ESTIMATING FUNCTIONS FOR INHOMOGENEOUS COX PROCESSES

Rasmus Waagepetersen

Department of Mathematics, Aalborg University, Fredrik Bajersvej 7G, DK-9220 Aalborg, Denmark (rw@math.aau.dk)

Abstract. Estimation methods are reviewed for inhomogeneous Cox processes with tractable first and second order properties. We illustrate the various suggestions by means of data examples.

Keywords: composite likelihood function, estimating function, intensity function, K-function, Markov chain Monte Carlo, maximum likelihood estimation, minimum contrast estimation, pair correlation, product density.

1 Introduction

Cox processes play an important role for modeling of a wide variety of spatial point patterns. In the literature on inference for spatial Cox processes, stationarity has often been a fundamental assumption. However, the examples in Section 2 illustrate that it is often important to take into account inhomogeneity due to varying observation conditions or spatially varying covariates.

Likelihood-based inference for Cox processes can be carried out using Markov chain Monte Carlo methods, see [12] and [14]. However, the Markov chain Monte Carlo approach is computationally demanding and not yet amenable for routine analyses by non-specialists. On the other hand, second order summary statistics like the K-function and the pair correlation function are often available in closed form. Minimum contrast estimation has therefore been a popular approach to parameter estimation where parameter estimates are chosen to minimize an integrated squared distance between non-parametric estimates of the summary statistic and its theoretical expression, see [8], [13], [6], or [12].

Minimum contrast estimation has mainly been applied to stationary Cox processes. In Section 3, however, we consider examples of inhomogeneous Cox processes which have both known intensity function and known pair correlation/K-function. Minimum contrast estimation and estimating functions based on the intensity function or the second order product density then provide computationally cheap alternatives to maximum likelihood estimation, see Section 4. A case study of likelihood-based inference is considered in Section 5.

2 Data examples

Figure 1 shows positions of 55 minke whales (*balaneoptera acutorostrata*) observed in a part of the North Atlantic near Spitzbergen. The whales are observed visually from a ship sailing along predetermined so-called transect lines. The point pattern can be thought of as an incomplete observation of all the whale positions, since it is only possible to observe whales within the vicinity of the ship. Moreover, whales within sighting distance may fail to be observed due to bad weather conditions or if they are diving. The probability of observing a whale is a decreasing function of the distance from the whale to the ship and

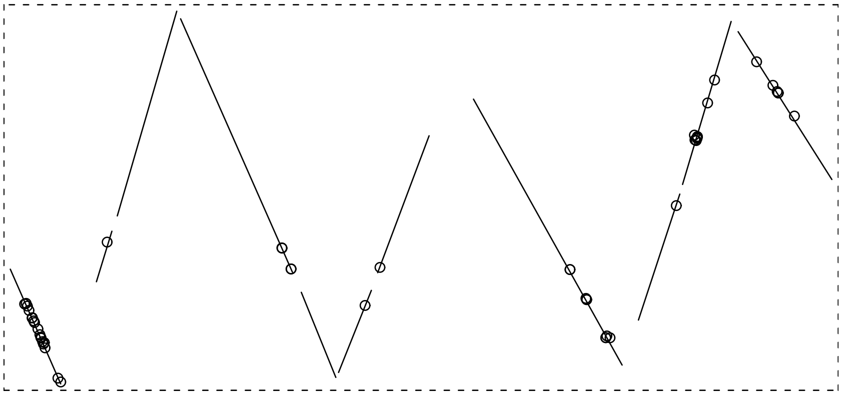

Fig. 1: Observed whales along transect lines. The enclosing rectangle is of dimensions 263 km by 116 km.

