A toolbox for fitting non-separable space-time log-Gaussian Cox models using R-INLA

Elias T. Krainski

Universidade Federal do Paraná
Departamento de Estatística
Laboratório de Estatística e Geoinformação

UseR!2019, July 9-12, 2019

Sumário

Burkitt data, splancs (Rowlingson and Diggle 1993)

- Observed time lymphoma cases from year 1960 to 1975 (16 years)

Burkitt data, `splancs` (Rowlingson and Diggle 1993)

- Observed time lymphoma cases from year 1960 to 1975 (16 years)

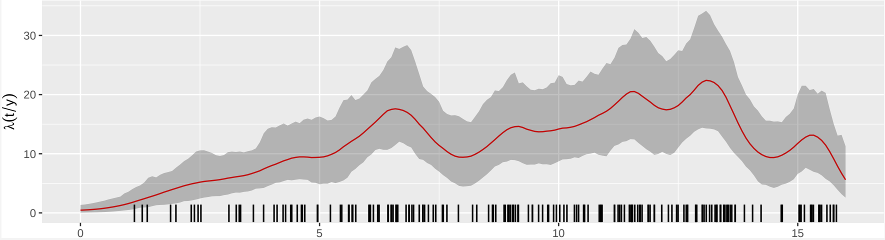
- **Model**: intensity function $\lambda(t)$ to describe how likely a case is to happen at time t?

Burkitt data, splancs (Rowlingson and Diggle 1993)

- Observed time lymphoma cases from year 1960 to 1975 (16 years)

- **Model**: intensity function $\lambda(t)$ to describe how likely a case is to happen at time t?

Fitted intensity function and 95% credibility interval



Observed vs expected number of cases

```
sum(dat1$year<5) # Number of observed cases (year < 5)
## [1] 27
head(pred.res, 5)
##    year mean   sd q0.025 median q0.975  smin smax   cv  var
## 1 0.000 0.45 0.39  0.035   0.37    1.3 0.018  1.8 0.86 0.15
## 2 0.083 0.49 0.40  0.043   0.40    1.4 0.021  1.8 0.83 0.16
## 3 0.167 0.53 0.42  0.059   0.45    1.5 0.025  1.8 0.80 0.18
## 4 0.250 0.57 0.44  0.078   0.49    1.6 0.031  1.9 0.76 0.19
## 5 0.333 0.63 0.46  0.092   0.52    1.8 0.043  2.0 0.73 0.21
sum(pred.res$mean[pred.res$year<5]/12) # Expected (year < 5)
## [1] 24
c(nrow(dat1), sum(pred.res$mean)/12) # Total Obs. and Exp.
## [1] 188 183
```

The data and modeling approaches

- Observed data: $\mathbf{y}$, n events, recorded as observed time points, $\mathbf{y} : t_1, t_2, ..., t_n$

The data and modeling approaches

- Observed data: $\mathbf{y}$, n events, recorded as observed time points, $\mathbf{y} : t_1, t_2, ..., t_n$

Descriptive approach

- **Kernel** intensity estimation

The data and modeling approaches

- Observed data: $\mathbf{y}$, n events, recorded as observed time points, $\mathbf{y} : t_1, t_2, \ldots, t_n$

Descriptive approach

- **Kernel** intensity estimation

Common approach

- Aggregate the data into discrete grid cells
- Model the number of cases in each grid cell t as Poisson(λ_t)
- Model the log(λ_t)
  - pick the **model** framework you like the most

The data and modeling approaches

- Observed data: $\mathbf{y}$, n events, recorded as observed time points, $\mathbf{y} : t_1, t_2, ..., t_n$

Descriptive approach

- **Kernel** intensity estimation

Common approach

- Aggregate the data into discrete grid cells
- Model the number of cases in each grid cell t as Poisson(λ_t)
- Model the log(λ_t)
    - pick the **model** framework you like the most

Point process modeling approach

- Treat the data as it is, no aggregation

Sumário

The log-Likelihood for a set of points

- Given the domain $[0, L]$ and $\mathbf{y}$

$$l(\lambda(.)|\mathbf{y}) = \boxed{L} - \log\left(\int_0^L \boxed{\lambda(t)} \partial t\right) + \sum_{i=1}^{n} \log(\lambda(\boxed{t_i}))$$

