1. 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](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 $\mu_i$.

(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 $\beta_2$ (under flat prior for $(\beta_1, \beta_2)$), 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.

**Solution:**

The posterior distribution is

$$p(\beta_1, \beta_2 \mid data) \propto p(\beta_1, \beta_2) \prod_{i=1}^{n} \text{Pois}(y_i \mid \mu_i)$$

$$= p(\beta_1, \beta_2) \prod_{i=1}^{n} \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}$$

$$= \prod_{i=1}^{n} \frac{(\exp(\beta_1 + \beta_2 x_i))^y e^{\exp(-\exp(\beta_1 + \beta_2 x_i))}}{y_i!}$$

$$:= S(\beta_1, \beta_2),$$

given flat prior and log link.

**Algorithm 1:** Metropolis-Hastings Algorithm for Poisson GLM with log link.

Start with initial values $\beta_1^{(0)}$ and $\beta_2^{(0)}$.

for iteration $m = 1, 2, \ldots$ until convergence do

1. Draw $\beta_1, \beta_2 \sim N_2((\hat{\beta}_1, \hat{\beta}_2)^{(m-1)}, dJ^{-1}(\hat{\beta}_1, \hat{\beta}_2))$, where $J^{-1}$ is the Fisher information, $d$ is the tuning parameter and $\hat{\beta}_1, \hat{\beta}_2$ are the MLEs;

2. Set $(\beta_1, \beta_2)^{(m)} = (\beta_1, \beta_2)^{*}$ if $U \sim Unif(0, 1) < \rho = \min \left\{ \frac{S(\beta_1, \beta_2)^{*}}{S((\beta_1, \beta_2)^{(m-1)})}, 1 \right\}$;

$(\beta_1, \beta_2)^{(m)} = (\beta_1, \beta_2)^{(m-1)}$ otherwise.

end

Like binomial GLM, an Metropolis-Hastings algorithm is used to simulate posterior sample for $(\beta_1, \beta_2)$. A random walk normal distribution proposal is used with the Fisher information evaluated at the MLEs $(\beta_1, \beta_2)$ factored by the tuning parameter $d$. The initial values for $(\beta_1, \beta_2)$ is set to $(0, 0)$. The total number of iterations is 110000.
Figure 1: Convergence of parameters of Poisson GLM in (a).

The first 10000 draws are burned out and I thin the chain by keeping every 20th draw and discard the rest. The 5000 draws are left for analysis. The tuning parameter is chosen so that the acceptance rate is around 40%.

<table>
<thead>
<tr>
<th></th>
<th>PostMean</th>
<th>2.5%Quantile</th>
<th>97.5%Quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>0.9677</td>
<td>0.5524</td>
<td>1.3759</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.0019</td>
<td>0.0013</td>
<td>0.0025</td>
</tr>
</tbody>
</table>

Table 1: Posterior summary for ($\beta_1, \beta_2$) of Poisson GLM in (a).

Figure 1 shows a good convergent behavior of the parameters $\beta_1$ and $\beta_2$. Posterior summary is in Table 1. The posterior means are close to the MLEs derived in homework 2. Figure 2 shows the estimated response mean as a function of the covariate. For a Poisson random variable, large mean implies large variance, and so the interval is wider when $x$ is large (larger $\mu$ as well because $\beta_2 > 0$). However, the equality of mean and variance makes the model inflexible that poorly explains the overdispersed data for any given $x$ as the 95% interval is too narrow to include many observed values.

To get the posterior predictive residuals, for $i = 1, 2, ..., n$, and for MCMC iteration $m = 1, ..., M$, we do the followings.

- Generate new (replicate) responses sample $z_i = \{z_i^{(m)}\}_{m=1}^M$, where $z_i^{(m)}$ sampled from $\text{Pois} \left( y_i | \mu_i = \exp(\beta_1^{(m)} + \beta_2^{(m)} x_i) \right)$ is the replicate response corresponding to $(y_i, x_i)$ and posterior sample $(\beta_1, \beta_2)^{(m)}$.
- Derive the posterior predictive residual $r_{i,pp}^{(m)}$ by $r_{i,pp}^{(m)} = y_i - z_i^{(m)}$. 

