

Partial Likelihood Analysis for Spatio-temporal Point Process Data

Irene Kaimi

i.kaimi@lancaster.ac.uk

Peter Diggle

p.diggle@lancaster.ac.uk



Introduction

Estimation of the parameters of interest of spatio-temporal point process models is performed using the partial likelihood method.

- Data: $(\mathbf{x}_i, t_i) \in A \times [0, T]$; $i = 1, \dots, n$, where \mathbf{x}_i is the location and t_i is the time of occurrence of an event of interest
- \mathcal{H}_t : the complete history of the process up to time t
- $\lambda(\mathbf{x}, t|\mathcal{H}_t)$: the conditional intensity for an event at location \mathbf{x} and time t .

The full log-likelihood function of the process is given by

$$L(\theta) = \sum_{i=1}^n \log \lambda(\mathbf{x}_i, t_i|\mathcal{H}_{t_i}) - \int_0^T \int_A \lambda(\mathbf{x}, t|\mathcal{H}_t) d\mathbf{x} dt. \quad (1)$$

Maximisation of Equation (1) with respect to the parameters of interest, θ , gives the maximum likelihood estimators (MLE's) of θ .

However,

1. the conditional intensity $\lambda(\mathbf{x}, t|\mathcal{H}_t)$ may be intractable
2. the double integral appearing in Equation (1) usually cannot be evaluated by applying routine methods.

We propose a partial likelihood as an alternative to the full likelihood for spatiotemporal point processes. Partial likelihood was originally introduced by Cox (1972) for applications in survival analysis. Diggle (2006) suggests that partial likelihood can also be used for spatiotemporal models that can be defined through their conditional intensity.

Partial Likelihood Formulation

To obtain the partial likelihood we condition on the times t_i , and then take into account the log-likelihood formed for the observed time-ordering of the events $i = 1, \dots, n$. Let \mathcal{R}_i the risk set at time t_i . The contribution of the i th event to the partial likelihood is

$$p_i = \frac{\lambda(\mathbf{x}_i, t_i|\mathcal{H}_{t_i})}{\int_{\mathcal{R}_i} \lambda(\mathbf{x}, t_i|\mathcal{H}_{t_i}) d\mathbf{x}} \quad (2)$$

Then, the partial likelihood is given by

$$L_p(\theta) = \sum_{i=1}^n \log p_i = \sum_{i=1}^n \log \lambda(\mathbf{x}_i, t_i|\mathcal{H}_{t_i}) - \sum_{i=1}^n \log \left\{ \int_{\mathcal{R}_i} \lambda(\mathbf{x}, t_i|\mathcal{H}_{t_i}) d\mathbf{x} \right\}. \quad (3)$$

For a *spatially continuous point process model*, for which points of the process can occur anywhere in the spatial region of interest A , the risk set \mathcal{R}_i is the whole spatial region A . On the other hand, for a *spatially discrete point process model* points of the process can only occur at a finite set of locations \mathbf{x}_j ; $j = 1, \dots, N$, for some $N \geq n$. Hence, the integral appearing in Equations (2) and (3) reduces to a sum and

$$L_p(\theta) = \sum_{i=1}^n \log \lambda(\mathbf{x}_i, t_i|\mathcal{H}_{t_i}) - \sum_{i=1}^n \log \left\{ \sum_{j \geq i} \lambda(\mathbf{x}_j, t_i|\mathcal{H}_{t_i}) \right\}. \quad (4)$$

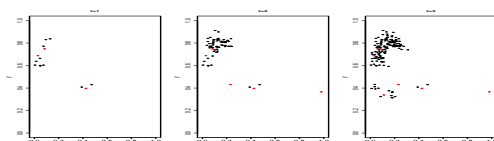
We present two examples of continuous point process models for which application of the full likelihood is intractable, whereas the partial likelihood method is applicable and can be used to conduct inference.

Infectious disease model

Simulation

An infectious disease model in $A \times [0, T]$; $A = [a, b] \times [c, d]$ may be described by the algorithm:

- The first event occurs at a random location and time $(\mathbf{x}_0, t_0) \in A \times [0, T]$.
- Progenitor event (\mathbf{x}_0, t_0) generates offspring at rate λ , within time δ since its birth. The locations of the offspring \mathbf{x}_{off} are determined by $\mathbf{x}_{\text{off}} = \mathbf{x}_{\text{par}} + \mathbf{w}$, where $\mathbf{w} \sim f$. f is a bivariate distribution, e.g. $\mathbf{w} \sim \text{BVN}((0, 0)^T, \Sigma)$, $\Sigma = \sigma^2 I$.
- Each first-generation offspring produces offspring of second generation within time δ from their parent events in the same way.
- Generations of events stop when time T is reached.
- Processes producing events following the same routine initiate by new progenitor events, occurring at rate ρ , at random locations $(\mathbf{x}, t) \in A \times [0, T]$.



