Maximum Likelihood Estimates for Spatial Pure Birth Processes

Nancy Lopes Garcia, *Universidade Estadual de Campinas* *

Thomas G. Kurtz, *University of Wisconsin - Madison* †

January 9, 2001

Abstract

We can write the likelihood function for the non-homogeneous pure birth process on a compact set through the method of projection introduced by Kurtz (1989) and studied by Garcia (1995), as the projection of the likelihood function. The fact that the projected likelihood can be interpreted as an expectation suggests that MCMC stochastic approximation could be useful in computing parameter estimates. In this case, we obtain a.s. convergence and distributional results for the convergence of the approximants to the maximum likelihood estimator.

Key words: Spatial pure birth process, projection method, Monte Carlo approximants.

AMS Classification: □

1 Introduction

Maximum likelihood estimation for pure birth process has been extensively studied for the one-dimensional case, see for example Keiding (1974), Beyer, Keiding and Simonsen (1976), Moran (1951). However, very little is known for the non-homogeneous spatial pure birth process, that

*Postal address: Departamento de Estatística, IMECC, Cidade Universitária "Zeferino Vaz", Caixa Postal 6065, 13.081-970 - Campinas, SP - BRAZIL

†Postal address: Department of Mathematics, University of Wisconsin, Madison Wisconsin 53705, U.S.A.

is, the birth rate is a function of the size of the population and the location of its elements. We shall discuss parametric maximum likelihood estimation under three different sampling schemes:

(A) Permanent observation in a fixed time interval $[0, T]$;

(B) Observation after a fixed period of time T;

(C) Observation until a fixed number of points k.

Note that in schemes (B) and (C), we just have the observation of the location of the points, but not the birth times.

In Section 2 we describe the projection method for the the spatial pure birth process and get the full likelihood function under scheme (A). In Section 3 we find the likelihood funtion for the process under scheme (B), which is find as the projection of the full likelihood. A MCMC procedure is described to get an approximant for the MLE estimator of a parameter. In Section 4, scheme (C) is studied and another Monte Carlo procedure is used in order to get approximants for the MLE. Section 5 is devoted to simulation studies to show the applicability of the methods described in the previous sections.

2 Pure Birth Process

A pure birth process can be obtained easily using the projection method introduced by Kurtz (1989) described by Garcia (1995). The basic idea of the projection method is the construction of point processes (say on $\mathbb{R}^d$) from higher-dimensional Poisson processes by constructing a random subset of the higher-dimensional space and projecting the points of the Poisson process lying in that set onto the lower-dimensional subspace. We identify a point process with the counting measure N given by assigning unit mass to each point, that is, $N(A)$ is the number of points in a set A. With this identification in mind, let $\mathcal{N}(\mathbb{R}^d)$ denote the collection of counting measures on $\mathbb{R}^d$.

Consider a spatial pure birth process $(\tilde{N})$ in a region $K \subset \mathbb{R}^d$. We specify this process in terms of a nonnegative function $\lambda : \mathbb{R}^d \times \mathcal{N}(\mathbb{R}^d) \to [0, \infty)$. The meaning of λ is that if the point configuration at time t is $n \in \mathcal{N}(\mathbb{R}^d)$, then the probability that a point is added to the configuration in a neighborhood of the point x having area ΔA in the next interval of length

Δt is approximately $\lambda(x,n)\Delta A \Delta t$.

Let N be a Poisson process on $K\times[0,\infty)$ with Lebesgue mean measure. We need to construct a family of random sets Γ_t and hence of point processes $\tilde{N}_t$ such that:

$$\begin{cases} \dot{\tau}(t,x) = \lambda(x,\tilde{N}_t) \\ \tau(0,x) = 0 \\ \tilde{N}_t(B) = N(\Gamma_t \cap B \times [0,\infty)) \\ \Gamma_t = \{(x,y); x \in \mathbb{R}^d, 0 \leq y \leq \tau(t,x)\} \end{cases} \tag{2.1}$$

with Γ_t a stopping set with respect to the filtration $\{\mathcal{F}_A\}$ ($\mathcal{F}_A = \sigma(N(B); B \subset A)$).

