On Parameter Estimation by Aggregated Poisson Observations

E. O. Accrachi, C. T. Aïdara and A. S. Dabye

Abstract. We consider the problem of parameter estimation by observations of inhomogeneous Poisson processes. The intensity function of the process is supposed to be a smooth function with respect to the unknown parameter. We propose a Chi-square statistic on the base of aggregated observations and we define a Minimum Chi-square Estimator with the help of this statistics. We show that this estimator is consistent and asymptotically normal. We discuss possible generalizations of the obtained results.

Introduction

Statistical inference for inhomogeneous Poisson processes provides interesting problems of mathematical statistics. It consists in inferring the unknown characteristics of the observed statistical model which is present in the intensity function of the inhomogeneous Poisson process. We can mention among others: Kutoyants [11, 12, 13], Karr [10], Kutoyants and Spokoiny [14], Aubry and Dabye [1], Ba and Dabye [2]. The most important aspects of inference are the theory of estimation and hypothesis testing. The method of Minumum Chi-square Estimation occupies an important place in this field. There are many applications in biology (Greenwood and Nikulin [6]), astrophysics (Mighell [15]), nuclear physics (Harris and Kandji [7]), as well as the references therein. The list is not exhaustive.

The Chi-square goodness of fit test was introduced in the work of Karl Pearson in the early twentieth century. Originally, this is a simple hypothesis testing problem. The study was made thanks to the statistics built by him

that bears his name. As a result, many of the contributions have been made later. Over the years, given the changing contexts, this method was applied to many different statistical models and has led to many other interesting tests. In the work by Greenwood and Nikulin [6], one can find a detailed summary of the history of the development of this method done by many researchers in this field.

The controversy between Pearson and Fisher regarding the distribution to be compared is one of the great stories in the history of statistics.

Nowadays we know that the results from this remarkable and fantastic discovery are only true for the case of simple null hypotheses. In 1924, Fisher proved that in the case of composite hypotheses, where we need to make estimates, the limit distribution of the Pearson's statistic is no longer valid. Moreover, this limit distribution depends on the method used to estimate the parameters. In particular, if the parameters are estimated by the Karl Pearson's statistic minimization method or by any other asymptotically equivalent method, then the limit distribution may still be Chi-square, but the number of degrees of freedom is reduced by the dimension of the estimated parameter. Chernoff and Lehmann proved in [4] that using the maximum likelihood method derived from raw data, the limit distribution of the Pearson's statistic changes dramatically. This distribution depends on the unknown parameter and therefore is unusable for the test. Once again, the problem of determining a free statistic (Asymptotic Distribution Free) arises.

We know that the method of the Maximum Likelihood Estimation is widely used and its sovereignty has been questioned by Berkson [3], who argues that the Minimum Chi-square, and not the Maximum Likelihood, is the basic principle of the statistical inference. In any application, there is a wide class of estimators that Berkson would classify as Minimum Chi-square estimators, whereas in the literature the definition of such a class is not unambiguous. Moreover, according to Berkson, the Pearson's Chi-square statistic formulation requires data from a discrete distribution, or data from a continuous distribution grouped into intervals (or more generally subsets) of the sample space. There is no doubt that the minimization of Pearson's Chi-square statistics almost invariably requires tedious and complex procedures, but current authors believe that the preference for maximum likelihood is perhaps more due to prejudice than to real computational problems.

Without going so far as to advocate the universal application of the estimation based on the minimization of Chi-square statistics, they feel that there are situations where it may be preferable to use other methods and, if necessary, to use modern powerful calculation techniques.

Shortly after, Greenwood and Nikulin [6] contributed, under some regularity conditions and for any continuous probability law, a limit distribution that does not depend on an unknown parameter of the statistic constructed

from the Maximum Likelihood Estimator (MLE). Also worth mentioning is Moore's significant contribution in [16], which summarized all these results. He showed, in particular, that the modifications made to Pearson's statistics are derived from Wald's approach [17].

Many researchers preferred to focus their work on the tests, rather than on the estimate of the Minimum Chi-square. The authors also discovered that there are almost no works on Poisson processes in continuous time.

In this paper, after introducing our statistical model of continuous time Poisson processes, we introduce a Chi-square type statistic and construct an estimator of the parameter ϑ with the help of this statistic. Subsequently, we study the asymptotic properties (the consistency and the asymptotic normality) of this estimator. We are advancing the perspectives, and propose possible generalizations of the obtained results.

