Numerical Approximations of Log Gaussian Cox Processes (Student Abstract)

Francois Buet-Golfouse,[1] Hans Roggeman,[2]

[1] Department of Mathematics, University College London
[2] Independent Researcher
ucahfbu@ucl.ac.uk, hraoyama@gmail.com

Abstract

This paper considers a multi-state Log Gaussian Cox Process ("LGCP") on a graph, where transmissions amongst states are calibrated using a non-parametric approach. We thus consider multi-output LGCPs and introduce numerical approximations to compute posterior distributions extremely quickly and in a completely transparent and reproducible fashion. The model is tested on historical data and shows very good performance.

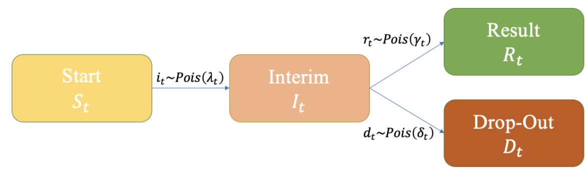

Figure 1: A simple Graph Log Gaussian Cox Process.

Introduction

Our paper builds on (Álvarez et al. 2010; Aglietti, Damoulas, and Bonilla 2019) and considers Log Gaussian Cox Processes to model the movements of agents across multiple states in a graph. Our viewpoint is mostly practical, as it is key to have access to closed-form and numerically stable expressions that can be implemented, run and checked extremely quickly by specialists and non-specialists alike.

Contributions First, we define a graph Log Gaussian Cox Process, which corresponds to estimating the probability structure of a probabilistic graphical model ("PGM") at the population level via a multi-output LGCP, and extend results from the single-state case. Second, we derive novel pseudo-closed form expressions for the posterior distribution of the graph LGCP, thus enabling quick and reproducible computations and avoiding the computational burden of existing techniques.

Graph Log Gaussian Cox Process

For the sake of simplicity, we limit ourselves to a basic directed graph representing the path of agents seeking to accomplish a given task. Indeed, we consider time series (over $t = 1, \cdots, T$ periods) of count data, S_t, I_t, R_t and D_t, representing the count of agents in, respectively, the Start, Interim, Result and Drop-Out states. In addition, we consider the *changes* in those quantities and denote those by lowercase letters, so that $i_t := I_t - I_{t-1}$ represents the number of newcomers in the interim state in period t. In keeping with the literature (as we generalise (Møller, Syversveen, and Waagepetersen 1998)), we posit Poisson-type dynamics for the daily changes in states' populations:

$$i_t \sim \text{Pois}(\lambda_t), r_t \sim \text{Pois}(\gamma_t), d_t \sim \text{Pois}(\delta_t), \quad (1)$$

where $\lambda_t, \gamma_t, \delta_t$ are (random) hazard rate functions. Importantly, this is a *macro*-level model in the sense that it ignores more refined segments in the population of agents or individual features and considers the *overall* population flows. However, our setup can be extended to more structured populations.

We now specify the functions $\lambda_t, \gamma_t, \delta_t$ as exponentials of GPs, or Gaussian processes (Rasmussen and Williams 2006):

- $\log \lambda_t = f_t^\lambda \sim \text{GP}(\mathbf{K}^\lambda)$
- $\log \gamma_t = f_t^\gamma \sim \text{GP}(\mathbf{K}^\gamma)$
- $\log \delta_t = f_t^\delta \sim \text{GP}(\mathbf{K}^\delta)$[1].

Thus, we model the latent features as GPs and use time as well as additional state-dependent features as inputs to the covariance matrix. [2] We denote by $\mathbf{f}^\xi$ the $T \times 1$ vector of latent features for each $\xi = \lambda, \gamma, \delta$, by $\mathbf{y}^\xi$ the related vector of counts[3] (i.e., new interim, new result or new drop-out agents), and the corresponding $T \times T$ kernel matrix by $\mathbf{K}^\xi$.

