Method of Moments Estimators and Multi-step MLE for Poisson Processes

Ali S. Dabye

Laboratoire d'Études et de Recherches en Statistique

et Développement, Université Gaston Berger, Sénégal

(joint work with Yu. A. Kutoyants and A. A. Gounoung)

SAPS XI, 17.07.2017

Introduction

This work is devoted to the problem of parameter estimation in the case of continuous time observations of inhomogeneous Poisson processes. The Poisson process is one of the main models in the description of the series of events in real applied problems in optical telecommunications, biology, physics, financial mathematics etc. Note that the intensity function entirely identifies the process and therefore the statistical inference is concerned this function only. We suppose that the intensity function of the observed Poisson process is a known function which depends on some unknown finite-dimensional parameter. We consider the problem of this parameter estimation in the asymptotic of large samples. We have to note that the estimation theory (parametric and non parametric) is well developed and there exists a large number of publications devoted to this class of problems.

The method of moments and One-step estimation procedure in the case of i.i.d. observations are well known too. Our goal is to apply the method of moments to the estimation of the parameters of inhomogeneous Poisson processes and to present a version of One-step and Multi-step procedures with the help of some preliminary estimators obtain on the small learning interval.

We are given n independent observations $X^n = (X_1, \ldots, X_n)$ of the Poisson processes $X_j = (X_j(t), t \in \mathbb{T})$ with the same intensity function $\lambda(\vartheta, t), t \in \mathbb{T}$. Here $\mathbb{T}$ is an interval of observations. It can be finite, say, $\mathbb{T} = [0, T]$ or infinite $\mathbb{T} = [0, \infty)$, $\mathbb{T} = (-\infty, \infty)$. The unknown parameter $\vartheta \in \Theta$, where the set Θ is an open, convex and bounded subset of $\mathcal{R}^d$.

Recall that the increments of the Poisson process (X_j is a counting process) on disjoint intervals are independent and for any $k = 0, 1, 2, \ldots$ and $t_1 < t_2$

$$\mathbf{P}_\vartheta\Big(X_j\left(t_2\right) - X_j\left(t_1\right) = k\Big) = \frac{\left[\int_{t_1}^{t_2} \lambda\left(\vartheta, s\right) \mathrm{d}s\right]^k}{k!} \exp\left\{\int_{t_1}^{t_2} \lambda\left(\vartheta, s\right) \mathrm{d}s\right\}.$$

Recall that

$$\mathbf{E}_\vartheta X_j\left(t\right) = \Lambda\left(\vartheta, t\right) = \int^t \lambda\left(\vartheta, s\right) \mathrm{d}s, \qquad t \in \mathbb{T}.$$

We have to estimate the true value of $\vartheta = \vartheta_0$ by the observations X^n and to describe the asymptotic ($n \to \infty$) properties of estimators. It is known that under regularity conditions the method of moments estimators in the case of i.i.d. observations of the random variables are consistent and asymptotically normal.

Our goal is to introduce the method of moments estimators (MME) in the case of observations of inhomogeneous Poisson processes. This method of estimation was introduced by Karl Pearson in 1894 in the case of observations of the i.i.d. random variables. Then it was extended to many other models of observations and widely used in applied problems. It seems that till now this method was not yet used for the estimation of the parameters of inhomogeneous Poisson processes. The maximum likelihood estimator (MLE) $\hat{\vartheta}_n$ (under regularity conditions) is consistent, asymptotically normal

$$\sqrt{n}\left(\hat{\vartheta}_n-\vartheta_0\right)\Longrightarrow\mathcal{N}\left(0,\mathbb{I}\left(\vartheta_0\right)^{-1}\right)$$

and asymptotically efficient (see, Kutoyants (1984), [9]). Here $\mathbb{I}\left(\vartheta_0\right)$ is the Fisher information matrix

$$\mathbb{I}\left(\vartheta_0\right)=\int_{\mathbb{T}}\dot{\lambda}\left(\vartheta_0,t\right)\dot{\lambda}\left(\vartheta_0,t\right)^{\tau}\lambda\left(\vartheta_0,t\right)^{-1}\mathrm{d}t.$$

Here and in the sequel dot means derivation with respect to (w.r.t.) ϑ and A^τ means the transpose of the vector (or matrix) A.

Recall that in the regular case the following lower bound (called Hajek-Le Cam) holds: for any estimator $\bar\vartheta_n$ and any $\vartheta_0 \in \Theta$ we have

$$\lim_{\nu\to 0} \varliminf_{n\to\infty} \sup_{|\vartheta-\vartheta_0|<\nu} n\mathbf{E}_\vartheta \left|\mathbb{I}\left(\vartheta_0\right)^{1/2}\left(\bar\vartheta_n-\vartheta\right)\right|^2 \geq d.$$

This bound allows us to define the asymptotically efficient estimator $\check\vartheta_n$ as estimator satisfying the equality

$$\lim_{\nu\to 0} \varliminf_{n\to\infty} \sup_{|\vartheta-\vartheta_0|<\nu} n\mathbf{E}_\vartheta \left|\mathbb{I}\left(\vartheta_0\right)^{1/2}\left(\check\vartheta_n-\vartheta\right)\right|^2 = d$$

for all $\vartheta_0 \in \Theta$.

If we verify that the moments of the MLE converge uniformly on ϑ then this proves the asymptotic efficiency of the MLE.

In the present work we introduce two classes of estimators. The first one is the class of the method of moments estimators and the second class is the Multi-step MLEs.

We show that the MMEs for many models of inhomogeneous Poisson processes are easy to calculate, but these estimators as usual are not asymptotically efficient. The MLEs are asymptotically efficient, but their calculation is often a difficult problem.

