The Asymptotics of Misspecified MLEs for Some Stochastic Processes: a Survey.

Yury A. Kutoyants

Le Mans University, Le Mans, France,

National Research University "MPEI", Moscow

Tomsk State University, Tomsk

SAPS-XI, Petergof, 17 July, 2017

Signal in WGN - Problem

Observed process is

$$X\left(t\right)=S\left(\vartheta ,t\right)+\varepsilon n\left(t\right), \qquad 0\leq t\leq T,$$

where $S\left(\vartheta ,t\right)$ is the signal and $n\left(t\right)$ is the gaussian noise. We have to estimate ϑ. The statistician supposes that the observed (theoretical) signal is $Q\left(\vartheta ,t\right)$ and the model is

$$X\left(t\right)=Q\left(\vartheta ,t\right)+\varepsilon m\left(t\right), \qquad 0\leq t\leq T.$$

where $m\left(t\right)$ is gaussian process. Therefore we have the problem of *misspecification*. We are mainly interested by the estimation of ϑ in the cases where the regularity conditions (smoothness) of the signals $S\left(\vartheta ,t\right)$ and $Q\left(\vartheta ,t\right)$ are different. The asymptotic corresponds to $\varepsilon\rightarrow 0$, i.e., we have *small noise* asymptotics.

Poisson Processes - Problem

We observe a periodic Poisson process $X^n = (X_t, 0 \leq t \leq n\tau)$ with known period $\tau > 0$ and intensity function

$$\lambda_* (\vartheta_0, t), \quad 0 \leq t \leq \tau, \quad \vartheta_0 \in \Theta.$$

Here ϑ_0 is the true value. The statistician supposes that the intensity function belongs to another parametric family

$$\lambda (\vartheta, t), \quad 0 \leq t \leq \tau, \quad \vartheta \in \Theta.$$

Here once more we are interested in the situations where the regularity conditions of these two families are different (*misspecifications in regularity*). The asymptotic corresponds to $n \to \infty$, i.e., we have *large samples* asymptotics.

Deterministic signal in WGN

Suppose that we observe a deterministic signal in WGN

$$X(t) = S(\vartheta_0, t) + \varepsilon\, n(t), \qquad 0 \le t \le T.$$

Here $S(\vartheta, t)$ is a "signal" and $n(t)$ is WGN and $0 < \varepsilon \ll 1$ is "small" parameter. Recall that the WGN $n(t)$ has properties

$$\mathbf{E} n(t) = 0, \qquad \mathbf{E} n(t)\, n(s) = \delta(t - s),$$

here $\delta(\cdot)$ is a *delta-function*.

The basic model can be rewritten as

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

where $W_t, 0 \le t \le T$ is a Wiener process and the WGN is defined as derivative $n\left(t\right) = \frac{\mathrm{d}W_t}{\mathrm{d}t}$. Of course, we put $X\left(t\right) = \frac{\mathrm{d}X_t}{\mathrm{d}t}$.

The likelihood ratio function is

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

The set $\Theta = (\alpha, \beta)$.

The maximum likelihood estimator (MLE) $\hat{\vartheta}_\varepsilon$ is defined by the MLEq

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

If this equation has many solutions, then we can take anyone as the MLE.

If the parameter ϑ is a r. v. with a density function $p\left(\vartheta\right), \vartheta \in \Theta$, then we can define the bayesian estimator (BE) $\tilde{\vartheta}_\varepsilon$ as follows

$$\tilde{\vartheta}_\varepsilon = \frac{\int_\alpha^\beta \theta p\left(\theta\right) V\left(\theta, X^T\right) \mathrm{d}\theta}{\int_\alpha^\beta p\left(\theta\right) V\left(\theta, X^T\right) \mathrm{d}\theta}.$$

Sometimes the BE can be used even in the situations where the parameter ϑ is not random. In that case we take as $p\left(\cdot\right)$ some continuous positive function and consider $\tilde{\vartheta}_\varepsilon$ as a method of construction of estimator. Recall that in the singular estimation problems the BE are asymptotically efficient and the MLE - not.

Introduce as well the trajectory fitting estimator (TFE) $\check{\vartheta}_\varepsilon$ (which can be called the minimum distance estimator (MDE)) defined by the relation

$$\int_0^T \left[X_t - \int_0^t S\left(\check{\vartheta}_\varepsilon, s\right) \mathrm{d}s\right]^2 \mathrm{d}t = \inf_{\theta\in\Theta} \int_0^T \left[X_t - \int_0^t S\left(\theta, s\right) \mathrm{d}s\right]^2 \mathrm{d}t.$$

Therefore we have three estimators: MLE $\hat{\vartheta}_\varepsilon$, BE $\tilde{\vartheta}_\varepsilon$ and TFE $\check{\vartheta}_\varepsilon$ and we are interested by the convergences

$$\frac{\hat{\vartheta}_\varepsilon - \vartheta_0}{\varphi_\varepsilon} \Longrightarrow \hat{u} \sim ?, \qquad \frac{\tilde{\vartheta}_\varepsilon - \vartheta_0}{\varphi_\varepsilon} \Longrightarrow \tilde{u} \sim ?, \qquad \frac{\check{\vartheta}_\varepsilon - \vartheta_0}{\psi_\varepsilon} \Longrightarrow \check{u} \sim ?$$

The scketch of the proof (approach developed by Ibragimov and Khasminskii): introduce the normalized likelihood ratio

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

The normalizing function $\varphi_\varepsilon\to 0$ is such that we have the convergence

$$Z_\varepsilon\left(\cdot\right)\Longrightarrow Z\left(\cdot\right),$$

where $Z\left(u\right),u\in R$ is some limit process.

Let us introduce the r.v.'s $\hat u$ and $\tilde u$ by the relations

$$Z\left(\hat u\right)=\sup_u Z\left(u\right),\qquad \tilde u=\frac{\int uZ\left(u\right)\mathrm{d}u}{\int Z\left(u\right)\mathrm{d}u}.$$

Now the limit distribution of the MLE is obtained as follows

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

Here we put $\vartheta=\vartheta_0+\varphi_\varepsilon u$.

For the bayesian estimator we have (once more we change the variables $\vartheta_u = \vartheta_0 + \varphi_\varepsilon u$):

$$\tilde{\vartheta}_\varepsilon = \frac{\int \theta p\left(\theta\right) V\left(\theta, X^T\right) \mathrm{d}\theta}{\int p\left(\theta\right) V\left(\theta, X^T\right) \mathrm{d}\theta} = \vartheta_0 + \varphi_\varepsilon \frac{\int_{\mathbb{U}_\varepsilon} up\left(\theta_u\right) V\left(\theta_u, X^T\right) \mathrm{d}u}{\int_{\mathbb{U}_\varepsilon} p\left(\theta_u\right) V\left(\theta_u, X^T\right) \mathrm{d}u}$$

$$= \vartheta_0 + \varphi_\varepsilon \frac{\int_{\mathbb{U}_\varepsilon} up\left(\theta_u\right) Z_\varepsilon\left(u\right) \mathrm{d}u}{\int_{\mathbb{U}_\varepsilon} p\left(\theta_u\right) Z_\varepsilon\left(u\right) \mathrm{d}u}.$$

Hence

$$\frac{\tilde{\vartheta}_\varepsilon - \vartheta_0}{\varphi_\varepsilon} = \frac{\int_{\mathbb{U}_\varepsilon} up\left(\theta_u\right) Z_\varepsilon\left(u\right) \mathrm{d}u}{\int_{\mathbb{U}_\varepsilon} p\left(\theta_u\right) Z_\varepsilon\left(u\right) \mathrm{d}u} \Longrightarrow \frac{\int_R uZ\left(u\right) \mathrm{d}u}{\int_R Z\left(u\right) \mathrm{d}u} = \tilde{u}.$$

Therefore

$$\frac{\hat{\vartheta}_\varepsilon - \vartheta_0}{\varphi_\varepsilon} \Longrightarrow \hat{u}, \qquad \frac{\tilde{\vartheta}_\varepsilon - \vartheta_0}{\varphi_\varepsilon} \Longrightarrow \tilde{u}.$$

No misspecification

Suppose that the signal $S(\vartheta, \cdot)$ is a **smooth** function of ϑ:

$$\mathrm{d}X_t = S(\vartheta_0, t)\,\mathrm{d}t + \varepsilon \mathrm{d}W_t, \qquad X_0 = 0, \qquad 0 \le t \le T$$

Introduce the normalized likelihood ratio

$$Z_\varepsilon(u) = \frac{V\left(\vartheta_0 + \frac{\varepsilon u}{\mathbb{I}(\vartheta_0)^{1/2}}, X^T\right)}{V(\vartheta_0, X^T)}, \quad u \in \mathbb{U}_\varepsilon = \left(\frac{\alpha - \vartheta_0}{\mathbb{I}(\vartheta_0)^{-1/2}\,\varepsilon}, \frac{\beta - \vartheta_0}{\mathbb{I}(\vartheta_0)^{-1/2}\,\varepsilon}\right)$$

We have the convergence (LAN)

$$Z_\varepsilon(u) \Longrightarrow Z(u) = \exp\left\{u\xi - \frac{u^2}{2}\right\}, \qquad u \in R.$$

Here $\xi \sim \mathcal{N}(0,1)$ and

$$\mathbb{I}(\vartheta_0) = \int_0^T \dot{S}(\vartheta_0, t)^2 \,\mathrm{d}t,$$

is the Fisher information and dot means derivation w.r.t. ϑ.

We have for the MLE $\hat{\vartheta}_\varepsilon$ and BE $\tilde{\vartheta}_\varepsilon$

$$\mathbb{I}(\vartheta_0)^{\frac{1}{2}} \varepsilon^{-1}\left(\hat{\vartheta}_\varepsilon - \vartheta_0\right) \Longrightarrow \xi, \qquad \mathbb{I}(\vartheta_0)^{\frac{1}{2}} \varepsilon^{-1}\left(\tilde{\vartheta}_\varepsilon - \vartheta_0\right) \Longrightarrow \xi,$$

$$\mathbf{E}_{\vartheta_0}\left|\frac{\hat{\vartheta}_\varepsilon - \vartheta_0}{\mathbb{I}(\vartheta_0)^{\frac{1}{2}}\varepsilon}\right|^p \longrightarrow \mathbf{E}\left|\xi\right|^p, \quad \mathbf{E}_{\vartheta_0}\left|\frac{\tilde{\vartheta}_\varepsilon - \vartheta_0}{\mathbb{I}(\vartheta_0)^{\frac{1}{2}}\varepsilon}\right|^p \longrightarrow \mathbf{E}\left|\xi\right|^p$$

and the relations

$$Z(\xi) = \sup_u Z(u), \qquad \xi = \frac{\int uZ(u)\,\mathrm{d}u}{\int Z(u)\,\mathrm{d}u}.$$

Therefore the MLE $\hat{\vartheta}_\varepsilon$ and BE $\tilde{\vartheta}_\varepsilon$ are consistent, asymptotically normal,we have the convergence of all polynomial moments and the both estimators are asymptotically efficient (Ibragimov-Khasminskii 1975).

