

Bayesian Hierarchical Poisson Regression Model for Overdispersed Count Data

Overview

Overdispersion occurs when count data appear more dispersed than expected under a reference model. Overdispersion can be caused by positive correlation among the observations, an incorrect model, an incorrect distributional specification, or incorrect variance functions. This example uses the MCMC procedure to fit a Bayesian hierarchical Poisson regression model to overdispersed count data. It displays how Bayesian hierarchical Poisson regression models are effective in capturing overdispersion and providing a better fit.

The SAS source code for this example is available as a text file attachment. In Adobe Acrobat, right-click the icon in the margin and select **Save Embedded File to Disk**. You can also double-click the icon to open the file immediately.

[source code](#)

Analysis

Count data frequently display overdispersion (more variation than expected from a standard parametric model). Breslow (1984) discusses these types of models and suggests several different ways to model them. Hierarchical Poisson models have been found effective in capturing the overdispersion in data sets with extra Poisson variation.

Hierarchical Poisson regression models are expressed as Poisson models with a log link and a normal variance on the mean parameter. More formally, a hierarchical Poisson regression model is written as

$$\begin{aligned} Y_{ij} | \lambda_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\ \log(\lambda_{ij}) &= \mathbf{X}_i \boldsymbol{\beta} + \gamma_{ij} \\ \gamma_{ij} &\sim \text{normal}(0, \sigma^2) \end{aligned}$$

for $i = 1, \dots, n$, $j = 1, \dots, J$, and $y \in \{0, 1, 2, \dots\}$.

Consider data collected by Margolin, Kaplan, and Zeiger (1981) and studied in Breslow (1984) from an Ames salmonella mutagenicity assay. [Table 1](#) reports the number of revertant colonies of TA98 salmonella (SALM) tested at six dose levels of quinoline (DOSE) with three replicate plates.

Table 1 Salmonella Data Set

Doses of Quinoline (μg per plate)					
0	10	33	100	333	1000
15	16	16	27	33	20
21	18	26	41	38	27
29	21	33	60	41	42

The following statements read the data into SAS and create the ASSAY data set. The variable $L_DOSE10 = \log(\text{DOSE} + 10)$ models the mutagenic effect of quinoline. The value 10 is chosen because it is the smallest nonzero treatment level.

```
data assay;
  input dose salm @@;
  l_dose10 = log(dose+10);
  plate = _n_;
  datalines;
  0 15 0 21 0 29
  10 16 10 18 10 21
  33 16 33 26 33 33
  100 27 100 41 100 60
  333 33 333 38 333 41
  1000 20 1000 27 1000 42
  ;
```

The data appear to be a likely candidate for a Bayesian hierarchical Poisson regression model with replicates. However, you usually begin with a standard analysis of a Poisson regression and evaluate any evidence for overdispersion.

Bayesian Poisson Regression Model

Suppose you want to fit a Bayesian Poisson regression model for the frequency of revertant colonies of TA98 Salmonella with density

$$\begin{aligned} \text{SALM}_i &\sim \text{Poisson}(\lambda_i) \\ \log(\lambda_i) &= \mathbf{X}_i \boldsymbol{\beta} \end{aligned} \tag{1}$$

for the $i = 1, \dots, 18$ plates, where $\boldsymbol{\beta}$ represents the regression parameters and \mathbf{X}_i is the vector of covariates written as $\mathbf{X}_i = \{1 \text{ DOSE}_i L_DOSE10_i\}$.

The likelihood function for each of the revertant colonies of salmonella and corresponding covariates is

$$p(\text{SALM}_i | \mathbf{X}_i, \boldsymbol{\beta}) = \text{Poisson}(\lambda_i) \tag{2}$$

where $p(\cdot | \cdot)$ denotes a conditional probability mass function. The Poisson density is evaluated at the specified value of SALM_i with a corresponding mean parameter λ_i . The three parameters, β_1 , β_2 , and β_3 , correspond to an intercept, the toxic effect for dose, and the mutagenic effect for dose, respectively.

