Principe de grandes déviations précises pour le processus empirique conditionnel

Cyrille Joutard

Principe de grandes déviations précises pour le processus empirique conditionnel

Cyrille Joutard

Département de Mathématiques, Université Montpellier 2, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France.
Université Montpellier 3.

Résumé: Considérant un cadre non paramétrique, nous établissons des résultats de grandes déviations précises pour le processus empirique conditionnel. En d'autres termes, nous donnons des développements complets des queues de probabilités. Pour prouver ceux-ci, nous nous servons d'un développement d'Edgeworth obtenu par une condition de Cramer. Notre résultat est similaire à celui du théorème de Bahadur-Rao obtenu pour la moyenne empirique (Bahadur and Rao 1960).
Mots-clés: processus empirique conditionnel, grandes déviations, estimation non paramétrique, développement d'Edgeworth, condition de Cramer.

Abstract: Considering a nonparametric framework, we state sharp large deviation results for the conditional empirical process. In other words, we provide full expansions for the tail probabilities. To prove these, we make use of an Edgeworth expansion obtained by a Cramer's condition. Our result is analogous to the Bahadur-Rao theorem on the sample mean (Bahadur and Rao 1960).
Key words: conditional empirical process, large deviation, nonparametric estimation, Edgeworth expansion, Cramer's condition.

Introduction. Let $(Y_1, \ldots, Y_n)$ be independent and identically distributed real random variables with zero mean and finite variance. It is well known that the sample mean $\bar{Y}_n = n^{-1} \sum_{j=1}^{n} Y_j$ converges to $\mathbb{E}Y_1 = 0$ in probability, as $n \to \infty$. Hence, if A is a Borel subset of $\mathbb{R}$ and $0 \notin A$, then $\mathbb{P}(\bar{Y}_n \in A)$ converges to 0. Furthermore, there exists a logarithmic equivalent of this deviation probability. In particular, for $a > 0$, we have the following result :

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\bar{X}_n \geq a) = -I(a).$$

where $I(a)$ is the Fenchel-Legendre dual of the log Laplace of X_1. The function I, which depends on the distribution of X_1, is known as the rate function. This result is a consequence of the large deviation principle (LDP) satisfied by $\bar{Y}_n$.
For the general definition of a large deviation principle, we refer to [8] or [9]. In some cases, statistical applications need sharper results than that provided by large deviations.

One of the first results of this kind is due to Bahadur and Rao [1]. Considering the sample mean $\bar{Y}_n$, they obtained the following expansions :

$$\mathbb{P}(\bar{Y}_n \geq c) = \frac{\exp(-nI(c))}{(2\pi n)^{1/2}\sigma_c\tau_c}\left[1 + \sum_{j=1}^{p}\frac{a_j}{n^j} + O\left(\frac{1}{n^{p+1}}\right)\right], \quad c > 0 \tag{1}$$

or

$$\mathbb{P}(\bar{Y}_n \leq c) = \frac{\exp(-nI(c))}{(2\pi n)^{1/2}\sigma_c\tau_c}\left[1 + \sum_{j=1}^{p}\frac{a_j}{n^j} + O\left(\frac{1}{n^{p+1}}\right)\right], \quad c < 0 \tag{2}$$

where $a_j \in \mathbb{R}$, and $\tau_c > 0$, $\sigma_c > 0$ are parameters depending on c.
The sample mean $\bar{Y}_n$ is said to satisfy a sharp large deviation principle (SLDP) whenever the result (1) or (2) holds for $c \in \mathbb{R}$ (see [3] or [4]). Note that Chaganty and Sethuraman [6] generalize the Bahadur and Rao theorem on the sample mean to an arbitrary sequence of random variables Z_n.
In this paper, we consider a nonparametric framework that we introduce hereafter. Let $((X_1, Y_1), \ldots, (X_n, Y_n))$ be an i.i.d. bivariate sample with common density function $f(x, y)$ and let (X, Y) be a pair of random variables having the same distribution as (X_1, Y_1). For a Borel set $A \subset \mathbb{R}$, the conditional empirical measure of $\{Y \in A\}$ given $X = x_0 \in \mathbb{R}$ is defined by

$$\hat{m}_n(A|x_0) = \frac{\sum_{j=1}^{n}\mathbb{I}_{\{Y_j\in A\}}K\left(\frac{x_0-X_j}{h_n}\right)}{\sum_{j=1}^{n}K\left(\frac{x_0-X_j}{h_n}\right)},$$

provided $\sum_{j=1}^{n}K\left(\frac{x_0-X_j}{h_n}\right) \neq 0$ (otherwise $\hat{m}_n(A|x_0) = 0$), where K is a positive function integrating to one and (h_n) is a sequence of positive real numbers tending to zero.
For $x_0 \in \mathbb{R}$, the conditional empirical process is then defined by

