# POLS 7050

Spring 2008

April 9, 2007

**Models for Event Count Data, I**

## An Introduction to Count Data

*Event count models* are models where the dependent variable is a count of events. So, we're considering a variable $Y_i$ where $Y \in 0, 1, 2, ....$ Event count variables are thus *nonnegative integers* - bounded at zero below, unbounded above.

Note a couple things that *aren't* count data:

- **Ordinal Data**
  - Items such as Likert scales may look like event counts, but they aren't...
  - Use ordered logit/probit instead.
- **Grouped binary data**
  - Data which are the number of "successess" (or failures) out of some known number of binary trials.
  - E.g.: the number of successful veto overrides in each Congress, or the number of failed coup attempts in a given nation.
  - Grouped binary data can be expressed as event counts, but are not event counts.
    - One should (generally) not use event count models for grouped binary data.
    - The exception is when there are relatively few successes, relative to the possible number of trials (see below).

Event counts have a few interesting properties that can make them hard to deal with. In particulary, they are:

- discrete (meaning they can only take on integer values), and (as noted above)
- strictly nonnegative.

both of these things can make using OLS on event counts a bad idea; King (1988) tells how and why, specifically. In general, using OLS to model event counts yields estimates which are:

- *inaccurate* (for example, OLS can yield negative predicted counts), and
- *inefficient* (because they fail to account for the heteroskedastic nature of event counts- more on this in a bit...)

# Event Count Data: Motivations

## Count Data and Poisson Processes

A good place to start thinking about event count data is with an abstract model of event counts. Suppose we are interested in studying events, and that those events occur over time. We might consider the *constant rate* at which events occur; call this rate $\lambda$. It's useful to think of $\lambda$ as the expected number of events in any particular time "period" of length $h$. Imagine further that the events in question are *independent*; that is, the occurrence of one event has no bearing on the probability that another will occur.

If the process that gives rise to the events in questions (what we'll call the event process) conforms to these assumptions, then it's pretty straightforward to show that as the length of the interval $h \to 0$,

- The probability of an event occurring in the interval $(t, t+h] = \lambda h$
- The probability of no event occurring in the interval $(t, t+h] = 1 - \lambda h$

Such a variable is what is known as a *Poisson process*: events occur independently with a constant probability equal to $\lambda$ times the length of the interval (that is, $\lambda h$).

Next, consider our outcome variable $Y_t$ as the number of events that have occurred in the interval $t$ of length $h$. For such a process, the probability that the number of events occurring in $(t, t+h]$ is equal to some value $y \in \{0, 1, 2, 3, ...\}$ is:

$$\Pr(Y_t = y) = \frac{\exp(-\lambda h)\lambda h^y}{y!} \tag{1}$$

If all the intervals are of the same length (and equal to 1), this reduces to:

$$\Pr(Y_t = y) = \frac{\exp(-\lambda)\lambda^y}{y!} \tag{2}$$

This is the way we typically see the *Poisson distribution* written. By this logic, the Poisson distribution is the limiting distribution for the number of independent (Poisson) events occurring in some fixed period of length $h$ (for Eq. 1) or 1 (for Eq. 2).

The assumptions underlying the event process – *constant arrival rates*, and *independence across events* – are key to deriving the Poisson distribution in this way. If we relax these assumptions (as we'll do next week), the resulting distribution(s) are not Poisson. We'll return to this point later in the course.

**Other Motivations**

There are many other ways of motivating the Poisson distribution. For example, we can think of the Poisson as the distribution for counts of "rare events." Consider a large number of Bernoulli (binary) trials, where the probability of an event in any one trial is small. In such a situation, the total number of events observed will follow a Poisson distribution.

Formally, for $M$ independent Bernoulli trials with (sufficiently small) probability of success $\pi$ and where $M\pi \equiv \lambda > 0$,[1] the probability of observing exactly $y$ total "successes" as the number of trials grows without limit is:

$$\begin{aligned}
\Pr(Y_i = y) &= \lim_{M\to\infty} \left[ \binom{M}{y} \left(\frac{\lambda}{M}\right)^y \left(1 - \frac{\lambda}{M}\right)^{M-y} \right] \\
&= \frac{\lambda^y e^{-\lambda}}{y!}
\end{aligned}$$

This was actually the original derivation of the Poisson distribution (by – who else? – Simeon-Denis Poisson, back in 1837). Cameron and Trivedi (1998, Ch. 1) call this the "Law of Rare Events" motivation for the Poisson distribution; see their book for other ways to motivate the Poisson.

