Bayesian Parameter Estimation for the Log-Gaussian Cox Process

P. Diggle†, B. Taylor†, T. Davies‡, B. Rowlingson†

† Lancaster University, UK ‡ Massey University, NZ

email: b.taylor1@lancaster.ac.uk

Log-Gaussian Cox Processes

Conditional on a spatially and temporally continuous Gaussian field, $\mathscr{Y}$, observations, X, are assumed to arise from a Poisson process with rate function,

$$\pi(X_t(s,t)|\mathscr{Y}_t(s,t),\xi) \sim \text{Poisson}\{\lambda(s)\lambda(t)\exp[\mathscr{Y}_t(s,t)]\}, \tag{1}$$

where $\lambda(s)$ and $\lambda(t)$ are respectively deterministic spatial and temporal components of the intensity. It is common to approximate $\mathscr{Y}$ by a discrete version, Y, see [7, 2].

In the above, $\xi = (\sigma, \phi, \mu, \theta)$ are the parameters of Y. Conditional on a realisation of the field, y_0, at time 0 say, the transition density of the field at time t, corresponding to the random variable Y_t is,

$$\pi(Y_t|y_0,\xi) \sim \text{N}[y_t; a(t)+b(t)y_0, G(t)]. \tag{2}$$

Under the assumptions of [2],

$$\begin{aligned} a(t) &= \mu[\mathbf{1} - \exp(-\theta t)], \\ b(t) &= \exp(-\theta t)\mathbb{I}, \\ G(t) &= [1-\exp(-2\theta t)]R, \end{aligned}$$

where $\mathbb{I}$ is identity matrix and R is the spatially stationary covariance on the grid and $\mathbf{1} = (1,\ldots,1)^T$. For any two s_1, s_2 on the grid, the spatial covariance between the two locations is:

$$R(s_1,s_2) = \sigma^2 r(-||s_1-s_2||/\phi),$$

for some correlation function r.

The Difficulty with Vanilla Metropolis-Hastings

Assume that observations $x_{1:T}$ have arisen conditional on the latent field $y_{1:T}$ which is parametrised by ξ. Bayesian inference about ξ is via the posterior, $\pi(\xi|x_{1:T})$, which is proportional to $\pi(x_{1:T}|\xi)\pi(\xi)$. A Metropolis-Hastings algorithm [6, 5] to sample from this posterior would work as follows.

Suppose the current value of the chain is ξ. Sample $\xi^\star$ from a density $q(\xi,\xi^\star)$ and accept the move (ie set $\xi \leftarrow \xi^\star$) with probability,

$$\min\left\{1, \frac{\pi(x_{1:T}|\xi^\star)\pi(\xi^\star)}{\pi(x_{1:T}|\xi)\pi(\xi)}\frac{q(\xi^\star,\xi)}{q(\xi,\xi^\star)}\right\}. \tag{3}$$

Unfortunately the expression,

$$\pi(x_{1:T}|\xi) = \int \pi(x_{1:T}|y_{1:T},\xi)\pi(y_{1:T}|\xi)\mathrm{d}y_{1:t}, \tag{4}$$

involves a high-dimensional and analytically intractable integral.

A Useful Sequential Monte Carlo Identity

Suppose that a weighted sample of size M from the discretised field conditional on the data,

$$S = \{y_{1:T}^{(j)}, W^{(j)}\}_{j=1}^M \sim \pi(y_{1:T}|x_{1:T},\xi),$$

is available; here the $W^{(j)}$ are normalised weights.

Consider an importance re-weighting step to move the sample S from being distributed as $\pi(y_{1:T}|x_{1:T},\xi)$, to being distributed as $\pi(y_{1:T}|x_{1:T},\xi^\star)$. The jth importance weight is proportional to,

$$w^{(j)} = \frac{\pi(x_{1:T}|y_{1:T}^{(j)},\xi^\star)}{\pi(x_{1:T}|y_{1:T}^{(j)},\xi)}\frac{\pi(y_{1:T}^{(j)}|\xi^\star)}{\pi(y_{1:T}^{(j)}|\xi)} = \frac{\pi(y_{1:T}^{(j)}|\xi^\star)}{\pi(y_{1:T}^{(j)}|\xi)}, \tag{5}$$

this gives $S^\star = \{y_{1:T}^{(j)}, W^{(j)}w^{(j)}\}_{j=1}^M \sim \pi(y_{1:T}|x_{1:T},\xi^\star)$.

