Complex point processes without grids or pain

Janine Illian, Daniel Simpson, Finn Lindgren, Håvard Rue, Sigrunn Sørbye

University of St Andrews
NTNU
University of Tromsø

February 5, 2011
Outline

Introduction—Koalas, Trees and Strange Boundaries

Working with a Continuous Random Field

Marks!

Conclusion
the reality of point process modelling...

- spatial point processes model the spatial organisation of individuals
- development of methodology has mainly been **theory-driven**
- only "point process experts" ever fit models – mainly as examples
  ⇒ models are rarely fitted to answer concrete questions
- little work on the practicality of fitting, comparing and assessing **complex and realistic models**
Introduction—Koalas, Trees and Strange Boundaries

Working with a Continuous Random Field Marks!

Conclusion

conservation study

- study conducted at the Koala Conservation Centre on Phillip Island, near Melbourne, Australia, 1993 - 2004
- \( \approx 20 \) koalas present in the reserve at all times throughout study; reserve enclosed by a koala-proof fence
- koalas feed on eucalyptus leaves which are toxic to most animals; koalas have adapted to this

Do the koalas feed selectively, i.e. do they choose trees with the least toxic/ most nutritious leaves?
spatial point pattern data

- all 915 trees in woodland individually numbered and mapped

spatial autocorrelation:

- trees are likely to aggregate in areas where soil nutrient levels are good
- no data on soil properties available
modelling the tree locations

The tree locations are modelled using a log Gaussian Cox process, i.e. a Cox process with random intensity

\[ \Lambda(s) = \exp\{\mu_1 + \beta_1 Z(s) + U(s)\}, \]

where \( \{Z(.)\} \) is a (possibly stationary and isotropic) Gaussian random field, \( s \in \mathbb{R}^2 \) and \( \{U(.)\} \) is an error field.

- conditional on \( \Lambda(.) \), Poisson process (spatial independence)
model fitting

- log-Gaussian Cox (LGC) processes are **latent Gaussian models**
- observations are independent given a latent field
  ⇒ we can use INLA to fit these models
- enables the fitting of realistically complex point process models; more on this complexity later...
The plot...

How should we grid this...?
Outline

Introduction—Koalas, Trees and Strange Boundaries

Working with a Continuous Random Field

Marks!

Conclusion
The likelihood (aka why point processes are hard to infer)

The likelihood *in the most boring case* is

$$\log(\pi(Y|\eta)) = |\Omega| - \int_{\Omega} \Lambda(s) \, ds + \sum_{s_i \in Y} \Lambda(s_i),$$

where $Y$ is the set of observed locations.

We cannot compute $\int_{\Omega} \exp(\eta(s)) \, ds$. 
What is usually done

- Take the region and construct a fine lattice.
- Bin the observations into the grid cells
- The number of points in cell \( \{ s_{ij} \} \) is conditionally Poisson, i.e.

\[
  y_{ij} | \eta_{ij} \sim Po(|s_{ij}| \exp(\eta_{ij})),
\]

where \( \eta_{ij} ' = Z(s_{ij}) \).
Is this approach satisfactory?

- You must use a very dense lattice.
- How well you treat your data is tied to how fine your lattice is.
- Binning *feels weird.*
A better solution: SPDEs

When you have a hammer, everything looks like a nail?

- The spatial Markov process (in this case) makes sense physically.
- You get the field everywhere \textit{irrespective of your spatial resolution}.
- You can sensibly incorporate boundaries
- You don’t need a covariance function!
- You can locally control the resolution of your spatial component. Coarse result:

\[
\sup_{f \in H^1(\Omega), \|f\|_{H^1} \leq 1} \mathbb{E} \langle f, x - x_h \rangle_{H^1}^2 \leq C h^2.
\]
That likelihood again

The random field is given by

\[ Z(s) = \sum_{i \in \text{Vertices}} x_i \phi_i(s). \]

