Electronic Journal of Statistics
Vol. 12 (2018) 80–106
ISSN: 1935-7524
https://doi.org/10.1214/17-EJS1385

On misspecifications in regularity and properties of estimators

Oleg V. Chernoyarov

Department of Electronics and Nanoelectronics, National Research Universtity "MPEI", Moscow, Russia

Voronezh State University, Voronezh, Russia
e-mail: chernoyarovov@mpei.ru

Yury A. Kutoyants

Department of Mathematics, Le Mans University Le Mans, France

Voronezh State University, Voronezh, Russia

Tomsk State University, Tomsk, Russia
e-mail: kutoyants@univ-lemans.fr

and

Andrei P. Trifonov

Department of Radio Physics, Voronezh State University, Voronezh, Russia
e-mail: trifonov@phys.vsu.ru

Abstract: The problem of parameter estimation by the continuous time observations of a deterministic signal in white Gaussian noise is considered. The asymptotic properties of the maximum likelihood estimator are described in the asymptotic of small noise (large signal-to-noise ratio). We are interested in the situation when there is a misspecification in the regularity conditions. In particular it is supposed that the statistician uses a discontinuous (change-point type) model of signal, when the true signal is continuously differentiable function of the unknown parameter.

MSC 2010 subject classifications: Primary 62M02, 62G10; secondary 62G20.
Keywords and phrases: Misspecification, change-point type model, maximum likelihood estimator, regularity conditions.

Received June 2017.

1. Introduction

Consider the problem of parameter estimation by the observations of the signals in White Gaussian Noise (WGN) model

$$\mathrm{d}X_t = S(\vartheta, t)\,\mathrm{d}t + \varepsilon \mathrm{d}W_t, \quad X_0 = x_0, \quad 0 \le t \le T.$$

Here $S(\vartheta, t)$ is deterministic signal and we have to estimate the parameter $\vartheta \in \Theta = (\alpha, \beta)$ by continuous time observations $X^T = (X_t, 0 \leq t \leq T)$. Without loss of generality we suppose that $x_0 = 0$. We are interested in the asymptotic behavior of estimators of this parameter as $\varepsilon \to 0$. It is known that if the signal $S(\cdot, t)$ is a smooth function of the first argument for any t with finite Fisher information

$$\mathrm{I}(\vartheta) = \int_0^T \dot{S}(\vartheta, t)^2 \, \mathrm{d}t, \tag{1.1}$$

then the maximum likelihood estimator $\hat{\vartheta}_\varepsilon$ is consistent, asymptotically normal with the rate of convergence ε, i.e. $\varepsilon^{-1}\big(\hat{\vartheta}_\varepsilon - \vartheta_0\big) \Rightarrow \mathcal{N}\big(0, \mathrm{I}(\vartheta_0)^{-1}\big)$ and asymptotically efficient [5], [11], [12]. Here and in the sequel dot means derivation w.r.t. ϑ.

The situation changes if the signal $S(\vartheta, t) = S(t - \vartheta)$, where $S(t)$ is discontinuous function in t, say, has a jump at the point $t = 0$. Then the Fisher information does not exist and the properties of the estimators are essentially different. For example, the MLE is consistent, has non Gaussian limit distribution with the rate of convergence ε^2, i.e., $\varepsilon^{-2}\big(\hat{\vartheta}_\varepsilon - \vartheta_0\big) \Rightarrow \hat{u}$ and asymptotically efficient are Bayesian estimators [6]. Here $\hat{u}$ is some non degenerate random variable.

Let us recall that there is always a gap between mathematical model chosen by a statistician to describe the results of observations and the model which corresponds exactly to these observations. Sometimes the difference is not important and the theoretical results are in good agreement with the real data and sometimes the difference can be essential. The difference between the models is determined by the Kullback-Leibler distance between the corresponding measures. We are interested in the situations where the L_2 difference between the signals can be very small and the same time the difference of the properties of estimators is sufficiently important. We show that even the rate of convergence of estimators is quite different for two such close models.

We are in the situation of *misspecification*. This misspecification concerns not only the choice of the signal in the family of models close to the model of real data, but we suppose that the regularity conditions assumed by the statistician are wrong. In particularly, the observed signal $S(\vartheta_0, t)$ can be *smooth* with respect to the unknown parameter ϑ_0, but the signal chosen by the statistician $M(\vartheta, t)$ is discontinuous. Our goal is to describe the properties of the corresponding pseudo-MLE $\hat{\vartheta}_\varepsilon$. Recall that this estimator converges to the value $\hat{\vartheta}_0$, which minimizes the Kullback-Leibler distance. Then we study its limit distribution and show that it converges to a non Gaussian limit law with the rate $\varepsilon^{2/3}$, which is different from the rate ε in smooth case and the rate ε^2 in discontinuous case.

This statement of the problem comes from the statistical radio physics. Suppose that we have to detect a rectangular signal (change point problem) with the help of the maximum likelihood algorithm. It is known that the received signal can not have exactly rectangular form because the electrical current in the circuit which produces the signal can not have a discontinuous changes and we have just a strongly increasing curve. The L_2 difference between rectangular and strongly increasing signals can be small and it is interesting to see what happens with the properties of the estimators in such situation. This approximation can be good or bad depending on the front of the signal and the level of signal-to-noise ratio.

We consider as well in some sense *inverse problem*, where the theoretical model is smooth and the real data model is discontinuous and we describe the asymptotics of the pseudo-MLE as $\varepsilon \to 0$. We show that in this case the estimator $\hat{\vartheta}_\varepsilon$ converges to the point $\hat{\vartheta}_0$ which minimizes the Kullback-Leibler distance and is asymptotically normal with the rate ε.

In the next section we remind some well-known properties of the estimators in smooth, cusp-type and discontinuous signals together with the limits of the corresponding normalized likelihood ratios. This will allow us to see the changes due to the different type of misspecifications. In particular, we show that in the case of misspecifications the limit likelihood ratios are in some sense "mixtures of the true likelihood ratios". The stochastic part is defined by the *theoretical* model and the deterministic part is as it has to be in *true* model.

At the end we describe the conditions on the misspecified model (discontinuous vs discontinuous) which allow nevertheless to prove the consistency (true) of the pseudo-MLE.

2. Preliminaries

Let us consider the problem of parameter estimation by the observations (in continuous time) of the deterministic signal in the presence of WGN of small intensity

$$\mathrm{d}X_t = S\left(\vartheta_0, t\right)\mathrm{d}t + \varepsilon \mathrm{d}W_t, \quad X_0 = 0, \quad 0 \le t \le T. \tag{2.1}$$

Here the unknown parameter $\vartheta_0 \in \Theta = (\alpha, \beta)$ and $S\left(\vartheta_0, \cdot\right) \in L_2\left(0, T\right)$. We suppose for the simplicity of exposition that this parameter is one-dimensional and that α and β are finite. We are interested in the behavior of the estimators of this parameter in the asymptotic of *small noise*, i.e., as $\varepsilon \to 0$. The likelihood ratio function is

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

The maximum likelihood estimator (MLE) $\hat{\vartheta}_\varepsilon$ and the Bayesian estimator (BE) $\tilde{\vartheta}_\varepsilon$ (for quadratic loss function) are defined by the relations

$$L\left(\hat{\vartheta}_\varepsilon, X^T\right) = \sup_{\vartheta \in \Theta} L\left(\vartheta, X^T\right), \qquad \tilde{\vartheta}_\varepsilon = \frac{\int_\Theta \vartheta p\left(\vartheta\right) L\left(\vartheta, X^T\right)\mathrm{d}\vartheta}{\int_\Theta p\left(\vartheta\right) L\left(\vartheta, X^T\right)\mathrm{d}\vartheta}. \tag{2.2}$$

Here $p\left(\cdot\right)$ is the density of the distribution *a priory.* It is supposed to be continuous and positive.

It is well known that in the case of the smooth w.r.t. ϑ signal $S(\vartheta,t)$ these estimators are consistent, asymptotically normal

$$\varepsilon^{-1}\left(\hat\vartheta_\varepsilon-\vartheta_0\right)\Longrightarrow\mathcal{N}\left(0,\mathrm{I}(\vartheta_0)^{-1}\right),\qquad \varepsilon^{-1}\left(\tilde\vartheta_\varepsilon-\vartheta_0\right)\Longrightarrow\mathcal{N}\left(0,\mathrm{I}(\vartheta_0)^{-1}\right),$$

with convergent polynomial moments of any order $p>0$

$$\lim_{\varepsilon\to0}\mathbf{E}_{\vartheta_0}\left|\frac{\hat\vartheta_\varepsilon-\vartheta_0}{\varepsilon}\right|^p=\mathrm{I}(\vartheta_0)^{-\frac{p}{2}}\,\mathbf{E}\left|\zeta\right|^p,\qquad \lim_{\varepsilon\to0}\mathbf{E}_{\vartheta_0}\left|\frac{\tilde\vartheta_\varepsilon-\vartheta_0}{\varepsilon}\right|^p=\mathrm{I}(\vartheta_0)^{-\frac{p}{2}}\,\mathbf{E}\left|\zeta\right|^p$$

and the both estimators are asymptotically efficient. Here $\mathrm{I}(\vartheta_0)$ is the Fisher information (1.1) and $\zeta\sim\mathcal{N}(0,1)$. The normalized likelihood ratio $Z_\varepsilon(u)$ has the following limit

$$Z_\varepsilon(u)=\frac{L(\vartheta_0+\varepsilon u,X)}{L(\vartheta_0,X)}\Longrightarrow Z(u)=\exp\left\{u\Delta-\frac{u^2}{2}\mathrm{I}(\vartheta_0)\right\},\quad u\in\mathcal{R}. \tag{2.3}$$

Here $\Delta\sim\mathcal{N}(0,\mathrm{I}(\vartheta_0))$. For the proofs see [5] or [7].

Suppose that the signal $S(\vartheta,t)$ has *cusp*-type singularity, say, $S(\vartheta_0,t)=|t-\theta_0|^\kappa$, where $0<\alpha<\vartheta_0<\beta<T$ and $\kappa\in(0,\frac12)$. Then the Fisher information is ∞ and we have a singular problem of parameter estimation. Introduce the Hurst parameter $H=\kappa+\frac12$ and double-side fractional Brownian motion (fBm) $W^H(u),u\in\mathcal{R}$. It is known that the MLE and BE are consistent, have limit distributions

$$\varepsilon^{-\frac{2}{H}}\left(\hat\vartheta_\varepsilon-\vartheta_0\right)\Longrightarrow\hat\xi,\qquad \varepsilon^{-\frac{2}{H}}\left(\tilde\vartheta_\varepsilon-\vartheta_0\right)\Longrightarrow\tilde\xi,$$

the polynomial moments converge: for any $p>0$

$$\lim_{\varepsilon\to0}\mathbf{E}_{\vartheta_0}\left|\frac{\hat\vartheta_\varepsilon-\vartheta_0}{\varepsilon^{\frac{2}{H}}}\right|^p=\mathbf{E}\left|\hat\xi\right|^p,\qquad \lim_{\varepsilon\to0}\mathbf{E}_{\vartheta_0}\left|\frac{\tilde\vartheta_\varepsilon-\vartheta_0}{\varepsilon^{\frac{2}{H}}}\right|^p=\mathbf{E}\left|\tilde\xi\right|^p$$

and the BE are asymptotically efficient. Here the random variables $\hat\xi$ and $\tilde\xi$ are defined by the relations

$$Z(\hat\xi)=\sup_{u\in\mathcal{R}}Z(u),\qquad \tilde\xi=\frac{\int_{\mathcal{R}}uZ(u)\,\mathrm{d}u}{\int_{\mathcal{R}}Z(u)\,\mathrm{d}u}, \tag{2.4}$$

where the process $Z(u),u\in\mathcal{R}$ is the limit of the normalized likelihood ratio

$$Z_\varepsilon(u)=\frac{L\left(\vartheta_0+\varepsilon^{\frac{2}{H}}u,X^T\right)}{L(\vartheta_0,X^T)}\Rightarrow Z(u)=\exp\left\{\gamma_{\vartheta_0}W^H(u)-\frac{\gamma_{\vartheta_0}^2}{2}|u|^{2H}\right\}.$$

Here $\gamma_{\vartheta_0}>0$ is some constant. The proofs can be found in [2]. Note that the case $\kappa\in\left(-\frac12,0\right)$ was studied in [8].

