H.1) $h_n \to 0$ et $nh_n \to \infty$ quand $n \to \infty$.
(K.1) $\int_{-\infty}^{\infty} |\exp\{tK(u)\} - 1|\,\mathrm{d}u < \infty$ pour tout $t \in \mathbb{R}$.
(K.2) $\int_{-\infty}^{\infty} |u(\exp\{tK(u)\} - 1)|\,\mathrm{d}u < \infty$ pour tout $t \in \mathbb{R}$.

(K.3) $\int_{-\infty}^{\infty} K(u)\exp\{tK(u)\}\,\mathrm{d}u < \infty$ pour tout $t \in \mathbb{R}$.
(G.1) g est derivable par rapport à la première composante.
(G.2) $\sup_{-\infty<u<\infty} \int_a^b |\frac{\partial}{\partial u} g(u,v)|\,\mathrm{d}v < \infty$.
(F.1) $\sup_{f\in\mathcal{F}} |f(t)| < \infty$ pour un certain $t \in [a,b]$.
(F.2) $\sup_{f\in\mathcal{F}} \sup_{|t-s|\leqslant\delta} |f(t)-f(s)| \to 0$ quand $\delta \to 0$.

2.1. Discussion des conditions

Les conditions (K) sont des conditions classiques rencontrées dans les résultats des grandes déviations en estimation non paramétrique. Les conditions (F) assurent la compacité de la famille $\mathcal{F}$ et constituent des conditions d'application du théorème d'Arzela–Ascoli.

Introduisons maintenant les fonctions qui serviront à exposer nos résultats. Pour $(\alpha,\beta,\gamma) \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}$, posons

$$l_1(\alpha,\beta) = \sum_{j=1}^{m} \int_{-\infty}^{\infty} \int_{t_{j-1}}^{t_j} \big(\exp\big\{(\alpha_j v + \beta_j)K(u)\big\} - 1\big) g(z_0,v)\,\mathrm{d}v\,\mathrm{d}u \quad \text{et}$$

$$l_2(\alpha,\beta,\gamma) = \sum_{j=1}^{m} \int_{-\infty}^{\infty} \int_{t_{j-1}}^{t_j} \big(\exp\big\{(\alpha_j v + \beta_j + \gamma)K(u)\big\} - 1\big) g(z_0,v)\,\mathrm{d}v\,\mathrm{d}u.$$

Pour $(x,y,z) \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}$, posons

$$l_1^*(x,y) = \sup_{(\alpha,\beta)\in\mathbb{R}^m\times\mathbb{R}^m} \big\{\big\langle (x,y),(\alpha,\beta)\big\rangle - l_1(\alpha,\beta)\big\} \quad \text{et}$$
$$l_2^*(x,y,z) = \sup_{(\alpha,\beta,\gamma)\in\mathbb{R}^m\times\mathbb{R}^m\times\mathbb{R}} \big\{\big\langle (x,y,z),(\alpha,\beta,\gamma)\big\rangle - l_2(\alpha,\beta,\gamma)\big\}.$$

Les fonctions l_1 et l_2 sont les fonctions limites des suites de transformées de Laplace associées au vecteurs (A_n, B_n) et $(A_n, B_n, W_n(\mathbf{1}))$, respectivement, alors que les fonctions l_1^* et l_2^* sont les transformées de Legendre des fonctions l_1 et l_2, respectivement.

Nos résultats sont présentés sous la forme de deux théorèmes etablissant les PGD pour le processus $W_n(f)$ et $U_n(f)$ et de corollaires présentant les déductions faites pour les estimateurs de la fonction de densité et de la fonction de régression :

Théorème 1. *On suppose vérifiées les conditions* (H.1), (K.1)–(K.3), (G.1)–(G.2) et (F.1)–(F.2). *Alors le processus W_n indexé par la famille de fonctions $\mathcal{F}$ satisfait un PGD dans $\mathcal{C}(\mathcal{F})$ avec la vitesse nh_n et une bonne fonction de taux*

$$J(Z) = \sup_{\delta>0} \liminf_{m\to\infty} \inf_{Z'\in B_{Z,\delta}} J_m(Z')$$

où

$$J_m(Z') = \inf\left\{ l_1^*(x,y){:}\ \sum_{j=1}^{m} x_j\alpha_j(f) + y_j\beta_j(f) = Z'(f) \text{ pour tout } f \in \mathcal{F}\right\}.$$