If K is compact, the system above consists of finitely many differential equations, and it is straightforward to see that there exists a unique solution of the system (2.1). Moreover, it gives an easy procedure to simulate pure birth processes. The projected process $\tilde{N}_t$ is the point process corresponding to a spatial pure birth process with space-time intensity given by λ. That is,

$$\tilde{N}_t(B) - \int_0^t \int_B \lambda(x,\tilde{N}_s) dx\, ds \tag{2.2}$$

is an $\{\mathcal{F}_{\Gamma_t}\}$-martingale measure.

Let $\tilde{N}(B\times[0,t]) = \tilde{N}_t(B)$. The distribution of $\tilde{N}$ restricted to $K\times[0,T]$ is absolutely continuous with respect to the distribution of the Poisson process on $K\times[0,T]$ with Lebesgue mean measure. The Radon-Nikodym derivative was obtained by Kurtz(1989) and is given by

$$L_T(n) = \exp\Big\{\int_{K\times[0,T]} \log\lambda(x,n_{s-}) n(dx\times ds) - \int_{K\times[0,T]} (\lambda(x,n_s)-1) dx\, ds\Big\}. \tag{2.3}$$

For example, suppose that $\lambda(x,n_s) = \exp\{-\int \rho(\beta, x-y) n_s(dy)\}$, with $\rho \in C^3$ and $\rho'(x)$ uniformely bounded for x bounded away from zero, being completely known. Let $(x_i,y_i), i = 1,\ldots,\tilde{N}_T(K)$ denote the points of $\tilde{N}$ in $K\times[0,T]$. Then L_T becomes

$$L_T(n) = \exp\Big\{-\sum_{i<j}\rho(\beta,x_i-x_j) - \int_{[0,T]}\int_K (\exp\{-\sum_i \rho(\beta,x-x_i)\mathbf{I}_{\{y_i<s\}}\}-1) dx\, ds\Big\} \tag{2.4}$$

and finding the value of β that maximizes $L_T(n)$ is straightforward.

3 Likelihood function for a process observed after a fixed period of time

If we observe the process $\tilde{N}_t$ at a fixed time $t = T$, we have interest in studying the Radon-Nikodym derivative of the distribution of $\tilde{N}_T$ with respect to the distribution of the Poisson process with mean measure Tm (T times Lebesgue measure) on K. Since the Poisson process on $K \times [0, T]$ with Lebesgue mean measure has the property that its projection on K is a Poisson process with mean measure Tm and, conditioned on the projection on K, the third coordinates of the points are independent and uniformly distributed on $[0, T]$, it follows (Kurtz, 1989) that the Radon-Nikodym derivative for $\tilde{N}_T$ can be obtained by "integrating out" the third coordinate in (2.4). In this case, "projected" Radon-Nikodym derivative is

$$
\begin{aligned}
&\hat{L}_T(x_1, \ldots, x_k) = \\
&= \quad T^{-k} \int_0^T \cdots \int_0^T \exp\Big\{ -\sum_{i<j} \rho(\beta, x_i - x_j) \\
&\qquad - \int_0^T \int_K (\exp\{-\sum_i \rho(\beta, x - x_i) \mathbf{I}_{\{y_i < s\}}\} - 1) dx\, ds \Big\} dy_1 \ldots dy_k. \qquad (3.1)
\end{aligned}
$$

Equation (3.1) gives a likelihood for the distribution of $\tilde{N}_T$ and hence provides the basis for maximum likelihood estimation of β. The k-fold integral in the right side of (3.1) makes the desired maximization difficult. For estimation purposes, we can see (3.1) as the likelihood function for a missing data problem. The EM algorithm (Dempster *et al.*, 1977) cannot be used in this case since it requires closed-form expressions for the conditional expectation in the E-step. Monte Carlo EM (Wei and Tanner, 1990; Guo and Thompson, 1991) can be used but it does not give an estimate for the error. Since the complete data likelihood is known, Monte Carlo maximum likelihood can be used (Thompson and Guo, 1991) and the Monte Carlo error can be estimated (Geyer, 1994).

