

# Bayesian Inference and Computation: Assignment 3

Deadline:

On the module's Canvas page you will find a comparative judgement dataset (csv file). Your assignment is to fit a model to data using a Bayesian approach. This will involve writing an MCMC algorithm.

## 1 Problem

Dar es Salaam is the largest city in Tanzania, where about 25% of the population lie in poverty (earning under \$21 a month). Identifying which areas in the city are in poverty can help. Dar es Salaam is made up of 452 subwards.

## 2 Data

The dataset consists of 75,078 comparisons, where judges were shown pairs of subwards and asked to choose which of the pair was more deprived. The first column displays the subward that was chosen to be the most deprived of the pair, the second column displays the subward that was chosen to be the least deprived.

## 3 Model

The subwards are labelled  $1, \dots, N$ , and subward  $i$  is assigned a value  $\lambda_i \in \mathbb{R}$ . Large positive values are associated with affluent subwards, large negative with negative subwards. When comparing areas  $i$  and  $j$ , the probability area  $i$  is judged to be more deprived than area  $j$  depends on the difference in their parameters

$$\text{logit}(\pi_{ij}) = \lambda_i - \lambda_j \iff \pi_{ij} = \frac{\exp(\lambda_i)}{\exp(\lambda_i) + \exp(\lambda_j)} \quad (i \neq j, 1 \leq i, j \leq N).$$

Under the assumption that each comparison is independent, if subwards  $i$  and  $j$  are compared  $N_{ij}$  times, then the number of times  $i$  is chosen to be more deprived than  $j$  is

$$Y_{ij} \sim \text{Bin}(N_{ij}, \pi_{ij}).$$

The likelihood function is the product of the density function for each pair of subwards:

$$\pi(\mathbf{y} \mid \boldsymbol{\lambda}) = \prod_{i=1}^N \prod_{j=1}^{i-1} \binom{N_{ij}}{y_{ij}} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{N_{ij} - y_{ij}},$$

where  $\boldsymbol{\lambda} = \{\lambda_1, \dots, \lambda_N\}$  is the set of subward parameters. Due to identifiability constraints,  $\lambda_1$  is unidentifiable and should be set to 0.

## 4 Task

Using a Normal prior distribution derive the posterior distribution  $\pi(\boldsymbol{\lambda} \mid \mathbf{y})$  and the full conditional distributions  $\pi(\lambda_i \mid \mathbf{y}, \lambda_1, \dots, \lambda_{i-1}, \lambda_{i+1}, \dots, \lambda_N)$ . Construct an MCMC algorithm to generate samples from the posterior distribution. Write an R script to run the MCMC algorithm and use it to generate samples from the posterior distribution.

You should submit two files:

1. A report of no more than two pages deriving the posterior distribution, displaying your MCMC algorithm and summarising the results.
2. A csv file containing the deprivation estimates for each subward.