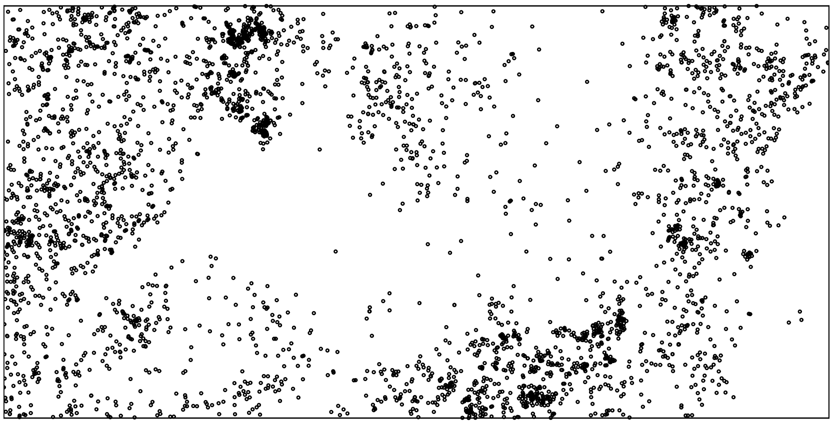

Fig. 2: Locations of Beilschmiedia pendula Lauraceae trees.

is effectively zero for distances larger than 2 km. The objective is to estimate the whale intensity. When estimating the uncertainty of the intensity estimate it is important to take into account that the whales tend to cluster around locations of high prey intensity. More details on the data set can be found in [14].

In studies of biodiversity of tropical rain forests, it is of interest to study whether the spatial patterns of the many different tree species can be related to spatial variations in environmental variables concerning topography and soil properties. Figure 2 shows positions of 3605 *Beilschmiedia pendula Lauraceae* trees in a 1000 m by 500 m rectangular observation window in the tropical rain forest of Barro Colorado Island. This data set is a part of a much larger data set containing positions of hundreds of thousands of trees belonging to thousands of species, see [4, 5, 9]. In addition to the tree positions, covariate information on altitude and norm of altitude gradient is available, see Figure 3, and the question is whether the intensity of *Beilschmiedia* trees may be viewed as a spatially varying function of the covariates. In the study of this question it is, as for the whales, important to take into account clustering — in this case due to tree reproduction by seed dispersal and possibly unobserved covariates.

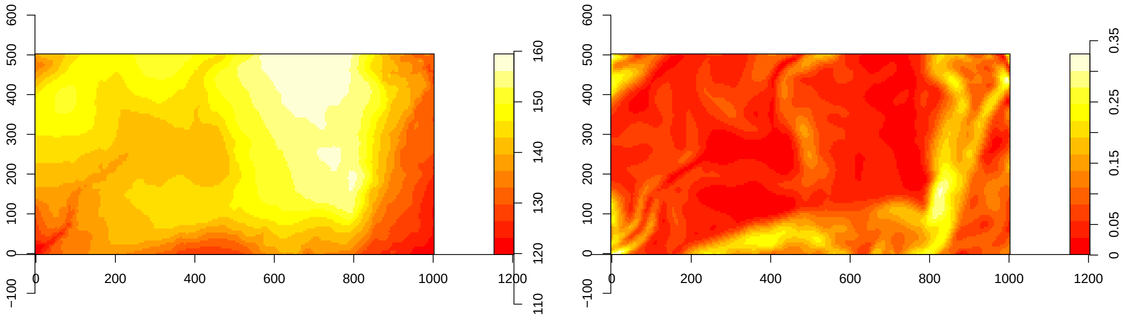

Fig. 3: Altitude (left plot) in meter and norm of altitude gradient (right plot).

3 Inhomogeneous Cox processes with known first and second order properties

In the following $\mathbf{X}$ denotes a Cox process on a subset S of the plane and driven by a non-negative random intensity function $\Lambda = (\Lambda(u))_{u \in S}$. That is, conditional on Λ, $\mathbf{X}$ is a Poisson process with intensity function Λ. In the following we consider two important classes of spatial Cox processes.

3.1 Log Gaussian Cox process

In analogy with random effect models one may define a random intensity function by

$$\log \Lambda(u) = z(u)\beta^{\mathsf{T}} + \Psi(u)$$

where $\Psi = (\Psi(u))_{u \in S}$ is a zero-mean Gaussian process. Then $\mathbf{X}$ is a log Gaussian Cox process [11].