  - $\boxed{L}$ is the size of the time domain
  - $\boxed{\lambda(t)}$ is the intensity at time t
  - $\boxed{t_i}$ are time coordinates of the observed cases

The log-Likelihood for a set of points

- Given the domain $[0, L]$ and $\mathbf{y}$

$$l(\lambda(.)|\mathbf{y}) = \boxed{L} - \log\left(\int_0^L \boxed{\lambda(t)}\partial t\right) + \sum_{i=1}^{n}\log(\lambda(\boxed{t_i}))$$

  - $\boxed{L}$ is the size of the time domain
  - $\boxed{\lambda(t)}$ is the intensity at time t
  - $\boxed{t_i}$ are time coordinates of the observed cases

Direct likelihood approximation, Simpson et al. (2016)

$$l(\lambda(.)|\mathbf{y}) \approx L - \sum_{j=1}^{m} w_j \log\lambda(t_j) + \sum_{i=1}^{n}\log(\lambda(t_i))$$

- m integration points $t_1, ..., t_m$, weights $w_1, ..., w_m$
  - if the grid is equaly spaced, $w_j = w_0$

The log-Gaussian Cox point process model

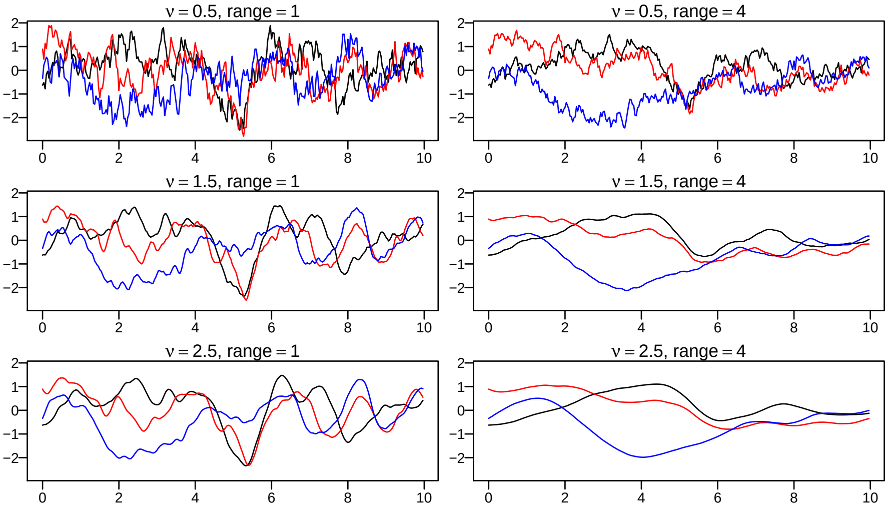
- Assumption (lGCpp): log of λ follows a Gaussian process
  - $\log(\lambda(.))$ is non-observable/latent

The log-Gaussian Cox point process model

- Assumption (lGCpp): log of λ follows a Gaussian process
  - $\log(\lambda(.))$ is non-observable/latent

Figure 1: A latent Gaussian model !!!

Matern 1D field samples (smoothness ν and range)



Our model framework: Latent Gaussian models (LGM)

- Bayesian hierarchical model

$\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}$	$\prod_i \pi(y_i \mid \eta_i, \boldsymbol{\theta})$	Several distributions!
$\boldsymbol{x} \mid \boldsymbol{\theta}$	$\pi(\boldsymbol{x} \mid \boldsymbol{\theta}) : \mathcal{N}(\mathbf{0}, \mathbf{Q}(\boldsymbol{\theta})^{-1})$	Gaussian (GMRF)!
$\boldsymbol{\theta}$	$\pi(\boldsymbol{\theta})$	Any distribution!

Our model framework: Latent Gaussian models (LGM)

- Bayesian hierarchical model

$\boldsymbol{y}\vert\boldsymbol{x},\boldsymbol{\theta}$	$\prod_i \pi(y_i\vert\eta_i,\boldsymbol{\theta})$	Several distributions!
$\boldsymbol{x}\vert\boldsymbol{\theta}$	$\pi(\boldsymbol{x}\vert\boldsymbol{\theta}) : \mathcal{N}(\mathbf{0},\mathbf{Q}(\boldsymbol{\theta})^{-1})$	Gaussian (GMRF)!
$\boldsymbol{\theta}$	$\pi(\boldsymbol{\theta})$	Any distribution!