For the MDE $\check{\vartheta}_\varepsilon$ we have the similar asymptotic normality

$$\varepsilon^{-1}\left(\check{\vartheta}_\varepsilon-\vartheta_0\right)\Longrightarrow\mathcal{N}\left(0,\mathbb{D}\left(\vartheta_0\right)\right),$$

here

$$\mathbb{D}\left(\vartheta\right)=\frac{\int_0^T\left(\int_s^T\int_0^t\dot{S}\left(\vartheta,v\right)\mathrm{d}v\mathrm{d}t\right)^2\mathrm{d}s}{\left(\int_0^T\left(\int_0^t\dot{S}\left(\vartheta,s\right)\right)^2\mathrm{d}s\right)^2}\geq\mathbb{I}\left(\vartheta\right)^{-1}.$$

The asymptotic efficiency is defined with the help of the following Hajek-Le Cam's (1972) lower bound. For all $\vartheta_0 \in \Theta$ we have

$$\lim_{\delta\to 0}\lim_{\varepsilon\to 0}\sup_{|\vartheta-\vartheta_0|<\delta}\varepsilon^{-2}\mathbf{E}_\vartheta\left|\bar\vartheta_\varepsilon-\vartheta\right|^2\geq \mathbb{I}\left(\vartheta_0\right)^{-1}.$$

Therefore we call an estimator ϑ^*_ε asymptotically efficient if for all $\vartheta_0 \in \Theta$

$$\lim_{\delta\to 0}\lim_{\varepsilon\to 0}\sup_{|\vartheta-\vartheta_0|<\delta}\varepsilon^{-2}\mathbf{E}_\vartheta\left|\vartheta^*_\varepsilon-\vartheta\right|^2= \mathbb{I}\left(\vartheta_0\right)^{-1}.$$

These results were generalized to a wide class of (colored) Gaussian noises $n\left(\cdot\right)$ in the work K. 1980. Using the theory of Reproducing Kernel Hilbert Space (RKHS) it was shown that if the regularity conditions are expressed in the terms of RKHS-norms, then the MLE and BE have the similar asymptotic properties.

If the signal is a **discontinuous** function

$$S(\vartheta,t) = h(t)\,1\!\!1_{\{t<\vartheta\}} + g(t)\,1\!\!1_{\{t\geq\vartheta\}}$$

where $h(t)$ and $g(t)$ are two different functions and the unknown parameter $\vartheta \in \Theta \subset (0,T)$. Introduce the normalized likelihood ratio

$$Z_\varepsilon(u) = \frac{V\left(\vartheta_0 + \frac{\varepsilon^2 u}{r(\vartheta_0)^2}, X^T\right)}{V(\vartheta_0, X^T)}, \quad u \in \mathbb{U}_\varepsilon = \left(\frac{\alpha-\vartheta_0}{r(\vartheta_0)^{-2}\,\varepsilon^2}, \frac{\beta-\vartheta_0}{r(\vartheta_0)^{-2}\,\varepsilon^2}\right).$$

We have the convergence

$$Z_\varepsilon(u) \Longrightarrow Z(u) = \exp\left\{W(u) - \frac{|u|}{2}\right\}, \qquad u \in R,$$

where $r(\vartheta) = g(\vartheta) - h(\vartheta)$ and $W(\cdot)$ is two-sided Wiener process.

This convergence allows us to prove that the MLE and BE have the rate of convergece ε^2 with different limit distributions:

$$r\left(\vartheta_0\right)^2 \varepsilon^{-2}\left(\hat{\vartheta}_\varepsilon-\vartheta_0\right) \Longrightarrow \hat{\zeta}, \qquad\qquad r\left(\vartheta_0\right)^2 \varepsilon^{-2}\left(\tilde{\vartheta}_\varepsilon-\vartheta_0\right) \Longrightarrow \tilde{\zeta},$$

where $\hat{\zeta}$ and $\tilde{\zeta}$ are the random variables defined as follows.

$$Z(\hat{\zeta}) = \sup_{u\in R} Z\left(u\right), \qquad\qquad \tilde{\zeta} = \frac{\int_R u\, Z\left(u\right)\,{\rm d}u}{\int_R Z\left(u\right)\,{\rm d}u}.$$

For the proof see Ibragimov, Khasminskii 1975.

We have $\mathbf{E}_{\vartheta_0}\hat{\zeta}^2 = 26$ (Terent'ev 1968) and $\mathbf{E}_{\vartheta_0}\tilde{\zeta}^2 \approx 19,3$ (Ibragimov, Khasminskii 1975, Golubev 1979, Rubin, Song 1995, Novikov, Kordzakhia 2013).

The asymptotic efficiency is defined with the help of the following lower bound : for all estimators $\bar\vartheta_\varepsilon$ and all $\vartheta_0 \in \Theta$ we have

$$\lim_{\delta\to 0}\lim_{\varepsilon\to 0}\sup_{|\vartheta-\vartheta_0|<\delta} r\left(\vartheta_0\right)^4 \varepsilon^{-4}\mathbf{E}_\vartheta \left|\bar\vartheta_\varepsilon-\vartheta\right|^2 \geq \mathbf{E}_{\vartheta_0}\tilde\zeta^2.$$

We call an estimator ϑ^*_ε asymptotically efficient if for all $\vartheta_0 \in \Theta$

$$\lim_{\delta\to 0}\lim_{\varepsilon\to 0}\sup_{|\vartheta-\vartheta_0|<\delta} r\left(\vartheta_0\right)^4 \varepsilon^{-4}\mathbf{E}_\vartheta \left|\vartheta^*_\varepsilon-\vartheta\right|^2 = \mathbf{E}_{\vartheta_0}\tilde\zeta^2.$$

For the proof see Ibragimov, Khasminskii 1975. The BE $\tilde\vartheta_\varepsilon$ is asymptotically efficient.

Let us mention as well the parameter estimation problem with **cusp-type** singularity. Suppose that the observed process is

$$\mathrm{d}X_t = \left[a\left|t-\vartheta\right|^{\kappa} + h\left(\vartheta,t\right)\right]\mathrm{d}t + \varepsilon \mathrm{d}W_t, \qquad X_0=0, \quad 0\leq t\leq T,$$

where $\kappa\in(0,\frac{1}{2})$. The function $h\left(\vartheta,t\right)$ is continuously differentiable w.r.t. ϑ. The unknown parameter $\vartheta\in\Theta=(\alpha,\beta)$, $0<\alpha<\beta<T$.

Introduce the Hurst parameter $H=\kappa+\frac{1}{2}$ and the normalized likelihood ratio

$$Z_\varepsilon\left(u\right)=\frac{V\left(\vartheta_0+\frac{\varepsilon^{\frac{1}{H}}u}{\Gamma^{\frac{1}{H}}},X^T\right)}{V\left(\vartheta_0,X^T\right)}, \quad u\in\mathbb{U}_\varepsilon=\left(\frac{\alpha-\vartheta_0}{\Gamma^{-\frac{1}{H}}\varepsilon^{\frac{1}{H}}},\frac{\beta-\vartheta_0}{\Gamma^{-\frac{1}{H}}\varepsilon^{\frac{1}{H}}}\right).$$

We have the convergence

$$Z_\varepsilon(u) \Longrightarrow Z(u) = \exp\left\{W^H(u) - \frac{|u|^{2H}}{2}\right\}, \qquad u \in R,$$

Here Γ is some constant and $W^H(\cdot)$ is double-side fractional Brownien motion (fBm).Therefore the MLE $\hat\vartheta_\varepsilon$ and BE $\tilde\vartheta_\varepsilon$ are consistent, have different limit distributions

$$\Gamma^{\frac{1}{H}}\varepsilon^{-\frac{1}{H}}\left(\hat\vartheta_\varepsilon - \vartheta_0\right) \Longrightarrow \hat\zeta_H, \qquad \Gamma^{\frac{1}{H}}\varepsilon^{-\frac{1}{H}}\left(\tilde\vartheta_\varepsilon - \vartheta_0\right) \Longrightarrow \tilde\zeta_H,$$

$$\mathbf{E}_{\vartheta_0}\left|\frac{\hat\vartheta_\varepsilon - \vartheta_0}{\Gamma^{-\frac{1}{H}}\varepsilon^{\frac{1}{H}}}\right| \longrightarrow \mathbf{E}\left|\hat\zeta_H\right|^p, \qquad \mathbf{E}_{\vartheta_0}\left|\frac{\tilde\vartheta_\varepsilon - \vartheta_0}{\Gamma^{-\frac{1}{H}}\varepsilon^{\frac{1}{H}}}\right| \longrightarrow \mathbf{E}\left|\hat\zeta_H\right|^p$$

(Chernoyarov, Dachian, K. 2015). Here

$$Z(\hat\zeta_H) = \sup_{u\in R} Z(u), \qquad \tilde\zeta_H = \frac{\int_R u\,Z(u)\,\mathrm{d}u}{\int_R Z(u)\,\mathrm{d}u}.$$