The main result of this work is the introduction of the Multi-step MLEs which are easy to calculate and which are asymptotically efficient. These Multi-step MLEs are calculated in several steps. For example, One-step MLE is calculated as follows. First we fix the *learning observations* $X^N = (X_1, \ldots, X_N)$, where $N = \left[n^\delta\right]$ with $\delta \in \left(\frac{1}{2}, 1\right)$. Here $[a]$ is the entier part of a. By the observations X^N we construct the MME ϑ_N^* and then with the help of it we introduce the One-step MLE by the equality

$$\vartheta_n^\star = \vartheta_N^* + \frac{1}{n}\mathbb{I}\left(\vartheta_N^*\right)^{-1} \sum_{j=N+1}^{n} \int_{\mathbb{T}} \frac{\dot\lambda\left(\vartheta_N^*, t\right)}{\lambda\left(\vartheta_N^*, t\right)} \left[\mathrm{d}X_j\left(t\right) - \lambda\left(\vartheta_N^*, t\right)\mathrm{d}t\right].$$

It is shown that this estimator is asymptotically normal

$$\sqrt{n}\left(\vartheta_n^\star - \vartheta_0\right) \Longrightarrow \mathcal{N}\left(0, \mathbb{I}\left(\vartheta_0\right)^{-1}\right)$$

and is asymptotically efficient.

Method of Moments for Poisson processes

Let us construct the method of moments estimator in the case of observations of inhomogeneous Poisson process. We have n independent observations $X^n = (X_1, ..., X_n)$ of the Poisson processes $X_j = (X_j(t), t \in \mathbb{T})$ with the intensity function $(\lambda\left(\theta, t\right), t \in \mathbb{T})$. The unknown parameter $\theta \in \Theta \subset \mathbb{R}^d$. Here Θ is an open, convex, bounded set.

In the construction of the method of moments estimator we follow the same way as in the construction of MME in the case of i.i.d. random variables.

Introduce the vector-function $\mathbf{g}(s) = (g_1(s), ..., g_d(s)), t \in \mathbb{T}$ and the vector of integrals $\mathbf{I}^{(d)} = (I_1, \ldots, I_d)$, where

$$I_l = \int_{\mathbb{T}} g_l(s)\, \mathrm{d}X_1(s), \qquad l = 1, \ldots, d.$$

We have

$$\mathbb{E}_\theta \mathbf{I}^{(d)} = \int_{\mathbb{T}} \mathbf{g}(s)\, \lambda(\theta, s)\, \mathrm{d}s.$$

Let us denote $\mathbf{M}(\vartheta) = \mathbb{E}_\theta \mathbf{I}^{(d)}$ and suppose that the function $\mathbf{g}(\cdot)$ is such that the equation $\mathbf{M}(\vartheta) = \mathbf{a}$ for all $\vartheta \in \Theta$ has a unique solution $\vartheta = \mathbf{M}^{-1}(\mathbf{a}) = \mathbf{H}(\mathbf{a})$. Here $\mathbf{H}(\mathbf{a})$ is the inverse function for $\mathbf{M}(\cdot)$.

The method of moments estimator ϑ_n^* is defined by the equation

$$\vartheta_n^* = \mathbf{H}\left(\mathbf{a}_n\right),$$

where

$$\mathbf{a}_n = \frac{1}{n}\sum_{j=1}^{n}\int_{\mathbb{T}} \mathbf{g}\left(s\right) \mathrm{d}X_j\left(s\right)$$

Introduce the *Regularity conditions* $\mathcal{R}_0$ *:*

- *For any* $\nu > 0$ *and any* $\vartheta_0 \in \Theta$

$$\inf_{|\vartheta-\vartheta_0|>\nu} \left|\mathbf{M}\left(\vartheta\right) - \mathbf{M}\left(\vartheta_0\right)\right| > 0.$$

- *The vector-function* $\mathbf{H}\left(\cdot\right)$ *is continuously differentiable.*

Introduce the matrix

$$\mathbb{D}\left(\vartheta\right)=\frac{\partial\mathbf{H}\left(\vartheta\right)}{\partial\vartheta}\mathbb{G}\left(\vartheta\right)\frac{\partial\mathbf{H}\left(\vartheta\right)}{\partial\vartheta}^{T}.$$

Here the matrices

$$\left(\frac{\partial\mathbf{H}\left(\vartheta\right)}{\partial\vartheta}\right)_{lk}=\frac{\partial H_l\left(\vartheta\right)}{\partial\vartheta_k},\qquad \mathbb{G}\left(\vartheta\right)_{l,k}=\int_{\mathbb{T}}g_l\left(s\right)g_k\left(s\right)\lambda\left(\vartheta,s\right)\mathrm{d}s.$$

Theorem 1 *Suppose that the vector-function* $\mathbf{g}\left(\cdot\right)$ *is such, that the regularity conditions* $\mathcal{R}_0$ *are fulfilled. Then the MME* ϑ_n^* *is consistent and asymptotically normal*

$$\sqrt{n}\left(\vartheta_n^*-\vartheta_0\right)\Longrightarrow\mathcal{N}\left(0,\mathbb{D}\left(\vartheta_0\right)\right). \qquad (1)$$

Proof. By the Law of Large Numbers

$$a_{l,n} = \frac{1}{n}\sum_{j=1}^{n}\int_{\mathbb{T}} g_l\left(s\right)\mathrm{d}X_j\left(s\right) \longrightarrow \int_{\mathbb{T}} g_l\left(s\right)\lambda\left(\vartheta_0, s\right)\mathrm{d}s, \qquad l = 1, ..., d$$

and hence by the well-known Continuous Mapping Theorem $\mathbf{H}\left(\mathbf{a}_n\right) \longrightarrow \mathbf{H}\left(\mathbf{a}_0\right) = \vartheta_0$. Here we put $\mathbf{a}_0 = \mathbf{M}\left(\vartheta_0\right)$. To show asymptotic normality we write

$$\sqrt{n}\left(\vartheta_n^* - \vartheta_0\right) = \sqrt{n}\left(\mathbf{H}\left(\mathbf{a}_n\right) - \mathbf{H}\left(\mathbf{a}_0\right)\right) = \sqrt{n}\left(\mathbf{H}\left(\mathbf{a}_0 + b_n\eta_n\right) - \mathbf{H}\left(\mathbf{a}_0\right)\right)$$

where $b_n = n^{-1/2}$ and the vector

$$\eta_n = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\int_{\mathbb{T}} \mathbf{g}\left(s\right)\left[\mathrm{d}X_j\left(s\right) - \lambda\left(\vartheta_0, s\right)\mathrm{d}s\right].$$

By the Central Limit Theorem

$$\eta_n \Longrightarrow \mathcal{N}\left(0, \mathbb{G}\left(\vartheta\right)\right).$$

The asymptotic normality (1) now follows from this convergence and the presentation

$$\sqrt{n}\left(\vartheta_n^* - \vartheta_0\right) = \frac{\partial \mathbf{H}\left(\vartheta\right)}{\partial \vartheta}\eta_n\left(1 + o\left(1\right)\right).$$

Recall that the vector-function $H\left(a\right)$ is continuously differentiable.