- Precision matrix $\mathbf{Q}(\boldsymbol{\theta})$ is sparse
  - Gaussian distributions with sparse precision are Gaussian Markov random fields (GMRFs)
    - allows fast computations

Our model framework: Latent Gaussian models (LGM)

- Bayesian hierarchical model

$\boldsymbol{y}\|\boldsymbol{x},\boldsymbol{\theta}$	$\prod_i \pi(y_i\|\eta_i,\boldsymbol{\theta})$	Several distributions!
$\boldsymbol{x}\|\boldsymbol{\theta}$	$\pi(\boldsymbol{x}\|\boldsymbol{\theta}) : \mathcal{N}(\mathbf{0},\mathbf{Q}(\boldsymbol{\theta})^{-1})$	Gaussian (GMRF)!
$\boldsymbol{\theta}$	$\pi(\boldsymbol{\theta})$	Any distribution!

- Precision matrix $\mathbf{Q}(\boldsymbol{\theta})$ is sparse
  - Gaussian distributions with sparse precision are Gaussian Markov random fields (GMRFs)
    - allows fast computations

- **INLA**: Integrated Nested Laplace Approximations, Rue, Martino, and Chopin (2009) and Rue et al. (2017)

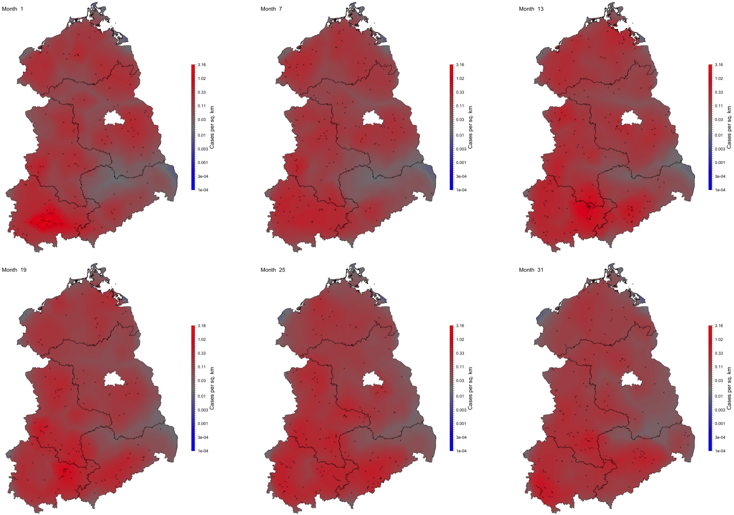
Sumário

The idea into practice, using inlabru

```
# data
dat1 <- data.frame(year = burkitt$t/365.25)
# domain (0, 16) and integration points
mesh.t <- inla.mesh.1d(seq(-3, 20, 1/12))
# assumed GMRF model (Matern) definition
model <- year ~ 0 + Intercept +
  spde1D(map = year,
         model = inla.spde2.pcmatern(
           mesh = mesh.t,
           prior.sigma = c(1, 0.5), # P(sigma>1) = 0.5
           prior.range = c(0.1, 0.01)))# P(range<0.1) = 0.01
# fit the model
fit.model1 <- lgcp(model, dat1)
```

Sumário

Space-time example: fox rabie cases



What do we need?

- space-time integration points
  - temporal ones and spatial ones

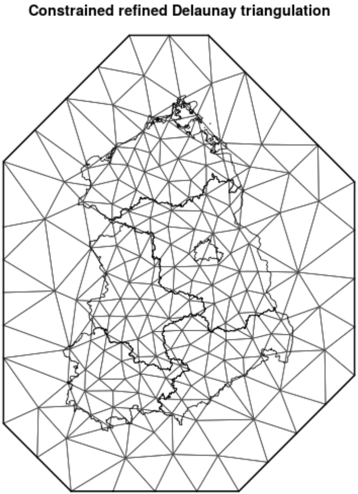
What do we need?