Figure 2: Point and interval estimates for the estimated response mean $\hat{\mu}$ in (a). The solid red line is the posterior mean. Shaded areas between two dashed purple lines are 95% interval estimates. Black points are observed data.

Hence the posterior predictive residual distribution is summarized by the sample $\{r_{i,pp}^{(m)}\}_{m=1}^M$.

The posterior predictive residual distribution for each observation is shown in Figure 3. Observations are numbered by ascending order of length of rolls. The residuals are of large variations, but many interquartile ranges do not cover zero, showing that zero is at the tails of residual distributions. When $x$ is large, the corresponding residual distributions are more far away from zero and have more variations. We may consider a more reasonable model to have a better fit.

(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

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

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 $E(Y_i \mid \beta_1, \beta_2, \lambda)$ and $\text{Var}(Y_i \mid \beta_1, \beta_2, \lambda)$, 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 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). Discuss model checking results based on posterior predictive residuals.

Regarding the priors, you can use again the flat prior for $(\beta_1, \beta_2)$, but perform prior sensitivity analysis for $\lambda$ considering different proper priors, including $p(\lambda) = (\lambda + 1)^{-2}$.

Solution:
Figure 3: Posterior predictive residual boxplot. Observation is numbered by ascending order of length of rolls.
In the second stage of the hierarchical model, the distribution of $\mu_i \mid \gamma_i, \lambda$ is $Ga(\lambda, \lambda/\gamma_i)$ ($Ga$ stands for gamma distribution), and so

$$E(\mu_i \mid \gamma_i, \lambda) = \frac{\lambda}{\lambda/\gamma_i} = \gamma_i$$
$$Var(\mu_i \mid \gamma_i, \lambda) = \frac{\lambda}{(\lambda/\gamma_i)^2} = \gamma_i^2/\lambda.$$

Therefore, given $log(\gamma_i) = \beta_1 + \beta_2x_i$,

$$E(Y_i \mid \beta_1, \beta_2, \lambda) = E[E(Y_i \mid \mu_i) \mid \gamma_i, \lambda] = E(\mu_i \mid \gamma_i, \lambda) = \gamma_i = \exp(\beta_1 + \beta_2x_i).$$

For standard Poisson GLM, since $log(\mu_i) = \beta_1 + \beta_2x_i$,

$$E(Y_i \mid \beta_1, \beta_2) = \mu_i = \exp(\beta_1 + \beta_2x_i).$$

Hence $E(Y_i \mid \beta_1, \beta_2, \lambda) = E(Y_i \mid \beta_1, \beta_2) = \exp(\beta_1 + \beta_2x_i).$

Under hierarchical Poisson GLM,

$$Var(Y_i \mid \beta_1, \beta_2, \lambda) = E[Var(Y_i \mid \mu_i) \mid \gamma_i, \lambda] + Var[E(Y_i \mid \mu_i) \mid \gamma_i, \lambda]$$
$$= E(\mu_i \mid \gamma_i, \lambda) + Var(\mu_i \mid \gamma_i, \lambda)$$
$$= \gamma_i + \frac{\gamma_i^2}{\lambda}$$
$$= \gamma_i \left(1 + \frac{\gamma_i}{\lambda}\right)$$
$$= \exp(\beta_1 + \beta_2x_i) \left(1 + \frac{\exp(\beta_1 + \beta_2x_i)}{\lambda}\right).$$

However, under standard Poisson GLM, since the mean and variance are the same for Poisson distribution,

$$Var(Y_i \mid \beta_1, \beta_2) = \mu_i = \exp(\beta_1 + \beta_2x_i).$$

Because $\frac{\exp(\beta_1 + \beta_2x_i)}{\lambda} > 0$ for all $\lambda > 0, \beta_1 \in \mathbb{R}, \beta_2 \in \mathbb{R}$, $Var(Y_i \mid \beta_1, \beta_2, \lambda) > Var(Y_i \mid \beta_1, \beta_2)$. Hence, hierarchical Poisson GLM is an overdispersion model that looses the constraint of equality of mean and variance and allows for flexible mean-variance relationship.