The intensity function

$$\log \rho(u) = z(u)\beta^{\mathsf{T}} + c(u,u)/2 \tag{1}$$

is log linear where $c(u,v) = \mathrm{E}\Psi(u)\Psi(v)$ denotes the covariance function of Ψ. The covariance function c and the pair correlation function g are in a one-to-one correspondence as

$$g(u,v) = \exp(c(u,v))$$

and higher-order product densities are nicely expressed in terms of ρ and g [11].

For the tropical rain forest trees in Figure 2, we consider later on inference for a log Gaussian Cox process with $z(u) = (1, z_2(u), z_3(u))$ where $z_2(u)$ and $z_3(u)$ denote the altitude and gradient covariates at u. An exponential covariance function $c(u,v) = \sigma^2 \exp(-\|u-v\|/\phi)$ is used for the Gaussian process, where σ and ϕ are positive parameters.

3.2 Shot-noise Cox process

A shot noise Cox process $\mathbf{X}$ has

$$\Lambda(u) = \sum_{(c,\gamma) \in \mathbf{\Phi}} \gamma k(c,u) \tag{2}$$

where $\mathbf{\Phi}$ is a Poisson process on $\mathbf{R}^2 \times (0, \infty)$ and $k(c, \cdot)$ is a density for a two-dimensional continuous random variable [10]. The process is an example of a Poisson cluster process, since $\mathbf{X}$ is distributed as the superposition (i.e. union) of independent Poisson processes $\mathbf{X}_{(c,\gamma)}$ with intensity functions $\gamma k(c, \cdot)$, $(c, \gamma) \in \mathbf{\Phi}$, where we interpret $\mathbf{X}_{(c,\gamma)}$ as a cluster with centre c and mean number of points γ.

A special case is a Neyman-Scott process $\mathbf{X}$, where the centre points form a stationary Poisson process with intensity κ and the γ's are all equal to a positive parameter α. If furthermore $k(c, \cdot)$ is a bivariate normal density with mean c and covariance matrix $\omega^2 I$, then $\mathbf{X}$ is a Thomas process. A Neyman-Scott process is stationary with intensity $\rho = \alpha\kappa$, and the Thomas process is also isotropic with

$$g(r) = 1 + \exp\left(-r^2/(4\omega^2)\right)/(4\pi\kappa\omega^2), \quad r > 0. \tag{3}$$

In [14], the positions of minke whales in Figure 1 are modeled as an independent thinning of a shot noise Cox process. Letting $p(u)$ denote the probability of observing a whale at location u, the process of observed whales is a Cox process too with random intensity function $\Lambda(u) = p(u) \sum_{(c,\gamma)\in\mathbf{\Phi}} \gamma k(c, u)$. The cluster centres c are assumed to form a stationary Poisson process, and given the c's, the γ's are i.i.d. gamma random variables with unit scale parameter. The kernel $k(c, \cdot)$ is the density of $N_2(c, \omega^2 I)$ restricted to $c + [-3\omega, 3\omega]^2$.

Spatial covariates may naturally be introduced using a multiplicative model

$$\Lambda(u) = \exp\left(z(u)\beta^{\mathsf{T}}\right) \sum_{(c,\gamma)\in\mathbf{\Phi}} \gamma k(c, u) \tag{4}$$

[15]. A nice feature of the latter model is that the pair correlation function of $\mathbf{X}$ is the same for (2) and (4), i.e. it does not depend on the parameter β. In addition to the log Gaussian Cox process model for the tropical trees, we consider an inhomogeneous Thomas process of the form (4) where the random intensity function of a stationary Thomas process is multiplied by $\exp(\beta_2 z_2(u) + \beta_3 z_3(u))$ where $z_2(u)$ and $z_3(u)$ are specified as for the log Gaussian Cox process model. The intensity function of the inhomogeneous Thomas process is

$$\rho(u) = \kappa\alpha \exp\left(\beta_2 z_2(u) + \beta_3 z_3(u)\right) \tag{5}$$

while the pair correlation function is equal to (3).