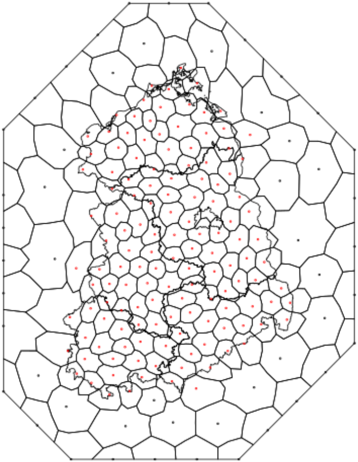
- space-time integration points
  - temporal ones and spatial ones
- space-time Gaussian process
  - separable: time Kronecker space
  - non-separable: **spacetime** driven process

What do we need?

- space-time integration points
  - temporal ones and spatial ones
- space-time Gaussian process
  - separable: time Kronecker space
  - non-separable: **spacetime** driven process
- code!

Adding spatial integration points





Space-time non-separable process

- Consider the stochastic iterated heat equation, Krainski (2018)

$$\left(\gamma_t \frac{d}{dt} + L^{\alpha_s/2}\right)^{\alpha_t} u(t) = \dot{W}_{\gamma_e^2 L^{\alpha_e}}(t).$$

- differential (system of) equations with respect to time
- L: Spatial Matern
- smoothness parameters $(\alpha_t, \alpha_s, \alpha_e)$
  - flexible model, consider some cases
- scale parameters $(\gamma_t, \gamma_s, \gamma_e) > 0$

Space-time non-separable process

- Consider the stochastic iterated heat equation, Krainski (2018)

$$\left(\gamma_t \frac{d}{dt} + L^{\alpha_s/2}\right)^{\alpha_t} u(t) = \dot{W}_{\gamma_e^2 L^{\alpha_e}}(t).$$

- differential (system of) equations with respect to time
- L: Spatial Matern
- smoothness parameters $(\alpha_t, \alpha_s, \alpha_e)$
  - flexible model, consider some cases
- scale parameters $(\gamma_t, \gamma_s, \gamma_e) > 0$

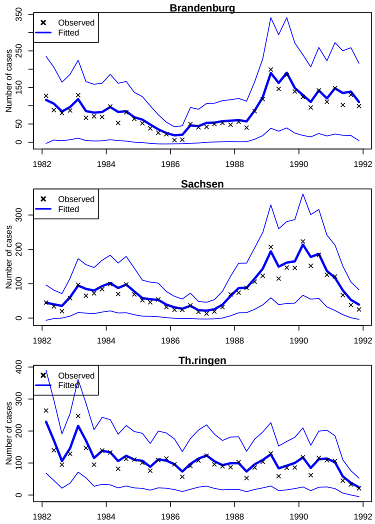
Advantage of using SPDEs

- sparse precision matrix representation!

Code

- consider the INLA package, Rue et al. (2017)
  - define the precision matrix using the `rgeneric` model
- consider the efficient PARDISO library, Kourounis, Fuchs, and Schenk (2018)

Preliminar result: Observed and expected cases, big areas per time



Work in progress

- wrap it into a function for the user
  - dozens lines of code write now
- parameter interpretation $\rightarrow$ define new ones such as
  - spatial range, temporal range, marginal variance

Sumário

References

Kourounis, D., A. Fuchs, and O. Schenk. 2018. "Towards the Next Generation of Multiperiod Optimal Power Flow Solvers." *IEEE Transactions on Power Systems* PP (99): 1–10. doi:10.1109/TPWRS.2017.2789187.

Krainski, Elias Teixeira. 2018. "Statistical Analysis of Space-Time Data: New Models and Applications." PhD thesis, Norwegian University of Science; Technology.

Rowlingson, B. S., and P. J. Diggle. 1993. "Splancs: Spatial Point Pattern Analysis Code in S-Plus." *Computers & Geosciences* 19 (5): 627–55. doi:https://doi.org/10.1016/0098-3004(93)90099-Q.

Rue, H., S. Martino, and N. Chopin. 2009. "Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations (with Discussion)." *Journal of the Royal Statistical Society, Series B* 71 (2): 319–92.

Rue, H., A. I. Riebler, S. H. Sørbye, J. B. Illian, D. P. Simpson, and F. K. Lindgren. 2017. "Bayesian Computing with INLA: A Review." *Annual Review of Statistics and Its Application* 4: 395–421.

Simpson, D. P., J. B. Illian, F. Lindren, S. H Sørbye, and H. Rue. 2016. "Going Off Grid: Computationally Efficient Inference for Log-Gaussian Cox Processes." *Biometrika* 103 (1): 49–70.