STAT7630: Bayesian Statistics
Lecture Slides # 12

Bayesian Count Regression Models
Chapter 12 (Poisson & Negative Binomial Regression)

Elvan Ceyhan
Department of Mathematics & Statistics
Auburn University

Fall 2024,
Updated: November, 2024

Outline

Count Regression Models

- Poisson Regression Model
- Negative Binomial Regression Model

Regression for Count Data

- We consider a regression model where the response variable Y takes on count values, such as $0, 1, 2, 3, \ldots$
- When the count values in the dataset are relatively large, the response (conditional on predictors) may be approximately normally distributed, allowing the application of Normal-response models from Chapter 9.
- But, if the counts $Y_1, Y_2, \ldots, Y_n$ are small to moderate, it is inappropriate to treat the responses as normally distributed. Small counts are highly discrete and often exhibit skewness, requiring alternative modeling approaches.

Outline

A Better Regression Model for Count Responses

- A suitable regression model for count-valued responses is the **Poisson regression model**, which assumes:

$$Y_i \mid \lambda_i \overset{\text{ind}}{\sim} \text{Pois}(\lambda_i)$$

- This model expresses the conditional mean for the i-th individual as:

$$\mathbf{E}(Y_i \mid \lambda_i) = \lambda_i$$

Setup of Poisson Regression Model

- Recall that the Poisson mean must be strictly positive.
- To ensure $\mathbf{E}(Y_i \mid \lambda_i) = \lambda_i$ remains positive, we model $\log(\lambda_i)$ as a linear combination of predictor variables:

$$\log(\lambda_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{k-1} X_{i,k-1}$$

- Consequently, the model for the mean response given the predictors is:

$$\mathbf{E}(Y_i \mid \mathbf{X}) = \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_{k-1} X_{i,k-1})$$

- This model was introduced in Chapter 6 with the sparrow offspring data.

Example of Poisson Regression Model

- Consider the dataset where the individuals are high school students, sourced from the UCLA Advanced Research and Computing website.
- The response variable is the **number of awards** a student has received for academic performance.
- This count-valued response takes values $0, 1, 2, 3, \ldots$, with most values in the dataset being relatively small.
- A key predictor variable (X_1) is the student's **math exam score**.

Example of Poisson Regression Model (Continued)

- Additionally, we include a **categorical predictor** that identifies the track the student is on, with three categories: *General*, *Academic*, and *Vocational*.
- This categorical variable is coded using two dummy variables:

$$X_2 = \begin{cases} 1 & \text{if student is on academic track} \\ 0 & \text{otherwise} \end{cases}$$

$$X_3 = \begin{cases} 1 & \text{if student is on vocational track} \\ 0 & \text{otherwise} \end{cases}$$

- The *general* track serves as the baseline category, and the coefficients of X_2 and X_3 are interpreted relative to this baseline.

Equation for the Poisson Regression Model

- The model equation is given by:

$$\mathbf{E}(Y_i \mid \mathbf{X}) = \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3})$$

- This expresses the expected counts as a **nonlinear function** of the predictors, distinguishing it from the normal regression model.
- With Poisson data, note that the **variance of the response equals the mean**, implying that as the mean response increases, the variability of the responses around the regression curve also increases.
- This contrasts with the normal regression model, where constant variance of $Y \mid \mathbf{X}$ is assumed.

Priors in the Poisson Regression Model

- In regression models with non-normal responses, such as Poisson regression, conjugate priors for the regression coefficients (the β's) are typically unavailable.
- However, we can still assign **independent normal priors** to each β_j, $j = 0, 1, 2, \ldots, k-1$, similar to the approach used in the sparrow data example.
- If we have a prior belief regarding the direction of a coefficient, we may set the prior mean to a positive or negative value accordingly; otherwise, a mean of 0 may be appropriate.
- Specifying a **large prior variance** reflects less certainty about our prior knowledge, allowing the data to have greater influence.

Fitting the Poisson Regression Model

- Since we do not use conjugate priors for the β's, we rely on **MCMC methods** to sample from the posterior distribution, specifically employing the **Metropolis-Hastings** algorithm.
- This can be implemented in `R` as done previously for the sparrow data, or alternatively, we can use the `stan_glm` function from the `rstanarm` package to automate the Metropolis-Hastings procedure.
- It is still crucial to perform our usual **MCMC diagnostics** and, if necessary, apply **remedial actions** to ensure model reliability.
- Refer to the provided `R` examples for details on fitting the model.

Interpretations of Estimated Parameters

- The posterior estimate of β_1 is approximately 0.07 (this value may vary slightly based on the choice of priors and the specific MCMC run).
- For a fixed level of track, the expected number of awards earned increases by a factor of $e^{0.07} = 1.07$ for each one-point increase in math test score.
- The posterior estimate of β_2 is around 1.03 (again, this value may vary depending on the priors and MCMC details).
- For students on the academic track, the expected number of awards earned is $e^{1.03} = 2.8$ times that of students on the general track, holding math test score constant.

