6 Exercises, week 6

The exercise solutions have to be returned at the latest on Sunday April 30'th.

- Pen and paper exercises: you can scan the solutions and combile them into pdf or write them with Latex/word/... Compile all the answers into one pdf file
- Computer exercises: Report the answers to no-coding parts of exercises (if any) in pdf and compile with answers to pen and paper exercises. Additionally, send also the code used to solve the exercises. Note!
  - Only code should be returned. **Do not send data files**!
  - Write and comment the code so that it can be run by using your code only and the data provided in the course web pages.
  - If the lecturer cannot understand or run your code you will not get points from coding part even if the results were correct.
- zip all files into one folder to reduce the size of submission.

Send the zipped files to jarno.vanhatalo@helsinki.fi.

For basic properties and results concerning Gaussian distributions and processes see e.g.
`https://en.wikipedia.org/wiki/Multivariate_normal_distribution`
`http://www.gaussianprocess.org/gpml/chapters/`

6.1 Repeated vs. one time sampling, 1 point

Consider species distribution modeling. Let $\tilde{\pi}(\mathbf{s})$ denote the probability that a species is present at location $\mathbf{s}$ and $\tilde{\pi}_0$ denote the probability to observe the species given it is present. Consider we have presence absence observations and two types of sampling schemes: a) one sampling event per location, b) many sampling events per location. We assume further that the species presence/absence status does not change between the sampling events. Derive the observation model for both sampling schemes and show that only with b) we can infer both the probability of species presence and the observation probability.

6.2 Sampling from Poisson processes, 2 points

a) Consider generating a realization from a homogenous Poisson process in domain D (that is $\lambda(\mathbf{s}) = \lambda$). Proof that this can be done by first sampling $n' \sim \text{Poisson}(\lambda|D|)$ and then, given, n' sampling n' locations uniformly from D.

b) So called thinning method to obtain samples from nonhomogenous Poisson process is the following. Let $\lambda_{max} = \max_{\mathbf{s}} \lambda(\mathbf{s}), s \in D$. Generate a point pattern from homogenous Poisson process using the constant intensity λ_{max}. For each point generated, do a rejection step so that, for $\mathbf{s}_i$ draw $U_i \sim \mathrm{Unif}(0,1)$ and retain $\mathbf{s}_i$ if $U_i < \lambda \mathbf{s}_i / \lambda_{max}$. Show that with this procedure the remaining points follow a nonhomogeneous Poisson process with $\lambda(\mathbf{s})$.

Hint! Derive the distribution of number of points $N(B)$ for any $B \subset D$ and show that these satisfy the properties of the Poisson process.

6.3 Log Gaussian cox process, 1 point

Load the coal mine disaster data coal.txt from the course web page. The data contains the dates of coal mine explosions that killed ten or more men in Britain between 15 March 1851 and 22nd March 1962. Model the data with log Gaussian Cox process. The covariance function of the Gaussian process should be squared exponential with length-scale 10 years and magnitude 1. Plot the posterior median and 95% credible interval of the intensity function between years 1851 and 1963.

6.4 Partially observed realizations from Poisson process, 1 point

Consider that we want to infer the posterior of the intensity function of a log Gaussian Cox process over domain D. However, we have observed the points only in finite number of disjoint regions $B_i, i = 1, \ldots, n$. Moreover, the probability of observing the points in these regions might not be one and it might vary between the regions so that the observation probability within a region can be considered to be constant. Define a model for the data.