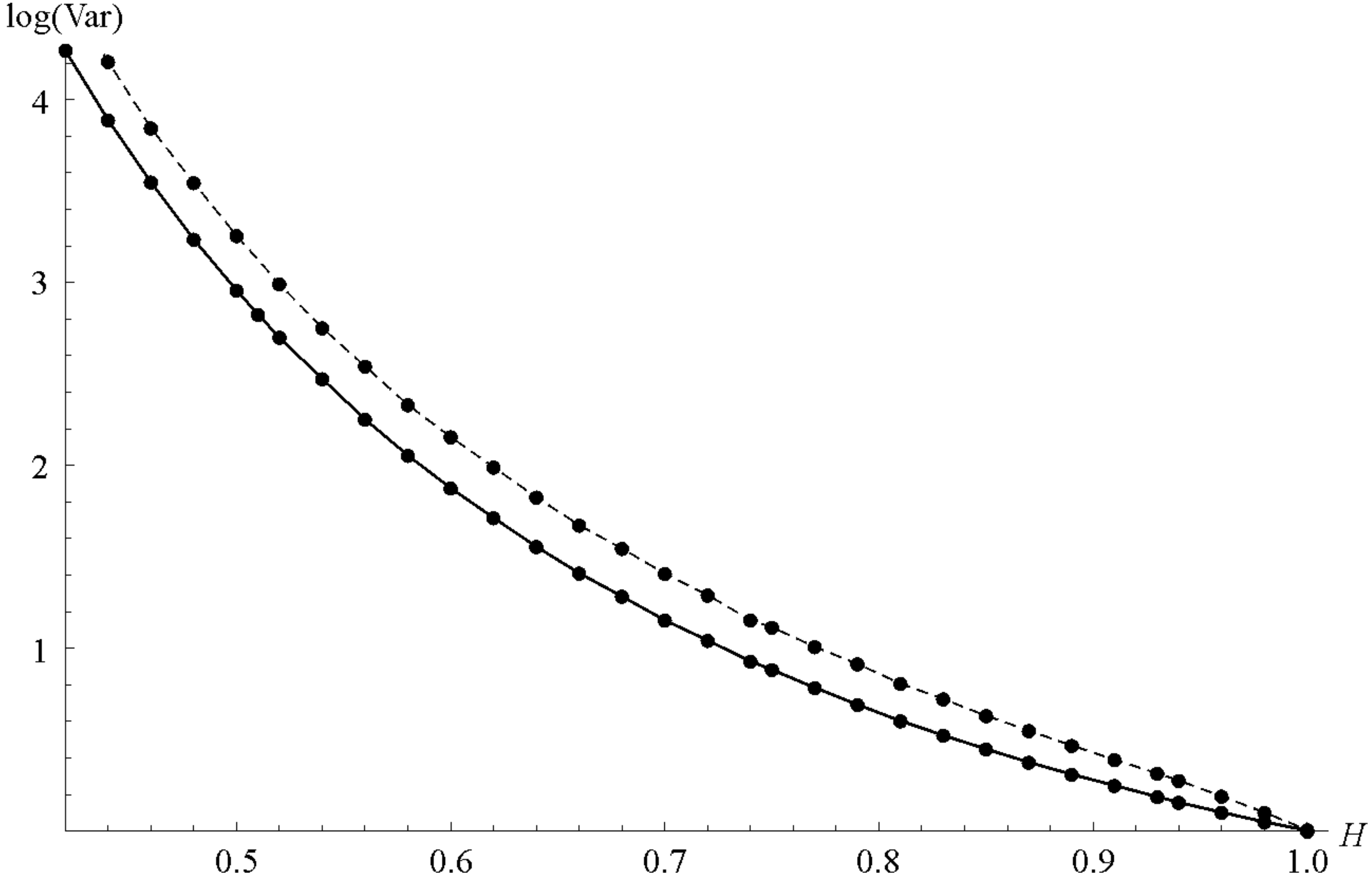
The asymptotic efficiency is defined with the help of the following lower bound : for all estimators $\bar\vartheta_\varepsilon$ and all $\vartheta_0 \in \Theta$ we have

$$\lim_{\delta\to 0}\lim_{\varepsilon\to 0}\sup_{|\vartheta-\vartheta_0|<\delta}\Gamma^{\frac{2}{H}}\varepsilon^{-\frac{2}{H}}\mathbf{E}_\vartheta\left|\bar\vartheta_\varepsilon-\vartheta\right|^2 \geq \mathbf{E}_{\vartheta_0}\tilde\zeta_H^2.$$

We call an estimator ϑ_ε^* asymptotically efficient if for all $\vartheta_0 \in \Theta$

$$\lim_{\delta\to 0}\lim_{\varepsilon\to 0}\sup_{|\vartheta-\vartheta_0|<\delta}\Gamma^{\frac{2}{H}}\varepsilon^{-\frac{2}{H}}\mathbf{E}_\vartheta\left|\vartheta_\varepsilon^*-\vartheta\right|^2 = \mathbf{E}_{\vartheta_0}\tilde\zeta_H^2.$$

The asymptotically efficient are the BE only.



Values of $\ln \mathbf{E}_\vartheta \tilde{\zeta}_H^2$ (solid line) and $\ln \mathbf{E}_\vartheta \hat{\zeta}_H^2$ (dashed line) for $H \in \left[\frac{2}{5}, 1\right]$ (Novikov, Kordzakhia, Ling, 2014).

Three types of regularity

- Smooth

$$\frac{\hat{\vartheta}_\varepsilon - \vartheta_0}{\varepsilon^\nu} \Longrightarrow \hat{u}_1, \qquad \nu = 1,$$

- Cusp

$$\frac{\hat{\vartheta}_\varepsilon - \vartheta_0}{\varepsilon^\nu} \Longrightarrow \hat{u}_2, \qquad 1 < \nu < 2,$$

- Change-point

$$\frac{\hat{\vartheta}_\varepsilon - \vartheta_0}{\varepsilon^\nu} \Longrightarrow \hat{u}_3, \qquad \nu = 2.$$

Poisson processe, no misspecifications

We observe Poisson process $X^n = (X_t, 0 \leq t \leq n\tau)$ of intensity function $\lambda(\vartheta, t) = \lambda(t - \vartheta)$. The function $\lambda(t)$ can be smooth, cusp-type and discontinuous, i.e., we have three types of regularity:

- Smooth (K. 1978)

$$n^{\frac{\nu}{2}} \left(\hat{\vartheta}_n - \vartheta_0\right) \Longrightarrow \hat{v}_1, \qquad \nu = 1,$$

- Cusp $\lambda(\vartheta, t) = a\left|t - \vartheta\right|^{\kappa} + h(t)$ (Dachian 2003)

$$n^{\frac{\nu}{2}} \left(\hat{\vartheta}_n - \vartheta_0\right) \Longrightarrow \hat{v}_2, \qquad 1 < \nu < 2,$$

- Change-point $\lambda(\vartheta, t) = h(t)\, 1\!\!\mathrm{I}_{\{t<\vartheta\}} + g(t)\, 1\!\!\mathrm{I}_{\{t \geq \vartheta\}}$ (K. 1980)

$$n^{\frac{\nu}{2}} \left(\hat{\vartheta}_n - \vartheta_0\right) \Longrightarrow \hat{v}_3, \qquad \nu = 2.$$

Misspecifications

Suppose that the observed process (*real model*) is

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

where ϑ_0 is the true value of unknown parameter. The statistician uses the *theoretical model*

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

with $Q\left(\vartheta, \cdot\right) \in L_2\left(0, T\right)$. The likelihood ratio (misspecified) is

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

The (pseudo) MLE $\hat{\vartheta}_\varepsilon$ is defined by the equation

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

To understand what is the limit of the MLE we write the likelihood ratio as follows

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

where we denoted as $\left\|\cdot\right\|$ the $L_2\left(0, T\right)$ norm.