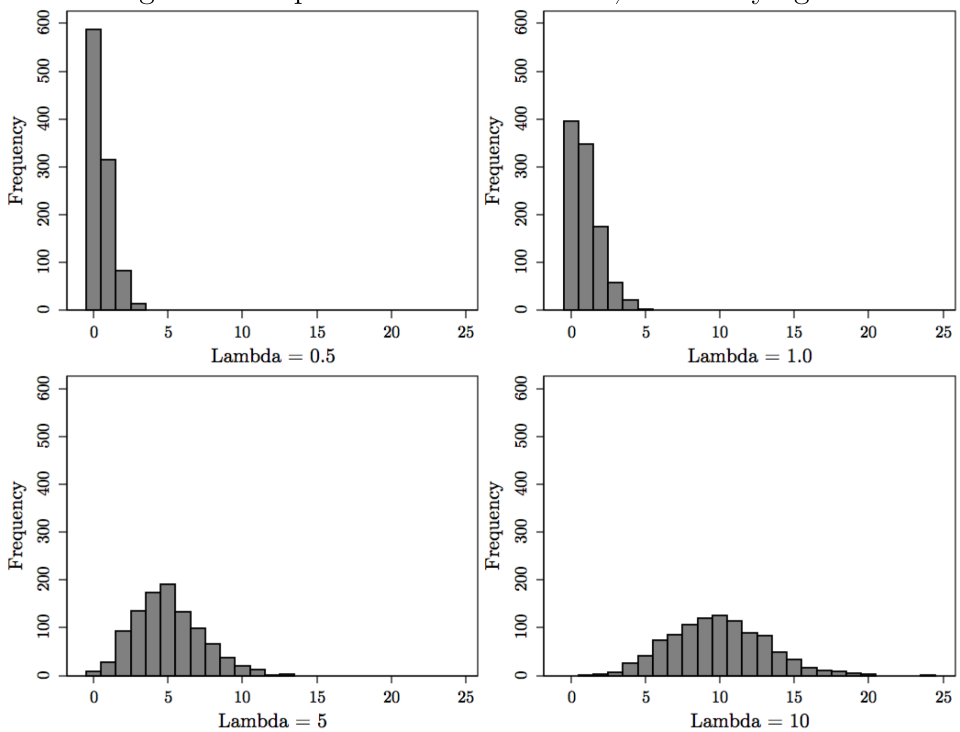
**The Poisson Distribution: Characteristics**

What is this odd thing we call the Poisson distribution, anyway? The Poisson distribution has several important traits:

- It is a discrete probability distribution, with support on the non-negative integers.
- The "rate" $\lambda$ can also be interpreted as the expected number of events during an observation period $t$. In fact, for a Poisson variate $Y$, $\text{E}(Y) = \lambda$.
- As $\lambda$ increases, several interesting things happen:
  1. The *mean/mode* of the distribution gets bigger (no shock there).
  2. The *variance* of the distribution gets larger as well. This also makes sense: since the variable is bounded from below, its variability will necessarily get larger with its mean. In fact, in the Poisson, the mean equals the variance (that is, $\text{E}(Y) = \text{Var}(Y) = \lambda$).
  3. The distribution becomes more Normal-looking (and, in fact, becomes more Normal, period).

  All of these things are illustrated in Figure 1.

Figure 1: Empirical Poisson Variates, with Varying $\lambda$s




Note as well that the Poisson distribution...

- ...is not preserved under affine transformations – that is, affine transformations of Poisson variates are not themselves (necessarily) Poisson variates as well.
- ...is preserved under addition (convolution) provided that the components are independent. That is, for two Poisson variates $X \sim \text{Poisson}(\lambda_X)$ and $Y \sim \text{Poisson}(\lambda_Y)$, $Z = X + Y \sim \text{Poisson}(\mu_{X+Y})$ *iff* $X$ and $Y$ are *independent*. (See e.g. Winkelmann 1997, Chapter 2 for proofs). However,
- ...the same is not true for differences of Poisson variates; see Johnson et al. (2005), §4.12.3 for details.

We could talk more about the Poisson distribution's interesting history, but . . . let's not.