1 Statistical model and the Minimum Chi-square Estimator

Suppose that we observe n samples of an inhomogeneous Poisson process $X_j = \{X_j(t), 0 \leq t \leq T\}$, $j = 1, \ldots, n$. The mean and intensity functions are denoted as $\Lambda(\vartheta, \cdot)$ and $\lambda(\vartheta, \cdot)$ respectively, i.e.

$$\mathbb{E}_\vartheta X_j(t) = \Lambda(\vartheta, t) = \int_0^t \lambda(\vartheta, s)ds, \qquad \Lambda(\vartheta, T) < \infty.$$

The parameter $\vartheta \in \Theta = (\alpha, \beta)$, where $|\alpha| + |\beta| < \infty$. Here ϑ is the unknown parameter and $\mathbb{E}_\vartheta$ is the mathematical expectation.

Our goal is to construct an estimator of ϑ by aggregated data.

We introduce a partition $\mathcal{A}$ of the interval $[0, T]$ in the following way: we choose numbers $a_0, \ldots, a_L$ such that

$$0 = a_0 < a_1 < \ldots < a_{L-1} < a_L = T,$$

and we obtain the intervals

$$A_1 = [a_0, a_1), A_2 = [a_1, a_2), \ldots, A_L = [a_{L-1}, a_L].$$

Note that these intervals verify the following relations

$$A_l \cap A_m = \emptyset, \quad l \neq m, \qquad \cup_{l=1}^L A_l = [0, T], \quad \Lambda(\vartheta, A_l) = \int_{A_l} \lambda(\vartheta, s)\, ds.$$

Here and in the sequel, in the notation "$\Lambda(\vartheta, A_l)$", Λ is the intensity measure. Let us denote as well

$$X_j^l = X_j(a_l) - X_j(a_{l-1}) = \int_{a_{l-1}}^{a_l} dX_j(t), \ l = 1, \ldots, L, \ j = 1, \ldots, n,$$

then $X_j = \left(X_j^l, \quad l = 1, \ldots, L\right)$, $j = 1, \ldots, n$, are independent Poisson random vectors which verify

$$\mathbb{P}\left(X_j^l = k\right) = \frac{\left(\Lambda\left(\vartheta, [a_{l-1}, a_l)\right)\right)^k}{k!} \exp\left\{-\Lambda\left(\vartheta, [a_{l-1}, a_l)\right)\right\}.$$

Our goal is to introduce the Chi-square type statistic and to construct an estimator of the parameter ϑ with the help of this statistic.

Let us put

$$\mathbb{T}_n(\vartheta) = \sum_{l=1}^{L} \frac{\left(\hat{\Lambda}_n(A_l) - \Lambda(\vartheta, A_l)\right)^2}{\Lambda(\vartheta, A_l)},$$

where

$$\hat{\Lambda}_n(A_l) = \frac{1}{n}\sum_{j=1}^{n}\left(X_j(a_l) - X_j(a_{l-1})\right) = \frac{1}{n}\sum_{j=1}^{n} X_j^l.$$

We have to estimate ϑ by the observations $X^n = \left(X_j, \, j = 1, \ldots, n\right)$. We denote ϑ_0 the true value of the parameter. The derivative of any function $f(\vartheta, t)$ $(\vartheta \in \Theta, \, t \in [0, T])$ with respect to ϑ will be always denoted $\dot{f}(\vartheta, t)$.

Let us introduce the following functions

$$\rho\left(\vartheta_0, \vartheta\right) = \sum_{l=1}^{L} \frac{\left(\Lambda(\vartheta, A_l) - \Lambda(\vartheta_0, A_l)\right)^2}{\Lambda(\vartheta_0, A_l)}, \qquad \psi(\vartheta_0, \gamma) = \inf_{|\vartheta - \vartheta_0| > \gamma} \sqrt{\rho\left(\vartheta_0, \vartheta\right)},$$

$$\delta_l = \hat{\Lambda}_n(A_l) - \Lambda(\vartheta_0, A_l), \qquad \varphi_l(\vartheta) = \Lambda(\vartheta_0, A_l) - \Lambda(\vartheta, A_l).$$

Define the *Minimum Chi-square Estimator* (MCE) $\vartheta_n^\star$ which minimizes the function $\mathbb{T}_n(\vartheta)$ as a solution of the equation

$$\mathbb{T}_n(\vartheta_n^\star) = \inf_{\vartheta \in \Theta} \mathbb{T}_n(\vartheta).$$

If this equation has more than one solution then any of them can be taken as the MCE. In the sequel we will write

$$\vartheta_n^\star = \arg\inf_{\vartheta \in \Theta} \mathbb{T}_n(\vartheta).$$

We are interested in asymptotic properties of the MCE for large samples $(n \to \infty)$. We need the following conditions.

Conditions $\mathcal{C}$.

$\mathcal{C}_1$: $\varkappa_* = \min_{1 \leq l \leq L} \inf_{\vartheta \in \Theta} \Lambda(\vartheta, A_l) > 0$.

$\mathcal{C}_2$: $\psi(\vartheta_0, \gamma) > 0$ for all $\vartheta_0 \in \Theta$ and $\gamma > 0$.

