Consistent parametric estimation of the intensity of a spatial-temporal point process.

Frederic Paik Schoenberg

Department of Statistics, University of California, Los Angeles, CA 90095–1554, USA.

phone: 310-794-5193

fax: 310-472-3984

email: frederic@stat.ucla.edu

Postal address: UCLA Dept. of Statistics

8142 Math-Science Building

Los Angeles, CA 90095–1554, USA.

Abstract

We consider conditions under which parametric estimates of the intensity of a spatial-temporal point process are consistent. Although the actual point process being estimated may not be Poisson, an estimate involving maximizing a function that corresponds exactly to the log-likelihood if the process is Poisson is consistent under certain simple conditions. A second estimate based on weighted least squares is also shown to be consistent under quite similar assumptions. The conditions for consistency are simple and easily verified, and examples are provided to illustrate the extent to which consistent estimation may be achieved. An important special case is when the point processes being estimated are in fact Poisson, though other important examples are explored as well.

Key words: maximum likelihood estimation, intensity function, weighted least squares estimation, consistency, Poisson process, conditional intensity.

1 Introduction.

Maximum likelihood estimates (MLEs) have been extensively used in point process inference for decades, at least partly because of their known asymptotic properties. The consistency and asymptotic normality of the maximum likelihood estimate of the intensity of a stationary point process on the line are elementary (see e.g. Cox and Lewis 1966), and for the parameters governing the conditional intensity of an arbitrary stationary point process on the line, the consistency, asymptotic normality and efficiency of the MLE were proven by Ogata (1978).

There have since been a host of similar proofs, generalizing the important results in

Ogata (1978) to more general point processes under various conditions, of which we name a few. The case of non-stationary Poisson processes on the line was investigated by Kutoyants (1984) and more recently by Helmers and Zitikis (1999), and nice summaries of results for general non-stationary point processes on the line were given by Karr (1986) and Andersen et al. (1993). Regarding higher-dimensional point processes, conditions for the consistency and asymptotic normality of the MLE were derived by Brillinger (1975) for stationary multivariate Poisson processes, by Rathbun and Cressie (1994) for the case of non-stationary Poisson processes in $\mathbf{R}^d$, by Krickeberg (1982) for such processes in locally compact Hausdorff spaces, by Nishayama (1995) for a class of sequential marked point processes, and by Rathbun (1996a) for non-stationary spatial-temporal point processes. Jensen (1993) derived the asymptotic normality of the MLE for spatial Gibbs point processes under conditions similar to those in Rathbun (1996a), and Rathbun (1996b) established conditions for consistent estimation in the case of a spatial modulated Poisson process with partially observed covariates. Using simulations, Huang and Ogata (1999) assessed the relative efficiency of the MLE, the maximum pseudo-likelihood estimator and an approximate maximum likelihood estimator for spatial processes with strong interactions.

The results above are very important since point processes are commonly modeled via their conditional intensities, with parameters estimated by maximum likelihood. However, in some cases one may wish to estimate the unconditional intensity or mean measure, i.e. the expected value of the conditional intensity, assuming it exists. (Hereafter we refer to the unconditional intensity simply as the *intensity*). The present paper explores the problem of parametric estimation of the intensity of an arbitrary simple spatial-temporal point process, keeping assumptions about the point process and its conditional intensity to a minimum.

Estimating point process intensities may be important in applications, for at least three reasons. First, fewer assumptions about the higher-order properties of the point process are required. The assumptions required for the proofs listed above of the consistency of the MLE for the conditional intensity of a point process are unfortunately quite stringent, involving multiple restrictions on the derivatives of the conditional intensity. These conditions can be extremely difficult to verify in applications, and, as we show in the succeeding Sections, are not required for the purpose of estimating the intensity consistently. Second, whereas the conditional intensity uniquely characterizes the finite-dimensional distributions of any simple point process (see e.g. Daley and Vere-Jones, 1988), the intensity uniquely determines the mean number of points such a process has in any measurable subset of its domain. Hence accurate estimation of the intensity is critical in cases where the mean behavior of a point process is of interest. In research on wildfires and earthquakes, for example, the estimation of background rates is overwhelmingly important in hazard estimation (see e.g. Ogata 1998, Peng 2003), and while the second-order properties (clustering, inhibition) are important especially for short-term hazard forecasts, it is often of interest to obtain background rate estimates that do not depend on the assumption of a particular model for the full conditional intensity. Third, in some cases the parametric form of the intensity may be more readily suggested than that of the conditional intensity. Often a functional form for the intensity may be inferred by examining nonparametric intensity estimates, such as those produced by smoothing the point process using kernels, splines, or wavelets (see e.g. Vere-Jones 1992, Brillinger 1998). With regard to point process models for wildfires, for instance, while second-order properties remain a subject of considerable debate, the estimation of the background rate is typically performed by simply smoothing over previous events; similar methods are used in seismology (Ogata 1998, Ogata et al. 2001, Peng 2003, Schoenberg 2003). Little formal justification has been given for these methods of background rate estimation, which seem sensible only if the process is approximately Poisson. Clarification of what is meant by "approximately Poisson", i.e. identification of conditions under which parametric estimates of background rates made under the assumption that processes are Poisson are in fact consistent, even when the processes studied are not Poisson, is a concern of the present paper.

The current paper explores two simple estimates of the intensity. The consistency of the Poisson maximum likelihood estimate (PMLE), defined as the MLE of the intensity if the process was Poisson, is demonstrated under more general and much simpler assumptions than those pertaining to the consistency of the MLE. Only a slight variant of these conditions is needed to establish the consistency of the weighted least squares estimator (WLSE) as well. The simplicity of these assumptions, which can readily be verified in applications, may greatly facilitate an analysis of when the PMLE and WLSE are consistent, and equally importantly, when they are not. Note that in the case where the point process being estimated is Poisson, the intensity and conditional intensity are the same, as are the PMLE and MLE; hence for this situation our results represent a proof of the consistency of the MLE under conditions that are easily verifiable, without restrictions on the derivatives of the intensity function.