Suppose the following prior distributions are placed on the parameters, where $\pi(\cdot)$ indicates a prior distribution:

$$\pi(\beta_1), \pi(\beta_2), \pi(\beta_3) = \text{normal}(0, \sigma^2 = 1000)$$

The diffuse normal prior expresses your lack of knowledge about the regression parameters.

Using Bayes' theorem, the likelihood function and prior distributions determine the posterior distribution of β_1 , β_2 , and β_3 . PROC MCMC obtains samples from the desired posterior distribution. You do not need to specify the exact form of the posterior distribution.

The goodness-of-fit Pearson chi-square statistic χ_P^2 as given in McCullagh and Nelder (1989) is calculated to assess model fit:

$$\chi_P^2 = \sum_{i=1}^n \frac{[\text{SALM}_i - E(\text{SALM}_i)]^2}{V(\text{SALM}_i)} \quad (3)$$

First, let $E(\cdot)$ and $V(\cdot)$ represent an expectation and a variance, respectively, for a Poisson likelihood $E(\text{SALM}_i) = V(\text{SALM}_i) = \lambda_i$, where λ_i is defined in Equation 1. If there is no overdispersion, the Pearson statistic approximately equals the number of observations in the data set minus the number of parameters in the model.

The following SAS statements use the likelihood function and diffuse prior distributions to fit the Bayesian Poisson regression model and calculate the Pearson chi-square statistic:

```
ods graphics on;
proc mcmc data=assay outpost=postout thin=5 seed=1181 nbi=10000 nmc=10000
  propcov=quanew monitor=(_parms_ Pearson) dic;
  parms beta1 2 beta2 0 beta3 0;
  prior beta: ~ normal(0,var=1000);
  lambda = exp(beta1 + beta2*dose + beta3*1_dose10);
  model salm ~ poisson(lambda);
  if plate eq 1 then Pearson = 0;
  Pearson = Pearson + ((salm - lambda)**2/lambda);
run;
ods graphics off;
```

The PROC MCMC statement invokes the procedure and specifies the input data set. The OUTPOST= option specifies the output data set for posterior samples of parameters. The THIN= option controls the thinning of the Markov chain and specifies that one of every 3 samples is kept. Thinning is often used to reduce the correlations among posterior sample draws. The SEED= option specifies a seed for the random number generator (the seed guarantees the reproducibility of the random stream). The NBI= option specifies the number of burn-in iterations. The NMC= option specifies the number of posterior simulation iterations. The PROPCOV=QUANEW option uses the estimated inverse Hessian matrix as the initial proposal covariance matrix. The MONITOR=(_PARMS_ PEARSON) option outputs analysis on the _PARMS_ (which is shorthand for all model parameters) and the Pearson chi-square statistic. The DIC option specifies that the deviance information criterion (DIC) is requested.

