Записки научных
семинаров ПОМИ
Том 363, 2009 г.

Y. A. Kutoyants

ON PROPERTIES OF ESTIMATORS IN NONREGULAR SITUATIONS FOR POISSON PROCESSES

Abstract. We consider the problem of parameter estimation by observations of inhomogeneous Poisson process. It is well-known that if the regularity conditions are fulfilled then the maximum likelihood and bayesian estimators are consistent, asymptotically normal, and asymptotically efficient. These regularity conditions can be roughly presented as follows: a) the intensity function of observed process belongs to known parametric family of functions, b) the model is identifiable, c) the Fisher information is positive continuous function, d) the intensity function is sufficiently smooth with respect to the unknown parameter, e) this parameter is an interior point of the interval. We are interested in the properties of estimators when these regularity conditions are not fulfilled. More precisely, we preset a review of the results which correspond to the rejection of these conditions one by one and we show how the properties of the MLE and Bayesian estimators change. The proofs of these results are essentially based on some general results by Ibragimov and Khasminskii.

1. Introduction

7 We start with the classical model of i.i.d. observations. Let $X_1, \cdots, X_n$ be independent and identically distributed random variables with the density function $f_*(x)$. We suppose that $f_*(x) = f(\vartheta, x)$, where $f(\cdot,\cdot)$ is a known function depending on the unknown parameter $\vartheta \in \Theta = (\alpha, \beta)$. We have to estimate ϑ and to describe the properties of estimators in the asymptotic of large samples ($n \to \infty$). We discuss below two estimators: maximum likelihood and bayesian. Let us introduce the likelihood function $L_n(\vartheta, X^n) = \prod_{j=1}^n f(\vartheta, X_j)$. Then the maximum likelihood estimator (MLE) $\hat\vartheta_n$ and bayesian estimator (BE) $\tilde\vartheta_n$ (for quadratic loss function and density a priori $p(\cdot)$) are defined by the equations

$$L_n\left(\hat\vartheta_n, X^n\right) = \sup_{\vartheta\in\Theta} L_n(\vartheta, X^n), \quad \tilde\vartheta_n = \frac{\int_\alpha^\beta \theta p(\theta) L_n(\vartheta, X^n)\,\mathrm{d}\theta}{\int_\alpha^\beta p(\theta) L_n(\vartheta, X^n)\,\mathrm{d}\theta}. \qquad (1)$$

It is well-known that if the *conditions of regularity* are fulfilled then these estimators are consistent, asymptotically normal

$$\sqrt{n}\left(\hat{\vartheta}_n-\vartheta\right)\Longrightarrow\mathcal{N}\left(0,\mathrm{I}\left(\vartheta\right)^{-1}\right),\quad \sqrt{n}\left(\tilde{\vartheta}_n-\vartheta\right)\Longrightarrow\mathcal{N}\left(0,\mathrm{I}\left(\vartheta\right)^{-1}\right),$$

and asymptotically efficient. The density $f(\vartheta,x)$ has to take place in Fisher information as

$$\mathrm{I}\left(\vartheta\right)=\int\frac{\dot{f}\left(\vartheta,x\right)^2}{}\mathrm{d}\,\mu\left(x\right)$$

is the Fisher information. The proofs you can find in any book on asymptotical statistics, e.g., Ibragimov and Khasminski (1981).

These regularity conditions can be roughly described as follows

- *The density $f_*(x)$ of the observed r.v.'s belongs to the parametric family, i.e., there exists a value $\vartheta_0\in\Theta=(\alpha,\beta)$ such that $f_*(x)=f(\vartheta_0,x)$.*
- *The function $f(\vartheta,x)$ is one or more times differentiable w.r.t. ϑ with certain majoration of the derivatives.*
- *The Fisher information $\mathrm{I}(\vartheta)$ is positive function.*
- *The Fisher information $\mathrm{I}(\vartheta)$ is continuous function.*
- *The model is identifiable: if $\vartheta_1\neq\vartheta_2$ then $f(\vartheta_1,x)\neq f(\vartheta_2,x)$.*
- *The true value ϑ_0 is an interior point of the set Θ, i.e., $\vartheta_0\neq\alpha$ and $\vartheta_0\neq\beta$.*
- *We can observe all values of the random variables X_j.*
- *The statistical model is fixed and can not be chosen in some optimal way.*

Of course, this list is not exhaustive and the other conditions can be mentioned too. We are interested by the properties of estimators, when the similar regularity conditions are not fulfilled for some models of continuous time stochastic processes. More precisely, we replace these regularity conditions by other conditions and study the properties of estimators under these new conditions. This approach allows to understand better the role of each regularity condition in the properties of estimators. As the model of observations in this work we take inhomogeneous Poisson process. The similar work concerning parameter estimation for ergodic diffusion processes was already published (see [7]), but it seems that the more detailed exposition of the proofs will be useful and it is given here.

2. Regular case

We observe n independent trajectories $X^n = (X_1, \ldots, X_n)$, where $X_j = \{X_j(t), 0 \le t \le \tau\}$, of a Poisson process $X^\tau = \{X(t), 0 \le t \le \tau\}$ of intensity function $\lambda_* = \{\lambda_*(t), 0 \le t \le \tau\}$, i.e., $X(0) = 0$, the increments on disjoint intervals are independent and

$$\mathbf{P}\{X(t) = k\} = \frac{\Lambda(t)^k}{k!} \exp\{-\Lambda(t)\}, \quad \Lambda(t) = \int_0^t \lambda_*(s)\, \mathrm{d}s.$$

The Poisson process sometimes is defined as a series of events $0 < t_1 < t_2 < \ldots < t_M < T$ and $X(t)$, $0 \le t \le \tau$ is the corresponding counting process, i.e., $X(t)$ is equal to the number of events observed up to time t. The process X^τ is càdlàg (right continuous with left limits at every point t).

The same model of observation we obtain in the case of τ-periodic Poisson process $X^{T_n} = \{X(s), 0 \le s \le T_n\}$, if the intensity function $\lambda_*(s)$ is τ-periodic and $T_n = \tau n$. Then we can cut the trajectory X^{T_n} on n pieces $X_j(t) = X(t + (j-1)\tau) - X((j-1)\tau)$, $0 \le t \le \tau$ with $j = 1, \ldots, n$. We suppose that the period τ is known (does not depend on ϑ). As the increments of the Poisson process are independent, this model coincides with the mentioned above one.

The statistician can suppose that this intensity function belongs to some parametric class of functions, i.e., $\lambda_* = \lambda_\vartheta$, where $\lambda_\vartheta = \{\lambda(\vartheta, t), 0 \le t \le \tau\}$ with $\vartheta \in \Theta = (\alpha, \beta)$. Therefore he (or she) obtains the problem of estimation of the parameter ϑ by the observations X^n of the Poisson process of intensity function λ_ϑ, $\vartheta \in \Theta$.

We suppose that the intensity is bounded positive function and hence the likelihood ratio function for this parametric family is

$$L(\vartheta, X^n) = \exp\left\{\sum_{j=1}^n \int_0^\tau \ln \lambda(\vartheta, t)\, \mathrm{d}X_j(t) - n \int_0^\tau [\lambda(\vartheta, t) - 1]\, \mathrm{d}t\right\} \quad (2)$$

and the MLE $\hat\vartheta_n$ and BE $\tilde\vartheta_n$ for quadratic loss function and prior density $p(\theta), \theta \in \Theta$ (positive, continuous on Θ) are defined by the same equations (1).