The structure of this paper is as follows. After formally introducing the PMLE in Section 2, Section 3 summarizes previous results on the MLE and then gives simpler conditions and a simple proof of the consistency of the PMLE. Section 4 provides similar conditions for the consistency of the WLSE. Several examples and counterexamples are given in Section 5 to demonstrate the need for the conditions in the previous Sections and to clarify under what conditions consistent estimation of the intensity is achievable, and Section 6 summarizes the results and lists some directions for future research.

2 Preliminaries

Following Brémaud (1981), we consider a spatial-temporal point process to be a measurable mapping from a filtered probability space $(\Omega, \mathcal{F}, P)$ onto Φ, the collection of all boundedly finite counting measures on the spatial-temporal domain $\mathcal{S} \times [0, \infty)$. The filtration $\mathcal{F} = \{\mathcal{F}_t\}_{t \geq 0}$ is assumed to be increasing and right continuous, and the spatial domain $\mathcal{S}$ any measurable space equipped with measure $\mu_{\mathcal{S}}$ defined on the Borel subsets of $\mathcal{S}$. Let $\mathcal{B}$ denote the Borel subsets of space-time $\mathcal{S} \times [0, \infty)$. For any spatial-temporal subset $B \in \mathcal{B}$ the random variable $N(B)$ represents the number of points in B.

Let $\mu_{\mathbf{R}}$ denote Lebesgue measure on the real (time-) line, and let $\mu_{\mathcal{B}}$ denote the product measure $\mu_{\mathcal{S}} \times \mu_{\mathbf{R}}$ on space-time. Assuming it exists, an *$\mathcal{F}$-conditional intensity* of N may be defined as an integrable, non-negative, real-valued, $\mathcal{F}$-predictable process λ^* such that, for any other non-negative $\mathcal{F}$-predictable process $Y(s,t)$,

$$E \int_B Y(s,t) \lambda^*(s,t) d\mu_{\mathcal{B}} = E \int_B Y(s,t) dN,$$

for $B \in \mathcal{B}$. Let the *intensity* $\lambda(s,t)$ denote the expectation with respect to P of $\lambda^*(s,t)$, provided it exists.

In what follows we consider sequences of point processes, $\{N_T\}$, $T = 1, 2, \ldots$, where only the points of N_T occurring from time 0 to time T, over all of $\mathcal{S}$, may be observed. We assume throughout that each process N_T has a conditional intensity λ_T^* whose expectation λ_T exists and is known up to a fixed parameter vector θ, within a complete separable metric space Θ of possibilities.

The (partial) log-likelihood function for N_T is conventionally expressed in terms of the conditional intensity λ_T^* as:

$$\int_{\mathcal{S}} \int_0^T \log \lambda_T^*(s,t) dN_T(s,t) - \int_{\mathcal{S}} \int_0^T \lambda_T^*(s,t) d\mu_{\mathcal{B}}(s,t).$$

When a functional form for λ_T^* is known, the parameters governing λ_T^* are typically estimated using the MLE, i.e. the value of the parameters maximizing the log-likelihood function above.

If N_T is a Poisson process, then λ_T^* and λ_T are identical, so in this case the log-likelihood $L_T(\theta)$ may be written

$$\int_{\mathcal{S}} \int_0^T \log \lambda_T(s,t;\theta) dN_T(s,t) - \int_{\mathcal{S}} \int_0^T \lambda_T(s,t,\theta) d\mu_{\mathcal{B}}(s,t). \tag{1}$$

Hence the estimator $\hat{\theta}$ maximizing (1) may be called the Poisson maximum likelihood estimator (PMLE) of θ. In the next section we examine the case where $\hat{\theta}$ is used to estimate θ even though N may not be Poisson.

3 Asymptotic Properties of the PMLE

As mentioned in the introduction, several authors have proven the consistency and asymptotic normality of the MLE for the parameters governing the conditional intensity of a point process. These proofs typically proceed in standard fashion by writing a Taylor expansion of $dL_T(\theta)/d\theta_j$, where $L_T(\theta)$ is the log-likelihood function (1), yielding the approximation

$$\partial L_T(\hat{\theta})/\partial \theta_j \approx \partial L_T(\theta^*)/\partial \theta_j + \sum_{k=1}^{K} (\theta_k^* - \hat{\theta}_k) I_T^{jk}(\theta^*),$$

where θ_j is the jth coordinate of θ, θ^* is the true parameter vector being estimated, and $I_T^{jk}(\theta) = \partial^2 L_T(\theta)/\partial \theta_j \partial \theta_k$ is the jk element of the Fisher information matrix. The asymptotic results then follow by observing that $\partial L_T(\theta^*)/\partial \theta_j$ are local square integrable martingales and invoking the martingale central limit theorem. For details see e.g. Theorems VI.1.1 and VI.1.2 of Andersen et al., 1993.

Conditions are required, however, to ensure that the remainder terms in the Taylor approximation are negligible and that the assumptions for the martingale central limit theorem are met. For instance, Rathbun (1996a) considers conditions on the first and second partial derivatives of the conditional intensity and invokes a result of Sweeting (1980) to prove the consistency and asymptotic normality of the MLE. Andersen et al. (1993) consider conditions slightly stronger than those of Rathbun (1996a), including conditions on the third partial derivatives of the conditional intensity. However, even Rathbun's conditions can be difficult to check in applications, as noted on page 62 of Rathbun (1996a).

The proofs of Rathbun (1996a) and Andersen et al. (1993) for the consistency and asymptotic normality of the MLE for the parameters governing λ^* extend readily to the use of the PMLE as an estimate of the parameters governing λ. For estimating the parameters governing the conditional intensity λ^*, the MLE is generally substantially more efficient than the PMLE, which essentially discards information on the second and higher-order properties of the point process. Indeed, in many cases certain parameters governing λ^* are not even identifiable from the intensity, λ. However, for the simpler case of estimating only the parameters governing λ, fewer conditions are needed and much simpler machinery is required. Below we present assumptions for the consistency of the PMLE which we believe are considerably simpler and easier to verify than those referred to above.