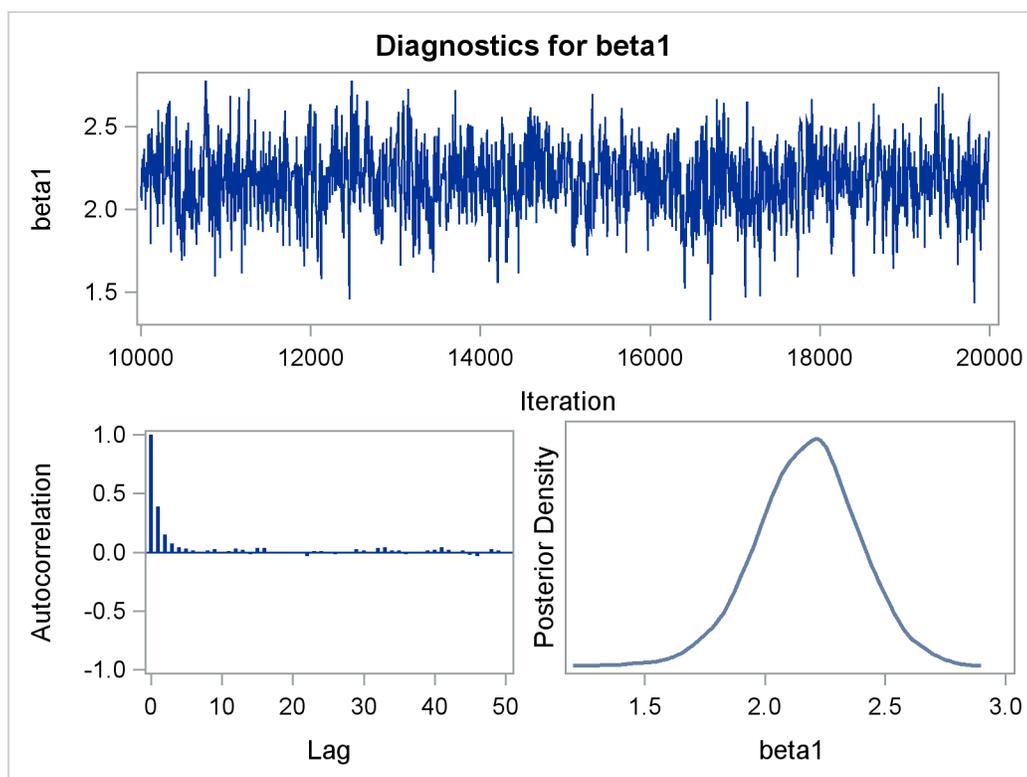
The PARMS statement specifies the parameters in the model and assigns initial values to each of them. The PRIOR statements specify priors for all the parameters. The notation beta: in the PRIOR statement

is shorthand for all variables that start with `beta`. The shorthand notation is not necessary, but it keeps your code succinct. The `LAMBDA` assignment statement calculates λ_i for each observation in the Poisson model, as given in Equation 1. The `MODEL` statement specifies the Poisson function for `SALM`.

The next two lines of statements use `SALM` and `LAMBDA` (which are the expected value and the variance of the Poisson distribution, respectively) to calculate the Pearson chi-square statistic. The `IF` statement sets the value of `PEARSON` to 0 at the top of the data set (that is, when the value of the data set variable `PLATE` is 1). As `PROC MCMC` cycles through the data set at each iteration, the procedure cumulatively adds the Pearson chi-square statistic over each value of `SALM`. By the end of the data set, you obtain the Pearson chi-square statistic, as defined in Equation 3.

Figure 1 displays diagnostic plots to assess whether the Markov chains have converged.

Figure 1 Bayesian Poisson Model Diagnostic Plots for β_1



The trace plot in Figure 1 indicates that the chain appears to have reached a stationary distribution. It also has good mixing and is dense. The autocorrelation plot indicates low autocorrelation and efficient sampling. Finally, the kernel density plot shows a unimodal shape of posterior marginal distribution for β_1 . The remaining diagnostic plots (not shown here) similarly indicate good convergence in the other parameters.

`PROC MCMC` produces formal diagnostic tests by default. They are omitted here since informal checks on the chains, autocorrelation, and posterior density plots display stabilization and convergence.

Figure 2 reports summary and interval statistics for each parameter's posterior distribution. `PROC MCMC` calculates the sampled value of the Pearson chi-square statistic at each iteration and produces its corresponding posterior summary statistics.

Figure 2 Posterior Model Summary of Poisson Regression

The MCMC Procedure						
Posterior Summaries						
Parameter	N	Mean	Standard Deviation	Percentiles		
				25%	50%	75%
beta1	2000	2.1768	0.2110	2.0391	2.1834	2.3142
beta2	2000	-0.00101	0.000233	-0.00117	-0.00101	-0.00085
beta3	2000	0.3184	0.0549	0.2817	0.3191	0.3549
Pearson	2000	49.3974	3.4673	46.8851	48.4551	50.7996

Posterior Intervals				
Parameter	Alpha	Equal-Tail Interval		HPD Interval
beta1	0.050	1.7421	2.5790	1.7467 2.5798
beta2	0.050	-0.00148	-0.00056	-0.00146 -0.00054
beta3	0.050	0.2136	0.4288	0.2169 0.4305
Pearson	0.050	45.5715	59.2921	45.3427 56.2309

With $n = 18$ and three model parameters, the sampled value 49.3974 of the Pearson chi-square statistic is much greater than the desired $18 - 3 = 15$, providing evidence of overdispersion. Although you have already captured the mutagenic and toxic effects, a Bayesian hierarchical Poisson model can also capture the observed excess variation due to the replicate plates.