Assuming $p(\lambda, \beta_1, \beta_2) = p(\lambda)p(\beta_1, \beta_2)$ and if $p(\beta_1, \beta_2) \propto 1$ and $p(\lambda) = (\lambda + 1)^{-2}$, the target posterior distribution is

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

Now we derive full conditionals.
• Full conditional for $\mu_i$, $i = 1, ..., n$.
\[
p(\mu_i | \cdots) \propto \mu_i^{y_i} e^{-\mu_i} \mu_i^{\lambda - 1} \exp \left( -\frac{\lambda}{\gamma_i} \mu_i \right) \\
= \mu_i^{y_i + \lambda - 1} \exp \left( -\frac{\lambda}{\gamma_i} \mu_i \right) \\
\sim Ga \left( \mu_i | y_i + \lambda, 1 + \frac{\lambda}{\exp(\beta_1 + \beta_2 x_i)} \right).
\]

• Full conditional for $\lambda$.
\[
p(\lambda | \cdots) \propto \prod_{i=1}^{n} \frac{1}{\Gamma(\lambda)} \left( \frac{\lambda}{\gamma_i} \right)^\lambda \mu_i^{\lambda - 1} \exp \left( -\frac{\lambda}{\gamma_i} \mu_i \right) (\lambda + 1)^{-2}.
\]

Let $w = \log(\lambda) \in (-\infty, \infty)$. Then $\lambda = e^w$ and $d\lambda/dw = e^w$. Therefore,
\[
p(w | \cdots) \propto \prod_{i=1}^{n} \frac{1}{\Gamma(e^w)} \left( e^w \exp(\beta_1 + \beta_2 x_i) \right)^{e^w} \mu_i^{e^w - 1} \exp \left( -\frac{e^w}{\exp(\beta_1 + \beta_2 x_i)} \mu_i \right) (e^w + 1)^{-2 e^w} \\
:= h(w).
\]

Hence,
\[
\ell(w | \cdots) := \log p(w | \cdots) \propto \sum_{i=1}^{n} \log \left[ Ga \left( e^w, \frac{e^w}{\exp(\beta_1 + \beta_2 x_i)} \right) \right] + w - 2 \log(1 + e^w).
\]

• Full conditional for $(\beta_1, \beta_2)$.
\[
p(\beta_1, \beta_2 | \cdots) \propto \prod_{i=1}^{n} \frac{1}{\Gamma(\lambda)} \left( \frac{\lambda}{\gamma_i} \right)^\lambda \mu_i^{\lambda - 1} \exp \left( -\frac{\lambda}{\gamma_i} \mu_i \right) \\
= \prod_{i=1}^{n} \frac{1}{\Gamma(\lambda)} \left( \frac{\lambda}{\exp(\beta_1 + \beta_2 x_i)} \right)^\lambda \mu_i^{\lambda - 1} \exp \left( -\frac{\lambda}{\exp(\beta_1 + \beta_2 x_i)} \mu_i \right) \\
:= S(\beta_1, \beta_2)
\]

An MCMC algorithm for posterior simulation is described in Algorithm 2. We direct sample $\mu_i$ from its full conditional distribution. To get a sample of $\lambda$, we do transformation of variables that $w = \log(\lambda)$, and then use random walk normal proposal to simulate $w$ and hence $\lambda$. For $(\beta_1, \beta_2)$, we follow the same Metropolis-Hastings step as standard Poisson GLM in (a). The initial values for $(\beta_1, \beta_2)$ and lambda are (0, 0), and 1000, respectively. The total number of iterations is 1100000. The first 100000 draws are burned out and I thin the chain by keeping every 200th draw and discard the rest. The 5000 draws are left for analysis. The tuning parameters $v$ and $d$ are chosen so that the acceptance rate for $w$ and $(\beta_1, \beta_2)$ are both around 38%.