4 Simulation-free estimation procedures

In this section $\mathbf{x}$ denotes an observation of $\mathbf{X} \cap W$ where W is a bounded observation window.

4.1 Estimation of the intensity function

Suppose that the intensity function ρ_θ of the point process $\mathbf{X}$ is known on closed form and depends on the unknown parameter θ, cf. (1) and (5). Consider a finite partitioning $C_i,\ i \in I$ of the observation window W into disjoint cells C_i of small areas $|C_i|$, and let u_i denote a representative point in C_i. Let $N_i = \mathbf{1}[\mathbf{X} \cap C_i \neq \emptyset]$ and $p_i(\theta) = \mathrm{P}_\theta(N_i = 1)$. Then $p_i(\theta) \approx \rho_\theta(u_i)|C_i|$ and the composite likelihood based on the $N_i,\ i \in I$, is

$$\prod_{i\in I} p_i(\theta)^{N_i}(1 - p_i(\theta))^{(1-N_i)} \approx \prod_i (\rho_\theta(u_i)|C_i|)^{N_i}(1 - \rho_\theta(u_i)|C_i|)^{1-N_i}.$$

We neglect the factors $|C_i|$ in the first part of the product, since they cancel when we form likelihood ratios. In the limit, under suitable regularity conditions and when the cell sizes $|C_i|$ tend to zero, the log composite likelihood becomes

$$\sum_{u \in \mathbf{x}} \log \rho_\theta(u) - \int_W \rho_\theta(u)\,\mathrm{d}u$$

which coincides with the log likelihood function in the case of a Poisson process with intensity function ρ_θ. The corresponding estimating function is given by the derivative

$$\psi_1(\theta) = \sum_{u \in \mathbf{x}} \mathrm{d} \log \rho_\theta(u)/\mathrm{d}\theta - \int_W (\mathrm{d} \log \rho_\theta(u)/\mathrm{d}\theta) \rho_\theta(u)\,\mathrm{d}u. \tag{6}$$

By the Campbell theorem $\psi_1(\theta) = 0$ is an unbiased estimating equation, and it can easily be solved using e.g. `spatstat` [2] provided ρ_θ is on log linear form.

For both the log Gaussian Cox process model and the inhomogeneous Thomas process model proposed for the tropical tree data, the intensity function is of the form $\exp(z(u)(\tilde{\beta}_1, \beta_2, \beta_3)^\mathsf{T})$ where $\tilde{\beta}_1 = \sigma^2/2 + \beta_1$ in the log Gaussian Cox process case and $\tilde{\beta}_1 = \log(\kappa\alpha)$ for the inhomogeneous Thomas process. Using the estimating function (6) and `spatstat`, we obtain $(\hat{\tilde{\beta}}_1, \hat{\beta}_2, \hat{\beta}_3) = (-4.99, 0.02, 5.84)$. The estimate of course coincides with the MLE under the Poisson process with the same intensity function. Estimates of the clustering parameters (σ^2, ϕ) respectively (κ, ω) may be obtained using minimum contrast estimation, see Section 4.3.

4.2 Asymptotic normality of regression parameter estimates

For an inhomogeneous Neyman-Scott process with random intensity function of the form (4) it is easy to establish asymptotic normality of the estimating function (6) in the asymptotic setup of increasing κ, i.e. increasing intensity of the mother points [15]. Asymptotic normality of the parameter β then follows using standard asymptotic results for estimating functions. Simulation studies in [15] suggest that the approximate normality is valid even for moderate values of κ.

Hence for the inhomogeneous Thomas model used for the tropical tree data, $(\hat{\beta}_2, \hat{\beta}_3)$ is asymptotically normal as $\tilde{\beta}_1 \to \infty$ and we obtain approximate 95% confidence intervals $[-0.02, 0.06]$ and $[0.89, 10.80]$ for β_2 and β_3, respectively. Under the Poisson process model much more narrow approximate 95% confidence intervals $[0.02, 0.03]$ and $[5.34, 6.34]$ are obtained (again the asymptotics is for increasing $\tilde{\beta}_1$).