$\mathcal{C}_3$: The function $\Lambda(\vartheta, \cdot)$ is differentiable with respect to $\vartheta \in \Theta$, and

$$\inf_{\vartheta \in \Theta} \mathcal{D}(\vartheta) > 0, \qquad \mathcal{D}(\vartheta) = \sum_{l=1}^{L} \frac{\dot{\Lambda}^2(\vartheta, A_l)}{\Lambda(\vartheta, A_l)}.$$

2 Consistency

Theorem 1 *Suppose that the conditions $\mathcal{C}_1$ and $\mathcal{C}_2$ hold. Then the MCE $\vartheta_n^\star$ is consistent. More precisely,*

$$\mathbb{P}_{\theta_0}\left\{|\vartheta_n^\star - \vartheta_0| > \gamma\right\} \ \leq \ \frac{4\,\Lambda(\vartheta_0, T)}{n\,\varkappa_*\,\psi(\vartheta_0, \gamma)^2}\ \xrightarrow[n\to+\infty]{}\ 0. \tag{1}$$

Proof. Using the properties of the norm

$$\|a + b\| \geq \|a\| - \|b\| \text{ and } \|a + b\| \leq \|a\| + \|b\|,$$

for any $\gamma > 0$ we can write

$$\mathbb{P}_{\theta_0}\left\{|\vartheta_n^\star - \vartheta_0| > \gamma\right\} = \mathbb{P}_{\theta_0}\left\{\inf_{|\vartheta-\vartheta_0|\leq\gamma}\sqrt{\mathbb{T}_n(\vartheta)} > \inf_{|\vartheta-\vartheta_0|>\gamma}\sqrt{\mathbb{T}_n(\vartheta)}\right\}$$

$$= \mathbb{P}_{\theta_0}\left\{\inf_{|\vartheta-\vartheta_0|\leq\gamma}\left\|\frac{\delta_l}{\Lambda(\vartheta, A_l)} + \frac{\varphi_l(\vartheta)}{\Lambda(\vartheta, A_l)}\right\| > \inf_{|\vartheta-\vartheta_0|>\gamma}\left\|\frac{\delta_l}{\Lambda(\vartheta, A_l)} + \frac{\varphi_l(\vartheta)}{\Lambda(\vartheta, A_l)}\right\|\right\}$$

$$\leq \mathbb{P}_{\theta_0}\left\{\inf_{|\vartheta-\vartheta_0|\leq\gamma}\left\|\frac{\delta_l}{\Lambda(\vartheta, A_l)} + \frac{\varphi_l(\vartheta)}{\Lambda(\vartheta, A_l)}\right\|\right.$$

$$\left. > \inf_{|\vartheta-\vartheta_0|>\gamma}\left(\left\|\frac{\varphi_l(\vartheta)}{\sqrt{\Lambda(\vartheta, A_l)}}\right\| - \left\|\frac{\delta_l}{\sqrt{\Lambda(\vartheta, A_l)}}\right\|\right)\right\}.$$

Here and in the sequel $\|\mathbf{a}\|$ means $\|(a_1, \ldots, a_L)\|$, where $\|.\|$ is the Euclidian norm in $\mathcal{R}^{\mathcal{L}}$. Further,

$$\mathbb{P}_{\theta_0}\left\{|\vartheta_n^\star - \vartheta_0| > \gamma\right\}$$

$$\leq \mathbb{P}_{\theta_0}\left\{\inf_{|\vartheta-\vartheta_0|\leq\gamma}\left\|\frac{\delta_l}{\sqrt{\Lambda(\vartheta, A_l)}}\right\| + \inf_{|\vartheta-\vartheta_0|\leq\gamma}\left\|\frac{\varphi_l(\vartheta)}{\sqrt{\Lambda(\vartheta, A_l)}}\right\|\right.$$

$$\left. > \inf_{|\vartheta-\vartheta_0|>\gamma}\left(\left\|\frac{\varphi_l(\vartheta)}{\Lambda(\vartheta, A_l)}\right\| - \left\|\frac{\delta_l}{\sqrt{\varkappa_*}}\right\|\right)\right\}$$

$$\leq \mathbb{P}_{\theta_0}\left\{2\left\|\frac{\delta_l}{\sqrt{\varkappa_*}}\right\| > \inf_{|\vartheta-\vartheta_0|>\gamma}\left\|\frac{\varphi_l(\vartheta)}{\sqrt{\Lambda(\vartheta, A_l)}}\right\|\right\} = \mathbb{P}_{\theta_0}\left\{2\left\|\frac{\delta_l}{\sqrt{\varkappa_*}}\right\| > \psi(\vartheta_0, \gamma)\right\},$$

since

$$\inf_{|\vartheta-\vartheta_0|\leq\gamma}\left\|\frac{\varphi_l(\vartheta)}{\sqrt{\Lambda(\vartheta, A_l)}}\right\| = 0.$$

