

Homework 5

Ex1 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 , as well as point and interval estimates for the response mean as 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

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

where $\log(\gamma_i) = \beta_1 + \beta_2 x_i$. (Here, $\Gamma(u) = \int_0^\infty t^{u-1} \exp(-t) dt$ is the Gamma function.) Derive the expression for $E(Y_i | \beta_1, \beta_2, \lambda)$ and $Var(Y_i | \beta_1, \beta_2, \lambda)$, 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 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.)

Sol 1 (a) Given the flat prior of β , the posterior distribution of β are

$$Pr(\beta|x, y) \propto \pi(\beta)L(\beta; x, y) \quad (1)$$

$$= \prod_{i=1}^n \frac{\lambda_i^{y_i} \exp(-\lambda_i)}{y_i!} \quad (2)$$

where $\lambda_i = \mu_i = \exp(\beta_1 + \beta_2 x_i)$. Using the Metropolis Hasting algorithm, we obtain the posterior distributions of β . After picking up reasonable number of samples and burn-in, the empirical posterior distributions are plotted in Figure 1. Then the point

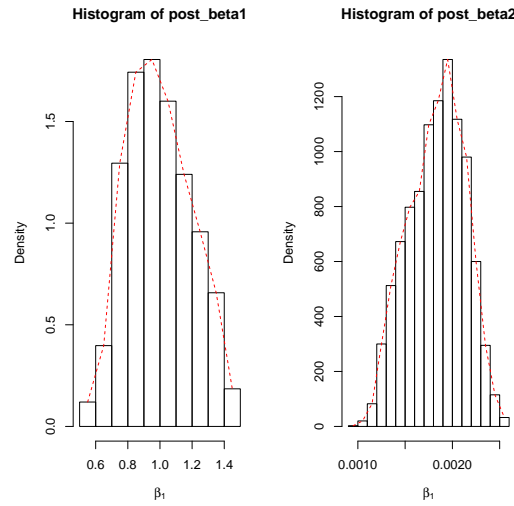


Figure 1: Empirical posterior distributions of β_1 and β_2

and 95% centered interval estimates are shown in Figure 2 Since this case has 32

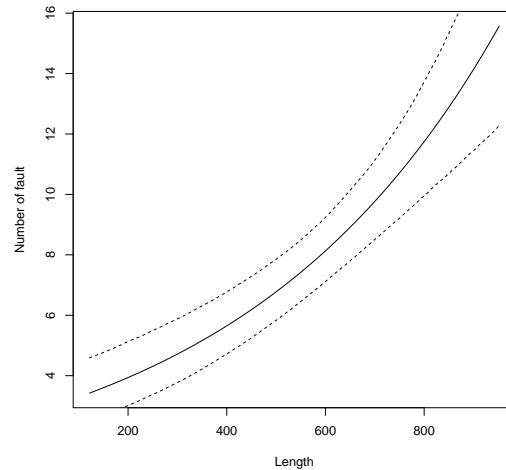


Figure 2: The response mean curve

posterior predictive residual distributions, it is difficult to show all distributions

here. Therefore, for each covariate, we first sort observation \mathbf{y} by the increasing order in \mathbf{x} , then generate boxplots of those posterior predictive distributions in Figure 3. It shows that many posterior predictive residual distributions deviate from

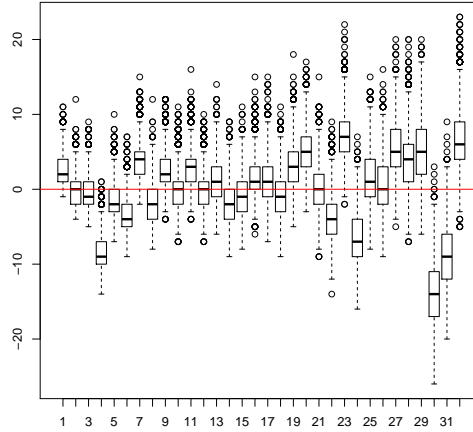


Figure 3: Boxplots for all posterior predictive residuals

zero. And as length increases, the deviation for the residuals is more significant. In other words, those residuals violate the zero mean assumption. Moreover, only 15 y_i located in the 50% center interval estimates of their corresponding posterior predictive distribution. So, the model checking shows this model is not ideal.

(b) The expressions for $E(Y_i|\beta_1, \beta_2, \lambda)$ and $Var(Y_i|\beta_1, \beta_2, \lambda)$ are

$$E(Y_i|\beta_1, \beta_2, \lambda) = E(E(Y_i|\mu_i)) \quad (3)$$

$$= \gamma_i \quad (4)$$

$$= \exp(\beta_1 + \beta_2 x_i) \quad (5)$$

and

$$Var(Y_i|\beta_1, \beta_2, \lambda) = Var(E(Y_i|\mu_i)) + E(Var(Y_i|\mu_i)) \quad (6)$$

$$= \frac{\gamma_i^2}{\lambda} + \gamma_i \quad (7)$$

$$= \frac{\exp^2(\beta_1 + \beta_2 x_i)}{\lambda} + \exp(\beta_1 + \beta_2 x_i) \quad (8)$$

As for the non-hierarchical, the mean and variance are $E(Y_i|\beta_1, \beta_2) = \exp(\beta_1 + \beta_2 x_i)$ and $Var(Y_i|\beta_1, \beta_2) = \exp(\beta_1 + \beta_2 x_i)$.