Figure 3 reports the calculated DIC (Spiegelhalter et al. 2002) for the Bayesian Poisson regression model. The DIC is a model assessment tool and a Bayesian alternative to Akaike's or Bayesian information criterion. The DIC can be applied to non-nested models and models that have data which are not independent and identically distributed. When models with different DIC values are compared, a smaller DIC indicates a better fit to the data set. The DIC for this model is 141.882.

Figure 3 DIC of Poisson Regression

Deviance Information Criterion	
Dbar (posterior mean of deviance)	139.070
Dmean (deviance evaluated at posterior mean)	136.258
pD (effective number of parameters)	2.812
DIC (smaller is better)	141.882

Bayesian Hierarchical Poisson Regression Model

In overdispersed Poisson regression, the parameter estimates do not vary much from the Poisson model, but the estimated variance is inflated. Draper (1996) considers Bayesian hierarchical Poisson regression models

for this type of data with density

$$\begin{aligned}SALM_{ij}|\lambda_{ij} &\sim \text{Poisson}(\lambda_{ij}) \\ \log(\lambda_{ij}) &= \mathbf{X}_i\boldsymbol{\beta} + \gamma_{ij} \\ \gamma_{ij} &\sim \text{normal}(0, \sigma^2)\end{aligned}\tag{4}$$

for the $i = 1, \dots, n$ treatment levels and $j = 1, \dots, J$ replicates. The likelihood function is given in Equation 2, but now there are additional parameters in the model. You add random effects γ'_{ij} s with a normal prior with mean 0 and variance σ^2 .

Suppose the priors are placed on the regression and variance parameters as follows:

$$\begin{aligned}\pi(\beta_1), \pi(\beta_2), \pi(\beta_3) &\sim \text{normal}(0, \sigma^2 = 1000) \\ \pi(\sigma^2) &\sim \text{igamma}(\text{shape} = 0.01, \text{scale} = 0.01)\end{aligned}$$

The mean and variance, $E(\text{SALM}_i)$ and $V(\text{SALM}_i)$, respectively, for the Pearson chi-square statistic in the hierarchical Poisson regression model are calculated using the iterated expectation and variance functions as

$$\begin{aligned}E(\text{SALM}_i) &= E[E(\text{SALM}_i|\Lambda_i)] \\ &= \exp\left\{\mathbf{X}_i\boldsymbol{\beta} + \frac{\sigma^2}{2}\right\} \\ &= \mu_i\end{aligned}\tag{5}$$

$$\begin{aligned}V(\text{SALM}_i) &= E[V(\text{SALM}_i|\Lambda_i)] + V[E(\text{SALM}_i|\Lambda_i)] \\ &= \mu_i + (e^{\sigma^2} - 1)\mu_i^2\end{aligned}\tag{6}$$

The following SAS statements use the likelihood function and prior distributions to fit the Bayesian Poisson hierarchical regression model and calculate the Pearson chi-square statistic.

```
ods graphics on;
proc mcmc data=assay outpost=postout thin=5 seed=248601 nbi=10000 nmc=100000
  monitor=(beta1-beta3 Pearson) dic;
  array llambda[18];

  parms beta1 2 beta2 0.01 beta3 0.5;
  parms s2 1;
  parms (llambda:) 1;

  prior beta: ~ normal(0, var=1000);
  prior s2 ~ igamma(0.01, s=0.01);
  w = beta1 + beta2*dose + beta3*1_dose10;
  if plate eq 1 then lp = 0;
  lp = lp + lpdfnorm(llambda[plate], w, sqrt(s2));
  prior llambda: ~ general(lp);
  lambda = exp(llambda[plate]);
  model salm ~ poisson(lambda);
```

```

mu = exp(w + s2/2);
sigma2 = mu + (exp(s2) - 1)*(mu**2);
if plate eq 1 then Pearson = 0;
Pearson = Pearson + ((salm - mu)**2/(sigma2));
run;
ods graphics off;

```

The ARRAY statement associates a name with a list of variables and constants. The ARRAY statement enables easy referencing of the variables and corresponding values in the procedure. In this example, the ARRAY statement is used to create an array of parameters for the $\log(\lambda_{ij})$ values of all 18 observations.