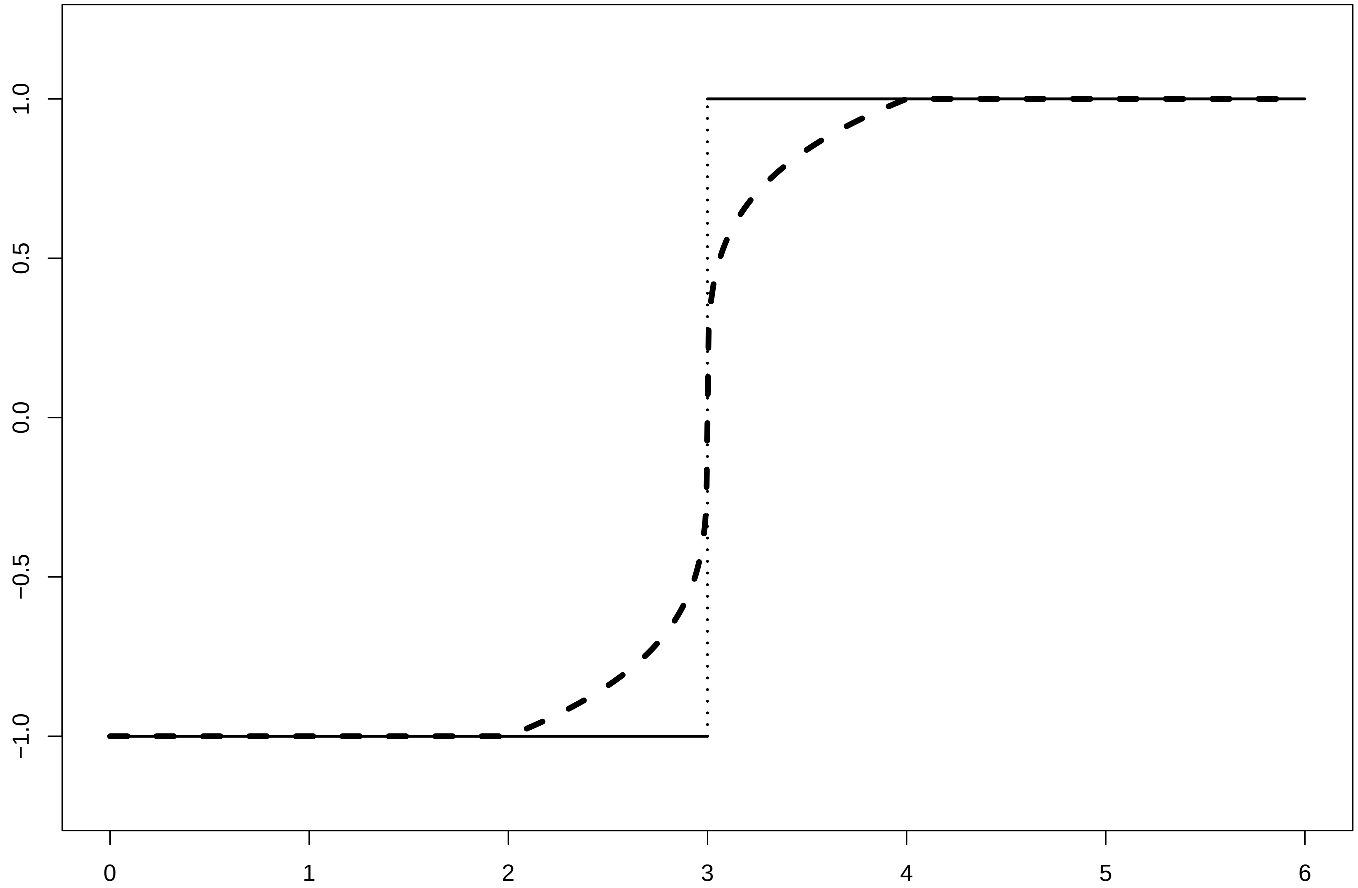


Figure 1: Theoretical (dashed line) and real (continuous line) signals

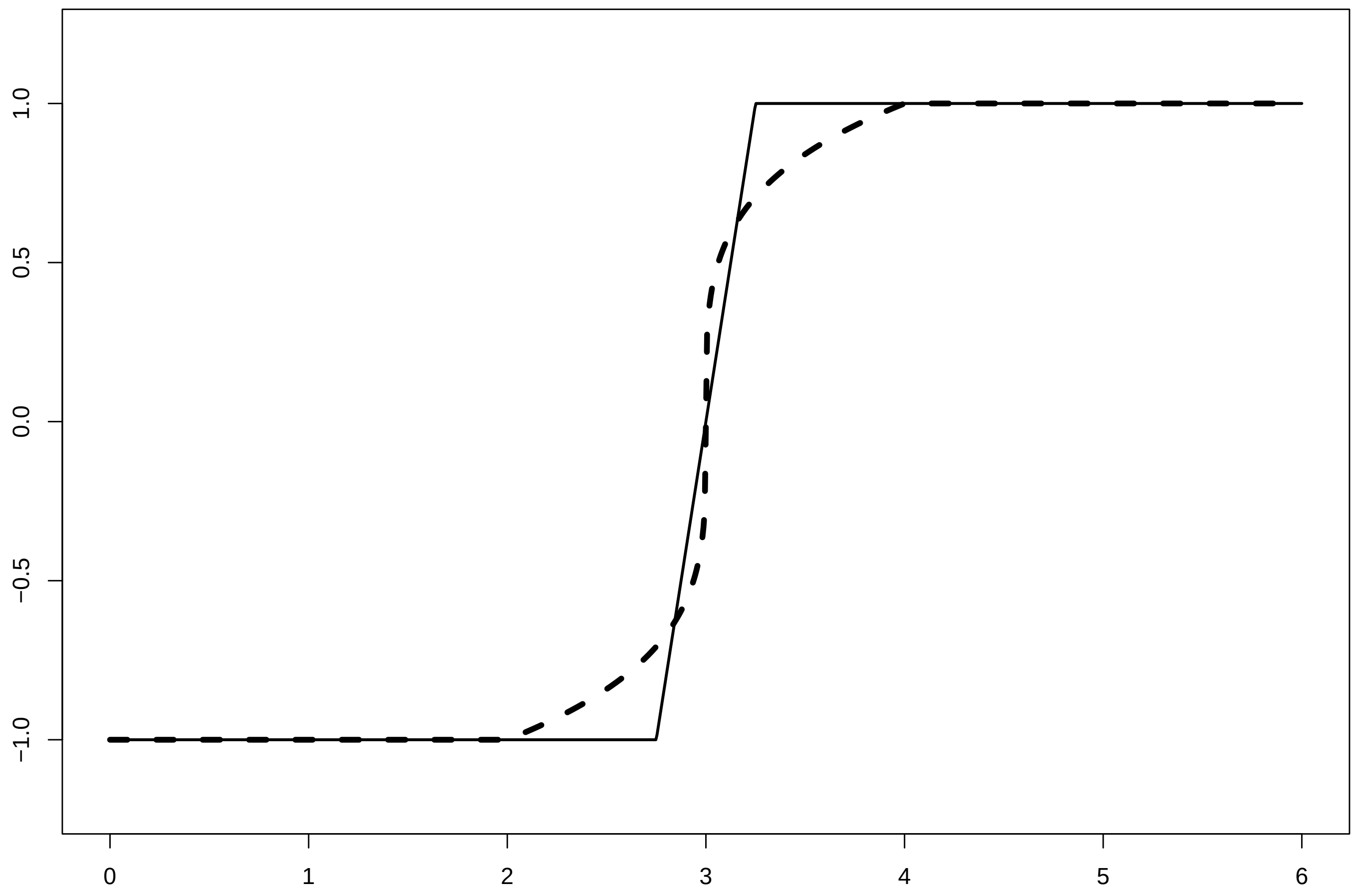


Figure 2: Theoretical (dashed line) and real (continuous line) signals

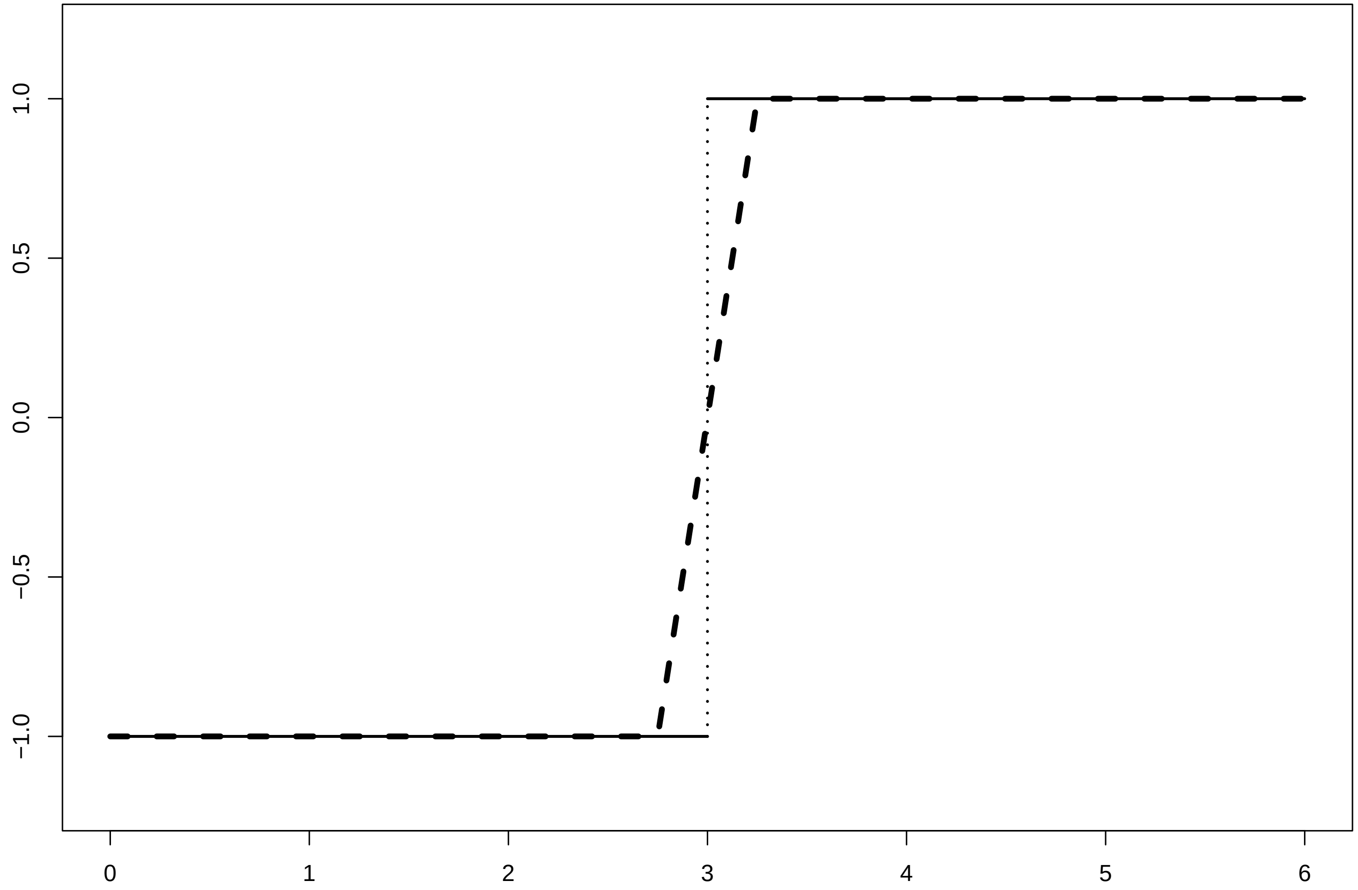


Figure 3: Theoretical (dashed line) and real (continuous line) signals

It can be easily verified that under mild regularity conditions we have the convergence in probability

$$\sup_{\vartheta\in\Theta}\left|\varepsilon^2\ln V\left(\vartheta,X^T\right)-\frac{1}{2}\left\|Q\left(\vartheta,\cdot\right)-S\left(\vartheta_0,\cdot\right)\right\|^2+\frac{1}{2}\left\|S\left(\vartheta_0,\cdot\right)\right\|^2\right|\longrightarrow 0.$$

Hence if we suppose that the equation

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

has a unique solution $\hat{\vartheta}$, then we obtain the well-known result that in the case of misspecification the MLE $\hat{\vartheta}_\varepsilon$ converges to the value $\hat{\vartheta}$, which minimizes the Kullback-Leibler distance.

Discontinuous vs smooth

Here we consider the situation where the true model of observations

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

has a signal $S\left(\vartheta_0, t\right)$ is a smooth w.r.t. ϑ function but the theoretical model chosen by statistician

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

contains the signal

$$Q\left(\vartheta, t\right) = h\left(t\right) \mathbb{I}_{\{t<\vartheta\}} + g\left(t\right) \mathbb{I}_{\{t\geq\vartheta\}},$$

which is a discontinous function of time with the jump at the point ϑ.

The unknown parameter $\vartheta \in \Theta = (\alpha, \beta)$ with $0 < \alpha < \beta < T$. We observe a trajectory $X^T = (X_t, 0 \leq t \leq T)$ and we want to estimate ϑ_0. Introduce the pseudo-likelihood ratio

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

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

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

Let us introduce the following notations:

$$\delta\left(t\right)=h\left(t\right)-g\left(t\right),\qquad \Phi\left(\vartheta\right)=\int_0^T\left[Q\left(\vartheta,t\right)-S\left(\vartheta_0,t\right)\right]^2\mathrm{d}t,$$

$$\gamma\left(\vartheta\right)=\frac{\ddot{\Phi}(\vartheta)}{2},\qquad \hat{Z}\left(u\right)=\exp\left\{\delta(\hat{\vartheta})W\left(u\right)-\frac{\gamma(\hat{\vartheta})}{2}u^2\right\},\quad u\in R,$$

$$\hat{u}=\arg\sup_{u\in R}\left[\delta(\hat{\vartheta})W\left(u\right)-\frac{\gamma(\hat{\vartheta})}{2}u^2\right],\qquad \hat{u}_0=\arg\sup_{v\in R}\left[w\left(v\right)-\frac{v^2}{2}\right].$$

Here dot means differentiating w.r.t. ϑ and $w\left(v\right),v\in R$ is double-sided Wiener process.

Conditions $\mathcal{M}$.

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

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

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*

$$\frac{\hat\vartheta_\varepsilon-\hat\vartheta}{\varepsilon^{2/3}}\Longrightarrow \hat u=\left(\frac{\delta(\hat\vartheta)}{\gamma(\hat\vartheta)}\right)^{2/3}\hat u_0,$$

and for any $p>0$

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

The proof in Chernoyarov, K., Trifonov 2015.