In the hierarchical model, the parameters are $\boldsymbol{\mu}, \beta_1, \beta_2$ and λ . The posterior distri-

bution of those parameters are

$$\Pr(\boldsymbol{\mu}, \beta_1, \beta_2, \lambda | \mathbf{y}, \mathbf{x}) \quad (9)$$

$$\propto \pi(\boldsymbol{\mu}, \beta_1, \beta_2, \lambda) \prod_{i=1}^n \text{Pois}(y_i | \mu_i) \text{Gamma}(\mu_i | \lambda, \frac{\gamma_i}{\lambda}) \quad (10)$$

where $\gamma_i = \exp(\beta_1 + \beta_2 x_i)$, λ is the shape parameter and $\frac{\gamma_i}{\lambda}$ is the scale parameter. Besides, the posterior predictive distribution of a new response y_0 is $\Pr(y_0 | \mathbf{x}, \mathbf{y}, x_0) = \int \Pr(y_0 | \beta_1, \beta_2, \lambda, x_0) \Pr(\beta_1, \beta_2, \lambda | \mathbf{x}, \mathbf{y}) d(\beta_1, \beta_2, \lambda)$. And it can be estimated by $\hat{\Pr}(y_0 | \mathbf{x}, \mathbf{y}, x_0) = \frac{1}{G} \sum_{g=1}^G \Pr(y_0 | \beta_1^{(g)}, \beta_2^{(g)}, \lambda^{(g)}, x_0)$, where $\beta_1^{(g)}, \beta_2^{(g)}, \lambda^{(g)}$ are sampled from the posterior distribution. So $y_0^{(g)}$ can be sampled from $\Pr(y_0 | \beta_1^{(g)}, \beta_2^{(g)}, \lambda^{(g)}, x_0)$ where $\beta_1^{(g)}, \beta_2^{(g)}, \lambda^{(g)}$ are sampled from the posterior predictive distribution.

Assuming $\boldsymbol{\mu}$ have flat prior, then MCMC algorithm implement recursively update parameters as follows:

- 1 Given $\beta_1, \beta_2, \lambda$, sample the $\boldsymbol{\mu}$ from the posterior distribution by $\mu_i | \beta_1, \beta_2, \lambda, \mathbf{y}, \mathbf{x} \sim \text{Gamma}(\lambda + y_i, \frac{\gamma_i}{\lambda} + 1)$.
- 2 Sample the λ in log scale to guarantee that $\lambda > 0$ using M-H method.
- 3 Then sample the β_1 and β_2 using M-H method.

After selecting reasonable tuning parameters and burn-in, the posterior distributions of β_1, β_2 and λ are shown in Figure 4 Given the posterior samples of parameters,

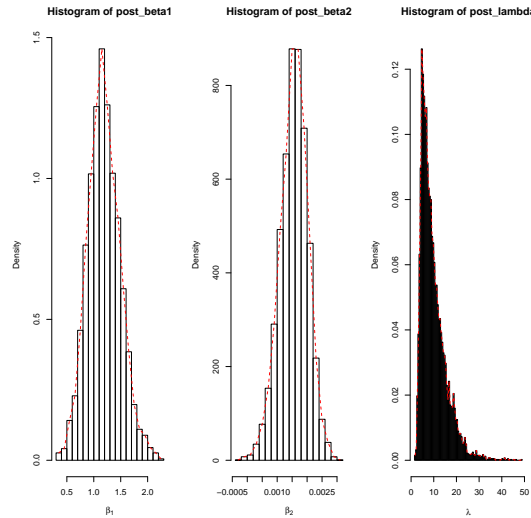


Figure 4: Empirical posterior distributions of β_1, β_2 and λ

we show the point and 95% centered interval estimates fro the response mean in Figure 5. Based on the same procedures as in (a), we plot the boxplots for the posterior predictive residuals following the increasing order in \mathbf{x} . From Figure 6,

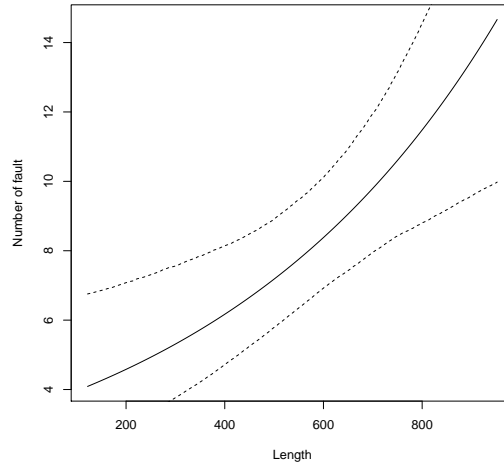


Figure 5: The response mean curve

compared with the non-hierarchical model, more posterior predictive residuals are close to 0.