The full likelihood can be written as

$$
\begin{aligned}
&L(\beta, \mathbf{x}, \mathbf{y}) = \qquad (3.2) \\
&= \quad \exp\{-\sum_{i<j} \rho(\beta, x_i - x_j) - \int_{[0,T]} \int_K (\exp\{-\sum_i \rho(\beta, x - x_i) \mathbf{1}_{\{y_i < s\}}\} - 1) dx\, ds\}
\end{aligned}
$$

and the projected likelihood

$$L(\beta, \mathbf{x}) = \frac{1}{T^k} \int_{[0,T]^k} L(\beta, \mathbf{x}, \mathbf{y}) d\mathbf{y}. \tag{3.3}$$

Observe that we can rewrite (3.3) as

$$L(\beta, \mathbf{x}) = L(\beta_0, \mathbf{x}) \Big[\int_{[0,T]^k} \frac{L(\beta, \mathbf{x}, \mathbf{y})}{L(\beta_0, \mathbf{x}, \mathbf{y})} g(\mathbf{y}|\mathbf{x}, \beta_0) d\mathbf{y} \Big] \Big[\int_{[0,T]^k} g(\mathbf{y}|\mathbf{x}, \beta_0) d\mathbf{y} \Big]^{-1} \tag{3.4}$$

where $g(\mathbf{y}|\mathbf{x}, \beta_0)$ is the conditional distribution of $\mathbf{Y}$ given $\mathbf{X} = \mathbf{x}$ and β_0. Thus, if $\{\mathbf{Y}^{(j)}, j = 1, \ldots, m\}$ are drawn from an ergodic Markov chain with stationary distribution $g(\mathbf{y}|\mathbf{x}, \beta_0)$, a Monte Carlo approximant for (3.4) is

$$L_m(\beta, \mathbf{x}) = L(\beta_0, \mathbf{x}). \frac{1}{m} \sum_{j=1}^{m} \frac{L(\beta, \mathbf{x}, \mathbf{Y}^{(j)})}{L(\beta_0, \mathbf{x}, \mathbf{Y}^{(j)})}. \tag{3.5}$$

Note that we have to generate a sample from the distribution

$$g(\mathbf{y}|\mathbf{x}, \beta_0) \propto \frac{L(\beta_0, \mathbf{x}, \mathbf{y})}{L(\beta_0, \mathbf{x})} \tag{3.6}$$

which is known up to a constant. Let the log-likelihood be

$$l(\beta, \mathbf{x}) = \log L(\beta, \mathbf{x}) - \log L(\beta_0, \mathbf{x})$$

and its Monte Carlo approximant

$$l_m(\beta, \mathbf{x}) = \log L_m(\beta, \mathbf{x}) - \log L(\beta_0, \mathbf{x}). \tag{3.7}$$

Denote $\hat{\beta}_m = \arg\ \max l_m(\beta)$ and $\hat{\beta} = \arg\ \max l(\beta)$, the maximum likelihood estimator of β.

Theorem 3.8 *If the parameter space is compact and does not contain zero and $\hat{\beta}$ is unique, then $\hat{\beta}_m \to \hat{\beta}$ a.s. as $m \to \infty$.*

Proof: Since the maps $\beta \mapsto L(\beta, \mathbf{x}, \mathbf{y})$ are continuous, it is enough to prove that l_m converges uniformly to l on compact sets a.s.. For this, it is enough to prove that $|l_m'(\beta, \mathbf{x})| \leq C(\mathbf{x})$ for all β. To do this note that