Let us change the variables $u = rv$, then we have the equality in distribution

$$\delta(\hat\vartheta)W(u) - \frac{\gamma(\hat\vartheta)}{2}u^2 = \delta(\hat\vartheta)\sqrt{r}w\left(v\right) - \frac{\gamma(\hat\vartheta)r^2}{2}v^2$$

$$= \delta(\hat\vartheta)\sqrt{r}\left(w\left(v\right) - \frac{\gamma(\hat\vartheta)r^{3/2}}{\delta(\hat\vartheta)}\frac{v^2}{2}\right) = \frac{\delta(\hat\vartheta)^{4/3}}{\sigma(\hat\vartheta)^{1/3}}\left(w\left(v\right) - \frac{v^2}{2}\right),$$

where we put $r = \left(\delta(\hat\vartheta)/\gamma(\hat\vartheta)\right)^{2/3}$. Hence $\hat u = \left(\frac{\delta(\hat\vartheta)}{\gamma(\hat\vartheta)}\right)^{2/3}\hat u_0$, where

$$Z_0\left(\hat u_0\right) = \sup_u Z_0\left(u\right), \qquad Z_0\left(u\right) = \exp\left\{w\left(u\right) - \frac{u^2}{2}\right\}$$

and

$$\frac{\hat\vartheta_\varepsilon - \hat\vartheta}{\varepsilon^{2/3}} \Longrightarrow \hat u = \left(\frac{\delta(\hat\vartheta)}{\gamma(\hat\vartheta)}\right)^{2/3}\hat u_0,$$

Note that as $\hat{\vartheta}$ is the point of minimum of the function $\Phi(\vartheta)$ we have the equality

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

which is equivalent to

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

Of course, this is a necessary condition only. If this equation has no solution, say,

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

then $\hat{\vartheta} = \alpha$.

In these two cases the behavior of the estimator $\hat{\vartheta}_\varepsilon$ can be studied as it was done in K. 1994, Section 2.8. If we have the equality

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

for 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 such situations 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 K. 1994, Section 2.7.

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 see the behavior of TFE $\check{\vartheta}_\varepsilon$ defined by the relation

$$\check{\vartheta}_\varepsilon = \arg\inf_{\vartheta\in\Theta} \int_0^T \left[X_t - \int_0^t Q\left(\vartheta, s\right)\,\mathrm{d}s\right]^2 \mathrm{d}t.$$

Note that the function

$$q\left(\vartheta, t\right) = \int_0^t Q\left(\vartheta, s\right)\,\mathrm{d}s$$

has continuous in $L_2\left[0, T\right]$ derivative w.r.t. ϑ.

Therefore if we suppose that the function

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

has a unique minimum at the point $\check\vartheta\in\Theta$, then it can be shown that TFE is asymptotically normal

$$\frac{\check\vartheta_\varepsilon-\check\vartheta}{\varepsilon}=\frac{\int_0^T W_t\,\dot q\left(\check\vartheta,t\right)\mathrm{d}t}{\int_0^T\dot q\left(\check\vartheta,t\right)^2\mathrm{d}t}\left(1+o\left(1\right)\right)\Longrightarrow\mathcal{N}\left(0,\mathbb{D}\left(\vartheta_0,\check\vartheta\right)\right).$$

Therefore this estimator is asymptotically normal with the good rate ε. The details of the proof can be found in the Section 7.4 in K. 1994.

Example 1. Suppose that the observed process is

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

where $\vartheta_0 \in \Theta = (\alpha, \beta)$, $0 < \alpha < \beta < T$ and the theoretical model is

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

The pseudo-likelihood ratio is the function

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

because $\mathrm{sgn}\,(t-\vartheta)^2 = 1$.

Note that

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

Hence the MLE $\hat{\vartheta}_\varepsilon$ defined by the relation

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

in this misspecified parameter estimation problem is consistent.

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

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

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

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

This is a particular case covered by the Theorem 1.

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* we can obtain different rates of convergence of estimators. Put $\vartheta = \vartheta_0 + \varepsilon^{\frac{2}{2\kappa+1}} u$. Then the corresponding calculations provides us

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

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

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

Therefore choosing different $\kappa > \frac{1}{2}$ we can obtain any rate $\varepsilon^{\gamma}, \gamma < 1$ of convergence of pseudo-MLE (Chernoyarov, Dachian, K. [1]).

Bayes estimators

The estimator is

$$\tilde{\vartheta}_\varepsilon = \frac{\int_\alpha^\beta \vartheta p\left(\vartheta\right) V\left(\vartheta, X^T\right) \mathrm{d}\vartheta}{\int_\alpha^\beta p\left(\vartheta\right) V\left(\vartheta, X^T\right) \mathrm{d}\vartheta},$$

where $p\left(\vartheta\right), \alpha < \vartheta < \beta$ is continuous positive density of the distribution of the random variable ϑ.

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

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

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

Hence

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

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

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

Open Problems.

1. Calculate

$$\mathbf{E}\,\hat{u}_0^2, \qquad \hat{u}_0 = \arg\sup_{v\in R}\left\{w\,(v) - \frac{v^2}{2}\right\}$$

2. To prove the asymptotics

$$\frac{\tilde{\vartheta}_\varepsilon - \hat{\vartheta}}{\varepsilon^{2/3}} \approx \hat{u} = \left(\frac{\delta(\hat{\vartheta})}{\gamma(\hat{\vartheta})}\right)^{2/3} \hat{u}_0,$$

This means that as usual in regular estimation problems the asymptotic behavior of the BE is asymptotically equivalent to that of the MLE.

Smooth vs discontinuous

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

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

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

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

with the "smooth" signal $Q(\vartheta, \cdot)$. The likelihood ratio $L\left(\vartheta, X^T\right)$ and the pseudo-MLE $\hat{\vartheta}_\varepsilon$ are defined by the same relations. As before, we are interested by the asymptotic behavior of $\hat{\vartheta}_\varepsilon$ as $\varepsilon \to 0$.

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

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

where

$$Q(\vartheta, t) = \frac{t - \vartheta}{\delta} 1\!\!\mathrm{I}_{|t-\vartheta| \leq \delta} + \operatorname{sgn}(t - \vartheta)$$

to estimate the parameter $\vartheta \in \Theta = (\alpha, \beta)$, where $0 < \alpha < \beta < T$.

It is easy to see that the function

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

atteints its minimum at the point $\hat{\vartheta} = \vartheta_0$. Therefore the pseudo-MLE

$$\hat{\vartheta}_\varepsilon \longrightarrow \vartheta_0.$$

It has the Gaussian distribution

$$\frac{\hat{\vartheta}_\varepsilon - \vartheta_0}{\varepsilon} \sim \mathcal{N}(0, D)$$

and the rate of convergence is ε.

General case. Introduce the conditions of regularity.

Conditions $\mathcal{R}$.

1. *The functions $h(\cdot)$ and $g(\cdot)$ are bounded and for all $\vartheta \in [\alpha, \beta]$ we have $h(\vartheta) \neq g(\vartheta)$.*
2. *The function*

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

has a unique minimum at the point $\hat{\vartheta} \in \Theta$.

3. *The function $Q(\vartheta, t) \in \mathcal{C}_b^2$.*
4. *The function*

$$\ddot{\Phi}(\hat{\vartheta}) = 2\int_0^T \ddot{Q}(\hat{\vartheta}, t)\left[Q(\hat{\vartheta}, t) - S(\vartheta_0, t)\right] \mathrm{d}t + \int_0^T \dot{Q}(\hat{\vartheta}, t)^2 \,\mathrm{d}t > 0.$$

Let us denote

$$\mathbb{I}\left(\vartheta\right)=\int_{0}^{T}\dot{Q}(\vartheta,t)^{2}\,\mathrm{d}t,\qquad \mathbb{D}\left(\vartheta_{0}\right)^{2}=\ddot{\Phi}(\hat{\vartheta})^{-2}\mathbb{I}(\hat{\vartheta}).$$

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

$$\frac{\hat{\vartheta}_{\varepsilon}-\hat{\vartheta}}{\varepsilon}\Longrightarrow\hat{u}\sim\mathcal{N}\left(0,\mathbb{D}\left(\vartheta_{0}\right)^{2}\right),$$

and for any $p>0$

$$\lim_{\varepsilon\rightarrow0}\mathbf{E}_{\vartheta_{0}}\left|\frac{\hat{\vartheta}_{\varepsilon}-\hat{\vartheta}}{\varepsilon}\right|^{p}=\mathbf{E}_{\vartheta_{0}}\left|\hat{u}\right|^{p}.$$

Discontinuous versus discontinuous

The observed model is discontinuous and the statistician knows this but takes the wrong signals before and after the jump, then nevereless it is possible to have the consistent estimation.

Problem. The *theoretical model* is

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

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

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

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

$$q(\vartheta) > \frac{g(\vartheta) - h(\vartheta)}{2}, \qquad \vartheta \in \Theta, \tag{1}$$

then for $\vartheta < \vartheta_0$

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

For $\vartheta > \vartheta_0$ under condition

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

we obtain the similar inequality

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

Therefore

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

and we obtain the following result.

Proposition 2 *If the conditions* (1) *and* (2) *are fulfilled then the pseudo-MLE* $\hat{\vartheta}_\varepsilon$ *is consistent.*

It can be shown that

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

For the details see the similar problem in Section 5.3, K. 1994.

Cusp vs smooth

Suppose that the model of observations choosen by the statistician (*theoretical model*) is

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

The signal $Q\left(\vartheta, t\right)$ is supposed to be

$$Q\left(\vartheta, t\right) = a\left|t - \vartheta\right|^{\kappa}, \qquad 0 \leq t \leq T,$$

where $\kappa \in (0, \frac{1}{2})$ and $\vartheta \in \Theta = (\alpha < \vartheta < \beta)$. As before we suppose that $0 < \alpha < \beta < T$.

The observed process (*real model*) is

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

where $\vartheta_0 \in \Theta$ is the true value and $S\left(\vartheta, \cdot\right)$ is sufficiently smooth.

Introduce the function

$$\Phi(\vartheta,\hat\vartheta)=\left\|Q(\vartheta,\cdot)-S\left(\vartheta_0,\cdot\right)\right\|^2-\left\|Q(\hat\vartheta,\cdot)-S\left(\vartheta_0,\cdot\right)\right\|^2$$

and the conditions of regularity:

Condition $\mathcal{M}$.