Let θ^* denote the true value of the parameter vector being estimated, and let $\hat{\theta}_T$ denote the PMLE of θ^*. Given any value of θ^*, we assume that there exists a function $\phi(T)$ such that the following conditions hold for sufficiently large T:

(A1) The parameter space Θ admits a finite partition of compact subsets $\Theta_T^1, ..., \Theta_T^J$ such that $\lambda_T(s,t;\theta)$ is continuous as a function of θ within each subset Θ_T^j.

(A2) For all $\theta \in \Theta$, $V\left[\int_{\mathcal{S}}\int_0^T \log\lambda_T(s,t;\theta)dN_T(s,t)\right] = o\left(\phi(T)^2\right)$.

(A3) Given any neighborhood U of θ^*, there exists $\gamma_1 > 0$ so that for all sufficiently large T, there is a subset of $\mathcal{S} \times [0,T]$ of $\mu_{\mathcal{B}}$-measure at least $\gamma_1\phi(T)$ on which $\lambda_T(s,t;\theta^*)$ and $|\log\lambda_T(s,t;\theta^*) - \log\lambda_T(s,t;\theta)|$ are uniformly bounded away from zero on $\Theta \cap U^c$, the complement of U.

Note that condition (A1) implies that the parameter space Θ is compact; this and the continuity property in (A1) ensure that a maximum of the Poisson likelihood (1) exists within Θ (see e.g. 4.16 of Rudin, 1976). We also note in passing that assumption (A1) allows λ_T to contain a finite number of discontinuities as a function of θ and that for the case where the parameter space Θ contains only finitely many elements, the continuity condition in (A1) is not required for Theorem 3.1 below. Assumption (A2) controls the variance of the process N. Assumption (A3) ensures that, on a sufficiently large portion of space-time, $\lambda_T(s,t;\theta^*)$ is positive and sufficiently distinct from $\lambda_T(s,t;\theta)$ for θ outside a neighborhood of θ^*. (A3) precludes the case, for example, where $\lambda(s,t;\theta)$ does not depend on θ at all, or where $\lambda(s,t;\theta^*) = 0$ everywhere. Section 5 provides examples to illustrate why these assumptions are needed to establish the consistency of the PMLE.

Theorem 3.1. Under assumptions A1-A3, the PMLE $\hat{\theta}$ is consistent.

Proof.

Consider the value θ^* fixed. We seek to show that $\forall \epsilon > 0$, for any neighborhood U of θ^*, for all sufficiently large T,

$$P\left(\hat{\theta}_T \notin U\right) < \epsilon. \tag{2}$$

Fix $\epsilon > 0$ and U. We first show that there exists $\delta > 0$ such that for sufficiently large T,

$$EL_T(\theta^*)/\phi(T) - \sup_{\theta \in U^c} EL_T(\theta)/\phi(T) \geq \delta, \tag{3}$$

where now $L_T(\theta)$ is defined by (1).

Observe that

$$\begin{aligned}
EL_T(\theta^*) \quad & - \quad \sup_{\theta \in U^c} EL_T(\theta) \\
&= E \int_{\mathcal{S}} \int_0^T \log \lambda_T(s,t;\theta^*) dN_T(s,t) - \int_{\mathcal{S}} \int_0^T \lambda_T(s,t;\theta^*) d\mu_{\mathcal{S}}(s) dt \\
&\quad - \sup_{\theta \in U^c} \left\{ E \int_{\mathcal{S}} \int_0^T \log \lambda_T(s,t;\theta) dN_T(s,t) - \int_{\mathcal{S}} \int_0^T \lambda_T(s,t;\theta) d\mu_{\mathcal{S}}(s) dt \right\} \\
&= \int_{\mathcal{S}} \int_0^T \log \lambda_T(s,t;\theta^*) \lambda_T(s,t;\theta^*) d\mu_{\mathcal{S}}(s) dt - \int_{\mathcal{S}} \int_0^T \lambda_T(s,t;\theta^*) d\mu_{\mathcal{S}}(s) dt \\
&\quad - \sup_{\theta \in U^c} \left\{ \int_{\mathcal{S}} \int_0^T \log \lambda_T(s,t;\theta) \lambda_T(s,t;\theta^*) d\mu_{\mathcal{S}}(s) dt - \int_{\mathcal{S}} \int_0^T \lambda_T(s,t;\theta) d\mu_{\mathcal{S}}(s) dt \right\} \\
&= \inf_{\theta \in U^c} \left\{ \int_{\mathcal{S}} \int_0^T \lambda_T(s,t;\theta^*) \left[\log \lambda_T(s,t;\theta^*) - \log \lambda_T(s,t;\theta) - 1 + \frac{\lambda_T(s,t;\theta)}{\lambda_T(s,t;\theta^*)} \right] d\mu_{\mathcal{S}}(s) dt \right\} \\
&= \inf_{\theta \in U^c} \left\{ \int_{\mathcal{S}} \int_0^T \lambda_T(s,t;\theta^*) \left[\exp(\psi(s,t,\theta)) - \psi(s,t,\theta) - 1 \right] d\mu_{\mathcal{S}}(s) dt \right\},
\end{aligned}$$

where $\psi(s,t,\theta) = \log \lambda_T(s,t;\theta) - \log \lambda_T(s,t;\theta^*)$.

Assumption (A3) ensures that there exist some positive constants $\gamma_1, \gamma_2, \gamma_3$ such that for sufficiently large T, for all s,t in a subset of $\mathcal{S} \times [0,T]$ with measure at least $\gamma_1 \phi(T)$, $\lambda_T(s,t;\theta^*) > \gamma_2$ and $|\psi(s,t,\theta)| > \gamma_3$. Let $\gamma_4 = \min\{\exp(\gamma_3) - \gamma_3 - 1, \exp(\gamma_3) + \gamma_3 - 1\}$. Recalling that $\gamma_3 > 0$ and that the inequality $\exp(x) \geq x+1$ has equality iff. $x = 0$, (see e.g. Abramowitz 1964), it follows that $\gamma_4 > 0$.

Hence, for sufficiently large T, $EL_T(\theta^*) - \sup_{\theta \in U^c} EL_T(\theta) \geq \gamma_1\gamma_2\gamma_4\phi(T)$, which establishes (3) for $\delta = \gamma_1\gamma_2\gamma_4$.