Suppose now that the function $S(\vartheta,t)$ has discontinuity, say,

$$S(\vartheta_0,t)=h(t)\,\mathbb{1}_{\{t<\vartheta_0\}}+g(t)\,\mathbb{1}_{\{t\geq\vartheta_0\}},\qquad 0<\alpha<\vartheta_0<\beta<T,$$

where $h(t)\neq g(t)$ for $t\in(\alpha,\beta)$.

It is known that the MLE and BE are consistent, have limit distributions

$$\varepsilon^{-2}\left(\hat\vartheta_\varepsilon-\vartheta_0\right)\Longrightarrow\hat\eta,\qquad \varepsilon^{-2}\left(\tilde\vartheta_\varepsilon-\vartheta_0\right)\Longrightarrow\tilde\eta,\tag{2.5}$$

the polynomial moments converge: for any $p>0$

$$\lim_{\varepsilon\to 0}\mathbf{E}_{\vartheta_0}\left|\frac{\hat\vartheta_\varepsilon-\vartheta_0}{\varepsilon^2}\right|^p=\mathbf{E}\left|\hat\eta\right|^p,\qquad \lim_{\varepsilon\to 0}\mathbf{E}_{\vartheta_0}\left|\frac{\tilde\vartheta_\varepsilon-\vartheta_0}{\varepsilon^2}\right|^p=\mathbf{E}\left|\tilde\eta\right|^p$$

and the BE are asymptotically efficient. Here the random variables $\hat\eta$ and $\tilde\eta$ are defined by the relations (2.4), where the process $Z(u),u\in\mathcal{R}$ is the limit of the normalized likelihood ratio

$$Z_\varepsilon(u)=\frac{L\left(\vartheta_0+\varepsilon^2u,X^T\right)}{L\left(\vartheta_0,X^T\right)}\Longrightarrow Z(u)=\exp\left\{\delta(\vartheta_0)\,W(u)-\frac{\delta(\vartheta_0)^2\left|u\right|}{2}\right\}.\tag{2.6}$$

Here $\delta(\vartheta)=h(\vartheta)-g(\vartheta)$ and $W(\cdot)$ is a double-sided Wiener process. For the proofs see [6] or [7].

We presented here three models of observations with different types of regularity and the normalized likelihood ratios $Z_\varepsilon(u)$ have three different rates of normalizing. The solutions of all these problems were obtained by the same method of the study of the properties of estimators developed by Ibragimov and Khasminskii [7]. Let us recall here the main steps. Suppose that for some normalizing function $\varphi_\varepsilon\to 0$ the random process

$$Z_\varepsilon(u)=\frac{L\left(\vartheta_0+\varphi_\varepsilon u,X^T\right)}{L\left(\vartheta_0,X^T\right)},\qquad u\in\mathrm{U}_\varepsilon=\left(\frac{\alpha-\vartheta_0}{\varphi_\varepsilon},\frac{\beta-\vartheta_0}{\varphi_\varepsilon}\right)$$

converges to a non degenerate random process $Z(u),u\in\mathcal{R}$, i.e.,

$$Z_\varepsilon(\cdot)\Longrightarrow Z(\cdot).\tag{2.7}$$

Introduce the random variables $\hat u$ and $\tilde u$ by the relations like (2.4)

$$Z(\hat u)=\sup_{u\in\mathcal{R}}Z(u),\qquad \tilde u=\frac{\int_{\mathcal{R}}uZ(u)\,\mathrm{d}u}{\int_{\mathcal{R}}Z(u)\,\mathrm{d}u}.\tag{2.8}$$

The limit distribution of the normalized MLE $\hat u_\varepsilon=\varphi_\varepsilon^{-1}\left(\hat\vartheta_\varepsilon-\vartheta_0\right)$ can be obtained as follows:

$$\mathbf{P}_{\vartheta_0}\left(\frac{\hat\vartheta_\varepsilon-\vartheta_0}{\varphi_\varepsilon}<x\right)=\mathbf{P}_{\vartheta_0}\left(\hat\vartheta_\varepsilon<\vartheta_0+\varphi_\varepsilon x\right)$$

$$
\begin{aligned}
&= \mathbf{P}_{\vartheta_0}\left\{\sup_{\vartheta<\vartheta_0+\varphi_\varepsilon x} L\left(\vartheta,X^T\right) > \sup_{\vartheta\geq\vartheta_0+\varphi_\varepsilon x} L\left(\vartheta,X^T\right)\right\}\\
&= \mathbf{P}_{\vartheta_0}\left\{\sup_{\vartheta<\vartheta_0+\varphi_\varepsilon x} \frac{L\left(\vartheta,X^T\right)}{L\left(\vartheta_0,X^T\right)} > \sup_{\vartheta\geq\vartheta_0+\varphi_\varepsilon x} \frac{L\left(\vartheta,X^T\right)}{L\left(\vartheta_0,X^T\right)}\right\}\\
&= \mathbf{P}_{\vartheta_0}\left\{\sup_{u<x,u\in \mathrm{U}_\varepsilon} Z_\varepsilon\left(u\right) > \sup_{u\geq x,u\in \mathrm{U}_\varepsilon} Z_\varepsilon\left(u\right)\right\}\\
&\longrightarrow \mathbf{P}_{\vartheta_0}\left\{\sup_{u<x} Z\left(u\right) > \sup_{u\geq x} Z\left(u\right)\right\} = \mathbf{P}_{\vartheta_0}\left(\hat u < x\right). \qquad (2.9)
\end{aligned}
$$

Here we put $\vartheta=\vartheta_0+\varphi_\varepsilon u$ and used the convergence (2.7). Therefore we proved that

$$
\frac{\hat\vartheta_\varepsilon-\vartheta_0}{\varphi_\varepsilon} \quad \Longrightarrow \quad \hat u.
$$

For the Bayesian estimator we have:

$$
\begin{aligned}
\tilde\vartheta_\varepsilon &= \frac{\int_\Theta \theta p\left(\theta\right) L\left(\theta,X^T\right)\mathrm{d}\theta}{\int_\Theta p\left(\theta\right) L\left(\theta,X^T\right)\mathrm{d}\theta} = \vartheta_0+\varphi_\varepsilon \frac{\int_{\mathrm{U}_\varepsilon} u p\left(\theta_u\right) L\left(\theta_u,X^T\right)\mathrm{d}u}{\int_{\mathrm{U}_\varepsilon} p\left(\theta_u\right) L\left(\theta_u,X^T\right)\mathrm{d}u}\\
&= \vartheta_0+\varphi_\varepsilon \frac{\int_{\mathrm{U}_\varepsilon} u p\left(\theta_u\right) \frac{L\left(\theta_u,X^T\right)}{L\left(\vartheta_0,X^T\right)}\,\mathrm{d}u}{\int_{\mathrm{U}_\varepsilon} p\left(\theta_u\right) \frac{L\left(\theta_u,X^T\right)}{L\left(\vartheta_0,X^T\right)}\,\mathrm{d}u} = \vartheta_0+\varphi_\varepsilon \frac{\int_{\mathrm{U}_\varepsilon} u p\left(\theta_u\right) Z_\varepsilon\left(u\right)\mathrm{d}u}{\int_{\mathrm{U}_\varepsilon} p\left(\theta_u\right) Z_\varepsilon\left(u\right)\mathrm{d}u},
\end{aligned}
$$

where once more we changed the variables $\vartheta_u=\vartheta_0+\varphi_\varepsilon u$. Hence

$$
\frac{\tilde\vartheta_\varepsilon-\vartheta_0}{\varphi_\varepsilon} = \frac{\int_{\mathrm{U}_\varepsilon} u p\left(\theta_u\right) Z_\varepsilon\left(u\right)\mathrm{d}u}{\int_{\mathrm{U}_\varepsilon} p\left(\theta_u\right) Z_\varepsilon\left(u\right)\mathrm{d}u} \Longrightarrow \frac{\int_{\mathcal R} u Z\left(u\right)\mathrm{d}u}{\int_{\mathcal R} Z\left(u\right)\mathrm{d}u} = \tilde u. \qquad (2.10)
$$

Here $p\left(\vartheta_0+\varphi_\varepsilon u\right)\to p\left(\vartheta_0\right)$.

We are interested in the following problem of misspecification. Suppose that model of observations chosen by the statistician is

$$
\mathrm{d}X_t = M\left(\vartheta,t\right)\mathrm{d}t+\varepsilon \mathrm{d}W_t, \quad X_0=0, \quad 0\leq t\leq T
$$

and the real data model (2.1) are different. Especially we study the situations where the regularity conditions of these models do not coincide. For example, the signal $S\left(\vartheta,t\right)$ is a smooth function of ϑ (regular case), but the statistician supposes that the observed model has singularities of cusp or discontinuous types.

We presente in this work one model of observations with three different conditions of regularity. We study the properties of the estimators when the misspecifications are smooth-discontinuous and discontinuous-smooth and discontinuous-discontinuous. The similar statements related to the second model were studied in [2].

The proofs of all presented below results are based on the technique developed by Ibragimov and Khasminskii [7], but there is the following difference. As usual in misspecification problems the pseudo-likelihood ratio $V\left(\vartheta, X^T\right)$ does not satisfy the equality $\mathbf{E}_{\vartheta_0} V\left(\vartheta, X^T\right) = 1$. The limit of the MLE $\hat\vartheta_\varepsilon$ is the point $\hat\vartheta$ which minimizes the corresponding Kullback-Leibler distance. Moreover, the normalized pseudo-likelihood ratio

$$Z_\varepsilon(u) = \frac{V\left(\hat\vartheta + \varphi_\varepsilon u, X^T\right)}{V\left(\hat\vartheta, X^T\right)}, \qquad u \in \mathrm{U}_\varepsilon = \left(\frac{\alpha - \hat\vartheta}{\varphi_\varepsilon}, \frac{\beta - \hat\vartheta}{\varphi_\varepsilon}\right)$$

has no non degenerate limit $Z(u)$. We show that with the "correct" normalizing functions φ_ε we obtain the limits like $Z_\varepsilon(u) \to \infty$. That is why in the problems below we propose two normalizing functions φ_ε and ψ_ε such that the random process

$$\hat Z_\varepsilon(u) = \left[Z_\varepsilon(u)\right]^{\psi_\varepsilon}, \qquad u \in \mathrm{U}_\varepsilon$$

converges to some non degenerate random process $\hat Z(u)$. The limit distributions of the MLEs are obtained with the help of the convergence similar to (2.9), where we use the equality

$$\begin{aligned} \mathbf{P}_{\vartheta_0}&\left\{\sup_{u<x, u\in \mathrm{U}_\varepsilon} Z_\varepsilon(u) > \sup_{u\geq x, u\in \mathrm{U}_\varepsilon} Z_\varepsilon(u)\right\} \\ &= \mathbf{P}_{\vartheta_0}\left\{\sup_{u<x, u\in \mathrm{U}_\varepsilon} \hat Z_\varepsilon(u) > \sup_{u\geq x, u\in \mathrm{U}_\varepsilon} \hat Z_\varepsilon(u)\right\} \\ &\longrightarrow \mathbf{P}_{\vartheta_0}\left\{\sup_{u<x} \hat Z(u) > \sup_{u\geq x} \hat Z(u)\right\} = \mathbf{P}_{\vartheta_0}\left(\hat u < x\right). \end{aligned} \tag{2.11}$$

This last convergence provides us the limit

$$\frac{\hat\vartheta_\varepsilon - \hat\vartheta}{\varphi_\varepsilon} \quad\Longrightarrow\quad \hat u$$

with a different $\hat u$.

