Lab 4: Binomial geostatistical model

April 19, 2021

Summary

In the lab session, you will learn the following:

- how to analyse prevalence data.
- develop a binomial geostatistical model
- perform spatial prediction of the prevalance at observed and the unobserved locations.

Analysis of river blindness data in Liberia (Binomial Modelling)

1. Load the river blindness data, `LiberiaRemoData.csv`. in the MBG app. To open the MBG app type the following in the R console: `shiny::runGitHub(repo="MBGapp", username= "olatunjijohnson", ref="main", subdir = "inst/MBGapp")`.

2. Check for possible correlations between prevalence (or a reasonable transformation of prevalence) and `elevation`. Is there a reasonable linear relationships between the outcome and `elevation`? What transformations of `elevation` might make the relationship reasonably linear?

3. Fit the following binomial geostatistical model to the data:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 \log d(x_i) + \beta_2 x_{i,1} + S(x_i) + Z_i, \tag{1}$$

   where $d(x_i)$ is the elevation at location x_i; $x_{i,1}$ is the east-west ordinate (`long`) of the location x_i; $S(x_i)$ is a stationary Gaussian process with exponential correlation function and Z_i are i.i.d. Gaussian variables.

4. Interpret the model parameters.

5. Perform prediction at spatial resolution of 5km.

6. Visualise the following:

   - Predicted mean prevalence. What are the areas with the highest prevalence?
   - Standard errors of the predictions.
   - Visualise the probability that the prevalence exceeds the threshold of 20%. What happens to the exceedance probabilities if you change this threshold upward?

Analysis of river blindness data in Liberia (Linear Modelling)

1. Consider an outcome variable $\widetilde{Y}_i$, the empirical logit transformation of prevalence, given by
$$\widetilde{Y}_i = \log\left(\frac{Y_i + 0.5}{n_i - Y_i + 0.5}\right), \tag{2}$$
where Y_i is the number of positive cases (`npos`) out of a total of n_i individuals tested (`ntest`) at location x_i.

2. Fit linear geostatistical models to the empirical logit $\widetilde{Y}$ as follows.
$$\widetilde{Y}_i = \beta_0 + \beta_1 \log(d_i) + \beta_2 x_{i,1} + S(x_i) + Z_i. \tag{3}$$
where d_i is the elevation at location x_i; $x_{i,1}$ is the east-west ordinate (`long`) of the location x_i; $S(x_i)$ is a stationary Gaussian process with exponential correlation function and Z_i are i.i.d. Gaussian variables.

3. Use the model above for spatial predictions, using the same grid as in the previous section.

4. Compare the linear model to the corresponding Binomial model regarding
   (a) the parameter estimates
   (b) mean prevalence
   (c) standard errors of prevalence
   (d) the probability that prevalence exceeds the threshold of 20%

Exercises in R

The `Lab4.R` script provides the code to perform the analysis of the prevalence data from river blindness. Using this code and the code provided in the previous labs perform steps 1-4 for the analysis of the empirical logit of prevalence.