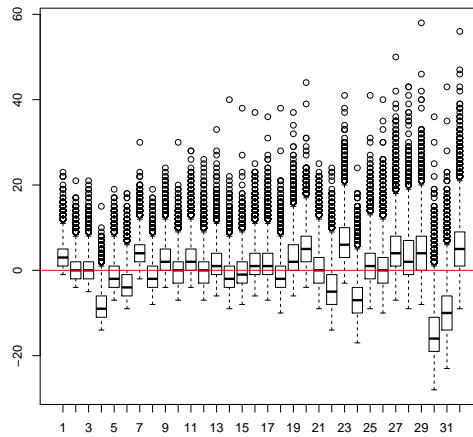


Figure 6: Boxplots for all posterior predictive residuals

Replacing the flat prior for λ by the prior $p(\lambda) = (\lambda + 1)^{-2}$. The empirical posterior distributions of parameters are shown in Figure 7. Comparing the posterior distributions between Figure 4 and Figure 7, the difference with respect to the parameter λ is a little significant. The posterior distribution of λ with flat prior is relatively flatter than that with prior $p(\lambda) = (\lambda + 1)^{-2}$, which means the prior sensitivity is not very significant and posterior distribution shows that λ is around 7. And comparing the empirical posterior predictive residuals Figure 6 and 7, the difference is not significant either. Perhaps, the model with prior $p(\lambda) = (\lambda + 1)^{-2}$ is a little better than the model with flat prior.

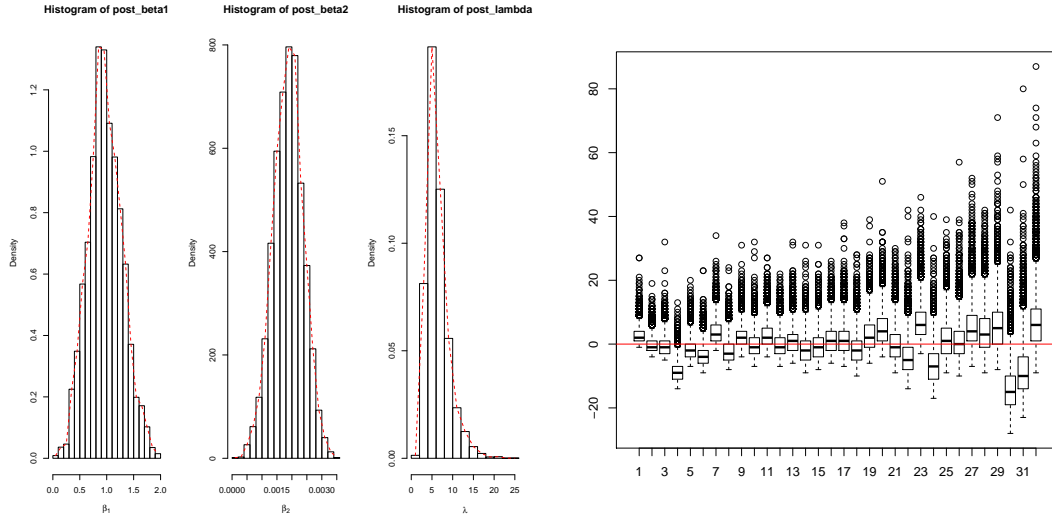


Figure 7: Empirical posterior distributions of β_1, β_2 and λ and boxplots for posterior predictive residuals

- (c) Looking through the boxplots of posterior predictive residuals, hierarchical model is better than non-hierarchical model, because the posterior predictive residuals in hierarchical model are more close to 0.

Quantitatively, we use the quadratic loss L measure for the formal comparison of the two models. (considering all flat priors)

The loss function is

$$L_k(M_j) = \sum_{i=1}^n Var^{M_j}(y_i^{rep}|data) + \frac{k}{k+1} \sum_{i=1}^n (y_i^{rep} - E^{M_j}(y_i^{rep}|data))^2 \quad (11)$$

The quadratic loss L measures are summarized in the Table 1.

The Table 1 shows that loss functions for hierarchical model is significantly smaller

Table 1: Quadratic loss L measures for both non-hierarchical and hierarchical models

Model	Non-hierarchical model	Hierarchical model
$k/(k+1) \rightarrow 0$	897.72	689.80
$k/(k+1) \rightarrow 1$	55242.79	39616.07

than those for non-hierarchical model, no matter how to select the regularity parameter k . So quantitatively the hierarchical model is better than the non-hierarchical model.