Ville Väänänen / 63527M

ville.vaananen@aalto.fi
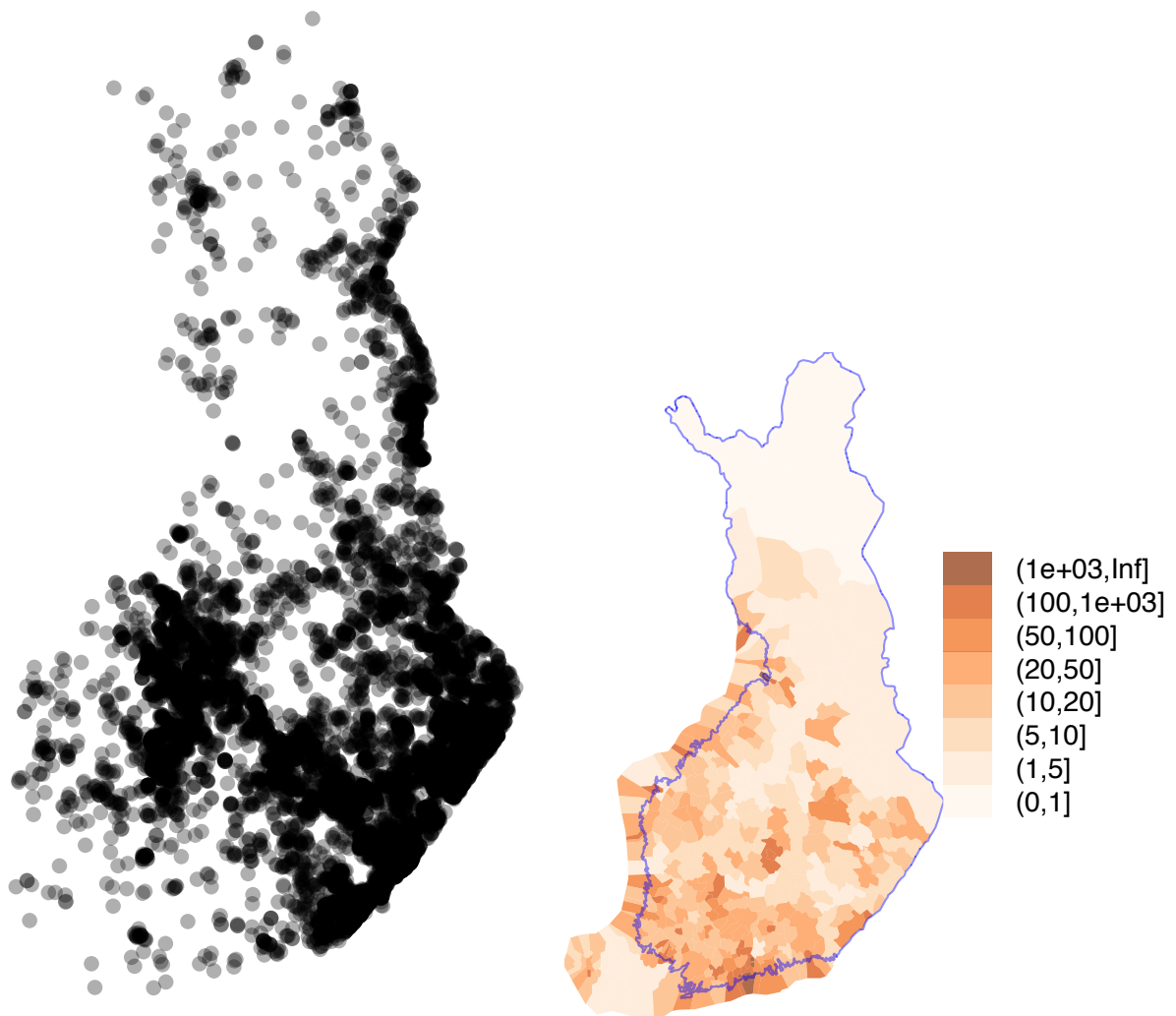
Bear observations in Finland in 2010

S-114.4202 Special Course in Computational Engineering II

March 26, 2012

1   Data description

The bear observations in the year 2010 are plotted in figure 1a and the population density of the municipalities (the number of people in the municipality per its land area) is plotted in figure 1b. As can be seen, there is some clustering which has to be associated with increased probability of reporting due population density. Thus an interesting problem is, how should we take into account the probability for an animal to get observed in a given area? Is it directly proportional to population density? Is there some other factors associated such as the amount of hunters in the area?