We propose two examples of misspecified models which admit the consistent estimation of the unknown parameter.

3. Main results

We suppose that the observed process (*real model*) is

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

where $\vartheta_0 \in \Theta = (\alpha, \beta)$ is the true value of unknown parameter and $S(\vartheta, \cdot) \in L_2(0, T)$. If we use the *theoretical model*

$$\mathrm{d}X_t = M(\vartheta, t)\,\mathrm{d}t + \varepsilon \mathrm{d}W_t, \quad X_0 = 0, \quad 0 \leq t \leq T,$$

with $M(\vartheta,\cdot) \in L_2(0,T)$, then the (pseudo-) likelihood ratio (misspecified) is

$$V\left(\vartheta, X^T\right) = \exp\left\{\frac{1}{\varepsilon^2}\int_0^T M(\vartheta,t)\,\mathrm{d}X_t - \frac{1}{2\varepsilon^2}\int_0^T M(\vartheta,t)^2\,\mathrm{d}t\right\},\ \vartheta\in\Theta \quad (3.2)$$

and the pseudo-MLE $\hat{\vartheta}_\varepsilon$ is defined by the equation

$$V\left(\hat{\vartheta}_\varepsilon, X^T\right) = \sup_{\vartheta\in\Theta} V\left(\vartheta, X^T\right). \quad (3.3)$$

If this equation has more than one solution, then any one of them can be taken as $\hat{\vartheta}_\varepsilon$. To understand what is the limit of the pseudo-MLE we write the likelihood ratio as follows

$$\begin{aligned}
&\varepsilon^2 \ln V\left(\vartheta, X^T\right)\\
&\quad = \varepsilon\int_0^T M(\vartheta,t)\,\mathrm{d}W_t - \frac{1}{2}\int_0^T \Big[M(\vartheta,t)^2 - 2M(\vartheta,t)\,S(\vartheta_0,t)\Big]\,\mathrm{d}t\\
&\quad = \varepsilon\int_0^T M(\vartheta,t)\,\mathrm{d}W_t - \frac{1}{2}\left\|M(\vartheta,\cdot) - S(\vartheta_0,\cdot)\right\|^2 + \frac{1}{2}\left\|S(\vartheta_0,\cdot)\right\|^2,
\end{aligned}$$

where we denoted as $\|\cdot\|$ the $L_2(0,T)$ norm. It can be easily verified that under mild regularity conditions we have the convergence

$$\begin{aligned}
&\sup_{\vartheta\in\Theta}\left|\varepsilon^2\ln V\left(\vartheta,X^T\right) + \frac{1}{2}\left\|M(\vartheta,\cdot) - S(\vartheta_0,\cdot)\right\|^2 - \frac{1}{2}\left\|S(\vartheta_0,\cdot)\right\|^2\right|\\
&\qquad = \varepsilon\sup_{\vartheta\in\Theta}\left|\int_0^T M(\vartheta,t)\,\mathrm{d}W_t\right| \longrightarrow 0
\end{aligned}$$

as $\varepsilon\to 0$. Hence if we suppose that the equation

$$\inf_{\vartheta\in\Theta}\left\|M(\vartheta,\cdot) - S(\vartheta_0,\cdot)\right\| = \left\|M(\hat{\vartheta},\cdot) - S(\vartheta_0,\cdot)\right\|$$

has a unique solution $\hat{\vartheta}\in\bar{\Theta} = [\alpha,\beta]$, then we obtain the following well-known result: in the case of misspecification the pseudo-MLE $\hat{\vartheta}_\varepsilon$ converges to the value $\hat{\vartheta}$, which minimizes the Kullback-Leibler distance

$$\hat{\vartheta}_\varepsilon \longrightarrow \hat{\vartheta}. \quad (3.4)$$

Of course, if we have no misspecification, then $\hat{\vartheta}_\varepsilon \longrightarrow \vartheta_0$. It is interesting to note that in general case $\hat{\vartheta}\neq\vartheta_0$ but sometimes $\hat{\vartheta}=\vartheta_0$ and we consider the conditions of the consistency in such situations (see Examples 1 and 2 below).

3.1. Discontinuous versus smooth

Here we consider the situation where the true model (described by the signal $S(\vartheta,\cdot)$) is smooth w.r.t. ϑ but the theoretical model chosen by statistician has discontinuous signal $M(\vartheta,\cdot)$.

We start with one simple example. This example allows us to see that with the wrong models it is possible sometimes to have the consistency of the pseudo-MLEs. Moreover we can see as well the correct choice of the functions φ_ε and ψ_ε.

Example 1. Suppose that the observed process is (3.1), where the signal

$$S\left(\vartheta_0,t\right)=\delta^{-1}\left(t-\vartheta_0\right)\mathbb{1}_{\left\{\left|t-\vartheta_0\right|<\delta\right\}}+\operatorname{sgn}\left(t-\vartheta_0\right)\mathbb{1}_{\left\{\left|t-\vartheta_0\right|\geq\delta\right\}} \tag{3.5}$$

and $\vartheta_0\in\Theta=\left(\alpha,\beta\right)$, $0<\alpha<\beta<T$. Here $\alpha>\delta$ and $\beta<T-\delta$. The theoretical model is

$$\mathrm{d}X_t=\operatorname{sgn}\left(t-\vartheta\right)\mathrm{d}t+\varepsilon\mathrm{d}W_t,\quad X_0=0,\quad 0\leq t\leq T,\quad \vartheta\in\Theta. \tag{3.6}$$

This means that we observe the process (3.1) but we suppose that the observations are (3.6) and try to estimate the parameter ϑ_0 of the model (3.1) using the (wrong) model (3.6). These two signals $S\left(\vartheta_0,t\right)$ and $M\left(\vartheta,t\right)$ are presented on the Figure 1 below, where $\vartheta_0=3$ and $\vartheta=3$.

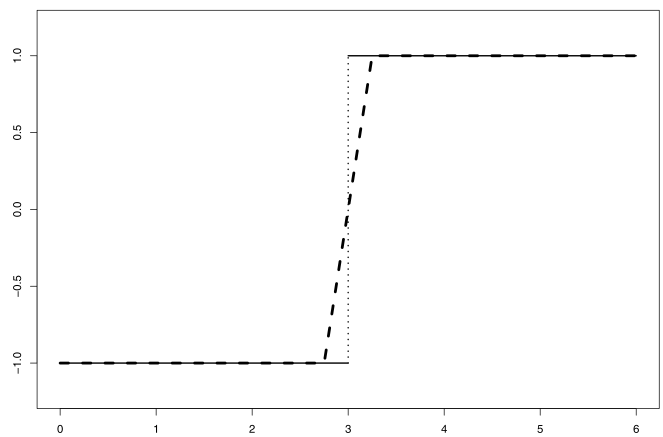

FIG 1. *Real $S\left(3,t\right)$ (dashed line) and theoretical $M\left(3,t\right)$ signals*

Introduce the pseudo-likelihood function

$$V\left(\vartheta,X^T\right)=\exp\left\{\frac{1}{\varepsilon^2}\int_0^T\operatorname{sgn}\left(t-\vartheta\right)\mathrm{d}X_t-\frac{T}{2\varepsilon^2}\right\},\qquad \vartheta\in\Theta$$

and define the pseudo-MLE $\hat\vartheta_\varepsilon$ by the relation

$$\hat\vartheta_\varepsilon=\arg\sup_{\vartheta\in\Theta}\int_0^T\operatorname{sgn}\left(t-\vartheta\right)\mathrm{d}X_t.$$

Note that

$$\hat\vartheta=\arg\inf_{\vartheta\in\Theta}\int_0^T\left[\operatorname{sgn}\left(t-\vartheta\right)-S\left(\vartheta_0,t\right)\right]^2\mathrm{d}t=\vartheta_0.$$

Hence the pseudo-MLE $\hat{\vartheta}_\varepsilon$ defined in this misspecified parameter estimation problem according to (3.4) is consistent.

To study its rate of convergence and the limit distribution we introduce the normalized likelihood ratio

$$Z_\varepsilon(u) = \frac{V\left(\vartheta_u, X^T\right)}{V\left(\vartheta_0, X^T\right)}, \qquad u \in \mathrm{U}_\varepsilon,$$

where $\vartheta_u = \vartheta_0 + \varphi_\varepsilon u$. Here $\varphi_\varepsilon \to 0$ will be chosen later and

$$\mathrm{U}_\varepsilon = \left(\frac{\alpha - \vartheta_0}{\varphi_\varepsilon}, \frac{\beta - \vartheta_0}{\varphi_\varepsilon}\right) \longrightarrow (-\infty, \infty)$$

as $\varepsilon \to 0$.

The substitution of the observation process in the likelihood ratio yields us the following expression (we suppose that $u > 0$)

$$\begin{aligned}
\ln Z_\varepsilon(u) &= \frac{1}{\varepsilon}\int_0^T \left[\operatorname{sgn}\left(t - \vartheta_0 - \varphi_\varepsilon u\right) - \operatorname{sgn}\left(t - \vartheta_0\right)\right] \mathrm{d}W_t \\
&\quad + \frac{1}{\varepsilon^2}\int_0^T \left[\operatorname{sgn}\left(t - \vartheta_0 - \varphi_\varepsilon u\right) - \operatorname{sgn}\left(t - \vartheta_0\right)\right] S\left(\vartheta, t_0\right) \mathrm{d}t \\
&= -\frac{2}{\varepsilon}\int_0^T \mathbb{1}_{\{\vartheta_0 < t < \vartheta_0 + \varphi_\varepsilon u\}} \mathrm{d}W_t - \frac{2}{\varepsilon^2}\int_0^T \mathbb{1}_{\{\vartheta_0 < t < \vartheta_0 + \varphi_\varepsilon u\}} S\left(\vartheta_0, t\right) \mathrm{d}t \\
&= -\frac{2}{\varepsilon}\left[W_{\vartheta_0 + \varphi_\varepsilon u} - W_{\vartheta_0}\right] - \frac{2}{\varepsilon^2}\int_{\vartheta_0}^{\vartheta_0 + \varphi_\varepsilon u} S\left(\vartheta_0, t\right) \mathrm{d}t \\
&= \frac{2\sqrt{\varphi_\varepsilon}}{\varepsilon} W_+(u) - \frac{\varphi_\varepsilon^2}{\delta\varepsilon^2} u^2 = \frac{2\sqrt{\varphi_\varepsilon}}{\varepsilon}\left[W_+(u) - \frac{\varphi_\varepsilon^{3/2}}{\delta\varepsilon}\frac{u^2}{2}\right],
\end{aligned} \tag{3.7}$$

where we denoted the Wiener process

$$W_+(u) = \varphi_\varepsilon^{-1/2}\left[W_{\vartheta_0 + \varphi_\varepsilon u} - W_{\vartheta_0}\right], \qquad u \in \left[0, \frac{\beta - \vartheta_0}{\varphi_\varepsilon}\right).$$