Example 1. Suppose that the intensity function is

$$\lambda(\theta,t)=\sum_{l=1}^{d}\theta_l h_l(t)+\lambda_0,\qquad t\in\mathbb{T}.$$

Introduce the vector-function $\mathbf{g}(\cdot)$ and the corresponding integrals $\mathbf{I}^{(d)}$. The vector $\mathbf{M}(\vartheta)=\mathring{A}\vartheta+\lambda_0\mathbf{G}$, where

$$\mathring{A}_{kl}=\int_{\mathbb{T}} g_k(t)\,h_l(t)\,\mathrm{d}t,\qquad G_k=\int_{\mathbb{T}} g_k(t)\,\mathrm{d}t$$

in obvious notations. Hence we can write

$$\vartheta=\mathring{A}^{-1}\left[\mathbf{M}(\vartheta)-\lambda_0\mathbf{G}\right]=\mathring{A}^{-1}\left[\mathbf{a}-\lambda_0\mathbf{G}\right]=\mathbf{H}(\mathbf{a}).$$

Therefore the MME ϑ_n^* is given by the equality

$$\vartheta_n^* = \mathring{A}^{-1}\left[\mathbf{a}_n - \lambda_0 \mathbf{G}\right] = \mathring{A}^{-1}\frac{1}{n}\sum_{j=1}^{n}\int_{\mathbb{T}} \mathbf{g}\left(t\right)\left[\mathrm{d}X_j\left(t\right) - \lambda_0 \mathrm{d}t\right]. \qquad (2)$$

This estimator by the Theorem 1 is consistent and asymptotically normal. To simplify its calculation we can take such functions $\mathbf{g}\left(\cdot\right)$ that the matrix $\mathring{A}$ became diagonal.

Example 2. Suppose that the inhomogeneous Poisson processes $X^{(n)}$ are observed on the time interval $\mathbb{T} = [0, \infty)$ and have the intensity function

$$\lambda(\vartheta, t) = \frac{t^{\beta-1}\alpha^{\beta}}{\Gamma(\beta)} \exp(-\alpha t), t \geq 0,$$

i.e., we have Poisson processes with the Gamma intensity function. The unknown parameter is $\vartheta = (\alpha, \beta)$. We know, that

$$M_1(\vartheta) = \int_0^\infty t\ \lambda(\vartheta, t)\,\mathrm{d}t = \frac{\beta}{\alpha},$$
$$M_2(\vartheta) = \int_0^\infty t^2 \lambda(\vartheta, t)\,\mathrm{d}t = \frac{\beta(\beta+1)}{\alpha^2}.$$

Hence, if we take $g(t) = (g_1(t), g_2(t)) = \left(t, t^2\right)$, then the system $\mathbf{M}(\vartheta) = \mathbf{a}$ has the unique solution

$$\alpha = \frac{a_1}{a_2 - a_1^2}, \qquad \beta = \frac{a_1^2}{a_2 - a_1^2}.$$

Therefore the MME $\vartheta_n^* = (\alpha_n^*, \beta_n^*)$ is

$$\alpha_n^* = \frac{\frac{1}{n}\sum_{j=1}^{n}\int_0^\infty t\mathrm{d}X_j\left(t\right)}{\left(\frac{1}{n}\sum_{j=1}^{n}\int_0^\infty t^2\mathrm{d}X_j\left(t\right) - \left(\frac{1}{n}\sum_{j=1}^{n}\int_0^\infty t\mathrm{d}X_j\left(t\right)\right)^2\right)}, \tag{3}$$

$$\beta_n^* = \frac{\left(\frac{1}{n}\sum_{j=1}^{n}\int_0^\infty t\mathrm{d}X_j\left(t\right)\right)^2}{\left(\frac{1}{n}\sum_{j=1}^{n}\int_0^\infty t^2\mathrm{d}X_j\left(t\right) - \left(\frac{1}{n}\sum_{j=1}^{n}\int_0^\infty t\mathrm{d}X_j\left(t\right)\right)^2\right)}. \tag{4}$$

This estimator is consistent and asymptotically normal.

The similar example can be considered and in the case of observations on $\mathbb{T} = (-\infty, +\infty)$ and the Gaussian intensity function with $\vartheta = \left(\alpha, \sigma^2\right)$:

$$\lambda\left(\vartheta, t\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\left(t-\alpha\right)^2}{2\sigma^2}\right\}, \quad t \in \mathbb{R}.$$

Further examples and the convergence of moments of these estimators can be found in *Kutoyants (Introduction to statistics of Poisson Processes (2018). To appear.*

One-Step MLE

The One-step MLE was introduced by Fisher (1925). This One-step procedure allows to improved a consistent estimator $\bar\vartheta_n$ up to asymptotically efficient (One-step MLE) $\vartheta_n^\star$. We consider the similar construction in the case of inhomogeneous Poisson processes. Suppose that the observations $X^n = (X_1, ..., X_n)$ are Poisson processes with the intensity function $\lambda(\vartheta, t), t \in \mathbb{T}$.

