ON AN ESTIMATION PROBLEM FOR TYPE I CENSORED SPATIAL POISSON PROCESSES

J. HURT, P. LACHOUT, * *Charles University, Prague*
D. PFEIFER, ** *Universität Hamburg*

Abstract

In this paper we consider the problem of estimating the intensity of a spatial homogeneous Poisson process if a part of the observations [quadrat counts] is censored. The actual problem has occurred during a court case when one of the authors was a referee for the defense.

TYPE I CENSORING; SPATIAL POISSON PROCESS; INTENSITY ESTIMATION
AMS CLASSIFICATION: PRIMARY: 62F12; SECONDARY: 62N25, 90B25

1. Introduction

In this paper we consider a spatial homogeneous Poisson process ξ with unknown intensity $\mu > 0$ which is to be estimated by quadrat counts. However, not all of the information is present; rather, only those quadrats are counted for which the number of points does not exceed a fixed number $K > 0$. Such a situation actually arose when one of the authors was a referee for the defense in a recent court case. Toner dust particles produced by a copy machine were counted on a critical document by electron microscopy, but the intensity of the underlying Poisson process was estimated on the basis of type I censored quadrats only which did not contain more than $K = 4$ observed particles. [The non-mathematical reasoning of the laboratory for this kind of censoring was to exclude what they called "systematic errors".] Since the resulting underestimate for the intensity of the Poisson process was unfavorable for the accused, however, a "naive" bias correction for the estimate was suggested, which could be understood and accepted also by non-mathematically trained judges. In this paper, we want to show that this "naive" intensity estimator is related to the maximum-likelihood estimator for type I censored data, and investigate the asymptotic properties of these estimators.

---

* Postal address: Department of Finance and Insurance, Charles University, Sokolovska 83, CZ-18675 Prague 8, Czech Republic.
First author supported by Grant Agency of Czech Republic, Grant No. 201/97/1176; second author supported by project CEZ:J13/98:113 200 008 and by the Grant Agency of Czech Republic, Grant No. 201/99/0264.

** Postal address: Universität Hamburg, Institut für Mathematische Stochastik, Bundesstr. 55, D–20146 Hamburg, Germany. Supported partly by the *Verein zur Förderung der Versicherungswissenschaften in Hamburg e.V.*

2. A "naive" estimator for type I censored data

The original problem can be reformulated in mathematical terms as follows (see [1] or [3] for a survey on censorization): suppose $X_1, \ldots, X_n$ are independent Poisson distributed random variables with parameter $\lambda > 0$, distributed as X, corresponding to the counts $\xi(A_1), \ldots, \xi(A_n)$ in disjoint quadrats $A_1, \ldots, A_n$ of equal area $\lambda = \mu\, \mathrm{m}(A_i)$, where m denotes Lebesgue measure. Observed are the couples $(W_i, I_i),\ i = 1, \ldots, n$ with $W_i = \min\{X_i, K+1\}$ and $I_i = \mathbf{1}\{X_i \leq K\}$. Let $N = \sum_{i=1}^n I_i,\ S = \sum_{i=1}^n I_i W_i$. Obviously, the estimator

$$\hat{\lambda}_1 = \begin{cases} \dfrac{S}{N}, & N > 0 \\ \infty, & N = 0 \text{ (i.e. all observations are censored)} \end{cases}$$

for λ which was used in the court case is biased, with

$$E\big[\hat{\lambda}_1 \mid \hat{\lambda}_1 < \infty\big] = E[X \mid X \leq K] = \frac{\sum_{i=0}^{K-1} \dfrac{\lambda^{i+1}}{i!}}{\sum_{i=0}^{K} \dfrac{\lambda^i}{i!}} = \lambda - \frac{\lambda^{K+1}}{K! \sum_{i=0}^{K} \dfrac{\lambda^i}{i!}} < \lambda - \frac{\lambda^{K+1}}{K!} e^{-\lambda} < \lambda.$$

