Poisson Regression

Ryan Godwin

ECON 7010 - University of Manitoba

Abstract. These lecture notes introduce Maximum Likelihood Estimation (MLE) of a Poisson regression model.

1 Motivating the Poisson Regression Model

There are many instances in Econometrics where the variable that we want to explain is a count variable, i.e. $y = 0, 1, 2, ...$ Examples of some cases are:

· calls at a call-centre
· number of customers
· doctor visits
· bank failures
· insurance claims
· patents

1.1 Why not use OLS?

OLS will not work very well here. The dependent variable y is not continuous, and $y \geq 0$. OLS ignores this information (so it will be inefficient). In addition, OLS will not be able to provide very useful predictions (e.g. the probability of a certain count occuring).

Instead, a count data model should be used. The Poisson model is a possibility, however, it rarely describes data well in real-world applications. The Poisson model is likely so prevalent due to its relationship to more complicated count data models, rather than for its practical usefulness.

2 The Probability Mass Function (PMF) for a Poisson Distributed Random Variable

If y follows a Poisson distribution, then:

$$P(y = y_i \mid \lambda) = \frac{\lambda^{y_i}}{e^{\lambda} y_i!} \quad ; \quad y = 0, 1, 2, ... \quad ; \quad \lambda > 0. \tag{1}$$

The mean and variance of this distribution is λ (this equi-dispersion property proves too restrictive for most applications). Provided an estimate for λ, this PMF may be used to infer probabilities of events (e.g. $P(y \geq 6)$).

2.1 Maximum Likelihood Estimation for the Poisson Distribution

Assuming independence of the y_i's, the joint log-likelihood is:

$$l(\lambda \mid y_i) = \sum_{i=1}^{n} (y_i \log \lambda - \lambda - \log y_i!) \tag{2}$$

Or:

$$l(\lambda \mid y_i) = \sum_{i=1}^{n} y_i \log \lambda - n\lambda - \sum_{i=1}^{n} \log y_i! \tag{3}$$

Taking the derivative of (3) with respect to λ we get:

$$\frac{\partial l}{\partial \lambda} = \frac{\sum_{i=1}^{n} y_i}{\lambda} - n \tag{4}$$

Setting (4) equal to zero for the first order condition, and solving for λ, yields:

$$\tilde{\lambda} = \bar{y} \tag{5}$$

In order to verify that (5) is the MLE for λ we take the second derivative of (3) with respect to λ:

$$\frac{\partial^2 l}{\partial \lambda^2} = -\frac{\sum_{i=1}^{n} y_i}{\lambda^2} \tag{6}$$

Since (6) is negative, the log-likelihood is concave and (5) solves for the global maximum. Note that (6) is the Hessian matrix, H, however, since the Poisson distribution has only one parameter (λ) the Hessian is scalar.

2.2 The Variance of $\tilde{\lambda}$

The variance of an MLE may be found by taking the inverse of the negative of the expected Hessian matrix (the matrix of second order derivatives and cross derivatives of the log-likelihood). In the present context:

$$var(\tilde{\lambda}) = [-E(H)]^{-1} = \frac{\lambda^2}{\sum E(y_i)} = \frac{\lambda^2}{n\lambda} = \frac{\lambda}{n} \tag{7}$$

Using the *invariance* property of MLEs, an MLE for the variance of $\tilde{\lambda}$ is found by substituting $\tilde{\lambda}$ into (7):

$$\widetilde{var(\tilde{\lambda})} = \frac{\tilde{\lambda}}{n} \tag{8}$$

2.3 Exercise - Flying-bomb Hits on London During WWII

The following data is on number of bomb hits in south London during WWII (Feller, 1957). The city was divided into 576 areas, and the number of areas hit exactly y times was counted. What does the assumption of independence of the data imply here?

Table 1. Observed and Expected Counts of Bomb Hits

Hits	0	1	2	3	4	5+
Observed	229	211	93	35	7	1
Expected	228	211	98	30	7	1

What is $\tilde{\lambda}$? What is $\widetilde{var}(\tilde{\lambda})$? How are the "Expected" values in the table calculated?