Regularity Conditions:

1. *There exists* $\vartheta_0 \in \Theta$ *such that* $\lambda_*(t) = \lambda(\vartheta_0, t), 0 \le t \le \tau$.

2. *The function* $\sqrt{\lambda(\vartheta,t)}, 0 \leq t \leq \tau$ *has two continuous bounded derivatives with respect to* ϑ.
3. *The Fisher information*

$$0 < \mathrm{I}(\vartheta) = \int_0^\tau \frac{\dot\lambda(\vartheta,t)^2}{\lambda(\vartheta,t)}\,\mathrm{d}t < \infty.$$

4. *The Fisher information* $\mathrm{I}(\vartheta)$ *is continuous function.*
5. *The condition of identifiability is fulfilled: for any* $\nu > 0$

$$\inf_{|\theta-\vartheta_0|>\nu}\int_0^\tau \left[\sqrt{\lambda(\vartheta,t)}-\sqrt{\lambda(\vartheta_0,t)}\right]^2\,\mathrm{d}t > 0.$$

6. *The parameter* ϑ_0 *is an interior point of the set* $\Theta = (\alpha,\beta)$.
7. *The process* $X_j(t)$ *is observed on the whole interval* $[0,\tau]$.
8. *The model of observed process is fixed, i.e., in the statement of the problem the intensity function* λ_ϑ *is given (can not be chosen by the statistician).*

Of course, Condition 2 implies 4, but we present both of them, because we consider below the case, when 4 is not fulfilled. The properties of estimators are described in the following theorem.

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

$$\sqrt{n}\left(\hat\vartheta_n-\vartheta_0\right)\Longrightarrow\mathcal{N}\left(0,\frac{1}{\mathrm{I}(\vartheta_0)}\right),\qquad \sqrt{n}\left(\tilde\vartheta_n-\vartheta_0\right)\Longrightarrow\mathcal{N}\left(0,\frac{1}{\mathrm{I}(\vartheta_0)}\right)$$

asymptotically efficient and the moments of these estimators converge too.

The proof can be found in [6], Theorems 2.4 and 2.5.

This proof is essentialy based on the general results obtained by Ibragimov and Khasminskii [5], which we present below in a bit more general situation, than we need for this theorem. Let us denote by $Z_n(u)$ the normalized likelihood ratio process

$$Z_n(u) = \frac{L(\vartheta_0+\varphi_n u, X^n)}{L(\vartheta_0,X^n)},\quad u\in\mathbb{U}_n=\left(\frac{(\alpha-\vartheta_0)}{\varphi_n},\frac{(\beta-\vartheta_0)}{\varphi_n}\right),$$

where $\varphi_n \to 0$ and the rate of this convergence is such that $Z_n(u)$ has some non degenerate limit (in distribution) $Z(u)$. Below we suppose that in the bayesian case (ϑ is a random variable) the loss function is quadratic and the density *a priory* $p(\vartheta)$, $\vartheta \in (\alpha, \beta)$ is continuous positive function. Let us define the random variables $\hat{u}$ and $\tilde{u}$ by the relations

$$Z(\hat{u}) = \sup_{u \in \mathcal{R}} Z(u), \quad \tilde{u} = \frac{\int_{\mathcal{R}} u\, Z(u)\, \mathrm{d}u}{\int_{\mathcal{R}} Z(u)\, \mathrm{d}u}. \tag{3}$$

The study of the likelihood ratio $Z_n(\cdot)$ allows to describe the properties of estimators (maximum likelihood and bayesian) and this is illustrated by the following theorem.

Theorem 2 (Ibragimov, Khasminskii)**.** *Suppose that the following conditions are fulfilled*

1. *There exist constants $a > 1$ and $B > 0$ such that for all $u \in \mathbb{U}_n$*

$$\mathbf{E}_\vartheta \left| Z_n^{1/2}(u_2) - Z_n^{1/2}(u_1) \right|^2 \le B\, |u_2 - u_1|^a. \tag{4}$$

2. *There exist constants $\kappa > 0$ and $\gamma > 0$ such that for all $u \in \mathbb{U}_n$*

$$\mathbf{E}_\vartheta Z_n^{1/2}(u) \le e^{-\kappa |u|^\gamma}. \tag{5}$$

3. *The marginal distributions*

$$(Z_n(u_1), \ldots, Z_n(u_k)) \Longrightarrow (Z(u_1), \ldots, Z(u_k))$$

and $Z(\cdot)$ attains with probability 1 its maximal value at a unique point $\hat{u}$.

Then, the MLE $\hat{\vartheta}_n$ *and* BE $\tilde{\vartheta}_n$ *are consistent,*

$$\varphi_n^{-1}\left(\hat{\vartheta}_n - \vartheta\right) \Longrightarrow \hat{u}, \quad \varphi_n^{-1}\left(\tilde{\vartheta}_n - \vartheta\right) \Longrightarrow \tilde{u},$$

and for any $p > 0$,

$$\mathbf{E}_\vartheta \left| \frac{\hat{\vartheta}_n - \vartheta}{\varphi_n} \right|^p \longrightarrow \mathbf{E}_\vartheta |\hat{u}|^p, \quad \mathbf{E}_\vartheta \left| \frac{\tilde{\vartheta}_n - \vartheta}{\varphi_n} \right|^p \longrightarrow \mathbf{E}_\vartheta |\tilde{u}|^p.$$

For the proof (essentially more general results) see [5, Theorems 3.1.1 and 3.2.1]. Note that in the case of bayesian estimators it is sufficient that the parameter $a > 0$.

In the regular case of the Theorem 1, the sequence $\varphi_n = n^{-1/2}$ and the limit process is

$$Z(u) = \exp\left\{u\,\zeta(\vartheta_0) - \frac{u^2}{2}\,\mathrm{I}(\vartheta_0)\right\}, \quad u \in \mathcal{R},$$

where $\zeta(\vartheta_0) \sim \mathcal{N}(0, \mathrm{I}(\vartheta_0))$. Hence

$$\hat{u} = \frac{\zeta(\vartheta_0)}{\mathrm{I}(\vartheta_0)} \quad \sim \quad \mathcal{N}\left(0, \mathrm{I}(\vartheta_0)^{-1}\right).$$

To check the conditions (4) and (5) in the case of inhomogeneous Poisson processes we use the following estimates (below ϑ_0 is the true value and $\vartheta_i = \vartheta_0 + \frac{u_i}{\sqrt{n}}$)

$$\begin{aligned}
\mathbf{E}_{\vartheta_0}\left|Z_n^{1/2}(u_2) - Z_n^{1/2}(u_1)\right|^2 &= 2 - 2\mathbf{E}_{\vartheta_0}\left[Z_n(u_2)\,Z_n(u_1)\right]^{1/2} \\
&= 2 - 2\mathbf{E}_{\vartheta_1}\left[\frac{Z_n(u_2)}{Z_n(u_1)}\right]^{1/2} \\
&= 2 - 2\exp\left\{-\frac{n}{2}\int_0^\tau \left[\sqrt{\lambda(\vartheta_2, t)} - \sqrt{\lambda(\vartheta_1, t)}\right]^2 \mathrm{d}\,t\right\} \\
&\le n\int_0^\tau \left[\sqrt{\lambda(\vartheta_2, t)} - \sqrt{\lambda(\vartheta_1, t)}\right]^2 \mathrm{d}\,t \qquad (6)
\end{aligned}$$

and $\left(\vartheta_u = \vartheta_0 + \frac{u}{\sqrt{n}}\right)$

$$\begin{aligned}
&\mathbf{E}_{\vartheta_0} Z_n^{1/2}(u) \\
&= \left(\mathbf{E}_{\vartheta_0}\exp\left\{\frac{1}{2}\int_0^\tau \ln\frac{\lambda(\vartheta_u, t)}{\lambda(\vartheta_0, t)}\mathrm{d}\,X(t) - \frac{1}{2}\int_0^\tau \left[\lambda(\vartheta_u, t) - \lambda(\vartheta_0, t)\right]\mathrm{d}\,t\right\}\right)^n \\
&= \exp\left\{-\frac{n}{2}\int_0^\tau \left[\sqrt{\lambda(\vartheta_u, t)} - \sqrt{\lambda(\vartheta_0, t)}\right]^2 \mathrm{d}\,t\right\}. \qquad (7)
\end{aligned}$$

