Workshop 2
Note: Before starting this practical, you might want to spend (about) 15/20min. looking at the JAGS manual (available on Learn). If you already did it, then go directly to question 1.

1. Caries study: Describing caries experience in Flanders (adapted from Lesaffre and Lawson, 2012, p. 37)

The Signal-Tandmobiel study is a longitudinal oral health intervention study involving a sample of 4468 children. A random sample was taken by selecting primary schools at random and therein all children from the first class. The children were examined in 1996 by 16 trained dentists (examiners) and annually thereafter for 6 years. Here, we look at the caries experience on primary teeth of the first year of the study; hence, the data of 7-year old children are evaluated here. Caries experience on primary teeth is classically measured by the dmft-index. This score represents the number of primary teeth that are decayed (d), missing due to extraction for caries reasons (m) or filled (f) because of caries. It varies from 0 (no caries experience) to 20 (all primary teeth affected). We will analyse a subsample of the data formed by the dmft-index of 100 children (dataset dmft.Rdata is available on Learn). As a natural candidate for modelling the dmft-index we will use a Poisson distribution, i.e.,

$$f(y;\theta) = \frac{e^{-\theta}\theta^{y}}{y!}, \quad y = 0, 1, 2 \ldots \quad \theta > 0.$$

Additionally, the following prior information is available:

- The review paper of Vanobbergen et al. (2001) reported an average dmft-index of 4.1 obtained in a study based on 109 seven-year-old children and conducted in Lie`ge in 1983, while an average of 1.39 was obtained around Ghent on 200 five-year-old children examined in 1994.
- It is known that oral hygiene had improved considerably in Flanders in the recent years.

The authors stated, and leveraging conjugacy properties, that a Gamma(a, b) prior distribution for θ

with shape a = 3 and rate b = 1 seems to adequately represent the aforementioned knowledge.

(i) Conduct some exploratory data analysis. What do you think about the suitability of the Poisson model for this dataset?

(ii) Using exact calculations, determine the posterior mean, standard deviation and 95% credible interval for θ.

(iii) Repeat part (ii), but using rejection sampling. You might want to consider as proposal distribu- tion an exponential distribution with mean

equal to the mean of the data. Comment about the efficiency of the algorithm.

(iv) Repeat part (ii), but using sampling importance resampling.

(v) Repeat part (ii), but using JAGS. If now we change the prior to an approximate Jeffrey's prior, say a Gamma(1/2, 0.001), do the results change?

2. Analysis of binomial data: revisiting the drug example

The aim of this question is to re-do most part sof questions 1 and 2 from the first practical but now using JAGS. Remember the context: a new drug is being considered for relief of chronic pain, with the success rate θ being the proportion of patients experiencing pain relief. According to past information, a Beta(9.2, 13.8) prior distribution was suggested. This drug had 15 successes out of 20 patients.

(i) Compute the posterior mean, standard deviation and a 95% credible interval. Compare with the exact results.

(ii) What is the probability that the true success rate is greater than 0.6. Compare with the exact result.

(iii) Suppose 40 more patients were entered into the study. What is the chance that at least 25 of them experience pain relief? Compare with the exact result.

(iv) Conduct the 'prior/data compatibility check', i.e., calculate the predictive probability of observ- ing at least 15 successes under this prior. Compare with the exact result.

(v) In practical 1 we have then considered a mixture prior, where it was supposed that most drugs (95%) are assumed to come from the stated Beta(9.2, 13.8) prior, but there is a small chance that the drug might be a 'winner'. 'Winners' were assumed to have a Beta(12, 3) prior distribution. What is now the chance that the response rate is greater than 0.6? Compare with the exact result.

(vi) Under this mixture prior, what is the posterior predictive probability that at least 25 out of 40 new patients experience pain relief?

(vii) For this mixture prior, repeat the prior/data compatibility test performed previously. Are the data more compatible with this mixture prior? Compare with the exact result.