AMS 274 Homework #5 **Mary Silva**
UCSC 13 December 2018

a) We fit a Bayesian Poisson GLM with the logarithmic link, $\log(\mu_i) = \beta_1 + \beta_2 x_i$. The posterior distributions for β_1 and β_2 (under a flat prior for (β_1, β_2)) are summarized in table 1 and figure 1.

	mean	sd	HPD 2.5%	HPD 97.5%
β_0	0.965088	0.211464	0.563064	1.384523
β_1	0.001934	0.000304	0.001328	0.002520

Table 1: Summary of posterior distribution for β_1 and β_2.

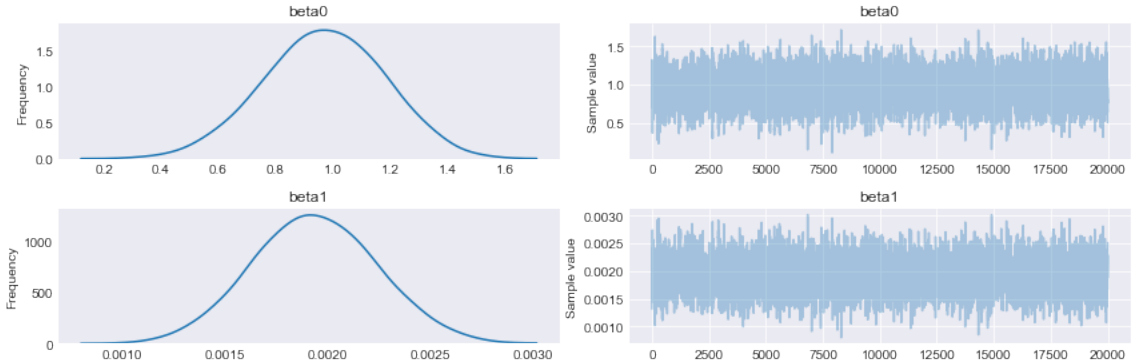

Figure 1: Posterior density and trace plot for β_1 and β_2.

Next, in figure 2, we plot point and interval estimates for the response mean as a function of the covariate (over a grid of covariate values).

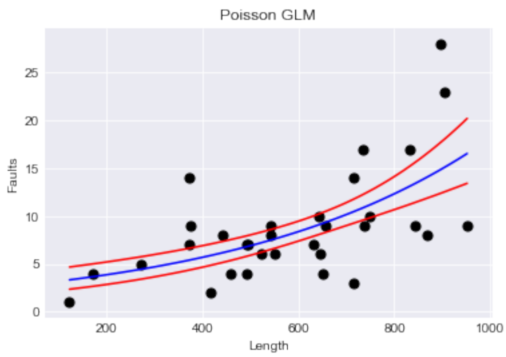

Figure 2: Point and interval estimates for the response mean as a function of the covariate. The red lines are the 95% prediction interval, the y-axis represents the response mean for Faults.

Next, we obtain samples from the posterior and compare the inferred means to the actual sample mean (Figure 3). Also in Figure 3, we plot the posterior predictive residual distribution for each observation, plotted in ascending order of length of rolls. The residuals do not cover 0 in many places of the plot. We see that zero is at the tails of the residual distributions. When x is large, the residual distributions are further from zero. Thus, we consider a hierarchical model.

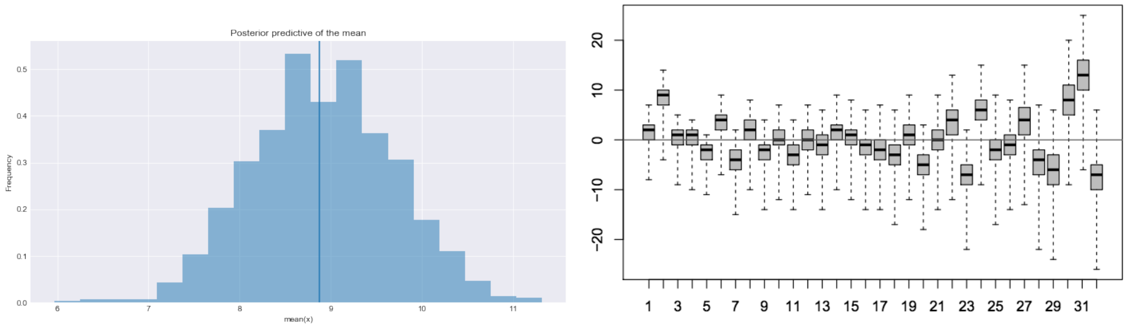

Figure 3: (Left) Posterior predictive of the mean. (Right) Posterior predictive residual boxplot with observations in ascending order of x along the x-axis.