With $\Theta_T^1, ..., \Theta_T^J$ defined as in assumption (A1), fix J elements $\theta^1 \in \Theta_T^1, ..., \theta^J \in \Theta_T^J$. By assumption (A2), for each such value θ^j,

$$V\left[\frac{L_T(\theta^j)}{\phi(T)}\right] = V\left[\frac{1}{\phi(T)} \int_{\mathcal{S}} \int_0^T \log \lambda_T(s,t;\theta^j) dN_T\right] \to 0.$$

Thus for each j, $[L_T(\theta^j) - EL_T(\theta^j)]/\phi(T)$ has mean zero and variance converging to zero, so from Chebyshev's inequality

$$\frac{L_T(\theta^j) - EL_T(\theta^j)}{\phi(T)} \underset{T\to\infty}{\overset{p}{\to}} 0. \tag{4}$$

$$\tag{5}$$

Since by assumption (A1) the function $\lambda_T(s,t;\theta)$ is continuous with respect to θ on Θ_T^j, so is the function

$$\frac{L_T(\theta) - EL_T(\theta)}{\phi(T)} = \frac{\int_{\mathcal{S}} \int_0^T \log \lambda_T(s,t;\theta) dN_T - \int_{\mathcal{S}} \int_0^T \log \lambda_T(s,t;\theta) \lambda_T(s,t;\theta^*) d\mu_{\mathcal{S}}(s) dt}{\phi(T)}.$$

Thus the compactness of Θ_T^j implies that $[L_T(\theta) - EL_T(\theta)]/\phi(T) \underset{T\to\infty}{\overset{p}{\to}} 0$ uniformly on Θ_T^j. Since $\Theta = \bigcup_{j=1}^J \Theta_T^j$, $[L_T(\theta) - EL_T(\theta)]/\phi(T) \underset{T\to\infty}{\overset{p}{\to}} 0$ uniformly on all of Θ.

Hence there is a $\delta > 0$ such that for sufficiently large T,

$$P\left(\sup_{\theta}[L_T(\theta) - EL_T(\theta)]/\phi(T) \geq \delta/2\right) < \epsilon/2. \tag{6}$$

Let $\check{\theta}_T$ denote a (possibly non-unique) value of θ maximizing $L_T(\theta)$ among $\theta \in U^c$, i.e. $L_T(\check{\theta}_T) \geq L_T(\theta), \forall\theta \notin U$. Putting together (3) and (6) yields, for sufficiently large T,

$$\begin{aligned}
P\left(\hat{\theta}_T \notin U\right) &= P\left(L_T(\check{\theta}_T) \geq \sup_{\theta \in U} L_T(\theta)\right) \\
&\leq P\left(L_T(\check{\theta}_T) \geq L_T(\theta^*)\right) \\
&\leq P\left(L_T(\check{\theta}_T) - EL_T(\check{\theta}_T) \geq \delta\phi(T)/2\right) + P\left(EL_T(\check{\theta}_T) - EL_T(\theta^*) > -\delta\phi(T)\right) \\
&\quad + P\left(EL_T(\theta^*) - L_T(\theta^*) \geq \delta\phi(T)/2\right) \\
&< \epsilon/2 + 0 + \epsilon/2,
\end{aligned}$$

establishing (2).

Assumptions (A1-A4) are by no means minimal, but they are quite straightforward to verify, in contrast to the conditions in previous results regarding maximum likelihood estimation. In particular, no conditions on the derivatives of λ are required. We remark below in particular about two of the three assumptions for Theorem 3.1.

Remark 3.2. Like previous authors (e.g. Ogata 1978, Andersen et al. 1993, Rathbun and Cressie 1994, Rathbun 1996a), we assume that the parameter space Θ is compact; this assumption is implicit in (A1). Note however that in certain cases the proof of consistency in Theorem 3.1 remains valid even when the parameter space is not compact. In particular, assumption (A1) may be discarded for the purposes of Theorem 3.1 if it may be shown directly that the parameter space Θ contains compact subsets Θ_T with $P(\hat{\theta}_T \notin \Theta_T) \underset{T\to\infty}{\to} 0$, with λ_T continuous as a function of θ on Θ_T; in such cases one may further replace Θ with Θ_T in conditions (A2) and (A3). This feature may be relevant in applications where often it is quite natural to consider the domain for each estimated parameter to be the whole real line $\mathbf{R}$ or in some cases the half-line $\mathbf{R}^+$, rather than some compact subset thereof.

Remark 3.3. Note that assumption (A3) is quite a bit stronger than what is minimally necessary for Theorem 3.1. (A3) is only used in the proof of relation (3) and hence may be discarded for processes where this inequality can be proven directly.

4 Weighted Least Squares Estimates

The parameters θ^* governing the intensity of a spatial-temporal point process can alternatively be estimated by weighted least squares (WLS). Here the estimator $\tilde{\theta}_T$ is chosen to minimize the quadratic variation:

$$Q_T(\theta) = \sum_{i=1}^{I_T} w_i^T \left[N_T(B_i^T) - E\{N_T(B_i^T); \theta\} \right]^2, \tag{7}$$

where for given T, the sets $\{B_1^T, \ldots, B_{I_T}^T\}$ form a partition of the product space $\mathcal{S} \times [0, T]$, and the weights w_i^T are non-negative constants. Often in practice the weights are chosen so that w_i^T is inversely proportional to an estimate of the variance of $N_T(B_i^T)$. Here $E\{N_T(B_i^T); \theta\} = \int_{B_i^T} \lambda(s, t; \theta) d\mu_{\mathcal{S}}(s) dt$; with this notation, $EN_T(B_i^T) = E\{N_T(B_i^T); \theta^*\}$. For simplicity, assume that for each T, the number of bins I_T in the partition is finite.

We consider the following replacements for assumptions (A2-A3):

(B2) For all θ in Θ, $\max_i V \left[\int_{B_i^T} dN_T \right] = o\left(\phi(T)^2\right)$.