[1]Formally, holding $\lambda$ constant as $M \to \infty$ requires that $\pi \to 0$.

## The Poisson Regression Model

If we assume our event count is distributed according to a Poisson distribution, then the next thing we probably want to know is the effect of covariates $\mathbf{X}_i$ on (the expected value of) $Y_i$. Since we know that $E(Y) > 0$, we need to restrict the "link" to be positive; the exponential is the standard one for this.

So, we have

$$\mathrm{E}(Y_i) \equiv \lambda_i = \exp(\mathbf{X}_i\boldsymbol{\beta}) \tag{3}$$

Incorporating this into Eq. (2), this yields a probability model that looks like:

$$\Pr(Y_i = y|\mathbf{X}_i, \boldsymbol{\beta}) = \frac{\exp[-\exp(\mathbf{X}_i\boldsymbol{\beta})][\exp(\mathbf{X}_i\boldsymbol{\beta})]^y}{y!} \tag{4}$$

Making the usual assumptions about the conditional independence of our $N$ observations, (4) then yields a pretty simple likelihood:

$$\mathrm{L} = \prod_{i=1}^{N} \frac{\exp[-\exp(\mathbf{X}_i\boldsymbol{\beta})][\exp(\mathbf{X}_i\boldsymbol{\beta})]^{Y_i}}{Y_i!} \tag{5}$$

and an equally simple log-likelihood:

$$\ln \mathrm{L} = \sum_{i=1}^{N} \{-\exp(\mathbf{X}_i\boldsymbol{\beta}) + Y_i\mathbf{X}_i\boldsymbol{\beta} - \ln(Y_i!)\} \tag{6}$$

where the last term $-\ln(Y_i!)$ can be omitted because it doesn't vary with $\boldsymbol{\beta}$.

It's actually pretty easy to demonstrate that this log-likelihood is globally concave, and so estimation is really easy and reliable using more-or-less any optimizer we care to adopt. We'll talk about some tinkering with this model a bit later, but for now, this will be our focus for the day.

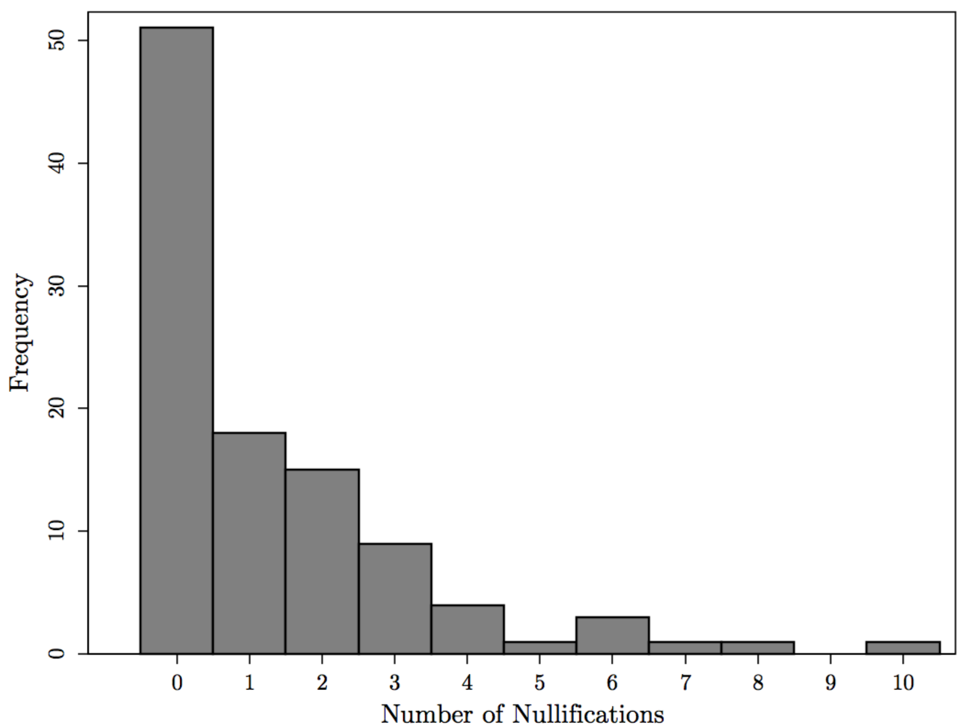
## An Example: Judicial Review of Congressional Acts