Therefore if we take

$$\varphi_\varepsilon = (\delta\varepsilon)^{2/3}, \qquad \psi_\varepsilon = \frac{\varepsilon^{2/3}}{2\delta^{1/3}},$$

then we can write

$$\hat{Z}_\varepsilon(u) = \left(Z_\varepsilon(u)\right)^{\psi_\varepsilon} = \exp\left\{W_+(u) - \frac{u^2}{2}\right\}, \qquad u \in \left[0, \frac{\beta - \vartheta_0}{\varphi_\varepsilon}\right).$$

For the negative u we obtain a similar representation

$$\hat{Z}_\varepsilon(u) = \left(Z_\varepsilon(u)\right)^{\psi_\varepsilon} = \exp\left\{W_-(-u) - \frac{u^2}{2}\right\}, \qquad u \in \left(\frac{\alpha - \vartheta_0}{\varphi_\varepsilon}, 0\right],$$

where $W_-(u), u \geq 0$ is a Wiener process independent of $W_+(u), u \geq 0$. Hence if we denote $W(\cdot)$ a two-sided Wiener process, then

$$\hat{Z}_\varepsilon(u) = (Z_\varepsilon(u))^{\psi_\varepsilon} = \exp\left\{W(u) - \frac{u^2}{2}\right\}, \qquad u \in \mathrm{U}_\varepsilon.$$

Now the properties of the pseudo-MLE $\hat{\vartheta}_\varepsilon$ follow from the relations (2.9) and (2.11) as follows:

$$\begin{aligned}
\mathbf{P}_{\vartheta_0}\left(\frac{\hat{\vartheta}_\varepsilon - \vartheta_0}{(\delta\varepsilon)^{2/3}} < x\right) &= \mathbf{P}_{\vartheta_0}\left(\hat{\vartheta}_\varepsilon < \vartheta_0 + (\delta\varepsilon)^{2/3} x\right) \\
&= \mathbf{P}_{\vartheta_0}\left\{\sup_{\vartheta < \vartheta_0 + \varphi_\varepsilon x} V\left(\vartheta, X^T\right) > \sup_{\vartheta \geq \vartheta_0 + \varphi_\varepsilon x} V\left(\vartheta, X^T\right)\right\} \\
&= \mathbf{P}_{\vartheta_0}\left\{\sup_{\vartheta < \vartheta_0 + \varphi_\varepsilon x} \frac{V\left(\vartheta, X^T\right)}{V\left(\vartheta_0, X^T\right)} > \sup_{\vartheta \geq \vartheta_0 + \varphi_\varepsilon x} \frac{V\left(\vartheta, X^T\right)}{V\left(\vartheta_0, X^T\right)}\right\} \\
&= \mathbf{P}_{\vartheta_0}\left\{\sup_{u < x, u \in \mathrm{U}_\varepsilon} Z_\varepsilon(u) > \sup_{u \geq x, u \in \mathrm{U}_\varepsilon} Z_\varepsilon(u)\right\} \\
&= \mathbf{P}_{\vartheta_0}\left\{\sup_{u < x, u \in \mathrm{U}_\varepsilon} \hat{Z}_\varepsilon(u) > \sup_{u \geq x, u \in \mathrm{U}_\varepsilon} \hat{Z}_\varepsilon(u)\right\} = \mathbf{P}_{\vartheta_0}\left(\hat{u}_\varepsilon < x\right), \quad (3.8)
\end{aligned}$$

where we denoted $\hat{u}_\varepsilon$ the solution of the following equation

$$W(\hat{u}_\varepsilon) - \frac{\hat{u}_\varepsilon^2}{2} = \sup_{u \in \mathrm{U}_\varepsilon}\left(W(u) - \frac{u^2}{2}\right).$$

Let us denote $\hat{Z}(u) = \exp\left\{w(u) - u^2/2\right\}, u \in \mathcal{R}$, where $w(\cdot)$ is two-sided Wiener process and note that for $u \in \mathrm{U}_\varepsilon$ we have (in distribution) the relation $\hat{Z}_\varepsilon(u) = \hat{Z}(u)$. We say that this equality is in distribution because the Wiener processes in $Z_\varepsilon(u)$ and $Z(u)$ are different and the Wiener process $W(\cdot)$ depends on ε. It can be shown that

$$\mathbf{P}_{\vartheta_0}(\hat{u}_\varepsilon < x) \longrightarrow \mathbf{P}_{\vartheta_0}(\hat{u} < x),$$

i.e.,

$$\hat{u}_\varepsilon \Longrightarrow \hat{u} = \arg\sup_{u \in R}\left(w(u) - \frac{u^2}{2}\right)$$

and we have the convergence of moments. Therefore we obtain the following

Proposition 1. *The pseudo-MLE $\hat{\vartheta}_\varepsilon$ in this problem is consistent, converges in distribution*

$$\frac{\hat{\vartheta}_\varepsilon - \vartheta_0}{(\delta\varepsilon)^{2/3}} \Longrightarrow \hat{u}$$

and the moments converge: for any $p > 0$

$$\lim_{\varepsilon\to 0}\mathbf{E}_{\vartheta_0}\left|\frac{\hat\vartheta_\varepsilon-\vartheta_0}{\left(\delta\varepsilon\right)^{2/3}}\right|^p=\mathbf{E}\left|\hat u\right|^p.$$

The proof follows from more general result of the Theorem 1 below.

Return now to the general smooth model of observations

$$\mathrm{d}X_t=S\left(\vartheta_0,t\right)\mathrm{d}t+\varepsilon\mathrm{d}W_t,\quad X_0=0,\quad 0\le t\le T \tag{3.9}$$

and the discontinuous theoretical model

$$\mathrm{d}X_t=M\left(\vartheta,t\right)\mathrm{d}t+\varepsilon\mathrm{d}W_t,\quad X_0=0,\quad 0\le t\le T, \tag{3.10}$$

where the signal

$$M\left(\vartheta,t\right)=h\left(t\right)1\!\!1_{\{t<\vartheta\}}+g\left(t\right)1\!\!1_{\{t\ge\vartheta\}}.$$

The unknown parameter $\vartheta\in\Theta=(\alpha,\beta)$ with $0<\alpha<\beta<T$. We observe a trajectory $X^T=(X_t,0\le t\le T)$ of the solution of the equation (3.9) and we want to estimate ϑ_0 supposing that the observed process is (3.10). Therefore we introduce the pseudo-likelihood ratio

$$\begin{aligned}V\left(\vartheta,X^T\right)=\exp\Bigg\{&\frac{1}{\varepsilon^2}\int_0^\vartheta h\left(t\right)\mathrm{d}X_t+\frac{1}{\varepsilon^2}\int_\vartheta^T g\left(t\right)\mathrm{d}X_t\\&-\frac{1}{2\varepsilon^2}\int_0^\vartheta h\left(t\right)^2\mathrm{d}t-\frac{1}{2\varepsilon^2}\int_\vartheta^T g\left(t\right)^2\mathrm{d}t\Bigg\},\quad \vartheta\in\Theta\end{aligned}$$

and define the pseudo-MLE $\hat\vartheta_\varepsilon$ by the equation (3.3).

Let us introduce the following notations:

$$\begin{aligned}\delta\left(t\right)&=h\left(t\right)-g\left(t\right),\qquad \Phi\left(\vartheta\right)=\int_0^T\left[M\left(\vartheta,t\right)-S\left(\vartheta_0,t\right)\right]^2\mathrm{d}t,\\ \ddot\Phi\left(\vartheta\right)&=2\left[h\left(\vartheta\right)-S\left(\vartheta_0,\vartheta\right)\right]\left[\dot h\left(\vartheta\right)-S'\left(\vartheta_0,\vartheta\right)\right]\\&\quad-2\left[g\left(\vartheta\right)-S\left(\vartheta_0,\vartheta\right)\right]\left[\dot g\left(\vartheta\right)-S'\left(\vartheta_0,\vartheta\right)\right],\\ \gamma(\hat\vartheta)&=\frac{\ddot\Phi(\hat\vartheta)}{2},\qquad \hat\vartheta\in\Theta,\\ \hat Z\left(u\right)&=\exp\left\{\delta(\hat\vartheta)W\left(u\right)-\frac{\gamma(\hat\vartheta)}{2}u^2\right\},\qquad u\in\mathcal{R},\\ Z^o\left(v\right)&=\exp\left\{w\left(v\right)-\frac{v^2}{2}\right\},\qquad v\in\mathcal{R}\end{aligned}$$

$$\hat{u}=\arg\sup_{u\in\mathcal{R}}\left[\delta(\hat{\vartheta})W\left(u\right)-\frac{\gamma(\hat{\vartheta})}{2}u^2\right],\qquad \hat{v}=\arg\sup_{u\in\mathcal{R}}\left[w\left(v\right)-\frac{v^2}{2}\right].$$

Here dot means differentiating w.r.t. ϑ, prime means differentiating w.r.t. t, $W\left(u\right), u\in R$ and $w\left(v\right), v\in R$ are two-sided Wiener processes.

Note that

$$\hat{u}=\hat{v}\left(\frac{\delta(\hat{\vartheta})}{\gamma(\hat{\vartheta})}\right)^{\frac{2}{3}}. \tag{3.11}$$

Indeed, let us put $u=rv$. Then we can write

$$\begin{aligned}\delta(\hat{\vartheta})W\left(u\right)-\frac{\gamma(\hat{\vartheta})}{2}u^2&=\sqrt{r}\delta(\hat{\vartheta})w\left(v\right)-\frac{\gamma(\hat{\vartheta})r^2}{2}v^2\\&=\sqrt{r}\delta(\hat{\vartheta})\left[w\left(v\right)-\frac{\gamma(\hat{\vartheta})r^{\frac{3}{2}}}{2\delta(\hat{\vartheta})}v^2\right]=\sqrt{r}\delta(\hat{\vartheta})\left[w\left(v\right)-\frac{v^2}{2}\right]\end{aligned}$$

if we put $r=\delta(\hat{\vartheta})^{\frac{2}{3}}\gamma(\hat{\vartheta})^{-\frac{2}{3}}$. Here $w\left(v\right)=r^{-1/2}W\left(rv\right)$. This proves (3.11).

Conditions $\mathcal{M}$.

1. $\inf_{t\in\Theta}\delta\left(t\right)>0$.
2. *The equation*

$$\int_0^{\hat{\vartheta}}\left[h\left(t\right)-S\left(\vartheta_0,t\right)\right]^2{\rm d}t+\int_{\hat{\vartheta}}^T\left[g\left(t\right)-S\left(\vartheta_0,t\right)\right]^2{\rm d}t=\inf_{\vartheta\in\Theta}\Phi\left(\vartheta\right)$$

   has a unique solution $\hat{\vartheta}=\hat{\vartheta}\left(\vartheta_0\right)\in\Theta$.
3. *The functions $h\left(t\right), g\left(t\right)$ and $S\left(\vartheta,t\right)$ are continuously differentiable w.r.t. $t\in\Theta$.*
4. $\inf_{\vartheta\in\Theta}\ddot{\Phi}\left(\vartheta\right)>0$.

The properties of the pseudo-MLE $\hat{\vartheta}_\varepsilon$ are described in the following theorem.

