J. Japan Statist. Soc.
Vol. 38 No. 1 2008 51–63

ON REGULAR AND SINGULAR ESTIMATION FOR ERGODIC DIFFUSION

Yury A. Kutoyants*

The asymptotic properties of the maximum likelihood and bayesian estimators of finite dimensional parameters of any statistical model depend strongly on the regularity conditions. It is well-known that if these conditions are fulfilled then the estimators are consistent, asymptotically normal and asymptotically efficient. These regularity conditions are of the following type: the model is sufficiently smooth w.r.t. the unknown parameter, the Fisher information is a positive continuous function, the model is correct and identifiable and the unknown parameter is an interior point of the parameter set. In this work we present a review of the properties of these estimators in the situations when these regularity conditions are not fulfilled. The presented results allow us to better understand the role of regularity conditions. As the model of observations we consider the one-dimensional ergodic diffusion process.

Key words and phrases: Asymptotic properties, maximum likelihood estimators, misspecified, misspecified models, non identifiable model, regularity conditions, singular estimation problem.

1. Introduction

This work is devoted to the clarification of the role of the set of regularity conditions in the problems of parameter estimation. Particularly, we are interested by the following question: how will the asymptotic properties of the classical estimators (maximum likelihood and bayesian) change if we change one of these conditions? The exposition is illustrated on the model of the ergodic diffusion process, but we will have similar effects if we take any other well-known model of observations. First, we review the regularity conditions and properties of estimators in the i.i.d. case, and then we consider the model of the ergodic diffusion process in detail.

Let $X_1, \ldots, X_n$ be independently and identically distributed random variables with the density function $f(\vartheta, x)$. The parameter $\vartheta \in \Theta = (\alpha, \beta)$ is supposed to be unknown and we have to estimate it and to describe the properties of estimators in the asymptotic of large samples ($n \to \infty$). Let us introduce the likelihood function $L_n(\vartheta, X^n)$, the maximum likelihood estimator (MLE) $\hat{\vartheta}_n$ and the bayesian estimator (BE) $\tilde{\vartheta}_n$ (for quadratic loss function and density a priori $p(\cdot)$)

$$L_n(\hat{\vartheta}_n, X^n) = \sup_{\vartheta \in \Theta} L_n(\vartheta, X^n), \qquad \tilde{\vartheta}_n = \frac{\int_\alpha^\beta \theta p(\theta) L_n(\vartheta, X^n) \mathrm{d}\theta}{\int_\alpha^\beta p(\theta) L_n(\vartheta, X^n) \mathrm{d}\theta}.$$

It is well-known that if the conditions of regularity are fulfilled then these esti-

Accepted September 29, 2007.
*Laboratory of Statistics and Processes, University of Maine, Le Mans, 72085, France.

mators are consistent, asymptotically normal

$$\sqrt{n}(\hat\vartheta_n-\vartheta)\Rightarrow\mathcal{N}(0,\mathrm{I}(\vartheta)^{-1}),\qquad \sqrt{n}(\tilde\vartheta_n-\vartheta)\Rightarrow\mathcal{N}(0,\mathrm{I}(\vartheta)^{-1}),\tag{1.1}$$

and asymptotically efficient. Here $\mathrm{I}(\vartheta)$ is the Fisher information. The proofs are found in any book on asymptotical statistics, e.g., Cramér (1946), Ibragimov and Khasminskii (1981), van der Vaart (1998) etc. These regularity conditions can be roughly described as follows

- The density $f(x)$ of the observed r.v.'s belongs to the prescribed parametric family, i.e., there exists a value $\vartheta_0\in\Theta$ such that $f(x)=f(\vartheta_0,x)$.
- The function $f(\vartheta,x)$ is one or more times differentiable w.r.t. ϑ with certain majoration of the derivatives.
- The Fisher information $\mathrm{I}(\vartheta)$ is a positive function.
- The Fisher information $\mathrm{I}(\vartheta)$ is a continuous function.
- The model is identifiable, i.e., $f(\vartheta_1,x)\neq f(\vartheta_2,x)$ if $\vartheta_1\neq\vartheta_2$.
- The true value ϑ_0 is an interior point of the set Θ, i.e., $\vartheta_0\neq\alpha$ and $\vartheta_0\neq\beta$.
- The observed variables are indeed independent.

Note that if at least one of these conditions is not fulfilled, then we have no more asymptotic normality (1.1). The behavior of the estimators in the situations without regularity conditions (more precisely, with different conditions) always attracted the attention of statisticians. There is a huge literature concerning the properties of estimators in non regular situations, but usually these conditions are studied separately. Here we treat all of them together.

In the present work we describe in detail the properties of the MLE and BE of a one-dimensional parameter in the situations when the similar regularity conditions are rejected one by one. This approach, by our mind allows a better understanding of the particular role of each of these conditions. As a model of observations, we take the one-dimensional ergodic diffusion process

$$\mathrm{d}X_t=S(\vartheta,X_t)\mathrm{d}t+\sigma(X_t)\mathrm{d}W_t,\qquad X_0,\quad 0\le t\le T,$$

and we want to see how the properties of the MLE and BE depend on the properties of the function $S(\vartheta,x)$. Surely, the properties of the same estimators, but for the different statistical models, will be quite close to the case mentioned here for ergodic diffusion. The regularity conditions in parameter estimation problems for a wide class of stochastic processes (with discrete and continuous time) can be found in Taniguchi and Kakizawa (2000).