As our running example, we'll consider the question of judicial review – specifically, the number of Acts of Congress struck down as unconstitutional by the U.S. Supreme Court in every Congress (that is, every two years). Our response variable $Y_i$ is the number of Acts of Congress overturned ("nullified") by the Supreme Court in each Congress, from 1789-1996 ($N = 104$). We'll consider two covariates:

- The *mean tenure* (`tenure`) of the Supreme Court's justices ($\bar{X} = 10.4, \sigma = 3.4, \mathrm{E}(\hat{\beta}) > 0$).
- Whether (1) or not (0) there was *unified government* (`unified`) ($\bar{X} = 0.83, \mathrm{E}(\hat{\beta}) < 0$).

The obligatory histogram looks like this:

Figure 2: Number of Congressional Acts Nullified by the Supreme Court, by Congress, 1789-1996



And here are the results, from Stata's -poisson- command:

```
. poisson nulls tenure unified

Iteration 0:   log likelihood = -189.53751
Iteration 1:   log likelihood = -189.53751

Poisson regression                                Number of obs   =        104
                                                  LR chi2(2)      =      14.27
                                                  Prob > chi2     =     0.0008
Log likelihood = -189.53751                       Pseudo R2       =     0.0363

------------------------------------------------------------------------------
   nulls |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
  tenure |   .0958868   .0256271    3.742   0.000     .0456585     .146115
 unified |   .1434999   .2327122    0.617   0.537    -.3126077    .5996074
   _cons |  -.8776214   .3712678   -2.364   0.018    -1.605293   -.1499499
------------------------------------------------------------------------------
```

Note how quickly the results converged, thanks in large part to the aforementioned fact that the likelihood is globally concave.

## Model-Checking and Goodness-of-Fit

We'll go into the question of whether the Poisson is a "good" model for the data at more length next time. For now, just recognize that all of the usual ML-based approaches for global model fit are available (if we're using Stata that's the `fitstat` command):

```
. fitstat

Measures of Fit for poisson of nulls

Log-Lik Intercept Only:      -196.673   Log-Lik Full Model:           -189.538
D(101):                       379.075   LR(2):                          14.271
                                        Prob > LR:                       0.001
McFadden's R2:                  0.036   McFadden's Adj R2:               0.021
ML (Cox-Snell) R2:              0.128   Cragg-Uhler(Nagelkerke) R2:      0.131
AIC:                            3.703   AIC*n:                         385.075
BIC:                          -90.008   BIC':                           -4.982
BIC used by Stata:            393.008   AIC used by Stata:             385.075
```

Here:

- The LR test is the usual $F$-type test against the global null,
- The IAC and BIC are the usual information criteria scores, and
- The various pseudo-$R^2$s are what they are.

## Model Interpretation

### Signs-n-Significance

The signs of the $\hat{\boldsymbol{\beta}}$s indicate the effect on the expected number of counts. So, positive signs indicate positive effects – higher values of $X$ correspond to higher counts – while the opposite is true for negative signs. Here, `tenure` seems to "matter" – in that higher mean Supreme Court tenures are associated with greater numbers of nullifications – but unified/divided government doesn't.

### Incident Rate Ratios

As described here, the Poisson model is a log-linear model not unlike the various flavors of logit we talked about before. This means that (surprise, surprise!) there is something like an odds-ratio interpretation of Poisson regression coefficients, just as there is for those models.

In the Poisson case, the quantity of interest is known as the *incidence rate* – that is, $\hat{\lambda}$. The natural way to compare two observations, then, is the *incidence rate ratio* (or *IRR*). For e.g. a binary covariate $X_D$, we can think of the IRR as the ratio

$$\frac{\hat{\lambda}|X_D = 1}{\hat{\lambda}|X_D = 0} = \frac{\exp(\hat{\beta}_0 + \bar{\mathbf{X}}\hat{\boldsymbol{\beta}} + (X_D = 1)\hat{\beta}_{X_D})}{\exp(\hat{\beta}_0 + \bar{\mathbf{X}}\hat{\boldsymbol{\beta}} + (X_D = 0)\hat{\beta}_{X_D})} = \exp(\hat{\beta}_{X_D})$$

That is, we can tell the relative change in the incidence rate for a one-unit change in any given variable $X_k$ by simply exponentiating its coefficient estimate $\hat{\beta}_k$.

