AMS 274 – Generalized Linear Models (Fall 2016)

Homework 5 (due by 12pm, Tuesday December 6)

Consider the data set from homework 2, problem 3 on the incidence of faults in the manufacturing of rolls of fabric:

`http://www.stat.columbia.edu/~gelman/book/data/fabric.asc`

where the first column contains the length of each roll, which is the covariate with values x_i, and the second column contains the number of faults, which is the response with values y_i and means μ_i.

(a) Fit a Bayesian Poisson GLM to these data, using a logarithmic link, $\log(\mu_i) = \beta_1 + \beta_2 x_i$. Obtain the posterior distributions for β_1 and β_2 (under a flat prior for (β_1, β_2)), as well as point and interval estimates for the response mean as a function of the covariate (over a grid of covariate values). Obtain the distributions of the posterior predictive residuals, and use them for model checking.

(b) Develop a hierarchical extension of the Poisson GLM from part (a), using a gamma distribution for the response means across roll lengths. Specifically, for the second stage of the hierarchical model, assume that

$$\mu_i \mid \gamma_i, \lambda \overset{ind.}{\sim} \frac{1}{\Gamma(\lambda)} \left(\frac{\lambda}{\gamma_i}\right)^{\lambda} \mu_i^{\lambda-1} \exp\left(-\frac{\lambda}{\gamma_i}\mu_i\right) \qquad \mu_i > 0;\;\; \lambda > 0, \gamma_i > 0,$$

where $\log(\gamma_i) = \beta_1 + \beta_2 x_i$. (Here, $\Gamma(u) = \int_0^{\infty} t^{u-1}\exp(-t)\mathrm{d}t$ is the Gamma function.)

Derive the expressions for $\mathrm{E}(Y_i \mid \beta_1, \beta_2, \lambda)$ and $\mathrm{Var}(Y_i \mid \beta_1, \beta_2, \lambda)$, and compare them with the corresponding expressions under the non-hierarchical model from part (a). Develop an MCMC method for posterior simulation providing details for all its steps. Derive the expression for the posterior predictive distribution of a new (unobserved) response y_0 corresponding to a specified covariate value x_0, which is not included in the observed x_i. Implement the MCMC algorithm to obtain the posterior distributions for β_1, β_2 and λ, as well as point and interval estimates for the response mean as a function of the covariate (over a grid of covariate values). Discuss model checking results based on posterior predictive residuals.

Regarding the priors, you can use again the flat prior for (β_1, β_2), but perform prior sensitivity analysis for λ considering different proper priors, including $p(\lambda) = (\lambda + 1)^{-2}$.

(c) Based on your results from parts (a) and (b), provide discussion on empirical comparison between the two models. Moreover, use the *quadratic loss L measure* for formal comparison of the two models, in particular, to check if the hierarchical Poisson GLM offers an improvement to the fit of the non-hierarchical GLM. (Provide details on the required expressions for computing the value of the model comparison criterion.)