The convergent behavior of the parameters $\beta_1$, $\beta_2$ and $\lambda$ shown in Figure 4 looks healthy after a million of iterations. Posterior summary is in Table 2. The posterior means of $\beta_1$ and $\beta_2$ are close to the means under the standard Poisson GLM. The posterior mean of $\lambda$ is 8.842, which is not that big and so it shows that overdispersion is captured by the hierarchical Poisson GLM.

To get the response mean curve, for each grid value of the covariate $x_j$, $j = 1, ..., J$, and for each posterior sample, $\lambda^{(m)}, \beta_1^{(m)}, \beta_2^{(m)}$, $m = 1, ..., M$, we sample $\mu_j^{(m)}$ from
Figure 4: Convergence of parameters of hierarchical Poisson GLM in (b).

Estimated response mean (Hierarchical)

Figure 5: Point and interval estimates for the estimated response mean \( \hat{\mu} \) in (b). The solid red line is the posterior mean. Shaded areas between two dashed purple lines are 95% interval estimates. Black points are observed data.
Algorithm 2: MCMC Algorithm for Hierarchical Poisson GLM with log link in (b).

Start with initial values $\lambda^{(0)}, \beta_1^{(0)}$ and $\beta_2^{(0)}$.

for iteration $m = 1, 2, \ldots$ until convergence do

1. for $i = 1, \ldots, n$ do
   draw $\mu_i^{(m)} \sim Ga \left( \mu_i \mid y_i + \lambda^{(m-1)} \right) \frac{\lambda^{(m-1)}}{\exp(\beta_1^{(m-1)} + \beta_2^{(m-1)} x_i)}$
   end

2. Draw $w^* \sim N(w^{(m-1)}, v)$, where $v$ is a tuning parameter.

3. Set $w^{(m)} = w^*$ and $\lambda^{(m)} = \lambda^* = e^{w^*}$ if $U \sim Unif(0, 1) < \rho_w = \min \left\{ \frac{h(w^*)}{h(w^{(m-1)})}, 1 \right\}$;
   $w^{(m)} = w^{(m-1)}$ and $\lambda^{(m)} = \lambda^{(m-1)}$ otherwise.

4. Draw $(\beta_1, \beta_2)^* \sim N_2((\beta_1, \beta_2)^{(m-1)}, dJ^{-1}(\hat{\beta}_1, \hat{\beta}_2))$, where $J^{-1}$ is the Fisher information,
   $d$ is the tuning parameter and $\hat{\beta}_1, \hat{\beta}_2$ are the MLEs;

5. Set $(\beta_1, \beta_2)^{(m)} = (\beta_1, \beta_2)^*$ if $U \sim Unif(0, 1) < \rho = \min \left\{ \frac{s(\beta_1, \beta_2)^*}{s(\beta_1, \beta_2)^{(m-1)}}, 1 \right\}$;
   $(\beta_1, \beta_2)^{(m)} = (\beta_1, \beta_2)^{(m-1)}$ otherwise.

end

<table>
<thead>
<tr>
<th>PostMean</th>
<th>2.5%Quantile</th>
<th>97.5%Quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>1.0003</td>
<td>0.4097</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.0019</td>
<td>0.0010</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>8.8416</td>
<td>3.2492</td>
</tr>
</tbody>
</table>

Table 2: Posterior summary for $(\beta_1, \beta_2, \lambda)$ of Hierarchical Poisson GLM in (b).

$Ga(\lambda^{(m)}, \lambda^{(m)} / \exp(\beta_1^{(m)} + \beta_2^{(m)} x_j))$. Then for each sample $\{\mu_j^{(m)}\}_{m=1}^M$, we compute its
mean as the point estimate for response mean, and 2.5% and 97.5% quantiles for the
interval estimate.