$$\frac{\partial}{\partial \beta} l_m(\beta, \mathbf{x}) =$$

$$
\begin{aligned}
&= \Big(\sum_{j=1}^m \frac{L(\beta,\mathbf{x},\mathbf{Y}^{(j)})}{L(\beta_0,\mathbf{x},\mathbf{Y}^{(j)})}\Big)^{-1}\sum_{j=1}^m \frac{\partial}{\partial\beta}\frac{L(\beta,\mathbf{x},\mathbf{Y}^{(j)})}{L(\beta_0,\mathbf{x},\mathbf{Y}^{(j)})} \\
&= \Big(\sum_{j=1}^m \frac{L(\beta,\mathbf{x},\mathbf{Y}^{(j)})}{L(\beta_0,\mathbf{x},\mathbf{Y}^{(j)})}\Big)^{-1} \times \\
&\quad \sum_{j=1}^m \frac{L(\beta,\mathbf{x},\mathbf{Y}^{(j)})}{L(\beta_0,\mathbf{x},\mathbf{Y}^{(j)})}\frac{\partial}{\partial\beta}[-\sum_{i<j}\rho(\beta,x_i-x_j)-\int_{[0,T]}\int_K(\exp\{-\sum_i\rho(\beta,x-x_i|)\mathbf{1}_{\{y_i<s\}}\}-1)dx\,ds\}] \\
&= \Big(\sum_{j=1}^m \frac{L(\beta,\mathbf{x},\mathbf{Y}^{(j)})}{L(\beta_0,\mathbf{x},\mathbf{Y}^{(j)})}\Big)^{-1} \times \\
&\quad \sum_{j=1}^m \frac{L(\beta,\mathbf{x},\mathbf{Y}^{(j)})}{L(\beta_0,\mathbf{x},\mathbf{Y}^{(j)})}[-\sum_{i<j}\frac{\partial}{\partial\beta}\rho(\beta,x_i-x_j)+\int_{[0,T]}\int_K(\exp\{-\sum_i\rho(\beta,x-x_i)\mathbf{1}_{\{y_i<s\}}\} \\
&\qquad\qquad -\sum_i\rho(\beta,x-x_i)\mathbf{1}_{\mathbf{y}_i<s\}})dx\,ds\}] \\
&\le \sup_\beta[\sum_{i<j}\frac{\partial}{\partial\beta}\rho(\beta,x_i-x_j)+\int_{[0,T]}\int_K(\exp\{-\sum_i\rho(\beta,x-x_i)\mathbf{1}_{\{y_i<s\}}\} \\
&\qquad\qquad -\sum_i\rho(\beta,x-x_i)\mathbf{1}_{\{y_i<s\}})dx\,ds\}]<\infty
\end{aligned}
\tag{3.9}
$$

if β is bounded away from zero.

The convergence $\hat{\beta}_m \to \hat{\beta}$ a.s. follows immediately from Theorem 4, Geyer (1994). □.

Theorem 3.10 *Under conditions of Theorem 3.8, there exists a positive constant $\sigma^2(\hat{\beta})$ depending on $\hat{\beta}$ such that*

$$\frac{\sqrt{m}(\hat{\beta}_m-\hat{\beta})}{\sigma(\hat{\beta})} \xrightarrow{\mathcal{L}} N(0,1)$$

as $m\to\infty$.

Proof. We have to verify the conditions of Theorem 7, Geyer (1994). This theorem states that if:

(a) $m^{1/2}\nabla l_m(\beta) \xrightarrow{\mathcal{L}} N(0,A(\beta))$ for some covariance matrix $A(\beta)$.

(b) $B(\beta)=-\nabla^2 l(\beta)$ is positive definite.

(c) $\nabla^3 l_m(\beta)$ is bounded in probability uniformly in a neighborhood of $\hat{\beta}$.

Then $-\nabla^2 l_m(\hat{\beta}_m)\to B(\hat{\beta})$ in probability and

$$n^{1/2}(\hat{\beta}_m-\hat{\beta}) \xrightarrow{\mathcal{L}} N(0,B(\hat{\beta})^{-1}A(\hat{\beta})B(\hat{\beta})^{-1}).$$

Note that, in our case, we have a one-dimensional parameter and

$$\sigma^2(\hat{\beta})=B(\hat{\beta})^{-2}A(\hat{\beta}). \tag{3.11}$$

All the conditions except (a) are fairly straightforward and can be verified. In our case, we have

$$B = \frac{\left[\int_{[0,T]^k} \frac{\partial}{\partial\beta} L(\beta, \mathbf{x}, \mathbf{y}) d\mathbf{y}\right]^2 - \int_{[0,T]^k} L(\beta, \mathbf{x}, \mathbf{y}) d\mathbf{y} \int_{[0,T]^k} \frac{\partial^2}{\partial\beta^2} L(\beta, \mathbf{x}, \mathbf{y}) d\mathbf{y}}{\left[\int_{[0,T]^k} L(\beta, \mathbf{x}, \mathbf{y}) d\mathbf{y}\right]^2}$$

is positive and can be estimated using the same Metropolis-Hastings algorithm used for computing $\hat{\beta}_m$. Moreover, $\nabla^3 l_m(\beta)$ is bounded for β bounded away from zero. Therefore, Theorem 3.10 can be used to get

$$m^{1/2}(\hat{\beta}_m - \hat{\beta}) \xrightarrow{\mathcal{L}} N(0, \frac{A(\hat{\beta})}{B(\hat{\beta})^2})$$