(B3) Given any neighborhood U of θ^*, there exist constants $\nu_1, \nu_2, \nu_3 > 0$ so that for sufficiently large T, a fraction of at least $\nu_1 \phi(T)$ of the bins B_i^T have product measure $\mu_{\mathcal{B}}(B_i^T)$

at least $\nu_2/\sqrt{w_i^T}$ and the property that either $\lambda_T(s,t;\theta) - \lambda_T(s,t;\theta^*) > \nu_3$ or $\lambda_T(s,t;\theta) - \lambda_T(s,t;\theta^*) < -\nu_3$ for all $s,t \in B_i^T$ and all $\theta \in U^c$.

Assumption (B2) guarantees that the process N is not too volatile. Like (A3), assumption (B3) ensures that outside neighborhoods U of θ^*, λ_T is uniformly bounded away from its value at θ^* within a sufficient fraction of adequately-sized (and adequately-weighted) bins. As with assumptions (A2-A3), these assumptions are relatively easy to verify.

Theorem 4.1. Assuming (A1) and (B2-B3), the WLSE $\tilde{\theta}_T$ is consistent.

Proof.

$$\begin{aligned}
Q_T(\theta) &= \sum_i w_i^T \left[N_T(B_i^T) - E\{N_T(B_i^T);\theta\}\right]^2 \\
&= \sum_i w_i^T \left[N_T(B_i^T)^2 - 2N_T(B_i^T)E\{N_T(B_i^T);\theta\} + (E\{N_T(B_i^T);\theta\})^2\right].
\end{aligned}$$

Taking expectations yields

$$EQ_T(\theta) = \sum_i w_i^T \left[EN_T(B_i^T)^2 - 2EN_T(B_i^T)E\{N_T(B_i^T);\theta\} + (E\{N_T(B_i^T);\theta\})^2\right]. \tag{8}$$

Fix θ^* and a neighborhood U around it. Letting $\delta = \nu_1\nu_2^2\nu_3^2 > 0$, from (8) and (B4) one obtains, for sufficiently large T,

$$\begin{aligned}
&\inf_\theta \frac{1}{I_T\phi(T)} \left[EQ_T(\theta) - EQ_T(\theta^*)\right] \\
&\quad= \inf_{\theta\in U^c} \frac{1}{I_T\phi(T)} \sum_i w_i^T \left[(E\{N_T(B_i^T);\theta\})^2 - 2EN_T(B_i^T)E\{N_T(B_i^T);\theta\} + (EN_T(B_i^T))^2\right] \\
&\quad= \inf_{\theta\in U^c} \frac{1}{I_T\phi(T)} \sum_i w_i^T \left[E\{N_T(B_i^T);\theta\} - EN_T(B_i^T)\right]^2 \\
&\quad= \inf_{\theta\in U^c} \frac{1}{I_T\phi(T)} \sum_i w_i^T \left[\int_{t\in B_i} (\lambda_T(s,t;\theta) - \lambda_T(s,t;\theta^*))\, d\mu_{\mathcal{S}}(s)dt\right]^2 \\
&\quad\geq \frac{1}{I_T\phi(T)} \nu_1\phi(T)I_T[\nu_2\nu_3]^2 \\
&\quad= \delta.
\end{aligned} \tag{9}$$

Note that assumption (B2) implies that $V[Q_T(\theta)] = o\left(I_T^2\phi(T)^2\right)$ by the Cauchy-Schwarz inequality. Thus $[Q_T(\theta) - EQ_T(\theta)]/I_T\phi(T) \underset{T\to\infty}{\overset{p}{\to}} 0$ for all θ in Θ, and assumption (A1) ensures that this convergence is uniform over all θ in Θ, just as in Theorem 3.1.

Hence if $\grave{\theta}_T$ denotes the WLSE of θ among $\grave{\theta}_T \in U^c$, then for sufficiently large T,

$$
\begin{aligned}
P\left(\tilde{\theta}_T \notin U\right) &\leq P\left(Q_T(\grave{\theta}_T) \geq Q_T(\theta^*)\right) \\
&\leq P\left(Q_T(\grave{\theta}_T) - EQ_T(\grave{\theta}_T) \leq -\delta I_T\phi(T)/2\right) + P\left(EQ_T(\grave{\theta}_T) - EQ_T(\theta^*) < \delta I_T\phi(T)\right) \\
&\quad + P\left(EQ_T(\theta^*) - Q_T(\theta^*) \leq -\delta I_T\phi(T)/2\right) \\
&< \epsilon/2 + 0 + \epsilon/2,
\end{aligned}
$$

which completes the proof.

5 Examples and Counterexamples

Some examples may help to clarify when the conditions for consistent estimation are satisfied. Our first two examples consider the Poisson case, where $\lambda = \lambda^*$ and the PMLE and MLE are equivalent.

Example 5.1. Suppose N_T is a sequence of spatial-temporal versions of the cyclic Poisson process, studied for example by Helmers and Zitikis (1999) and Helmers et al. (2003). That is, suppose N_T is Poisson with separable intensity function

$$\lambda_T(s,t;\theta) = f(s;\theta)g(t;\theta), \tag{10}$$

where $f, g > 0$, Θ a compact subset of $\mathbf{R}^K$ for some positive integer K, and g is any integrable cyclic function with (possibly unknown) period τ, i.e. $g(t;\theta) = g(t + j\tau;\theta)$, for all t and any integer j. Let f and g be continuous in θ with $|\log fg|$ bounded for each θ by some constant

B_θ, and suppose $\alpha := \int_{\mathcal{S}} f(s;\theta^*)d\mu_{\mathcal{S}}(s) < \infty$ and $\beta := \int_0^\tau g(t;\theta^*)dt < \infty$. Finally, suppose that condition (A3) holds with $\phi(T) = T$; note for example that one only needs $f(s;\theta^*) > c_1 > 0$ for s in some non-null subset of $\mathcal{S}$, and $g(t;\theta^*) > c_2 > 0$ for t in some non-null subset of $[0,\tau)$, in order to ensure that $\lambda_T(s,t;\theta^*)$ is uniformly bounded away from zero on a subset of $\mu_{\mathcal{B}}$-measure at least $\gamma_1 T$. Note also that most parameterizations in which $\lambda_T(s,t;\theta)$ is monotonic in θ (e.g. an exponentially parameterized intensity, or a linearly parameterized intensity, as in Ex. 6.2 of Rathbun and Cressie) satisfy the condition in (A3) regarding a lower bound on $|\log\lambda_T(s,t;\theta^*) - \log\lambda_T(s,t;\theta)|$ for $\theta \in U^c$.

