

A bioassay experiment using the logistic link

Gelman *et al.* discuss the following model (in Section 3.7 of the book). To test the toxicity of a new drug, a standard approach is to perform bioassay experiments on animals. Various dose levels of a chemical compound are given to batches of animals and a binary outcome (survival or death) is recorded for each animal. We consider the following data.

| Dose (log g / ml) | Number of deaths | Number of animals |
|-------------------|------------------|-------------------|
| -0.86 | 0 | 5 |
| -0.30 | 1 | 5 |
| -0.05 | 3 | 5 |
| 0.73 | 5 | 5 |

Let y_i be the number of deaths out of n_i at a dose level x_i (expressed in the units log (g / ml)). We model the outcomes y_i (number of deaths) as the values of binomially distributed random variables Y_i , which are independent given θ , where

$$[Y_i | \theta] \stackrel{\text{ind}}{\sim} \text{Bin}(n_i, p_i), \quad i = 1, \dots, n$$

and the probability p_i is given by

$$\text{logit } p_i = \alpha + \beta x_i.$$

The parameter vector $\theta = (\alpha, \beta)$, and p_i is a function of θ . We take the flat prior, where α and β are independently normally distributed with large variances.

Since this is a standard model (a logistic regression model), you can get the MLE (maximum likelihood estimate) and its approximate covariance matrix (the inverse of the negative of the Hessian matrix of the log-likelihood) by using standard software. In R this goes as follows.

```
x <- c(-0.86, -0.3, -0.05, 0.73)
n <- c(5, 5, 5, 5)
y <- c(0, 1, 3, 5)
response <- cbind(y, n - y)
fit <- glm(response ~ x, family = binomial)
# This is the MLE:
coef(fit)
# This is the approximate covariance matrix:
vcov(fit)
```

You should get approximately the same kind of results for the posterior expectation and posterior covariance from the Bayesian model.

Your task is to explore the posterior distribution using a Metropolis–Hastings sampler you write yourself. Suitable proposal distributions could be one of the following.

- a) The independent M–H sampler which proposes from multivariate t distribution. Use a small value for the degrees of freedom parameter (e.g., four), and select the center of the t distribution to be the MLE and select the scale matrix of the t distribution to be proportional to the approximate covariance matrix of the MLE.
- b) The RWM algorithm where the distribution of the increment is a multivariate normal (or multivariate t) where the covariance matrix is proportional to the approximate covariance matrix of the MLE.

Your tasks are as follows.

- Write a formula for the likelihood.
- Make a contour plot of the posterior density.
- Select one MCMC algorithm. Tune your selected algorithm so that it seems to give a good representation of the posterior with a moderate sample size. Then fix the proposal distribution.
- Draw a simulated sample from the posterior density on top of the contour plot. Estimate posterior summaries such as the posterior expectation and the posterior covariance matrix.
- A particular parameter of interest is the LD-50, the value of the dose where the probability of death is 50 %. It is given by $\phi = -\alpha/\beta$. Plot a histogram of the posterior distribution of ϕ and calculate a central 95 % credible interval for it.