Theorem 1. *Let the conditions $\mathcal{M}$ be fulfilled then the estimator $\hat{\vartheta}_\varepsilon$ converges to the value $\hat{\vartheta}$, has the limit distribution*

$$\frac{\hat{\vartheta}_\varepsilon-\hat{\vartheta}}{\varepsilon^{2/3}}\Longrightarrow\hat{u}, \tag{3.12}$$

and for any $p>0$

$$\lim_{\varepsilon\to0}\mathbf{E}_{\vartheta_0}\left|\frac{\hat{\vartheta}_\varepsilon-\hat{\vartheta}}{\varepsilon^{2/3}}\right|^p=\mathbf{E}_{\vartheta_0}\left|\hat{u}\right|^p. \tag{3.13}$$

Proof. As before we study the normalized pseudo-likelihood ratio process

$$Z_\varepsilon\left(u\right)=\frac{V\left(\hat{\vartheta}+\varphi_\varepsilon u,X^T\right)}{V\left(\hat{\vartheta},X^T\right)},\qquad u\in\mathrm{U}_\varepsilon,$$

where $\varphi_\varepsilon = \varepsilon^{2/3}$. We have $(u > 0)$

$$\begin{aligned}
\ln Z_\varepsilon(u) &= \frac{1}{\varepsilon^2}\int_0^T \left[M\left(\hat\vartheta+\varphi_\varepsilon u,t\right) - M\left(\hat\vartheta,t\right)\right]\mathrm{d}X_t \\
&\quad - \frac{1}{2\varepsilon^2}\int_0^T \left[M\left(\hat\vartheta+\varphi_\varepsilon u,t\right)^2 - M\left(\hat\vartheta,t\right)^2\right]\mathrm{d}t \\
&= \frac{1}{\varepsilon}\int_{\hat\vartheta}^{\hat\vartheta+\varphi_\varepsilon u} \left[h(t) - g(t)\right]\mathrm{d}W_t \\
&\quad - \frac{1}{2\varepsilon^2}\int_{\hat\vartheta}^{\hat\vartheta+\varphi_\varepsilon u} \left(\left[h(t) - S(\vartheta_0,t)\right]^2 - \left[g(t) - S(\vartheta_0,t)\right]^2\right)\mathrm{d}t \\
&= \frac{\delta(\hat\vartheta)\sqrt{\varphi_\varepsilon}}{\varepsilon}\left[\frac{W_{\hat\vartheta+\varphi_\varepsilon u} - W_{\hat\vartheta}}{\sqrt{\varphi_\varepsilon}}\right] - \frac{\Phi\left(\hat\vartheta+\varphi_\varepsilon u\right) - \Phi\left(\hat\vartheta\right)}{2\varepsilon^2} + o(1) \\
&= \frac{\sqrt{\varphi_\varepsilon}}{\varepsilon}\delta(\hat\vartheta)W_+(u) - \frac{\varphi_\varepsilon^2 u^2}{4\varepsilon^2}\ddot\Phi(\hat\vartheta) + o(1) \\
&= \varepsilon^{-2/3}\left[\delta(\hat\vartheta)W_+(u) - \frac{\ddot\Phi(\hat\vartheta)}{2}\,\frac{u^2}{2}\right] + o(1).
\end{aligned}$$

Here we introduced the Wiener process

$$W_+(u) = \frac{W_{\hat\vartheta+\varphi_\varepsilon u} - W_{\hat\vartheta}}{\sqrt{\varphi_\varepsilon}}, \quad u \in \left[0, \frac{\beta-\hat\vartheta}{\varepsilon^{2/3}}\right),$$

and used in the expansion of $\Phi(\hat\vartheta+\varphi_\varepsilon u)$ the equality $\dot\Phi(\hat\vartheta) = 0$.

For the negative values $u < 0$ we obtain the similar representation

$$\ln Z_\varepsilon(u) = \varepsilon^{-2/3}\left[\delta(\hat\vartheta)W_-(-u) - \frac{\ddot\Phi(\hat\vartheta)}{2}\,\frac{u^2}{2}\right] + o(1)$$

with the independent Wiener process $W_-(u), u \geq 0$.

Let us put $\psi_\varepsilon = \varphi_\varepsilon = \varepsilon^{2/3}$ and introduce the random process

$$\hat Z_\varepsilon(u) = \left(Z_\varepsilon(u)\right)^{\psi_\varepsilon} = \exp\left\{\delta(\hat\vartheta)W(u) - \frac{\gamma(\hat\vartheta)}{2}u^2 + o(1)\right\}, \quad u \in \mathrm{U}_\varepsilon. \quad (3.14)$$

We define $Z_\varepsilon(u)$ linearly decreasing to zero on the interval

$$\left[\frac{\beta-\hat\vartheta}{\varepsilon^{2/3}}, \frac{\beta-\hat\vartheta}{\varepsilon^{2/3}} + 1\right]$$

and increasing from zero to $\hat Z_\varepsilon\left(\frac{\alpha-\hat\vartheta}{\varepsilon}\right)$ on the interval

$$\left[\frac{\alpha-\hat\vartheta}{\varepsilon^{2/3}} - 1, \frac{\alpha-\hat\vartheta}{\varepsilon^{2/3}}\right].$$

Further we put $\hat{Z}_\varepsilon\left(u\right)=0$ for

$$u\notin\left[\frac{\alpha-\hat{\vartheta}}{\varepsilon^{2/3}}-1,\frac{\beta-\hat{\vartheta}}{\varepsilon^{2/3}}+1\right].$$

Now the process $\hat{Z}_\varepsilon\left(u\right)$ is defined for all $u\in\mathcal{R}$. Note that this process is continuous with probability 1.

Let us denote by $\mathbf{Q}_{\vartheta_0,\varepsilon}$ the measure induced by this process in the space $\mathcal{C}_0\left(\mathcal{R}\right)$ of continuous functions decreasing to zero at infinity. The corresponding measurable space we denote as $\left(\mathcal{C}_0\left(\mathcal{R}\right),\mathcal{B}\right)$, where $\mathcal{B}$ is Borelian σ-algebra. By $\mathbf{Q}_{\vartheta_0}$ we denote the measure of the limit process $\hat{Z}\left(\cdot\right)$. According to the approach of Ibragimov-Khasminskii we have to verify the uniform on compacts $\mathrm{K}\subset\Theta$ weak convergence

$$\mathbf{Q}_{\vartheta_0,\varepsilon}\quad\Longrightarrow\quad\mathbf{Q}_{\vartheta_0}. \tag{3.15}$$

The next lemmas provide us the convergence of finite-dimensional distributions and two properties of the random process $\hat{Z}_\varepsilon\left(\cdot\right)$ allowing to prove the tightness of the family of measures $\{\mathbf{Q}_{\vartheta_0,\varepsilon},\vartheta\in\Theta\}$. As the pseudo-MLE $\hat{u}_\varepsilon=\varphi_\varepsilon^{-1}(\hat{\vartheta}_\varepsilon-\hat{\vartheta})$ is a continuous functional on the space $\left(\mathcal{C}_0\left(\mathcal{R}\right),\mathcal{B}\right)$, i.e.,

$$\hat{u}_\varepsilon=\Psi(\hat{Z}_\varepsilon)=\arg\sup_{u\in\mathcal{R}}\hat{Z}_\varepsilon\left(u\right)$$

we obtain from (3.15) the convergence $\hat{u}_\varepsilon\Longrightarrow\hat{u}$. Remind that the process $\hat{Z}_\varepsilon\left(u\right)$ is defined on $\mathcal{R}$.

From the representation (3.14) we obtain immediately the first lemma. □

Lemma 1. *The finite-dimensional distributions of $\hat{Z}_\varepsilon\left(\cdot\right)$ converge, i.e.,: for any set $u_1,\ldots,u_k$ and any $k=1,2,\ldots$*

$$\left(\hat{Z}_\varepsilon\left(u_1\right),\ldots,\hat{Z}_\varepsilon\left(u_k\right)\right)\Longrightarrow\left(\hat{Z}\left(u_1\right),\ldots,\hat{Z}\left(u_k\right)\right). \tag{3.16}$$

This convergence is uniform in ϑ on compacts $\mathrm{K}\subset\Theta$.

We need the following elementary estimate

Lemma 2. *There exists a constant $\kappa>0$ such that*

$$\Phi\left(\vartheta\right)-\Phi(\hat{\vartheta})\geq\kappa\left(\vartheta-\hat{\vartheta}\right)^2. \tag{3.17}$$

Proof. As the point $\hat{\vartheta}$ is a unique minimum of the function $\Phi\left(\vartheta\right)$, we can write for any $\nu>0$

$$m\left(\nu\right)=\inf_{\left|\vartheta-\hat{\vartheta}\right|>\nu}\Phi\left(\vartheta\right)-\Phi(\hat{\vartheta})>0.$$

Hence for $\left|\vartheta-\hat{\vartheta}\right|>\nu$

$$\Phi\left(\vartheta\right)-\Phi(\hat{\vartheta})\geq m\left(\nu\right)\geq m\left(\nu\right)\frac{\left(\vartheta-\hat{\vartheta}\right)^2}{\left(\beta-\alpha\right)^2}.$$

Further, for the values $\left|\vartheta-\hat\vartheta\right|\leq \nu$ by Taylor expansion we have

$$\Phi\left(\vartheta\right)-\Phi(\hat\vartheta)=\frac{1}{2}\ddot\Phi(\hat\vartheta)\left(\vartheta-\hat\vartheta\right)^2\left(1+o\left(1\right)\right).$$

Therefore for sufficiently small ν we can write

$$\Phi\left(\vartheta\right)-\Phi(\hat\vartheta)\geq\frac{1}{4}\ddot\Phi(\hat\vartheta)\left(\vartheta-\hat\vartheta\right)^2.$$

Taking

$$\kappa=\min\left(\frac{m\left(\nu\right)}{\left(\beta-\alpha\right)^2},\frac{1}{4}\ddot\Phi(\hat\vartheta)\right)$$

we obtain (3.17).

This estimate allows us to verify the boundedness of all moments of the pseudo likelihood ratio process. □

Lemma 3. *For any $p>0$ there exist constants $c>0$ and $d>0$ such that for all $|u|\geq d$*

$$\mathbf{E}_{\vartheta_0}\hat Z_\varepsilon^p\left(u\right)\leq e^{-cu^2}. \tag{3.18}$$

Proof. Indeed, we have

$$\mathbf{E}_{\vartheta_0}\hat Z_\varepsilon\left(u\right)^p=\exp\left\{\frac{p^2\varepsilon^{-2/3}}{2}\int_{\hat\vartheta}^{\hat\vartheta+\varphi_\varepsilon u}\delta\left(t\right)^2{\rm d}t-p\varepsilon^{-4/3}\left[\Phi(\hat\vartheta+\varphi_\varepsilon u)-\Phi(\hat\vartheta)\right]\right\}.$$

Now the estimate (3.18) follows from the relations

$$\varepsilon^{-2/3}\int_{\hat\vartheta}^{\hat\vartheta+\varphi_\varepsilon u}\delta\left(t\right)^2{\rm d}t\leq\sup_{t\in\Theta}\delta\left(t\right)^2\left|u\right|,$$
$$\varepsilon^{-4/3}\left[\Phi(\hat\vartheta+\varphi_\varepsilon u)-\Phi(\hat\vartheta)\right]\geq\kappa u^2,$$

where we used (3.17). Therefore we obtain the estimate (3.18) with some $c>0$ and $d>0$. □

Lemma 4. *For any $N>0$ and $|u_1|<N,|u_2|<N$ we have the estimate*

$$\mathbf{E}_{\vartheta_0}\left|\hat Z_\varepsilon\left(u_2\right)-\hat Z_\varepsilon\left(u_1\right)\right|^4\leq C\left(1+N^2\right)\left|u_2-u_1\right|^2 \tag{3.19}$$

with some constant $C>0$.

