Convergence of posteriors for discretized log Gaussian Cox processes

Rasmus Waagepetersen*
Aalborg University

Abstract

In Markov chain Monte Carlo posterior computation for log Gaussian Cox processes (LGCPs) a discretization of the continuously indexed Gaussian field is required. It is demonstrated that approximate posterior expectations computed from discretized LGCPs converge to the exact posterior expectations when the cell sizes of the discretization tends to zero. The effect of discretization is studied in a data example.

Keywords: Bayesian inference, discretization, log Gaussian Cox process, Monte Carlo, point processes, posterior.

1 Introduction

Log Gaussian Cox processes (LGCPs) are Cox point processes with a log Gaussian random intensity function and were independently introduced in astronomy by Coles and Jones (1991) and in statistics by Møller, Syversveen and Waagepetersen (1998). LGCPs have been used for modelling of a wide range of biological point data including point patterns of positions of trees, weed plants, and geographically referenced disease cases, see Møller et al. (1998), Brix and Møller (2001), Brix and Diggle (2001), and Beneš et al. (2002). In these papers the posterior distribution of the unobserved Gaussian field or the random intensity function is computed using Markov chain

*Address for correspondence: Department of Mathematical Sciences, Fredrik Bajers Vej 7G, DK-9220 Aalborg.

Monte Carlo (MCMC) methods where the Gaussian field is approximated by a step function obtained by discretization. In this paper it is verified that approximate posteriors for discretized LGCPs converge to the exact posteriors when the cell sizes for the discretization tends to zero. The dependence of posterior results on the choice of discretization is investigated in a data example with a point pattern of positions of Scots pine saplings.

2 Posterior for discretized LGCPs

A Cox process X with random intensity function Z is a point process such that $X|Z$ is a Poisson process with intensity function Z. For an LGCP X on a set $S \subseteq \mathbb{R}^d$ the random intensity function is given by $Z = \exp(Y)$ where Y is a Gaussian field on S for which the random intensity measure

$$\mu(B) = \int_B \exp(Y(s))\mathrm{d}s$$

is well-defined and finite almost surely for all bounded $B \subseteq S$ (this is e.g. the case if Y is almost surely continuous). In the sequel we let x denote an observed realization of $X \cap W$ where $W \subseteq S$ is an observation window with positive volume $|W| > 0$ and we assume that S is compact. The conditional density of x given Y with respect to a unit rate Poisson process on W is

$$p(x|Y) = \exp\left(|W| - \int_W \exp\big(Y(s)\big)\mathrm{d}s\right) \prod_{\xi \in x} \exp\big(Y(\xi)\big). \tag{1}$$

The posterior distribution of Y given $X \cap W = x$ is specified as follows. For $n \geq 1$ and distinct $s_1, \ldots, s_n \in S$, let $f_{(s_1,\ldots,s_n)}$ denote the density of $(Y(s_1), \ldots, Y(s_n))$. The posterior density of $(Y(s_1), \ldots, Y(s_n))$ given $X \cap W = x$ is given by

$$f_{(s_1,\ldots,s_n)}((y_1, \ldots, y_n)|x) \propto \mathbb{E}[p(x|Y)|Y(s_1) = y_1, \ldots, Y(s_n) = y_n] f_{(s_1,\ldots,s_n)}((y_1, \ldots, y_n)). \tag{2}$$

The posterior of the process Y given $X \cap W = x$ is then given by the consistent set of finite-dimensional posterior distributions with densities of the form (2) for $n \geq 1$ and distinct $s_1, \ldots, s_n \in S$.

The posterior (2) is not analytically tractable and application of Monte Carlo methods is not possible since the conditional density (1) depends on

$Y_W = \{Y(s) : s \in W\}$ which consists of an infinite number of random variables. In practice we therefore approximate Y by a random step function Y^D: let $D = \{C_s\}_{s\in I}$ be a partition of S into disjoint Borel sets C_s indexed by a finite set $I \subset S$ so that $s \in C_s, s \in I$. The step function Y^D then takes constant value $Y(s)$ within C_s.

Consider a measurable functional $h : \mathbb{R}^S \to \mathbb{R}$ where $\mathbb{R}^S$ denotes the set of real valued functions on S. The posterior expectation

$$\mathbb{E}[h(Y)|x] = \mathbb{E}[h(Y)p(x|Y)]/\mathbb{E}[p(x|Y)] \tag{3}$$

is then approximated by

$$\mathbb{E}[h(Y^D)|x] = \mathbb{E}[h(Y^D)p(x|Y^D)]/\mathbb{E}[p(x|Y^D)]. \tag{4}$$

A Monte Carlo estimate of the expectation (4) can be computed using conditional simulations of the finite-dimensional Gaussian vector $(Y(s))_{s\in I}$ given x with conditional density

$$f_D((y(s))_{s\in I}|x) \propto p(x|y^D)f_{(s)_{s\in I}}((y(s))_{s\in I}), \quad (y(s))_{s\in I} \in \mathbb{R}^I.$$

Simulations from the above density can be obtained as in Møller et al. (1998) using MCMC.