4.3 Minimum contrast estimation of clustering parameters

The solid curve in Figure 4 shows an estimate of the K-function for the tropical rain forest trees obtained using

$$\hat{K}(r) = \sum_{u,v \in \mathbf{x}} \frac{\mathbf{1}[0 < \|u - v\| \le r]}{\hat{\rho}(u)\hat{\rho}(v)|W \cap W_{u-v}|}$$

[1] with $\hat{\rho}$ given by the estimated parametric intensity function from Section 4.1.

For the inhomogeneous Thomas process, a minimum contrast estimate $(\hat{\kappa}, \hat{\omega}) = (8\mathrm{e}{-5}, 20)$ is obtained by minimizing

$$\int_0^{100} (\hat{K}(r)^{1/4} - K(r; \kappa, \omega)^{1/4})^2 \mathrm{d}r \tag{7}$$

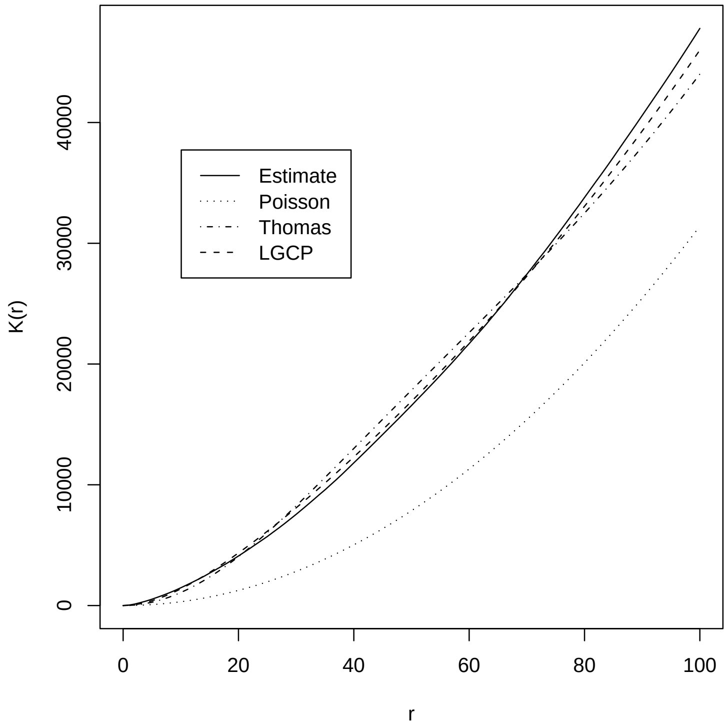

Fig. 4: Estimated K-function for tropical rain forest trees and theoretical K-functions for fitted Thomas, log Gaussian Cox, and Poisson processes.

where

$$K(r;\kappa,\omega) = \pi r^2 + (1 - \exp(-r^2/(4\omega)^2))/\kappa$$

is the theoretical expression for the K-function. For the log Gaussian Cox process, we calculate instead the theoretical K-function

$$K(r;\sigma,\phi) = 2\pi \int_0^r s \exp\left(\sigma^2 \exp(-s/\phi)\right) \mathrm{d}s$$

using numerical integration, and obtain the minimum contrast estimate $(\hat{\sigma}, \hat{\phi}) = (1.33, 34.7)$. The estimated theoretical K-functions are shown in Figure 4.

Minimum contrast estimation is computationally very easy. A disadvantage is the need to choose certain tuning parameters like the upper limit 100 and the exponent 1/4 in the integral (7). Typically, these parameters are chosen on an ad hoc basis.