- In the example, this yields an estimated IRR for the `unified` variable is equal to $\exp(0.143) = 1.15$.
- Substantively, this means that the incidence rate under unified government is about 1.15 times that under divided government (that is $\lambda_{\text{Unified}} = 1.15 \times \lambda_{\text{Divided}}$), which is not a large change.
- Similarly, a one-year change in the `tenure` variable corresponds to an estimated IRR of $\exp(0.096) = 1.10$. That means that increasing the average tenure of the Court increases the estimated incidence rate by a factor of 1.1.
- More generally, for a $\delta$-unit change in $X_k$ is $\exp(\delta\hat{\beta}_k)$.

- Thus, a 10-year change in `tenure` corresponds to an estimated IRR of $\exp(10 \times 0.096) = \exp(0.96) = 2.61$.
- That is, a Court with an average `tenure` of $t$ years will be expected to have an incidence of judicial review roughly 2.6 times that of a Court with a mean `tenure` equal to $t - 10$.

- As with odds ratios in logits, Stata will report IRRs for you automatically; simply specify the `-irr-` option to the `-poisson-` command:

```
. poisson, irr

Poisson regression                                Number of obs   =        104
                                                  LR chi2(2)      =      14.27
                                                  Prob > chi2     =     0.0008
Log likelihood = -189.53751                       Pseudo R2       =     0.0363
------------------------------------------------------------------------------
   nulls |      IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
  tenure | 1.100634   .0282061    3.74   0.000     1.046717    1.157329
 unified | 1.154307   .2686212    0.62   0.537     .7315369    1.821404
------------------------------------------------------------------------------
```

Note that this also gives you the estimated standard errors and confidence intervals associated with the IRRs. As with logit, etc., IRRs are a nice way of verbally describing your results.

**Predicted Counts**

In a Poisson regression model, the predicted count is just $\exp(\bar{\mathbf{X}}\hat{\boldsymbol{\beta}})$, which – let's face it – is pretty easy to calculate. E.g., for a "typical" / modal case (that is, a unified government in which the average Court tenure is ten years), we get a predicted count of:

$$\begin{aligned} \mathrm{E}(Y|\bar{\mathbf{X}}_i) &= \exp[-0.878 + (0.096 \times 10) + (0.143 \times 1)] \\ &= \exp(0.225) \\ &= 1.25 \end{aligned}$$

From this, you can calculate the change in expected counts by calculating the predicted count for different values of $\mathbf{X}_i$ and taking the difference.

- E.g., the expected count for the same Congress with an average Court `tenure` of 20 years is $\exp(1.185) = 3.27$.
- So, a ten-year increase in tenure corresponds to a $(3.27 - 1.25) \approx$ 2-act increase in judicial review.

- Note that $\frac{3.27}{1.25} = 2.61$, *which is the same as the IRR for a ten-year change reported above.*
- Be sure to include measures of uncertainty here as well (Clarify can be of some help in this respect...).

Predicted counts can be interesting either for *within-sample* or *out-of-sample* predictions. To graph out-of-sample predictions as a function of continuous covariates, we adopt the same "dummy dataset" strategy we've been using for logit, etc.:

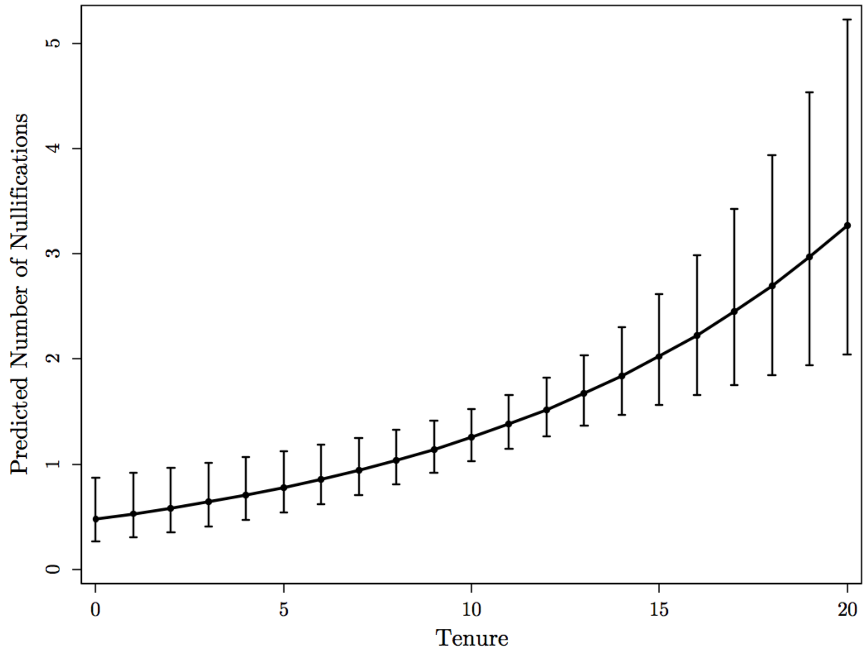
```
. clear

. set obs 21

. gen unified = 1

. gen tenure=_n-1

. save MLENotes10Sim.dta

. use MLENotes10real.dta, clear

. poisson nulls unified tenure
```