if condition (a) is satisfied. This condition involves a Markov chain Central Limit Theorem. The Kipnis-Varadhan theorem (Kipnis and Varadhan, 1986) requires reversibility and summabiblity of the autocovariances. It is pointed by P. Green in Besag (1986) that a Metropolis-Hastins algorithm can always be designed so that the Markov chain is reversible. The summability condition is hard to verify, but it is related only to the Metropolis-Hastings generation procedure, not the maximum likelihood problem. Note that

$$\nabla l_m(\beta) = \frac{\sum_{j=1}^m \frac{\partial}{\partial\beta} L(\beta, \mathbf{x}, \mathbf{Y}^{(j)}) / L(\beta_0, \mathbf{x}, \mathbf{Y}^{(j)})}{\sum_{j=1}^m L(\beta, \mathbf{x}, \mathbf{Y}^{(j)}) / L(\beta_0, \mathbf{x}, \mathbf{Y}^{(j)})} \tag{3.12}$$

where $\{\mathbf{Y}^{(j)}, j = 1, \ldots, m\}$ are drawn from a Markov chain with stationary distribution $g(\mathbf{y}|\mathbf{x}, \beta_0)$ defined by (3.6). Then

$$\frac{1}{m} \sum_{j=1}^m \frac{L(\beta, \mathbf{x}, \mathbf{Y}^{(j)})}{L(\beta_0, \mathbf{x}, \mathbf{Y}^{(j)})} \to \frac{1}{T^k} \int_{[0,T]^k} \frac{L(\beta, \mathbf{x}, \mathbf{y})}{L(\beta_0, \mathbf{x}, \mathbf{y})} g(\mathbf{y}|\mathbf{x}, \beta_0)\, d\mathbf{y} \tag{3.13}$$

and

$$\frac{1}{m} \sum_{j=1}^m \frac{\frac{\partial}{\partial\beta} L(\beta, \mathbf{x}, \mathbf{Y}^{(j)})}{L(\beta_0, \mathbf{x}, \mathbf{Y}^{(j)})} \to \frac{1}{T^k} \int_{[0,T]^k} \frac{\frac{\partial}{\partial\beta} L(\beta, \mathbf{x}, \mathbf{y})}{L(\beta_0, \mathbf{x}, \mathbf{y})} g(\mathbf{y}|\mathbf{x}, \beta_0)\, d\mathbf{y} \tag{3.14}$$

almost surely as $m \to \infty$. Let

$$\gamma_t = \gamma_{-t} = \text{Cov}\Big(\frac{\frac{\partial}{\partial\beta} L(\beta, \mathbf{x}, \mathbf{Y}^{(j)})}{L(\beta_0, \mathbf{x}, \mathbf{Y}^{(j)})}, \frac{\frac{\partial}{\partial\beta} L(\beta, \mathbf{x}, \mathbf{Y}^{(j+t)})}{L(\beta_0, \mathbf{x}, \mathbf{Y}^{(j+t)})} \Big)$$

and $\gamma^2 = \sum_{t=-\infty}^{\infty} \gamma_t$. If $\gamma^2 < \infty$ we have (Kipnis and Varadhan, 1986) that

$$m^{1/2} \nabla l_m(\beta) \xrightarrow{\mathcal{L}} N(0, A)$$

where

$$A = \gamma^2 \Big[\int_{[0,T]^k} \frac{L(\beta, \mathbf{x}, \mathbf{y})}{L(\beta_0, \mathbf{x}, \mathbf{Y}^{(j)})} g(\mathbf{y}|\mathbf{x}, \beta_0) \Big]^2. \tag{3.15}$$

Both factors in (3.15) can be estimated, the second factor by the Metropolis-Hastings algorithm as

$$\Big(\frac{1}{m}\sum_{j=1}^{m}\frac{L(\beta,\mathbf{x},\mathbf{Y}^{(j)})}{L(\beta_0,\mathbf{x},\mathbf{Y}^{(j)})}\Big)^2$$

and γ^2 by standard time series methods (Geyer, 1992). It is important to notice that $m^{-1}A(\beta)/B(\beta)^2$ is the estimated variance of the Monte Carlo method in calculating $\hat{\beta}$, i.e., the error involved in using the Monte Carlo estimate $\hat{\beta}_m$ to approximate the exact MLE $\hat{\beta}$. The error involved in using the exact MLE to approximate the true parameter β is given, under the usual regularity conditions, by the inverse Fisher information, which is estimated by $B(\beta)^{-1}$. □