2. Regular case

Let us suppose that we observe in continuous time a trajectory $X^T=\{X_t,0\le t\le T\}$ of a diffusion process

$$\mathrm{d}X_t=S_*(X_t)\mathrm{d}t+\sigma_*(X_t)\mathrm{d}W_t,\qquad X_0,\quad 0\le t\le T,\tag{2.1}$$

where $W_t,0\le t\le T$ is the standard Wiener process, $S_*(x)$ and $\sigma_*(x)^2$ are the trend and diffusion coefficients and X_0 is the initial value. Statistician can suppose that the diffusion coefficient is a known function, because $\sigma_*(x)^2$ can be estimated without error by continuous time observations, and in a wide class of problems it can be supposed as well that the trend coefficient belongs to some parametric class of functions, i.e., $S_*(x) = S(\vartheta, x)$ where $S(\vartheta, x)$ is a known function depending on the unknown parameter $\vartheta \in \Theta$.

Therefore we obtain the problem of estimation of the parameter ϑ by the observations X^T of the stochastic process

$$\mathrm{d}X_t = S(\vartheta, X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t, \qquad X_0, \quad 0 \le t \le T. \tag{2.2}$$

We suppose that equation (2.2) has a unique weak solution and the measures $\{\boldsymbol{P}_\vartheta^{(T)}, \vartheta \in \Theta\}$ are equivalent (see the conditions in Liptser and Shiryayev (2001)). Here $\boldsymbol{P}_\vartheta^{(T)}$ is the measure induced by the process (2.2) in the measurable space $(\mathcal{C}_T, \mathcal{B}_T)$ of continuous on $[0, T]$ functions (space of its realizations). We suppose that

$$\lim_{|x|\to\infty} \operatorname{sgn}(x)\frac{S(\vartheta, x)}{\sigma(x)^2} \le -\gamma, \qquad \inf_x \sigma(x)^2 \ge \kappa \tag{2.3}$$

where $\gamma > 0$, $\kappa > 0$ do not depend on ϑ. By this condition the process (2.2) is positive recurrent and has ergodic properties with invariant density function

$$f(\vartheta, x) = \frac{1}{G(\vartheta)\sigma(x)^2} \exp\left\{2\int_0^x \frac{S(\vartheta, y)}{\sigma(y)^2}\mathrm{d}y\right\},$$

i.e., for any integrable function $h(x)$ the law of large numbers holds: with probability 1

$$\frac{1}{T}\int_0^T h(X_t)\mathrm{d}t \to \boldsymbol{E}_\vartheta h(\xi) = \int_R h(x) f(\vartheta, x)\mathrm{d}x.$$

Here $G(\vartheta)$ is the normalizing constant and ξ is the random variable with the stationary density function $f(\vartheta, x)$.

The likelihood ratio function is

$$L(\vartheta, X^T) = \exp\left\{\int_0^T \frac{S(\vartheta, X_t)}{\sigma(X_t)^2}\mathrm{d}X_t - \frac{1}{2}\int_0^T \frac{S(\vartheta, X_t)^2}{\sigma(X_t)^2}\mathrm{d}t\right\}$$

and the MLE $\hat{\vartheta}_T$ and BE $\tilde{\vartheta}_T$ for quadratic loss function and prior density $p(\theta)$, $\theta \in \Theta$ are

$$L(\hat{\vartheta}_T, X^T) = \sup_{\theta\in\Theta} L(\theta, X^T), \qquad \tilde{\vartheta}_T = \int_\alpha^\beta \theta p(\theta \mid X^T)\mathrm{d}\theta,$$

where $p(\vartheta \mid X^T)$ is density a posteriori calculated as usual by the formula

$$p(\vartheta \mid X^T) = \frac{p(\vartheta)L(\vartheta, X^T)}{\int_\alpha^\beta p(\theta)L(\theta, X^T)\mathrm{d}\theta}.$$

Below we suppose that the prior density $p(\theta)$ is positive on Θ continuous function.

Let us introduce the following

Regularity Conditions.

(i) The trend coefficient $S_*(x)$ of the observed process (2.1) indeed belongs to the parametric family $\{S(\vartheta, x), \vartheta \in \Theta\}$, i.e., there exists $\vartheta_0 \in \Theta$ (true value) such that $S_*(x) = S(\vartheta_0, x)$ and $\sigma(x) = \sigma_*(x)$.

(ii) The following condition of identifiability is fulfilled: for any $\nu > 0$

$$\inf_{|\theta-\vartheta_0|>\nu} \boldsymbol{E}_{\vartheta_0} \left(\frac{S(\theta,\xi) - S(\vartheta_0,\xi)}{\sigma(\xi)}\right)^2 > 0. \tag{2.4}$$

(iii) The Fisher information is

$$\mathrm{I}(\vartheta) = \boldsymbol{E}_\vartheta \left(\frac{\dot S(\vartheta,\xi)}{\sigma(\xi)}\right)^2 > 0,$$

where $\dot S(\vartheta,\xi)$ is differentiated w.r.t. ϑ.

(iv) The function $S(\vartheta, x)$ has two continuous bounded (in ϑ and x) derivatives w.r.t. ϑ and $\sigma(x)$ is bounded.

(v) The parameter ϑ_0 is an interior point of the set $\Theta = (\alpha, \beta)$.

(vi) The observed diffusion process is ergodic (recurrent positive), i.e., it has finite unique invariant measure.

Note that by the condition (iv), the Fisher information $\mathrm{I}(\vartheta)$ is a continuous function of ϑ.

These regularity conditions allow us to describe the asymptotic ($T \to \infty$) properties of estimators.