4.4 An estimating function based on the second order product density

The estimating function (6) is based on the intensity function which is also known as the first order product density. Alternatively, we may consider an estimating function based on the second order product density $\rho_\theta^{(2)}(u,v) = \rho_\theta(u)\rho_\theta(v)g_\theta(u,v)$:

$$\psi_2(\theta) = \sum_{u,v\in\mathbf{x}}^{\neq} \mathrm{d}\log\rho_\theta^{(2)}(u,v)/\mathrm{d}\theta - \int_{W^2} \left(\mathrm{d}\log\rho_\theta^{(2)}(u,v)/\mathrm{d}\theta\right) \rho_\theta^{(2)}(u,v)\, \mathrm{d}u\mathrm{d}v. \tag{8}$$

This is the score of a limit of composite log likelihood functions based on Bernouilli observations $N_{ij} = \mathbf{1}[\mathbf{X} \cap C_i \neq \emptyset, \mathbf{X} \cap C_j \neq \emptyset]$, $i \neq j$. The integral in (8) typically must be evaluated using numerical integration. In the stationary case, [7] considers a related unbiased estimating function, where the integral is replaced by the number of pairs of distinct points times $\log \int_{W^2} \rho^{(2)}(u, v)\, \mathrm{d}u\mathrm{d}v$.

The estimating function (8) allows simultaneous estimation of the parameters $(\tilde{\beta}_1, \beta_2, \beta_3)$ and (κ, ω) for the inhomogeneous Thomas process. For fixed values of ω, $\psi_2(\theta) = 0$ is solved with respect to the other parameters using Newton-Raphson (Newton-Raphson for all the parameters jointly turns out to be numerically unstable). We then search for an approximate solution with respect to ω within a finite set of ω-values. The resulting estimates of $(\tilde{\beta}_1, \beta_2, \beta_3)$ and (κ, ω) are respectively $(-5.00, 0.02, 5.73)$ and $(0.00007, 30)$. The estimate of ω differs considerably from the minimum contrast estimate while the remaining estimates are quite similar to those obtained previously for the inhomogeneous Thomas process. The numerical computation of ψ_2 and its derivatives is quite time consuming and the whole process of solving $\psi_2(\theta) = 0$ takes about one hour.

In a small scale simulation study, ψ_2 provided a slightly more efficient estimate of (β_2, β_3) than ψ_1. On the other hand, the minimum contrast estimate of (κ, ω) was more efficient than the ψ_2 estimate.

5 Likelihood-based inference for North Atlantic whales

For a Cox process, the likelihood function is given by

$$L(\theta) = \mathbf{E}_\theta f(\mathbf{x}|\Lambda)$$

where $f(x|\Lambda)$ is the Poisson process density of $\mathbf{X} \cap W$ given Λ and the expectation is with respect to the distribution of Λ which we assume is parametrized by θ. In general, the only way to compute likelihood ratios or likelihood function derivatives, is to use Monte Carlo based on conditional simulations of $(\Lambda(u))_{u \in W}$ given $\mathbf{X} \cap W = \mathbf{x}$ [12, 14].

For the shot noise Cox process model for the whales in Section 3.2, the unknown parameters are the intensity κ of the cluster centres, the mean number α of whales per cluster, and the standard deviation ω of the Gaussian density. Since the kernel function k is assumed to have bounded support, $(\Lambda(u))_{u \in W}$ is given in terms of a finite Poisson process $\Phi \cap B \times (0, \infty)$ where B is chosen so that points (c, γ) with c outside B do not contribute to $(\Lambda(u))_{u \in W}$. To implement Monte Carlo estimation of the likelihood function it is then required to simulate $\Phi \cap B \times (0, \infty)$ given $\mathbf{X} \cap W = \mathbf{x}$. This is done in [14] using a certain Gibbs sampler strategy where birth-death updates are used to update $\Phi \cap B \times (0, \infty)$ conditional on *all* whales (both the observed whales $\mathbf{X} \cap W = \mathbf{x}$ and those which were missed) and where the unobserved whales are sampled jointly from the Poisson process distribution of the unobserved whales given $\Phi \cap B \times (0, \infty)$ and $\mathbf{X} \cap W = \mathbf{x}$.