Théorème 2. *Sous les hypothèses du Théorème* 1, *le processus vectoriel U_n satisfait un PGD dans $\mathcal{C}(\mathcal{F}) \times \mathbb{R}$ avec la vitesse nh_n et une bonne fonction de taux*

$$I(Z,t) = \sup_{\delta>0} \liminf_{m\to\infty} \inf_{(Z',t')\in B_{(Z,t),\delta}} I_m(Z',t')$$

où

$$I_m(Z',t') = \inf\left\{ l_2^*(x,y,t'){:}\ \sum_{j=1}^{m} x_j\alpha_j(f) + y_j\beta_j(f) = Z'(f) \text{ pour tout } f \in \mathcal{F}\right\}.$$

Corollaire 1. *On suppose vérifiées les conditions* (H.1) *et* (K.1)–(K.3) *et* (G.1)–(G.2). *Alors* $W_n(\mathbf{1})$ *satisfait un PGD dans* $\mathbb{R}$ *avec la vitesse* nh_n *et la bonne fonction de taux* J *définie par*

$$R(t) = \sup_{\gamma\in\mathbb{R}}\left\{t\gamma - g_X(z_0)\int_{-\infty}^{\infty}\big(\exp\{\gamma K(u)\} - 1\big)\,\mathrm{d}u\right\}.$$

Corollaire 2. *On suppose vérifiées les conditions du Théorème* 2. *Alors le processus* $m_n(f, z_0)$ *indexé par la famille des fonctions* $\mathcal{F}$ *satisfait un PGD dans* $\mathcal{C}(\mathcal{F})$ *avec la vitesse* nh_n *et une fonction de taux définie par*

$$I_2(Z) = \inf\big\{I(tZ, t)\colon\ t \in \mathbb{R}^*\big\}.$$

3. Application à la sélection de modèles pour le critère des grandes déviations

On considère un processus V_n défini sur $\Omega\times\mathcal{F}\times T$ et prenant des valeurs dans $\mathbb{R}$, où T est une partie compacte de $\mathbb{R}$ telle que pour tout $n \geqslant 1$ et pour presque tout $\omega\in\Omega$, l'application $(f, z) \to V_n(\omega, f, z)$ soit continue. On suppose que ce processus vérifie un PGD fonctionnel dans $\mathcal{C}(\mathcal{F}\times T)$. Pour une partie A de $\mathcal{C}(T)$ et une sous famille $\mathcal{G}$ de $\mathcal{F}$, posons

$$\Omega_{n,A,\mathcal{G}} = \big\{\omega\in\Omega\colon\ V_n(\omega,\cdot,\cdot)\in\Lambda_{A,\mathcal{G}}\big\}$$

où

$$\Lambda_{A,\mathcal{G}} = \big\{Z\in\mathcal{C}(\mathcal{F}\times T)\colon\ \forall f\in\mathcal{G},\ Z(f,\cdot)\in A\big\}.$$

La proposition suivante est un outil permettant de faire de la sélection de modèles. Avant de l'énnoncer, on définit d'abord ce que l'on entend par ensemble de continuité. Un ensemble O est dit ensemble de continuité d'une fonction I si l'on a $\inf_{x\in\mathring{O}} I(x) = \inf_{x\in\overline{O}} I(x)$, où $\mathring{O}$ est l'intérieur de O et $\overline{O}$ sa fermeture.

Proposition. *On suppose que* $\Lambda_{A,\mathcal{G}_1}$ *et* $\Lambda_{A,\mathcal{G}_2}$ *sont des ensembles de continuité pour la fonction* I. *Alors on a*

$$\lim_{n\to\infty}\epsilon_n\log P(\Omega_{n,A,\mathcal{G}_i}) = -\inf\big\{I(Z)\colon\ Z\in\Lambda_{A,\mathcal{G}_i}\big\} := -I_{A,\mathcal{G}_i},\quad i = 1, 2.$$

La comparaison des performances des modèles $(A,\mathcal{G}_1)$ et $(A,\mathcal{G}_2)$ pour le critère des grandes déviations se fait en comparant les quantités $I_{A,\mathcal{G}_1}$ et $I_{A,\mathcal{G}_2}$. On dira alors que le modèle $(A,\mathcal{G}_1)$ est meilleur que le modèle $I_{A,\mathcal{G}_2}$ si $I_{A,\mathcal{G}_1}\leqslant I_{A,\mathcal{G}_2}$.