Theorem 1. *Let the conditions of regularity be fulfilled, then the MLE and BE are consistent, asymptotically normal*

$$\sqrt{T}(\hat\vartheta_T - \vartheta_0) \Rightarrow \mathcal{N}(0, \mathrm{I}(\vartheta_0)^{-1}), \qquad \sqrt{T}(\tilde\vartheta_T - \vartheta_0) \Rightarrow \mathcal{N}(0, \mathrm{I}(\vartheta_0)^{-1}),$$

and the moments of these estimators converge and they are asymptotically efficient.

The proof can be found in Kutoyants (2004). It is based on the general result by Ibragimov and Khasminskii (1981). The regularity conditions allow us to check the conditions of the Theorems 3.1.1 (MLE) and 3.2.1 (BE) in Ibragimov and Khasminskii (1981) and therefore to provide the mentioned properties of the MLE and BE. Here we recall the main steps of the proofs. Let us denote the normalized likelihood ratio process by $Z_T(u)$ and the limit process by $Z(u)$:

$$Z_T(u) = \frac{L\left(\vartheta_0 + \frac{u}{\sqrt{T}}, X^T\right)}{L(\vartheta_0, X^T)}, \qquad Z(u) = \exp\left\{u\zeta(\vartheta_0) - \frac{u^2}{2}\mathrm{I}(\vartheta_0)\right\} \tag{2.5}$$

where $\zeta(\vartheta_0) \sim \mathcal{N}(0, \mathrm{I}(\vartheta_0))$. Suppose that we already proved the weak convergence of the stochastic process

$$Z_T(\cdot) \Rightarrow Z(\cdot) \tag{2.6}$$

in the space of continuous on R functions vanishing to infinity. Then the asymptotic normality of the MLE can be obtained in the following way.

$$
\begin{aligned}
&\boldsymbol{P}_{\vartheta_0}\{\sqrt{T}(\hat\vartheta_T-\vartheta_0)<x\}\\
&\quad=\boldsymbol{P}_{\vartheta_0}\left\{\sup_{\sqrt{T}(\theta-\vartheta_0)<x} L(\vartheta,X^T)>\sup_{\sqrt{T}(\theta-\vartheta_0)\ge x} L(\vartheta,X^T)\right\}\\
&\quad=\boldsymbol{P}_{\vartheta_0}\left\{\sup_{u<x} Z_T(u)>\sup_{u\ge x} Z_T(u)\right\}\to\boldsymbol{P}_{\vartheta_0}\left\{\sup_{u<x} Z(u)>\sup_{u\ge x} Z(u)\right\}\\
&\quad=\boldsymbol{P}_{\vartheta_0}\left(\frac{\zeta(\vartheta_0)}{\mathrm{I}(\vartheta_0)}<x\right),\qquad\text{i.e.}\quad \sqrt{T}(\hat\vartheta_T-\vartheta_0)\Rightarrow\mathcal{N}(0,\mathrm{I}(\vartheta_0)^{-1}).
\end{aligned}
\tag{2.7}
$$

For the bayesian estimators we first change the variable $\theta=\vartheta_0+u/\sqrt{T}\equiv\vartheta_u$ in the integrals below

$$
\tilde\vartheta_T=\frac{\int_\alpha^\beta \theta p(\theta)L(\theta,X^T)\mathrm{d}\theta}{\int_\alpha^\beta p(\theta)L(\theta,X^T)\mathrm{d}\theta}=\vartheta_0+\frac{1}{\sqrt{T}}\,\frac{\int_{U_T} up(\vartheta_u)L(\vartheta_u,X^T)\mathrm{d}u}{\int_{U_T} p(\vartheta_u)L(\vartheta_u,X^T)\mathrm{d}u},
$$

where $U_T=(\sqrt{T}(\alpha-\vartheta_0),\sqrt{T}(\beta-\vartheta_0))$. Then using the convergence $p(\vartheta_u)\to p(\vartheta_0)$ and (2.3) and the notation $Z_T(u)=L(\vartheta_u,X^T)/L(\vartheta_0,X^T)$ we can write

$$
\begin{aligned}
\boldsymbol{P}_{\vartheta_0}\{\sqrt{T}(\tilde\vartheta_T-\vartheta_0)<x\}&=\boldsymbol{P}_{\vartheta_0}\left\{\frac{\int_{U_T} up(\vartheta_u)Z_T(u)\mathrm{d}u}{\int_{U_T} p(\vartheta_u)Z_T(u)\mathrm{d}u}<x\right\}\\
&\to\boldsymbol{P}_{\vartheta_0}\left\{\frac{\int_R uZ(u)\mathrm{d}u}{\int_R Z(u)\mathrm{d}u}<x\right\}=\boldsymbol{P}_{\vartheta_0}\left(\frac{\zeta(\vartheta_0)}{\mathrm{I}(\vartheta_0)}<x\right)
\end{aligned}
\tag{2.8}
$$

and elementary calculus yields the equality

$$
\int_R uZ(u)\mathrm{d}u=\int_R ue^{u\zeta(\vartheta_0)-(u^2/2)\mathrm{I}(\vartheta_0)}\mathrm{d}u=\frac{\zeta(\vartheta_0)}{\mathrm{I}(\vartheta_0)}\int_R Z(u)\mathrm{d}u.
$$

We do not discuss here the technical details of the proof of the convergence (2.7), (2.8). The reader can find it in Ibragimov and Khasminskii (1981). We mention here their method because we use it below in almost all non regular estimation problems by proving the corresponding convergence of the likelihood ratio process $Z_T(\cdot)$ to the limit $Z(\cdot)$ which is different in each problem.

