Consider the data set on the incidence of faults in the manufacturing of rolls of fabric:


(recall that this data set was also studied as part of homework set 2). 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 $\mu_i$.

(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 $\beta_1$ and $\beta_2$, 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

$$\mu_i | \gamma_i, \lambda \sim \frac{1}{\Gamma(\lambda)} \left( \frac{\lambda}{\gamma_i} \right)^\lambda \mu_i^{\lambda-1} \exp \left( -\frac{\lambda}{\gamma_i} \mu_i \right) \quad \mu_i > 0; \quad \lambda > 0, \gamma_i > 0,$$

where $\log(\gamma_i) = \beta_1 + \beta_2 x_i$. (Here, $\Gamma(u) = \int_0^\infty t^{u-1} \exp(-t) dt$ is the Gamma function.) Derive the expressions for $E(Y_i | \beta_1, \beta_2, \lambda)$ and $\text{Var}(Y_i | \beta_1, \beta_2, \lambda)$, 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 $\beta_1$, $\beta_2$ and $\lambda$, as well as point and interval estimates for the response mean as a function of the covariate (over a grid of covariate values). Finally, obtain the distributions of the posterior predictive residuals.

(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.)

Note: For both the standard Poisson glm and its hierarchical extension, you may work with a flat prior for $(\beta_1, \beta_2)$. Under the hierarchical glm, you may use the (proper) prior, $p(\lambda) = (\lambda + 1)^{-2}$, for $\lambda > 0$. 

---

AMS 274 – Generalized Linear Models (Spring 2010)

Homework 5 (due by 12pm, Friday June 4)

Please email me a pdf file with your hwk; alternatively, you can slide your hwk under my office door. The hwk papers will be collected from my office at noon on Friday June 4th, so please note that the due date above to submit the hwk is a strict one.