Condition $\mathcal{P}_0$. *We have a (preliminary) estimator* $\bar\vartheta_n$, *which is consistent and such that* $\sqrt{n}\left(\bar\vartheta_n - \vartheta_0\right)$ *is bounded in probability.*

Introduce the learning observations $X^N = (X_1, ..., X_N)$, $N = \left[n^\delta\right]$, $\delta \in \left(\frac{1}{2}, 1\right)$ and the One-step MLE

$$\vartheta_n^\star = \bar\vartheta_N + \frac{\mathbb{I}\left(\bar\vartheta_N\right)^{-1}}{n} \sum_{j=N+1}^{n} \int_{\mathbb{T}} \frac{\dot\lambda\left(\bar\vartheta_N, t\right)}{\lambda\left(\bar\vartheta_N, t\right)} \left[\mathrm{d}X_j\left(t\right) - \lambda\left(\bar\vartheta_N, t\right)\mathrm{d}t\right].$$

Here $\bar\vartheta_N$ is the preliminary estimator constructed by the first N observations.

Regularity conditions $\mathcal{L}_0$:

- *The function* $\ell\left(\vartheta, t\right) = \ln\lambda\left(\vartheta, t\right)$ *has three continuous bounded derivatives. w.r.t.* ϑ
- *The Fisher information matrix* $\mathbb{I}\left(\vartheta\right)$ *is uniformly on* $\vartheta \in \Theta$ *non degenerated:*

$$\inf_{\vartheta\in\Theta} \inf_{|\mu|=1} \mu^\tau \mathbb{I}\left(\vartheta\right)\mu > 0.$$

Here $\mu \in \mathbb{R}^d$.

Theorem 2 *Suppose that the conditions $\mathcal{P}_0$ and $\mathcal{L}_0$ are fulfilled. Then the One-step MLE $\vartheta_n^\star$ is asymptotically normal*

$$\sqrt{n}\left(\vartheta_n^\star - \vartheta_0\right) \Longrightarrow \mathcal{N}\left(0, \mathbb{I}\left(\vartheta_0\right)^{-1}\right).$$

We have the equality

$$\begin{aligned}\sqrt{n}\left(\vartheta_n^\star - \vartheta_0\right) &= \sqrt{n}\left(\bar\vartheta_N - \vartheta_0\right) + \\ &\quad + \mathbb{I}\left(\bar\vartheta_N\right)^{-1} \frac{1}{\sqrt{n}} \sum_{j=N+1}^{n} \int_{\mathbb{T}} \dot\ell\left(\bar\vartheta_N, t\right)\left[\mathrm{d}X_j\left(t\right) - \lambda\left(\vartheta_0, t\right)\mathrm{d}t\right] \\ &\quad + \mathbb{I}\left(\bar\vartheta_N\right)^{-1} \frac{n-N}{\sqrt{n}} \int_{\mathbb{T}} \dot\ell\left(\bar\vartheta_N, t\right)\left[\lambda\left(\vartheta_0, t\right) - \lambda\left(\bar\vartheta_N, t\right)\right]\mathrm{d}t.\end{aligned}$$

As $\bar{\vartheta}_N \longrightarrow \vartheta_0$ we can write

$$\mathbb{I}\left(\bar{\vartheta}_N\right)^{-1}\frac{1}{\sqrt{n}}\sum_{j=N+1}^{n}\int_{\mathbb{T}}\dot{\ell}\left(\bar{\vartheta}_N,t\right)\left[\mathrm{d}X_j\left(t\right)-\lambda\left(\vartheta_0,t\right)\mathrm{d}t\right]$$
$$=\mathbb{I}\left(\vartheta_0\right)^{-1}\frac{1}{\sqrt{n}}\sum_{j=N+1}^{n}\int_{\mathbb{T}}\dot{\ell}\left(\vartheta_0,t\right)\left[\mathrm{d}X_j\left(t\right)-\lambda\left(\vartheta_0,t\right)\mathrm{d}t\right]+o\left(1\right).$$

By the Central Limit Theorem

$$\frac{\mathbb{I}\left(\vartheta_0\right)^{-1}}{\sqrt{n}}\sum_{j=N+1}^{n}\int_{\mathbb{T}}\dot{\ell}\left(\vartheta_0,t\right)\left[\mathrm{d}X_j\left(t\right)-\lambda\left(\vartheta_0,t\right)\mathrm{d}t\right]\Longrightarrow\mathcal{N}\left(0,\mathbb{I}\left(\vartheta_0\right)^{-1}\right).$$

Let us consider the remainder

$$
\begin{aligned}
R_n = &\sqrt{n}\left(\bar\vartheta_N-\vartheta_0\right)\\
&\quad+\mathbb{I}\left(\bar\vartheta_N\right)^{-1}\frac{n-N}{\sqrt{n}}\int_{\mathbb{T}}\dot\ell\left(\bar\vartheta_N,t\right)\left[\lambda\left(\vartheta_0,t\right)-\lambda\left(\bar\vartheta_N,t\right)\right]{\rm d}t\\
&=\sqrt{n}\left(\bar\vartheta_N-\vartheta_0\right)\mathbb{I}\left(\bar\vartheta_N\right)^{-1}\left[\mathbb{I}\left(\bar\vartheta_N\right)-\int_{\mathbb{T}}\dot\lambda(\bar\vartheta_N,t)^\tau\dot\ell\left(\bar\vartheta_N,t\right){\rm d}t\right]\\
&\quad+\sqrt{n}\left(\bar\vartheta_N-\vartheta_0\right)O\left(\frac{N}{n}\right)+O\left(\sqrt{n}\left(\bar\vartheta_N-\vartheta_0\right)^2\right)=o\left(1\right),
\end{aligned}
$$

where we used the equality

$$\mathbb{I}\left(\bar\vartheta_N\right) = \int_{\mathbb{T}} \dot\lambda(\bar\vartheta_N, t)^\tau \dot\ell\left(\bar\vartheta_N, t\right) \mathrm{d}t$$

and the Taylor expansion at the point $\bar\vartheta_N$:

$$\begin{aligned}\lambda\left(\vartheta_0, t\right) - \lambda\left(\bar\vartheta_N, t\right) &= -\int_0^1 \dot\lambda\left(\bar\vartheta_N + s\left(\bar\vartheta_N - \vartheta_0\right), t\right)^\tau \left(\bar\vartheta_N - \vartheta_0\right) \mathrm{d}s \\ &= -\dot\lambda\left(\bar\vartheta_N, t\right)^\tau \left(\bar\vartheta_N - \vartheta_0\right) + O\left(\left(\bar\vartheta_N - \vartheta_0\right)^2\right).\end{aligned}$$

Remind that $\sqrt{n}O\left(\left(\bar\vartheta_N - \vartheta_0\right)^2\right) \sim \sqrt{n}O\left(n^{-\delta}\right) = o\left(1\right)$.