It turns out that the ratio of the normalising constants of the densities $\pi(y_{1:T}|x_{1:T},\xi^\star)$ and $\pi(y_{1:T}|x_{1:T},\xi)$ can be estimated unbiasedly by the sum of the unnormalised importance weights, that is,

$$\frac{\widehat{\pi(x_{1:T}|\xi^\star)}}{\pi(x_{1:T}|\xi)} = \sum_{j=1}^M W^{(j)}w^{(j)}, \tag{6}$$

see [3]. This unbiased estimate will replace the exact version in equation (3) and motivates the new algorithm.

Importance-Weighted MCMC

The proposed algorithm (IWMCMC) proceeds from an initial choice of parameters, ξ, say. Using the MALA algorithm of [2], sample $\{y_{1:T}^{(j)}, W^{(j)}\}_{j=1}^M \sim \pi(y_{1:T}|x_{1:T},\xi)$, where $W^{(j)} = 1/M$ for all j. The choice of the initial parameters should be such that the MALA chain mixes well. IWMCMC proceeds as follows:

1. Propose a move from ξ to $\xi^\star$ from q.
2. Accept the move with probability,

$$\min\left\{1, \frac{\pi(\xi^\star)}{\pi(\xi)}\frac{q(\xi^\star,\xi)}{q(\xi,\xi^\star)}\sum_{j=1}^M W^{(j)}w^{(j)}\right\},$$

where $w^{(j)}$ is as defined in (5).

3. If the move *IS accepted*:
   - Update the parameter $\xi \leftarrow \xi^\star$.
   - Compute new normalised importance weights,

   $$W^{(j)} \leftarrow W^{(j)}w^{(j)} \Big/ \textstyle\sum_{i=1}^M W^{(i)}w^{(i)}$$

   where the Ws on the RHS of the above are the old values.
4. Else if the move is *NOT accepted*:
   - Update the parameter $\xi \leftarrow \xi$, ie stay put.
5. The sample $\{y_{1:T}^{(j)}, W^{(j)}\}_{j=1}^M \sim \pi(y_{1:T}|x_{1:T},\xi)$. Go to 1.

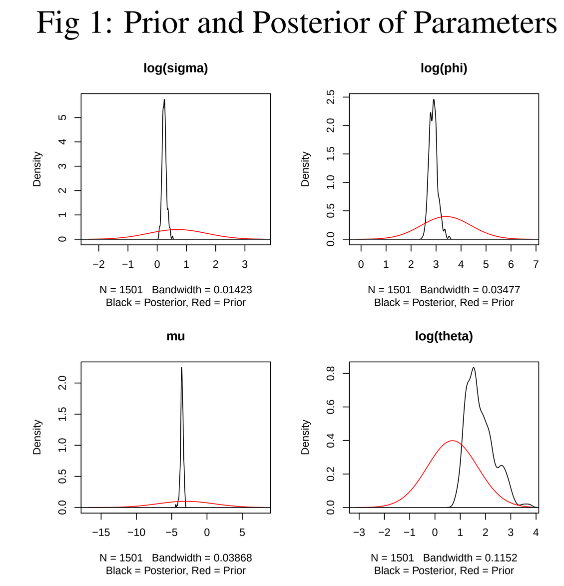
Simulation Study

Data were simulated from a discrete approximation to a log-Gaussian Cox process with parameters,

$$\xi = \{\sigma = 2,\ \phi = 30,\ \mu = -3,\ \theta = 2\}, \tag{7}$$

on an observation window of dimension 100×100 units and for a time period of length 11 days. The MALA algorithm was run for 700000 iterations, 200000 of which were discarded as burn-in; the chain was thinned by 100 and the last three days of data were used to estimate the parameters.

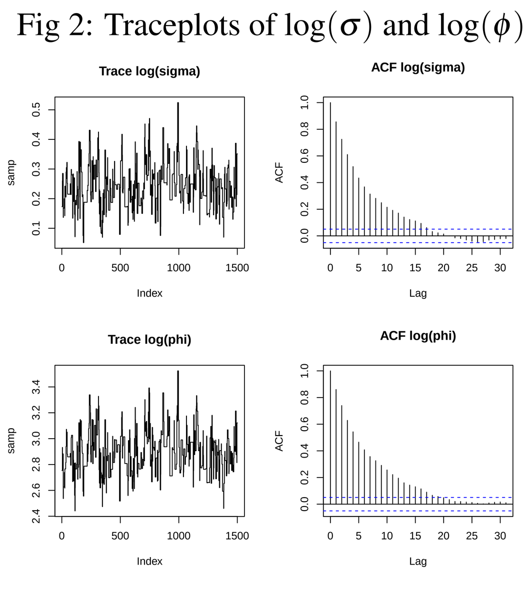
Fig 1: Prior and Posterior of Parameters



IWMCMC was run for 3000 iterations, discarding the first 1500 as burn-in. Figures 1–3 show plots of the prior and posterior of the parameters as well as trace and autocorrelation plots. MCMC for the parameters took place on the log scale for for σ, ϕ and θ.

