A comparison of spatial point process models in epidemiological applications

Jesper Møller

Aalborg University

1 Introduction

As pointed out in Richardson (2001) (for short, "SR"), point process models provide a natural setting in epidemiological applications. The aims of this contribution are to review the advantages and problems when using such models for aggregated data, compare the properties of different model classes, clarify some of the properties of these model classes in connection to epidemiological applications (which sometimes seem to have been misunderstood), propose some alternative models which are not covered in SR, and point out some open problems. The focus is on three particular model classes: the Heikkinen and Arjas (1998) models, log Gaussian Cox (LGC) processes (Møller et al., 1998), and shot-noise G Cox (SNGC) processes (Brix, 1999). Poisson/gamma models (Daley and Vere-Jones, 1988; Wolpert and Ickstadt, 1998) are special cases of SNGC processes.

2 Brief description of models

Let us start by recalling the notation and definitions used for these models. As in SR, the point process Y of exact locations of cases observed in a bounded planar region R, is assumed to be a Cox process driven by a random intensity function $\Lambda(s),\ s \in R$. That is $Y|\Lambda$ is a Poisson process with intensity function Λ: conditionally on Λ, the number of cases $Y(A)$ in a subregion $A \subseteq R$ is Poisson distributed with mean $\int_A \Lambda(s)ds$, and the number of cases $Y(A_1), \ldots, Y(A_n)$ in disjoint subregions $A_1, \ldots, A_n$ are independent. When dealing with aggregated data $Y(A_1), \ldots, Y(A_n)$, we assume $A_1, \ldots, A_n$ to be a subdivision of R.

In Heikkinen and Arjas (1998), $\Lambda = \sum_k \Lambda_k \mathbf{1}_{A_k}$ where $\{A_k\}$ is the Voronoi tessellation generated by a point process of nuclei $\{g_k\} \subset R$, i.e. A_k is the set of points in R closer to g_k than to any other nuclei. The nuclei follow a homogeneous Poisson process on R, and conditionally on $\{g_k\}$, $\{\log \Lambda_k\}$ is modelled by a conditional autoregression (Besag, 1974).

For a LGC model, $\Lambda(s) = \exp(Z(s))$ where Z is a Gaussian process with mean function $\mu(s) = EZ(s)$ and covariance function $c(s,t) = Cov(Z(s), Z(t))$. To obtain a well-defined Cox process, smoothness properties of c are required (see

Møller et al., 1998). Note that stationarity of Z is not required, but it may be convenient to assume stationarity and isotropy of c so that $c(s,t)$ depends only on the distance $||s-t||$.

The construction of a SNGC process is more complicated, but simplifying the setting a bit, it boils down to the following: $\Lambda(s) = \sum_j k(s,u_j)\gamma_j$ where k is a kernel (for simplicity we assume that $k(\cdot,u)$ is a density function for a continuous random variable), and $\{(u_j,\gamma_j)\} subset E \times [0,\infty)$ where E is a given planar region. Typically in applications, $E = R$, or in order to reduce edge effects, E is much larger than R. Further, $\{(u_j,\gamma_j)\}$ is a Poisson point process with an intensity measure of the product form,

$$\nu(A\times B) = (\alpha(A)/\Gamma(1-\kappa)) \times \int_B \gamma^{-\kappa-1}\exp(-\tau\gamma)d\gamma, \quad A\subseteq E,\ B\subseteq[0,\infty),$$

where $\kappa < 1$ and $\tau \geq 0$ are parameters with $\tau > 0$ if $\kappa \leq 0$, and α is a nonnegative and nonzero Radon measure. If $\kappa < 0$, we obtain a kind of modified Neyman-Scott process as $\{u_j\}$ is a Poisson process with intensity measure $(\tau/-\kappa)\alpha$, and $\{u_j\}$ is independent of the "marks" $\{\gamma_j\}$, which in turn are mutually independent and follow a common Gamma distribution $\Gamma(-\kappa,\tau)$ (in a usual Neyman-Scott process, all marks are equal and deterministic). The situation is less simple for $\kappa \geq 0$ as there are infinite many points $\{u_j\}$. For $\kappa = 0$, we have a Poisson/gamma model. As noticed in Wolpert and Ickstadt (1998), we may extend the model by replacing the parameter τ with a positive function $\tau(u), u\in E$, and redefining

$$\nu(A\times B) = (1/\Gamma(1-\kappa)) \int_A\int_B \gamma^{-\kappa-1}\exp(-\tau(u)\gamma)\alpha(du)d\gamma, \quad A\subseteq E,\ B\subseteq[0,\infty).$$

Extensions to time-space models are studied in Brix and Møller (1998) and Brix and Diggle (2000) for LGC processes, and in Brix and Chadoeuf (1998) for SNGC processes.

3 General properties of the models

The Heikkinen-Arjas and SNGC models can be used for nonparametric Bayesian modelling. LGC processes provide easily interpretable Cox process models with varied degrees of smoothing in the intensity surface $\exp(Z)$, as determined by the smoothness properties of μ and c. Below we discuss the general properties of the varied models, particularly how flexibility they are, and how they may be checked.