Therefore we obtained the representation

$$\sqrt{n}\left(\vartheta_n^\star - \vartheta_0\right) = \frac{\mathbb{I}\left(\vartheta_0\right)^{-1}}{\sqrt{n}} \sum_{j=N+1}^{n} \int_{\mathbb{T}} \dot\ell\left(\vartheta_0, t\right)\left[\mathrm{d}X_j\left(t\right) - \lambda\left(\vartheta_0, t\right)\mathrm{d}t\right] + o\left(1\right)$$

which proves the theorem.

Remark 1. If we suppose that the moments of the preliminary estimator are bounded, say,

$$\mathbf{E}_{\vartheta_0}\left|\bar\vartheta_n-\vartheta_0\right|^p\leq C$$

where $p\geq 2$ and $C>0$ does not depend on n, then the presented proof allows to verify that the moments of the One-step MLE are bounded too and that $\vartheta_n^\star$ is asymptotically efficient.

In all examples below the MLEs have no explicit expression.

Example 1. Suppose that the intensity function is

$$\lambda\left(\vartheta,t\right)=\sum_{l=1}^{d}\vartheta_l h_l\left(t\right)+\lambda_0, t\in\mathbb{T}$$

and ϑ_n^* is the MME. The Fisher information matrix is

$$\mathbb{I}\left(\vartheta\right)_{lk}=\int_{\mathbb{T}}\frac{h_l\left(t\right)h_k\left(t\right)}{h\left(t\right)^{\tau}\vartheta+\lambda_0}\,\mathrm{d}t,\qquad l,k=1,\ldots,d$$

and the One-step MLE in this case is

$$\vartheta_n^{\star}=\vartheta_N^*$$
$$+\frac{\mathbb{I}\left(\vartheta_N^*\right)^{-1}}{n}\sum_{j=N+1}^{n}\int_{\mathbb{T}}\frac{h\left(t\right)}{h\left(t\right)^{\tau}\vartheta_N^*+\lambda_0}\left[\mathrm{d}X_j\left(t\right)-h\left(t\right)^{\tau}\vartheta_n^*\mathrm{d}t-\lambda_0\mathrm{d}t\right].$$

Here $N=\left[n^{\delta}\right]$ and $\delta\in\left(\frac{1}{2},1\right)$. By the Theorem 2 this estimator is consistent and asymptotically normal.

Example 2. Suppose that the intensity function is

$$\lambda(\vartheta, t) = \frac{t^{\beta-1}\alpha^{\beta}\exp(-\alpha t)}{\Gamma(\beta)}, \qquad t \geq 0,$$

where the unknown parameter is $\vartheta = (\alpha, \beta)$. Once more we have a situation, where the explicit calculation of the MLE is impossible. The preliminary estimator can be the MME $\vartheta_n^* = (\alpha_n^*, \beta_n^*)$ (see (3) and (4)).

The vector $\dot{l}(\vartheta,t)=\left(\frac{\beta}{\alpha}-t,\ln(\alpha t)-\frac{\dot\Gamma(\beta)}{\Gamma(\beta)}\right)$ and the Fisher information matrix $\mathbb{I}(\vartheta)=(\mathbb{I}_{lk}(\vartheta))_{2\times2}$ is

$$\mathbb{I}_{11}(\vartheta)=\frac{\beta}{\alpha^2},\qquad \mathbb{I}_{12}(\vartheta)=-\frac{1}{\alpha},\quad \mathbb{I}_{22}(\vartheta)=\frac{\ddot\Gamma(\beta)\,\Gamma(\beta)-\dot\Gamma(\beta)^2}{\Gamma(\beta)^2}.$$

Hence the One-step MLE is

$$\vartheta_n^\star=\vartheta_N^*+\mathbb{I}(\vartheta_N^*)^{-1}\frac{1}{n}\sum_{j=N+1}^{n}\int_{\mathbb{T}}\dot{l}(\vartheta_N^*,t)\left[\mathrm{d}X_j(t)-\lambda(\vartheta_N^*,t)\,\mathrm{d}t\right]$$

and this estimators is asymptotically normal with the limit covariance matrix $\mathbb{I}(\vartheta_0)^{-1}$.

One-step MLE-process

Suppose that we have the same model of observations of n independent inhomogeneous Poisson processes: $X^n = (X_1, ..., X_n)$ with the intensity function $\lambda(\vartheta, t), t \in \mathbb{T}$, where ϑ is unknown parameter. Our goal is to construct an estimator process $\vartheta_n^\star = \left(\vartheta_{k,n}^\star, k = 1, \ldots, n\right)$, where the estimator $\vartheta_{k,n}^\star$ satisfies the following conditions

1. *The estimator $\vartheta_{k,n}^\star$ is based on the first k observations $X^{(k)}$.*
2. *The calculation of this estimator has to be relatively simple.*
3. *The estimator $\vartheta_{k,n}^\star$ is asymptotically efficient.*

Note that the MLE $\hat{\vartheta}_{k,n}$ defined by the relations

$$V\left(\hat{\vartheta}_{k,n}, X^k\right) = \sup_{\vartheta\in\Theta} V\left(\vartheta, X^k\right), \qquad k = 1, ..., n \tag{5}$$

satisfies the conditions (1) and (3), but not (2). The likelihood ratio function $V\left(\vartheta, X^k\right), \vartheta \in \Theta$ is

$$V\left(\vartheta, X^k\right) = \exp\left\{\sum_{j=1}^{k} \int_{\mathbb{T}} \ln\lambda\left(\vartheta, t\right) \mathrm{d}X_j\left(t\right) - k\int_{\mathbb{T}} \left[\lambda\left(\vartheta, t\right) - 1\right] \mathrm{d}t\right\}.$$

Remind that the solutions of the equations (5) in the case of non linear intensity functions $\lambda(\vartheta, \cdot)$ can be computationally difficult problems. This is typical situation of ”on-line” estimation.