Let for abbreviation denote

$$a(\lambda, K) = E[X \mid X \leq K],\ b(\lambda, K) = E[X \mid X > K].$$

Then obviously,

$$b(\lambda, K) = \lambda \frac{\sum_{i=K}^{\infty} \dfrac{\lambda^i}{i!}}{\sum_{i=K+1}^{\infty} \dfrac{\lambda^i}{i!}} = \lambda + \frac{\lambda^{K+1}}{K! \sum_{i=K+1}^{\infty} \dfrac{\lambda^i}{i!}} > \lambda + \frac{\lambda^{K+1}}{K!} e^{-\lambda} > \lambda,$$

and

$$a(\lambda, K) P(X \leq K) + b(\lambda, K) P(X > K) = \lambda.$$

The "naive" estimator $\hat{\lambda}_2$ under censoring is then given implicitly as solution of the equation

$$\frac{S + (n-N) b(\lambda, K)}{n} = \lambda. \tag{1}$$

The idea behind the "naive" estimator is as follows: S gives the number of observed particles without censoring, and $(n-N)b(\lambda, K)$ is close to the expected value for the number of censored particles when λ is known, so $S + (n-N)b(\lambda, K)$ is approximately equal to the total number of counted particles, say T, with T/n being the "classical" unbiased estimate for λ. Note that in general, there is no explicit solution for $\hat{\lambda}_2$; however, since $b(\lambda, K)$ is strictly increasing with λ and convex with $\lim_{\lambda \to \infty} b(\lambda, K)/\lambda = 1,\ \lim_{\lambda \to 0} b(\lambda, K) = 0$, there is always a unique positive solution for λ in case of $N > 0$, which can easily be calculated using e.g. computer algebra systems. For $N = 0$, i.e. all observations are censored, we put $\hat{\lambda}_2 = \infty$.

In the case of toner dust particles, $S = 159$ uncensored particles were reported within $N = 234$ quadrats of 0.5 mm^2 each, with a total number of $n = 240$ quadrats. From these figures, one obtains

$$\hat{\lambda}_1 = 0.679487, \quad \hat{\lambda}_2 = 0.791128$$

which means that $\hat{\lambda}_2$ [and therefore possibly also λ] is actually considerably larger than $\hat{\lambda}_1$.

3. The maximum-likelihood approach

Let us first observe that for the joint distributions of W_i and I_i, we have in general

$$P(W_i = w, I_i = 1) = \begin{cases} P(X_i = w) & \text{for } w \leq K \\ 0 & \text{otherwise,} \end{cases}$$

$$P(W_i = w, I_i = 0) = \begin{cases} P(X_i > K) & \text{for } w = K + 1 \\ 0 & \text{otherwise.} \end{cases}$$

The frequency function f - i.e. the counting density - for each pair (W_i, I_i) is hence given by

$$f(w, j) = \left[P_\lambda(X_i = w)\right]^j \left[P_\lambda(X_i > K)\mathbf{1}\{w = K+1\}\right]^{1-j}, \; w = 0, \ldots, K+1, \; j \in \{0, 1\},$$

where P_λ denotes the underlying parametric probability measure. Given the observations $W_i = w_i, \; I_i = j_i, \; i = 1, \ldots, n$, the likelihood function $L(\lambda; \mathbf{w}, \mathbf{j})$ with $\mathbf{w} = (w_1, \ldots, w_n), \; \mathbf{j} = (j_1, \ldots, j_n)$ can hence be written as

$$\begin{aligned} L(\lambda; \mathbf{w}, \mathbf{j}) &= \prod_{i=1}^{n} f(w_i, j_i) = \prod_{i=1}^{n} \left[P_\lambda(X_i = w_i)\right]^{j_i} \left[P_\lambda(X_i > K)\mathbf{1}\{w_i = K + 1\}\right]^{1-j_i} \\ &= \prod_{i \in \mathrm{K}} P_\lambda(X_i = w_i) \prod_{i \in \mathrm{K}^c} P_\lambda(X_i > K) \end{aligned}$$

or, in terms of random variables,

$$L(\lambda; \mathbf{W}, \mathbf{I}) = G(\lambda, K)^{n-N} e^{-N\lambda} \lambda^S \prod_{W_i \leq K} (W_i!)^{-1}$$

with the index set $\mathrm{K} = \{i \mid w_i \leq K, \; i = 1, \ldots, n\} = \{i \mid j_i = 1, \; i = 1, \ldots, n\}$, survival function $G(\lambda, K) = P_\lambda(X_i > K)$ and $\mathbf{W} = (W_1, \ldots, W_n), \; \mathbf{I} = (I_1, \ldots, I_n)$. Note that $\sum_{W_i \leq K} W_i = S$ and $\sum_{W_i \leq K} I_i = N$.