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 $\mu_i|\gamma_i, \lambda \sim Gamma(\lambda, \lambda\gamma_i^{-1})$, where $\log(\gamma_i) = \beta_1 + \beta_2 x_i$.

Under the Bayesian Poisson Hierarchical GLM,

$$
\begin{aligned}
E\left(\mu_i|\gamma_i, \lambda\right) &= \frac{\lambda}{\lambda/\gamma_i} = \gamma_i \\
Var\left(\mu_i|\gamma_i, \lambda\right) &= \frac{\lambda}{(\lambda/\gamma)^2} = \frac{\gamma^2}{\lambda}
\end{aligned}
$$

Thus,

$$
E\left(Y_i|\beta_1, \beta_2, \lambda\right) = E\left[E(Y_i|\mu_i)|\gamma_i, \lambda\right] = E(\mu_i|\gamma_i, \lambda) = \gamma_i = \exp(\beta_1 + \beta_2 x_i)
$$

Since $E\left(Y_i|\beta_1, \beta_2\right) = \mu_i = \exp(\beta_1 + \beta_2 x_i)$, then

$$
E\left(Y_i|\beta_1, \beta_2, \lambda\right) = E\left(Y_i|\beta_1, \beta_2\right) = \exp(\beta_1 + \beta_2 x_i)
$$

Also

$$
\begin{aligned}
Var(Y_i|\beta_1, \beta_2, \lambda) &= E[Var(Y_i|\mu_i)|\gamma_i, \lambda] + Var[E(Y_i|\mu_i)|\gamma_i, \lambda] \\
&= E(\mu_i|\gamma_i, \lambda) + Var(\mu_i|\gamma_i, \lambda) \\
&= \gamma_i + \frac{\gamma_i^2}{\lambda} \\
&= \exp(\beta_1 + \beta_2 x_i)\left(1 + \frac{\exp(\beta_1 + \beta_2 x_i)}{\lambda}\right)
\end{aligned}
$$

Under standard poisson GLM, the mean and variance are the same, $Var(Y_i|\beta_1, \beta_2) = \mu_i = \exp(\beta_1 + \beta_2 x_i)$. Since

$$
\frac{\exp(\beta_1 + \beta_2 x_i)}{\lambda} > 0
$$

for all $\lambda > 0, \beta_1, \beta_2 \in \mathcal{R}$, then $Var(Y_i|\beta_1, \beta_2, \lambda) > Var(Y_i|\beta_1, \beta_2)$. Thus, hierarchical Poisson GLM is an overdispersion model that loses the constraint of equality of mean and variance. Hierarchical GLM allows more flexible mean-variance relationship. *Bedrick,* *E.J., Christensen, R. and Johnson, W. (1996).*

Next, we develop an MCMC method for posterior simulation providing details for all its steps and 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.

Suppose $p(\lambda, \beta_1, \beta_2) = p(\lambda)p(\beta_1, \beta_2)$ and assigning flat priors to $\boldsymbol{\beta}$, and

$$p(\lambda) = \frac{1}{(\lambda+1)^2}$$

Then the posterior distribution becomes

$$p(\mu_1, ..., \mu_n, \lambda, \beta_1, \beta_2|\text{data}) \propto \left(\prod_{i=1}^{n} \frac{\mu_i^{y_i} \exp -\mu_i}{y_i!}\right) \left(\prod_{i=1}^{n} \frac{1}{\Gamma(\lambda)} \left(\frac{\lambda}{\gamma_i}\right)^{\lambda} \mu_i^{\lambda-1} \exp\left(-\frac{\lambda}{\gamma_i}\mu_i\right)\right) \frac{1}{(\lambda+1)^2}$$

The full conditionals then become

$$(\mu_i|\cdot) \sim \text{Gamma}\left(\mu_i|y_i+\lambda, \frac{\lambda}{\exp(\beta_1+\beta_2 x_i)}\right)$$
$$p(\lambda|\cdot) \propto \left(\prod_{i=1}^{n} \frac{1}{\Gamma(\lambda)} \left(\frac{\lambda}{\gamma_i}\right)^{\lambda} \mu_i^{\lambda-1} \exp\left(-\frac{\lambda}{\gamma_i}\mu_i\right)\right) \frac{1}{(\lambda+1)^2}$$