The construction of such estimator-process is very close to the given above construction of the One-step MLE. Introduce the same learning observations $X^N=(X_1,...,X_N)$, where $N=[n^\delta]$, with $\delta\in\left(\frac{1}{2},1\right)$ and suppose that we have a preliminary estimator $\bar\vartheta_N$ such that $\sqrt{N}\left(\bar\vartheta_N-\vartheta_0\right)$ is bounded in probability (condition $\mathcal{P}_0$).

The One-step MLE-process is

$$\vartheta^\star_{k,n}=\bar\vartheta_N+\mathbb{I}\left(\bar\vartheta_N\right)^{-1}\frac{1}{k}\sum_{j=N+1}^{k}\int_{\mathbb{T}}\dot\ell\left(\bar\vartheta_N,t\right)\left[{\rm d}X_j\left(t\right)-\lambda\left(\bar\vartheta_N,t\right){\rm d}t\right],$$

where $k=N+1,...,n$.

Theorem 3 *Suppose that the conditions $\mathcal{P}_0$ and $\mathcal{L}_0$ are fulfilled. Then the One-step MLE-process $\vartheta_n^\star = \left(\vartheta_{k,n}^*, k = N+1, ..., n\right)$ is consistent and asymptotically normal*

$$\sqrt{k}\left(\vartheta_{k,n}^\star - \vartheta_0\right) \Longrightarrow \mathcal{N}\left(0, \mathbb{I}\left(\vartheta_0\right)^{-1}\right)$$

where we put $k = [sn]$. Here $s \in (0, 1]$.

Proof. There is no need to present a new proof because it is a slight modification of the given above proof of the Theorem 2.

Two-step MLE-process

The one step MLE-process presented in the preceding section allows us to calculate the values $\vartheta^\star_{k,n}$ for $k = N+1, ..., n$, where $N = \left[n^\delta\right]$ with $\delta \in (\frac{1}{2}, 1]$. Therefore we have no estimators for $k = 1, ..., N$.

It is interesting to reduce the learning interval and to start the estimation process earlier. Let us see how it can be done with the learning interval $X^N = (X_1, \ldots, X_N)$ with $N = \left[n^\delta\right]$ and $\delta \in \left(\frac{1}{3}, \frac{1}{2}\right]$. We suppose that a preliminary estimator $\bar\vartheta_N$ is given.

Then we define the second preliminary estimator

$$\bar\vartheta_{k,n} = \bar\vartheta_N + \mathbb{I}\left(\bar\vartheta_N\right)^{-1} \frac{1}{k} \sum_{j=N+1}^{k} \int_{\mathbb{T}} \dot\ell\left(\bar\vartheta_N, t\right) \left[{\rm d}X_j\left(t\right) - \lambda\left(\bar\vartheta_N, t\right) {\rm d}t \right],$$

and the Two-step MLE-process is defined by the relation

$$\vartheta^{\star\star}_{k,n} = \bar\vartheta_{k,n} + \mathbb{I}\left(\bar\vartheta_N\right)^{-1} \frac{1}{k} \sum_{j=N+1}^{k} \int_{\mathbb{T}} \dot\ell\left(\bar\vartheta_N, t\right) \left[{\rm d}X_j\left(t\right) - \lambda\left(\bar\vartheta_{k,n}, t\right) {\rm d}t \right],$$

where $k = N + 1, ..., n$.

Let us show that it is asymptotically normal

$$\sqrt{k}\left(\vartheta^{\star\star}_{k,n}-\vartheta_0\right)\Longrightarrow \mathcal{N}\left(0,\mathbb{I}\left(\vartheta_0\right)^{-1}\right).$$

Here $k=[sn]$ and $s\in(0,1]$. We have

$$\begin{aligned}\sqrt{k}\left(\vartheta^{\star\star}_{k,n}-\vartheta_0\right)&=\sqrt{k}\left(\bar\vartheta_{k,n}-\vartheta_0\right)+\\ &+\mathbb{I}\left(\bar\vartheta_N\right)^{-1}\frac{1}{k}\sum_{j=N+1}^{k}\int_{\mathbb{T}}\dot\ell\left(\bar\vartheta_N,t\right)\left[\mathrm{d}X_j\left(t\right)-\lambda\left(\vartheta_0,t\right)\mathrm{d}t\right]\\ &+\mathbb{I}\left(\bar\vartheta_N\right)^{-1}\frac{(k-N)}{k}\int_{\mathbb{T}}\dot\ell\left(\bar\vartheta_N,t\right)\left[\lambda\left(\vartheta_0,t\right)-\lambda\left(\bar\vartheta_{k,n},t\right)\right]\mathrm{d}t.\end{aligned}$$