To maximize L, it thus suffices to maximize $\ln L$ which is up to a term independent of λ equal to

$$L^*(\lambda; \mathbf{W}, \mathbf{I}) = (n - N) \ln G(\lambda, K) - N\lambda + S \ln \lambda,$$

with

$$\frac{\partial}{\partial\lambda}L^*(\lambda;\mathbf{W},\mathbf{I}) = (n-N)\frac{G'(\lambda,K)}{G(\lambda,K)} - N + \frac{S}{\lambda} = (n-N)\frac{b(\lambda,K)}{\lambda} - n + \frac{S}{\lambda}.$$

Here we have used the relationship

$$G'(\lambda,K) = G(\lambda,K-1) - G(\lambda,K), \text{ i.e. } \frac{G'(\lambda,K)}{G(\lambda,K)} = \frac{b(\lambda,K)}{\lambda} - 1.$$

Equating the partial derivative of the log-likelihood function to zero hence gives the equivalent expression

$$\frac{S+(n-N)b(\lambda,K)}{n} = \lambda, \tag{2}$$

which corresponds precisely to equation (1). The corresponding maximum-likelihood estimator is thus identical to the naive estimator above.

4. Asymptotic properties of the (naive) ML-estimator

In this section we study the asymptotic distribution for the properly normalized estimator sequences obtained form equations (1) and (2). The argumentation here follows closely the general scheme introduced in [2]. Let

$$F(\lambda,K) = P(X_i \le K) = e^{-\lambda}\sum_{j=0}^{K}\frac{\lambda^j}{j!},$$

$$G(\lambda,K) = P(X_i > K) = 1 - F(\lambda,K) = e^{-\lambda}\sum_{j=K+1}^{\infty}\frac{\lambda^j}{j!},$$

$$H(\lambda,S,N) = (n-N)\frac{b(\lambda,K)}{\lambda} - n + \frac{S}{\lambda} = (n-N)\frac{F(\lambda,K)-F(\lambda,K-1)}{G(\lambda,K)} - N + \frac{S}{\lambda}.$$

Note that here $H(\lambda,S,N) = \frac{\partial}{\partial\lambda}L^*(\lambda;\mathbf{W},\mathbf{I})$ is just the partial derivative of the log-likelihood function above. Further,

$$E(N) = n\,F(\lambda,K),\ E(S) = n\,\lambda F(\lambda,K-1).$$

For the sequel, let λ_0 denote the true underlying Poisson parameter. Then

$$\frac{1}{n}E\big[H(\lambda,S,N)\big] = \frac{G(\lambda_0,K)}{G(\lambda,K)}\big(F(\lambda,K)-F(\lambda,K-1)\big) - \big(F(\lambda_0,K)-F(\lambda_0,K-1)\big),$$

which is independent of n. Denote this expression as $H^*(\lambda,\lambda_0,K)$. In particular,

$$E\big[H(\lambda_0,S,N)\big] = H^*(\lambda_0,\lambda_0,K) = 0.$$

By the SLLN, we also have

$$\lim_{n\to\infty}\frac{1}{n}H(\lambda,S,N) = H^*(\lambda,\lambda_0,K) \quad \text{a.s.}$$

Since the functions $H(\lambda, S, N)$ and $H^*(\lambda, \lambda_0, K)$ are decreasing and continous in λ with $H^*(\lambda_0, \lambda_0, K) = 0$, we further have $\lim_{n\to\infty} \hat{\lambda}_2 = \lambda_0$ a.s., i.e. the (naive) ML-estimator sequence is strongly consistent. Next, we see that $H^*(\lambda, \lambda_0, K)$ is differentiable w.r.t. λ, giving

$$\frac{\partial}{\partial\lambda} H^*(\lambda, \lambda_0, K) = -\frac{\lambda_0}{\lambda^2} F(\lambda_0, K-1) - \left(\frac{F(\lambda, K) - 2F(\lambda, K-1) + F(\lambda, K-2)}{G(\lambda, K)} + \frac{\big(F(\lambda, K) - F(\lambda, K-1)\big)^2}{G^2(\lambda, K)}\right) G(\lambda_0, K) = H^{**}(\lambda, \lambda_0, K).$$