3. Misspecified model

Suppose now that the parametric model $\{S(\vartheta,x),\sigma(x)\}$ does not correspond to the observed process (1.1), but the statistician nevertheless uses this model to estimate the parameter ϑ (no true model). Note that in any applied problem the mathematical model and the real model of data never coincide. Sometimes this difference is small but in some situations this difference can become important and we have to see what happens with the estimators. Therefore we are in the situation when the condition (i) is not fulfilled and we suppose that the other conditions of regularity are fulfilled with corresponding modification of the notation. It can be shown that the MLE and BE converge to the value

$$\vartheta_* = \arg \inf_{\theta\in\Theta} \boldsymbol{E}_* \left(\frac{S(\theta,\xi_*)-S_*(\xi_*)}{\sigma(\xi_*)}\right)^2,$$

where ξ_* has an invariant distribution of the ergodic diffusion (1.1). Moreover they are asymptotically normal

$$\sqrt{T}(\hat{\vartheta}_T-\vartheta_*) \Rightarrow \mathcal{N}(0,D_*^2), \qquad \sqrt{T}(\tilde{\vartheta}_T-\vartheta_*) \Rightarrow \mathcal{N}(0,D_*^2).$$

As usual in classical statistics, the MLE and BE converge to the value ϑ_* which minimizes the Kullback-Leibler distance. Here this corresponds to the best mean square choice of the parametric model. It is interesting to note that if $S_*(x) = S(\vartheta_0,x)$ and $\sigma_*(x)\neq\sigma(x)$, then nevertheless $\vartheta_*=\vartheta_0$, i.e., both estimators are consistent, but not asymptotically efficient, $D_*^2\neq \mathrm{I}(\vartheta_0)^{-1}$ (see McKeague (1984), Yoshida (1990) or Kutoyants (2004) Section 2.6.1). Otherwise, these estimators are not consistent because the true value does not exist. The process (1.1) can be written as a contaminated version of (2.1), i.e.,

$$\mathrm{d}X_t = S(\vartheta_0,X_t)\mathrm{d}t + h(X_t)\mathrm{d}t + \sigma_*(X_t)\mathrm{d}W_t, \qquad X_0, \quad 0\leq t\leq T, \tag{3.1}$$

where $h(x)$ is an unknown function. Hence

$$\vartheta_* = \arg \inf_{\theta\in\Theta} \boldsymbol{E}_* \left(\frac{S(\theta,\xi_*)-S(\vartheta_0,\xi_*)-h(\xi_*)}{\sigma(\xi_*)}\right)^2.$$

There exists a class of parameter estimation problems with contamination which nevertheless make it possible to have a consistent MLE and BE. We discuss here two possibilities. The first one is applied if we know the support A of the function $h(x)$ and we know as well that the model is identifiable in the following sense:

$$\vartheta_0 = \arg \inf_{\theta\in\Theta} \boldsymbol{E}_* \left(\frac{S(\theta,\xi_*)-S(\vartheta_0,\xi_*)}{\sigma(\xi_*)}\right)^2 1_{\{\xi_*\in A^c\}}.$$

Then we modify the likelihood ratio as follows

$$\ln L(\vartheta,X^T) = \int_0^T \frac{S(\vartheta,X_t)}{\sigma(X_t)^2} 1_{\{X_t\in A^c\}}\mathrm{d}X_t - \frac{1}{2}\int_0^T \frac{S(\vartheta,X_t)^2}{\sigma(X_t)^2} 1_{\{X_t\in A^c\}}\mathrm{d}t$$

and using this pseudo likelihood we construct the MLE and BE. It can be shown that these estimators are consistent and asymptotically normal, but not asymptotically efficient.

Another possibility to have a consistent MLE and BE is to consider discontinuous trends. Suppose that the trend coefficient $S(\vartheta,x)=S(x-\vartheta)$ is a discontinuous function, i.e., there exists a point x_* such that $S(x_*+)-S(x_*-)\neq 0$, then for a wide class of functions $h(x)$ it is possible to have the equality $\vartheta_*=\vartheta_0$ and therefore the MLE and BE are consistent. Moreover, $T(\hat{\vartheta}_T - \vartheta_0)$ and $T(\tilde{\vartheta}_T - \vartheta_0)$ have non degenerate limits. For example, suppose that

$$\mathrm{d}X_t = -\mathrm{sgn}(X_t - \vartheta_0)\mathrm{d}t + h(X_t)\mathrm{d}t + \mathrm{d}W_t, \qquad X_0, \quad 0 \le t \le T. \tag{3.2}$$

The MLE and BE we construct on the base of the model (3.2) with $h(x) \equiv 0$, and then we substitute the observations (3.2) (containing $h(x)$). Then if we suppose that $\sup_x |h(x)| < 1$, then $\vartheta_* = \vartheta_0$ and these estimators are consistent (see Kutoyants (2004) Section 3.4). Note that if we allow $h(x)$ to take the values ± 1 then the situation with $X_t = X_0 + W_t$ is not excluded ($h(x) = \mathrm{sgn}(x - \theta_0)$) and the consistent estimation is impossible. We see that the MLE and BE in singular estimation problems can be much more robust than in the regular case.