We can write for some $\gamma>0$, which we chose later

$$
\begin{aligned}
n^\gamma\left(\bar\vartheta_{k,n}-\vartheta_0\right) &= n^\gamma\left(\bar\vartheta_N-\vartheta_0\right)\\
&\quad+\mathbb{I}\left(\bar\vartheta_N\right)^{-1}\frac{n^\gamma}{k}\sum_{j=N+1}^{k}\int_{\mathbb{T}}\dot\ell\left(\bar\vartheta_N,t\right)\left[\mathrm{d}X_j\left(t\right)-\lambda\left(\vartheta_0,t\right)\mathrm{d}t\right]\\
&\quad+\mathbb{I}\left(\bar\vartheta_N\right)^{-1}\frac{n^\gamma\left(k-N\right)}{k}\int_{\mathbb{T}}\dot\ell\left(\bar\vartheta_N,t\right)\left[\lambda\left(\vartheta_0,t\right)-\lambda\left(\bar\vartheta_N,t\right)\right]\mathrm{d}t\\
&= n^\gamma\left(\bar\vartheta_N-\vartheta_0\right)\left[J-\left(1-\frac{N}{k}\right)\mathbb{I}\left(\bar\vartheta_N\right)^{-1}\int_{\mathbb{T}}\dot\ell\left(\bar\vartheta_N,t\right)\lambda(\tilde\vartheta,t)\mathrm{d}t\right]\\
&\quad+\mathbb{I}\left(\bar\vartheta_N\right)^{-1}\frac{n^\gamma}{k}\sum_{j=N+1}^{k}\int_{\mathbb{T}}\dot\ell\left(\bar\vartheta_N,t\right)\left[\mathrm{d}X_j\left(t\right)-\lambda\left(\vartheta_0,t\right)\mathrm{d}t\right]\\
&= O\left(n^\gamma\left|\bar\vartheta_N-\vartheta_0\right|^2\right)+O\left(\frac{N}{k}\right)\\
&\quad+\mathbb{I}\left(\bar\vartheta_N\right)^{-1}\frac{n^\gamma}{k}\sum_{j=N+1}^{k}\int_{\mathbb{T}}\dot\ell\left(\bar\vartheta_N,t\right)\left[\mathrm{d}X_j\left(t\right)-\lambda\left(\vartheta_0,t\right)\mathrm{d}t\right].
\end{aligned}
$$

If we take $\gamma < \delta$ then we have

$$n^{\gamma} n^{-\delta} \left(n^{\frac{\delta}{2}} \left| \bar{\vartheta}_N - \vartheta_0 \right) \right|^2 \longrightarrow 0.$$

Further, as $\gamma < \delta \leq \frac{1}{2}$ we have

$$\frac{n^{\gamma}}{k} \sum_{j=N+1}^{k} \int_{\mathbb{T}} \dot{\ell} \left(\bar{\vartheta}_N, t \right) \left[\mathrm{d}X_j \left(t \right) - \lambda \left(\vartheta_0, t \right) \mathrm{d}t \right]$$
$$= \frac{n^{\gamma - \frac{1}{2}}}{\sqrt{sk}} \sum_{j=N+1}^{k} \int_{\mathbb{T}} \dot{\ell} \left(\bar{\vartheta}_N, t \right) \left[\mathrm{d}X_j \left(t \right) - \lambda \left(\vartheta_0, t \right) \mathrm{d}t \right] = o \left(n^{\gamma - \frac{1}{2}} \right) \to 0.$$

Hence for $\gamma < \delta$

$$n^{\gamma} \left(\bar{\vartheta}_{k,n} - \vartheta_0 \right) \longrightarrow 0.$$

Therefore

$$\sqrt{k}\left(\vartheta^{\star\star}_{k,n}-\vartheta_0\right)=O\left(\sqrt{k}\left|\bar\vartheta_{k,n}-\vartheta_0\right|\,\left|\bar\vartheta_N-\vartheta_0\right|\right)$$

$$+\,\mathbb{I}\left(\bar\vartheta_{k,n}\right)^{-1}\frac{1}{\sqrt{k}}\sum_{j=N+1}^{k}\int_{\mathbb{T}}\dot\ell\left(\bar\vartheta_N,t\right)\left[\mathrm{d}X_j\left(t\right)-\lambda\left(\vartheta_0,t\right)\mathrm{d}t\right].$$

We see that if we take $\frac{1}{2}-\gamma-\frac{\delta}{2}<0$ then

$$\sqrt{k}\left|\bar\vartheta_{k,n}-\vartheta_0\right|\left|\bar\vartheta_N-\vartheta_0\right|$$

$$=n^{\frac{1}{2}}n^{-\gamma}n^{-\frac{\delta}{2}}\left(n^{\gamma}\left|\bar\vartheta_{k,n}-\vartheta_0\right|\right)n^{\frac{\delta}{2}}\left|\bar\vartheta_N-\vartheta_0\right|\to 0$$

Therefore if $\delta \in \left(\frac{1}{3}, \frac{1}{2}\right)$, then we can take such γ, that $\gamma < \delta$ and $\gamma > \frac{1-\delta}{2}$. Finally we obtain

$$\begin{aligned}
&\sqrt{k}\left(\vartheta_{k,n}^{\star\star} - \vartheta_0\right) \\
&\qquad = \frac{\mathbb{I}\left(\vartheta_0\right)^{-1}}{\sqrt{k}} \sum_{j=N+1}^{k} \int_{\mathbb{T}} \dot{\ell}\left(\vartheta_0, t\right)\left[\mathrm{d}X_j\left(t\right) - \lambda\left(\vartheta_0, t\right)\mathrm{d}t\right] + o\left(1\right) \\
&\qquad \Longrightarrow \mathcal{N}\left(0, \mathbb{I}\left(\vartheta_0\right)^{-1}\right).
\end{aligned}$$

Therefore we proved the following theorem

Theorem 4 *Let the conditions $\mathcal{P}_0$ and $\mathcal{L}_0$ be fulfilled. Then the Two-step MLE-process $\left(\vartheta_{k,n}^{\star\star}, k = N+1, ..., n\right)$ is asymptotically normal*

$$\sqrt{k}\left(\vartheta_{k,n}^{**} - \vartheta_0\right) \Longrightarrow \mathcal{N}\left(0, \mathbb{I}\left(\vartheta_0\right)^{-1}\right).$$

Here $k = [sn]$.