Deriving the Posterior Distribution

For $\xi = \lambda, \gamma, \delta$, since $\mathbf{y}_t^\xi | \xi_t \sim \text{Pois}(\xi_t)$ for $t = 1, \cdots, T$, the posterior distribution of the latent feature vector $\mathbf{f}^\xi$, con-

[1] We use $\text{GP}(\mathbf{K}^\xi)$ as shorthand notation for a Gaussian Process with zero mean whose kernel matrix entries depend on a chosen kernel and a given input feature vector. We could consider a multi-output GP introducing dependence amongst the different states.

[2] For the sake of simplicity, we choose an RBF kernel and a zero mean function for the GPs but obtained very similar results with a Matern 3/2 kernel. We also explored a multi-output LGCP.

[3] For instance, $\mathbf{y}^\lambda = [i_1, \cdots, i_T]^T$.

ditional on the observed vector of counts $\mathbf{y}^\xi$, is simply

$$p(\mathbf{f}^\xi|\mathbf{y}^\xi) \propto \exp\left({\mathbf{y}^\xi}^T\mathbf{f}^\xi - \sum_{i=1}^{T} e^{f_i^\xi} - \frac{1}{2}\left(\mathbf{f}^\xi - \mathbf{m}^\xi\right)^T {\mathbf{K}^\xi}^{-1}\left(\mathbf{f}^\xi - \mathbf{m}^\xi\right)\right), \quad (2)$$

as indicated in (Diggle et al. 2013). Because this posterior distribution is not tractable, one must resort to numerical or probabilistic techniques to derive it.

Local Laplace: A Closed-Form Approximation

We introduce a new technique related to the Laplace approximation, but allowing to bypass any optimisation step to derive closed-form expression. Indeed, the posterior distribution $p(\mathbf{f}^\xi|\mathbf{y}^\xi)$ can be approximated by:

$$p(\mathbf{f}^\xi|\mathbf{y}^\xi) = \phi_{\boldsymbol{\mu}^\xi,\boldsymbol{\Sigma}^\xi}(\mathbf{f}^\xi), \quad (3)$$

where $\phi_{\boldsymbol{\mu}^\xi,\boldsymbol{\Sigma}^\xi}$ is the multivariate Gaussian density with the parameters (where $\hat{\mathbf{f}}_t^\xi = \log(\mathbf{y}_t^\xi)$ and $\mathbf{H}^\xi = \mathrm{diag}(\mathbf{y}^\xi)$):

$$\begin{aligned} \boldsymbol{\mu}^\xi &= \boldsymbol{\Sigma}^\xi\left({\mathbf{K}^\xi}^{-1}\mathbf{m}^\xi + \mathbf{H}^\xi\hat{\mathbf{f}}^\xi\right) \\ \boldsymbol{\Sigma}^\xi &= \left({\mathbf{K}^\xi}^{-1} + \mathbf{H}^\xi\right)^{-1}. \end{aligned}$$

Total Count Probabilities

Supposing now that we have derived a (Gaussian) posterior distribution, say $f_t^\xi = \log\xi_t \sim N(\mu_{\xi,t}, \sigma_{\xi,t}^2)$, and since $\mathbf{y}_t^\xi|\xi_t \sim \mathrm{Pois}(\xi_t)$, we can derive an approximate distribution of $\mathbf{y}_t^\xi$, for $\xi = \lambda, \gamma, \delta$:

$$\mathbb{P}\left(\mathbf{y}_t^\xi = k\right) = \frac{\Gamma(\alpha_t^\xi + k)}{\Gamma(k+1)\Gamma(\alpha_t^\xi)}\left(\frac{\beta_t^\xi}{\beta_t^\xi + 1}\right)^{\alpha_t^\xi}\left(\frac{1}{\beta_t^\xi + 1}\right)^k, \quad (4)$$

where $\alpha_t^\xi = \frac{1}{e^{\sigma_{\xi,t}^2} - 1}$ and $\beta_t^\xi = \frac{e^{-\left(\mu_{\xi,t} + \frac{\sigma_{\xi,t}^2}{2}\right)}}{e^{\sigma_{\xi,t}^2} - 1}$, thanks to matching moments of Gamma and lognormal distributions. Importantly, this is a (simple) negative binomial distribution. Thus, combining our proposed local Laplace approximation with this expression implies that all quantities of interest are available in closed-form and do not require any optimisation.