3 Convergence of posteriors for discretized log Gaussian Cox processes

Let D^k, $k \geq 1$, with associated index sets I^k and cells $\{C_s^k\}_{s\in I^k}$ define a sequence of partitions of S as in Section 2 so that the cell sizes tend uniformly to zero, i.e. for all $r > 0$ there is a K so that $k \geq K$ implies $C_s^k \subseteq b(s,r)$, $s \in I^k$, where $b(s,r)$ is the ball with centre s and radius r. In Theorem 1 below we verify that the approximate expectation $\mathbb{E}[h(Y^{D^k})|x]$ converges to the exact posterior expectation (3) when k tends to infinity.

Denote by m the mean function of Y and let $\delta = Y - m$ be the zero mean "residual" process which is said to be homogeneous if the covariance $\mathbb{E}\delta(u)\delta(v)$ only depends on $u - v$ for $u, v \in S$. In the following lemma used in the proof of Theorem 1 we require that S is compact and that δ is homogeneous and almost surely continuous. This is typically the case in applications of LGCPs, see Møller et al. (1998).

Lemma 1. *Suppose that S is compact and $\delta = (\delta(s))_{s\in S}$ is homogeneous and almost surely continuous with constant variance σ^2. For any $l \in \mathbb{R}$ and $\psi^2 > \sigma^2$,*

$$\mathbb{E}\exp\big(l\sup_{s\in S}\delta(s) + (\sup_{s\in S}\delta(s))^2/(2\psi^2)\big) < \infty.$$

Proof. Let $0 < \epsilon < (2\sigma^2)^{-1} - (2\psi^2)^{-1}$ and let $\delta_{\text{sup}} = \sup_{s\in S}\delta(s)$. Since $\lim_{t\to\infty} t^{-2}\log P(\delta_{\text{sup}} > t) = -(2\sigma^2)^{-1}$ (see e.g. p. 160 in Adler, 1981) there exists a $T > 0$ such that

$$P(\delta_{\text{sup}} > t) < \exp(t^2(\epsilon - (2\sigma^2)^{-1}))$$

for all $t > T$. Let $F(t) = \exp(lt + t^2/(2\psi^2))$, $\tilde{F}(t) = \exp(t^2/(2\psi^2)$, $t_0 = -l\psi^2$, note that $\delta_{\text{inf}} = \inf_{s\in S}\delta(s)$ is distributed as $-\delta_{\text{sup}}$, and assume without loss of generality that $l \geq 0$. Thereby

$$\begin{aligned}&\mathbb{E}[F(\delta_{\text{sup}})1(\delta_{\text{sup}} < t_0)] \leq \mathbb{E}[\tilde{F}(\delta_{\text{sup}})1(\delta_{\text{sup}} < 0)]\\ &= \mathbb{E}[\tilde{F}(\delta_{\text{inf}})1(\delta_{\text{inf}} > 0)] \leq \mathbb{E}[F(\delta_{\text{sup}})1(\delta_{\text{sup}} \geq t_0)].\end{aligned}$$

Consider next $\mathbb{E}[F(\delta_{\text{sup}})1(\delta_{\text{sup}} \geq t_0)] = \lim_{b\to\infty}\int_{t_0}^b F(t)G(\mathrm{d}t)$ where $G(t) = 1 - P(\delta_{\text{sup}} > t)$. By partial integration, for $b > t_0$,

$$\int_{t_0}^b F(t)G(\mathrm{d}t) = F(b)(G(b)-1) - F(t_0)G(t_0) + F(t_0) + \int_{t_0}^b P(\delta_{\text{sup}} > t)F(\mathrm{d}t).$$

Here

$$\lim_{b\to\infty}|F(b)(G(b)-1)| \leq \lim_{b\to\infty}\exp(lb + b^2(\epsilon - (2\sigma^2)^{-1} + (2\psi^2)^{-1})) = 0$$

and $\lim_{b\to\infty}\int_{t_0}^b P(\delta_{\text{sup}} > t)F(\mathrm{d}t)$ is bounded by

$$(F(T) - F(t_0)) + \int_T^\infty \exp\big(lt + t^2(\epsilon - (2\sigma^2)^{-1} + (2\psi^2)^{-1})\big)(l + t/\psi^2)\mathrm{d}t < \infty.$$