The regularity conditions allow to obtain the low and upper estimates

$$c\left|\vartheta_2-\vartheta_1\right|^2 \leq \int\limits_0^\tau \left[\sqrt{\lambda\left(\vartheta_2,t\right)}-\sqrt{\lambda\left(\vartheta_1,t\right)}\right]^2 \mathrm{d}\,t \leq C\left|\vartheta_2-\vartheta_1\right|^2 \qquad (8)$$

which provide immediately (4) and (5). Using the direct expansion of the functions

$$\lambda\left(\vartheta_0+\frac{u}{\sqrt{n}},t\right)=\lambda\left(\vartheta_0,t\right)+\frac{u}{\sqrt{n}}\dot{\lambda}\left(\vartheta_0,t\right)+o\left(\frac{u}{\sqrt{n}}\right)$$

and $\ln\lambda\left(\vartheta_0+\frac{u}{\sqrt{n}},t\right)$ we obtain the following representation of the likelihood ratio

$$Z_n\left(u\right)=\exp\left\{u\Delta_n\left(\vartheta_0,X^n\right)-\frac{u^2}{2}\mathrm{I}\left(\vartheta_0\right)+r_n\right\},$$

where

$$\Delta_n\left(\vartheta_0,X^n\right)=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\int\limits_0^\tau\frac{\dot{\lambda}\left(\vartheta_0,t\right)}{\lambda\left(\vartheta_0,t\right)}\left[\mathrm{d}\,X_j\left(t\right)-\lambda\left(\vartheta_0,t\right)\mathrm{d}\,t\right]\Longrightarrow\zeta\left(\vartheta_0\right)$$

and $r_n\to 0$. This representation provides the convergence of the marginal distributions of the process $Z_n\left(\cdot\right)$ to the marginal distributions of the process $Z\left(\cdot\right)$. Therefore all conditions of the Theorem 2 are fulfilled and the MLE and BE are consistent, asymptotically normal. Let us remind how the weak convergence of the likelihood ratio process provides these properties of estimators.

Suppose that we already have the weak convergence of the stochastic processes

$$Z_n\left(\cdot\right)\Longrightarrow Z\left(\cdot\right) \qquad (9)$$

in the space of continuous on $\mathcal{R}$ functions vanishing in infinity. Then according to [5] the asymptotic normality of the MLE can be obtained by the following way.

$$
\begin{aligned}
&\mathbf{P}\left\{\sqrt{n}\left(\hat\vartheta_n-\vartheta_0\right)<x\right\}\\
=&\mathbf{P}\left\{\sup_{\sqrt{n}(\theta-\vartheta_0)<x} L\left(\vartheta,X^n\right)>\sup_{\sqrt{n}(\theta-\vartheta_0)\geq x} L\left(\vartheta,X^n\right)\right\}\\
=&\mathbf{P}\left\{\sup_{\sqrt{n}(\theta-\vartheta_0)<x}\frac{L\left(\vartheta,X^n\right)}{L\left(\vartheta_0,X^n\right)}>\sup_{\sqrt{n}(\theta-\vartheta_0)\geq x}\frac{L\left(\vartheta,X^n\right)}{L\left(\vartheta_0,X^n\right)}\right\}\\
=&\mathbf{P}\left\{\sup_{u<x} Z_n\left(u\right)>\sup_{u\geq x} Z_n\left(u\right)\right\}\longrightarrow\mathbf{P}\left\{\sup_{u<x} Z\left(u\right)>\sup_{u\geq x} Z\left(u\right)\right\}\\
=&\mathbf{P}\left(\frac{\zeta\left(\vartheta_0\right)}{\mathrm{I}\left(\vartheta_0\right)}<x\right),\quad\text{i.e.}\quad\sqrt{n}\left(\hat\vartheta_n-\vartheta_0\right)\Longrightarrow\mathcal{N}\left(0,\frac{1}{\mathrm{I}\left(\vartheta_0\right)}\right).
\end{aligned}
\tag{10}
$$

where we put $\vartheta=\vartheta_0+u/\sqrt{n}$.

For the BE we change the variable $\theta=\vartheta_0+u/\sqrt{n}\equiv\vartheta_u$

$$
\tilde\vartheta_n=\frac{\int_\alpha^\beta\theta p\left(\theta\right)L\left(\theta,X^n\right)\mathrm{d}\,\theta}{\int_\alpha^\beta p\left(\theta\right)L\left(\theta,X^n\right)\mathrm{d}\,\theta}=\vartheta_0+\frac{1}{\sqrt{n}}\frac{\int_{\mathbb{U}_n} up\left(\vartheta_u\right)L\left(\vartheta_u,X^n\right)\mathrm{d}\,u}{\int_{\mathbb{U}_n} p\left(\vartheta_u\right)L\left(\vartheta_u,X^n\right)\mathrm{d}\,u},
$$

Then using the convergence $p\left(\vartheta_u\right)\rightarrow p\left(\vartheta_0\right)$ (according to [5]) we can write

$$
\mathbf{P}_{\vartheta_0}\left\{\sqrt{n}\left(\tilde\vartheta_n-\vartheta_0\right)<x\right\}=\mathbf{P}\left\{\frac{\int_{\mathbb{U}_n} u\,p\left(\vartheta_u\right)Z_n\left(u\right)\,\mathrm{d}\,u}{\int_{\mathbb{U}_n} p\left(\vartheta_u\right)Z_n\left(u\right)\,\mathrm{d}\,u}<x\right\}
$$

$$
\longrightarrow\mathbf{P}\left\{\frac{\int_R u\,Z\left(u\right)\,\mathrm{d}\,u}{\int_R Z\left(u\right)\,\mathrm{d}\,u}<x\right\}=\mathbf{P}\left(\frac{\zeta\left(\vartheta_0\right)}{\mathrm{I}\left(\vartheta_0\right)}<x\right)\tag{11}
$$

because the elementary calculus yield the equality

$$
\int_R u\,Z\left(u\right)\,\mathrm{d}\,u=\int_R u\,e^{u\zeta(\vartheta_0)-\frac{u^2}{2}\mathrm{I}\left(\vartheta_0\right)}\,\mathrm{d}\,u=\frac{\zeta\left(\vartheta_0\right)}{\mathrm{I}\left(\vartheta_0\right)}\int_R Z\left(u\right)\,\mathrm{d}\,u.
$$

Hence

$$\sqrt{n}\left(\tilde\vartheta_n-\vartheta_0\right)\Longrightarrow \mathcal{N}\left(0,\frac{1}{\mathrm{I}\left(\vartheta_0\right)}\right).$$

Moreover, by Theorem 2,

$$\left(n\mathrm{I}\left(\vartheta_0\right)\right)^{\frac{p}{2}}\mathbf{E}_{\vartheta_0}\left|\hat\vartheta_n-\vartheta_0\right|^p\longrightarrow \mathbf{E}\left|\zeta\right|^p,$$

$$\left(n\mathrm{I}\left(\vartheta_0\right)\right)^{\frac{p}{2}}\mathbf{E}_{\vartheta_0}\left|\tilde\vartheta_n-\vartheta_0\right|^p\longrightarrow \mathbf{E}\left|\zeta\right|^p,$$

where $\zeta\sim\mathcal{N}\left(0,1\right)$.