4. Non identifiable model

Suppose that we have the same model for the different values of the parameter, i.e., $S(\vartheta_1, x) = S(\vartheta_l, x)$, $l = 2, \ldots, k$, where $\vartheta_l \neq \vartheta_i$, $l \neq i$ and $\vartheta_l, \vartheta_i \in \Theta$ (too many true models). It is well-known that the MLE converges to the set $\{\vartheta_1, \ldots, \vartheta_k\}$ of all *true values*, e.g., see Bagchi and Borkar (1984). It is possible to study this convergence as follows. Introduce a partition $\Theta = \cup_{l=1}^k \Theta_l$ such that $\vartheta_l \in \Theta_l$ and $\Theta_l \cap \Theta_i = \emptyset$. The identifiability conditions (in each set Θ_l) are: for any (small) $\nu > 0$

$$\inf_{\theta \in \Theta_l, |\theta - \vartheta_l| > \nu} \boldsymbol{E}_{\vartheta_1} \left(\frac{S(\theta, \xi) - S(\vartheta_l, \xi)}{\sigma(\xi)} \right)^2 > 0, \qquad l = 1, 2, \ldots, k.$$

Let us introduce the Gaussian vector $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_k)$ with zero mean and covariance matrix $\varrho = (\varrho_{li})$

$$\varrho_{li} = \boldsymbol{E}(\zeta_l \zeta_i) = (\mathrm{I}(\vartheta_l)\mathrm{I}(\vartheta_i))^{-1/2} \boldsymbol{E}_{\vartheta_1} \left(\frac{\dot{S}(\vartheta_l, \xi)\dot{S}(\vartheta_i, \xi)}{\sigma(\xi)^2} \right)$$

where the Fisher informations $\mathrm{I}(\vartheta_l) = \boldsymbol{E}_{\vartheta_1}(\frac{\dot{S}(\vartheta_l, \xi)}{\sigma(\xi)})^2 > 0$, $l = 1, 2, \ldots, k$. Define two random variables: discrete and continuous $\hat{\vartheta} = \sum_{l=1}^k \vartheta_l 1_{\{H_l\}}$, $\tilde{\vartheta} = \sum_{l=1}^k \vartheta_l Q_l$, where

$$H_l = \left\{ \omega : |\zeta_l| > \max_{i \neq l} |\zeta_i| \right\}, \qquad Q_l = \frac{p(\vartheta_l)\mathrm{I}(\vartheta_l)^{-1/2} e^{\zeta_l^2/2}}{\sum_{i=1}^k p(\vartheta_i)\mathrm{I}(\vartheta_i)^{-1/2} e^{\zeta_i^2/2}}.$$

It can be shown that the MLE and BE have the following limits: $\hat{\vartheta}_T \Rightarrow \hat{\vartheta}$ and $\tilde{\vartheta}_T \Rightarrow \tilde{\vartheta}$. Moreover

$$\sqrt{T}(\hat{\vartheta}_T - \hat{\vartheta}) \Rightarrow \hat{\zeta}, \qquad \sqrt{T}(\tilde{\vartheta}_T - \hat{\vartheta}) \Rightarrow \tilde{\zeta}, \tag{4.1}$$

where $\hat{\zeta} = \sum_{l=1}^k \zeta_l \mathrm{I}(\vartheta_l)^{-1/2} 1_{\{H_l\}}$. The random variable $\tilde{\zeta}$ can be calculated too, but its expression is too cumbersome. Note that the random variables $\hat{\vartheta}$ and $\tilde{\vartheta}$ are not defined on the same (as the estimators) probability space and the exact expression describing the limit distributions of the estimators is a bit more complicated, than (4.1) (see details in Kutoyants (2004) Section 2.6.2).

The proof is based on the weak convergence of the vector of processes

$$\boldsymbol{Z}_T(\boldsymbol{u}) = (Z_T^{(1)}(u_1), \ldots, Z_T^{(k)}(u_k)), \qquad Z_T^{(l)}(u_l) = \frac{L\left(\vartheta_l + \frac{u_l}{\sqrt{T}}, X^T\right)}{L(\vartheta_l, X^T)}$$

to the limit process $\boldsymbol{Z}(\boldsymbol{u}) = (Z^{(1)}(u_1), \ldots, Z^{(k)}(u_k))$, where

$$Z^{(l)}(u_l) = \exp\left\{u_l \zeta_l \mathrm{I}(\vartheta_l)^{1/2} - \frac{u_l^2}{2}\mathrm{I}(\vartheta_l)\right\}, \qquad l = 1, \ldots, k.$$

5. Null Fisher information

Suppose that $\mathrm{I}(\vartheta_0) = 0$. This means that at one point ϑ_0 (true value) the function $\dot S(\vartheta_0, x) = 0$ for all $x \in R$. Moreover, we consider a more general case and suppose that the function $S(\vartheta, x)$ is $k+1$ times continuously differentiable w.r.t. ϑ with bounded derivatives $S^{(l)}(\vartheta, x)$, $l = 1, \ldots, k$ satisfying the equalities $S^{(l)}(\vartheta_0, x) = 0$, $l = 1, \ldots, k-1$ and

$$\mathrm{I}_k(\vartheta_0) = \boldsymbol{E}_{\vartheta_0}\left(\frac{S^{(k)}(\vartheta_0, \xi)}{k!\sigma(\xi)}\right)^2 > 0.$$

Introduce the random variable $\zeta_k \sim \mathcal{N}(0, \mathrm{I}_k(\vartheta_0)^{-1})$. Then we have two cases: if k is odd then

$$T^{1/2k}(\hat\vartheta_T - \vartheta_0) \Rightarrow (\zeta_k)^{1/k},$$

and if k is even, we then suppose that the function $S(\vartheta, x)$ at the point ϑ_0 has different derivatives from the left and from the right $\dot S(\vartheta_0^{\pm}, x)$, ζ_k^+ and ζ_k^- denote the corresponding Gaussian variables and put $\zeta_k^+ = \max(0, \zeta_k^+, \zeta_k^-)$. We have

$$T^{1/2k}(\hat\vartheta_T - \vartheta_0) \Rightarrow (\zeta_k^+)^{1/k}.$$

The proofs can be found in Kutoyants (2004), Section 2.6.3. A similar statement in the problems of parameter estimation in the asymptotics of small noise was considered in Kutoyants (1994), Theorems 2.9 and 2.10. The limit expressions for the Bayesian estimators are more complicated.

