Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

On the asymptotic properties of Bayes-type estimators with general loss functions

Teppei Ogihara *

The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190–8562, Japan
PRESTO, Japan Science and Technology Agency, Japan
School of Multidisciplinary Sciences, SOKENDAI (The Graduate University for Advanced Studies), Japan

A R T I C L E I N F O

Article history:
Received 21 June 2017
Received in revised form 7 March 2018
Accepted 6 June 2018
Available online xxxx

Keywords:
Asymptotic normality
Bayes-type estimation
Convergence of moments
Diffusion processes
Jump diffusion processes
Polynomial-type large deviation inequalities

A B S T R A C T

We study the asymptotic behavior of Bayes-type estimators and give sufficient conditions to obtain the asymptotic limit distribution of the estimation error. We assume so called polynomial-type large deviation inequalities and prove the asymptotic equivalence of the estimation errors of Bayes-type and M-estimators by virtue of Ibragimov–Has'minskiĭ theory. The results can be applied to several cases such as diffusion processes and jump diffusion processes. In this paper, we focus on the application to ergodic diffusion processes and ergodic jump diffusion processes, demonstrating asymptotic normality and convergence of moments for Bayes-type estimators with general loss functions.

1. Introduction

The theory of random fields of likelihood ratios is a powerful tool to investigate the asymptotic behavior of Bayes-type estimators. This theory was initiated by Ibragimov and Has'minskiĭ (1972, 1973, 1981), who applied it to statistical models of regular independent and identically distributed (i.i.d.) observations and Gaussian white-noise models. After that, Kutoyants applied Ibragimov–Has'minskiĭ theory to other statistical models, including diffusion-type and point processes. See Kutoyants (1984, 1994) for the details. Yoshida (2006, 2011) introduced polynomial-type large deviation inequalities and gave a scheme to obtain asymptotic properties of the M-estimator and the Bayes-type estimator under certain moment conditions for the contrast function and its derivatives. This scheme can be applied to many classes of statistical model and gives the consistency, asymptotic (mixed) normality and convergence of moments for quasi-maximum likelihood estimators and Bayes-type estimators. See Yoshida (2006, 2011) for an application to ergodic diffusion processes, Ogihara and Yoshida (2011) for ergodic jump diffusion processes, Masuda (2010) for Ornstein–Uhlenbeck processes driven by heavy-tailed symmetric Lévy processes, Uchida and Yoshida (2013) for diffusion processes observed in a fixed interval, Ogihara and Yoshida (2014) for diffusion processes with nonsynchronous observations.

One of the most important motivations for the study of quasi-maximum likelihood estimators and Bayes-type estimators is that these estimators are asymptotically efficient in several models. For statistical models of regular i.i.d. observations, we can obtain minimax theorems for estimation errors, and hence can define the concept of asymptotic efficiency of estimators. Since the maximum likelihood estimator and the Bayes estimator attain minimax bounds, these estimators are

* Correspondence to: The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190–8562, Japan.
E-mail address: ogihara@ism.ac.jp.

https://doi.org/10.1016/j.jspi.2018.06.001
0378-3758/

asymptotically efficient. See Ibragimov and Has'minskiĭ (1981) for details. We also have the asymptotic efficiency of quasi-maximum likelihood estimators and Bayes-type estimators for some statistical models of diffusion processes with discrete observations. Jeganathan (1983) extended the minimax theorems to statistical models that satisfy the local asymptotic mixed normality (LAMN) property. Moreover, Gobet (2001) proved the LAMN property for models of diffusion processes observed in a fixed interval and the estimators proposed in Genon-Catalot and Jacod (1994) have asymptotic minimal variance. Gobet (2002) proved local asymptotic normality for statistical models of ergodic diffusion processes, and Ogihara (2015) gives the LAMN property and asymptotic efficiency of the quasi-maximum likelihood estimator and the Bayes-type estimator proposed in Ogihara and Yoshida (2014) for diffusion processes with nonsynchronous observations in a fixed interval.

Convergence of moments is another important asymptotic property used in the study of asymptotic expansions of estimators and information criteria. Yoshida (2006, 2011) applied polynomial-type large deviation inequalities to the Bayes-type estimator with a quadratic loss function. This estimator can be obtained as a ratio of certain integrals with respect to the parameter if the loss function is quadratic, and hence its asymptotic behavior can be specified by using polynomial-type large deviation inequalities. Along another line, Ibragimov and Has'minskiĭ (1981) considered the asymptotic behavior of Bayes-type estimators with a wider class of loss functions. Though their results apply mainly to models of i.i.d. observations, we can apply their ideas to models satisfying polynomial-type large deviation inequalities. This allows us to derive asymptotic properties of Bayes-type estimators with a wider class of loss functions, which is the subject of this paper.

