Polynomial type large deviation inequalities: M and Bayes estimators with applications to estimation for stochastic differential equations based on sampled data *

Nakahiro YOSHIDA

Graduate School of Mathematical Sciences, University of Tokyo,

3-8-1 Komaba, Meguro-ku, Tokyo 153, Japan.

e-mail: nakahiro@ms.u-tokyo.ac.jp

August 25, 2005

Analysis of the likelihood ratio is the first key step to investigate the performance of statistics. After the initiation of the local asymptotic normality by Le Cam and Hájek, a new paradime of analysis was established by Ibragimov and Has'minskii [4, 5, 6]. That is, all asymptotic properties of statistics in likelihood analysis could be reduced in the convergence of the random field formed by the likelihood ratios. This program was successfully implemented mainly for i.i.d. settings and white Gaussian noise models.

We shall recall Ibragimov-Has'minskii's result briefly. Let $\mathcal{E}^\epsilon = \{\mathcal{X}^\epsilon, \mathcal{A}^\epsilon, P^\epsilon_\theta, \theta \in \Theta\}$ be a sequence of statistical experiments with $\epsilon \in (0,1]$. Θ denotes a parameter space in $\mathbb{R}^m$. $\varphi(\epsilon)$ is a positive normalizing factor tending to zero as $\epsilon \downarrow 0$. For a $\theta_0 \in \Theta$, define a random field Z_ϵ by

$$Z_\epsilon(u) = \frac{dP^\epsilon_{\theta_0+\varphi(\epsilon)u}}{dP^\epsilon_{\theta_0}}(X^\epsilon)$$

for $u \in \mathbb{R}^m$. The following is a simplified version of their result[1]

Theorem 1. *(Ibragimov-Has'minskii [4, 5, 6]) Suppose that Z_ϵ satisfies the following conditions:*

(i) *There exist $\alpha > m$ and $k \geq \alpha$ such that for some constant C,*

$$E^\epsilon\left[|Z_\epsilon(u_2)^{1/k} - Z_\epsilon(u_1)^{1/k}|^k\right] \leq C|u_2-u_1|^\alpha \qquad (\forall u_1, u_2, \epsilon). \tag{1}$$

(ii) *For some $\gamma > 0$ and $c > 0$,*

$$E^\epsilon\left[Z_\epsilon^{1/2}(u) \geq e^{-c|u|^\gamma}\right] \leq e^{-c|u|^\gamma}. \tag{2}$$

(iii) *Finite-dimensional convergence: $Z_\epsilon \to^{d_f} Z$, where Z is a $\hat{C}(\mathbb{R}^m)$-valued random variable.*[2]

Then $(P^\epsilon)^{Z_\epsilon} \to \mathcal{L}\{Z\}$. Moreover,

$$P^\epsilon\left[\sup_{u:|u|\geq r} Z_\epsilon(u) > e^{-c_1 r^\gamma}\right] \leq e^{-c_1 r^\gamma}.$$

*The main part of this work was presented at Kyushu University in 2004. This work was in part supported by the Research Fund for Scientists of the Ministry of Science, Education and Culture No. 16500173, and by Cooperative Research Program of the Institute of Statistical Mathematics.

[1] P^ϵ denotes $P^\epsilon_{\theta_0}$. E^ϵ is the expectation with respect to P^ϵ.

[2] $\hat{C}(\mathbb{R}^m)$ is the sapce of continuous functions that tends to zero at the infinity.

If $\hat{u}$ uniquely attains the maximum of $Z(u)$, then for any sequence of the maximum likelihood estimator $\hat{\theta}_\epsilon$ for θ, $\hat{u}_\epsilon := \varphi(\epsilon)^{-1}(\hat{\theta}_\epsilon - \theta_0) \to^d \hat{u}$ and moreover

$$E^\epsilon\left[f(\hat{u}_\epsilon)\right] \to E\left[f(\hat{u})\right]$$

for every $f \in C_\uparrow(\mathbb{R}^m)$.[3]

Convergence of moments or the estimate of the error probability of an estimator plays an essential role in key steps of theoretical statistics. For mean bias correction to an estimator, we need the exsistence of the mean of it. The expected value of the plug-in functional

$$E^\epsilon\left[\log\ \text{likelihood}\ (\hat{u}_\epsilon)\right]$$

is necessary in prediction theory and in construction of information criteria. For example, the number of parameters appering in AIC as the correction term is nothing but the mean square of $\hat{u}_\epsilon$ [4] It is impossible to develop the higer-order asymptotic theory without the following type estimate:

$$\mathbf{P}\left[|\hat{\theta}_n - \theta_0| > n^{-\beta}\right] \leq n^{-N}$$

for $\beta < 1/2$. Also, for the same estimates for Bayes estimators, we need an estimate of the large deviation probability for the likelihood ratio random field. The necessity of the polynomial type large deviation inequalities in the theory of asymptotic expansion of the estimators for stochastic processes motivates this article (Yoshida [29], Sakamoto and Yoshida [19, 20]).