Example. Let

$$\mathrm{d}X_t = [(\vartheta - a)^3 X_t^2 - X_t^3]\mathrm{d}t + \mathrm{d}W_t$$

then $\mathrm{I}_l(a) = 0$, $l = 1, 2$ and $\mathrm{I}_3(a) = \boldsymbol{E}_a \xi^4 > 0$. Hence we have $T^{1/6}(\hat\vartheta_T - a) \Rightarrow \zeta_3^{1/3}$.

6. Discontinuous Fisher information

Suppose that the function $S(\vartheta,x)$ has at the point ϑ_0 two different derivatives, from the left $\dot S(\vartheta_0^-,x)$ and from the right $\dot S(\vartheta_0^+,x)$ such that $\mathrm{I}(\vartheta_0^-)\neq \mathrm{I}(\vartheta_0^+)$ and all the other conditions of regularity are fulfilled. Then the MLE is consistent, but it is no longer asymptotically normal. Let us introduce a Gaussian vector $\zeta=(\zeta_-,\zeta_+)$ with mean zero, $\boldsymbol{E}\zeta_-^2=\boldsymbol{E}\zeta_+^2=1$ and the covariance

$$\boldsymbol{E}(\zeta_-\zeta_+)=(\mathrm{I}(\vartheta_0^-)\mathrm{I}(\vartheta_0^+))^{-1/2}\boldsymbol{E}_{\vartheta_0}\left(\frac{\dot S(\vartheta_0^-,\xi)\dot S(\vartheta_0^+,\xi)}{\sigma(\xi)^2}\right).$$

Then it can be shown that the MLE $\hat\vartheta_T$ is consistent, and $\sqrt{T}(\hat\vartheta_T-\vartheta_0)\Rightarrow\hat\zeta$, but its limit distribution is a mixture of three random variables:

- $\hat\zeta=\zeta_-\mathrm{I}(\vartheta_0^-)^{-1/2}$ if $\zeta_-<0,\ \zeta_+<0$ or $\zeta_-<0,\ \zeta_+>0$ and $|\zeta_-|>|\zeta_+|$;
- $\hat\zeta=0$ if $\zeta_->0,\ \zeta_+<0$;
- $\hat\zeta=\zeta_+\mathrm{I}(\vartheta_0^+)^{-1/2}$ if $\zeta_->0,\ \zeta_+>0$ or $\zeta_-<0,\ \zeta_+>0$ and $|\zeta_-|<|\zeta_+|$.

These properties follow from the form of the limit likelihood ratio process

$$Z(u)=\begin{cases}\exp\left\{u\zeta_-\mathrm{I}(\vartheta_0^-)^{1/2}-\dfrac{u^2}{2}\mathrm{I}(\vartheta_0^-)\right\}, & u\le 0\\[2ex] \exp\left\{u\zeta_+\mathrm{I}(\vartheta_0^+)^{1/2}-\dfrac{u^2}{2}\mathrm{I}(\vartheta_0^+)\right\}, & u>0.\end{cases}$$

We see that there is an atom at the point 0. The proof can be carried out in a maner similar to Kutoyants (2004), Section 2.6.3. The BE $\tilde\vartheta_T$ has a different limit distribution

$$\sqrt{T}(\tilde\vartheta_T-\vartheta_0)\Rightarrow\tilde\zeta=\frac{\int_R uZ(u)\mathrm{d}u}{\int_R Z(u)\mathrm{d}u}.$$

Example. Suppose that

$$\mathrm{d}X_t=(\vartheta-a)[X_t 1_{\{\vartheta<a\}}-X_t^3 1_{\{\vartheta\ge a\}}]\mathrm{d}t+\mathrm{d}W_t,$$

then $\mathrm{I}(a-)=\boldsymbol{E}_a\xi^2$ and $\mathrm{I}(a+)=\boldsymbol{E}_a\xi^6$ and the MLE has the above mentioned limit distribution.

7. Border of the parameter set

If the true value ϑ_0 is on the border of the parameter set $\Theta=(\alpha,\beta)$, say, $\vartheta_0=\alpha$, then the MLE is consistent, but

$$\sqrt{T}(\hat\vartheta_T-\alpha)\Rightarrow\frac{\zeta}{\mathrm{I}(\alpha)}1_{\{\zeta\ge 0\}},\qquad \zeta\sim\mathcal{N}(0,\mathrm{I}(\alpha)).$$

Of course, here $\mathrm{I}(\alpha)=\mathrm{I}(\alpha^+)$. The estimator is asymptotically half-normal with an atom at 0, i.e., with probability 0, 5 it takes the value 0. This follows from the form of the limit likelihood ratio:

$$Z(u)=\exp\left\{u\zeta-\frac{u^2}{2}\mathrm{I}(\alpha)\right\},\qquad u\ge 0.$$