- We can use the specific form of \( Z(s) \) to approximate the likelihood.
- If we have covariates, it’s probably worthwhile putting them in this form (Least squares + penalty term).
Approximating the integral

\[
\int_{\Omega} \exp(\eta(s)) \, ds = \sum_i \int_{T_i} \exp \left( \sum_j x_j \phi_j(s) \right) \, ds \\
\approx \sum_i \tilde{w}_i \exp(x_i).
\]

where the weights are \( \tilde{w}_i = \int_{\Omega} \phi_i(s) \, ds \).
The approximate likelihood

The approximate likelihood is

$$\log(\pi(\mathbf{Y} | \eta)) \approx |\Omega| - \sum_{i \in \text{Vertices}} \tilde{w}_i \exp(x_i) + \sum_{s_i \in \mathbf{Y}} \exp(\eta(s_i))$$

In order to evaluate this, we also need the "evaluation map" $A$, which is the matrix that maps the vertices to the observed points. This is easy.
INLA call

```r
formula = yy ~ 1 + f(idx,model = "spde",
    spde.prefix = fmesh$prefix ,n=spde$n,
    param =c(T0_guess,0.1,K0_guess,0.1,NA,NA,NA,NA ))

r=inla(formula,family=c("poisson","poisson"),
    data = data,E = c(0.0*weights,weights))
```

Full inference take 17 seconds on my laptop.
Posterior log intensity
Outline

Introduction—Koalas, Trees and Strange Boundaries

Working with a Continuous Random Field

Marks!

Conclusion
complexity – marked point pattern

foliage collection and analysis
  ▶ leaf samples taken from each eucalyptus tree and analysed for palatability

palatability: combination of toxins and nutrients based on previous studies

spatial autocorrelation: palatability likely to not be independent of spatial pattern:
  ▶ in areas with high soil nutrient levels, nutrients in leaves high
marked point pattern – complexity

**koala tree visitation**

- tree use by individual koalas collected at monthly intervals between 1993 and March 2004
- entire reserve searched for koalas
- identities of all koalas found and of the trees occupied were recorded

**spatial autocorrelation**: koala visits likely to not be independent of **spatial pattern** and **palatability**:

- koalas move very little and are more likely to favour areas with higher tree density
- koalas are likely to favour trees with high palatability
in summary this suggest:

- **tree locations** depend on (unobserved) soil nutrients levels and local clustering
- **palatability** depends on spatial pattern (through soil nutrients levels)
- **koala visitation** depends on spatial pattern, palatability
- spatial point pattern data, to be modelled with a (marked) **spatial point process**

two types of marks:

1. palatability of leaves ("leaf marks")
2. koala use of trees (depends on palatability) ("koala marks")
modelling the leaf marks

The leaf marks $m_L$ are modelled using an intensity marked log Gaussian Cox process, where the marks are modelled as:

$$m_L(x_i|\mathcal{K}(x_i)) = \mu_2 + \beta_2 Z(x_i) + V(x_i),$$

where $x_i \in Y$, $Y$ unmarked point process, $\{Z(.)\}$ is as above and $\{V(.)\}$ is an i. i.d. normal field.

- conditional on $\mathcal{K}(.)$, marks independent
modelling the koala marks

The koala marks $m_K$ are modelled using a hierarchically and intensity marked Cox process, where the marks are modelled as:

$$m_K(x_i) | \mathcal{N}(x_i) \sim \text{Poisson}(\mu_3 + \beta_3 Z(x_i) + \beta_4 \cdot m_L(x_i) + W(x_i)),$$

where $x_i$, $\{Z(.)\}$ as above and $\{W(.)\}$ another Gaussian random field (with zero mean).

- conditional on $\mathcal{N}(.)$, marks independent
The posterior mean
Outline

Introduction—Koalas, Trees and Strange Boundaries

Working with a Continuous Random Field

Marks!

Conclusion
Spatiotemporal?

- Lots of interesting applications.
- Inference through a simple extension of the SPDEs.
- Avoid ‘grids’ in time.
- Space Markov + Time Markov = Possible
Closing Thoughts

- This just works.