Assumption (A1) is obviously satisfied since f and g are continuous in θ over all of Θ, and since N_T is Poisson,

$$\begin{aligned}
V_T(\theta) &:= Var\left[\int_{\mathcal{S}}\int_0^T \log\lambda_T(s,t;\theta)dN_T(s,t)\right] \qquad (11)\\
&= E\left[\int_{\mathcal{S}}\int_0^T \log\lambda_T(s,t;\theta)dN_T(s,t)\right]^2 - \left[E\int_{\mathcal{S}}\int_0^T \log\lambda_T(s,t;\theta)dN_T(s,t)\right]^2\\
&= \int_{\mathcal{S}}\int_0^T [\log\lambda_T(s,t;\theta)]^2\lambda_T(s,t;\theta^*)d\mu_{\mathcal{S}}(s)dt + \left[\int_{\mathcal{S}}\int_0^T \log\lambda_T(s,t;\theta)\lambda_T(s,t;\theta^*)d\mu_{\mathcal{S}}(s)dt\right]^2\\
&\quad - \left[\int_{\mathcal{S}}\int_0^T \log\lambda_T(s,t;\theta)\lambda_T(s,t;\theta^*)d\mu_{\mathcal{S}}(s)dt\right]^2\\
&= \int_{\mathcal{S}}\int_0^T [\log\lambda_T(s,t;\theta)]^2\lambda_T(s,t;\theta^*)d\mu_{\mathcal{S}}(s)dt\\
&\leq B_\theta^2\alpha\beta(1+T/\tau)\\
&= o(T^2),
\end{aligned}$$

so condition (A2) is satisfied with $\phi(T) = T$.

Example 5.2. Let N_T be a sequence of spatial-temporal Poisson processes, but not necessarily cyclical or separable. Suppose Θ is a compact subset of $\mathbf{R}^K$, that $\log \lambda_T(s,t;\theta)$ is continuous in θ and bounded in absolute value by some constant $B_\theta < \infty$, and that the space $\mathcal{S}$ has finite, positive measure $\mu_{\mathcal{S}}(\mathcal{S})$. Finally, suppose that λ_T is parameterized such that for θ outside any neighborhood U of θ^*, $|\log \lambda_T(s,t;\theta^*) - \log \lambda_T(s,t;\theta)|$ is uniformly bounded away from zero.

Then

$$\begin{aligned} V_T(\theta) &\leq B_\theta^2 V\left[\int_{\mathcal{S}}\int_0^T dN_T(s,t)\right] \\ &= B_\theta^2 E N_T(\mathcal{S}\times[0,T]) \\ &\leq B_\theta^2 \exp(B_\theta) T \mu_{\mathcal{S}}(\mathcal{S}), \end{aligned}$$

so requirement (A2) is fulfilled with $\phi(T) = T$. (A1) is satisfied by assumption, and (A3) is satisfied since $\lambda_T(s,t;\theta^*) > \exp(-B_\theta^*) > 0$ on $\mathcal{S}\times[0,T]$ which has measure $T\mu_{\mathcal{S}}(\mathcal{S})$.

Example 5.3. Suppose that N_T are spatial-temporal versions of the Isham and Westcott (1979) self-correcting point process, as described by Rathbun (1996a). Such processes have conditional intensity

$$\lambda_T^*(s,t;\theta) = \exp\left[f(s,t;\theta) - cN_T\left\{b(s,r)\times[0,t)\right\}\right], \tag{12}$$

where $b(s,r)$ is a ball of radius r around location s, and $c > 0$. As in the previous example, suppose that Θ is compact and $0 < \mu_{\mathcal{S}}(\mathcal{S}) < \infty$, and that f is continuous in θ with $|f(s,t;\theta)| \leq B_{\theta^*} < \infty$ and $|f(s,t;\theta^*) - f(s,t;\theta)| \geq B_U > 0$ outside any neighborhood U of θ^*. Processes obeying (12) are called self-correcting since the more (fewer) points N happens to place near location s, the more the conditional intensity λ_T^* adjusts by decreasing (resp., increasing) the rate at which points accumulate near s thereafter. Hence the variance of $N_T\{b(s,r) \times [0,t)\}$ is actually smaller than that of the Poisson process with equivalent intensity (see Isham and Westcott 1979 or Vere-Jones and Ogata 1984), and (A2) and (A3) are easily satisfied with $\phi(T) = T$ as in Example 5.2.

Note that not all the parameters in the conditional intensity are in general identifiable by the intensity $\lambda(s,t) = \exp\{f(s,t)\}E[exp(-cN_T\{b(s,r) \times [0,t)\})]$, which is typically difficult to formulate analytically for these types of processes. The case where $f(s,t) = \alpha + \beta t$ is discussed e.g. by Ogata and Vere-Jones (1984). Even this simple case is non-stationary with λ non-convergent (see p. 337 of Isham and Westcott, 1979) and difficult to formulate analytically; λ is the derivative with respect to t of the mean function $\mu(t) := EN(\mathcal{S}, t)$, which is governed by equation (15) of Isham and Westcott (1979). We certainly do not suggest the PMLE in favor of the MLE for the case where the form of the self-correcting model given above is known. The purpose of this example is merely to show that, for the purpose of estimating the parameters θ governing the intensity λ, the process need not be Poisson in order for the PMLE to be consistent.