$$\hat{\alpha}_n(y|x_0) = \sqrt{nh_n}(\hat{m}_n((-\infty, y]|x_0) - m(y|x_0)) \tag{3}$$

where $m(y|x_0) = \mathbb{P}(Y \leq y|X = x_0)$.
This paper deals with the SLDP satisfied by $\hat{\alpha}_n(y_0|x_0)$, for fixed (x_0, y_0).
Recently, there have been several works studying large deviations in the context of nonparametric function estimation. Louani [15, 16] obtained large deviation results for the nonparametric estimators of density and regression functions for real i.i.d. observations. Gao [10] studied some problems of large and moderate deviations for the density estimate when the observations are d-dimensional and i.i.d. Worms [17] worked on large and moderate deviations of the nonparametric estimators of density and regression functions for d-dimensional and dependent observations. Considering non independent observations, Liangzhen [14] obained large deviation principles in $L_1(\mathbb{R}^d)$ for the density estimate. Assuming i.i.d. observations, Berrahou [5] proved a pointwise large deviation principle for the delta-sequence method density estimator. Also considering an i.i.d. framework, Joutard

[13] proved results of sharp large deviations for the nonparametric estimators of density and regression functions (for $p = 0$). Concerning the conditional empirical process, strong convergence and asymptotic normality have been established by Stute [18]. Later, Louani [16] studied large deviations for the conditional empirical process. Here, we are interested with problems of sharp large deviations. Analogously to the Bahadur and Rao theorem on the sample mean [1], we establish asymptotic expansions for the conditional empirical process of the same kind of (1) and (2) (that is, of arbitrary order $p \geq 0$). For this, we need to specify the rate of convergence to zero of the bandwidth h_n. Other assumptions upon the density function $f(x, y)$ and the kernel K are also assumed. The proof follows the same approach as Bahadur and Rao [1]. In particular, an Egdeworth expansion of arbitrary order is necessary. For the statement of this expansion, we use the same kind of arguments as Hall [11] who proved an Egdeworth expansion for the kernel density estimator. A Parseval's formula is also used. Note that our proof can be easily adapted to deal with the case of the kernel estimators of density and regression functions.

Bibliographie

[1] R. Bahadur and R. Rao (1960). On deviations of the sample mean. *Ann. Math. Statist.*, **31**, 1015–1027.

[2] O.E. Barndoff-Nielsen and D.R. Cox (1989). *Asymptotic techniques for uses in statistics.* Chapman and Hall (Londres).

[3] B. Bercu, F. Gamboa and M. Lavielle (2000). Sharp large deviations for Gaussian quadratic forms with applications. *ESAIM:P&S*, **3**, 1–29.

[4] B. Bercu and A. Rouault (2002). Sharp large deviations for the Ornstein-Uhlenbeck process. *SIAM Theory of Probability and Its Applications*, **46**, 1–19.

[5] N. Berrahou (2003). Principe de grandes déviations pour l'estimateur de la densité par la méthode des delta-suites. *C. R. Math.Acad. Sci. Paris*, **337** (5), 347–352 (in French).

[6] N.R Chaganty and J. Sethuraman (1993). Strong large deviation and local limit theorems. *Ann. Probab.*, 21, 1671–1690.

[7] H. Cramer (1970). *Random variables and probability distributions.* Cambridge University Press.

[8] A. Dembo and O. Zeitouni (1998). *Large deviations techniques and applications.* Springer, New York. Second edition.

[9] J.D. Deuschel and D.W. Stroock (1989). *Large deviations.* Boston: Academic Press.

[10] F. Gao (2003). Moderate deviations and large deviations for Kernel density estimators. *J. Theor. Probab.*, **16**, No.2, 401–418.

[11] P. Hall (1992). *The Bootstrap and Edgeworth Expansion.* Springer-Verlag, New York.

[12] J.L. Jensen (1995). *Saddlepoint approximations.* Oxford University Press.

[13] C. Joutard (2006). Sharp large deviations in nonparametric estimation. *J. Nonparametric Stat.* **18**(3), 293–306.

[14] L. Liangzhen (2006). Large deviations of kernel density estimator in $L_1(\mathbb{R}^d)$ for reversible Markov processes. *Bernoulli*, **12**(1), 65–83.
[15] D. Louani (1998). Large deviations limit theorems for the kernel density estimator. *Scand. J. of Statist.*, **25**, 243–253.
[16] D. Louani (1999). Some large deviations limit theorems in conditional non-parametric statistics. *Statistics*, **33**, 171–196.
[17] J. Worms (2001). Moderate deviations of some dependent variables. II: Some kernel estimators. *Math. Methods Stat.*, **10**, No.2, 161–193.
[18] W. Stute (1986). Conditional empirical process. *Ann. Stat.*, **14**, 891–901.