Thus, $\mathbb{E}F(\delta_{\text{sup}}) \leq 2\mathbb{E}[F(\delta_{\text{sup}})1(\delta_{\text{sup}} \geq t_0)]$ is finite. $\square$

Theorem 1. *Let S and δ be as in Lemma 1 and assume that m is bounded and continuous on $S\setminus A$ where $|A| = 0$ and $x \cap A = \emptyset$. Assume further that $h : \mathbb{R}^S \to \mathbb{R}$ is a measurable functional with (H1) $h(y) \leq K\exp\big(l\sup_{s\in S}y(s) + (\sup_{s\in S}y(s))^2/(2\psi^2)\big)$ for some $K > 0$, $l \in \mathbb{R}$, $\psi^2 > \sigma^2$, and all piecewise continuous $y \in \mathbb{R}^S$, and where (H2) $\lim_{k\to\infty}h(Y^{D^k}) = h(Y)$ almost surely. Then*

$$\lim_{k\to\infty}\mathbb{E}[h(Y^{D^k})|x] = \mathbb{E}[h(Y)|x].$$

Proof. From the assumptions on S, m, and δ it follows that Y is almost surely Riemann integrable and m is continuous at the points in x. Combining this with the assumption (H2) yields

$$\lim_{k\to\infty} h(Y^{D^k})p(x|Y^{D^k}) = h(Y)p(x|Y)$$

almost surely. Further, by (H1),

$$\begin{aligned}
&h(Y^{D^k})p(x|Y^{D^k}) \\
&\leq \exp(|W|)K\exp\big(\mathrm{card}(x)\sup_{s\in S} Y(s) + l\sup_{s\in S} Y(s) + (\sup_{s\in S} Y(s))^2/(2\psi^2)\big) \\
&\leq \mathrm{const}\exp\big(\sup_{s\in S}\delta(s)(\mathrm{card}(x) + 2\sup_{s\in S} m(s) + l) + (\sup_{s\in S}\delta(s))^2/(2\psi^2)\big) \quad (5)
\end{aligned}$$

where $\mathrm{card}(x)$ is the number of points in x. The result now follows by dominated convergence since the upper bound in (5) has finite expectation by Lemma 1. □

Remark 1. *In applications one is typically interested in computing the posterior probability of some event (in which case $h(\cdot)$ is an indicator function), posterior moments of $Y(s)$ or $\exp(Y(s))$, $s \in S$, or the posterior expectation of $\mu(S) = \int_S \exp(Y(s))\mathrm{d}s$. All these cases are covered by Theorem 1.*

Remark 2. *Suppose that e.g. a linear model $m(s) = z(s)\beta^{\mathsf{T}}$ is specified for the mean function where $\beta \in \mathbb{R}^p$ and $z(s) = (z_1(s), \ldots, z_p(s))$ is a p-dimensional covariate function where each of the $z_i(\cdot)$ satisfies the assumption on m in Theorem 1. If we specify a prior p for $\theta = (\beta, \sigma^2)$ with compact support Θ then a simple extension of Theorem 1 yields that the approximate posterior $\mathbb{E}[h(Y^{D^k}, \theta)|x]$ converges to $\mathbb{E}[h(Y, \theta)|x]$ whenever $h(Y^{D^k}, \theta)$ is bounded as in (H1) with $\sigma^2 < \psi^2$ for $(\beta, \sigma^2) \in \Theta$.*

4 Example

In the Bayesian data analysis in Beneš et al. (2002) it was observed that the approximate posterior results were rather sensitive to the choice of discretization. Here we consider the effect of discretization in another data example from Møller et al. (1998) (see also Penttinen, Stoyan and Henttonen, 1992, and Stoyan and Stoyan, 1994) with positions of 126 Scots pine saplings observed within a square observation window W which is rescaled to the unit square, see Figure 1 (left).

