

Statistical Modelling

PROJECT 4

Rules

- A report of maximum 8 pages (including figures, comments, references, etc.) is required together with a script of the R-code.
- The report must be upload on Blackboard two days before your oral exam (until March 2021).
- It must be understood that this is an individual project and that your answers are solely the result of your personal work.

Introduction

The World Development Indicators from the World Bank contain over a thousand annual indicators of economic development from hundreds of countries around the world. The complete dataset is freely available at the website www.kaggle.com/worldbank/world-development-indicators.

A subset of indicators for 10 countries of Europe is contained in the file `data_project4.txt`. Such dataset includes the following variables:

- **CountryName**: name of the countries
- **CountryCode**: code identifying the country
- **days**: time required to start a business (days)
- **Nprocedures**: number of start-up procedures to register a business
- **newBusiness**: number of new businesses registered

We want to study how the time required to start a business in the European countries changes according to a set of available information.

Tasks

1. Make some exploratory analysis to provide a detailed description of the data.
2. Assume and write down a Poisson regression model for the time required to start a business as a function of the number of procedures to register a business and the number of new businesses.
3. Specify the hyperparameters of the prior distributions and comment your choices.
4. Write a BUGS code to implement a MCMC to approximate the posterior distribution of the parameters of interests.
5. Run the algorithm using OpenBUGS and the library R2OpenBUGS (or any other BUGS software) and check the convergence of the chains by means of suitable diagnostics tools. Choose the posterior sample size G , a suitable thinning and a burn in.
6. Based on the MCMC output provide Bayesian estimates of the regression coefficients. Also compute 95% credible interval for the above parameters. Comment the results; what can you say about the effect of the covariates on the time required to start a business?
7. Obtain a sample from the posterior predictive distribution of the time required to start a business given the average number of procedures and new businesses.