Spatial locations of a realisation of the infectious disease point process model $\lambda = 10$, $\rho h_0 = 5$, $\kappa = 0.001$ and $\delta = 0.1$. • progenitor events and • offspring

Model

Conditional intensity: $\lambda(\mathbf{x}, t|\mathcal{H}_t) = \sum_{j=1}^{n(t)} \lambda_j(\mathbf{x}, t_j) \mathbb{I}[(t - t_j) < \delta] + \rho$

- $\lambda_j(\mathbf{x}, t_j) = \lambda f(\mathbf{x} - \mathbf{x}_j)$; $f = f(\mathbf{x} - \mathbf{x}_j) = \frac{1}{2\pi\kappa} \exp\left(-\frac{1}{2\kappa}(\|\mathbf{x} - \mathbf{x}_j\|)^2\right)$
- ρ is the rate of spontaneous cases

Evaluate the integral:

1. Analytically: $\int_A f(\|\mathbf{x} - \mathbf{x}_j\|) d\mathbf{x} = \Pr(a \leq X_1 \leq b) \times \Pr(c \leq X_2 \leq d)$
2. Using a quadrature method; quadrature weights = areas of Voronoi tiles

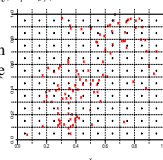
Then the integral can be approximated by a definite weighted sum,

$$\int_A \lambda(\mathbf{x}, t_i|\mathcal{H}_{t_i}) d\mathbf{x} \approx \sum_{k=1}^K w_k \lambda(\mathbf{u}_k, t_i|\mathcal{H}_{t_i}),$$

* \mathbf{u}_k ; $k = 1, \dots, K$: the design points, used to form a partition of the spatial region A , according to the Voronoi tessellation,

* w_k : the area of the tile that contains the point \mathbf{u}_k ,

* $\sum_{k=1}^K w_k = |A|$.



Results

Mean values and associated CIs for the MPLEs of κ and $\tau = \rho/\lambda$ based on 100 realisations of approximate size 100 of the infectious disease process model with $\lambda = 10$, $\rho = 5$, $\kappa = 0.001$ and $\delta = 0.1$.

	κ	$\tau = \rho/\lambda$
exact value	0.0010 (0.0009, 0.0013)	0.47 (0.29, 0.78)
10 × 10 grid	0.0009 (0.0007, 0.0013)	0.46 (0.25, 0.85)
25 × 25 grid	0.0010 (0.0009, 0.0013)	0.47 (0.29, 0.78)

Bird nesting at the Delta dEbre in Spain

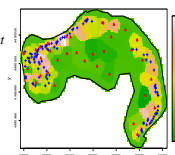
Data

- Settlement pattern of common terns on 21 small islets at the Ebro Delta Natural Park in Spain, which were regularly inspected at two-day intervals during the 2000 breeding season
- Data: locations and arrival times (\mathbf{x}_i, t_i) of nesting birds
- Birds known to prefer higher ground for nesting
- No two nests can coexist within minimum separation distance, $d_0 \approx 0.25$ m

Question of interest: Does the spatio-temporal pattern of nesting sites show any evidence of spatial interaction beyond the minimum separation distance?

In order to include the covariate elevation in the model, we approximate the surface by assigning each point on the islet to the elevation of the point closest to it, which requires no parametric assumptions.

Information about elevation for the largest islet (islet 84). Coast boundaries (—), spatial locations of the nests (+), and other locations for which elevation is recorded (Δ).



Estimated surface for islet 84 based on the set of available elevations.

Model

Conditional intensity: $\lambda(\mathbf{x}, t|\mathcal{H}_t) = \lambda_0(t) \exp\{\beta z(\mathbf{x})\} g(\mathbf{x}, t|\mathcal{H}_t)$

Results

The likelihood ratio tests between a number of models support:

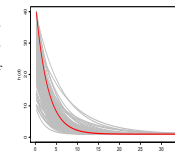
- $g(\mathbf{x}, t|\mathcal{H}_t) = h\left(\min_{j, t_j < t} (\|\mathbf{x}_j - \mathbf{x}\|)\right)$, hence evidence only for spatial dependence between nearest neighbours,
- h describes the relationship between events in space. Its form allows for minimum separation distance and for exponentially decaying interaction between nests

$$h(u) = \begin{cases} 0, & u \leq d_0 \\ 1 + \theta \exp\left\{-\frac{(u-d_0)}{\phi}\right\}, & u > d_0, \end{cases} \quad (5)$$

where ϕ and θ are the parameters of interest,

- $z(\mathbf{x})$ the elevation at location \mathbf{x} , and β the size of the log-linear effect which is found to be significant (p-value $< 10^{-6}$).
- $\hat{\beta}_{\text{MPILE}} = 0.05$, Monte Carlo standard error equal to 6.5×10^{-4} .

Monte Carlo test for goodness of fit of the spatial kernel h for the birds' data for islet 84, using model \mathcal{M}_1 and $\kappa = 1$: data (—); 99 simulated datasets of the process corresponding to the model (—). The estimated spatial kernel gives evidence of positive interaction between nests for distances up to 15 m.



References

- Cox, D. (1972). Regression models and life-tables (with discussion). *Journal of Royal Statistics Society Series B* **34**, 187–220.
- Diggle, P. (2006). Spatio-temporal point processes, partial likelihood, foot and mouth disease. *Statistical Methods in Medical Research* **15** (4), 325–336.