Figure 5 shows the estimated response mean as a function of the covariate under hier-
archical Poisson GLM. Comparing to the curve under standard Poisson GLM, we find
that the two point estimate curves are similar, but the interval under the hierarchical
model is much wider, which is more reasonable because it accounts for large variation
of the data set by covering most of the data points. Note that both point estimate and
interval estimate curves under the hierarchical model are not smoothed curves. Instead
of a deterministic relationship $\log(\mu) = \beta_1 + \beta_2 x$ as in standard Poisson GLM, both
estimate curves are derived from sampling, and hence some stochastic randomness is
unavoidable.

The joint posterior predictive distribution of a new (unobserved) $y_0$ and $\mu_0$ correspond-
ing to a new specified covariate value $x_0$ not included in the observed $x_i$ is

$$p((y_0, \mu_0), \mu_1, \ldots, \mu_n, \lambda, \beta_1, \beta_2 \mid data, x_0) = Pois(y_0 \mid \mu_0)Ga \left( \mu_0 \mid \lambda, \frac{\lambda}{\exp(\beta_1 + \beta_2 x_0)} \right)$$

$$\times p(\mu_1, \ldots, \mu_n, \lambda, \beta_1, \beta_2 \mid data).$$

Hence

$$p(y_0 \mid data, x_0) = \int Pois(y_0 \mid \mu_0)Ga \left( \mu_0 \mid \lambda, \frac{\lambda}{\exp(\beta_1 + \beta_2 x_0)} \right) p(\lambda, \beta_1, \beta_2 \mid data) \, d\mu_0 \, d\lambda \, d\beta_1 \, d\beta_2.$$
Figure 6: Posterior predictive residual boxplot under hierarchical Poisson GLM in (b). Observation is numbered by ascending order of length of rolls.

To get this distribution, for MCMC iteration $m = 1, \ldots, M$, we do the followings.

- Sample new $\mu_0^{(m)}$ from $Ga \left( \mu_0 \middle| \lambda^{(m)}, \frac{\lambda^{(m)}}{\exp(\beta_1^{(m)} + \beta_2^{(m)} x_0)} \right)$
- Given $\mu_0^{(m)}$, sample $y_0^{(m)}$ from $Pois(y_0 \middle| \mu_0^{(m)})$.

The sample $\{y_0^{(m)}\}_{m=1}^{M}$ forms the posterior predictive distribution of $y_0$.

To get the posterior predictive residuals, for MCMC iteration $m = 1, \ldots, M$, we do the followings.

- For each $i = 1, \ldots, n$, generate replicate response for the observation $y_i$, $z_i^{(m)}$, from $Pois(y_i \middle| \mu_i^{(m)})$, where $\{\mu_i^{(m)}\}_{m=1}^{M}$ is the posterior sample for $\mu_i$.
- For each $i = 1, \ldots, n$, the posterior predictive residual is $r_{i,pp}^{(m)} = y_i - z_i^{(m)}$.

Hence $\{r_{i,pp}^{(m)}\}_{m=1}^{M}$ is the posterior predictive residual sample for observation $y_i$, $i = 1, \ldots, n$.

Figure 6 shows the posterior predictive residual boxplot under hierarchical Poisson GLM. Figure 3 has 17 interquartile ranges that do not cover zero, but Figure 6 only has three. The residuals under hierarchical Poisson GLM have higher chance being around zero, showing a stronger predicting power, though with large uncertainty due to large variation of the observed data.
To perform prior sensitivity analysis, I consider three more priors on $\lambda$, (1) $Ga(\lambda \mid 1, 1)$; (2) $Ga(\lambda \mid 1, 0.001)$; and (3) $Ga(\lambda \mid 1000000, 1000)$. (1) and (3) have the same variance value 1, but (1) has mean 1 while (3) has mean 1000. (2) and (3) share the same mean 1000, but (2) has variance 1000000 while (3) has variance 1. Therefore, (1) and (3) are relatively informative, and (2) and (3) are assumed a priori that the hierarchical model is quite similar to the standard model since $\lambda$ has large mean 1000.