As in Møller et al. (1998) we assume that Y has constant mean μ and exponential covariance function $c(r) = \sigma^2 \exp(-r/\alpha)$, $r > 0$, where $\sigma^2 > 0$ is the variance and $\alpha > 0$ is the correlation parameter. Møller et al. (1998) obtained so-called minimum contrast estimates $\hat{\mu} = 3.88$, $\hat{\sigma} = 1.38$, and $\hat{\alpha} = 1/33$. In the following we fix α at the minimum contrast estimate and consider Bayesian inference for $\theta = (\mu, \sigma)$. According to Remark 2 the prior

$$p((\mu, \sigma)) = \sigma^{-1} 1\big[\sigma \in [0.001, 10], \mu \in [-10, 10]\big]$$

for μ and σ is chosen to have a compact support. In the following we let $S = W = [0, 1] \times [0, 1]$ and compute approximate posterior expectations and variances for $\int_S \exp(Y(s))\mathrm{d}s$, μ, σ, and $Y(s_i)$ for four locations $s_1 = (0.25, 0.25)$, $s_2 = (0.25, 0.75)$, $s_3 = (0.75, 0.25)$, and $s_4 = (0.75, 0.75)$.

In the computations we consider 9 partitions D^k, $k = 1, \ldots, 9$, where S is divided into square cells C_s^k indexed by the set I^k of cell center points and of sidelength $1/n_k$, where n_k takes the values 9, 17, 33, 65, 81, 97, 129, 193, 257 for $k = 1, \ldots, 9$. As an example, the partition D^4 is shown in the left plot in Figure 1. Simulations for the Monte Carlo estimation are obtained using the Langevin-Hastings MCMC algorithm described in Møller et al. (1998) extended with Metropolis-Hastings updates of μ and σ.

Monte Carlo estimates of the approximate posterior means and variances for the various statistics are plotted against the discretization index k in Figure 1 (right), Figure 2, and Figure 3. The size of the Monte Carlo error is indicated by small horizontal lines in the plots giving lower and upper 95% confidence limits for the Monte Carlo estimates. The confidence intervals are quite wide for the last two discretizations where smaller MCMC samples were used due to long computing time. The Monte Carlo standard errors are computed using Geyer (1992)'s initial monotone sequence estimate.

Except in the case of $\int_S \exp(Y(s))\mathrm{d}s$ the approximate posterior means and variances seem rather sensitive to the choice of discretization. For practical purposes values of $n_k \geq n_5 = 81$ might be judged to suffice since the approximate values of the posterior characteristics are reasonable stable for discretizations finer than n_5. Incidentally, the posterior means of μ and σ for the finest discretizations are rather close to the minimum contrast estimates.

By Theorem 1, approximate posterior expectations for discretized LGCPs will converge to the exact posterior expectations. However, the data example illustrates that approximate posterior results obtained with a given discretization should be interpreted with care. Preferably one should study the

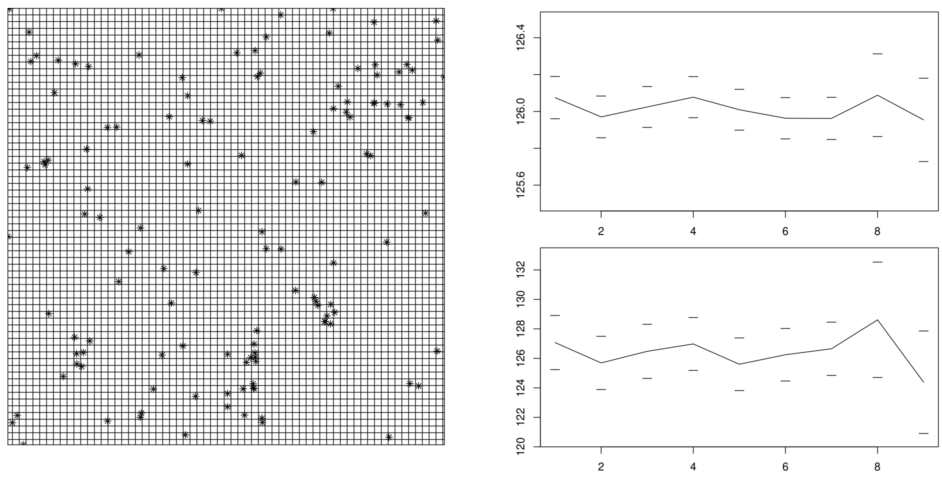

Figure 1: Left: positions of 126 scots pine saplings. The grid shows the partition D^4 used for approximate posterior computation. Right: Monte Carlo estimates of posterior means (top) and posterior variances (bottom) for $\int_S \exp(Y^{D^k}(s))\mathrm{d}s$ plotted against discretization index k. The small horizontal lines indicate confidence intervals for the Monte Carlo estimates.