Let us denote

$$\Psi(\zeta) = \int_0^\infty e^{u\zeta-(u^2/2)\mathrm{I}(\alpha)}\mathrm{d}u$$

then the limit $\tilde{\zeta}$ of the BE: $\sqrt{T}(\tilde{\vartheta}_T - \alpha) \Rightarrow \tilde{\zeta}$ can be written as $\tilde{\zeta} = [\ln \Psi(\zeta)]'_\zeta$. The direct calculation provides

$$\tilde{\zeta} = \frac{1}{\sqrt{\mathrm{I}(\alpha)}}\left(\zeta_* + \left(\int_{-\zeta_*}^\infty e^{-(1/2)(u^2-\zeta_*^2)}\mathrm{d}u\right)^{-1}\right), \qquad \zeta_* \sim \mathcal{N}(0,1).$$

If the true value ϑ_0 is out of the set Θ (wrong choice of the set Θ), then the properties of the estimators correspond to the situation described in Section 3, but the value ϑ_* can be on the border of the set Θ.

8. Cusp type singularity

Let us suppose that the observed process is

$$\mathrm{d}X_t = [a|X_t - \vartheta|^\kappa - X_t]\mathrm{d}t + \mathrm{d}W_t, \qquad X_0, \quad 0 \le t \le T$$

where $\kappa \in (0, \frac{1}{2})$. Then the trend coefficient is not differentiable at one point $x = \vartheta$ and the Fisher information $\mathrm{I}(\vartheta) = \infty$. To describe the properties of the MLE and BE we introduce the limit likelihood ratio process

$$Z(u) = \exp\left\{W^H(u) - \frac{|u|^{2H}}{2}\right\}$$

where $W^H(\cdot)$ is a double sided fractional Brownian motion and two random variables $\hat{\zeta}$ and $\tilde{\zeta}$ as follows

$$Z(\hat{\zeta}) = \sup_u Z(u), \qquad \tilde{\zeta} = \frac{\int_R uZ(u)\mathrm{d}u}{\int_R Z(u)\mathrm{d}u}. \tag{8.1}$$

Here $H = \kappa + \frac{1}{2}$ (the Hurst parameter). Put $\gamma_\vartheta = \Gamma_\vartheta^{1/H}$ where

$$\Gamma_\vartheta^2 = \frac{a^2\sin^2(2\pi\kappa)}{G(\vartheta)} \frac{\Gamma(1+\kappa)\Gamma\left(\frac{1}{2}-\kappa\right)}{2^{2(\kappa-1)}\sqrt{\pi}(2\kappa+1)}.$$

We show that

$$T^{1/2H}(\hat{\vartheta}_T - \vartheta) \Rightarrow \frac{\hat{\zeta}}{\gamma_\vartheta}, \qquad T^{1/2H}(\tilde{\vartheta}_T - \vartheta) \Rightarrow \frac{\tilde{\zeta}}{\gamma_\vartheta}$$

and the BE are asymptotically efficient. The proof can be found in Dachian and Kutoyants (2003) or in Kutoyants (2004), Section 3.2. The i.i.d. r.v.'s model was considered by Prakasa Rao (1968). The wide class of singular estimation problems (including cusp type) was studied by Ibragimov and Khasminskii (1981), Chapter VI.

9. Discontinuous trend coefficient

Let us suppose that the observed process is

$$\mathrm{d}X_t = -\mathrm{sgn}(X_t - \vartheta)\mathrm{d}t + \mathrm{d}W_t, \qquad X_0, \quad 0 \le t \le T.$$

Then the limit likelihood ratio is

$$Z(u) = \exp\left\{W(u) - \frac{|u|}{2}\right\}.$$

If we denote by $\hat{\zeta}$ and $\tilde{\zeta}$ the random variables defined by the equations (8.1) but with this limit process, then the MLE and BE have the following limits

$$T(\hat{\vartheta}_T - \vartheta) \Rightarrow \frac{\hat{\zeta}}{4}, \qquad T(\tilde{\vartheta}_T - \vartheta) \Rightarrow \frac{\tilde{\zeta}}{4} \tag{9.1}$$

and bayesian estimators are asymptotically efficient. We see that the rate of convergence is much better than in all other cases considered above. The proof you can find in Kutoyants (2004), Section 3.4, where the more general case of discontinuous trend coefficients is also discussed.

It is interesting to note that the same rate and the same limit distributions we obtain in the problem of delay estimation by the observations of Itô process

$$\mathrm{d}X_t = -\gamma X_{t-\vartheta}\mathrm{d}t + \mathrm{d}W_t, \qquad X_0, \quad 0 \le t \le T,$$

where $\vartheta \in (0, \frac{\pi}{2\gamma})$. This process has ergodic properties and for the estimators we obtain the convergence similar to (9.1)

$$T(\hat{\vartheta}_T - \vartheta) \Rightarrow \frac{\hat{\zeta}}{\gamma^2}, \qquad T(\tilde{\vartheta}_T - \vartheta) \Rightarrow \frac{\tilde{\zeta}}{\gamma^2}.$$

For the proof see Küchler and Kutoyants (2000) and Kutoyants (2004), Section 3.3. Note as well that the values $\boldsymbol{E}\hat{\zeta}^2 = 26$ and $\boldsymbol{E}\tilde{\zeta}^2 = 16\zeta(3) \sim 19,3$ were calculated by Terent'ye (1968) and Rubin and Song (1995) respectively. Here $\zeta(s)$ is the Rieman zeta function: $\zeta(s) = \sum_{k=1}^{\infty} k^{-s}$.