Hence, we obtaine

$$
\begin{aligned}
\mathbb{P}_{\theta_0}\left\{|\vartheta_n^\star - \vartheta_0| > \gamma\right\} &\leq \mathbb{P}_{\theta_0}\left\{2\left\|\frac{\delta_l}{\sqrt{\varkappa_*}}\right\| > \psi(\vartheta_0, \gamma)\right\} \\
&= \mathbb{P}_{\theta_0}\left\{\frac{2}{\sqrt{\varkappa_*}}\left(\sum_{l=1}^{L}\left(\hat{\Lambda}_n(A_l) - \Lambda(\vartheta, A_l)\right)^2\right)^{\frac{1}{2}} > \psi(\vartheta_0, \gamma)\right\} \\
&\leq \frac{4}{\varkappa_* \, \psi(\vartheta_0, \gamma)^2}\mathbb{E}_{\vartheta_0}\left(\sum_{l=1}^{L}\left(\hat{\Lambda}_n(A_l) - \Lambda(\vartheta_0, A_l)\right)^2\right) \\
&= \frac{4}{\varkappa_* \, \psi(\vartheta_0, \gamma)^2}\sum_{l=1}^{L}\mathbb{V}_{\vartheta_0}\left(\frac{1}{n}\sum_{j=1}^{n}X_j^l\right) = \frac{4}{n\,\varkappa_* \, \psi(\vartheta_0, \gamma)^2}\sum_{l=1}^{L}\Lambda(\vartheta_0, A_l) \\
&= \frac{4\,\Lambda(\vartheta_0, T)}{n\,\varkappa_* \, \psi(\vartheta_0, \gamma)^2}\xrightarrow[n\to+\infty]{} 0,
\end{aligned}
$$

where we used the Tchebychev inequality and the fact that

$$
\sum_{l=1}^{L}\Lambda(\vartheta_0, A_l) = \sum_{l=1}^{L}\int_{a_{l-1}}^{a_l}\lambda(\vartheta_0, s)\,ds = \Lambda(\vartheta_0, T).
$$

The theorem is proved. □

Remark 1 *For the proof of consistency there is no condition on the regularity of the model. This means that the obtained result is valid for the discontinuous intensity functions $\lambda(.)$.*

3   Asymptotic normality

In this section we prove the weak convergence to a normal law of the MCE $\vartheta_n^\star$. By " $\Rightarrow$ " we denote the convergence in distribution. The following theorem is valid.

Theorem 2 *If the conditions $\mathcal{C}$ are fulfilled, then the MCE $\vartheta_n^\star$ is asymptotically normal, that is*

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

Proof. Let us recall that the function $\mathbb{T}_n(\vartheta)$ is defined by

$$
\mathbb{T}_n(\vartheta) = \sum_{l=1}^{L}\frac{\left(\hat{\Lambda}_n(A_l) - \Lambda(\vartheta, A_l)\right)^2}{\Lambda(\vartheta, A_l)}
$$

and let us remind that the MCE $\vartheta_n^\star$ is a solution of the equation

$$\vartheta_n^\star = \arg\inf_{\vartheta\in\Theta} \mathbb{T}_n(\vartheta).$$

The function $\mathbb{T}_n(\vartheta)$ is differentiable with respect to ϑ. The MCE $\vartheta_n^\star$ minimizes the function $\mathbb{T}_n(\vartheta)$ and is consistent for ϑ_0 in the open set Θ. So the MCE is in the open set Θ with a probability converging to 1, and with the same probability the derivative

$$\frac{\partial}{\partial\vartheta}\mathbb{T}_n(\vartheta_n^\star) = 0. \tag{2}$$

The calculation of the derivative $\mathbb{T}_n(\vartheta)$ with respect to ϑ gives us

$$\frac{\partial}{\partial\vartheta}\mathbb{T}_n(\vartheta) = -2\sum_{l=1}^{L} \frac{\left(\hat{\Lambda}_n(A_l) - \Lambda(\vartheta,A_l)\right)\dot{\Lambda}(\vartheta,A_l)}{\Lambda(\vartheta,A_l)}$$

$$-\sum_{l=1}^{L} \frac{\left(\hat{\Lambda}_n(A_l) - \Lambda(\vartheta,A_l)\right)^2\dot{\Lambda}(\vartheta,A_l)}{\Lambda(\vartheta,A_l)^2}$$

$$= -\sum_{l=1}^{L} \frac{\left(\hat{\Lambda}_n(A_l) - \Lambda(\vartheta,A_l)\right)\dot{\Lambda}(\vartheta,A_l)}{\Lambda(\vartheta,A_l)^2} \left(\hat{\Lambda}_n(A_l) + \Lambda(\vartheta,A_l)\right).$$