Results and Benchmarking Exercise

Techniques such variational inference ("VI") or Markov Chain Monte Carlo ("MCMC") are usually employed (Rasmussen and Williams 2006; Teng, Nathoo, and Johnson 2017) to derive posterior distributions. To test our analytical approach, we have thus run MCMC, VI, the Laplace approximation and our proposal on the "Road casualties in Great Britain" data (Harvey and Durbin 1986)[4]. (For the sake of brevity, only the number d_t of killed drivers is shown here).

[4]Being a "driver", a "killed driver" and a "killed van driver" are considered as states.

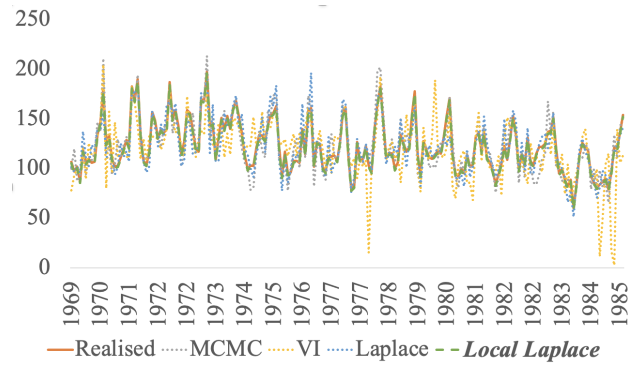

Figure 2: Time series of new deceased drivers d_t per quarter in Britain, 1969-1986, realised and estimated via different techniques.

Discussion

LGCPs are useful tools for modelling, as combining agent dynamics on a graph with a Bayesian non-parametric calibration results in a flexible and yet parsimonious tool. We have shown that such a framework yielded very good empirical results. We have introduced a number of numerical approximations that allow fast, reproducible and robust computations of the posterior distribution, and thus the distribution of the agent population on the graph.

References

Aglietti, V.; Damoulas, T.; and Bonilla, E. 2019. Efficient Inference in Multi-task Cox Process Models. *Proceedings of Machine Learning Research*, 89. ArXiv: 1805.09781.

Diggle, P. J.; Moraga, P.; Rowlingson, B.; and Taylor, B. M. 2013. Spatial and Spatio-Temporal Log-Gaussian Cox Processes: Extending the Geostatistical Paradigm. *Statistical Science*, 28(4): 542–563. Publisher: Institute of Mathematical Statistics.

Harvey, A. C.; and Durbin, J. 1986. The Effects of Seat Belt Legislation on British Road Casualties: A Case Study in Structural Time Series Modelling. *Journal of the Royal Statistical Society. Series A (General)*, 149(3): 187–227.

Møller, J.; Syversveen, A. R.; and Waagepetersen, R. P. 1998. Log Gaussian Cox Processes. *Scandinavian Journal of Statistics*, 25(3): 451–482.

Rasmussen, C. E.; and Williams, C. K. I. 2006. *Gaussian Processes for Machine Learning*. Cambridge, Mass: MIT Press. ISBN 978-0-262-18253-9.

Teng, M.; Nathoo, F.; and Johnson, T. D. 2017. Bayesian computation for Log-Gaussian Cox processes. *Journal of Statistical Computation and Simulation*, 87(11).

Álvarez, M.; Luengo, D.; Titsias, M.; and Lawrence, N. D. 2010. Efficient Multioutput Gaussian Processes through Variational Inducing Kernels. *Proceedings of Machine Learning Research*, 9: 25–32.