This allows for a stochastic expansion of $H(\lambda, S, N)$ as

$$H(\lambda, S, N) = \sum_{i=1}^{n} Y_i + Z_n(\lambda) + n(\lambda - \lambda_0)\left(H^{**}(\lambda_0, \lambda_0, K) + R_n(\lambda)\right),$$

where

$$\begin{aligned} Y_i &= -I_i + \frac{W_i I_i}{\lambda_0} + \frac{F(\lambda_0, K) - F(\lambda_0, K-1)}{G(\lambda_0, K)}(1 - I_i), \\ Z_n(\lambda) &= \left(\frac{1}{\lambda} - \frac{1}{\lambda_0}\right)(S - n\lambda_0 F(\lambda_0, K-1)) \\ &\quad + \left(\frac{F(\lambda, K) - F(\lambda, K-1)}{G(\lambda, K)} - \frac{F(\lambda_0, K) - F(\lambda_0, K-1)}{G(\lambda_0, K)}\right)\big(n - N - nG(\lambda_0, K)\big), \end{aligned}$$

and $R_n(\lambda) \to 0$ for $\lambda \to \lambda_0$ a.s. Observe $E(Y_i) = 0$ and

$$\begin{aligned} Var(Y_i) &= E(Y_i^2) \\ &= E(I_i) + \frac{E(W_i^2 I_i)}{\lambda_0^2} + \frac{\big(F(\lambda_0, K) - F(\lambda_0, K-1)\big)^2}{G^2(\lambda_0, K)} E(1 - I_i) - \frac{2}{\lambda_0} E(W_i I_i) \\ &= F(\lambda_0, K) + \frac{\lambda_0 F(\lambda_0, K-1) + \lambda_0^2 F(\lambda_0, K-2)}{\lambda_0^2} \\ &\quad + \frac{\big(F(\lambda_0, K) - F(\lambda_0, K-1)\big)^2}{G^2(\lambda_0, K)} G(\lambda_0, K) - \frac{2\lambda_0 F(\lambda_0, K-1)}{\lambda_0} \\ &= F(\lambda_0, K) + \left(\frac{1}{\lambda_0} - 2\right) F(\lambda_0, K-1) + F(\lambda_0, K-2) \\ &\quad + \frac{\big(F(\lambda_0, K) - F(\lambda_0, K-1)\big)^2}{G(\lambda_0, K)} = -H^{**}(\lambda_0, \lambda_0, K) > 0. \end{aligned} \tag{3}$$

From the CLT, we finally obtain

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i \xrightarrow[n\to\infty]{d} \mathcal{N}(0, Var(Y_1)), \qquad \frac{1}{\sqrt{n}} Z_n(\lambda) \xrightarrow[n\to\infty]{d} 0.$$

Hence,

$$\sqrt{n}(\hat{\lambda}_2 - \lambda_0) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i + \frac{1}{\sqrt{n}} Z_n(\lambda)}{H^{**}(\lambda_0, \lambda_0, K) + R_n(\lambda)} \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \sigma^2)$$

where

$$\sigma^2 = \frac{Var(Y_1)}{H^{**}(\lambda_0, \lambda_0, K)^2} = \frac{1}{Var(Y_1)}$$

(cf. equation (3)).

5. Simulation studies

The following table contains some results of simulation studies, which were performed with a Poisson intensity of $\lambda_0 = 0.8$. 100 and 500 samples each of size 240 were generated, with censoring at the level $K = 4$. The table contains the empirical quantites corresponding to the estimates $\hat{\lambda}_2$.

empirical	sample size 100	sample size 500
mean	0.800751	0.794921
variance	0.00326	0.00344
standard deviation	0.0571	0.0587
skewness	0.1357	0.0941
kurtosis	2.472	2.901
s.e. of sample mean	0.00571	0.00262
median	0.794053	0.79833
1^{st} quartile	0.76487	0.7542
3^{rd} quartile	0.84834	0.8357
min	0.675	0.6167
max	0.93823	0.9924

Note that by relation (3), the corresponding asymptotic standard deviation $s = \sigma/\sqrt{n}$ with $n = 240$ is here given by $s = 0.057743$ which is quite close to the observed standard deviations of 0.0571 and 0.0587, resp. On the basis of the normal approximation for the estimator $\hat{\lambda}_2$ we thus obtain an asymptotic 95% confidence interval for the true λ_0 in the initial example as $0.791128 \pm 1.96 * 0.057743$ or (0.6795, 0.9043). Note that the lower interval value here is even larger than the originally used estimate $\hat{\lambda}_1$.

References

[1] T.R. Fleming, D.P. Harrington: *Counting Processes and Survival Analysis.* Wiley, N.Y. 1991.

[2] L.A. Klimko and P.I. Nelson: *On conditional least-squares estimation for stochastic processes.* Ann.Statist. 6,3 (1978), 629 - 642.

[3] S.I. Resnick: *Adventures in Stochastic Processes.* Birkhäuser, Boston 1992.