*(output omitted)*

```
. use MLENotes10Sim.dta

. predict xb, xb

. predict se, stdp

. gen counthat=exp(xb)

. gen ub=exp(xb+(1.96*se))

. gen lb=exp(xb-(1.96*se))

. twoway (connected counthat tenure, lcolor(black) lpattern(solid) lwidth(medthick)
msymbol(smcircle) mcolor(black) msize(small)) (rcap ub lb tenure, lcolor(black)
msize(small)), ytitle(Predicted Number of Nullifications) xtitle(Tenure) legend(off)
graphregion(margin(vsmall))
```

which yields Figure 3. This demonstrates the increase in the expected number of nullifications associated with a change in `tenure` across its full range.

Figure 3: Predicted Number of Congressional Acts Nullified by the Supreme Court, by `tenure`




**Predicted Probabilities**

Beyond expected counts, we might also be interested in the probability that a particular observation $Y_i$ takes on a particular count value $y$. We can get this predicted probability by plugging the **X** values for that observation, and the estimates of $\hat{\beta}$, into the basic Poisson probability statement:

$$\widehat{\Pr(Y_i = y|\mathbf{X}_i, \hat{\boldsymbol{\beta}})} = \frac{\exp[-\exp(\mathbf{X}_i\hat{\boldsymbol{\beta}})][\exp(\mathbf{X}_i\hat{\boldsymbol{\beta}})]^y}{y!}$$

For example, for the "typical" above case, what are the probabilities of counts equalling 0,1,2, or 3?

$$\begin{aligned}\widehat{\Pr(Y_i = 0|\mathbf{X}_i, \hat{\boldsymbol{\beta}})} &= \frac{[\exp(-1.25)](1.25)^0}{0!} \\ &= \frac{(0.287)(1)}{1} \\ &= 0.287\end{aligned}$$

$$
\begin{aligned}
\widehat{\Pr(Y_i = 1|}\mathbf{X}_i, \hat{\boldsymbol{\beta}}) &= \frac{[\exp(-1.25)](1.25)^1}{1!} \\
&= \frac{(0.287)(1.25)}{1} \\
&= 0.359
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\Pr(Y_i = 2|}\mathbf{X}_i, \hat{\boldsymbol{\beta}}) &= \frac{[\exp(-1.25)](1.25)^2}{2!} \\
&= \frac{(0.287)(1.563)}{2} \\
&= 0.224
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\Pr(Y_i = 3|}\mathbf{X}_i, \hat{\boldsymbol{\beta}}) &= \frac{[\exp(-1.25)](1.25)^3}{3!} \\
&= \frac{(0.287)(1.953)}{6} \\
&= 0.093
\end{aligned}
$$

Adding these, you get 0.963, which tells you that these outcomes account for most of the potential outcomes at this level of the covariates. Obviously, it would be possible to calculate this for a larger range, and to automate this using Stata or whatever, and to graph them.

Finally, and unsurprisingly, most of the other user-written postestimation commands we've bumped into this semester will also interact happily with `-poisson-`. Those include `-test-`, `-mfx-`, Clarify, `-spost-`, and so forth. In our book, Long also suggests calculating the *mean predicted probability* of each possible count, across all observations, as a measure that can be used to assess model fit.