Since it is difficult to evaluate the components of the score function and observed information corresponding to the parameter ω, [14] compute the profile likelihood function $l_p(\omega) = \max_{(\kappa, \alpha)} \log L(\theta)$ for a finite set of ω-values $\omega_1, \ldots, \omega_m$. This is done by cumulating log likelihood ratios $\log L(\hat{\theta}_{l+1}) - \log L(\hat{\theta}_l)$ where $\hat{\theta}_l = (\hat{\kappa}_l, \hat{\alpha}_l, \omega_l)$ and $(\hat{\kappa}_l, \hat{\alpha}_l) = \arg\max_{(\kappa, \alpha)} \log L(\kappa, \alpha, \omega_l)$ is obtained using Newton-Raphson. The profile likelihood function is shown in Figure 1 (left plot) and gives $\hat{\omega} = \omega_5 = 0.6$ with corresponding values $\hat{\kappa}_5 = 0.025$ and $\hat{\alpha}_5 = 2.4$ found using Newton-Raphson. These numbers yield an estimated whale intensity of 0.06 whales per km^2 with a 95% parametric bootstrap confidence interval $[0.03, 0.08]$. Figure 1 (right plot) shows the fitted L-function; note the high variability

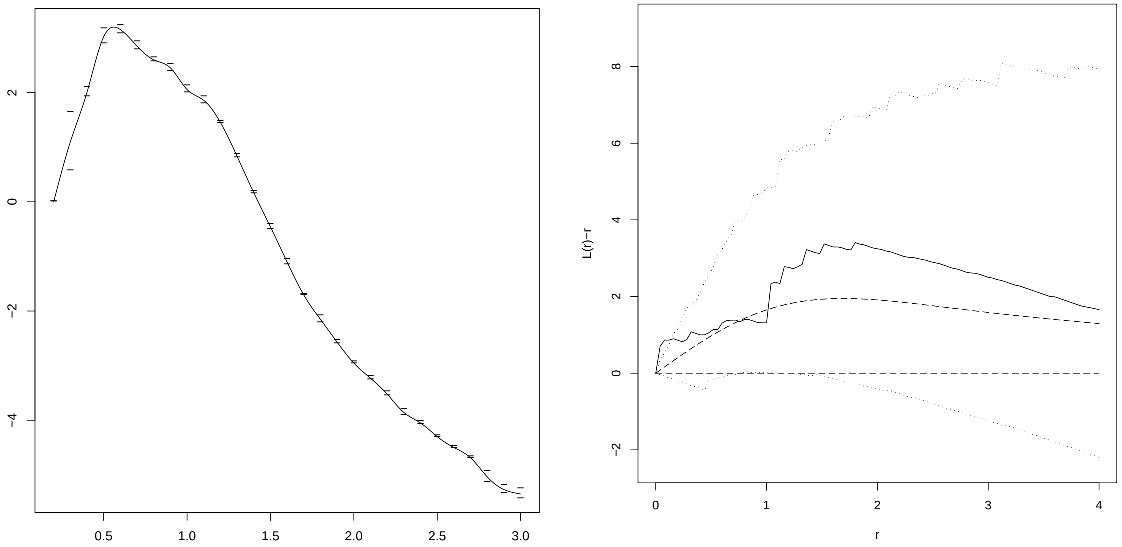

Figure 1: Fitting a shot noise Cox process model to the North Atlantic whales data set. Left: profile log likelihood function $l_p(\omega) = \max_{(\kappa,\alpha)} \log L(\theta)$ obtained by cumulating estimated log likelihood ratios, see text. The small horizontal bars indicate 95% Monte Carlo confidence intervals for the log likelihood ratios. Right: non-parametric estimate of $L(r) - r$ (solid line), 95% confidence envelopes based on simulations of fitted shot noise Cox process (dotted lines), $L(r) - r = 0$ for a Poisson process (lower dashed line), and $L(r) - r > 0$ for the fitted shot noise Cox process (upper dashed line).

of the non-parametric estimate of the L-function, cf. the envelopes computed from simulations of the fitted model. For this particular example, the computation of the profile likelihood function is very time consuming and Monte Carlo error occasionally caused negative definite estimated observed information matrices.