Example 5.4. As mentioned in Remark 3.2, in certain circumstances one may apply Theorem 3.1 even though the parameter space is not compact. A simple case is where $\Theta = \mathbf{R}$ and N is a stationary Poisson process with $\lambda(s,t;\theta) = \exp(\theta)$. For any T, there is positive probability that no points have yet been observed up to time T, in which case no maximum of the Poisson likelihood over $\mathbf{R}$ exists, i.e. the PMLE $\hat{\theta}_T$ is not defined. However, if $N(\mathcal{S} \times [0,T]) \geq 1$ then $\hat{\theta} = \log\{N(\mathcal{S} \times [0,T])\} - \log T \geq -\log T$. Thus, letting

$\Theta_T = [-\log T, \log T]$, one sees that

$$\begin{aligned} P(\hat{\theta}_T \notin \Theta_T) &= P\{N(\mathcal{S} \times [0,T]) = 0\} + P\{N(\mathcal{S} \times [0,T]) > T^2\} \\ &\leq P\{N(\mathcal{S} \times [0,T]) = 0\} + E\{N(\mathcal{S} \times [0,T])\}/T^2 \\ &= \exp\{-\exp(\theta^*)\mu_{\mathcal{S}}(\mathcal{S})T\} + \exp(\theta^*)\mu_{\mathcal{S}}(\mathcal{S})/T \\ &\underset{T\to\infty}{\longrightarrow} 0. \end{aligned}$$

Since λ is continuous in θ one may appeal to Remark 3.2, and (A2) and (A3) are readily satisfied on Θ_T with $\phi(T) = T$ since $V_T(\theta) = (\theta^*)^2 \exp(\theta^*)\mu_{\mathcal{S}}(\mathcal{S})T$ and since λ_T and $|\log \lambda_T(s,t;\theta^*) - \log \lambda_T(s,t;\theta)|$ are obviously uniformly bounded below on $\Theta_T \cap U^c$ for any neighborhood U of θ^*.

Example 5.5. For point processes with rapidly increasing intensities, assumption (A3) will typically not be satisfied, but in such cases one may appeal to Remark 3.3. For example, if N is Poisson and λ_T increases exponentially, i.e. if λ_T is separable as in (10) with $g(t) \propto \exp(\theta_K t)$, and with $\theta_K^* > 0$, f any continuous function of θ such that $0 < \int_{\mathcal{S}} f(s)d\mu_{\mathcal{S}}(s) < \infty$, and Θ compact, then (A1), (A2) and (3) are satisfied with $\phi(T) = \exp(\theta_K^* T)$. Indeed, as in Example 5.1, $V_T(\theta) = \int_{\mathcal{S}}\int_0^T [\log \lambda_T(s,t;\theta)]^2 \lambda_T(s,t;\theta^*)d\mu_{\mathcal{S}}(s)dt$, which now equals $\int_{\mathcal{S}}\int_0^T [\log f(s) + \theta_K t]^2 f(s)\exp(\theta_K^* t)d\mu_{\mathcal{S}}(s)dt = o(\{\exp(\theta_K^* T)\}^2)$, establishing (A2). From the equations following (3) in Theorem 3.1, for any neighborhood U of θ^*, and with $\gamma_4 > 0$ as in Theorem 3.1,

$$\begin{aligned} EL_T(\theta^*) - \sup_{U^c} EL_T(\theta) &\geq \inf_{\theta \in U^c} \left\{ \int_{\mathcal{S}} \int_0^T \lambda_T(s,t;\theta^*)[\gamma_4] d\mu_{\mathcal{S}}(s)dt \right\} \\ &= \int_{\mathcal{S}} \int_0^T f(s;\theta^*)\exp(\theta_K^* t)\gamma_4 d\mu_{\mathcal{S}}(s)dt \end{aligned}$$

$$= \gamma_4 \int_{\mathcal{S}} f(s) d\mu_{\mathcal{S}}(s) \{\phi(T) - 1\},$$

which immediately yields (3).

The next two examples illustrate limitations on the possibilities for consistent estimation of the intensity.

Example 5.6. Suppose N_T are observations on $\mathcal{S} \times [0, T]$ of a *finite* point process, N, i.e. a process that, with positive probability, may contain only finitely many points on $\mathcal{S} \times [0, \infty)$. In this case consistent estimation of θ is unachievable, as is well known (see for instance Example 6.3 of Rathbun and Cressie 1994). An example is when N is a spatial-temporal Poisson process with separable intensity as in (10), with $g(t; \theta) \propto \exp(\theta_i t)$, where $\theta_i < 0$. One may inquire which of the conditions (A1-A3) are not met in this case. Since the integral $\lim_{T \to \infty} \int_{\mathcal{S}} \int_0^T \lambda_T(s, t; \theta^*) < \infty$, the set on which λ_T is uniformly bounded away from zero must have finite product measure; hence $\phi(T) = O(1)$ in condition (A3). However, since N is a Poisson process $V_T(\theta)$ is non-decreasing as a function of T, so assumption (A2) is violated: $V_T(\theta)/\phi(T)^2 \not\to 0$ as $T \to \infty$.

Example 5.7. If λ_T has a change-point governed by a parameter in θ, then consistent estimation of θ is typically unachievable. For a simple example let N_T be Poisson with

$$\lambda_T(s, t; \theta) = \theta_1 \mathbf{1}_{\{t \le \theta_2\}} + \theta_3 \mathbf{1}_{\{t > \theta_2\}},$$

and where $0 < \mu_{\mathcal{S}}(\mathcal{S}) < \infty$. Assumption (A3) is violated, since for any ϵ with $0 < \epsilon < \theta_2^*$, $\lambda_T(s, t; (\theta_1^*, \theta_2^*, \theta_3^*)) = \lambda_T(s, t; (\theta_1^*, \theta_2^* - \epsilon, \theta_3^*))$ for $t > \theta_2^*$. Thus for any small neighborhood U

of θ^*, the set of (s,t) on which $|\log \lambda_T(s,t;\theta^*) - \log \lambda_T(s,t;\theta)|$ is uniformly bounded away from zero for $\theta \in U^c$ has $\mu_{\mathcal{B}}$-measure less than $\theta_2^* \mu_{\mathcal{S}}(\mathcal{S}) = O(1)$, which violates (A3) since $V_T(\theta) = O(T)$. In general if λ_T is governed by a parameter θ_i that does not affect $\lambda_T(s,t;\theta)$ on a set of (s,t) with infinite measure, then consistent estimation of θ_i is unachievable, so assumption (A3) is needed to ensure that such cases are excluded.