3. Misspecified model

Suppose now that the parametric family $\{\lambda_\vartheta,\vartheta\in\Theta\}$ does not correspond to the observed process X^n, i.e., the value $\vartheta_0\in\Theta$ such that $\lambda_*=\lambda_{\vartheta_0}$ does not exist, but the statistician nevertheless uses this model to estimate the parameter ϑ (*no true model* case), i.e., he (or she) calculates the likelihood ratio function by (5), where X^n are observations of the Poisson process of intensity function $\lambda_*\left(\cdot\right)$. It can be shown that the MLE and BE converge to the value

$$\vartheta_*=\arg\inf_{\theta\in\Theta}\int_0^\tau\left[\frac{\lambda\left(\vartheta,t\right)}{\lambda_*\left(t\right)}-1-\ln\frac{\lambda\left(\vartheta,t\right)}{\lambda_*\left(t\right)}\right]\lambda_*\left(t\right)\,\mathrm{d}t, \tag{12}$$

which minimizes the Kullback–Liebler distance between the measure $\mathbf{P}_*$, which corresponds to the observed process with intensity λ_* and the parametric family $\{\mathbf{P}_\vartheta,\vartheta\in\Theta\}$. Note that if $\lambda_*\left(t\right)=\lambda\left(\vartheta_0,t\right)$, $0\le t\le\tau$, then $\vartheta_*=\vartheta_0$, i.e., the both estimators are consistent.

Moreover if ϑ_* is an interior point of the set Θ, then these estimators are asymptotically normal:

$$\sqrt{n}\left(\hat\vartheta_n-\vartheta_*\right)\Longrightarrow\mathcal{N}\left(0,D_*^2\right),\quad \sqrt{n}\left(\tilde\vartheta_n-\vartheta_*\right)\Longrightarrow\mathcal{N}\left(0,D_*^2\right).$$

Here $D_*^2=d_*^2\,\mathrm{I}_*^{-2}$ with

$$d_*^2=\int_0^\tau\frac{\dot\lambda\left(\vartheta_*,t\right)^2}{\lambda\left(\vartheta_*,t\right)^2}\lambda_*\left(t\right)\,\mathrm{d}t,\quad \mathrm{I}_*=d_*^2+\int_0^\tau\ddot\lambda\left(\vartheta_*,t\right)\left[1-\frac{\lambda_*\left(t\right)}{\lambda\left(\vartheta_*,t\right)}\right]\mathrm{d}t.$$

Note that in this case the pseudo-LR function $Z_n(u)$ constructed on the base of the wrong parametric model has a different limit

$$Z_n(u) = \frac{L\left(\vartheta_* + u/\sqrt{n}, X^T\right)}{L\left(\vartheta_*, X^T\right)} \Longrightarrow Z(u) = \exp\left\{u\,\zeta_* - \frac{u^2}{2}\,\mathrm{I}_*\right\}$$

where $\zeta_* \sim \mathcal{N}\left(0, d_*^2\right)$. The details of this proof can be found in [6]. See as well Yoshida and Hayashi [9].

We are interested here by a different problem. The intensity of observed process $\lambda_*(t)$ can be written as contaminated version of the parametric model $\lambda_*(t) = \lambda(\vartheta_0, t) + h(t)$, $0 \leq t \leq T$, where $h(\cdot)$ (contamination) is unknown function. Hence $\vartheta_* = \vartheta_*(h)$ is the point of the minimum of the Kullback–Leibler distance (12). We can put the following question:

$$\textbf{when}\quad \vartheta_* = \vartheta_0?$$

i.e., when nevertheless the MLE and BE are consistent?

We consider two situations. The first one (**smooth**), when the support $\mathbb{A} \subset [0,\tau]$ of the function $h(\cdot)$ is known and $\mathbb{A}^c = [0,\tau] \setminus \mathbb{A} \neq \emptyset$. We can modify the likelihood ratio and write it as

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

i.e., we exclude the observations on $\mathbb{A}$ and define the MLE $\hat{\vartheta}_n$ and BE $\tilde{\vartheta}_n$ with the help of this function (we call them pseudo-MLE and pseudo-BE). Then we have to check if the set of intensity functions $\{\lambda(\vartheta, t), t \in \mathbb{A}^c, \vartheta \in \Theta\}$ satisfies the correspondingly modified regularity conditions. For example, the Fisher information

$$\mathrm{I}_*(\vartheta) = \int_{\mathbb{A}^c} \frac{\dot{\lambda}(\vartheta, t)^2}{\lambda(\vartheta, t)}\ \mathrm{d}\,t > 0$$

and the condition of identifiability: for any $\nu > 0$

$$\inf_{|\theta - \vartheta_0| > \nu} \int_{\mathbb{A}^c} \left[\sqrt{\lambda(\vartheta, t)} - \sqrt{\lambda(\vartheta_0, t)}\right]^2\ \mathrm{d}\,t > 0.$$

If these conditions are fulfilled, then the estimators $\hat{\vartheta}_n$ and $\tilde{\vartheta}_n$ converge to the true value (are consistent) and are asymptotically normal.

Discontinuous intensity functions. Suppose that intensity of the observed process is

$$\lambda_*(t) = [g_1(t) + h_1(t)]\ 1_{\{t<\vartheta_0\}} + [g_2(t) + h_2(t)]\ 1_{\{t\geq\vartheta_0\}},$$

where $g_1(\cdot) < g_2(\cdot)$ are known positive functions and the functions $h_1(\cdot)$, $h_2(\cdot)$ are unknown. We have to estimate the time ϑ_0 of switching of intensity function (change point estimation problem). The MLE and BE are constructed on the base of the model with

$$\lambda(\vartheta, t) = g_1(t)\ 1_{\{t<\vartheta\}} + g_2(t)\ 1_{\{t\geq\vartheta\}}, \quad 0 \leq t \leq \tau,$$

with the likelihood ratio function (2), i.e. as if $h_i(t) \equiv 0$, but the observations X^n used in (2) contain, of course, $h_i(\cdot)$. The Kullback–Leibler distance (12) for $\vartheta < \vartheta_0$ is

$$\begin{aligned} J_{KL}(\vartheta) = &\int_0^{\vartheta} \left[\frac{g_1(t)}{g_1(t)+h_1(t)} - 1 - \ln\frac{g_1(t)}{g_1(t)+h_1(t)}\right][g_1(t)+h_1(t)]\,\mathrm{d}\,t \\ &+ \int_{\vartheta}^{\vartheta_0} \left[\frac{g_1(t)}{g_2(t)+h_1(t)} - 1 - \ln\frac{g_2(t)}{g_1(t)+h_1(t)}\right][g_1(t)+h_1(t)]\,\mathrm{d}\,t \\ &+ \int_{\vartheta_0}^{\tau} \left[\frac{g_2(t)}{g_2(t)+h_2(t)} - 1 - \ln\frac{g_2(t)}{g_2(t)+h_2(t)}\right][g_2(t)+h_2(t)]\,\mathrm{d}\,t \end{aligned}$$

and the similar expression we have for $\vartheta > \vartheta_0$. It is easy to see that if the functions $h_i(\cdot)$ satisfy the following condition

$$0 < g_1(t) + h_1(t) < \frac{g_2(t) - g_1(t)}{\ln\frac{g_2(t)}{g_1(t)}} < g_2(t) + h_2(t), \tag{13}$$

then

$$\left.\frac{\mathrm{d}\,J_{KL}(\vartheta)}{\mathrm{d}\,\vartheta}\right|_{\vartheta<\vartheta_0} < 0, \quad \text{and} \quad \left.\frac{\mathrm{d}J_{KL}(\vartheta)}{\mathrm{d}\,\vartheta}\right|_{\vartheta>\vartheta_0} > 0.$$

Hence the minimum of this function is reached at the point $\vartheta_* = \vartheta_0$ and this provides the consistency of the estimators $\hat\vartheta_n$ and $\tilde\vartheta_n$. If we denote

$x = g_2(t)/g_1(t)$, $h_i = h_i(t)/g_1(t)$, then we obtain the following *regions of consistency* for h_i

$$h_1 < \frac{x-1}{\ln x} - 1, \quad h_2 > \frac{x-1}{\ln x} - x.$$

It is important to note that the values of h_i can be sufficiently large.