Proof. Let us denote

$$a_t=\varepsilon^{-1/3}\delta\left(t\right),\qquad b_t=-\varepsilon^{-2/3}\delta\left(t\right)\left[h\left(t\right)+g\left(t\right)-2S\left(\vartheta_0,t\right)\right],$$
$$G\left(t\right)=\exp\left\{\int_{\hat\vartheta+\varphi_\varepsilon u_1}^{t}a_s{\rm d}W_s+\int_{\hat\vartheta+\varphi_\varepsilon u_1}^{t}b_s{\rm d}s\right\}$$

Note that

$$G\left(\hat\vartheta+\varphi_\varepsilon u_2\right)=\frac{\hat Z_\varepsilon\left(u_2\right)}{\hat Z_\varepsilon\left(u_1\right)}.$$

The process $G\left(t\right)$ has stochastic differential

$$\mathrm{d}G\left(t\right)=G\left(t\right)\left[b_t+\frac{a_t^2}{2}\right]\mathrm{d}t+G\left(t\right)a_t\mathrm{d}W_t,\quad G\left(\hat\vartheta+\varphi_\varepsilon u_1\right)=1.$$

Therefore

$$G\left(\hat\vartheta+\varphi_\varepsilon u_2\right)=1+\int_{\hat\vartheta+\varphi_\varepsilon u_1}^{\hat\vartheta+\varphi_\varepsilon u_2}G\left(t\right)\left[b_t+\frac{a_t^2}{2}\right]\mathrm{d}t+\int_{\hat\vartheta+\varphi_\varepsilon u_1}^{\hat\vartheta+\varphi_\varepsilon u_2}G\left(t\right)a_t\mathrm{d}W_t.$$

We write

$$\begin{aligned}
\mathbf{E}_{\vartheta_0}\left|\hat Z_\varepsilon\left(u_2\right)-\hat Z_\varepsilon\left(u_1\right)\right|^4&=\mathbf{E}_{\vartheta_0}\hat Z_\varepsilon\left(u_1\right)^4\left|G\left(\hat\vartheta+\varphi_\varepsilon u_2\right)-1\right|^4\\
&=\mathbf{E}_{\vartheta_0}\hat Z_\varepsilon\left(u_1\right)^4\left|\int_{\hat\vartheta+\varphi_\varepsilon u_1}^{\hat\vartheta+\varphi_\varepsilon u_2}G\left(t\right)\left[b_t+\frac{a_t^2}{2}\right]\mathrm{d}t+\int_{\hat\vartheta+\varphi_\varepsilon u_1}^{\hat\vartheta+\varphi_\varepsilon u_2}G\left(t\right)a_t\mathrm{d}W_t\right|^4\\
&\leq C_1\mathbf{E}_{\vartheta_0}\hat Z_\varepsilon\left(u_1\right)^4\left|\int_{\hat\vartheta+\varphi_\varepsilon u_1}^{\hat\vartheta+\varphi_\varepsilon u_2}G\left(t\right)\left[b_t+\frac{a_t^2}{2}\right]\mathrm{d}t\right|^4\\
&\qquad+C_2\mathbf{E}_{\vartheta_0}\hat Z_\varepsilon\left(u_1\right)^4\left|\int_{\hat\vartheta+\varphi_\varepsilon u_1}^{\hat\vartheta+\varphi_\varepsilon u_2}G\left(t\right)a_t\mathrm{d}W_t\right|^4\\
&\leq C_1\left(u_2-u_1\right)^3\varphi_\varepsilon^3\int_{\hat\vartheta+\varphi_\varepsilon u_1}^{\hat\vartheta+\varphi_\varepsilon u_2}\mathbf{E}_{\vartheta_0}\hat Z_\varepsilon\left(u_1\right)^4G\left(t\right)^4\left|b_t+\frac{a_t^2}{2}\right|^4\mathrm{d}t\\
&\qquad+C_2\left(\mathbf{E}_{\vartheta_0}\hat Z_\varepsilon\left(u_1\right)^8\mathbf{E}_{\vartheta_0}\left|\int_{\hat\vartheta+\varphi_\varepsilon u_1}^{\hat\vartheta+\varphi_\varepsilon u_2}G\left(t\right)a_t\mathrm{d}W_t\right|^8\right)^{1/2}.
\end{aligned}$$

For stochastic integral we have the estimate

$$\begin{aligned}
\mathbf{E}_{\vartheta_0}\left|\int_{\hat\vartheta+\varphi_\varepsilon u_1}^{\hat\vartheta+\varphi_\varepsilon u_2}G\left(t\right)a_t\mathrm{d}W_t\right|^8&\leq C\mathbf{E}_{\vartheta_0}\left(\int_{\hat\vartheta+\varphi_\varepsilon u_1}^{\hat\vartheta+\varphi_\varepsilon u_2}G\left(t\right)^2a_t^2\mathrm{d}t\right)^4\\
&\leq\left(u_2-u_1\right)^3\varphi_\varepsilon^3\int_{\hat\vartheta+\varphi_\varepsilon u_1}^{\hat\vartheta+\varphi_\varepsilon u_2}a_t^8\,\mathbf{E}_{\vartheta_0}G\left(t\right)^8\,\mathrm{d}t.
\end{aligned}$$

Further

$$\mathbf{E}_{\vartheta_0}G\left(t\right)^8=\exp\left\{32\int_{\hat\vartheta+\varphi_\varepsilon u_1}^{t}a_s^2\,\mathrm{d}s-8\varepsilon^{-4/3}\left[\Phi\left(t\right)-\Phi\left(\hat\vartheta+\varphi_\varepsilon u_1\right)\right]\right\},$$

$$\mathbf{E}_{\vartheta_0}\hat{Z}_\varepsilon\left(u_1\right)^8 = \exp\left\{32\int_0^{\hat{\vartheta}+\varphi_\varepsilon u_1} a_s^2\,\mathrm{d}s - 8\varepsilon^{-4/3}\Phi\left(\hat{\vartheta}+\varphi_\varepsilon u_1\right)\right\}$$

Hence

$$\begin{aligned}\mathbf{E}_{\vartheta_0}\left|\hat{Z}_\varepsilon\left(u_2\right)-\hat{Z}_\varepsilon\left(u_1\right)\right|^4 &\leq C\left|u_2-u_1\right|^4 + C\left|u_2-u_1\right|^2\\ &\leq C\left(1+N^2\right)\left|u_2-u_1\right|^2\end{aligned}$$

for $|u_1| \leq N$ and $|u_2| \leq N$.

Now the properties (3.12) and (3.13) of the pseudo-MLE $\hat{\vartheta}_\varepsilon$ follow from the Lemmae 1, 3, 4 and the Theorem 1.10.1 in [7]. □

Remark 1. Note that as $\hat{\vartheta} \in \Theta$ is the point of minimum of the function $\Phi\left(\vartheta\right)$ we have the equality

$$\dot{\Phi}(\hat{\vartheta}) = \left[h(\hat{\vartheta}) - S(\vartheta_0,\hat{\vartheta})\right]^2 - \left[g(\hat{\vartheta}) - S(\vartheta_0,\hat{\vartheta})\right]^2 = 0, \tag{3.20}$$

which is equivalent to

$$\left(h(\hat{\vartheta}) - g(\hat{\vartheta})\right)\left[h(\hat{\vartheta}) + g(\hat{\vartheta}) - 2S(\vartheta_0,\hat{\vartheta})\right] = 0.$$

Hence the point $\hat{\vartheta}$ satisfies to the equality

$$S(\vartheta_0,\hat{\vartheta}) = \frac{h(\hat{\vartheta}) + g(\hat{\vartheta})}{2}.$$

The equation

$$S(\vartheta_0,t) = \frac{h(t)+g(t)}{2}, \qquad \alpha < t < \beta \tag{3.21}$$

can have many solutions corresponding to the local extremes of the function $\Phi\left(t\right), t \in \Theta$. If the equation (3.21) has no solution, say,

$$S(\vartheta_0,t) < \frac{h(t)+g(t)}{2}, \qquad \alpha < t < \beta, \tag{3.22}$$

then $\hat{\vartheta} = \alpha$. Otherwise $\hat{\vartheta} = \beta$. In these two cases the behavior of the estimator $\hat{\vartheta}_\varepsilon$ can be studied as it was done in [9], Section 2.8. If we have the equality

$$S(\vartheta_0,t) = \frac{h(t)+g(t)}{2}, \qquad a \leq t \leq b,$$

on some interval $[a,b]$, then any point of this interval can be taken as $\hat{\vartheta}$. We do not study here the properties of $\hat{\vartheta}_\varepsilon$ in this situation and in the situation when the function $\Phi\left(\vartheta\right), \alpha < \vartheta < \beta$ has two or more points of minimum. Note that such study can be done by the same way as in [9], Section 2.7.

Example 2. Choosing different smooth signals in the class

$$\mathcal{S} = \left\{ S\left(t-\vartheta\right) = \operatorname{sgn}\left(t-\vartheta\right)\left|t-\vartheta\right|^{\kappa}, \kappa > \frac{1}{2} \right\}$$

and the same *theoretical model* (3.6) we can obtain different rates of convergence of estimators. Note that once more we have $\hat{\vartheta} = \vartheta_0$ and the pseudo-MLE $\hat{\vartheta}_\varepsilon$ is consistent. The functions φ_ε and ψ_ε can be chosen as follows.

Let us fix some $\kappa \in (\frac{1}{2}, \infty)$. Then the corresponding calculations like (3.7) provide us the expression $(u > 0)$

$$\ln Z_\varepsilon\left(u\right) = \frac{2\sqrt{\varphi_\varepsilon}}{\varepsilon}\left[W_+\left(u\right) - \frac{\varphi_\varepsilon^{\frac{1}{2}+\kappa}}{\varepsilon}\frac{u^{1+\kappa}}{\left(1+\kappa\right)}\right].$$

Therefore if we put

$$\varphi_\varepsilon = \varepsilon^{\frac{2}{2\kappa+1}}, \qquad \psi_\varepsilon = \frac{\varepsilon^{\frac{2\kappa}{2\kappa+1}}}{2},$$

then

$$\hat{Z}_\varepsilon\left(u\right) = \exp\left\{W\left(u\right) - \frac{\left|u\right|^{1+\kappa}}{1+\kappa}\right\}, \qquad u \in \mathrm{U}_\varepsilon$$

and the pseudo-MLE $\hat{\vartheta}_\varepsilon$ satisfies the relations

$$\frac{\hat{\vartheta}_\varepsilon - \vartheta_0}{\varepsilon^{\frac{2}{2\kappa+1}}} = \hat{u}_\varepsilon = \arg\sup_{u\in\mathrm{U}_\varepsilon}\left[W\left(u\right) - \frac{\left|u\right|^{1+\kappa}}{1+\kappa}\right] \Longrightarrow \hat{u},$$

where

$$\hat{u} = \arg\sup_{u\in\mathcal{R}}\left[w\left(u\right) - \frac{\left|u\right|^{1+\kappa}}{1+\kappa}\right].$$

Therefore choosing different $\kappa > \frac{1}{2}$ we can obtain any rate $\varepsilon^\gamma, \gamma < 1$ of convergence of pseudo-MLE:

$$\frac{\hat{\vartheta}_\varepsilon - \vartheta_0}{\varepsilon^\gamma} \Longrightarrow \hat{u}.$$

We do not present here the details of the proof of this convergence, but it can be obtained following the same arguments as in the proof of the Theorem 1.

We see that the $\hat{\vartheta}_\varepsilon$ has a "bad" rate of convergence. Note that for other estimators the rate can be better.

