Bayesian analysis of space shuttle Challenger data

The space shuttle Challenger and its crew of seven people was destroyed soon after its launch on Jan 28, 1986. The disaster was due to leakage of gas from one of the fuel tanks. This was caused by damage to one of the six field-joint O-rings, which are insulating rings made of rubber. At the time of the disaster the temperature was 31 degrees Fahrenheit, which was unusually low (for Florida). There is data available on the number of field-joint O-rings showing signs of damage and the launch temperature on 23 previous flights.

We analyze the Challenger data using a logistic regression model, which assumes that conditionally on the parameters, the number of damaged O-rings in each of the flights is independently binomially distributed with probability which depends on the launch temperature. The formulation of this model is

$$\begin{aligned} [Y_i \mid \beta_0, \beta_1] &\overset{\text{ind}}{\sim} \mathrm{Bin}(6, \pi_i), \\ \mathrm{logit}(\pi_i) &= \beta_0 + \beta_1\,(x_i - m), \qquad i = 1, \dots, n, \end{aligned}$$

where $n = 23$ and y_i is the number of damaged O-rings and x_i is the launch temperature on the ith flight and $m = 69.6$ is the mean temperature of the previous flights.

Let Y^* be the number of damaged O-rings at the launch temperature $x^* = 31$. We assume that $Y = (Y_1, \dots, Y_n)$ and Y^* are conditionally independent given the parameters and that $p(y^* \mid \beta_0, \beta_1)$ is the binomial distribution with sample size 6 and "success" probability $\mathrm{logit}^{-1}(\beta_0 + \beta_1(x^* - m))$.

Since formulating the prior directly for the coefficients (β_0, β_1) is difficult, we instead formulate a prior for "success´´ (actually, failure) probabilities corresponding to the two temperatures 60 and 80 which are wide apart from each other, but still well within the observed range of the covariates. Our prior model for these probability parameters is that they are independent and uniformly distributed on the interval $(0, 1)$. In the case of the original model this means that the quantities α_1 and α_2 defined by

$$\alpha_1 = \mathrm{logit}^{-1}(\beta_0 + \beta_1\,(60 - m)), \qquad \alpha_2 = \mathrm{logit}^{-1}(\beta_0 + \beta_1\,(80 - m)).$$

are independent and uniform $\mathrm{Uni}(0, 1)$. Since (β_0, β_1) can be expressed uniquely in terms of (α_1, α_2), this implies a prior also for (β_0, β_1).

Carry out the following tasks.

- Solve (β_0, β_1) in terms of (α_1, α_2).

- Produce a scatter plot on (β_0, β_1)-plane, where you show points simulated from the prior. (Simulate values for (α_1, α_2) and solve for (β_0, β_1).)
- Produce a scatter plot from the posterior distribution of (β_0, β_1) using a MCMC sampler which you program yourself. For this you need to work out, how to evaluate the prior density and the likelihood. (E.g., use a RWM algorithm.)
- Plot the predictive distribution $y^* \mapsto p(y^* \mid y)$, where $y^* = 0, 1, \ldots, 6$. (Advice: in R you could implement this step by calculating the average of vectors `dbinom(0:6, size = 6, prob = pi.s[i])`, where `pi.s[i]` is the i'th sample for the "success" probability corresponding to temperature 31.)