1. *The parameter $\kappa\in\left(0,\frac{1}{2}\right)$.*
2. *The function $S\left(\vartheta,t\right)\in\mathcal{C}^2_\vartheta$.*
3. *The function $\Phi(\vartheta,\hat\vartheta)$ for all $\vartheta_0\in\Theta$ has a unique minimum at the point $\hat\vartheta=\hat\vartheta\left(\vartheta_0\right)$.*
4. *It's second derivative*

$$\gamma(\hat\vartheta)\equiv\left.\frac{\partial^2\Phi(\vartheta,\hat\vartheta)}{\partial\vartheta^2}\right|_{\vartheta=\hat\vartheta}>0$$

for all $\vartheta_0\in\Theta$.

Let us denote

$$\hat{Z}\left(u\right)=\exp\left\{aW^{H}\left(u\right)-\frac{\gamma(\hat{\vartheta})}{4}u^{2}\right\},\quad u\in R$$

$$\hat{Z}^{o}\left(u\right)=\exp\left\{w^{H}\left(v\right)-\frac{v^{2}}{2}\right\},\qquad v\in R$$

and define the random variables $\hat{\zeta},\hat{\zeta}_o$ by the relations

$$\hat{Z}(\hat{\zeta})=\sup_{u}\hat{Z}\left(u\right),\qquad \hat{Z}^{o}(\hat{\zeta}_{o})=\sup_{v}\hat{Z}^{o}\left(v\right).$$

Note that

$$\hat{\zeta}=\left(\frac{2a}{\gamma(\hat{\vartheta})}\right)^{\frac{H}{2H-1}}\hat{\zeta}_{o}.$$

Theorem 3 *Let the conditions $\mathcal{M}$ be fulfilled, then the estimator $\hat{\vartheta}_\varepsilon$ converges to the value $\hat{\vartheta}$, has the limit distribution*

$$\frac{\hat{\vartheta}_\varepsilon - \hat{\vartheta}}{\varepsilon^{\frac{2}{3-2\kappa}}} \Longrightarrow \hat{\zeta},$$

and for any $p > 0$

$$\lim_{\varepsilon\to 0} \mathbf{E}_{\vartheta_0} \left|\frac{\hat{\vartheta}_\varepsilon - \hat{\vartheta}}{\varepsilon^{\frac{2}{3-2\kappa}}}\right|^p = \mathbf{E}_{\vartheta_0} \left|\hat{\zeta}\right|^p = \left(\frac{2a}{\gamma(\hat{\vartheta})}\right)^{\frac{pH}{2H-1}} \mathbf{E}\left|\hat{\zeta}_o\right|^p .$$

For the proof see Chernoyarov, Dachian, K. 2015.

Discontinuous vs Cusp

Suppose that the model of observations choosen by the statistician is

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

where $\vartheta \in \Theta = (\alpha < \vartheta < \beta)$ and $0 < \alpha < \beta < T$. The observed process (*real model*) is

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

where $\vartheta_0 \in \Theta$ is the true value and

$$S\,(\vartheta_0, \cdot) = \mathrm{sgn}\,(t-\vartheta_0)\left[\left|t-\vartheta_0\right|^{\kappa} 1\!\!\mathrm{I}_{\{|t-\vartheta_0|\le 1\}} + 1\!\!\mathrm{I}_{\{|t-\vartheta_0|>1\}}\right]$$

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

Let us denote

$$\hat{Z}(u) = \exp\left\{W(u) - \frac{|u|^{\kappa+1}}{\kappa+1}\right\}, \quad u \in R$$

and define the random variable $\hat{\zeta}$ by the relations

$$\hat{Z}(\hat{\zeta}) = \sup_u \hat{Z}(u).$$

Below $H = \kappa + \frac{1}{2}$ is the Hurst constant.

Theorem 4 *The estimator $\hat{\vartheta}_\varepsilon$ is consistent has the limit distribution*

$$\frac{\hat{\vartheta}_\varepsilon - \vartheta_0}{\varepsilon^{\frac{1}{H}}} \Longrightarrow \hat{\zeta},$$

and for any $p > 0$ we have the convergence of p-moments.

Note that if $\kappa = \kappa_\varepsilon \to 0$, then the limit $(\varepsilon = 0)$ likelihood ratio coincides with the LR in the discontinuous case. The Kullback-Leibler distance between measures corresponding to the theoretical and real models is

$$
\begin{aligned}
D_{K-L} &= \frac{1}{2\varepsilon^2} \int_{-1}^{1} \left[Q\left(\vartheta_0, t\right) - S\left(\vartheta_0, t\right)\right]^2 \mathrm{d}t \\
&= \frac{2\kappa_\varepsilon^2}{\varepsilon^2 \left(\kappa_\varepsilon + 1\right)\left(2\kappa_\varepsilon + 1\right)} \approx \frac{2\kappa_\varepsilon^2}{\varepsilon^2}.
\end{aligned}
$$

Hence if $\kappa_\varepsilon = \varepsilon^{1+\gamma}$ with some $\gamma > 0$, then these two models (theoretical and real) are asymptotically indistingushible. The case $\kappa_\varepsilon = \varepsilon^{1-\gamma}$ merits to be studied.

Let us put $\kappa_\varepsilon = \varepsilon^{1-\gamma}$ and

$$\varphi_\varepsilon = \varepsilon^{\frac{1}{\kappa_\varepsilon + 1}}.$$

Then for the normalized likelihood ratio

$$\hat{Z}_\varepsilon\left(u\right) = \exp\left\{W\left(u\right) - \frac{\left|u\right|^{\kappa_\varepsilon+1}}{\kappa_\varepsilon + 1}\right\}, \qquad u \in \mathbb{U}_\varepsilon = \left(\frac{\alpha - \vartheta_0}{\varphi_\varepsilon}, \frac{\beta - \vartheta_0}{\varphi_\varepsilon}\right)$$

we obtain the convergence

$$\hat{Z}_\varepsilon\left(u\right) \Longrightarrow \hat{Z}\left(u\right) = \exp\left\{W\left(u\right) - \left|u\right|\right\}$$

Hence

$$\frac{\hat{\vartheta}_\varepsilon - \vartheta_0}{\varphi_\varepsilon} \Longrightarrow \hat{u}$$

with the corresponding $\hat{u}$.

Poisson Proceses. Misspecification.

Discontinuous versus smooth

Let us consider the following families: *theoretical* intensity function

$$\lambda\left(\vartheta,t\right)=\lambda\left(t\right)+\lambda_0\left(t\right)1\!\!\mathrm{I}_{\left\{t\geq\vartheta\right\}},\qquad \vartheta\in\left(\alpha,\beta\right),$$

where $0<\alpha<\beta<\tau$, $\inf_{\left\{\alpha<t<\beta\right\}}\lambda_0\left(t\right)>0$ and the *real* intensity function is

$$\lambda_*\left(\vartheta_0,t\right)=\lambda\left(t\right)+\lambda_0\left(t\right)\left(\frac{t-\vartheta_0}{\delta}+\frac{1}{2}\right)1\!\!\mathrm{I}_{\left\{\vartheta_0-\frac{\delta}{2}\leq t\leq\vartheta_0+\frac{\delta}{2}\right\}}$$
$$+\lambda_0\left(t\right)1\!\!\mathrm{I}_{\left\{t>\vartheta_0+\frac{\delta}{2}\right\}}.$$

Here $\vartheta_0\in\left(\alpha,\beta\right)$. For $\vartheta\in\left(\vartheta_0-\frac{\delta}{2},\vartheta_0+\frac{\delta}{2}\right)$ the both families have coinciding intensities outside of the interval $\left[\vartheta_0-\frac{\delta}{2}\leq t\leq\vartheta_0+\frac{\delta}{2}\right]$. We have to estimate the change-point parameter ϑ.

We suppose for simplicity that

$$\lambda\left(\vartheta,t\right)=\lambda+\lambda_0\,\mathbb{1}_{\{t\geq\vartheta\}},$$

$$\lambda_*\left(\vartheta_0,t\right)=\lambda+\lambda_0\left(\frac{t-\vartheta_0}{\delta}+\frac{1}{2}\right)\mathbb{1}_{\left\{\vartheta_0-\frac{\delta}{2}\leq t\leq\vartheta_0+\frac{\delta}{2}\right\}}+\lambda_0\,\mathbb{1}_{\left\{t>\vartheta_0+\frac{\delta}{2}\right\}}$$

respectively.

The difference between these two families is in the type of increasing of intensity function from the value λ to $\lambda+\lambda_0$. The first family corresponds to the *change-point* type intensity. Note that the technical devices have difficulties to provide a signal with intensity changing by a pure jump. The real systems have usually strongly increasing front but with finite rate.

The likelihood ratio (pseudo) is the discontinuous function

$$L\left(\vartheta, X^n\right) = \exp\left\{\sum_{j=1}^{n}\int_0^\tau \ln\lambda\left(\vartheta,t\right) \mathrm{d}X_j\left(t\right) - n\int_0^\tau \left[\lambda\left(\vartheta,t\right)-1\right]\mathrm{d}t\right\}$$

and the (pseudo) MLE $\hat\vartheta_n$ is defined by the same equation

$$L(\hat\vartheta_n, X^n) = \sup_{\vartheta\in\Theta} L\left(\vartheta, X^n\right).$$

To define the limit of the MLE we have to find the value $\hat\vartheta$ which minimizes the Kullback-Leibler distance

$$J_{K-L}(\hat\vartheta) = \inf_{\vartheta\in\Theta} J_{K-L}\left(\vartheta\right),$$

where

$$J_{K-L}\left(\vartheta\right) = \int_0^\tau \left[\lambda + \lambda_0 1\!\!\mathrm{I}_{\{t\geq\vartheta\}} - \lambda_*\left(\vartheta_0,t\right)\ln\left(\lambda + \lambda_0 1\!\!\mathrm{I}_{\{t\geq\vartheta\}}\right)\right]\mathrm{d}t.$$

Therefore the solution of the equation $\dot{J}_{K-L}(\hat{\vartheta}) = 0$ is

$$\hat{\vartheta} = \vartheta_0 - \frac{\delta}{2} + \delta \left(\frac{1}{\ln\left(1 + \frac{\lambda_0}{\lambda}\right)} - \frac{\lambda}{\lambda_0} \right).$$

The function $\Phi\left(x\right) = \left[\ln\left(1+x\right)\right]^{-1} - x^{-1} \in \left(0, \frac{1}{2}\right)$. We see that $\hat{\vartheta} = \hat{\vartheta}\left(\delta\right) \rightarrow \vartheta_0$ as $\delta \rightarrow 0$ and that $\hat{\vartheta} \in \left(\vartheta_0 - \frac{\delta}{2}, \vartheta_0\right)$. Note that the point $\hat{\vartheta}$ satisfies the equation

$$\lambda_*(\vartheta_0, \hat{\vartheta}) = \frac{\lambda_0}{\ln\left(1 + \frac{\lambda_0}{\lambda}\right)}.$$

Introduce the notations

$$\hat{Z}(u) = \exp\left\{W(u) - \frac{u^2}{2}\right\}, \qquad u \in \mathcal{R},$$

$$\hat{Z}(\hat{\xi}) = \sup_u \hat{Z}(u), \qquad a = \frac{\delta^{2/3}}{\left[\lambda_0 \ln\left(1 + \frac{\lambda_0}{\lambda}\right)\right]^{1/3}}$$

with double-sided Wiener process $W(\cdot)$, i.e.

$$W(u) = \begin{cases} W_1(u), & u \geq 0, \\ W_2(-u), & u \leq 0, \end{cases}$$

where $W_1(u), u \geq 0$ and $W_1(u), u \geq 0$ are two independent Wiener processes.