It can be shown that the rate of convergence is essentially better than in regular case, and $n\left(\hat{\vartheta}_n - \vartheta_0\right)$ converges in distribution to some random variable (see similar results in [3, 2]).

4. Nonidentifiable model

Suppose that we have the same model for the different values of the parameter, i.e., $\lambda(\vartheta_1, t) = \lambda(\vartheta_l, t)$, $l = 2, \ldots, k$, where $\vartheta_l \neq \vartheta_i$, $l \neq i$ and $\vartheta_l, \vartheta_i \in \Theta$ (too many true models). It is well-known that the MLE converges to the set $\{\vartheta_1, \ldots, \vartheta_k\}$ of all *true values.*

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

$$\varrho_{li} = \mathbf{E}(\zeta_l \zeta_i) = (\mathrm{I}(\vartheta_l)\,\mathrm{I}(\vartheta_i))^{-1/2} \int_0^\tau \frac{\dot{\lambda}(\vartheta_l, t)\,\dot{\lambda}(\vartheta_i, t)}{\lambda(\vartheta_i, t)} \mathrm{d}t$$

where the Fisher informations

$$\mathrm{I}(\vartheta_l) = \int_0^\tau \frac{\dot{\lambda}(\vartheta_l, t)^2}{\lambda(\vartheta_l, t)} \mathrm{d}t > 0, \quad l = 1, 2, \ldots, k.$$

Define two random variables: discrete and continuous $\hat{\vartheta} = \sum_{l=1}^{k} \vartheta_l \, 1_{\{\mathbb{H}_l\}}$ and $\tilde{\vartheta} = \sum_{l=1}^{k} \vartheta_l \, Q_l$, where (we suppose that $\mathbf{P}\{|\zeta_l| = |\zeta_i|\} = 0$)

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

It can be shown that the MLE and BE have the following limits:

$$\hat{\vartheta}_n \Longrightarrow \hat{\vartheta}, \quad \tilde{\vartheta}_n \Longrightarrow \tilde{\vartheta}.$$

Moreover

$$\sqrt{n}\left(\hat\vartheta_n-\hat\theta_n\right)\Longrightarrow\hat\zeta,\quad \sqrt{n}\left(\tilde\vartheta_n-\tilde\theta_n\right)\Longrightarrow\tilde\zeta,$$

where $\hat\theta_n,\tilde\theta_n$ are close to $\hat\vartheta,\tilde\vartheta$ random variables and

$$\hat\zeta=\sum_{l=1}^{k}\zeta_l\ \mathrm{I}\ (\vartheta_l)^{-1/2}\ 1_{\{\mathbb{H}_l\}}.$$

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

$$\mathbf{Z}_n(\mathbf{u})=\left(Z_n^{(1)}(u_1),\ldots,Z_n^{(k)}(u_k)\right),\quad Z_n^{(l)}(u_l)=\frac{L\left(\vartheta_l+\frac{u_l}{\sqrt{n}},X^T\right)}{L\left(\vartheta_l,X^T\right)}$$

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

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

(see details in [6, Section 4.2]).

Example. Let $\vartheta\in(0,3)$ and the intensity function

$$\lambda(\vartheta,t)=\left(\vartheta^3-3\vartheta^2+2\vartheta\right)\,t+(2\vartheta-3)\ t^2+1,\quad 0\le t\le 1$$

then $\lambda(1,t)=t^2+1$ and $\lambda(2,t)=t^2+1$. Hence we have

$$\hat\vartheta_n\Rightarrow\hat\vartheta=1_{\{|\zeta_1|>|\zeta_2|\}}+2_{\{|\zeta_1|\le|\zeta_2|\}}$$

and so on.

5. Null Fisher information

Suppose that $\mathrm{I}\ (\vartheta_0)=0$. This means that at one point ϑ_0 (true value) the function $\dot\lambda(\vartheta_0,t)=0$ for all $t\in[0,\tau]$. Moreover, suppose that the function $\lambda(\vartheta,t)$ is 4 times continuously differentiable w.r.t. ϑ with $\ddot\lambda(\vartheta_0,t)=0$ and

$$\mathrm{I}_3(\vartheta_0)=\int_0^\tau\frac{\dddot\lambda(\vartheta_0,t)^2}{(3!)^2\ \lambda(\vartheta_0,t)}\,\mathrm{d}t>0.$$

Introduce random variable $\zeta(\vartheta_0) \sim \mathcal{N}(0, \mathrm{I}_3(\vartheta_0))$. Then we have:

$$n^{1/6}\left(\hat{\vartheta}_n - \vartheta_0\right) \Longrightarrow \hat{u} = \left(\frac{\zeta(\vartheta_0)}{\mathrm{I}_3(\vartheta_0)}\right)^{1/3}.$$

The proof is based on the weak convergence

$$Z_n(u) = \frac{L\left(\vartheta_0 + \frac{u}{n^{1/6}}, X^n\right)}{L(\vartheta_0, X^n)} \Longrightarrow Z(u) = \exp\left\{u^3\zeta(\vartheta_0) - \frac{u^6}{2}\mathrm{I}_3(\vartheta_0)\right\}.$$

We have to check the conditions of the Theorem 2. Particularly estimates (8) are replaced by the estimates

$$c\left|\vartheta_2 - \vartheta_1\right|^6 \leq \int_0^\tau \left[\sqrt{\lambda(\vartheta_2, t)} - \sqrt{\lambda(\vartheta_1, t)}\right]^2 \mathrm{d}\, t \leq C\left|\vartheta_2 - \vartheta_1\right|^6.$$

The limit expression for the bayesian estimator is more complicated.

