STAT 535: Chapter 12: Bayesian Count Regression Models

David B. Hitchcock
E-Mail: hitchcock@stat.sc.edu

Spring 2022

Regression for Count Data

- We now consider the **regression model** in which a response variable Y takes on count values, such as $0, 1, 2, 3, \ldots$
- If the count values in the data set tend to be quite large, we might reasonably assume the response (given values of the predictors) to be approximately normal and use the methods of Chapter 9 for Normal-response models.
- However, if the sizes of the counts $Y_1, Y_2, \ldots, Y_n$ in our data set are small to moderate, then it doesn't make sense to treat the responses as normal (they would be highly discrete and quite possibly skewed).

A Better Regression Model for Count Responses

- A natural regression model for count-valued responses is the Poisson regression model, which assumes

$$Y_i|\lambda_i \overset{ind}{\sim} Pois\,(\lambda_i)$$

and models the conditional mean of the i-th individual as

$$E(Y_i|\lambda_i) = \lambda_i$$

Setup of Poisson Regression Model

- Note the the Poisson mean must be greater than 0.
- So to force $E(Y_i|\lambda_i) = \lambda_i$ to be positive, we actually relate $\log(\lambda_i)$ to the predictor variables:

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

- So the model for the mean response **given** the predictors is

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

- We saw this model in Chapter 6 with the sparrow offspring data.

Example of Poisson Regression Model

- Consider a data set in which the individuals are high school students (taken from the UCLA Advanced Research and Computing website).
- The response variable is the number of awards the student won for academic performance.
- This response takes values $0, 1, 2, 3 \ldots$ (most values in the data set are relatively small).
- One predictor variable (X_1) is the student's score on a math exam.

Example of Poisson Regression Model

- We also have a categorical predictor with three categories, which track the student is on, which could be "General", "Academic", or "Vocational".
- We code this 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" category is the **baseline** category, and the coefficients of the two dummy variables are interpreted relative to this category.

Equation for this Poisson Regression Model

- The model equation for this is:

$$E(Y_i|\boldsymbol{x}) = \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3})$$

- So the equation for the expected counts is a **nonlinear** function of the predictors (different from the normal regression model).
- Note that with Poisson data, the variance of the response equals the mean of the response.
- So as the mean response increases, the variability of the responses around the regression curve should increase as the predicted counts get larger (also different from the normal regression model where **constant variance** of $Y|\boldsymbol{x}$ is one of the key assumptions).

Priors in the Poisson Regression Model

- With regression models in which the response is non-normal, we generally do not have conjugate priors for the regression coefficients (the β's).
- We can still specify independent normal priors on each β_j, $j = 0, 1, 2, \ldots, k-1$, as we did in the sparrow data example.
- If we have a prior belief about the direction of the coefficient, we could set the prior mean to be positive or negative (otherwise we could set it to 0).
- A large prior variance would indicate less certainty about our prior knowledge.

Fitting the Poisson Regression Model

- Since do not use conjugate priors for the β's, we sample from the posterior using MCMC methods, specifically the Metropolis-Hastings method.
- We could code this using R as we did for the sparrow data, or we use use the `stan_glm` function in the `rstanarm` package to do the Metropolis-Hastings automatically.
- We would still want to do our usual MCMC diagnostics and (if necessary) remedial actions.
- See R examples for the fitting of the model.

Interpretations of Estimated Parameters

- The posterior estimate of β_1 is (around) 0.07 (it will change slightly depending on the exact type of priors chosen and even slightly based on the MCMC run).
- For a fixed level of track, the expected number of awards earned increased 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 (it will change slightly depending on the exact type of priors chosen and even slightly based on the MCMC run).
- The expected number of awards earned for a student on the academic track is $e^{1.03} = 2.8$ times the expected number of awards earned for a student on the general track (given the same level of math test score).

Checking Model Fit

- We can again check model fit using tools like the Mean Absolute Error (MAE).
- The `bayesrules` package offers nice built-in functions to calculate the MAE (and other measures of goodness-of-fit) for both the in-sample prediction performance and the out-of-sample (cross-validation) prediction performance.
- The model fit measures show a good fit of the Poisson model for the awards data.
- Often it is most useful to fit multiple models (i.e., with different sets of predictor variables) and to compare the models using the model-fit criteria.

Count Regression for Overdispersed Data

- Sometimes we have data is which the response variable is a count, but the Poisson regression model does not provide a good fit.
- Example: The `pulse` data frame in the `bayesrules` package has numerous variables measured on over 900 individuals (we will focus on three variables here):
- Y: number of books read in past year
- X_1: age in years
- X_2: Categorical: $X_2 = 1$ if person would rather be "wise but unhappy", $X_2 = 0$ if person would rather be "happy but unwise"

Problems with Poisson Regression with Overdispersed Data

- We might initially try a Poisson regression of Y on X_1 and X_2.
- We can fit this, but the posterior predictive analysis shows that the model is a poor fit (the posterior predictive distribution does not match the actual data at all).
- Some summary calculations show that the variance is much greater than the mean for this data set.
- The Poisson regression model assumes that given a set of predictor values, the true mean of Y should equal the variance of Y.
- For the "books" data, we see that within subsets of the data having similar predictor values, the variance greatly exceeds the mean.

Overdispersion in Data

- When the variance of a count variable exceeds the mean, we (loosely) say there is **overdispersion**.
- The book gives a general definition pertaining to lack of model fit: A random variable Y is overdispersed if the observed variability in Y exceeds the variability expected 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 when Y is overdispersed.
- The Negative Binomial distribution is also a good model for count data, since its support is $y = 0, 1, 2, \ldots$, but it does not assume $E(Y) = var(Y)$.
- In fact, for the negative binomial, $E(Y) < var(Y)$.

Form of the Negative Binomial Probability Function

- The probability function for the negative binomial has a couple of different parametrizations. One version with a parameter μ for the mean and another parameter r that is the "reciprocal dispersion" is:

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

- Here,

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

- So for r large, $E(Y) \approx var(Y)$, but for r small, $var(Y)$ is allowed to be much larger than $E(Y)$.

Fitting a Negative Binomial Regression Model

- The negative binomial regression model can be fit most easily using the `stan_glm` function in the `rstanarm` package, by specifying `family=neg_binomial_2`
- As with the Poisson regression, we model

$$E\left(Y_i|\boldsymbol{x}\right) = \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2})$$

  so that the expected counts produced by the model will be nonnegative.
- We set up the priors similarly as in the Poisson regression example.
- The plots and numerical statistics to check model fit can be obtained similarly to how they were with Poisson regression.
- Posterior predictions of the response variable for one or more individuals can also be obtained.

Substantive Conclusions with the `pulse` Regression Analysis

- See the R examples for the Negative Binomial regression analysis of the "books" data set.
- Again, the exact values of the estimated β's will vary based on the MCMC run.
- Age may not be an important predictor of number of books read.
- The estimated coefficient of `wise_unwise` is around 0.265, so the expected number of books read is 1.3 times more for people who prefer to be wise but unhappy than for people who prefer to be happy but unwise (holding age constant), since $e^{0.265} = 1.3$.
- The 95% credible interval for β_2 is completely above 0, so it is highly likely that preference for wisdom over happiness is positively associated with number of books read.

A Quick Model Comparison

- We might consider a couple of other models: Maybe a model without age as a predictor, or maybe a model with age, "wise_unwise", **and** their interaction.
- The `loo` function will calculate the ELPD criterion for each model.
- The code on the course website shows that the model with both predictors and their interaction has the highest ELPD and could be considered the best of these three models (although the ELPD values are very close).