Therefore, the equation (2) becomes

$$\sum_{l=1}^{L} \frac{\left(\hat{\Lambda}_n(A_l) - \Lambda(\vartheta_n^\star,A_l)\right)\dot{\Lambda}(\vartheta_n^\star,A_l)}{\Lambda(\vartheta_n^\star,A_l)^2} \left(\hat{\Lambda}_n(A_l) + \Lambda(\vartheta_n^\star,A_l)\right) = 0. \tag{3}$$

Recall that

$$\hat{\Lambda}_n(A_l) = \frac{1}{n}\sum_{j=1}^{n} X_j^l$$

and, according to the weak law of large numbers,

$$\hat{\Lambda}_n(A_l) \to \mathbb{E}_{\vartheta_0} X_j^l = \Lambda(\vartheta_0,A_l). \tag{4}$$

As the MCE is consistent and the functions $\Lambda(\vartheta,A_l)$, $\dot{\Lambda}(\vartheta,A_l)$ are continues, the equation (3) can be rewritten as

$$\sum_{l=1}^{L} \frac{\left(\hat{\Lambda}_n(A_l) - \Lambda(\vartheta_n^\star,A_l)\right)\dot{\Lambda}(\vartheta_n^\star,A_l)}{\Lambda(\vartheta_n^\star,A_l)^2} \left(2\Lambda(\vartheta_0,A_l)\right)\left(1+o(1)\right)$$

$$= 2\sum_{l=1}^{L} \frac{\left(\hat{\Lambda}_n(A_l) - \Lambda(\vartheta_n^\star,A_l)\right)\dot{\Lambda}(\vartheta_0,A_l)}{\Lambda(\vartheta_0,A_l)} \left(1+o(1)\right) = 0$$

where $o(1)$ is convergence in probability.

Let us introduce the random variable

$$\zeta_n(\vartheta_0, A_l) = \sqrt{n}\Big(\hat{\Lambda}_n(A_l) - \Lambda(\vartheta_0, A_l)\Big)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(X_j(a_l) - X_j(a_{l-1}) - \Lambda(\vartheta_0, A_l)\Big)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(X_j^l - \Lambda(\vartheta_0, A_l)\Big) \implies \zeta_l \sim \mathcal{N}\left(0, \Lambda(\vartheta_0, A_l)\right). \qquad (5)$$

Adding and subtracting $\Lambda(\vartheta_0, A_l)$ to $\zeta_n(\vartheta_n^\star, A_l)$ and using the Taylor expansion, we obtain

$$\zeta_n(\vartheta_n^\star, A_l) = \sqrt{n}\Big(\hat{\Lambda}_n(A_l) - \Lambda(\vartheta_n^\star, A_l)\Big)$$

$$= \sqrt{n}\Big(\hat{\Lambda}_n(A_l) - \Lambda(\vartheta_0, A_l)\Big) - \sqrt{n}\Big(\Lambda(\vartheta_n^\star, A_l) - \Lambda(\vartheta_0, A_l)\Big)$$

$$= \zeta_n(\vartheta_0, A_l) - \sqrt{n}\Big(\vartheta_n^\star - \vartheta_0\Big)\dot{\Lambda}(\tilde{\vartheta}_n, A_l),$$

where $\tilde{\vartheta}_n$ is an intermediate value between ϑ_0 and $\vartheta_n^\star$. Then

$$\sum_{l=1}^{L} \frac{\Big(\zeta_n(\vartheta_0, A_l) - \sqrt{n}\big(\vartheta_n^\star - \vartheta_0\big)\dot{\Lambda}(\tilde{\vartheta}_n, A_l)\Big)\dot{\Lambda}(\vartheta_0, A_l)}{\Lambda(\vartheta_0, A_l)} \Big(1 + o(1)\Big) = 0.$$

Consider the following random variable

$$u_n^\star = \sqrt{n}\big(\vartheta_n^\star - \vartheta_0\big), \qquad \vartheta_n^\star = \vartheta_0 + \frac{u_n^\star}{\sqrt{n}}.$$

We can write

$$\sum_{l=1}^{L} \frac{\dot{\Lambda}(\vartheta_0, A_l)}{\Lambda(\vartheta_0, A_l)}\zeta_n(\vartheta_0, A_l)(1 + o(1)) - u_n^\star \sum_{l=1}^{L} \frac{\dot{\Lambda}(\vartheta_0, A_l)^2}{\Lambda(\vartheta_0, A_l)}\Big(1 + o(1)\Big) = 0$$

and

$$u_n^\star = \left(\sum_{l=1}^{L} \frac{\dot{\Lambda}(\vartheta_0, A_l)^2}{\Lambda(\vartheta_0, A_l)}\right)^{-1} \sum_{l=1}^{L} \frac{\dot{\Lambda}(\vartheta_0, A_l)}{\Lambda(\vartheta_0, A_l)}\zeta_n(\vartheta_0, A_l)\Big(1 + o(1)\Big).$$