In this paper, we extend the results of asymptotic normality and convergence of moments of a Bayes-type estimator with a quadratic loss function given in Yoshida (2011) to a Bayes-type estimator with a general loss function. By changing the loss function, we can set a large penalty when the estimator is far from the true parameter value. Then, the Bayes-type estimator becomes sensitive to tail risk. Conversely, by reducing penalty for large errors, we obtain an estimator which is tolerant of such risk compared to the case of the quadratic loss function. If the loss function is quadratic, the Bayes-type estimator can be written as a ratio of some integrals with respect to parameters, as seen at the beginning of Section 5. In this case, the integrals can be estimated by dividing the domains of integration into ring domains and using tail probability estimates, such as [A1-k, p] below. However, we cannot use this approach to obtain asymptotic results for a Bayes-type estimator with a general loss function because we do not have an expression with integrals ratio. Here, we apply the idea used for the derivation of Theorem 8.2 in Ibragimov and Has'minskiĭ (1981) and prove the asymptotic equivalence of the estimation error of the Bayes-type and M-estimators. By doing this, we obtain the asymptotic distribution of the estimation error of the Bayes-type estimator if we have the asymptotic distribution of the M-estimator. In particular, we see that the asymptotic distribution for the Bayes-type estimator does not depend on the loss function. These results can be applied to ergodic diffusion processes, diffusion processes observed in a fixed interval, ergodic jump diffusion processes and diffusion processes with nonsynchronous observation; and we obtain asymptotic (mixed) normality and convergence of moments for Bayes-type estimators with general loss functions. We focus on the application to ergodic diffusion processes and ergodic jump diffusion processes in this paper.

This paper is organized as follows. In Section 2, we present the theory of random fields of likelihood ratios and polynomial-type large deviation inequalities, and we state our main results. Section 3 is devoted to the application of the main results to ergodic diffusion processes. The application to ergodic jump diffusion processes is studied in Section 4. The proofs are collected in Section 5.

2. Main results

We first introduce the Ibragimov–Has'minskiĭ theory of random fields of likelihood ratios. In several statistical models, the optimal convergence rates of parameters depend on the specific parameter. For example, in ergodic diffusion processes, the convergence rate of a quasi-maximum likelihood estimator for parameters used in diffusion coefficients is different from that for parameters used in drift coefficients, as will be shown in Section 3. Therefore, it is useful to consider a parameter space as a product space of individual parameter spaces and estimate these parameter spaces separately.

For $K \in \mathbb{N}$, let the parameter space $\Theta_k \subset \mathbb{R}^{d_k}$ be a bounded open set $(1 \le k \le K)$, and let $\Theta = \Theta_1 \times \Theta_2 \times \cdots \times \Theta_K \subset \mathbb{R}^d$, where $d = \sum_{k=1}^K d_k$. If $K \ge 2$, we assume Θ_k is a convex set for $1 \le k \le K$. Let $(\mathcal{X}, \mathcal{A}, \{P_\theta\}_\theta)$ be a statistical experiment. Observed data is an element of $\mathcal{X}$. For example, we set $\mathcal{X} = \mathbb{R}^m$ with some $m \in \mathbb{N}$ for discrete observations of diffusion processes, or $\mathcal{X} = C([0,\infty))$ for continuous observations. If a sequence $(\mathcal{X}_n, \mathcal{A}_n, \{P_{\theta,n}\}_\theta)_{n\in\mathbb{N}}$ of experiments are given, we regard it as one experiment $(\mathcal{X}, \mathcal{A}, \{P_\theta\}_\theta)$ by considering the product space: $\mathcal{X} = \prod_{n\in\mathbb{N}}\mathcal{X}_n$, $\mathcal{A} = \prod_{n\in\mathbb{N}}\mathcal{A}_n$, $P_\theta = \prod_{n\in\mathbb{N}}P_{n,\theta}$. Thus, we can deal with the sequence of observations. Let a contrast function $H_T : \Theta \times \mathcal{X} \to \mathbb{R}$ be a C^3 function with respect to $\theta \in \Theta$ and extended as a function on $\mathrm{clos}(\Theta) \times \mathcal{X}$ that is continuous with respect to $\theta \in \mathrm{clos}(\Theta)$ for a positive real number T, where $\mathrm{clos}(\Theta)$ represents the closure of Θ. By considering continuous index T, we can deal with a statistical model for continuous observations of diffusion process X_t on $[0, T]$ and its limit that $T \to \infty$. If contrast functions $\{H_n\}_{n\in\mathbb{N}}$ for discrete observations are given, we identify it with $\{H_T\}_{T>0}$ by $H_T = H_{[T]+1}$ for $T > 0$, where $[x]$ is the truncation of a real number x. We apply the same argument for Z_T^k, b_T^k and a_T^k below. We often omit the dependence of data for $H_T(\theta, x)$ and denote it by $H_T(\theta)$.

Let $\overline{\Theta}_k = \Theta_k \times \Theta_{k+1} \times \cdots \times \Theta_K$, $\overline{\theta}_k = (\theta_k, \theta_{k+1}, \ldots, \theta_K)$ and $\underline{\theta}_k = (\theta_1, \ldots, \theta_k)$ for any value $\theta = (\theta_1, \ldots, \theta_K) \in \Theta$. Let a_T^k be an invertible matrix of size d_k, $b_T^k = (\lambda_{\min}((a_T^k)^\top a_T^k))^{-1} \to \infty$ as $T \to \infty$ and $\lambda_{\max}((a_T^k)^\top a_T^k) \le C_1 (b_T^k)^{-1}$ for some $C_1 > 0$, where $\top$ indicates matrix transposition and $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ represent, respectively, the maximum and minimum eigenvalues of a matrix A.

Please cite this article in press as: Ogihara T., On the asymptotic properties of Bayes-type estimators with general loss functions. J. Statist. Plann. Inference (2018), https://doi.org/10.1016/j.jspi.2018.06.001.