6 Discussion

Although maximum likelihood estimation for point processes is very common, examples of applications where the assumptions necessary to establish the consistency of the MLE are verified are rather elusive. The general impression among applied researchers appears to be that asymptotic properties of the MLE such as consistency and asymptotic normality apply quite generally, and that verification of these properties for particular point process models is difficult and unnecessary.

The aim of the current paper is to show that, by contast, the PMLE and NLSE may be used to estimate the intensity consistently in situations where one is unwilling to make restrictive assumptions on the higher-order properties of the process. We show conditions under which the estimates of the intensity parameters constructed by assuming the observed point process is Poisson are consistent, even when the process in fact not Poisson. Further, the assumptions required for consistency of the PMLE and NLSE may readily be verified in practice.

Our results involve conditions on the rate of increase of the variance of the point process. Since variances are rather easy to check and are easily interpretable, it is possible that these conditions may be checked in applications without explicit assumption of a parametric form for the conditional intensity of the point process, but rather by examining sample variances directly, or by appealing to subject matter information or knowledge of the mechanism driving the point process. By contrast, it is difficult to see how the applied researcher could justify assumptions involving the second (and higher-order) partial derivatives of the conditional intensity of the point process without specifying the conditional intensity in detail.

Although our assumptions may be easily verifiable, they are by no means minimal, nor are our results optimally strong. In particular, only consistent estimation is investigated here. Similar conditions under which estimates may be shown to be asymptotically normal and/or efficient are important subjects for further research; such conditions likely will require assumptions pertaining at least to the first derivatives of the condtional intensity.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 9978318. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. Coordinating editor Michael Stein, two anonymous referees, and Tom Ferguson provided several useful suggestions.

References

Abramowitz, M., 1964. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. M. Abramowitz and I. Stegun (Eds.), U.S. Government Printing Office, Washington D.C.

Andersen, P.K., Borgan, O., Gill, R.D., Keiding, N., 1993. Statistical Models Based on Counting Processes. Springer, NY.

Brémaud, P., 1981. Point Processes and Queues. Springer, NY.

Brillinger, D.R., 1998. Some wavelet analyses of point process data. In: 31st Asilomar Conference on Signals, Systems and Computers, IEEE Computer Society, Los Alamitos, California, 1087–1091.

Cox, D.R., Lewis, P.A.W., 1966. The Statistical Analysis of Series of Events. Methuen, London.

Daley, D., Vere-Jones, D., 1988. An Introduction to the Theory of Point Processes. Springer, NY.

Helmers, R., Mangku, W., Zitikis, R., 2003. Consistent estimation of cyclic Poisson intensity functions. J. Multiv. Anal. 84, 19–39.

Helmers, R., Zitikis, R., 1999. On estimation of Poisson intensity functions. Ann. Inst. Statist. Math. 51, 265–280.

Huang, F., Ogata, Y., 1999. Improvements of the maximum pseudo-likelihood estimators in various spatial statistical models. J. Comp. Graph. Statist. 8, 510–530.

Isham, V., Westcott, M., 1979. A self-correcting point process. Stoch. Proc. Appl. 8, 335–347.

Jensen, J.L., 1993. Asymptotic normality of estimates in spatial point processes. Scand. J. Statist. 20, 97–109.

Karr, A.F., 1986. Point Processes and their Statistical Inference. Marcel Dekker, NY.

Krickeberg, K., 1982. Processus ponctuels en statistique. In: Hennequin, P.L. (Ed.), Ecole d'Eté de Probabilités de Saint-Flour X, International Cooperative Publishing House, Fairland, MD, 205–313.

Kutoyants, Y.A., 1984. Parameter Estimation for Stochastic Processes. Heldermann Verlag, Berlin.

Nishiyama, Y., 1995. Local asymptotic normality of a sequential model for market point processes and its applications. Ann. Inst. Statist. Math. 47, 195–209.

Ogata, Y., 1978. The asymptotic behaviour of maximum likelihood estimators for stationary point processes. Ann. Inst. Statist. Math. 30, 243–261.

Ogata, Y., 1998, Space-time point process models for earthquake occurrences. Ann. Inst. Statist. Math. 50, 379–402.

Ogata, Y., Katsura, K., Tanemura, M. 2001. Modelling of Heterogeneous Space-Time Seismic Activity and its Residual Analysis. Research Memorandum 808, Institute of Statistical Mathematics.

Ogata, Y., Vere-Jones, D. 1984. Inference for earthquake modles: a self-correcting model. Stoch. Proc. Appl. 17, 337–347.

Peng, R.D., 2003. Applications of Multi-dimensional Point Process Methodology to Wildfire Hazard Assessment. Ph.D. thesis, Department of Statistics, University of California, Los Angeles.

Rathbun, S.L., 1996a. Asymptotic properties of the maximum likelihood estimator for spatio-temporal point processes. J. Statist. Plan. Inf. 51, 55–74.

Rathbun, S.L., 1996b. Estimation of Poisson intensity using partially observed concomitant variables. Biometrics 52, 226–242.

Rathbun, S.L., Cressie, N. 1994. Asymptotic properties of estimators for the parameters of spatial inhomogeneous Poisson point processes. Adv. Appl. Probab. 26, 122–154.

Rudin, W., 1976. Principles of Mathematical Analysis, 3rd ed. McGraw-Hill, New York.

Schoenberg, F.P., 2003. Multi-dimensional residual analysis of point process models for earthquake occurrences. J. Amer. Statist. Assoc., in press.

Sweeting, T.J., 1980. Uniform asymptotic normality of the maximum likelihood estimator. Ann. Statist. 8, 1375–1381.

Vere-Jones, D., 1992. Statistical methods for the description and display of earthquake catalogs. In: Walden, A., Guttorp, P. (Eds.), Statistics in the Environmental and Earth Sciences, Edward Arnold, London, 220–236.

Vere-Jones, D., Ogata, Y. 1984. On the moments of a self-correcting process. J. Appl. Prob. 21, 335–342.