The first PARMS statement places all regression parameters in a single block and assigns them initial values. The second PARMS statement places the variance parameter in a separate block and assigns it an initial value of 1. The third PARMS statement places all the $\log(\lambda_{ij})$ parameters in a separate block and assigns them an initial value of 1.

The PRIOR statement on the β regression parameters does not change from the Poisson regression model. The second PRIOR statement specifies the prior distribution for σ^2 .

The W assignment statement calculates $\mathbf{X}_i\beta$ for each observation. The IF statement resets the value of LP to be zero at the top of the data set (that is, when the value of the data set variable PLATE is 1). As PROC MCMC cycles through the data set at each iteration, the procedure cumulatively adds the log density of the normally distributed $\log(\lambda_{ij})$ with mean W and variance σ^2 . The LP expression takes the value of the logarithm of the joint prior distribution of the λ_{ij} s. The next PRIOR statement indicates that the $\log(\lambda_{ij})$ is jointly distributed with LP as the value of the logarithm of the density. By the end of the data set, you obtain the sum of the log densities stored in the LP variable. The sum of the log densities is used as the input parameter in the GENERAL function in the PRIOR statement for the $\log(\lambda_{ij})$ parameters. The GENERAL function indicates that you are using a SAS statement to construct a nonstandard density or distribution.

The LAMBDA assignment statement calculates λ_{ij} in the hierarchical Poisson regression model, as given in Equation 4. The MODEL statement specifies the likelihood function for SALM, as given in Equation 2.

The next four lines of statements use SALM and the sampled values of β and σ^2 to calculate the Pearson chi-square statistic according to Equation 3. The moments are evaluated in the MU and SIGMA2 assignment variables according to Equation 5 and 6, respectively.

The diagnostic plot for β_1 is illustrated in Figure 4. It displays the desired convergence, low autocorrelation, and smooth unimodal marginal posterior density for the parameter. The remaining diagnostic plots (not shown here) similarly indicate good convergence in the other parameters.

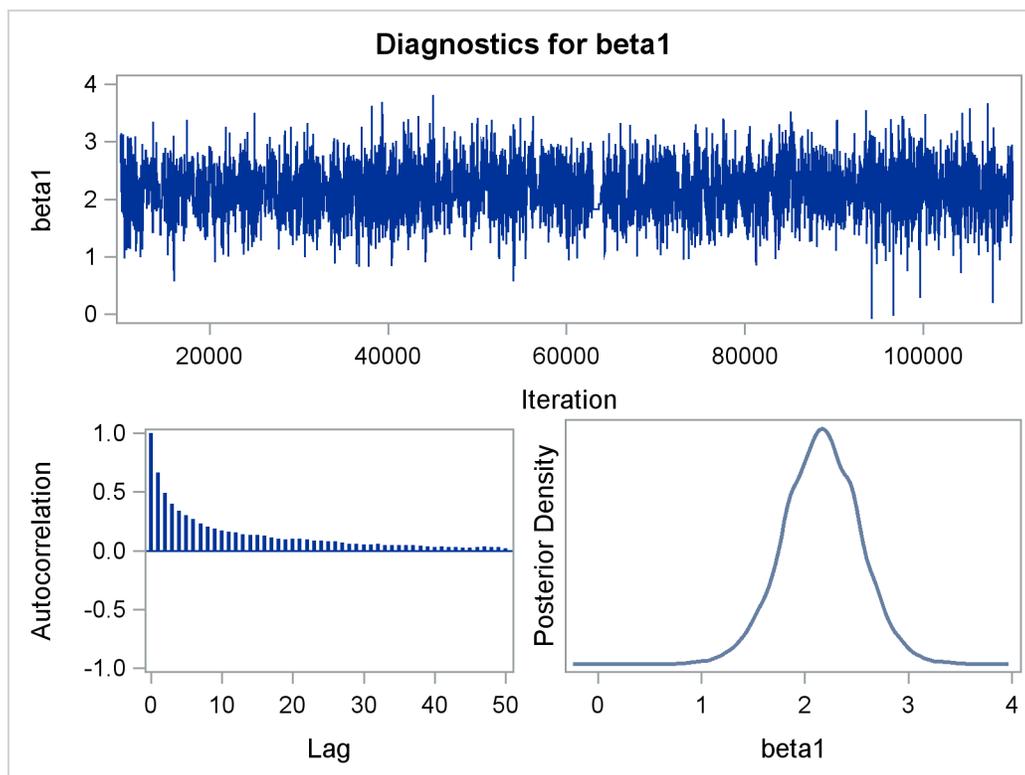
Figure 4 Bayesian Hierarchical Poisson Regression Model Diagnostic Plot for β_1 