Remind that the problem of delay estimation of the same stochastic process

$$\mathrm{d}X_t = -\gamma X_{t-\vartheta}\mathrm{d}t + \varepsilon\mathrm{d}W_t, \qquad X_0, \quad 0 \le t \le T,$$

but in the asymptotics of small noise ($\varepsilon \to 0$) became regular and both estimators are consistent, asymptotycaly normal (with regular rate of convergence) and asymptotically efficient (see Kutoyants (2007) for details).

10. Null-recurrent diffusion

Suppose that the condition (2.3) is not fulfilled and we consider an example of a null-recurrent diffusion process

$$\mathrm{d}X_t = -\vartheta\frac{X_t}{1+X_t^2}\mathrm{d}t + \sigma\mathrm{d}W_t, \qquad X_0, \quad 0 \le t \le T.$$

The MLE can be explicitly written as

$$\hat{\vartheta}_T = -\left(\int_0^T \frac{X_t^2}{(1+X_t^2)^2}\mathrm{d}t\right)^{-1} \int_0^T \frac{X_t}{1+X_t^2}\mathrm{d}X_t.$$

Then if $\frac{2\vartheta}{\sigma^2} > 1$, then the process is ergodic and $\hat{\vartheta}_T$ is $\sqrt{T}$-asymptotically normal, but if $-1 < \frac{2\vartheta}{\sigma^2} < 1$ then the diffusion process is null-recurrent and

$$T^{\gamma/2}(\hat{\vartheta}_T - \vartheta) \Rightarrow \frac{\zeta\eta^{\gamma/2}}{J_*(\vartheta)^{1/2}},$$

where $\gamma = 1/2 + \vartheta/\sigma^2$, $\zeta \sim \mathcal{N}(0,1)$ and η is independent of ζ stable r.v. with Laplace transform $\boldsymbol{E}e^{-p\eta} = e^{-p^\gamma}$. The constant $J_*(\vartheta)$ and proof of this result can be found in Höpfner and Kutoyants (2003) (see as well Kutoyants (2004) Section 3.5.1).

11. Conclusion

We presented here a collection of results which allow a better understanding of the role of each of the regularity conditions in the asymptotic properties of the MLE and BE. We see that these conditions are mainly independent and have to be verified for any particular statistical model. For example, if we have no global identifiability condition like (2.4) and say that *in the vicinity of the true value there exists a solution of the maximum likelihood equation which converges to the true value*, then this solution is not a consistent MLE, because the situation described in Section 4 above is not excluded, and in the vicinity of each value ϑ_l we have similar solutions, which converge to ϑ_l. Another interesting point is the "robustness" of the MLE in the case of the misspecified singular estimation problem illustrated by the model (3.2).

Acknowledgements

The author wishes to thank the Referee for pointing out the omission of the condition $\sigma(x) \in \mathcal{P}$ in Lemma 1.17, Kutoyants (2004). This condition is needed to prove (1.61) and has to be mentioned in the Regularity conditions $\boldsymbol{\mathcal{A}}$ there.

References

Bagchi, A. and Borkar, V. (1984). Parameter identification in infinite dimensional linear systems, *Stochastics*, **12**, 201–213.

Cramér, H. (1946). *Mathematical Methods of Statistics*, Princeton University Press.

Dachian, S. and Kutoyants, Yu. A. (2003). On cusp estimation of ergodic diffusion process, *J. Statist. Plann. Inference*, **117**, 153–166.

Höpfner, R. and Kutoyants, Yu. A. (2003). On a problem of statistical inference in null recurrent diffusions, *Stat. Inference Stoch. Process.*, **6**(1), 25–42.

Ibragimov, I. A. and Khasminskii, R. Z. (1981). *Statistical Estimation. Asymptotic Theory*, Springer-Verlag, New York.

Küchler, U. and Kutoyants, Yu. A. (2000). Delay estimation for some stationary diffusion-type processes, *Scand. J. Statist.*, **27**(3), 405–414.

Kutoyants, Yu. A. (1994). *Identification of Dynamical Systems with Small Noise*, Kluwer, Dordrecht.

Kutoyants, Yu. A. (2004). *Statistical Inference for Ergodic Diffusion Processes*, Springer Verlag, London.

Kutoyants, Yu. A. (2007). On delay estimation and testing for diffusion type processes, *Stochastic Differential Equations: Theory and Applications*, in honor of B. L. Rozovskii (eds. P. H. Baxendale and S. V. Lototsky), World Scientific, Singapore, 339–356.

Liptser, R. S. and Shiryayev, A. N. (2001). *Statistics of Random Processes*, *I*, *II*, 2nd ed., Springer-Verlag, New York.

McKeague, I. (1984). Estimation for diffusion processes under misspecified models, *J. Appl. Probab.*, **21**, 511–520.

Prakasa Rao, B. L. S. (1968). Estimation of the location of the cusp of a continuous density, *Ann. Math. Statist.*, **39**(1), 76–87.

Rubin, H. and Song, K.-S. (1995). Exact computation of the asymptotic efficiency of maximum likelihood estimators of a discontinuous signal in a Gaussian white noise, *The Annals of Statistics*, **23**(3), 732–739.

Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*, Springer-Verlag, New York.

Terent'yev, A. S. (1968). Probability distribution of a time location of an absolute maximum at the output of a synchronized filter, *Radioengineering and Electronics*, **13**(4), 652–657.

van der Vaart, A. W. (1998). *Asymptotic Statistics*, Cambridge University Press, Cambridge.

Yoshida, N. (1990). Asymptotic behavior of M-estimator and related random field for diffusion process, *Ann. Inst. Statist. Math.*, **42**(2), 221–251.