Further,

$$\sum_{l=1}^{L} \frac{\dot{\Lambda}(\vartheta_0, A_l)}{\Lambda(\vartheta_0, A_l)}\zeta_n(\vartheta_0, A_l) \implies \sum_{l=1}^{L} \frac{\dot{\Lambda}(\vartheta_0, A_l)}{\Lambda(\vartheta_0, A_l)}\zeta_l \sim \mathcal{N}\left(0, \mathcal{D}\left(\vartheta_0\right)\right),$$

and hence

$$u_n^\star = \sqrt{n}\left(\vartheta_n^\star - \vartheta_0\right) \implies \mathcal{N}\left(0, \mathcal{D}\left(\vartheta_0\right)^{-1}\right).$$

$\square$

Remark 2 *According to the theorem, the limit variance of the MCE is the same as the limit variance of the MLE. This is an important property of the MCE.*

The detailed analysis of the proof shows that we have the convergence of moments too and, in particular,

$$\lim_{n\to\infty} n\mathbb{E}_{\vartheta_0}\left(\vartheta_n^\star - \vartheta_0\right)^2 = \mathcal{D}\left(\vartheta_0\right).$$

Moreover, this convergence is uniform on compacts $K = [\alpha_1, \beta_1] \subset \Theta$:

$$\lim_{n\to\infty} \sup_{\vartheta_0 \in K} \left| n\mathbb{E}_{\vartheta_0}\left(\vartheta_n^\star - \vartheta_0\right)^2 - \mathcal{D}\left(\vartheta_0\right)\right| = 0. \qquad (6)$$

4   On asymptotic efficiency

Let us fix a partition $\mathcal{A}$ and denote $\left\{\mathbb{P}_\vartheta^{(n)}, \vartheta \in \Theta\right\}$ the family of measures induced in the space $\mathcal{R}^{nL}$ by the observations $X^n = (X_1, \ldots, X_n)$. The likelihood ratio formula is

$$L\left(\vartheta, X^n\right) = \exp\left\{\sum_{j=1}^n \sum_{l=1}^L X_j^l \ln \Lambda\left(\vartheta, A_l\right) - n\sum_{l=1}^L \Lambda\left(\vartheta, A_l\right) - \sum_{j=1}^n \sum_{l=1}^L \ln\left(X_j^l\right)\right\}.$$

We suppose that the conditions $\mathcal{C}$ are fulfilled and put $\varphi_n = n^{-1/2}$. Then the normalized likelihood ratio

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

admits the representation

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

Here

$$\Delta_n\left(\vartheta_0, X^n\right) = \frac{1}{\sqrt{n}}\sum_{j=1}^n \sum_{l=1}^L \frac{\dot{\Lambda}\left(\vartheta_0, A_l\right)}{\Lambda\left(\vartheta_0, A_l\right)}\left[X_j^l - \Lambda\left(\vartheta_0, A_l\right)\right],$$

$$\mathbb{I}\left(\vartheta_0\right) = \sum_{l=1}^L \frac{\dot{\Lambda}\left(\vartheta_0, A_l\right)^2}{\Lambda\left(\vartheta_0, A_l\right)}, \quad \Delta_n\left(\vartheta_0, X^n\right) \Longrightarrow \mathcal{N}\left(0, \mathbb{I}\left(\vartheta_0\right)\right)$$

and $\varepsilon_n \to 0$ (convergence in probability).

Therefore the family of measures $\left\{\mathbb{P}_\vartheta^{(n)}, \vartheta \in \Theta\right\}$ is locally asymptotically normal (LAN), and we have the following minimax lower bound on the risk of all the estimators $\bar{\vartheta}_n$

$$\lim_{\delta\to 0} \varliminf_{n\to\infty} \sup_{|\vartheta - \vartheta_0| \leq \delta} n\mathbb{E}_\vartheta\left(\bar{\vartheta}_n - \vartheta\right)^2 \geq \mathbb{I}\left(\vartheta_0\right)^{-1},$$

for all $\vartheta_0 \in \Theta$.

This lower bound allows us to define the class of asymptotically efficient estimators as estimators $\bar{\vartheta}_n$ which satisfy the equality

$$\lim_{\delta \to 0} \lim_{n \to \infty} \sup_{|\vartheta - \vartheta_0| \leq \delta} n \mathbb{E}_\vartheta \left(\bar{\vartheta}_n - \vartheta \right)^2 = \mathbb{I}\left(\vartheta_0\right)^{-1},$$

for all $\vartheta_0 \in \Theta$.