Checking Model Fit

- Model fit can be assessed using metrics such as the **Mean Absolute Error (MAE)**.
- The `bayesrules` package provides convenient functions for calculating MAE and other goodness-of-fit measures for both **in-sample** and **out-of-sample** (cross-validation) prediction performance.
- Model fit measures indicate that the Poisson model is a good fit for the awards data.
- It is often beneficial to fit multiple models with different sets of predictor variables and compare them using model-fit criteria to identify the most suitable model.

Outline

Count Regression Models

- Poisson Regression Model
- Negative Binomial Regression Model

Count Regression for Overdispersed Data

- In some cases, the response variable is a count, but the **Poisson regression model** may not provide an adequate fit.
- **Example**: The `pulse` dataset in the `bayesrules` package includes various variables measured on over 900 individuals. We focus here on three specific variables:
  - Y: Number of books read in the past year.
  - X_1: Age in years.
  - X_2: Categorical variable where $X_2 = 1$ if the person would prefer to be "wise but unhappy," and $X_2 = 0$ if they would prefer to be "happy but unwise."

Problems with Poisson Regression for Overdispersed Data

- An initial attempt at fitting a **Poisson regression** model of Y on X_1 and X_2 can be conducted.
- However, **posterior predictive analysis** reveals that this model provides a poor fit, as the posterior predictive distribution does not align with the observed data.
- Summary calculations indicate that the **variance is much greater than the mean** for this dataset.
- The Poisson regression model assumes that, given a set of predictor values, the mean of Y should equal its variance.
- For the "books" data, within subsets with similar predictor values, the **variance significantly exceeds the mean**, indicating overdispersion.

Overdispersion in Data

- When the variance of a count variable exceeds the mean, this is commonly referred to as **overdispersion**.
- The textbook provides a broader definition related to model fit: A random variable Y is **overdispersed** if the observed variability in Y exceeds the variability anticipated by the assumed probability model of Y.

Using the Negative Binomial to Account for Overdispersion

- The **Negative Binomial** probability model is a common alternative to the Poisson model when Y is overdispersed.
- Like the Poisson, the Negative Binomial distribution is well-suited for count data, as it is defined over $y = 0, 1, 2, \ldots$, but it relaxes the assumption that $\mathbf{E}(Y) = \text{Var}(Y)$.
- For the Negative Binomial distribution, it holds that $\mathbf{E}(Y) < \text{Var}(Y)$, accommodating scenarios where variance exceeds the mean.

Form of the Negative Binomial Probability Function

- The Negative Binomial distribution has several parametrizations. One common form uses μ for the mean and r as the "reciprocal dispersion" parameter:

$$f(y \mid \mu, r) = \binom{y+r-1}{y} \left(\frac{r}{\mu+r}\right)^r \left(\frac{\mu}{\mu+r}\right)^y, \text{ for } y = 0, 1, 2, \ldots$$

- Under this parametrization:

$$\mathbf{E}(Y \mid \mu, r) = \mu \quad \text{and} \quad \text{Var}(Y \mid \mu, r) = \mu + \frac{\mu^2}{r}.$$

- When r is large, $\mathbf{E}(Y) \approx \text{Var}(Y)$, closely resembling the Poisson distribution. For small r, however, **Var**(Y) can greatly exceed $\mathbf{E}(Y)$, allowing for overdispersion.

Fitting a Negative Binomial Regression Model

- The **negative binomial regression model** can be easily fitted using the `stan_glm` function in the `rstanarm` package by specifying `family = neg_binomial_2`.
- Similar to Poisson regression, we model the expected counts as:

$$\mathbf{E}(Y_i \mid \mathbf{X}) = \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}),$$

  ensuring nonnegative expected counts.
- Prior distributions for the coefficients are set up similarly to the Poisson regression example.
- Model fit diagnostics, such as plots and numerical statistics, are obtained in the same manner as with Poisson regression.
- **Posterior predictions** of the response variable can be generated for one or more individuals.

Substantive Conclusions from the Pulse Regression Analysis

- Refer to the `R` examples for the Negative Binomial regression analysis of the "books" dataset.
- Note that the exact values of the estimated β's will vary slightly with each MCMC run.
- **Age** does not appear to be a significant predictor of the number of books read.
- The estimated coefficient for the *wise vs. unwise* preference variable is approximately 0.265, suggesting that individuals who prefer to be "wise but unhappy" read **1.3 times more books** than those who prefer to be "happy but unwise" (holding age constant), as $e^{0.265} = 1.3$.
- The **95% credible interval** for β_2 lies entirely above 0, indicating a strong positive association between a preference for wisdom over happiness and the number of books read.

A Quick Model Comparison

- We may consider alternative models, such as:
  - A model without age as a predictor.
  - A model with age, *wise vs. unwise* preference, and their interaction term.
- The `loo` function can be used to calculate the **Expected Log Predictive Density (ELPD)** criterion for each model.
- According to the code provided on Canvas, the model including both predictors and their interaction achieves the highest ELPD, suggesting it as the best among these three models (though the ELPD values are quite close).