(a) Bear observations in 2010   (b) Population density

Figure 1: Bear observations in 2010 and the population density of Finland

2 Methods

Cox processes

Cox processes are a broad class of point processes that are nonhomogoneous Poisson processes given a *latent random intensity field* $\lambda(\mathbf{s})$. Throughout the report, we will denote vectors and vector valued functions with bold lowercase symbols and matrices with bold upper case symbols. We will denote a *spatial point process* defined in some bounded region $\Omega \subset \mathbb{R}^2$, also called the *observation window*, with $\mathcal{Y}(\mathbf{s})$ where $\mathbf{s} \in \Omega$. A realisation of a point process is called a *point pattern*

$$\mathbf{y} = \left[\mathcal{Y}(\mathbf{s}_1), \ldots, \mathcal{Y}(\mathbf{s}_n) \right]^T, \tag{1}$$

that can be thought of as an n dimensional random vector governed by a well defined multivariate distribution [1]. It is important to realise that the elements of $\mathbf{y}$ belong to some abstract sample space, i.e. bears or wolves in case of this report [1]. Also, (1) implicitly defines a matrix $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_n]$ of point locations.

Going back to Cox processes, given the intensity field, the number of points $N(D)$ inside a region $D \subset \Omega$ is a Poisson distributed random variable with mean $\Lambda(D) = \int_D \lambda(\mathbf{s}) \ \mathrm{d}\mathbf{s}$, i.e.

$$p\big(N(D) = k \mid \Lambda(D)\big) = \frac{\Lambda(D)^k}{k!} e^{-\Lambda(D)^k}. \tag{2}$$

In a log-Gaussian Cox process, the latent intensity field has the form

$$\lambda(\mathbf{s}) = \exp\big(Z(\mathbf{s})\big) \tag{3}$$

where Z is a *Gaussian random field* (also called a Gaussian process, GP) defined by a mean function $\mu(\mathbf{s})$ and a covariance function $Q(\mathbf{s}, \mathbf{s}')$ [2]. In this report a linear model will be used for the mean function, i.e

$$\big\langle \log \lambda(\mathbf{s}) \big\rangle = \boldsymbol{\beta}^T \mathbf{z}, \tag{4}$$

where typically $z_1(\mathbf{s}) = 1$, so that β_1 is the intercept. With this definition, possible covariate data can be included directly in $\mathbf{z}$.

Latent Gaussian models

Our aim is to be able to use a computationally efficient Bayesian procedure for inference in latent Gaussian models called INLA (or integrated nested Laplace approximations) [3]. Latent Gaussian models are a large set of flexible models, where given a realisation $\mathbf{x}$ of a latent (unobserved) Gaussian random field the observations $\mathbf{y}$ are independent and identically distributed. More precisely we

have

$$Z(\mathbf{s})|\boldsymbol{\theta} \sim \mathcal{GP}(\mu(\mathbf{s}, \boldsymbol{\theta}), Q(\mathbf{s}, \mathbf{s}', \boldsymbol{\theta})) \tag{5}$$

$$\mathbf{y}|Z(\mathbf{s}), \boldsymbol{\theta} \sim \prod_{k=1}^{n} p\big(y_k \mid Z(\mathbf{s}), \boldsymbol{\theta}\big) \tag{6}$$

Here we have introduced the hyperparameter $\boldsymbol{\theta}$. It is worth noting that in this vector notation $\mathbf{x}$ has to be thought of as belonging to an infinite dimensional Hilbert space.

Likelihood approximation

LGCP's can be thought of as latent Gaussian models and in their case the logarithm of the conditional likelihood function (6) can be written as [2, 1, 4] (going further, dependence on the hyperparameter will be suppressed but is assumed)

$$\log p(\mathbf{y} \mid Z) = |\Omega| - \int_{\Omega} e^{Z(\mathbf{s})} \; \mathbf{ds} + \sum_{k=1}^{n} Z(\mathbf{s}_k) \tag{7}$$

Unfortunately (7) cannot be evaluated without first forming some kind of approximation to the latent random field $Z(\mathbf{s}) = \log \lambda(\mathbf{s})$. With the help of such an approximation, the last term requires just the evaluation of the approximated field in the observed locations. Since the integral of the intensity function over the observation window (i.e. the mean number of points in the pattern) will remain analytically intractable it will be computed numerically [4].

A common approach for approximating the conditional likelihood is to partition the observation window to a regular grid and assume the field constant inside a grid cell. The value chosen for a cell can c_k be for example the value of the field in the cell's center point $\mathbf{c}_k$. Clearly in this case the number of points in a grid cell is Poisson distributed with mean $|c_k| \, e^{Z(\mathbf{c}_k)}$. As argued in [4] it is far from optimal to use the grid both for approximating the intensity field *and* the locations of the points y_k. A nicer alternative would be to form a finite dimensional *continous* approximation to the field, so that it could be evaluated at the exact point locations.