Note that the MLE $\hat{\vartheta}_n$, defined by the relation $L\left(\hat{\vartheta}_n, X^n\right) = \sup_{\vartheta \in \Theta} L\left(\vartheta, X^n\right)$ is one of the solutions of the equation

$$\sum_{j=1}^n \sum_{l=1}^L \frac{\dot{\Lambda}(\hat{\vartheta}_n, A_l)}{\Lambda(\hat{\vartheta}_n, A_l)} \left[X_j^l - \Lambda(\hat{\vartheta}_n, A_l) \right] = 0.$$

This estimator is consistent, asymptotically normal

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

and asymptotically efficient (see [11] and [9]).

Thanks to the convergence (6), we can write

$$\sup_{|\vartheta - \vartheta_0| \leq \delta} n \mathbb{E}_\vartheta \left(\bar{\vartheta}_n - \vartheta \right)^2 \longrightarrow \sup_{|\vartheta - \vartheta_0| \leq \delta} \mathcal{D}\left(\vartheta\right) \longrightarrow \mathcal{D}\left(\vartheta_0\right) = \mathbb{I}\left(\vartheta_0\right)^{-1}$$

as $n \to \infty$ and $\delta \to 0$. Therefore, the MCE $\vartheta_n^\star$ is asymptotically efficient.

5   Possible generalizations

1. The multidimensional case $\left(\boldsymbol{\vartheta} \in \boldsymbol{\Theta} \subset \mathbb{R}^d, d > 1\right)$ can be treated by the similar way. The only difference is in the definition of the matrix

$$\mathcal{D}\left(\boldsymbol{\vartheta}\right)_{r,m} = \sum_{l=1}^L \frac{\partial \Lambda(\boldsymbol{\vartheta}, A_l)}{\partial \vartheta_r} \frac{\partial \Lambda(\boldsymbol{\vartheta}, A_l)}{\partial \vartheta_m}, \qquad r, m = 1, \ldots, d.$$

The MCE in this case is asymptotically normal with the limit covariance matrix $\mathcal{D}(\boldsymbol{\vartheta}_0)^{-1}$.

2. Let us consider the following problem. Suppose we have the possibility to choose the partition $\mathcal{A} = (A_l, l = 1, \ldots, L)$ such that $\cup_{l=1}^L A_l = [0, T]$ and $A_l \cap A_m = \emptyset$ if $l \neq m$. The following question naturally arises: how to choose it in optimal way? We know that the limit variance of the MCE is $\mathcal{D}(\boldsymbol{\vartheta}_0, \mathcal{A})^{-1}$, where

$$\mathcal{D}(\boldsymbol{\vartheta}_0, \mathcal{A}) = \sum_{l=1}^L \frac{\dot{\Lambda}(\boldsymbol{\vartheta}_0, A_l)^2}{\Lambda(\boldsymbol{\vartheta}_0, A_l)}.$$

Therefore, in order to minimize $\mathcal{D}(\boldsymbol{\vartheta}_0, \mathcal{A})^{-1}$ we have to maximize $\mathcal{D}(\boldsymbol{\vartheta}_0, \mathcal{A})$. In other words, we have to find $\mathcal{A}^* = (A_l^\star, l = 1, \ldots, L)$ such that

$$\mathcal{D}^\star(\boldsymbol{\vartheta}_0, \mathcal{A}) = \sum_{l=1}^{L} \frac{\dot{\Lambda}(\boldsymbol{\vartheta}_0, A_l^\star)^2}{\Lambda(\boldsymbol{\vartheta}_0, A_l^\star)} \geq \mathcal{D}(\boldsymbol{\vartheta}_0, \mathcal{A}),$$

for any choice of the partition $\mathcal{A} = (A_l, l = 1, \ldots, L)$. Of course, the optimal choice $\mathcal{A}^* = \mathcal{A}^*(\boldsymbol{\vartheta}_0)$ depends on the unknown parameter $\boldsymbol{\vartheta}_0$. The partition $\mathcal{A}^*(\boldsymbol{\vartheta}_0)$ is obtained as a result of the corresponding problem of optimization. A similar problem for the MLE was considered by Kutoyants and Spokoiny [14] (see also Kutoyants [13], section 4.3). As in these works, the solution carried is out in two steps. First, we construct a preliminary consistent estimator $\bar{\theta}_N$ of the parameter $\boldsymbol{\vartheta}_0$ by one part of observations $X^N = (X_1, \ldots, X_N)$, where $N < n$. To do it we can take any $\mathcal{A}$ satisfying the conditions $\mathcal{C}$. Then, using this estimator, we can take the partition $\mathcal{A}^*(\bar{\boldsymbol{\vartheta}}_N)$ and prove the asymptotic "optimality" of this construction.