Let $w = \log(\lambda) \in (-\infty, \infty)$, then $\lambda = e^w$ and $d\lambda/dw = e^w$. Thus,

$$p(w|\cdot) \propto \left(\prod_{i=1}^{n} \frac{1}{\Gamma(e^w)} \left(\frac{e^w}{\exp(\beta_1+\beta_2 x_i)}\right)^{e^w} \mu_i^{e^w-1} \exp\left(-\frac{e^w}{\exp(\beta_1+\beta_2 x_i)}\mu_i\right)\right) \frac{e^w}{(e^w+1)^2} := h(w)$$

Probably mistakes up in here

Thus,

$$\log(p(w|\cdot)) \propto \sum_{i=1}^{n} \log\left[\text{Gamma}\left(e^w, \frac{e^w}{\exp(\beta_1+\beta_2 x_i)}\right)\right] + w - 2\log(1+e^w)$$

Finally, the full conditional for (β_1, β_2):

$$p(\beta_1, \beta_2|\cdot) \propto \left(\prod_{i=1}^{n} \frac{1}{\Gamma(\lambda)} \left(\frac{\lambda}{\exp(\beta_1+\beta_2 x_i)}\right)^{\lambda} \mu_i^{\lambda-1} \exp\left(-\frac{\lambda}{\exp(\beta_1+\beta_2 x_i)}\mu_i\right)\right) := G(\beta_1, \beta_2)$$

The MCMC sampling algorithm is outlined below

Algorithm 1: MCMC algorithm for Hierarchical GLM

Initialize $\lambda^{(0)}, \beta_1^{(0)}, \beta_2^{(0)}$
for $j = 1, ...,$ convergence :
1. **for** $i = 1, ..., n$:
   sample $\mu_i^{(j)} \sim \text{Gamma}\left(\mu_i | y_i + \lambda, \frac{\lambda}{\exp(\beta_1+\beta_2 x_i)}\right)$
2. Sample $w* \sim N(w^{(j-1)}, \nu)$ where ν is a tuning parameter.
3. **if** $U \sim Unif(0,1) < \rho_w = min\left\{\frac{h(w^*)}{h(w^{(j-1)})}, 1\right\}$:
   let $w^{(j)} = w^*$ **and** $\lambda^{(j)} = \lambda* = e^{w^*}$
   else:
   $w^{(j)} = w^{(j-1)}$ and $\lambda^{(j)} = \lambda^{(j-1)}$
4. Sample $(\beta_1, \beta_2)^* \sim N_2\left((\beta_1, \beta_2)^{(j-1)}, dJ^{-1}(\hat{\beta}_1, \hat{\beta}_2)\right)$, where J^{-1} is the Fisher Information, and $\hat{\boldsymbol{\beta}}$ are the MLE.
5. **if** $U \sim Unif(0,1) < \rho_w = min\left\{\frac{G((\beta_1,\beta_2)^*)}{G((\beta_1,\beta_2)^{(j-1)})}, 1\right\}$:
   $(\beta_1, \beta_2)^{(j)} = (\beta_1, \beta_2)^*$
   else:
   $(\beta_1, \beta_2)^{(j)} = (\beta_1, \beta_2)^{(j-1)}$

end.

Using the above algorithm, we obtain the following posterior estimates

	Mean	SD	2.5%	97.5%
β_1	1.006	0.29	0.43	1.58
β_2	0.0019	0.00	0.00	0.00
λ	8.911	6.26	3.34	22.31

Figure 4 shows the estimated response mean as a function of the covariate (over a grid of covariate values). Compared to the curve obtained with the standard GLM, this one appears to have a wider prediction interval. This is because the model accounts for large variation of the data. The mean point estimates for both models are nearly the same.

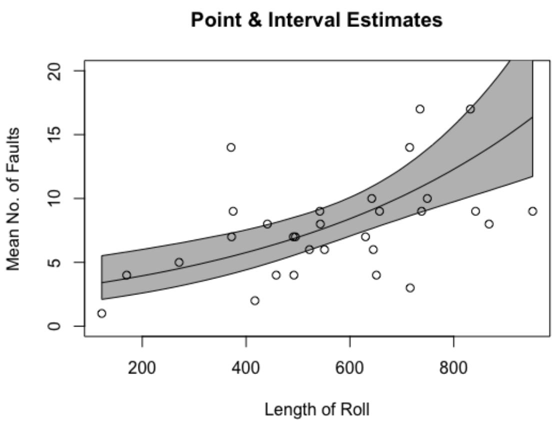

Figure 4: Point and 95% interval estimates for the estimated resonse mean for the hierarchical model.