Table 3 shows posterior summary of $\lambda$ under the prior $(1 + \lambda)^{-2}$ and three additional priors. The posterior distribution is sensitive to prior distributions. In particular, $Ga(\lambda \mid 1000000, 1000)$ prior dominates. Although not shown here, posterior estimates for $\beta_1$ and $\beta_2$ are similar under these four different prior distributions.

![Figure 7: Estimated response mean under different priors](image)

Table 3: Posterior estimates of $\lambda$ under different priors.

<table>
<thead>
<tr>
<th></th>
<th>PostMean</th>
<th>2.5% Quantile</th>
<th>97.5% Quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$</td>
<td>8.89</td>
<td>3.19</td>
<td>22.26</td>
</tr>
<tr>
<td>$\lambda$ (1)</td>
<td>4.54</td>
<td>2.23</td>
<td>7.96</td>
</tr>
<tr>
<td>$\lambda$ (2)</td>
<td>168.44</td>
<td>4.75</td>
<td>1664.31</td>
</tr>
<tr>
<td>$\lambda$ (3)</td>
<td>1000.00</td>
<td>998.00</td>
<td>1001.97</td>
</tr>
</tbody>
</table>

(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.)

Solution:
The standard Poisson GLM cannot capture overdispersion of the observed data. Its
posterior predictive residuals are of low probability being around zero, and its nar-
row interval for the response means treats many observed values as extreme values or
outliers. The hierarchical Poisson GLM overcomes these problems by allowing larger
variance in the model.

The quadratic loss $L$ measure is

$$
L_k = \sum_{i=1}^{n} \text{Var}(z_i | data) + \frac{k}{k+1} \sum_{i=1}^{n} (y_i - E(z_i | data))^2
$$

For the hierarchical model, we compute the measure as follows.

- For iteration $m = 1, ..., M$ and for $i = 1, ..., n$, sample $z_i^{(m)}$ from $Pois(y_i | \mu_i^{(m)})$,
  where $\{\mu_i^{(m)}\}_{m=1}^{M}$ is the posterior sample for $\mu_i$.
- Compute estimated $E(z_i | data) \approx \frac{1}{M} \sum_{m=1}^{M} z_i^{(m)}$.
- Compute estimated $\text{Var}(z_i | data) = \frac{1}{M} \sum_{m=1}^{M} (z_i^{(m)})^2 - \left( \frac{1}{M} \sum_{m=1}^{M} z_i^{(m)} \right)^2$.
- Compute $L_k^{\text{hier}}$ for any given $k$.

For the standard model, we compute the measure as follows.

- For iteration $m = 1, ..., M$ and for $i = 1, ..., n$, compute estimated $E(z_i | data) \approx \frac{1}{M} \sum_{m=1}^{M} \mu_i^{(m)}$, where $\mu_i^{(m)} = \exp(\beta_1^{(m)} + \beta_2^{(m)} x_i)$, and $\{\beta_1^{(m)}\}_{m=1}^{M} \{\beta_2^{(m)}\}_{m=1}^{M}$ are posterior samples for $\beta_1$ and $\beta_2$.
- Compute estimated $\text{Var}(z_i | data) = \frac{1}{M} \sum_{m=1}^{M} (z_i^{(m)})^2 - \left( \frac{1}{M} \sum_{m=1}^{M} \mu_i^{(m)} \right)^2$, where $z_i^{(m)} \sim Pois(y_i | \mu_i^{(m)})$.
- Compute $L_k^{\text{stan}}$ for any given $k$.

Figure 8 shows quadratic loss $L$ measures for standard and hierarchical Poisson model.
The standard Poisson GLM is better only when $k = 0$, the situation that $L_k$ measures
penalty term only and ignores goodness-of-fit term. It is not surprising because the
hierarchical model is an overdispersed model. But for all $k \geq 1$, the hierarchical model
has smaller quadratic loss than the standard model, which implies it has much better
goodness-of-fit than the standard model.
Figure 8: Quadratic loss L measures.