Kutoyants found that Ibragimov-Has'minskii's scheme could work for stochastic processes including diffusion type processes and point processes. Among his many results, Kutoyants established a complete theory for processes with small diffusions (Kutoyants [9]). Many applications of this methodology were also presented in Kutoyants [10]. We should note that the Ibragimov-Has'minskii-Kutoyants' theory was effectively used in derivation of asymptotic expansions ([26, 28]).

We can use (1) or another simpler continuity inequality in most cases in practice. Contrarilly, as the core of the theory, the large deviation inequality (2) is not easy to obtain for most of stochastic processes, even for nonlinear ergodic diffusions; asymptotic theory for small perturbed systems is special in this sense. As seen in Kutoyants [9], for ergodic diffusions, is necessary an explicit expression of the moment generating function of a kind of deviation in the statistical model, although it implies very strong results such as the convergence of moments once it is obtained; strong assumptions possibly produce strong results! Without such a strong assumption, the convergence of statistical random fields was proved in [25], however without moment convergence.

It is an important observation, as some of statisticians have been aware and really it was implicitly written in Ibragimov and Has'minskii's papers, that the exponential type large deviation inequality like (2) is much stronger than our use. The polynomial type large deviation inequality is sufficient to develop a theory. Here the polynomial type large deviation inequality means:

$$P^\epsilon\left[\sup_{u:|u|\geq r} Z_\epsilon(u) \geq r^{-N}\right] \leq \frac{C_N}{r^N} \qquad (r > 0)$$

because the rate of the convergence of the probability is of or faster than a polynomial order. Some exponential in place of r^{-N} on the left-hand side is very often possible.

Kutoyants [12] presented a polynomial type large deviation inequality for one-dimensional diffusions by means of the local time. The aim of the present article is to provide a polynomial type large deviation inequality in a more abstract, general setting of the partially locally asymptotically quadratic (PLAQ) sequence of experiments (without any special properties belonging to diffusion processes). By our results, for stochastic processes including nonlinear non-Gaussian time series models and semimartingales as well as multi-dimensional

[3]The set of continuous functions on $\mathbb{R}^m$ of at most polynomial growth.

[4]In the literature of asymptotic statistics, unformunately, such an expectation has been very often assumed to exist without any mathematical backing. Rigorous treatment of this problem under a certain integrability assumption can be seen, e.g., in Uchida and Yoshida [22].

diffusion processes, it is possible to obtain new convergence results, e.g., the convergence of moments of the M-estimator, and the asymptotic normality of the Bayes estimator and convergence of moments of it. Our results also provide the same consequences even for estimators based on sampled data from diffusions with/without jumps[5]. The grading of the parameters appearing later is necessary for this reason.

In this talk, we discuss asymptotic properties (consistency, asymptotic normality and convergence of moments) of M and Bayesian type estimators in estimation of a stochastic differential equation based on sampled data.

References

[1] Adams, R. A: Sobolev spaces. Pure and Applied Mathematics, Vol. 65. Academic Press, New York-London, 1975.

[2] Doukhan, P.: Mixing: properties and examples. Lect. Notes in Statisit. **85**, Springer 1995

[3] Doukhan, P., Oppenheim, G., Taqqu, M. S. (eds.): Theory and applications of long-range dependence. Birkhauser Boston, Inc., Boston, MA, 2003

[4] Ibragimov, I.A., Has'minskii, R.Z.: The asymptotic behavior of certain statistical estimates in the smooth case. I. Investigation of the likelihood ratio. (Russian) Teor. Verojatnost. i Primenen. 17 (1972), 469–486.

[5] Ibragimov, I.A., Has'minskii, R.Z.: Asymptotic behavior of certain statistical estimates. II. Limit theorems for a posteriori density and for Bayesian estimates. (Russian) Teor. Verojatnost. i Primenen. 18 (1973), 78–93.