Let us study the *trajectory fitting estimator* (TFE) ϑ_ε^* defined by the relation

$$\vartheta_\varepsilon^* = \arg\inf_{\vartheta\in\Theta}\int_0^T\left[X_t - m\left(\vartheta,t\right)\right]^2 \mathrm{d}t, \tag{3.23}$$

where

$$m(\vartheta,t)=\int_0^t M(\vartheta,s)\,\mathrm{d}s.$$

To describe the asymptotic properties of this estimator we need a slightly different regularity conditions. Suppose that the function

$$\Psi(\vartheta)=\int_0^T\left[m(\vartheta,t)-s(\vartheta_0,t)\right]^2\mathrm{d}t,\qquad \vartheta\in\Theta$$

has a unique minimum at the point $\vartheta^*\in\Theta$. Here

$$s(\vartheta_0,t)=\int_0^t S(\vartheta_0,v)\,\mathrm{d}v.$$

Then the TFE ϑ_ε^* under regularity conditions admits the representation

$$\frac{\vartheta_\varepsilon^*-\vartheta^*}{\varepsilon}=\frac{\int_0^T W_t\,\dot m(\vartheta^*,t)\,\mathrm{d}t}{\int_0^T \dot m(\vartheta^*,t)^2\,\mathrm{d}t}\left(1+o(1)\right).$$

Therefore this estimator is asymptotically normal with the rate ε. The details of the proof can be found in the Section 7.4 in [9].

3.2. Smooth versus discontinuous

Suppose that the true model (3.9) has discontinuous trend coefficient $S(\vartheta_0,t)$ of the following form

$$\mathrm{d}X_t=\left[h(t)\,\mathbb{1}_{\{t<\vartheta_0\}}+g(t)\,\mathbb{1}_{\{t\geq\vartheta_0\}}\right]\mathrm{d}t+\varepsilon\mathrm{d}W_t,\quad X_0=0,\quad 0\leq t\leq T,\tag{3.24}$$

where $\vartheta_0\in\Theta=(\alpha,\beta)$, $0<\alpha<\beta<T$, but the statistician uses the model

$$\mathrm{d}X_t=M(\vartheta,t)\,\mathrm{d}t+\varepsilon\mathrm{d}W_t,\quad X_0=0,\quad 0\leq t\leq T,\tag{3.25}$$

with the “smooth” signal $M(\vartheta,\cdot)$. The pseudo-likelihood ratio $V\left(\vartheta,X^T\right)$ and the pseudo-MLE $\hat\vartheta_\varepsilon$ are defined by the same relations (3.2), (3.3). As before, we are interested in the asymptotic behavior of $\hat\vartheta_\varepsilon$ as $\varepsilon\to0$.

To show that the situation is quite different we start with the example which is “symmetric” to the Example 1.

Example 3. Suppose that the observed process is

$$\mathrm{d}X_t=\operatorname{sgn}(t-\vartheta_0)\,\mathrm{d}t+\varepsilon\mathrm{d}W_t,\quad X_0=0,\quad 0\leq t\leq T,$$

and we use the model

$$\mathrm{d}X_t=M(\vartheta_0,t)\,\mathrm{d}t+\varepsilon\mathrm{d}W_t,\quad X_0=0,\quad 0\leq t\leq T,$$

with

$$M(\vartheta,t)=\delta^{-1}(t-\vartheta)\,\mathbb{1}_{\{|t-\vartheta|<\delta\}}+\operatorname{sgn}(t-\vartheta)\,\mathbb{1}_{\{|t-\vartheta|\geq\delta\}},$$

where the parameter $\vartheta\in\Theta=(\alpha,\beta)$. Here $0<\delta<\alpha<\beta<T-\delta$.

These two signals $S(\vartheta,t)=\operatorname{sgn}(t-\vartheta)$ and $M(\vartheta,t)$ are presented on the Figure 1.

It is easy to see that the function

$$\Phi(\vartheta)=\int_0^T\left[M(\vartheta,t)-\operatorname{sgn}(t-\vartheta_0)\right]^2\mathrm{d}t,\qquad \vartheta\in\Theta$$

attained its minimum at the point

$$\hat\vartheta=\vartheta_0$$

and therefore the pseudo-MLE is consistent

$$\hat\vartheta_\varepsilon\longrightarrow\vartheta_0.$$

More detailed analysis shows that it has asymptotically Gaussian distribution

$$\frac{\hat\vartheta_\varepsilon-\vartheta_0}{\varepsilon}\sim\mathcal{N}\left(0,D^2\right)$$

with the "regular" rate ε of convergence and some limit variance $D^2>0$.

Let us return to the problem with the equations (3.24), (3.25). Introduce the notations

$$\begin{aligned}
\Phi(\vartheta)&=\frac{1}{2}\int_0^T\left[M(\vartheta,t)-S(\vartheta_0,t)\right]^2\mathrm{d}t,\qquad \Phi(\hat\vartheta)=\inf_{\vartheta\in\Theta}\Phi(\vartheta),\\
\ddot\Phi(\hat\vartheta)&=\int_0^T\ddot M(\hat\vartheta,t)\left[M(\hat\vartheta,t)-S(\vartheta_0,t)\right]\mathrm{d}t+\int_0^T\dot M(\hat\vartheta,t)^2\,\mathrm{d}t,\\
\mathrm{I}(\vartheta)&=\int_0^T\dot M(\vartheta,t)^2\,\mathrm{d}t,\qquad \mathrm{D}(\vartheta_0)^2=\ddot\Phi(\hat\vartheta)^{-2}\mathrm{I}(\hat\vartheta).
\end{aligned}$$

and note that $\hat\vartheta=\hat\vartheta(\vartheta_0)$.

The conditions of regularity:

Conditions $\mathcal{R}$.

1. *The functions $h(\cdot)$ and $g(\cdot)$ are bounded and $\inf_{t\in\Theta}|h(t)-g(t)|>0$.*
2. *The function $\Phi(\vartheta),\vartheta\in\Theta$ has a unique minimum at the point $\hat\vartheta\in\Theta$.*
3. *The function $M(\vartheta,t)$ is two times continuously differentiable w.r.t. ϑ.*
4. *We have*

$$\inf_{\vartheta_0\in\Theta}\ddot\Phi(\hat\vartheta)>0.$$

Theorem 2. *Let the conditions $\mathcal{R}$ be fulfilled, then the estimator $\hat\vartheta_\varepsilon$ converges to the value $\hat\vartheta$, is asymptotically normal*

$$\frac{\hat\vartheta_\varepsilon - \hat\vartheta}{\varepsilon} \Longrightarrow \hat\xi \sim \mathcal{N}\left(0, \mathrm{D}\left(\vartheta_0\right)^2\right),$$

and for any $p > 0$

$$\lim_{\varepsilon\to 0} \mathbf{E}_{\vartheta_0} \left|\frac{\hat\vartheta_\varepsilon - \hat\vartheta}{\varepsilon}\right|^p = \mathbf{E}_{\vartheta_0} \left|\hat\xi\right|^p.$$

Proof. Using the Taylor expansion we can write for the pseudo-likelihood ratio

$$Z_\varepsilon\left(u\right) = \frac{V\left(\hat\vartheta + \varepsilon u, X^T\right)}{V\left(\hat\vartheta, X^T\right)}, \quad u \in \mathrm{U}_\varepsilon = \left(\frac{\alpha - \hat\vartheta}{\varepsilon}, \frac{\beta - \hat\vartheta}{\varepsilon}\right)$$

the presentation

$$\ln Z_\varepsilon\left(u\right) = u \int_0^T \dot M(\hat\vartheta, t) \mathrm{d}W_t - \frac{u^2}{2} \ddot\Phi(\hat\vartheta) + o\left(1\right).$$

Therefore, if we denote

$$Z\left(u\right) = \exp\left\{u \int_0^T \dot M(\hat\vartheta, t) \mathrm{d}W_t - \frac{u^2}{2} \ddot\Phi(\hat\vartheta)\right\}, \quad u \in R,$$

then we obtain the first lemma. □

Lemma 5. *We have the convergence of finite-dimensional distributions of the process $Z_\varepsilon\left(\cdot\right)$: for any set $u_1, \ldots, u_k$ and any $k = 1, 2, \ldots$*

$$\left(Z_\varepsilon\left(u_1\right), \ldots, Z_\varepsilon\left(u_k\right)\right) \Longrightarrow \left(Z\left(u_1\right), \ldots, Z\left(u_k\right)\right). \tag{3.26}$$

This convergence is uniform in ϑ on compacts $\mathrm{K} \subset \Theta$.

The next lemma can be proved following the same arguments as in the proof of the Lemma 2.

Lemma 6. *There exists a constant $\kappa > 0$ such that*

$$\Phi\left(\vartheta\right) - \Phi(\hat\vartheta) \geq \kappa \left(\vartheta - \hat\vartheta\right)^2. \tag{3.27}$$

Note that now the moments of $Z_\varepsilon\left(u\right)$ are no more bounded. Denote $\vartheta_u = \hat\vartheta + \varepsilon u$, then for any $\psi > 0$, we can write

$$\mathbf{E}_{\vartheta_0} Z_\varepsilon\left(u\right)^\psi$$
$$= \exp\left\{\left(\frac{\psi^2}{2\varepsilon^2} \int_0^T \left[M(\vartheta_u, t) - M(\hat\vartheta, t)\right]^2 \mathrm{d}t - \frac{\psi}{\varepsilon^2} \left[\Phi(\vartheta_u) - \Phi(\hat\vartheta)\right]\right)\right\}$$

$$\leq \exp\left\{\left(\frac{\psi^2}{2}\,M - \frac{\psi\kappa}{2}\right)\,u^2\right\} = 1, \tag{3.28}$$

where we denoted

$$M = \sup_{\vartheta\in\Theta}\int_0^T \dot M(\vartheta,t)^2 \mathrm{d}t$$

and put $\psi = M^{-1}\kappa$.

Therefore we introduce the following normalized likelihood ratio

$$\hat Z_\varepsilon\left(u\right) = Z_\varepsilon\left(u\right)^{\psi}, \qquad u\in \mathrm{U}_\varepsilon$$

and establish the properties of this process similar to $\hat Z_\varepsilon\left(\cdot\right)$ in Lemmae 3 and 4.

Lemma 7. *Suppose that the conditions $\mathcal{R}$ are fulfilled, then we have the estimates*

$$\mathbf{E}_{\vartheta_0}\hat Z_\varepsilon^{1/2}\left(u\right) \leq e^{-\kappa u^2}, \tag{3.29}$$

$$\mathbf{E}_{\vartheta_0}\left|\hat Z_\varepsilon^{1/2}\left(u_2\right) - \hat Z_\varepsilon^{1/2}\left(u_1\right)\right|^2 \leq C\left(1+N^2\right)\left(u_2-u_1\right)^2 \tag{3.30}$$

for $|u_1|<N, |u_2|<N$

Proof. The first estimate (3.29) we obtain immediately from (3.28). The proof of the second estimate (3.30) can be carried out like the proof of the relation (3.19).

The properties of the process $\hat Z_\varepsilon\left(\cdot\right)$ established in the Lemmas 5 and 7 allows to cite Theorem 1.10.1 in [7] and to obtain the announced in the Theorem 2 properties of the pseudo-MLE $\hat\vartheta_\varepsilon$. □

3.3. Discontinuous versus discontinuous

Let us remind that if the the observed model is discontinuous and the statistician knows this but takes the wrong signals before and after the jump, then nevertheless it is possible to have the consistent estimation. Consider the following problem of parameter estimation in the situation of misspecification. The *theoretical model* is

$$\mathrm{d}X_t = \left[h\left(t\right)\mathbb{1}_{\{t<\vartheta\}} + g\left(t\right)\mathbb{1}_{\{t\geq\vartheta\}}\right]\mathrm{d}t + \varepsilon \mathrm{d}W_t,\quad X_0=0,\quad 0\leq t\leq T,$$

where $\vartheta\in\Theta=(\alpha,\beta)$, $0<\alpha<\beta<T$. Suppose that $h\left(t\right)-g\left(t\right)>0$ for $t\in[\alpha,\beta]$. The observed stochastic process has a different equation

$$\mathrm{d}X_t = \left[\left[h\left(t\right)+q\left(t\right)\right]\mathbb{1}_{\{t<\vartheta_0\}} + \left[g\left(t\right)+r\left(t\right)\right]\mathbb{1}_{\{t\geq\vartheta_0\}}\right]\mathrm{d}t + \varepsilon \mathrm{d}W_t,\quad 0\leq t\leq T,$$

where $q\left(t\right)$ and $r\left(t\right)$ are some unknown functions.