2.4 Specification Testing for the Poisson Distribution

Goodness-of-fit tests for the Poisson distribution can be achieved by comparing the observed and expected counts. For example, consider the following statistic based on the Pearson statistic:

$$P = \sum_{i=1}^{n} \frac{\left(y_i - \tilde{\lambda}_i\right)^2}{\tilde{\lambda}_i} \tag{9}$$

If the Poisson model is specified correctly, then $E[P] = n$ (or $n - 1$ for a degrees-of-freedom correction). There are several other goodness-of-fit test statistics available based on this idea, and most follow a chi-square distribution. Rejection of the null hypothesis does not indicate the appropriate distribution, only that the Poisson model is misspecified (indicating the loss of some or all of MLs asymptotic properties).

The main limitation of the Poisson distribution in applications is it's property of equidispersion. Most count data are overdispersed, i.e. the variance exceeds the mean. Hence, there are several tests based on this restriction.

In many cases, there are other candidate distributions that the data may follow (e.g. negative binomial or zero-inflated Poisson), that *nest* the Poisson distribution. Wald, likelihood ratio, and score testing procedures may be used.

3 The Poisson Regression Model

One reason for overdispersion is unobserved heterogeneity. Heterogeneity can become observed by including explanatory variables (in applications this seldom accounts for overdispersion). A more important reason for including explanatory variables is to estimate how they are related to y.

In addition to the distribution assumption (1), and independence between observations, we will now assume:

$$E[y_i \mid x_i] = \lambda_i = \exp\left(x_i'\beta\right) \tag{10}$$

That is, the mean of y is conditional on x and can vary by individual or observation, etc. The specific form of the *link function* is somewhat arbitrary, but ensures that $\lambda_i > 0$. For example, consider the number of *doctor visits*. An individual's doctor visits may depend on age, underlying health conditions, genetics, and insurance status. The economist may be interested in moral hazard or adverse selection.

By substituting (10) into (1), multiplying across all observations (by independence of the data), and taking logs, we have the following joint log-likelihood function:

$$l(\beta \mid y_i, X_i) = \sum_{i=1}^{n} y_i X_i'\beta - \exp X_i'\beta - \log y_i! \tag{11}$$

The derivative of (11) with respect to the vector, β, is:

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^{n} (y_i - \exp X_i'\beta) X_i \tag{12}$$

Setting (12) equal to zero does not admit a closed form solution for β. Hence, numerical methods, such as Newton-Raphson, must be used to obtaining the ML estimate. Note that asymptotic standard errors for the βs can again be estimated by inverting the expected Hessian matrix.

3.1 Interpreting the βs

Due to the exponent in the link function, the βs do not have as simple of an interpretation as they do in OLS. For example, a one unit change in the j^{th} regressor leads to a *proportionate change* in $E[y_i \mid x_i]$ of β_j. Note that while standard errors for the βs can be estimated by inverting the Hessian, estimating standard errors of the semi-elasticities would require something called the *delta method*.

3.2 Illustrative Application - Bad Health

The data is from the German Health Survey, amended in Hilbe and Greene (2008). The variables are *numvisits* - number of visits to doctor during 1998, *badh* - equal to 1 if patient claims to be in bad health, *age* - age of patient. Enter the following code into R:

```
library(COUNT)
```

```
data(badhealth))
glmbadp <- glm(numvisit   badh + age, family=poisson, data=badhealth)
summary(glmbadp)
```

References

1. Greene, W. H. (2003). Econometric analysis. Pearson Education India.
2. Hilbe, Joseph M (2011), Negative Binomial Regression, Cambridge University Press Hilbe, J. and W. Greene (2008). Count Response Regression Models, in ed. C.R. Rao, J.P Miller, and D.C. Rao, Epidemiology and Medical Statistics, Elsevier Handbook of Statistics Series. London, UK: Elsevier.
3. Winkelmann, R. (2008). Econometric analysis of count data. Springer Science Business Media.