4 Likelihood function for a process observed until a fixed number of points are born

Theorem 3.8 and Theorem 3.10 give us nice asymptotic results but the summability condition is hard to verify depending on the Metropolis-Hastings algorithm used. However, a modification in the stopping time of observation can simplify the results. Instead of observing the process during a fixed period of time, we observe the process until a fixed number of particles are born (say, k), that is, we observe the process until a random time

$$\tau_k = \inf\{t \geq 0; \tilde{N}_t(K) = k\}$$

denoting $(x_i, y_i), i = 1, \ldots, k$, the points of $\tilde{N}$ in $K \times [0, \tau_k]$, then (2.3) becomes

$$L_k(\mathbf{x},\mathbf{y}) = \frac{dP}{dQ} = \exp\Big\{-\sum_{i<j}\rho(\beta, x_i - x_j) - \sum_{j=1}^{k}\int_K (\exp\{-\sum_{\sigma_i \leq j-1}\rho(\beta, x - x_i)\} - 1)y_j\Big\}$$

where σ is a permutation of $\{1, 2, \ldots, k\}$ such that $(x_{\sigma_1}, y_{(1)}), (x_{\sigma_2}, y_{(2)}), \ldots, (x_{\sigma_k}, y_{(k)})$ ($y_{(1)}, \ldots, y_{(k)}$ are the order statistics of $y_1, \ldots, y_k$). Under the unit Poisson process (Q) $y_1, y_2, \ldots, y_k$ are i.i.d. $\exp(m(K))$ random variables. Note that

$$\mathbb{E}^Q[e^{-uy_j}] = \int_0^\infty e^{-ux} m(K) e^{-m(K)x} dx = \frac{m(K)}{m(K) + u}.$$

Consequently,

$$\mathbb{E}[\exp\{-\sum_{j=1}^{k}\int_K (\exp\{-\sum_{\sigma_i \leq j-1}\rho(\beta, x - x_i)\} - 1)y_j\Big\}|\sigma] =$$

$$= \prod_{j=1}^{k} \mathbb{E}[\exp\{-\int_K (\exp\{-\sum_{\sigma_i \le j-1} \rho(\beta, x - x_i)\} - 1) y_j\Big\} | \sigma]$$

$$= \prod_{j=1}^{k} \frac{m(K)}{m(K) + \int_K (\exp\{-\sum_{\sigma_i \le j-1} \rho(\beta, x - x_i)\} - 1) dx}.$$

Under Q, σ is uniform random permutation and the projected Radon-Nikodym derivative is

$$\hat{L}_k(x_1, \ldots, x_k) = \mathbb{E}^Q[L_k(n)|\mathbf{x}]$$
$$= \exp\{-\sum_{i<j} \rho(\beta, x_i - x_j)\}\Big[\frac{1}{k!} \sum_{\sigma} \prod_{i=1}^{k} \frac{m(K)}{\int_K \exp\{-\sum_{\sigma_i \le j-1} \rho(\beta, x - x_i)\} dx}\Big] \tag{4.1}$$

If k is large, the calculation in (4.1) is prohibitive and one suggestion is to use Monte Carlo methods. Let $\sigma^{(1)}, \sigma^{(2)}, \ldots, \sigma^{(m)}$ be random permutations of $\{1, 2, \ldots, k\}$ and

$$\hat{L}_k^{(m)} = \exp\{-\sum_{i<j} \rho(\beta, x_i - x_j)\}\Big[\frac{1}{m} \sum_{l=1}^{m} \prod_{i=1}^{k} \frac{m(K)}{\int_K \exp\{-\sum_{\sigma_i^{(l)} \le j-1} \rho(\beta, x - x_i)\} dx}\Big]. \tag{4.2}$$