Proposition 3 *The pseudo-MLE $\hat\vartheta_n$ is "consistent"*

$$\mathbf{P}_{\vartheta_0} - \lim_{n\to\infty} \hat\vartheta_n = \hat\vartheta$$

and for any $p > 0$ the convergences

$$n^{1/3}\left(\hat\vartheta_n - \hat\vartheta\right) \Longrightarrow a\,\hat\xi, \qquad n^{p/3}\mathbf{E}_{\vartheta_0}\left|\hat\vartheta_n - \hat\vartheta\right|^p \longrightarrow a^p\mathbf{E}\left|\hat\xi\right|^p$$

hold.

Let us see what happens with the pseudo-BE in this situation. We have the representaton

$$\tilde{\vartheta}_n = \frac{\int_\alpha^\beta \theta p\left(\theta\right) L\left(\theta, X^n\right) \mathrm{d}\theta}{\int_\alpha^\beta p\left(\theta\right) L\left(\theta, X^n\right) \mathrm{d}\theta} = \hat{\vartheta} + \varphi_n \frac{\int_{\mathbb{U}_n} up\left(\theta_u\right) L\left(\theta_u, X^n\right) \mathrm{d}u}{\int_{\mathbb{U}_n} p\left(\theta_u\right) L\left(\theta_u, X^n\right) \mathrm{d}u},$$

where we changed the variables $\theta = \theta_u = \hat{\vartheta} + \varphi_n u$. Therefore we can write

$$\begin{aligned}\frac{\tilde{\vartheta}_n - \hat{\vartheta}}{\varphi_n} &= \frac{\int_{\mathbb{U}_n} up\left(\theta_u\right) Z_n\left(u\right) \mathrm{d}u}{\int_{\mathbb{U}_n} p\left(\theta\right) Z_n\left(u\right) \mathrm{d}u} = \frac{\int_{\mathbb{U}_n} u\, Z_n\left(u\right) \mathrm{d}u}{\int_{\mathbb{U}_n} Z_n\left(u\right) \mathrm{d}u} \left(1 + o\left(1\right)\right) \\ &= \frac{\int_{\mathbb{U}_n} u\hat{Z}_n\left(u\right)^{\sqrt{n\varphi_n\lambda_0\gamma}} \mathrm{d}u}{\int_{\mathbb{U}_n} \hat{Z}_n\left(u\right)^{\sqrt{n\varphi_n\lambda_0\gamma}} \mathrm{d}u} \left(1 + o\left(1\right)\right).\end{aligned}$$

The limit of the last ratio is an open problem. It probably coincides with the limit of the following exppression

$$\frac{\int_{\mathcal{R}} u\,\hat{Z}\,(u)^{\sqrt{n\varphi_n\lambda_0\gamma}}\,\mathrm{d}u}{\int_{\mathcal{R}} \hat{Z}\,(u)^{\sqrt{n\varphi_n\lambda_0\gamma}}\,\mathrm{d}u} \Longrightarrow \hat{\xi}$$

because the main contribution in the integral

$$\int_{\mathcal{R}} u\,\left(e^{W(u)-\frac{u^2}{2}}\right)^{\sqrt{n\varphi_n\lambda_0\gamma}}\,\mathrm{d}u$$

is given by the maximal values $W(\hat{\xi}) - \frac{\hat{\xi}^2}{2}$. Therefore the pseudo-MLE and pseudo-BE are asymptotically equivalent in this situation too.

Smooth vs discontinuous

Let us consider the opposite situation, i.e., we suppose that the *theoretical* intensity function is smooth:

$$\lambda(\vartheta,t)=\lambda+\delta^{-1}\lambda_0\left(t-\vartheta+\frac{\delta}{2}\right)\mathbb{I}_{\left\{\vartheta-\frac{\delta}{2}\leq t\leq\vartheta+\frac{\delta}{2}\right\}}+\lambda_0\mathbb{I}_{\left\{t>\vartheta+\frac{\delta}{2}\right\}},$$

where $0<\alpha<\beta<\tau$ and the *real* intensity function is discontinuous:

$$\lambda_*(\vartheta_0,t)=\lambda+\lambda_0\mathbb{I}_{\{t\geq\vartheta_0\}},\qquad \vartheta_0\in(\alpha,\beta).$$

The value $\hat\vartheta$ minimizes the Kullback-Leibler distance

$$\begin{aligned}J_{K-L}(\vartheta)=&\int_0^{\vartheta_0}\left[\lambda(\vartheta,t)-\lambda\ln\lambda(\vartheta,t)\right]\mathrm{d}t\\&+\int_{\vartheta_0}^{\tau}\left[\lambda(\vartheta,t)-(\lambda+\lambda_0)\ln\lambda(\vartheta,t)\right]\mathrm{d}t.\end{aligned}$$

Hence, if we suppose that $\hat{\vartheta} - \frac{\delta}{2} < \vartheta_0 < \hat{\vartheta} + \frac{\delta}{2}$ then $\hat{\vartheta}$ is solution of the equation

$$\dot{J}_{K-L}\left(\vartheta\right) = -\frac{\lambda_0}{\delta}\int_{\vartheta-\frac{\delta}{2}}^{\vartheta_0}\left[1-\frac{\lambda}{\lambda\left(\vartheta,t\right)}\right]\mathrm{d}t - \frac{\lambda_0}{\delta}\int_{\vartheta_0}^{\vartheta+\frac{\delta}{2}}\left[1-\frac{\lambda+\lambda_0}{\lambda\left(\vartheta,t\right)}\right]\mathrm{d}t = 0$$

Let us denote

$$D(\vartheta_0,\hat{\vartheta})^2 = \frac{\lambda_0^2}{2\delta^2}\int_{\hat{\vartheta}-\frac{\delta}{2}}^{\hat{\vartheta}+\frac{\delta}{2}}\frac{\lambda_*\left(\vartheta_0,t\right)}{\left[\lambda+\frac{\lambda_0}{\delta}\left(t-\hat{\vartheta}+\frac{\delta}{2}\right)\right]^2}\mathrm{d}t$$

and put $\varphi_n = (bn)^{-1/2}$, where $b = D(\vartheta_0,\hat{\vartheta})^2$.

Proposition 4 *The pseudo-MLE $\hat{\vartheta}_n$ is "cosistent"*

$$\mathbf{P}_{\vartheta_0} - \lim_{n\to\infty} \hat{\vartheta}_n = \hat{\vartheta}$$

and for any $p > 0$ we have the convergences

$$n^{1/2}\left(\hat{\vartheta}_n - \hat{\vartheta}\right) \Longrightarrow \mathcal{N}\left(0, D(\vartheta_0, \hat{\vartheta})^{-2}\right),$$

$$n^{p/2}\mathbf{E}_{\vartheta_0}\left|\hat{\vartheta}_n - \hat{\vartheta}\right|^p \longrightarrow D(\vartheta_0, \hat{\vartheta})^{-p}\mathbf{E}\left|\zeta\right|^p$$

hold. Here $\zeta \sim \mathcal{N}(0,1)$.

Discontinuous vs cusp

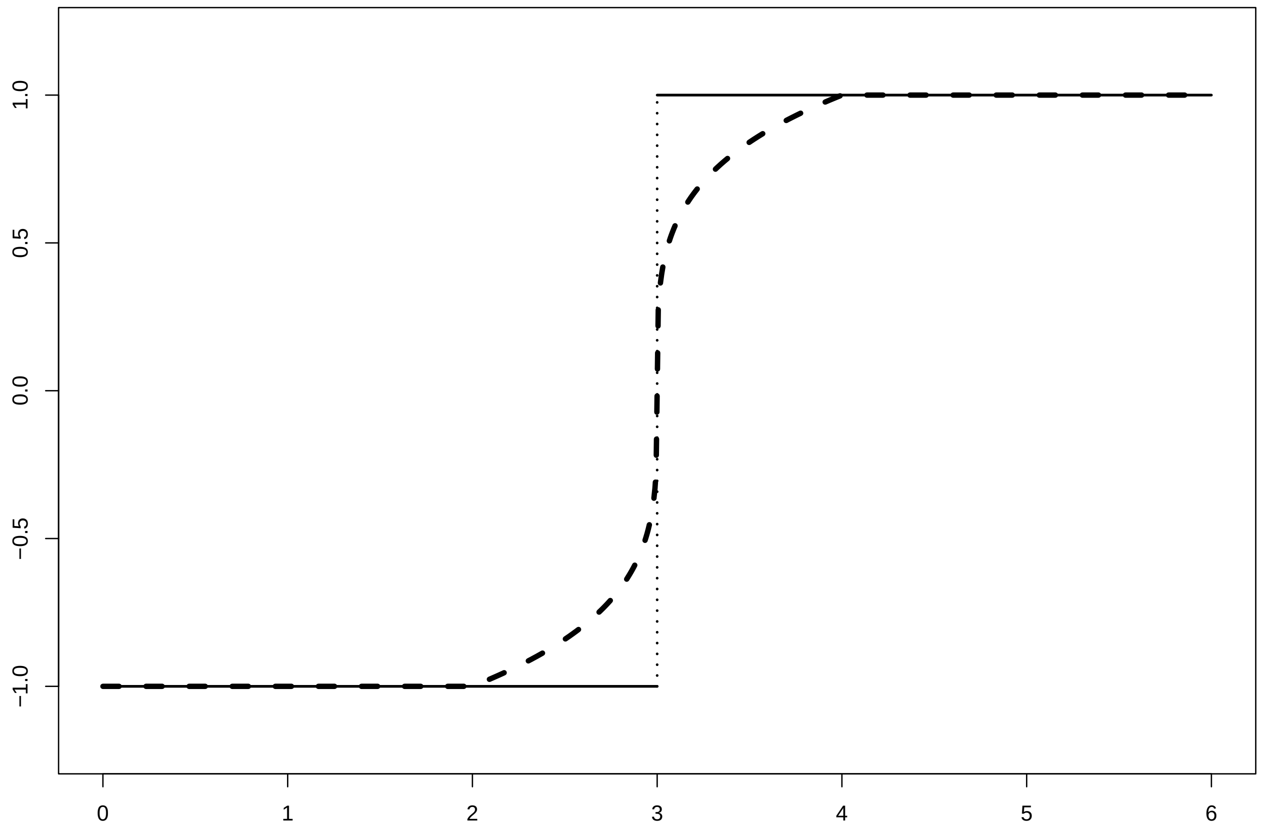
The next situation is in some sense the more close to the real problem. Suppose that the *theoretical* intensity function assumed by the statistician is of change-point type