From a computational point of view, the Bayesian approach is quite appealing. The need for computing the likelihood function is eliminated by a demarginalization strategy where the unknown random intensity function or cluster centre process is considered as an unknown parameter along with the original parameter θ. This simplifies computations, since the likelihood of the data given θ and the random intensity function is just a Poisson likelihood function. In [14], the unknown parameters κ, α, and ω are assumed to be a priori independent with uniform priors on bounded intervals for κ and ω and an informative $N(2, 1)$ (truncated at zero) prior for α (the whales are a priori believed to appear in small groups of 1-3 animals). The posterior means for κ, α, and ω are 0.027, 2.2, and 0.7, and the posterior mean of the whale intensity is identical to MLE. There is moreover close agreement between the 95% confidence interval and the 95 % central posterior interval $[0.04, 0.08]$ for the whale intensity.

Bayesian inference for a log Gaussian Cox process in the context of disease mapping is considered by [3].

6 Conclusion

Likelihood-based inference for spatial Cox processes is becoming increasingly feasible as cheap and powerful computers become available. However depending on the particular application, the Markov chain Monte Carlo approach may still require many ours of computing time. Hence there is still need to develop simple simulation-free estimation methods and to study asymptotic properties of the associated parameter estimates.

References

[1] Baddeley, A., Møller, J., and Waagepetersen, R.: *Non- and semi-parametric estimation of interaction in inhomogeneous point patterns.* Statistica Neerlandica **54** (2000), 329–350.

[2] Baddeley, A. and Turner, R.: *Spatstat: an R package for analyzing spatial point patterns.* Journal of Statistical Software **12** (2005), 1–42, URL: `www.jstatsoft.org`, ISSN: 1548-7660.

[3] Beneš, V., Bodlák, K., Møller, J., and Waagepetersen, R.: *A case study on point process modeling in disease mapping.* Image Analysis and Stereology **24** (2005), 159–168.

[4] Condit, R., Hubbell, S. P. and Foster, R. B.: *Changes in tree species abundance in a neotropical forest: impact of climate change.* Journal of Tropical Ecology **12**, 231–256.

[5] Condit, R.: *Tropical Forest Census Plots.* Springer-Verlag, Berlin, and R. G. Landes Company, Georgetown.

[6] Diggle, P. J.: *Statistical analysis of spatial point patterns.* Arnold, London, second edition.

[7] Guan, Y.: *A composite likelihood approach in fitting spatial point process models.* J. Amer. Statist. Assoc. (2006), to appear.

[8] Heinrich, L.: *Minimum contrast estimates for parameters of spatial ergodic point processes.* In: *Transactions of the 11th Prague Conference on Random Processes, Information Theory and Statistical Decision Functions.* Academic Publishing House, Prague 1992, 479-472.

[9] Hubbell, S. P. and Foster, R. B.: *Diversity of canopy trees in a neotropical forest and implications for conservation.* In: Sutton, S. L., Whitmore, T. C. and Chadwick, A. C. (eds.): *Tropical rain forests: ecology and management*, Blackwell Scientific Publications, 1983, 25–41.

[10] Møller, J.: *Shot noise Cox processes.* Advances in Applied Probability **35**, 4–26.

[11] Møller J., Syversveen, A.-R., and Waagepetersen, R.: *Log Gaussian Cox processes.* Scandinavian Journal of Statistics **25** (1998), 451–482.

[12] Møller, J. and Waagepetersen, R.: *Statistical inference and simulation for spatial point processes.* Chapman & Hall/CRC Press, Boca Raton, 2003.

[13] Stoyan, D.: *Statistical estimation of model parameters of planar Neyman-Scott cluster processes.* Metrika **39** (1992), 67–74.

[14] Waagepetersen, R and Schweder, T.: *Likelihood-based inference for clustered line transect data* Submitted, 2005.

[15] Waagepetersen, R.: *An estimating function approach to inference for inhomogeneous Neyman-Scott processes.* Submitted, 2005.