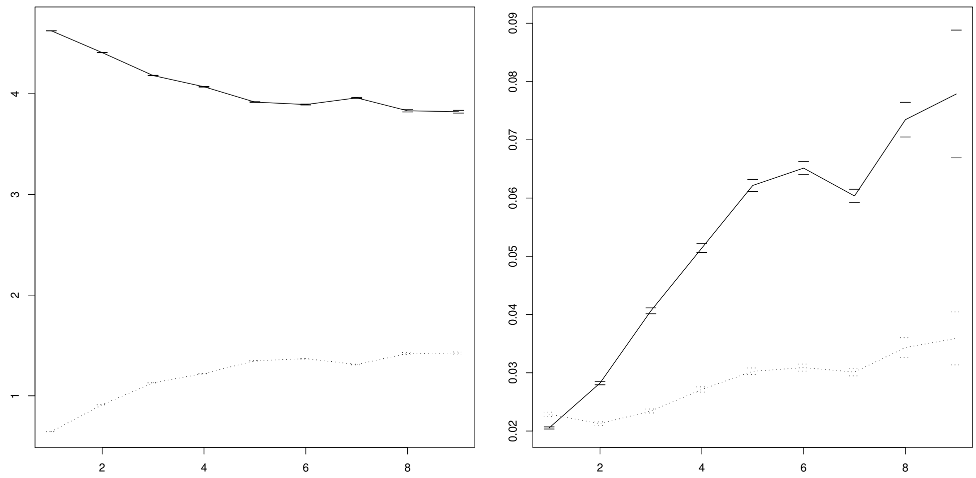

Figure 2: Monte Carlo estimates of approximate posterior means (left) and variances (right) for μ (solid) and σ (dotted) plotted against discretization index k. The small horizontal lines indicate confidence intervals for the Monte Carlo estimates.

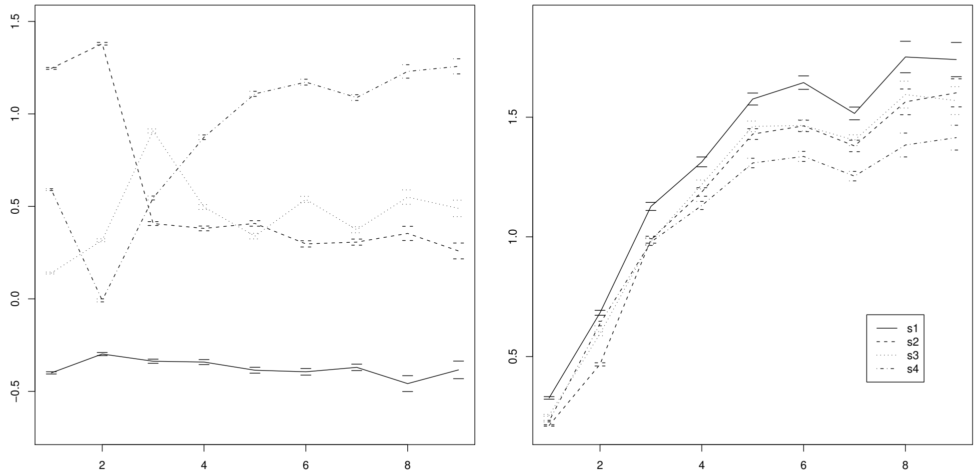

Figure 3: Monte Carlo estimates of posterior means (left) and variances (right) for $Y^{D^k}(s_i)$, $i = 1, \dots, 4$, plotted against discretization index k. The small horizontal lines indicate confidence intervals for the Monte Carlo estimates.

sensitivity to the choice of discretization by considering different discretizations.

Acknowledgments Helpful comments from reviewers of previous versions of the manuscript are gratefully acknowledged.

References

Adler, R. (1981). *The Geometry of Random Fields*, Wiley, New York.

Beneš, V., Bodlák, K., Møller, J. and Waagepetersen, R. P. (2002). Bayesian analysis of log Gaussian Cox process models for disease mapping, *Technical Report R-02-2001*, Department of Mathematical Sciences, Aalborg University.

Brix, A. and Diggle, P. J. (2001). Spatio-temporal prediction for log-Gaussian Cox processes, *Journal of the Royal Statistical Society Series B* **63**: 823–841.

Brix, A. and Møller, J. (2001). Space-time multitype log Gaussian Cox processes with a view to modelling weed data, *Scandinavian Journal of Statistics* **28**: 471–488.

Coles, P. and Jones, B. (1991). A lognormal model for the cosmological mass distribution, *Monthly Notices of the Royal Astronomical Society* **248**: 1–13.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo (with discussion), *Statistical Science* **8**: 473–483.

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

Penttinen, A., Stoyan, D. and Henttonen, H. M. (1992). Marked point processes in forest statistics, *Forest Science* **38**: 806–824.

Stoyan, D. and Stoyan, H. (1995). *Fractals, Random Shapes and Point Fields*, Wiley, Chichester.