Consider

$$d(\beta, \mathbf{x}) = \frac{[m(K)]^{k-1}}{k!} \sum_{\sigma} (\prod_{j=1}^{k-1} \int_K \exp\{-\sum_{\sigma_i \le j} \rho(\beta, x - x_i) dx\})^{-1} \tag{4.3}$$

and

$$d_m(\beta, \mathbf{x}) = \frac{[m(K)]^{k-1}}{m} \sum_{i=1}^{m} (\prod_{j=1}^{k-1} \int_K \exp\{-\sum_{\sigma_i^{(m)} \le j} \rho(\beta, x - x_i) dx\})^{-1} \tag{4.4}$$

then, for any fixed β

$$d_m(\beta, \mathbf{x}) \to d(\beta, \mathbf{x}) \text{ a.s..} \tag{4.5}$$

The log-likelihood, given the observation $\mathbf{x}$ can be written as

$$l(\beta, \mathbf{x}) = -\sum_{i<j} \rho(\beta, x_i - x_j) - \log d(\beta, \mathbf{x}) \tag{4.6}$$

and its Monte Carlo approximant as

$$l_m(\beta, \mathbf{x}) = -\sum_{i<j} \rho(\beta, x_i - x_j) - \log d_m(\beta, \mathbf{x}). \tag{4.7}$$

Then, for any fixed β

$$l_m(\beta, \mathbf{x}) \to l(\beta, \mathbf{x}) \text{ a.s.} \tag{4.8}$$

by (4.5). Let $\hat{\beta} = \arg\max_{\beta} l(\beta, \mathbf{x})$ and $\hat{\beta}_m = \arg\max_{\beta} l_m(\beta, x)$.

Theorem 4.9

$$\hat{\beta}_m \to \hat{\beta} \;\; a.s. \tag{4.10}$$

Proof. Since (4.5) and (4.8) hold for β on a dense set, we have

$$\hat{\beta}_m \to \hat{\beta} \text{ a.s.} \tag{4.11}$$

by Theorem 4, Geyer (1994). Note that the convergence here refers to the convergence of the Monte Carlo approximant to the maximum likelihood estimator of β. □

Theorem 4.12 *Given* $\mathbf{x}$, *we have*

$$\sqrt{m}(\hat{\beta} - \hat{\beta}_m) \xrightarrow{\mathcal{D}} N\Big(0, \frac{Var_\sigma(g_1(\sigma, \hat{\beta}, \mathbf{x}))}{\mathbb{E}(\frac{\partial}{\partial\beta} g_1(\sigma, \hat{\beta}, \mathbf{x}))\mathbb{E}(g_2(\sigma, \hat{\beta}, \mathbf{x})) - \mathbb{E}^2(g_1(\sigma, \hat{\beta}, \mathbf{x}))}\Big) \tag{4.13}$$

where

$$g_1(\sigma, \beta, \mathbf{x}) = \frac{\partial}{\partial\beta} g_2(\sigma, \beta, \mathbf{x}) \tag{4.14}$$

$$g_2(\sigma, \beta, \mathbf{x}) = \prod_{j=1}^{k-1} \int_K \exp\{-\sum_{\sigma_i \leq j} \rho(\beta, x - x_i) dx\})^{-1}. \tag{4.15}$$

Moreover, the expected values on the variance expression can be easily approximated by standard Monte Carlo methods.

Proof. Given $\mathbf{x}$

$$\sqrt{m}\Big(\frac{\partial}{\partial\beta} \log d_m(\beta, \mathbf{x}) - \frac{\partial}{\partial\beta} \log d(\beta, \mathbf{x})\Big) \xrightarrow{\mathcal{D}} N\Big(0, \frac{\mathrm{Var}_\sigma(g_1(\sigma, \beta, \mathbf{x}))}{\mathbb{E}^2_\sigma(g_2(\sigma, \beta, \mathbf{x}))}\Big) \tag{4.16}$$

as $m \to \infty$, where g_1 and g_2 are given by (4.14) and (4.15) respectively.

Since $\frac{\partial}{\partial\beta} l_m(\hat{\beta}_m, \mathbf{x}) = 0$ and $\frac{\partial}{\partial\beta} l(\hat{\beta}, \mathbf{x}) = 0$ we have from (4.16)

$$\sqrt{m}\Big(\frac{\partial}{\partial\beta} l_m(\hat{\beta}, \mathbf{x}) - \frac{\partial}{\partial\beta} l_m(\hat{\beta}_m, \mathbf{x})\Big) \xrightarrow{\mathcal{D}} N\Big(0, \frac{\mathrm{Var}_\sigma(g_1(\sigma, \beta, \mathbf{x}))}{\mathbb{E}^2_\sigma(g_2(\sigma, \beta, \mathbf{x}))}\Big). \tag{4.17}$$