Figure 5 reports summary and interval statistics for the parameters and for the Pearson chi-square statistic. The fit statistic in the Bayesian hierarchical Poisson regression model is greatly reduced with a value of 19.2576, which suggests a much better fit compared to its value of 49.43974 in the Poisson regression model. The value of the fit statistic is now much closer to the desired number of observations minus the number of parameters.

Compared to the results in Figure 2, the parameter estimates for intercept and treatment change a small amount, their standard errors are inflated, and the confidence intervals are wider. Thus the treatment effect of quinoline does not decrease with the Bayesian hierarchical Poisson regression model.

Figure 5 Posterior Summary of Bayesian Hierarchical Poisson Regression Model

The MCMC Procedure						
Posterior Summaries						
Parameter	N	Mean	Standard Deviation	Percentiles		
				25%	50%	75%
beta1	20000	2.1589	0.3718	1.9144	2.1628	2.4101
beta2	20000	-0.00100	0.000454	-0.00130	-0.00100	-0.00070
beta3	20000	0.3157	0.1015	0.2497	0.3144	0.3822
Pearson	20000	19.2576	7.3362	13.9696	18.3214	23.4289

Figure 5 continued

Parameter	Alpha	Posterior Intervals			
		Equal-Tail Interval		HPD Interval	
beta1	0.050	1.4122	2.8861	1.4368	2.9074
beta2	0.050	-0.00189	-0.00011	-0.00187	-0.00010
beta3	0.050	0.1144	0.5167	0.1217	0.5218
Pearson	0.050	7.8479	36.6477	6.9561	34.2420

Figure 3 reports the computed DIC for the Bayesian hierarchical Poisson regression model. The DIC for the hierarchical model is 123.652 and is smaller than the DIC for the Poisson regression model shown in Figure 3. A smaller value of DIC suggests a better fit; you see that the hierarchical model provides a better fit to the data.

Figure 6 DIC of Poisson Regression

Deviance Information Criterion	
Dbar (posterior mean of deviance)	110.605
Dmean (deviance evaluated at posterior mean)	97.558
pD (effective number of parameters)	13.047
DIC (smaller is better)	123.652

References

- Breslow, N. E. (1984), "Extra-Poisson Variation in Log-Linear Models," *Applied Statistics*, 33, 38–44.
- Draper, D. (1996), "Discussion of the Paper by Lee and Nelder," *Journal of the Royal Statistical Society, Series B*, 58, 662–663.
- Margolin, B. H., Kaplan, N., and Zeiger, E. (1981), "Statistical Analysis of the Ames Salmonella Microsome Test," *Proceedings of the National Academy of Science*, 76, 3779–3783.
- McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models*, Second Edition, London: Chapman & Hall.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van der Linde, A. (2002), "Bayesian Measures of Model Complexity and Fit," *Journal of the Royal Statistical Society, Series B*, 64(4), 583–616, with discussion.