[6] Ibragimov, I.A., Has'minskii, R.Z.: Statistical estimation: asymptotic theory. Springer, New York 1981

[7] Kessler, M.: Estimation of diffusion processes from discrete observations. Scand. J. Statist. **24**, 211-229 (1997)

[8] Kusuoka, S., Yoshida, N.: Malliavin calculus, geometric mixing, and expansion of diffusion functionals. Prob. Theory Related Fields **116**, 457-484 (2000)

[9] Kutoyants, Yu. A.: Parameter estimation for stochastic processes. Translated and edited by B.L.S.Prakasa Rao, Berlin: Herdermann 1984

[10] Kutoyants, Yu.: Identification of dynamical systems with small noise. Dordrecht Boston London: Kluwer 1994

[11] Kutoyants, Yu. A.: Statistical inference for spatial Poisson processes. Lect. Notes in Statistics **134**, Berlin Heiderberg New York London Paris Tokyo Hong Kong: Springer 1998

[12] Kutoyants, Yu. A.: Statistical inference for ergodic diffusion processes. Springer Series in Statistics. Springer-Verlag London, Ltd., London, 2004

[13] Meyn, S. P., Tweedie, R. L.: Stability of Markovian processes. I. Criteria for discrete-time chains. Adv. in Appl. Probab. **24**, 542–574 (1992)

[14] Meyn, S. P., Tweedie, R. L.: Stability of Markovian processes. II. Continuous-time processes and sampled chains. Adv. in Appl. Probab. **25**, 487–517 (1993)

[15] Meyn, S. P., Tweedie, R. L.: Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes. Adv. in Appl. Probab. **25**, 518–548 (1993)

[16] Prakasa-Rao, B. L. S.: Asymptotic theory for non-linear least square estimator for diffusion processes, Math. Operationsforsch. Statist. Ser. Stat. **14**, 195-209 (1983)

[5] The asymptotic normality of the M-estimator is already known for jump-diffusions; see Shimizu and Yoshida [21].

[17] Prakasa-Rao, B. L. S.: Statistical inference from sampled data for stochastic processes, Contemp. Math. **80**, 249-284 (1988)

[18] Rio, E.: Inegalites de moments pour les suites stationnaires et fortement melangeantes. (French) [Moment inequalities for stationary strongly mixing sequences] C. R. Acad. Sci. Paris Ser. I Math. 318, 355–360.(1994)

[19] Sakamoto, Y., Yoshida, N.: Higher order asymptotic expansions for a functional of a mixing process and applications to diffusion functionals. unpublished manuscript (1999)

[20] Sakamoto, Y., Yoshida, N.: Asymptotic expansion formulas for functionals of epsilon-Markov processes with a mixing property. Annals of the Institute of Statistical Mathematics, 56, 545-597 (2004)

[21] Shimizu, Y.; Yoshida, N.: Estimation of diffusion processes with jumps from discrete observations. (2002) preprint, to appear in Statistical Inference for Stochastic Processes

[22] Uchida, M., Yoshida, N.: Information criteria in model selection for mixing processes Statistical Inference for Stochastic Processes, 4, 73-98 (2001)

[23] Veretennikov, A. Yu. : Bounds for the mixing rate in the theory of stochastic equations. Theory Probab. Appl. **32**, 273-281 (1987)

[24] Veretennikov, A. Yu. : On polynomial mixing bounds for stochastic differential equations. Stoch. Processes Appl. **70**, 115-127 (1997)

[25] Yoshida, N.: Asymptotic behavior of M-estimator and related random field for diffusion process. Annals of the Institute of Statistical Mathematics, **42**, No. 2, 221-251 (1990)

[26] Yoshida, N: Asymptotic expansion for small diffusions via the theory of Malliavin-Watanabe. Prob. Theory Related Fields. **92**, 275-311 (1992)

[27] Yoshida, N: Estimation for diffusion processes from discrete observation. J. Multivar. Anal. **41**, 220-242 (1992)

[28] Yoshida,N.: Asymptotic expansion of Bayes estimators for small diffusions. Probab. Theory Relat. Fields **95**, 429-450 (1993)

[29] Yoshida, N.: Malliavin calculus and asymptotic expansion for martingales. Probab. Theory Relat. Fields **109**, 301-342 (1997)

[30] Yoshida, N.: Partial mixing and conditional Edgeworth expansion. Probab. Theory Related Fields 129, 559-624 (2004)

[31] Yoshida, N.: General M-estimation for stochastic differential equation with jumps by sampled data. (2005) in preparation

[32] Yoshida, N.: Asymptotic expansion for general M-estimation for stochastic differential equation with jumps. *Zenkin Tenkai 2003*, Hiroshima International University (2003)