Econometrics I - Application of MLE: count data

Ryan T. Godwin

University of Manitoba

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

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

LS can be inconsistent when the the dependent variable is a count, and does not provide a "fitted" model that is useful. Instead of LS, we could use a "count data" model. The most basic of count data models is the Poisson regression model.

Poisson distribution

When estimating a model by maximum likelihood, we first need to pick a suitable distribution for describing the y variable. We start with the Poisson distribution, which describes the number of events that will happen over some fixed interval (usually an interval of time). If the events are assumed to be independent from one another, and the times between events are exponentially distributed, then we get a Poisson distribution:

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

The mean and variance of this distribution is λ.[1] If we had an estimate for the mean λ then we would know everything about the counting process; e.g. the probability of more than 4 customers showing up in a day.

[1] This equi-dispersion property proves too restrictive for most applications.

Maximum likelihood estimation of the Poisson distribution

Suppose we have a sample of data, $\boldsymbol{y}$. The MLE will tell us the value of λ that is most likely to have "generated" the sample that we have observed.
The first step is to determine the *joint log likelihood*. Assuming independence of the y_is, the joint log-likelihood is:

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

or:

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

Now, taking the derivative of 2 with respect to λ we get:

$$\frac{\partial l}{\partial \lambda} = \frac{\sum_{i=1}^{n} y_i}{\lambda} - n \tag{3}$$

Setting 3 equal to zero for the FOC, and solving for λ, yields:

$$\tilde{\lambda} = \bar{y} \tag{4}$$

In order to verify that 4 is the MLE for λ we take the second derivative of 3:

$$\frac{\partial^2 l}{\partial \lambda^2} = -\frac{\sum_{i=1}^{n} y_i}{\lambda^2} \tag{5}$$

Since 5 is negative, the log-likelihood is concave and 4 solves for the global maximum. Note that 5 is the (scalar) Hessian matrix, H.

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:

$$\text{var}(\tilde{\lambda}) = [-E(H)]^{-1} = \frac{\lambda^2}{\sum E(y_i)} = \frac{\lambda^2}{n\lambda} = \frac{\lambda}{n} \tag{6}$$

Using the invariance property of MLEs, an MLE for the variance of $\tilde{\lambda}$ is found by substituting $\tilde{\lambda}$ into 6:

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

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?

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 $\widetilde{\text{var}(\tilde{\lambda})}$? How are the "Expected" values in the table calculated? How would you test the hypothesis that the expected number of bomb hits for an area is less than 1?

Specification testing for the Poisson distribution

Originally, we had to make the assumption that the data were Poisson distributed. The distributional assumption is an important first step, and one that can be tested. Similar to R^2 in the linear regression model, we can examine how well the model estimated by MLE "fits" the data. If the fit is good, the distributional assumption is generally considered to be good.

Goodness-of-fit tests for the Poisson distribution can be achieved by comparing the observed and expected counts. For example, consider the following χ^2 statistic:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

where O_i are the observed frequencies and E_i are the expected frequencies (obtained by plugging the MLE into the probability function). If the Poisson model is correctly specified, then the expected value of the above statistic is 0. If the above chi-square statistic becomes too large, we may reject the null that Poisson is the correct distribution, however, rejection does not indicate the appropriate distribution, only that the Poisson model is misspecified (and we lose 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.

The Poisson regression model

Once we include explanatory variables, or “regressors” into the model, it becomes a “regression model”. How do we accomplish this? A common modelling strategy is to “link” the mean of the distribution to regressors. We simply write an equation where the parameters of the distribution are determined by explanatory variables. Usually, one of the parameters of the distribution is the *mean*, and this is where we form the link. This allows, for example, different people with different characteristics to have different means. For Poisson, the link is usually:

$$E\left[y_i \mid X_i\right] = \lambda_i = \exp\left(X_i' \boldsymbol{\beta}\right) \tag{7}$$

That is, the mean of $\boldsymbol{y}$ is conditional on X and can vary by individual or observation. 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 7 into 1, multiplying across all observations (by independence of the data), and taking logs, we have the following joint log-likelihood function:

$$l\left(\boldsymbol{\beta} \mid y_i, X_i\right) = \sum_{i=1}^{n} y_i X_i' \boldsymbol{\beta} - \exp X_i' \boldsymbol{\beta} - \log y_i! \tag{8}$$

The derivative of 8 with respect to the vector $\boldsymbol{\beta}$ is:

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^{n} \left(y_i - \exp X_i'\beta\right) X_i \tag{9}$$

Setting 9 equal to zero does not admit a closed form solution for $\boldsymbol{\beta}$! Hence, numerical methods, such as Newton-Raphson, must be used to obtaining the ML estimate. Note that asymptotic standard errors for the $\boldsymbol{\beta}$ can again be estimated by inverting the expected Hessian matrix.

Interpreting the $\boldsymbol{\beta}$

Due to the exponent in the link function, the β do not have as simple of an interpretation as they do in LS (this is one of the reason people hesitate to depart from LS). For example, a one unit change in the jth regressor leads to a *proportionate* change in $\mathrm{E}[y_i \mid X_i]$. Note that while standard errors for the β can be estimated by inverting the Hessian, estimating standard errors of the semi-elasticities would require the *delta method*.