We study the conditions on $q\left(t\right)$ and $r\left(t\right)$ which allow the consistent estimation of the parameter ϑ_0.

The function $\Phi(\vartheta)$ for $\vartheta < \vartheta_0$ is

$$\Phi(\vartheta) = \int_0^{\vartheta} q(t)^2 \,\mathrm{d}t + \int_{\vartheta}^{\vartheta_0} \left[h(t) + q(t) - g(t)\right]^2 \mathrm{d}t + \int_{\vartheta_0}^{T} r(t)^2 \,\mathrm{d}t.$$

Hence

$$\begin{aligned}\frac{\mathrm{d}\Phi(\vartheta)}{\mathrm{d}\vartheta} &= q(\vartheta)^2 - \left[h(\vartheta) - g(\vartheta) + q(\vartheta)\right]^2 \\ &= -\left(h(\vartheta) - g(\vartheta)\right)\left[h(\vartheta) - g(\vartheta) + 2q(\vartheta)\right].\end{aligned}$$

If the function

$$q(\vartheta) > \frac{g(\vartheta) - h(\vartheta)}{2}, \qquad \vartheta \in \Theta, \tag{3.31}$$

then for $\vartheta < \vartheta_0$

$$\frac{\mathrm{d}\Phi(\vartheta)}{\mathrm{d}\vartheta} < 0.$$

For $\vartheta > \vartheta_0$ under condition

$$r(\vartheta) < \frac{h(\vartheta) - g(\vartheta)}{2} \tag{3.32}$$

we obtain the similar inequality

$$\frac{\mathrm{d}\Phi(\vartheta)}{\mathrm{d}\vartheta} > 0.$$

Therefore

$$\hat{\vartheta} = \arg\inf_{\vartheta\in\Theta} \Phi(\vartheta) = \vartheta_0$$

and we obtain the following result.

Proposition 2. *If the conditions (3.31) and (3.32) are fulfilled then the pseudo-MLE $\hat{\vartheta}_\varepsilon$ is consistent.*

It can be shown that

$$\frac{\hat{\vartheta}_\varepsilon - \vartheta_0}{\varepsilon^2} \Longrightarrow \xi$$

For the details see the similar problem in Section 5.3, [9]. The close problem of change-point detection for misspecified diffusion processes are studied in [1].

3.4. Discussion

There are several other interesting problems of misspecification in regularity, which can be studied by the proposed here approach.

One of them is to study the asymptotic behavior of the Bayesian estimator $\tilde{\vartheta}_\varepsilon$ (see (2.2)) in the situation described by the equations (3.24), (3.25).

It can be shown that $\tilde{\vartheta}_\varepsilon$ converges to the same value $\hat{\vartheta}$. Then using the notations of the section 3.1 we can write

$$\tilde{\vartheta}_\varepsilon = \frac{\int_\alpha^\beta \vartheta p\left(\vartheta\right) \frac{V\left(\vartheta, X^T\right)}{V\left(\hat{\vartheta}, X^T\right)} \mathrm{d}\vartheta}{\int_\alpha^\beta p\left(\vartheta\right) \frac{V\left(\vartheta, X^T\right)}{V\left(\hat{\vartheta}, X^T\right)} \mathrm{d}\vartheta} = \hat{\vartheta} + \varepsilon^{2/3} \frac{\int_{\mathrm{U}_\varepsilon} u\, p\left(\vartheta_u\right) Z_\varepsilon\left(u\right) \mathrm{d}u}{\int_{\mathrm{U}_\varepsilon} p\left(\vartheta_u\right) Z_\varepsilon\left(u\right) \mathrm{d}u},$$

where we changed the variables $\vartheta = \vartheta_u = \hat{\vartheta} + \varepsilon^{2/3} u$. Hence

$$\frac{\tilde{\vartheta}_\varepsilon - \hat{\vartheta}}{\varepsilon^{2/3}} \approx \frac{\int_{\mathrm{U}_\varepsilon} u\, Z_\varepsilon\left(u\right) \mathrm{d}u}{\int_{\mathrm{U}_\varepsilon} Z_\varepsilon\left(u\right) \mathrm{d}u} = \frac{\int_{\mathrm{U}_\varepsilon} u \left(\hat{Z}_\varepsilon\left(u\right)\right)^{2\varepsilon^{-2/3}} \mathrm{d}u}{\int_{\mathrm{U}_\varepsilon} \left(\hat{Z}_\varepsilon\left(u\right)\right)^{2\varepsilon^{-2/3}} \mathrm{d}u}$$

and the problem reduces to the study of the asymptotics of these two integrals in the situation, where

$$\hat{Z}_\varepsilon\left(u\right) = \exp\left\{\delta(\hat{\vartheta}) W\left(u\right) - \frac{\psi(\hat{\vartheta})}{2} u^2\right\} \left(1 + o\left(1\right)\right).$$

We can suppose that the detailed study will provide us the asymptotics

$$\frac{\tilde{\vartheta}_\varepsilon - \hat{\vartheta}}{\varepsilon^{2/3}} \approx \hat{u},$$

where $\hat{u}$ is as before the point of the maximum of the process $\delta(\hat{\vartheta}) W\left(u\right) - \frac{\psi(\hat{\vartheta})}{2} u^2$. This means that as usual in regular estimation problems the asymptotic behavior of the BE is similar to that of the MLE.

Another problem we obtain if we suppose that the observed process has a signal $M\left(\vartheta, \cdot\right)$ with a singularity of the *cusp-type* (theoretical model) but the observed process in reality has a *smooth* signal $S\left(\vartheta, \cdot\right)$, i.e.; *cusp versus smooth*. Say,

$$\mathrm{d}X_t = a\left|t - \vartheta\right|^\kappa \mathrm{d}t + \varepsilon \mathrm{d}W_t, \quad X_0 = 0, \quad 0 \le t \le T, \tag{3.33}$$

where $\kappa \in (0, \frac{1}{2})$.

Therefore the observed process is (3.9) but the statistician calculate the pseudo-likelihood ratio $V\left(\vartheta, X^T\right)$ and the pseudo-MLE $\hat{\vartheta}_\varepsilon$ following (3.2) and (3.3) respectively. It is clear that $\hat{\vartheta}_\varepsilon$ converges to the value

$$\hat{\vartheta} = \arg \inf_{\vartheta \in \Theta} \int_0^T \left[a\left|t - \vartheta\right|^\kappa - S\left(\vartheta_0, t\right)\right]^2 \mathrm{d}t,$$

which minimizes the Kullback-Leibner distance and we describe the limit distribution of $\varepsilon^{-\frac{2}{3-2\kappa}} \left(\hat{\vartheta}_\varepsilon - \hat{\vartheta}\right)$. For the details see [2].

The properties of the MLE and Bayesian estimators for the ergodic diffusion processes and inhomogeneous Poisson processes with cusp-type singularities are studied for example, in [4], [3]. For the general theory of the parameter estimation for different singular estimation problems see [7].

There is another class of problems related to the hypothesis testing in the situations of the misspecification in regularity conditions. Let us consider the hypothsis testing problem

$$\begin{aligned} &\mathcal{H}_0 \qquad \vartheta = \vartheta_0, \\ &\mathcal{H}_1 \qquad \vartheta > \vartheta_0, \end{aligned}$$

by the observations of the Example 1. This means that we suppose that the observations are (3.6) but the real signal is (3.5). The Wald's test based on the MLE $\hat\vartheta_\varepsilon$ is

$$\Psi_\varepsilon\left(X^T\right) = 1_{\left\{\varepsilon^{-2}\left(\hat\vartheta_\varepsilon - \vartheta_0\right) > c_\nu\right\}}.$$

The threshold c_ν is obtained as solution of the equation (no misspecification, observations are (3.6))

$$\lim_{\varepsilon\to 0} \mathbf{E}_{\vartheta_0} \Psi_\varepsilon\left(X^T\right) = \mathbf{P}\left(\hat\eta > c_\nu\right) = \nu.$$

Here $\nu \in (0,1)$ and the random variable $\hat\eta$ is the same as in (2.5) with $\delta\left(\vartheta_0\right) = 2$. If we use this test when the observed real signal is (3.5), then even under $\mathcal{H}_0$

$$\varepsilon^{-2}\left(\hat\vartheta_\varepsilon - \vartheta_0\right) = \varepsilon^{-4/3}\varepsilon^{-2/3}\left(\hat\vartheta_\varepsilon - \vartheta_0\right) \to \infty$$

and the hypothesis $\mathcal{H}_0$ will be always rejected.

Acknowledgment

This work was done under partial financial support of the grant of RSF number 15-11-10022. The authors are grateful to Pavel Chigansky for careful reading of the manuscript and the comments.

References

[1] Campillo, F., Le Gland, F., Kutoyants, Yu. A. (2000) Small noise asymptotics of the GLR test for off-line change detection in misspecified diffusion processes. *Stoch., Stoch. Reports.* 70, 109–129. MR1785066

[2] Chernoyarov, O. V., Dachian, S. Yu., Kutoyants, Yu. A. (2018) On parameter estimation for cusp-type signals. *Ann. Inst. Stat. Math.*, 70, 1, 39-62. MR3742817

[3] Dachian, S. Yu., (2003) Estimation of cusp location by Poisson observations. *Statist. Inference Stoch. Processes,* 6, 1, 1–14. MR1965181

[4] Dachian, S. Yu., Kutoyants, Yu. A. (2003) On cusp estimation of ergodic diffusion process. *J. Stat. Plann. Infer.*, 117, 153–166. MR2001147

[5] Ibragimov I. A., Has'minskii R. Z. (1974) An estimator of the parameter of a signal in Gaussian white noise, *Probl. Inform. Transm.*, 10, 31–46. MR0403126

[6] Ibragimov I. A., Has'minskii, R. Z. (1975) Parameter estimation for a discontinuous signal in white Gaussian noise. *Probl. Inform. ransm.*, 11, 203–212.

[7] Ibragimov, I. A. and Has'minskii, R. Z. (1981) *Statistical Estimation - Asymptotic Theory.* Springer-Verlag, New York. MR0620321

[8] Kordzakhia, N., Kutoyants, Y. A., Novikov, A. and Hin, L.-Y. (2017) On a representation of fractional Brownian motion and the limit distributions of statistics arising in cusp statistical models. Submitted. MR3403018

[9] Kutoyants, Yu. A. (1994) *Identification of Dynamical Systems with Small Noise.* Kluwer Academic Publisher, Dordrecht. MR1332492

[10] Kutoyants, Yu. A. (2017) The Asymptotics of Misspecified MLEs for Some Stochastic Processes: a Survey. *Stat. Inference Stoch. Process.*, 20, 3, 347–368. MR3694869

[11] Trifonov, A. P. (1984) *Detection of the Signals with Unknown Parameters. Theory of Signals Detection.* Radio i Svyaz, Moscow (in Russian). MR0759716

[12] Trifonov, A. P., Shinakov, Yu. S. (1986) *Simultaneous Detection of Signals and Estimation of their Parameters in the Presence of Noises.* Radio i Svyaz, Moscow (in Russian).