Example 4. Suppose that the intensity function of the observed inhomogeneous Poisson process is

$$\lambda(\vartheta, t) = A \sin(2\pi t + \vartheta) - \lambda_0, \qquad 0 \leq t \leq 1$$

where $\vartheta \in \Theta = (c_1, c_2)$, $0 < \alpha < \beta < 2\pi$ and $A < \lambda_0$. Let us take $g(t) = \cos(2\pi t)$ and note that

$$M(\vartheta) = \int_0^1 g(t)\,\lambda(\vartheta, t)\,\mathrm{d}t = \frac{A}{2}\cos(\vartheta), \qquad \vartheta = \arccos\left(\frac{2M(\vartheta)}{A}\right).$$

The MME is

$$\vartheta_n^* = \arccos\left(\frac{2}{An}\sum_{j=1}^n \int_0^1 \cos(2\pi t)\,\mathrm{d}X_j(t)\right).$$

The Fisher information

$$\mathbb{I} = \int_0^1 \frac{A^2 \cos^2\left(2\pi t\right)}{A\sin\left(2\pi t\right) + \lambda_0} \mathrm{d}t$$

does not depend on ϑ. Let us take $N = \left[n^{\frac{4}{9}}\right]$ and introduce the Two-step MLE-process as follows $(k = N+1, \ldots, n,)$

$$\bar\vartheta_{k,n} = \vartheta_N^* + \frac{1}{\mathbb{I}k} \sum_{j=N+1}^{k} \int_0^1 \frac{A\cos\left(2\pi t + \vartheta_N^*\right)}{A\sin\left(2\pi t + \vartheta_N^*\right) + \lambda_0} \,\mathrm{d}X_j\left(t\right),$$

$$\vartheta_{k,n}^{**} = \bar\vartheta_{k,n} + \frac{1}{\mathbb{I}k} \sum_{j=N+1}^{k} \int_0^1 \frac{A\cos\left(2\pi t + \vartheta_N^*\right)}{A\sin\left(2\pi t + \vartheta_N^*\right) + \lambda_0} \,\mathrm{d}X_j\left(t\right)$$

$$- \frac{k-N}{\mathbb{I}k} \int_0^1 \frac{\left[A\cos\left(2\pi t + \vartheta_N^*\right)\right]\left[A\sin\left(2\pi t + \bar\vartheta_{k,n}\right) + \lambda_0\right]}{A\sin\left(2\pi t + \vartheta_N^*\right) + \lambda_0} \,\mathrm{d}t$$

Thank you for your attention !!

References

[1] Albeverio, S., Laob, L.-J. and Zhaoc, X.-L. (2002) *Continuous Time Financial Market with a Poisson Process.* Springer, N.Y.

[2] Bar-David, I. (1969) Communication under Poisson regime. *IEEE Trans. Inform. Theory,* IT-15, 1, 31–37.

[3] Blæsild, P. and Granfeldt, J. (2003) *Statistics with Applications in Biology and Geology.* Chapman & Hall, London.

[4] Borovkov, A. A. (1998) *Mathematical Statistics.* Gordon & Breach, Amsterdam.

[5] Davies, R. (1977) Testing the hypothesis that a point process is Poisson. *Adv. Appl. Probab.* **9**, 724-746.

[6] Ibragimov I.A. and Khasminskii R. (1981) *Statistical*

Estimation - Asymptotic Theory. Springer-Verlag, New York.

[7] Kamatani, K. and Uchida, M. (2015) Hybrid Multi-step estimators for stochastic differential equations based on sampled data. *Stat. Inference Stoch. Process.* 18, 2, 177-204.

[8] Khasminskii, R. Z. and Kutoyants, Yu. A. (2016) "On parameter estimation of hidden telegraph process". (arXiv:1509.02704) To appear in Bernoulli.

[9] Kutoyants, Yu.A. (1984) *Parameter Estimation for Stochastic Processes.* Heldermann, Berlin.

[10] Kutoyants Yu.A. (1998) *Statistical Inference for Spatial Poisson Processes.* Springer-Verlag, N. Y.

[11] Kutoyants, Yu.A. (2014) On approximation of the backward stochastic differential equation. Small noise, large samples and high frequency cases. *Proc. Steklov Inst. Math.*, 287, 133-154.

[12] Kutoyants, Yu.A. (2015) On the Multi-step MLE-process for ergodic diffusion. To appear in *Stochastic Process. Appl.* (http://dx.doi.org/10.1016/j.spa.2016.10.007)

[13] Kutoyants Yu.A. (2016) *Introduction to Statistics of Poisson Processes*. To appear.

[14] Kutoyants, Yu.A. (2016) On approximation of BSDE and Multi-step MLE-processes. *Probab. Uncertain. Quant. Risk.* 1, 4, 23-41.

[15] Kutoyants, Yu.A. and Motrunich, A. (2016) On Multi-step MLE-process for Markov sequences. *Metrika*, 79, 6, 705-724.

[16] Le Cam, L. (1956) On the asymptotic theory of estimation and testing hypotheses. *Proc. 3rd Berkeley Symposium I*, 355-368.

[17] Lehmann, E.L. (1999) *Elements of Large-Sample Theory.* Springer, N.Y.

[18] Liptser, R. and Shiryayev, A. N. (2005) *Statistics of Random Processes*. 2nd ed, v.1, v.2, Springer, N.Y.

[19] Rigdon, S.E. and Basu, A.P. (2000) *Statistical Methods for the Reliability of Repairable Systems*, New York: John Wiley & Sons, Inc.

[20] Robinson, P.M. (1988) The stochastic difference between econometric statistics. *Econometrica.* 56, 3, 531-548.

[21] Sarkar, S.K. (2016) *Single Molecule Biophysics and Poisson Process Approach to Statistical Mechanics.* Morgan & Claypool, San Rafael CA.

[22] Snyder, D.R. and Miller, M.I. (1991) *Random Point Processes in Time and Space.* Springer, N. Y.

[23] Uchida, M. and Yoshida, N. (2012) Adaptive estimation of ergodic diffusion process based on sampled data. *Stoch. Proces. Appl.*, **122**, 2885-2924.

[24] Utsu, T., Ogata, Y., and Matsu'ura, R. (1995) The centenary of the Omori formula for a decay law of aftershock activity. *J. Phys. Earth*, 43, 1–33.