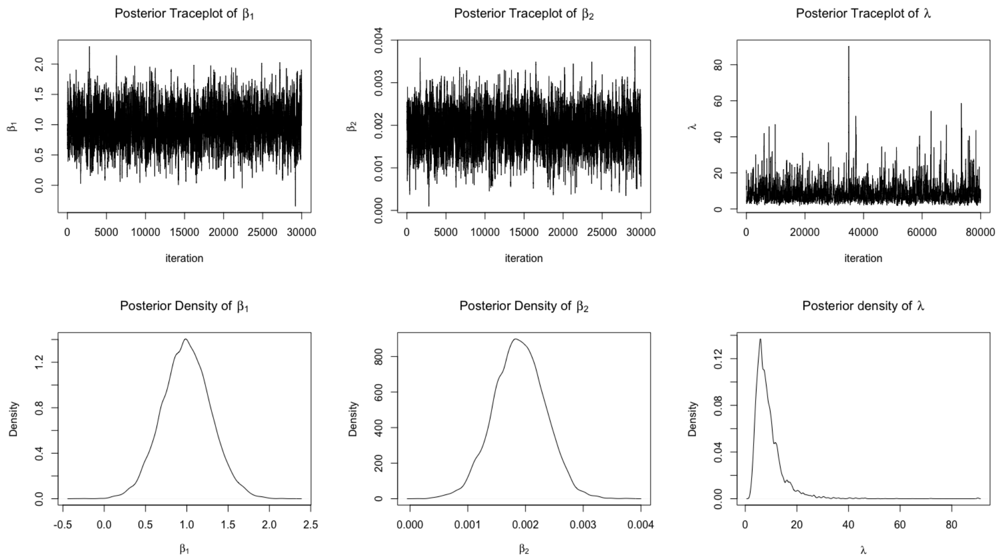

Figure 5: Convergence and density plots for the parameters of the hierarchical Poisson GLM.

Figure 5 shows convergence of the parameters and the posterior densities using the hierarchical Poisson GLM. Figure 6 shows the posterior predictive residual boxplot under hierarchical Poisson GLM. We see here that most of the interquartile ranges cover zero, unlike the case of the standard Poisson GLM in part (a). This implies that a hierarchical GLM has a higher predictive power, but due to larger variation, there's more uncertainty.

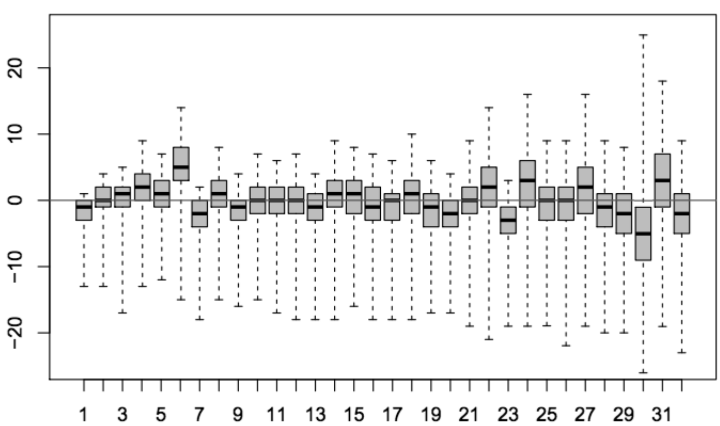

Figure 6: Boxplot of posterior predictive residuals with observations in ascending order of x along the x-axis

Figure 7: Quadratic loss L measures for both models.

c) Let $\boldsymbol{z} = (z_1, ...z_n)'$ denote the future response vector with the same sampling density $(y|\cdot)$. Gelfand & Ghosh (1998) and Ibraham et. al (2001) consider the quadratic loss L measure

$$L_q(k) = \sum_{i=1}^{n} Var(z_i|x_i, D) + k \sum_{i=1}^{n} (\mu_i - y_i)^2$$

Where $\mu_i = E(z_i|x_i, D)$. Essentially, this expression is the summation of some penalty term for model complexity and a goodness-of-fit term. A larger choice of k penalizes models that are more complex. Figure 7, shows the quadratic loss L measures for standard and hierarchical Poisson model. The standard Poisson GLM is only better when $k = 0$, which is when the Loss ignores the goodness of fit term. This is due to the fact that the hierarchical model is an overdispersed model. For $k \geq 1$, the hierarchical model has smaller quadratic loss, which leads us to the conclusion that the hierarchical model has a better fit than the model described in part (a).