Recall that two fundamental characteristics of the Cox process Y are the intensity function $\rho(s) = E\Lambda(s),\ s\in R$, and the pair correlation function $g(s,t) = E(\Lambda(s)\Lambda(t))/(\rho(s)\rho(t)),\ s,t\in R$ (see e.g. Stoyan et al., 1994). Nonparametric estimation of ρ is discussed in Diggle (1985). If $g(s,t)$ depends only on $s-t$, there is a simple relationship between the pair correlation function and the extension to inhomogeneous point processes of Ripley's K-function introduced in Baddeley et al. (2000), who also describe nonparametric estimation of K and g.

To the best of my knowledge, the question of model assessment is little explored for the Heikkinen-Arjas model. Heikkinen and Arjas (1998) stress that although Λ may be expected to be a smooth function but is modelled as a step function, smoothness may be obtained in the posterior mean $E(\Lambda(s)|Y)$. The specification of hyperparameters in the Heikkinen-Arjas model such as the intensity of nuclei and smoothness parameters in the conditional autoregession is a delicate matter. This is similar to the sensitivity of the choice of band width for nonparametric kernel estimation (Diggle, 1985). It is not clear to me to what extent the much more involved Heikkinen-Arjas construction is beneficial as compared to ordinary nonparametric kernel estimation. However the former may better identify rapid or sharp changes in the intensity surface, and posterior mean estimates, uncertainties and other things of interest may be determined by MCMC methods. The method may become even more useful if the Voronoi tessellation $\{A_k\}$ is replaced by one of Nicholls (1998) triangulation models.

LGC processes provide some flexibility in modelling aggregated point patterns when the aggregation is due to spatial heterogeneity. As discussed in Møller et al. (1998), certain Neyman-Scott processes have many properties in common with LGC processes. Note that the distribution of a LGC process restricted to a subregion is easily obtained, so problems with edge effects can be avoided, in particular when Y and $\exp(Z)|Y$ are to be simulated. Another very useful property is that the distribution of Z and hence Y is uniquely determined by the intensity and pair correlation functions, since $\rho(s) = \exp(\mu(s) + c(s,s)/2)$ and $g(s,t) = \exp(c(s,t))$. Due to these simple relationships, nonparametric estimates of ρ and g may suggest an appropriate choice of a parametric model class for μ and c, and model validation may conveniently be based on these and other summary statistics, cf. Møller et al. (1998) and Brix and Møller (1998). Since likelihood based methods are intractable for LGC processes, Møller et al. (1998) advocate the use of minimum contrast methods for estimating the unknown parameters of μ and c. Fully Bayesian analysis of LGC models is an interesting issue of research which is so far less explored.

Even more flexibility may be obtain by using SNGC processes. Closed form expressions of summary statistics are rare; an expression for the pair correlation function is known when the kernel k is Gaussian, see Brix (1999) and Wolpert and Ickstadt (1998). The distribution of a SNGC process restricted to a subregion is in general intractable, and some care may be needed when dealing with edge effects. It remains to compare the applicability of Poisson/gamma and other SNGC processes with LGC processes for Bayesian inference problems based on MCMC methods (see below). Possibly, this is easist for modified Neyman-Scott processes (i.e. when $\kappa < 0$) because of their simple construction.

4 Aggregation coherence and natural scaling

In epidemiological applications, $\Lambda(ds) = p(s)w(ds)$ where, as in SR, we assume w to be a known intensity measure for the individuals at risk, while p is a random risk function of primary interest. In general, when p is modelled in some way by using any of the models considered above, the joint and marginal distributions of aggregated data $Y(A_1), \ldots, Y(A_n)$ are intractable, there is no natural scaling at different levels of aggregation, and care is needed when results at different levels are to be compared, not at least when incorporating covariates. All we know is that given p, the $Y(A_i)$ are conditionally independent and Poisson distributed with means $\Lambda(A_i) = \int_{A_i} p(s)w(ds)$, and the generation and evaluation of such integrals will normally both require simulation methods and numerical techniques. On the other hand, aggregation coherence is not lost as long as we are treating the unobserved (Y, p) as a missing data problem.

SR discusses cases of a LGC process, $p(s) = \exp(Z(s))$, and a Poisson/gamma model, $p(s) = \sum_j \gamma_j k(s, u_j)$, noticing that restrictions are needed to ensure that the risk function takes values in $[0, 1]$ (recall that $p(s)$ is interpreted as a thinning probability). However, numeric evaluation of the integral $\int_{A_i} p(s)w(ds)$ appears not to be more complicated for many other forms of p, including the possibly more natural choice $p(s) = \exp(Z(s))/(1 + \exp(Z(s)))$ where it is ensured that $0 < p < 1$. Similarly, when models with covariates are constructed, from a computational view point it would not make much difference which kind of model is used.