Example. Let

$$\lambda(\vartheta, t) = \vartheta \sin^2(\vartheta t) + 2, \quad 0 \leq t \leq 1, \quad \vartheta \in (-1, 1)$$

then $\mathrm{I}_l(0) = 0, l = 1, 2$ and $\mathrm{I}_3(0) = \frac{1}{10}$. Hence

$$n^{1/6}\left(\hat{\vartheta}_n - 0\right) \Longrightarrow (10)^{1/6}\zeta^{1/3}, \qquad \zeta \sim \mathcal{N}(0, 1).$$

6. Discontinuous Fisher information

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

$$\mathbf{E}(\zeta_-\zeta_+) = \left(\mathrm{I}\left(\vartheta_0^-\right)\mathrm{I}\left(\vartheta_0^+\right)\right)^{-1/2} \int_0^\tau \frac{\dot{\lambda}\left(\vartheta_0^-, t\right)\dot{\lambda}\left(\vartheta_0^+, t\right)}{\lambda(\vartheta_0, t)} \mathrm{d}\, t.$$

Then with the help of Theorem 2 it can be shown that the MLE is consistent, and $\sqrt{n}\left(\hat\vartheta_n-\vartheta_0\right)\Rightarrow\hat\zeta$ but its limit distribution is a mixture of three random variables:

$$\hat\zeta=\begin{cases}\dfrac{\zeta_-}{\mathrm{I}\left(\vartheta_0^-\right)^{1/2}} & \text{if } \zeta_-<0,\ \ \zeta_+<0 \ \text{ or }\ \zeta_-<0,\ \ \zeta_+>0 \ \text{ and }\ |\zeta_-|>|\zeta_+| \\ 0 & \text{if } \zeta_->0,\ \ \zeta_+<0, \\ \dfrac{\zeta_+}{\mathrm{I}\left(\vartheta_0^+\right)^{1/2}} & \text{if } \zeta_->0,\ \ \zeta_+>0 \ \text{ or }\ \zeta_-<0,\ \ \zeta_+>0 \ \text{ and }\ |\zeta_-|<|\zeta_+|\end{cases}$$

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

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

We see that there is an atom at the point 0. This form of the limit likelihood ratio $Z\left(\cdot\right)$ provides as well the limit distribution of the bayesian estimates

$$\sqrt{n}\left(\tilde\vartheta_n-\vartheta_0\right)\Longrightarrow\tilde u=\frac{\int_{\mathcal{R}} uZ\left(u\right)\mathrm{d}\,u}{\int_{\mathcal{R}} Z\left(u\right)\mathrm{d}\,u}.$$

Example. Suppose that $\vartheta\in(0,2)$ and

$$\lambda\left(\vartheta,t\right)=\left(\vartheta-1\right)\left[3t\,1_{\{\vartheta<1\}}+5t^2\,1_{\{\vartheta\geq1\}}\right]+15,\quad 0\leq t\leq 1,$$

then $\mathrm{I}\,\left(1-\right)=\frac{1}{5}$ and $\mathrm{I}\,\left(1+\right)=\frac{1}{3}$ and the MLE has the mentioned above limit distribution.

7. Border of the parameter set

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

$$\sqrt{n}\left(\hat\vartheta_n-\alpha\right)\Longrightarrow\frac{\zeta\left(\alpha\right)}{\mathrm{I}\,\left(\alpha\right)}\,1_{\{\zeta\geq0\}},\quad \zeta\left(\alpha\right)\sim\mathcal{N}\left(0,\mathrm{I}\,\left(\alpha\right)\right).$$

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

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

For the BE we have the limit

$$\sqrt{n}\left(\tilde{\vartheta}_n - \alpha\right) \Longrightarrow \tilde{u} = \frac{\int_0^\infty uZ(u)\,\mathrm{d}\,u}{\int_0^\infty Z(u)\,\mathrm{d}\,u}$$

$$= \frac{1}{\sqrt{\mathrm{I}(\alpha)}}\left(\zeta_* + \left(\int_{-\zeta_*}^{\infty} e^{-\frac{1}{2}\left(u^2-\zeta_*^2\right)}\,\mathrm{d}\,u\right)^{-1}\right),$$

where $\zeta_* \sim \mathcal{N}(0,1)$.

To prove these results we have to check the conditions of the Theorem 2 for the likelihood ratio process

$$Z_n(u) = \frac{L\left(\frac{u}{\sqrt{n}}, X^n\right)}{L(0, X^n)}, \quad u \in \mathbb{U}_n = \left[0, \beta\sqrt{n}\right]$$

with the corresponding limit process.

8. Cusp type singularity

Let us suppose that the observed process has intensity function

$$\lambda(\vartheta, t) = a\left|t - \vartheta\right|^\kappa + \lambda_0, \quad 0 \leq t \leq T$$

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

$$Z_n(u) = \frac{L\left(\vartheta + \frac{u}{n^{1/2H}}, X^n\right)}{L(\vartheta, X^n)},$$

$$u \in \mathbb{U}_n = \left(n^{1/2H}(\alpha - \vartheta_0), n^{1/2H}(\beta - \vartheta_0)\right)$$

and the limit process

$$Z(u) = \exp\left\{\Gamma_\vartheta W^H(u) - \frac{|u|^{2H}}{2}\Gamma_\vartheta^2\right\}, \quad u \in \mathcal{R}.$$

Here $W^H(\cdot)$ is double sided fractional Brownian motion, $H = \kappa + \frac{1}{2}$ (Hurst parameter) and

$$\Gamma_\vartheta^2 = \frac{4a^2 \sin^2(2\pi\kappa)\,\mathrm{B}\,(1+\kappa, 1+\kappa)}{\lambda_0 \cos(\pi\kappa)},$$

where $\mathrm{B}\,(1+\kappa, 1+\kappa)$ is beta function.

We can check the conditions of the Theorem 2 and to show that the MLE and BE are consistent, have the following limits

$$n^{\frac{1}{2H}}\left(\hat\vartheta_n - \vartheta\right) \Longrightarrow \hat u, \quad n^{\frac{1}{2H}}\left(\tilde\vartheta_n - \vartheta\right) \Longrightarrow \tilde u,$$

where the random variables are defined by the same equations (3) and we have the corresponding convergence of moments. (For the proof see [4]).

9. Discontinuous intensity function

Let us suppose that the observed process $X^n = (X_1(\cdot), \ldots, X_n(\cdot))$, where $X_j(\cdot) = \{X_j(t), 0 \le t \le T\}$ has the intensity function $\lambda(t+\vartheta)$, $0 \le t \le T$ and the function $\lambda(y)$ is positive and continuously differentiable everywhere except at the point τ, that is $\lambda(\tau_+) - \lambda(\tau_-) = r \neq 0$. The set $\Theta = (\alpha, \beta) \subset (\tau - T, \tau)$. The likelihood ratio process (2) has discontinuous realizations and the MLE $\hat\vartheta_n$ is defined now by the following relation

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

The BE is defined as before.

The limit process $Z(u)$ for the normalized likelihood ratio

$$Z_n(u) = \frac{L\left(\vartheta + \frac{u}{n}, X^n\right)}{L(\vartheta, X^n)}, \quad \mathbb{U}_n = (n(\alpha - \vartheta_0), n(\beta - \vartheta_0))$$

is

$$Z(u) = \begin{cases} \exp\left\{\ln\frac{\lambda(\tau_+)}{\lambda(\tau_-)}\,\pi_+(u) - [\lambda(\tau_+) - \lambda(\tau_-)]\,u\right\}, & u \ge 0 \\ \exp\left\{\ln\frac{\lambda(\tau_-)}{\lambda(\tau_+)}\,\pi_-(-u) - [\lambda(\tau_+) - \lambda(\tau_-)]\,u\right\}, & u \le 0, \end{cases}$$

where $\pi_+(\cdot)$ and $\pi_-(\cdot)$ are independent Poisson processes of the intensity functions $\lambda(\tau_-)$ and $\lambda(\tau_+)$, respectively. Let us denote by $\hat u$ and $\tilde u$ the random variables defined by the equations

$$\max\left[Z(\hat u+), Z(\hat u-)\right] = \sup_{u\in\mathcal{R}} Z(u), \qquad \tilde u = \frac{\int_{\mathcal{R}} u\, Z(u)\, \mathrm{d}u}{\int_{\mathcal{R}} Z(u)\, \mathrm{d}u}.$$

Then the MLE and BE are consistent, have the following limits

$$n\left(\hat\vartheta_n - \vartheta\right) \Longrightarrow \hat u, \quad n\left(\tilde\vartheta_n - \vartheta\right) \Longrightarrow \tilde u,$$

and the convergence of all moments take place. It is shown that for all estimators we have a lower bound on the risks and the bayesian estimators are asymptotically efficient. For the proof see [6, Section 5.1].