This can be achieved by forming a piecewise linear basis function approximation to $Z(\mathbf{s})$

$$\hat{Z}(\mathbf{s}) = \sum_{j=1}^{m} z_j \phi_j(\mathbf{s}) \tag{8}$$

$$= \mathbf{z}^T \phi(\mathbf{s}) \tag{9}$$

The details of this approximation are laid out in [5] and only the results will be stated here. First of all, computational efficiency leads us to only consider *Markovian* Gaussian random fields, and it turns out that a subset of the Gaussian random fields with the *Matérn* covariance functions has the spatial Markov

property. The Matérn covariance functions are functions of only the distance $r = \|\mathbf{s} - \mathbf{s}'\|$ between two points

$$c_M(r) = \frac{\sigma^2}{2^{\nu-1}\Gamma\nu}(\kappa\, r)^\nu K_\nu(\kappa\, r), \tag{10}$$

where $K_\nu(\cdot)$ is the modified Bessel function of the second kind, $\nu > 0$ is the smoothing parameter, $\kappa > 0$ is the range parameter and σ^2 is the variance. The Matérn fields have the Markov property when $\alpha = \nu + \dfrac{d}{2}$ is an integer. In our case we have $d = 2$ and we will choose $\nu = 3$, so that $\alpha = 2$. The piecewise linear basis functions will be defined on a triangulated mesh of the observation window, so that the value of $\phi_i(\mathbf{s})$ is 1 on vertex i and goes linearly to zero on the neighboring vertices.

The integral in (7) is approximated with a sum using a simple deterministic numerical integration scheme with weights w_i and sigma points $\hat{\mathbf{s}}_i$, so that the approximation to the integral of some function f over the domain Ω is given by

$$\int_\Omega f(\mathbf{s})\ \mathrm{d}\mathbf{s} \approx \sum_{i=1}^p w_i\, f(\hat{\mathbf{s}}_i) \tag{11}$$

Examples of such numerical integration rules are the well known midpoint rule, trapezoid rule and and the Simpson's rule (f is interpolated with a polynomial of order 0, 1 or 2 respectfully).

Putting all this together, the approximation to the conditional log-likelihood is given as

$$\log p(\mathbf{y} \mid Z) \approx C - \sum_{i=1}^p w_i \exp\left(\mathbf{z}^T \boldsymbol{\phi}(\hat{\mathbf{s}}_i)\right) + \sum_{k=1}^n \mathbf{z}^T \boldsymbol{\phi}(\mathbf{s}_k) \tag{12}$$

$$= C - \mathbf{w}^T \exp\left(\mathbf{A}_1 \mathbf{z}\right) + \mathbf{1}^T \mathbf{A}_2 \mathbf{z}, \tag{13}$$

where $[\mathbf{A}_1]_{ij} = \phi_j(\hat{\mathbf{s}}_i)$, $[\mathbf{A}_2]_{kj} = \phi_j(\mathbf{s}_k)$ and C is an unimportant constant.

3 Results

4 Conclusion

References

[1] A Gelfand, PJ Diggle, and P Guttorp. *Handbook of Spatial Statistics*. CRC Press, 2010. ISBN: 9781420072877. URL: http://eprints.ncrm.ac.uk/2066/.

[2] Jesper Møller and Rasmus P. Waagepetersen. "Modern Statistics for Spatial Point Processes". In: *Scandinavian Journal of Statistics* 34.June 2006 (Sept. 2007), 070927154002001–??? ISSN: 0303-6898. DOI: 10.1111/j.1467-9469.2007.00569.x. URL: http://doi.wiley.com/10.1111/j.1467-9469.2007.00569.x.

[3]   Håvard Rue, Sara Martino, and Nicolas Chopin. "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71.2 (Apr. 2009), pp. 319–392. ISSN: 13697412. DOI: 10.1111/j.1467-9868.2008.00700.x. URL: http://doi.wiley.com/10.1111/j.1467-9868.2008.00700.x.

[4]   Daniel Simpson, Janine Illian, and Finn Lindgren. "Going off grid: Computationally efficient inference for log-Gaussian Cox processes". In: *Science And Technology* (2011). URL: http://adsabs.harvard.edu/abs/2011arXiv1111.0641S.

[5]   Daniel Simpson and Finn Lindgren. "Think continuous: Markovian Gaussian models in spatial statistics". In: *Arxiv preprint arXiv:1110.6796* (2011). URL: http://arxiv.org/abs/1110.6796.