Considering for instance the Poisson/gamma model $p(s) = \sum_j \gamma_j k(s, u_j)$, rather restrictive assumptions are needed to obtain a known distribution: assuming $\{B_l\}$ is a countable subdivision of E so that $\tau(u) = \tau_l$ on each B_l, and $k(s, u) = k_{il}$ on each $A_i \times B_l$, then $\{\Lambda(A_i)\}$ is a kind of moving average gamma field as $\Lambda(A_i) = w(A_i) \sum_l k_{il} \Gamma_l$, where $\Gamma_l \equiv \sum_{u_j \in B_l} \gamma_j \sim \Gamma(\alpha(B_l), \tau_l)$ and the Γ_l are independent.

5 Computational aspects

For all the abovementioned models we need to resort on simulations no matter if the distribution of (Λ, Y), or the conditional distribution of Λ (and possible hyperparameters) given Y, or the conditional distribution of Y, Λ (and possible hyperparameters) given aggregated data $Y(A_1), \ldots, Y(A_n)$ are considered. Since the MCMC algorithms for the different models are rather complicated, it would not make much sense to give the details here. Briefly, Heikkinen and Arjas (1998) use a reversible MCMC algorithm (Green, 1995); Møller et al. (1998) use two-dimensional fast Fourier transforms for circulant embeddings of covariance matrices, and for conditional simulations, Langevin-Hastings type updates; Wolpert and Ickstadt (1998) use an auxiliary variable method.

Notice that some kind of approximation is used in any of the papers: Heikkinen and Arjas (1998) "cheat" when they calculate Hastings ratios involving the normalizing constants in the conditional autoregressions corresponding to two models of different dimensions (the techniques in Rue (2001) may possibly overcome this problem); Møller et al. (1998) discretize the Gaussian field, using a fine grid; Wolpert and Ickstadt (1998) consider only those (u_j, γ_j) with $\gamma_j > \epsilon$, where $\epsilon > 0$ is a given threshold (e.g. the machine precision). Apart from the LGC case, the theoretical properties of the algorithms have so far not been studied in detail. In particular, Møller et al. (1998) establish geometric ergodicity for a truncated Langevin-Hastings algorithm for simulating $\Lambda|Y$.

6 Concluding remarks

All the considered models have their advantages and disadvantages. None makes life easy for spatial epidemiologists, and for that reason direct modelling of aggregated data as described in SR may be hard to beat. Treating the analysis of aggregated data as a missing data problem, MCMC methods may be feasible so that "aggregation coherence" is ensured. Moreover, computational problems with integrals and varied kinds of approximations/discretizations/truncations are typical in the existing MCMC algorithms. Alternative models such as modified Neyman-Scott processes seem yet not used in epidemiological applications, and they leave the hope for developing simpler and yet flexible models.

Acknowledgements

Supported by the European Union's research network "Statistical and Computational Methods for the Analysis of Spatial Data, ERB-FMRX-CT96-0096", by the Centre for Mathematical Physics and Stochastics (MaPhySto), funded by a grant from the Danish National Research Foundation, and by the Danish Natural Science Research Council.

References

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

Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society*, B, **36**, 192–225.

Brix, A. (1999) Generalized gamma measures and shot-noise Cox processes. *Advances in Applied Probability*, **31**, 929–953.

Brix, A. and Chadoeuf, F. (1998) Spatio-temporal modeling of weeds by shot-noise G Cox processes. Report 98-7, Department of Mathematics and Physics, Royal Veterinary and Agricultural University, Copenhagen.

Brix, A. and Diggle, P.J. (2000) Spatio-temporal prediction for log-Gaussian Cox processes. Manuscript, Department of Mathematics and Statistics, Lancaster University.

Brix, A. and Møller, J. (1998) Space-time multitype log Gaussian Cox processes with a view to modelling weed data. Research Report R-98-2012, Department of Mathematical Sciences, Aalborg University. To appear in *Scandinavian Journal of Statistics*.

Daley, D.J. and Vere-Jones, D. (1988) *An introduction to the theory of point processes*, Springer-Verlag, New York.

Diggle, P.J. (1985) A kernel method for smoothing point process data. *Applied Statistics*, **34**, 138–147.

Green, P.J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711–732.

Heikkinen, J. and Arjas, E. (1998) Non-parametric Bayesian estimation of a spatial Poisson intensity. *Scandinavian Journal of Statistics*, **25**, 435–450.

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

Nicholls, G.K. (1998) Bayesian image analysis with Markov chain Monte Carlo and coloured continuum triangulation models. *Journal of the Royal Statistical Society*, B, **60**, 643–659.

Richardson, S. (2001) Spatial models in epidemiological applications. To appear in *Highly structured stochastic systems*, Eds. Green, P.J., Hjort, N.L. and Richardson, S., Oxford University Press.

Rue, H. (2001) Fast sampling of Gaussian Markov random fields with applications. To appear in *Journal of the Royal Statistical Society*, B, **63**.

Stoyan, D., Kendall, W.S. and Mecke, J. (1995) *Stochastic geometry and its applications*, Wiley, New York, 2nd edition.

Wolpert, R.L. and Ickstadt, K. (1998) Poisson/gamma random field models for spatial statistics. *Biometrika*, **85**, 251–267.