10. Windows

Optimal windows. Suppose that we can have observations on some set $\mathbb{B}\subset[0,\tau]$ of Lebesgue measure $\mu(\mathbb{B})\leq\mu_* < \tau$ only. The family of such sets we denote as $\mathcal{F}_{\mu_*}$. Our goal is to find the best window $\mathbb{B}^*$ and estimator $\vartheta_n^* = \vartheta_n^*(\mathbb{B}^*)$ constructed by the observations X^n on this set $\mathbb{B}^*$, i.e., $X_j = \{X_j(t), t\in\mathbb{B}^*\}$. The best is understood as the minimizing the mean square error asymptotically

$$\inf_{\mathbb{B}\in\mathcal{F}_{\mu_*}} \inf_{\bar\vartheta_n} \mathbf{E}_\vartheta\left(\bar\vartheta_n(\mathbb{B}) - \vartheta\right)^2 \sim \mathbf{E}_\vartheta\left(\vartheta_n^*(\mathbb{B}^*) - \vartheta\right)^2.$$

If we fix the set $\mathbb{B}$, then we know that the MLE is asymptotically normal

$$\sqrt{n}\left(\hat\vartheta_n(\mathbb{B}) - \vartheta\right) \Longrightarrow \mathcal{N}\left(0, \mathrm{I}_{\mathbb{B}}(\vartheta)^{-1}\right), \quad \mathrm{I}_{\mathbb{B}}(\vartheta) = \int_{\mathbb{B}} \frac{\dot\lambda(\vartheta,t)^2}{\lambda(\vartheta,t)}\,\mathrm{d}t.$$

Therefore if we use the MLE then the best $\mathbb{B}^* = \mathbb{B}^*(\vartheta)$ corresponds to the solution of the following equation

$$\mathrm{I}_{\mathbb{B}^*}(\vartheta) = \sup_{\mathbb{B}\in\mathcal{F}_{\mu_*}} \mathrm{I}_{\mathbb{B}}(\vartheta).$$

To solve this equation we introduce the level sets $\mathbb{C}_{(\vartheta,r)}$ and function $\mu(\vartheta,r)$ as

$$\mathbb{C}_{(\vartheta,r)} = \left\{t:\ \frac{\dot\lambda(\vartheta,t)^2}{\lambda(\vartheta,t)} \geq r\right\}, \quad \mu(\vartheta,r) = \mu\left(\mathbb{C}_{(\vartheta,r)}\right).$$

Then we define $r_* = r_*(\vartheta)$ as solution of the equation $\mu(\vartheta,r) = \mu_*$. Now we put $\mathbb{B}^* = \mathbb{C}_{(\vartheta,r_*)}$. Of course, we are not obliged to use the MLE and moreover, this set $\mathbb{B}^*$ can not be used for construction of estimator because it depends on ϑ. Nevertheless it allows to introduce the lower bound on the risks of all couples (*set, estimator*): for any $\vartheta_0 \in \Theta$

$$\underline{\lim}_{\delta\to 0}\ \underline{\lim}_{n\to\infty}\ \inf_{\mathbb{B}\in\mathcal{F}_{\mu_*},\bar\vartheta_n}\ \sup_{|\vartheta-\vartheta_0|<\delta} n\,\mathbf{E}_\vartheta\left(\bar\vartheta_n(\mathbb{B}) - \vartheta\right)^2 \geq \mathrm{I}_{\mathbb{B}^*}(\vartheta_0)^{-1}.$$

To construct the asymptotically efficient in this sense couple we first some $\mathbb{B} \in \mathcal{F}_{\mu_*}$ and by observations $X^{\sqrt{n}} = \left\{X_j(\mathbb{B}), j = 1,\ldots, X_{[\sqrt{n}]}\right\}$ we (consistently) estimate ϑ using some estimator $\bar\vartheta_{\sqrt{n}}$. Then we introduce the observation window

$$\mathbb{B}_n^* = \left\{t:\ \frac{\dot\lambda\left(\bar\vartheta_{\sqrt{n}},t\right)^2}{\lambda\left(\bar\vartheta_{\sqrt{n}},t\right)} \geq r\left(\bar\vartheta_{\sqrt{n}},\mu_*\right)\right\}.$$

Now we construct the MLE $\hat\vartheta_{n-\sqrt{n}}$ and show that

$$\lim_{\delta\to 0}\ \lim_{n\to\infty}\ \sup_{|\vartheta-\vartheta_0|<\delta} n\,\mathbf{E}_\vartheta\left(\bar\vartheta_n(\mathbb{B}_n^*) - \vartheta\right)^2 = \mathrm{I}_{\mathbb{B}^*}(\vartheta_0)^{-1}.$$

For the conditions and proofs see Kutoyants and Spokoiny [8] or [6, Section 4.3].

Example. Let $\lambda(\vartheta,t) = [b + \vartheta\sin(\omega t)]^2$, $0 \leq t \leq \tau$, where $\tau = 2\pi/\omega$. Then the Fisher information is

$$\mathrm{I}_{\mathbb{B}}(\vartheta) = 4\int_{\mathbb{B}} [\sin(\omega t)]^2\,\mathrm{d}\,t.$$

Introduce

$$\mathbb{C}_r = \left\{t:\ 4\sin(\omega t)^2 \geq r\right\}, \quad r(\vartheta,\mu_*) = r_* = 4\,\sin^2\left(\frac{2\pi - \mu_*\omega}{4}\right),$$

where $(\mu_* < \tau)$. Then $\varphi = \arcsin\left(\frac{\sqrt{\tau_*}}{2}\right)$

$$\mathbb{B}^* = \left[\frac{\varphi}{\omega},\ \frac{\tau}{2} - \frac{\varphi}{\omega}\right] \ \cup\ \left[\frac{\tau}{2} + \frac{\varphi}{\omega},\ \tau - \frac{\varphi}{\omega}\right]$$
$$= \left[\frac{\tau - \mu_*}{4},\ \frac{\tau + \mu_*}{4}\right] \cup \left[\frac{3\tau - \mu_*}{4},\ \frac{3\tau + \mu_*}{4}\right].$$

Therefore the observations in the optimal window are

$$X^n = \left(X_1\left(\mathbb{B}^*\right), \dots, X_1\left(\mathbb{B}^*\right)\right) \quad \text{with} \quad X_j\left(\mathbb{B}^*\right) = \left\{X_j\left(t\right), t \in \mathbb{B}^*\right\}$$

and the asymptotically efficient estimator is the MLE $\hat{\vartheta}_n\left(\mathbb{B}^*\right)$.

Sufficient windows. The analysis of the proofs of the consistency of the MLE and BE in the discontinuous case (see, e.g., [6, Chapter 5]) shows that the main contribution to the likelihood ratio process is made by the observations near the jumps. Suppose that the intensity function is $\lambda\left(\vartheta, t\right) = \lambda\left(t - \vartheta\right)$, where $\vartheta \in (\alpha, \beta)$ and $0 < \alpha < \beta < \tau$. Suppose as well that the function $\lambda\left(s\right)$, $s \in (-\beta, \tau - \alpha)$ is discontinuous at some point τ_* and continuous on $[-\beta, \tau_*) \cup (\tau_*, \tau - \alpha]$. Then it is sufficient to keep the observations on the interval $\mathbb{B} = [\alpha + \tau_*, \beta + \tau_*]$ only, i.e., to use $X_j\left(\mathbb{B}\right) = \left\{X_j\left(t\right), \alpha + \tau_* \leq t \leq \beta + \tau_*\right\}$, $j = 1, \dots, n$ and the properties of the MLE and BE (consistency, limit distributions and convergence of moments) will be the same as in the case of complete observations on $[0, \tau]$.