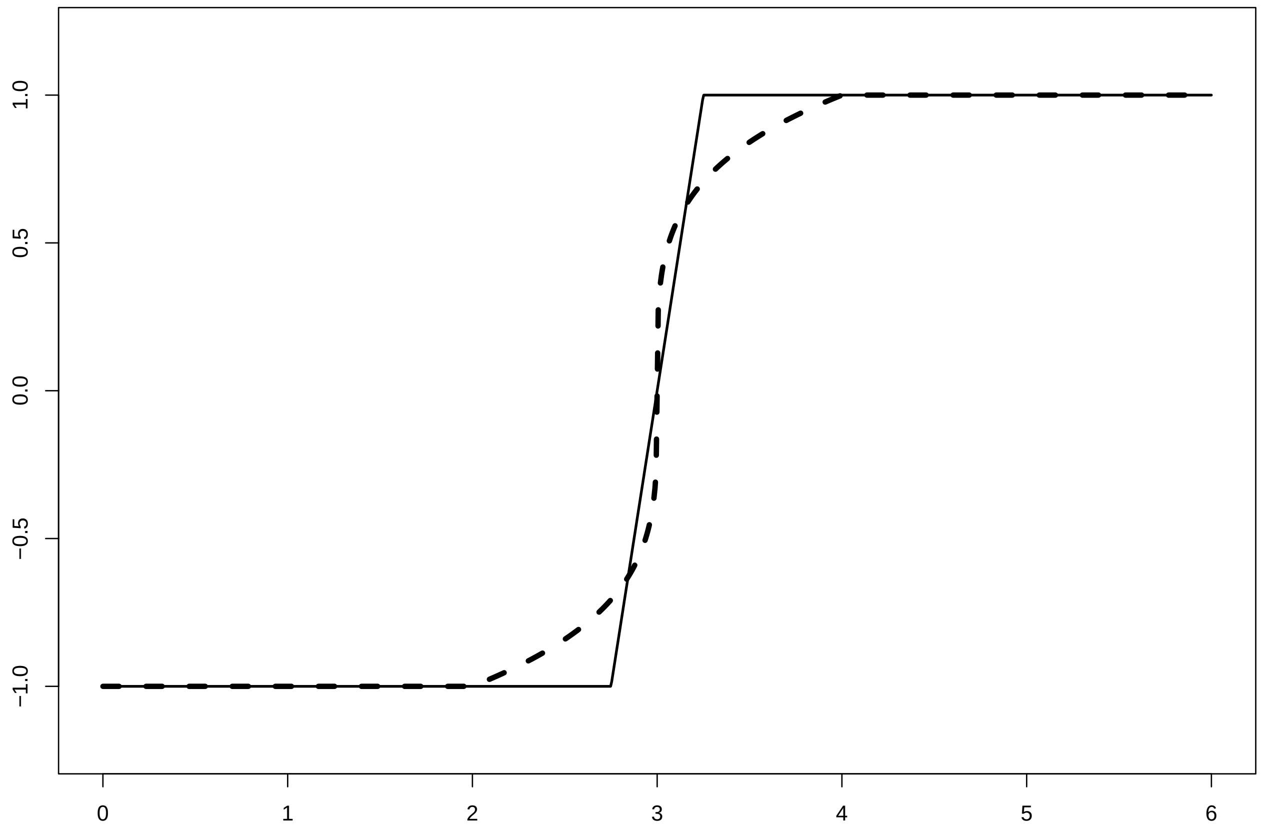
$$\lambda\left(\vartheta,t\right)=\lambda+\lambda_0 1\!\!1_{\{t>\vartheta\}}$$

but the *real* intensity function is of special cusp-type form

$$\lambda_*\left(\vartheta_0,t\right)=\lambda+\frac{\lambda_0}{2}\left[1+\operatorname{sgn}\left(t-\vartheta_0\right)\left|t-\vartheta_0\right|^{\kappa}\right]1\!\!1_{\{|t-\vartheta_0|\leq 1\}}+\lambda_0 1\!\!1_{\{t>\vartheta_0+1\}},$$

where $\kappa\in\left(0,\frac{1}{2}\right)$. The choice of κ close to 0 allows to have a good approximation of the indicator function.


1.0
0.5
0.0
−0.5
−1.0
0
1
2
3
4
5
6


1.0
0.5
0.0
−0.5
−1.0
0
1
2
3
4
5
6

The more interesting case corresponds to the values $\hat{\vartheta} \in (\vartheta_0 - 1, \vartheta_0 + 1)$. Then the value $\hat{\vartheta}$ which minimizes the Kullback-Leibler distance as before satisfies the equation

$$\lambda_*(\vartheta_0, \hat{\vartheta}) = \frac{\lambda_0}{\ln\left(1 + \frac{\lambda_0}{\lambda}\right)}.$$

Hence in the case $\hat{\vartheta} < \vartheta_0$ we obtain

$$\hat{\vartheta} = \vartheta_0 - \left[1 - 2\left(\frac{1}{\ln\left(1 + \frac{\lambda_0}{\lambda}\right)} - \frac{\lambda}{\lambda_0}\right)\right]^{\frac{1}{\kappa}}.$$

Note that if $\frac{\lambda_0}{\lambda}$ takes small values then $\hat{\vartheta} \approx \vartheta_0$ because as $x \to 0$

$$\frac{1}{\ln(1+x)} - \frac{1}{x} = \frac{x - \ln(1+x)}{x \ln(1+x)} = \frac{\frac{x^2}{2} + o\left(x^2\right)}{x\left(x - \frac{x^2}{2} + o\left(x^2\right)\right)} = \frac{1}{2} + o(1).$$

Below we put

$$b = \left(\frac{\kappa}{2}\left|\hat\vartheta - \vartheta_0\right|^{\kappa-1}\sqrt{\lambda_0 \ln\left(1+\frac{\lambda_0}{\lambda}\right)}\right)^{-2/3}.$$

Proposition 5 *The pseudo-MLE $\hat\vartheta_n$ is "cosistent"*

$$\mathbf{P}_{\vartheta_0} - \lim_{n\to\infty} \hat\vartheta_n = \hat\vartheta$$

and for any $p > 0$ we have the convergences

$$n^{1/3}\left(\hat\vartheta_n - \hat\vartheta\right) \Longrightarrow b\,\hat\xi,$$

$$n^{p/3}\mathbf{E}_{\vartheta_0}\left|\hat\vartheta_n - \hat\vartheta\right|^p \longrightarrow b^p\mathbf{E}\left|\hat\xi\right|^p$$

hold.

It is interesting to note that the rate $n^{-1/3}$ is due to the form of the Kullback-Leibler distance in the case of Poisson processes. The same intensities considered as signals observed in the white Gaussian noise provide a different rate. Indeed suppose that the observed process is

$$\mathrm{d}X_t = \lambda_*\left(\vartheta_0, t\right)\mathrm{d}t + \varepsilon_n \mathrm{d}W_t, \qquad 0 \leq t \leq \tau$$

and the statistician estimates ϑ on the base of the *theoretical* model

$$\mathrm{d}X_t = \lambda\left(\vartheta, t\right)\mathrm{d}t + \varepsilon_n \mathrm{d}W_t, \qquad 0 \leq t \leq \tau,$$

where the signals $\lambda_*\left(\cdot\right)$ and $\lambda\left(\cdot\right)$ are defined aboveand $\varepsilon_n = n^{-1/2}$.

Then the value $\hat{\vartheta}$ which minimizes the Kullback-Leibler distance is

$$\hat{\vartheta} = \arg \inf_{\vartheta \in \Theta} \int_0^{\tau} \left[\lambda\left(\vartheta, t\right) - \lambda_*\left(\vartheta_0, t\right)\right]^2 \mathrm{d}t = \vartheta_0.$$

The MLE in this problem has the following limit

$$n^{\frac{1}{\kappa+2}} \left(\hat{\vartheta}_n - \vartheta_0\right) \Longrightarrow \hat{\zeta} = \arg\sup_u \left\{W\left(u\right) - \frac{\left|u\right|^{\kappa+1}}{\kappa+1}\right\}.$$

References

[1] Chernoyarov, O.V., Dachian, S.Yu., Kutoyants, Yu.A. (2015) On prameter estimation for cusp-type signal. To appear in the Annales of the Institute of Statistical Mathematics.

[2] Chernoyarov, O.V., Kutoyants, Yu.A., Trifonov, A.P. (2015) On misspecification in regularity and properties of estimators. Subm.

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

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

[5] Höpfner, R., Kutoyants, Yu.A. (2012) On frequency estimation of periodic ergodic diffusion process. *Probl. Inform. Transm.*, 48, 2, 127-141.

[6] Huber, O.J. (1981) *Robust Statistics.* Wiley.

[7] Ibragimov I.A., Has'minskii R.Z. (1975) An estimator of the parameter of a signal in Gaussian white noise, *Probl. Inform. Transm.*, 10, 31-46.

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

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

[10] Kutoyants, Yu.A. (1977) Estimation of signal parameter in Gaussian noise. *Probl. Inform. Transm.*, 13, 4, 266-271.

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

[12] Liptser, R.S. and Shiryaev, A.N. (2001) *Statistics of Random Processes*, v. 1, 2-nd ed., Springer, N.Y.

[13] Novikov, A. and Kordzakhia, N. (2013) Pitman estimators.*Theory Probab. Appl.,* 57, 3, 521-529.

[14] Novikov, A., Kordzakhia, N. and Ling, T. (2014) On moments of Pitman estimators. *Theory Probab. Appl.,* 58, 4, 601-614.

[15] Rubin, H., Song, K. (1995) Exact computation of the asymptotic efficiency of MLE of a dicontinuous signal in a GWN. *Ann. Statist.* 23, 732–739.

[16] Stigler, S.M. (2010) The changing history of robustes. *American Statistician.* 64, 4, 277-281.

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