By Taylor's expansion

$$\begin{aligned} &\sqrt{m}\Big(\frac{\partial}{\partial\beta} l_m(\hat{\beta}, \mathbf{x}) - \frac{\partial}{\partial\beta} l_m(\hat{\beta}_m, \mathbf{x})\Big) = \\ &= \sqrt{m}(\hat{\beta} - \hat{\beta}_m) \frac{\partial^2}{\partial\beta^2} l_m(\hat{\beta}, \mathbf{x}) + R_{nm}(\hat{\beta}, \hat{\beta}_m)) \end{aligned} \tag{4.18}$$

with

$$R_{nm}(\hat{\beta},\hat{\beta}_m) = \sqrt{m}(\hat{\beta}-\hat{\beta}_m)^2\frac{\partial^3}{\partial\beta^3}l_m(\xi,\mathbf{x})$$

and

$$R_{nm}(\hat{\beta},\hat{\beta}_m) \to 0 \text{ a.s.} \tag{4.19}$$

as $m \to \infty$ by (4.11). Moreover,

$$\frac{\partial^2}{\partial\beta^2}l_m(\hat{\beta},\mathbf{x}) \to \frac{\partial^2}{\partial\beta^2}l(\hat{\beta},\mathbf{x}) \text{ a.s.} \tag{4.20}$$

as $m \to \infty$. By (4.17), (4.18), (4.19) and (4.20) we have the desired result. □

5 Simulation study

Suppose that we generate a pure birth process with space-time intensity given by

$$\lambda(x,n) = \exp\{-\int \beta^\alpha |x-y|^{-\alpha})n(dy)\}, \tag{5.1}$$

with $\alpha > 0$. In this case, using the projection method given by equation (2.1), it is very easy to simulate the pure birth process.

6 References

Besag, J. (1986) "On the statistical analysis of dirty pictures (with discussion)", *J. R. Stat. Soc.* B, **48**, pp. 259-302.

Beyer, J.E., Keiding, N. and Simonsen, W. (1976) "The exact behaviour of the maximum likelihood estimator in the pure birth process and the pure death process", *Scandinavian Journal of Statistics*, **3**, pp. 61–72.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) "Maximum likelihood from incomplete data via the EM algorith (with discussion) ", *J.R. Statit. Soc. B*, **39**, pp. 1–37.

Garcia, N.L. (1995) “Birth and Death Processes as Projections of Higher Dimensional Poisson Processes”, *Advances in Applied Probability*, **27**, No. 4, pp. 911–930.

Geyer, C.J. (1992) “Practical Markov Chain Monte Carlo”, *Statistical Science*, **7**, No. 4, pp. 473–511.

Geyer, C.J. (1994) “On the Convergence of the Monte Carlo Maximum Likelihood Calculations”, *J. R. Statist. Soc. B*, **56**, pp.261–274.

Guo, S.W. and Thompson, E.A. (1991) “ Monte Carlo estimation of variance component models for large complex pedigrees”, *IMA J. Math. Appl. Med. Biol.*, **8**, pp. 171–189.

Keiding, N. (1974) “Estimation in the birth process”, *Biometrika.* **61**, pp. 71–80.

Kipnis, C. and Varadhan, S.R.S. (1986), “Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions”, *Comm. Math. Phys.*, **104**, pp. 1–19.

Kurtz, T.G. (1989) “Stochastic Processes as Projections of Poisson Random Measures”, Special invited paper at IMS meeting, Washington, D.C.. Unpublished.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. Teller, E. (1953) “Equation of State Calculations by Fast Computing Machines”, *J. Chem. Phys.*, **21**, pp. 1087–1092.

Moran, P.A.P. (1951) “Estimation methods for evolutive processes”, *J. R. Stat. Soc.* B, **13**, pp. 141–146.

Thompson, E.A. and Guo, S.W. (1991) “Evaluation of likelihood ratios for complex genetic models”, *IMA J. Math. Appl. Med. Biol.*, **8**, pp. 149–169.

Wei, G.C.G. and Tanner, M.A. (1990) “ A Monte Carlo implementation of the EM algorithm and poor man's augmentation", *J. Am. Stat. Assoc.*, **85**, pp. 699–704.