A combination of the adaptive MCMC methods of [4] and [1] were used to automatically tune both the MALA as well as IWMCMC. The proposal kernel q was a Gaussian density centred on the current ξ; after 1000 iterations, the empirical covariance of the chain-to-date was used to inform the proposal variance.

Fig 2: Traceplots of $\log(\sigma)$ and $\log(\phi)$



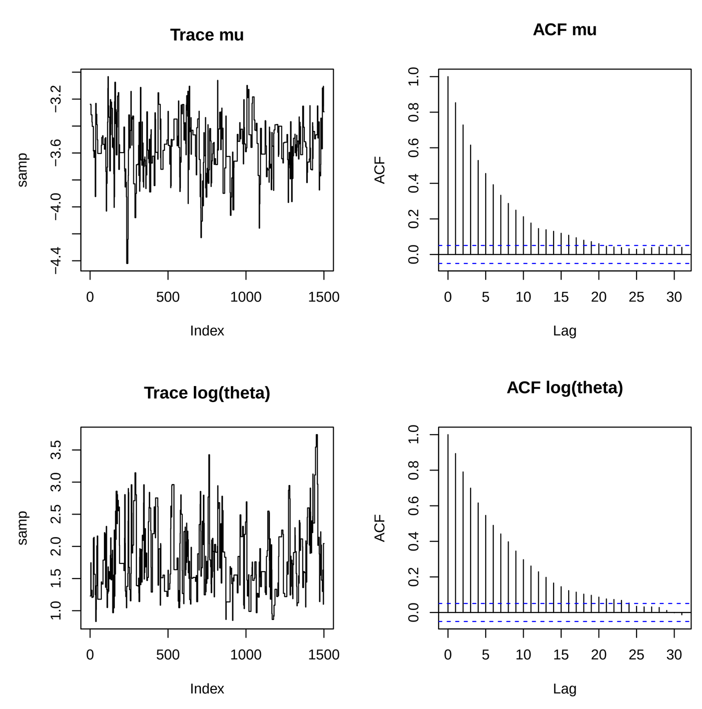
The posterior credible intervals for the parameters were:

Parameter	Median	0.025%	0.975%
σ	1.27	1.11	1.55
ϕ	18.05	13.37	28.31
μ	-3.51	-3.92	-3.03
θ	5.66	2.69	19.28

Conclusions and Further Work

This poster has introduced a new MCMC method for sampling from the posterior density of the parameters of a latent random variable given a set of realisations of a random process conditional on the unobserved latent variable. Early simulation results indicate that the method has potential to perform well for Bayesian inference with log-Gaussian Cox processes.

Fig 3: Traceplots of μ and $\log(\theta)$



Further work in this area will seek to explore the influence of the chosen parametrisation of $\mathscr{Y}$ on the performance of IWMCMC. Also of interest is the possibility of using particle-based methods; either as an in-line solution, or as an alternative to the proposed IWMCMC. A further avenue for research concerns the development of robust methods for choosing a good initial ξ; though various *ad-hoc* methods already exist, eg [2].

R Package

IWMCMC will be released as part of an R package, `lgcp`, for inference with log-Gaussian Cox processes.

References

[1] Christophe Andrieu and Johannes Thoms. A tutorial on adaptive MCMC. *Statistics and Computing*, 18(4):343–373, 2008.

[2] A. Brix and P. J. Diggle. Spatiotemporal prediction for log-gaussian cox processes. *Journal of the Royal Statistical Society, Series B*, 63(4):823–841, 2001.

[3] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006.

[4] Heikki Haario, Eero Saksman, and Johanna Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7:223–242, 1998.

[5] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[6] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[7] Jesper Møller, Anne Randi Syversveen, and Rasmus Plenge Waagepetersen. Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25(3):451–482, 1998.

†Address for correspondence: Department of Mathematics and Statistics, Fylde College, Lancaster University, Lancaster, LA1 4YF