3. Another interesting problem can be the choice of the partition in such a way that the contributions of the observations from different intervals to the limit variance of MCE are equal. This means that $\mathcal{A}$ is such that

$$d(\boldsymbol{\vartheta}_0) = \frac{\dot{\Lambda}(\boldsymbol{\vartheta}_0, A_1)^2}{\Lambda(\boldsymbol{\vartheta}_0, A_1)} = \frac{\dot{\Lambda}(\boldsymbol{\vartheta}_0, A_l)^2}{\Lambda(\boldsymbol{\vartheta}_0, A_l)}, \quad l = 2, \ldots, L.$$

Once more, the "optimal" $\mathcal{A}$ depends on the unknown parameter $\boldsymbol{\vartheta}_0$. The solution of this problem can be carried out in two steps, too. First, we construct, as above, a consistent estimator, and then we use it to define the sets.

4. The obtained results allow constructing the corresponding goodness-of-fit test with basic parametric hypothesis. The test can be asymptotically distribution free, but it is consistent against a special class of alternatives. As in classical case, the alternatives which give the same values of $\Lambda(\boldsymbol{\vartheta}_0, A_l)$ cannot be detected by the test.

References

[1] C. Aubry and A.S. Dabye, Asymptotic normality of the minimum distance estimators for Poisson process with a discontinuous intensity function, J. Statist. Plann. Inference, **99** (2001), 3–23.

[2] D.B. Ba and A.S. Dabye, On minimum L^p-distance estimation for inhomogeneous Poisson processes, Comm. Statist. Theory Methods, **45** (2016), No. 21, 6461–6470.

[3] J. Berkson, Minimum Chi-square, not maximum likelihood, Ann. Statist., **8** (1980), No. 3, 457–487.

[4] H. Chernoff and E.L. Lehmann, The use of maximum likelihood estimates in χ^2 tests for goodness of fit, Ann. Math. Statist., **25** (1954), 579–586.

[5] R.A. Fisher, On a property connecting the χ^2 measure of discrepancy with the method of maximum likelihood, Att de Congresso Internationale dei Mathematici, Bologna, **6** (1928), 94–100.

[6] P.E. Greenwood and M. Nikulin, *A Guide of Chi-Squared Testing*, John Wiley and Sons, New York, 1996.

[7] R.R. Harris and G.K. Kandji, On use of minimum chi-square estimation, Roy. Statist. Soc. Ser. D (The Statistician), **32** (1983), No. 4, 379–394.

[8] T. Hauschild and M. Jentschel, Comparison of maximum likelihood estimation and chi-square statistics applied to counting experiments, Nuclear Intruments and Methods in Physics research A, **457** (2000), 384–401.

[9] I.A. Ibragimov and R.Z. Khasminskii, *Statistical Estimation. Asymptotic Theory*, Springer, New York, 1981.

[10] A.F. Karr, *Point Processes and Their Statistical Inference* (second edition), Marcel Dekker, New York, 1991.

[11] Yu.A. Kutoyants, Parameter estimation of intensity of inhomogeneous Poisson processes, Problems of Control and Information Theory, **8** (1979), 137–149.

[12] Yu.A. Kutoyants, Multidimensional parameter estimation of intensity function of inhomogeneous Poisson process, Problems of Control and Information Theory, **11** (1982), 325–334.

[13] Yu.A. Kutoyants, *Statistical Inference for Spatial Poisson Processes*, Lecture. Notes in Statistics 134, Springer–Verlag, New York, 1998.

[14] Yu.A. Kutoyants and V. Spokoiny, Optimal choice of observation window for Poisson observations, Statist. Probab. Lett., **61** (1999), 17–29.

[15] K.J. Mighell, Parameter estimation in the astronomy with Poisson-distributed data. I. The χ^2_γ statistic. The Astrophysical Journal, **518** (1999), 380–393.

[16] D.S. Moore, Generalised inverses, Wald's method and the construction of chi-squared tests of fit, J. Amer. Statist. Assoc., **72** (1977), 131–137.

[17] A. Wald, Tests of statistical hypothesis concerning several parameters when the number of observations is large. Amer. Math. Soc. Trans., **54** (1943), 426–482.

Elhadji Ousseynouu Accrachi
Université Gaston Berger,
UFR SAT, LERSTAD, Saint–Louis, Sénégal.
elhadjiousseynouaccrachi@yahoo.fr

Cherif Ahmat Tidiane Aïdara
Université Gaston Berger,
UFR SAT, LERSTAD, Saint–Louis, Sénégal;
University of The Gambia,
Dpt. Physical & Natural Sciences, Banjul, Gambia.
cataidara@utg.edu.gm

Ali S. Dabye
Université Gaston Berger,
UFR SAT, LERSTAD, Saint–Louis, Sénégal.
dabye_ali@yahoo.fr