

UNIVERSITY OF EDINBURGH  
SCHOOL OF MATHEMATICS  
**Bayesian Data Analysis**

**Assignment 4**

---

**Salmonella data.**

Breslow (1984) analyses mutagenicity assay data on salmonella in which three plates have each been processed at various doses of quinoline (0,10,33,100,333,1000), and the number of colonies of TA98 salmonella subsequently measured.

Denote the dose by  $x_i$ ,  $i = 1, \dots, 6$ , and the number of colonies observed on plate  $j$  at dose  $x_i$  by  $y_{i,j}$ ,  $j = 1, 2, 3$ .

The theory suggests the following model for  $\mu_i = E y_{i,j}$ :

$$\log(\mu_i) = \alpha + \beta \log(x_i + 10) + \gamma x_i, \quad \text{with } \alpha, \beta, \gamma \in \mathbb{R}.$$

Consider the following likelihood:

$$y_{i,j} \mid \mu_i \sim \text{Pois}(\mu_i) \quad \text{independently (given) } \mu_i).$$

Use noninformative priors for  $\alpha, \beta, \gamma \in \mathbb{R}$  (e.g.  $N(0, 100^2)$ ).

The aim is to check the model fit for this data.

The strategy is to check the model fit using fitted values and residuals excluding the data for dose  $x_i = 100$  from the analysis, and then check the predictive model fit by comparing the predictive distribution for the number of colonies for the dose  $x_i = 100$  to the observed data.

*Model and data files for WinBUGS are available, as well as one file with initial values for each for the two models below. Please create a file with different initial values for the second chain.*

1. Run the model for doses 0, 10, 33, 333, 1000 (*the model and the data are given in files Assignment4model.odc and Assignment4data.odc*), check convergence and autocorrelation and produce posterior summaries for  $\alpha, \beta, \gamma$ , discarding the appropriate number of iterations and thinning as necessary (attach the plots justifying your choices). [2 marks]
2. Give the expressions for the mean and the variance of  $y_{i,j}$  given  $\mu_i$ , and hence give the expression for the Pearson (standardised) residuals. Edit the model specified in Q1 to add the Pearson (standardised) residuals (*deviance residuals for the Poisson data are already in the model file*). Please attach the modified code. [1 mark]
3. Produce a box plot for each of the deviance and Pearson residuals (*remember to monitor the residuals*). (*Use the box plot option from the Inference-Compare menu, with the corresponding residuals name as the node*).

Discuss the model fit, including whether the residuals are centred around 0, whether they are approximately between  $-2$  and  $2$ , whether there is a sudden jump between consecutive residuals, i.e. the evidence of outliers (*when they plotted ordered by their rank: right click on the plot, choose Properties/Special and tick "rank"*). [2 marks]

4. Produce the plot showing the model fit, with observations from the first plate. *(Remember to monitor  $\mu_i$ s. Use the model fit option from the Inference-Compare menu, with node mu, axis x, other y[,1] for the first plate).*

Repeat the process for the other two plates (editing "other" to y[,2] and y[,3], respectively).

Comment on the model fit for each plate, particularly on the uncertainty (i.e. whether majority of the observations lie within 95% credible intervals). [2 marks]

5. Obtain predictions of the number of colonies for dose  $x = 100$ . *You can use the model and the data files Assignment4modelPred.odc and Assignment4dataPred.odc. There, y.pred is the predicted number of colonies given dose xnew which is provided in the updated data file.*

Obtain the plot of the predictive distribution of y.pred (use option "density") or its posterior summaries, and discuss the evidence whether this predictive distribution predicts well the observed values of the number of colonies  $y_{new} = (27, 41, 60)$ . You can use the corresponding predictive p-values  $P(y.pred > y_{new}[j]|data)$ ,  $j = 1, 2, 3$  and their 95% credible intervals to support you the decision. *In the model file, these p-values are present as p[1], p[2] and p[3].*

*(Note that these p-values are one-sided, so the values very close either to 1 or to 0 indicate lack of fit).* [2 marks]

6. Make an overall conclusion about the evidence of the model fit to the data. [1 mark]
7. \* Discuss the evidence for the hypothesis  $\gamma = 0$ , taking into the account the MC error. *[The question is not compulsory but 0.5 mark will be added, up to the maximum of 10, for the correct answer.]*