Comme les résultats de grandes déviations obtenus dans cette Note sont restraints au cas ponctuel, la comparaison des performances des modèles qui peut être faite ici suppose l'ensemble T réduit au point z_0. En utilisant le Théorème 1, on peut comparer les modèles si $\Lambda_{A,\mathcal{G}1}$ et $\Lambda_{A,\mathcal{G}_2}$ sont des ensemble de continuité de J. L'exemple ci-dessous traite le cas d'une famille de fonctions affines et aboutit à la comparaison explicite de modèles.

Exemple. On considère la famille de l'ensemble des fonctions affines donnée par $\mathcal{F} = \{f : [a, b]\to\mathbb{R}\colon\ f(t) = xt + y, (x, y)\in\mathbb{R}^2\}$. Pour $f\in\mathcal{F}$, on a $W_n(f) = xA_n + yB_n$, avec

$$A_n = \frac{1}{nh_n}\sum_{i=1}^{n} Y_i K\left(\frac{X_i - z_0}{h_n}\right)\quad\text{et}\quad B_n = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{X_i - z_0}{h_n}\right).$$

Soit $h:\mathbb{R}^2\to\mathcal{L}(\mathbb{R}^2)$ la bijection canonique qui à (x', y') associe $\langle(x', y'), (\cdot,\cdot)\rangle$, où $\mathcal{L}(\mathbb{R}^2)$ désigne l'espace des formes linéaires sur $\mathbb{R}^2$. On remarque alors que $W_n = h(A_n, B_n)$ satisfait le même PGD que le processus (A_n, B_n). Au vu de ce qui a précédé, il est clair que le processus (A_n, B_n) satisfait un PGD dans $\mathbb{R}^2$ avec une bonne fonction de taux donnée par

$$l^*(x, y) = \sup_{(\alpha,\beta)\in\mathbb{R}^2}\big\{x\alpha + y\beta - l(\alpha,\beta)\big\}\quad\text{où}$$

$$l(\alpha,\beta)\} = \int_{-\infty}^{\infty}\int_a^b\big(\exp\big\{(\alpha v + \beta)K(u)\big\} - 1\big)g(z_0, v)\,\mathrm{d}u\,\mathrm{d}v.$$

Soit A une partie de $\mathbb{R}$ et soient $\mathcal{G}_1$ et $\mathcal{G}_2$ deux partie de $\mathbb{R}^2$. On a alors

$$\Lambda_{A,\mathcal{G}_i} = \left\{(x, y) \in \mathbb{R}^2\colon\ \forall(\alpha, \beta) \in \mathcal{G}_i, x\alpha + y\beta \in A\right\}, \quad i = 1, 2.$$

Pour A un ouvert convexe de $\mathbb{R}$ et $\mathcal{G}_1$ et $\mathcal{G}_2$ deux compacts dans $\mathbb{R}^2$, il suit que $\Lambda_{A,\mathcal{G}_1}$ et $\Lambda_{A,\mathcal{G}_2}$ sont des ouverts convexes de $\mathbb{R}^2$. Comme la fonction de taux l^* est convexe alors ces deux ensembles sont des ensembles de continuité de la fonction de taux l^*. Cela nous permet alors de comparer ces deux modèles.

Références

[1] A. Dembo, O. Zeitouni, Large Deviations Techniques and Applications, Springer, New York, 1998.

[2] U. Einmahl, D. Mason, An empirical process approach to the uniform consistency of kernel type function estimator, Theoretical Probab. 13 (2000) 1–37.

[3] D. Louani, PGD pour l'estimateur à noyau de la densité de probabilité, C. R. Acad. Sci. Paris, Ser. I 324 (1997) 569–572.

[4] D. Louani, Some large deviations limit theorems in conditional nonparametric statistics, Statistics 33 (1999) 171–196.