Moreover, if we have a consistent and asymptotically normal estimator $\bar{\vartheta}_n$ of ϑ (say, an estimator of the method of moments), then we can use the first $[\sqrt{n}]$ observations for preliminary estimation by $\bar{\vartheta}_{\sqrt{n}}$ of the window as $\mathbb{B}_n = \left[\bar{\vartheta}_{\sqrt{n}} - n^{-1/8}, \bar{\vartheta}_{\sqrt{n}} + n^{-1/8}\right]$, and then to construct the MLE $\hat{\vartheta}_{n-\sqrt{n}}$ and bayesian estimator $\tilde{\vartheta}_{n-\sqrt{n}}$. Note that $n^{1/4}\left(\bar{\vartheta}_{\sqrt{n}} - \vartheta\right) \Rightarrow \mathcal{N}\left(0, \sigma^2\right)$. Hence

$$\mathbf{P}_\vartheta\left\{\left|\bar{\vartheta}_{\sqrt{n}} - \vartheta\right| > n^{-1/8}\right\} = \mathbf{P}_\vartheta\left\{n^{1/4}\left|\bar{\vartheta}_{\sqrt{n}} - \vartheta\right| > n^{1/8}\right\} \longrightarrow 0.$$

Therefore we can have consistent and asymptotically efficient estimators constructed by observations in the window of vanishing size. In regular case such effect is difficult to wait.

Example. Let $\vartheta \in (\alpha, \beta) \subset (0, \tau)$ and

$$\lambda\left(\vartheta, t\right) = 2at + b\, 1_{\{t > \vartheta\}}, \quad 0 \leq t \leq \tau.$$

Then

$$\bar{\vartheta}_{\sqrt{n}} = \tau - \frac{1}{b}\left[\hat{\Lambda}_{\sqrt{n}}(\tau) - a\tau^2\right]$$

is consistent and asymptotically normal estimator of ϑ. Then we maximize the function

$$L(\vartheta, X^n) = \exp\left\{ \sum_{j=[\sqrt{n}]+1}^{n} \int_{\bar{\vartheta}_{\sqrt{n}}-n^{-1/8}}^{\bar{\vartheta}_{\sqrt{n}}+n^{-1/8}} \ln\lambda(\vartheta,t)\,\mathrm{d}X_j(t) - \left(n - [\sqrt{n}]\right) \int_{\bar{\vartheta}_{\sqrt{n}}-n^{-1/8}}^{\bar{\vartheta}_{\sqrt{n}}+n^{-1/8}} [\lambda(\vartheta,t) - 1]\,\mathrm{d}t \right\}$$

and construct the MLE. Note that the random variable $\bar{\vartheta}_{\sqrt{n}}$ is independent on $X_j, j = [\sqrt{n}] + 1, \dots, n$.

11. Rates of convergence

It is interesting to note that if we observe a periodic Poisson process $X^n = \{X(t), 0 \le t \le n\}$ with the intensity functions $\lambda(\vartheta + t)$ or $\lambda(\vartheta t)$, where $\lambda(\cdot)$ is periodic smooth function (phase and frequency modulations in the optical telecommunication theory), then we have $(n \to \infty)$

$$\mathbf{E}_\vartheta\left(\hat{\vartheta}_n - \vartheta\right)^2 \sim \frac{C}{n}, \quad \mathbf{E}_\vartheta\left(\hat{\vartheta}_n - \vartheta\right)^2 \sim \frac{C}{n^3},$$

respectively. If $\lambda(t)$ is discontinuous function then for the mentioned two cases of modulations we have the different rates

$$\mathbf{E}_\vartheta\left(\hat{\vartheta}_n - \vartheta\right)^2 \sim \frac{C}{n^2}, \quad \mathbf{E}_\vartheta\left(\hat{\vartheta}_n - \vartheta\right)^2 \sim \frac{C}{n^4}.$$

For the proofs, see [6].

Therefore it is natural to put the following question: *what is the maximal possible rate of convergence of the mean square error to zero?* Suppose that we can choose any function $\lambda(\vartheta,t), \vartheta \in [0,1], t \ge 0$ satisfying the only condition

$$0 \le \lambda(\vartheta,t) \le L_*,$$

where $L_* > 0$ is some given constant. We denote the class of such functions as $\mathcal{F}(L_*)$. It can be shown that

$$\inf_{\lambda(\cdot)\in\mathcal{F}(L_*)} \inf_{\bar\vartheta_n} \sup_{\vartheta\in[0,1]} \mathbf{E}_{\vartheta,\lambda} \left|\bar\vartheta_n - \vartheta\right|^2 = e^{-\frac{nL_*}{6}(1+o(1))},$$

i.e., the best rate is exponential. To prove this equality we need to prove two results. The first one is the lower bound for all $\lambda(\cdot) \in \mathcal{F}(L_*)$ and all estimators $\bar\vartheta_n$

$$\sup_{\vartheta\in[0,1]} \mathbf{E}_{\vartheta,\lambda} \left|\bar\vartheta_n - \vartheta\right|^2 \geq e^{-\frac{nL_*}{6}(1+o(1))},$$

and the second is to construct an intensity function $\lambda_*(\cdot) \in \mathcal{F}(L_*)$ and an estimator ϑ_n^* such that

$$\sup_{\vartheta\in[0,1]} \mathbf{E}_{\vartheta,\lambda_*} \left|\vartheta_n^* - \vartheta\right|^2 = e^{-\frac{nL_*}{6}(1+o(1))}.$$

For the proof, see [1].

References

1. M. V. Burnashev, Yu. A. Kutoyants, *On minimal α-mean error parameter transmission over Poisson channel.* — IEEE Transactions on Information Theory **47**, No. 6 (2001), 2505–2515.
2. A. Dabye, C. Farinetto, Yu. A. Kutoyants, *On Bayesian estimators in misspecified change-point problem for Poisson processes.* — Stat. Prob. Lett. **61**, No. 1 (2003), 17–30.
3. A. Dabye, Yu. A. Kutoyants, *On misspecified change-point problem for Poisson processes.* — J. Appl. Prob. **38A** (2001), 705–709.
4. S. Dachian, *Estimation of cusp location by Poisson observations.* — SISP **6**, No. 1 (2003), 1–14.
5. I. A. Ibragimov, R. Z. Khasminskii, *Statistical Estimation.* Springer, New York (1981).
6. Yu. A. Kutoyants, *Statistical Inference for Spatial Poisson Processes.* Springer, New York (1998).
7. Yu. A. Kutoyants, *On regular and singular estimation for ergodic diffusion.* — J. Japan Statist. Soc. Celebration Volume for H. Akaike **38** (2008), 51–63.
8. Yu. A. Kutoyants, V. Spokoiny, *Optimal choice of observation window for Poisson observations.* — Statist. Probab. Lett. **44** (1999), 291–298.
9. N. Yoshida, T. Hayashi, *On the robust estimation in Poisson processes with periodic intensities.* — AISM **42**, No. 3 (1990), 489–507.

Laboratoire de Statisque et Processus,
Université du Maine
E-mail: kutoyants@univ-lemans.fr

Поступило 10 декабря 2008 г.