## Note: Exposure and Offsets

Finally, we should touch briefly on the subject of *exposure*. As noted in (1) above, the general format for the Poisson distribution takes into account the extent of "exposure" of each subject $i$. In the Bernoulli / "rare events" formulation of the Poisson model, above, $Y_i$ is the number of events and $M_i$ is the number of "trials" (that is, the number of possible events). There, we assumed $M \to \infty$, but that isn't always the case.

Consider, for example, a model of the number of Supreme Court decisions in each term in which there is at least one dissenting vote/opinion (*a la* Caldeira and Zorn 1998). There's an upper limit on this number: the number of total decisions by the Court – obviously, we can't have $Y_i = 60$ if the Court only decided 59 cases in term $i$. This factor is often referred to as an observation's *exposure* (after a number of applications in biometrics).

Formally, if each of the observations doesn't have the same exposure (that is, if $M_i \neq M_j \, \forall \, i \neq j$), then the expected count is proportional to the exposure:

$$\mathrm{E}(Y_i|\mathbf{X}_i, M_i) = \lambda_i M_i$$

As a result, the amount of exposure needs to be accounted for in some way, or the model is misspecified. The easiest way to do this is to include $M_i$ as an *offset*:

$$\lambda_i = \exp[\mathbf{X}_i\beta + \ln(M_i)]$$

This amounts to including $\ln(M_i)$ among the right-hand-side variables, and constraining the coefficient to equal 1.0. Most software programs have straightforward ways to do this (e.g., in `Stata` , one uses the `-exposure-` option to `-poisson-`).

Alternatively one can simply include $\ln(M_i)$ among the covariates, and then test whether or not $\hat{\beta}_{\ln M} = 1.0$. This latter approach might even be of some substantive interest – for example, if we were modeling the aforementioned number of "dissent cases," it might be interesting to know whether, as the Court's workload increased, the ability/propensity for justices to cast dissenting votes or author dissenting opinions (say) decreased.

We can observe the potential significance of exposure by looking at some national-level data on 102 countries during the period from 1950-1985. These are data aggregated from "politically-relevant dyads" data, which means that they are sums or means of country-year data:

- `Ndyads` is the number of dyad-years which were aggregated to create each observation, ranging from five to 3249,
- `disputes` is the number of (interstate) dispute-years that country experienced during 1950-1985,
- `allies` is the number of (dyadic) ally-years each country had during 1950-1985, and
- `openness` is the country's mean trade openness (that is, the ratio $\frac{\text{Imports}_t + \text{Exports}_t}{\text{GDP}_t}$) across all 36 years in the data.

The interesting question is the aggregate association between trade openness and interstate conflict. Note, however, that the more a country is "in the data," the more possible disputes they could have gotten into:

```
* Data are aggregated dyadic data, 1950-1985...

. su Ndyads disputes allies openness

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      Ndyads |       114    179.3684    451.8623          5       3249
    disputes |       114    3.552632    7.714017          0         52
      allies |       114    63.91228    143.1086          0       1283
    openness |       102    .3920326    .2991905   .0317647   1.658911
```

Moreover – and not surprisingly – "exposure" is highly correlated with both disputes and with the number of allies one had:

```
. corr Ndyads disputes allies openness
(obs=102)
             |   Ndyads disputes   allies openness
-------------+------------------------------------
      Ndyads |   1.0000
    disputes |   0.8626   1.0000
      allies |   0.9200   0.8255   1.0000
    openness |  -0.0751  -0.1682  -0.1255   1.0000
```

An initial model that ignores "exposure" yields:

```
. poisson disputes allies openness

Poisson regression                                Number of obs   =        102
                                                  LR chi2(2)      =     339.71
                                                  Prob > chi2     =     0.0000
Log likelihood = -291.14533                       Pseudo R2       =     0.3684
------------------------------------------------------------------------------
    disputes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      allies |   .0025184   .0001159    21.73   0.000     .0022913    .0027455
    openness |  -1.114413   .2773631    -4.02   0.000    -1.658035   -.5707915
       _cons |    1.15595   .1117581    10.34   0.000      .936908    1.374992
------------------------------------------------------------------------------
```

Note that this model reflects the strong correlation (through exposure) of allies and disputes – countries that had a great deal of "exposure" get large numbers for both, suggesting (counterintuitively) that countries with larger numbers of allies are also *more* likely to get into disputes.

Once we include an exposure term in the model, however, the results change:

```
. poisson disputes allies openness, exposure(Ndyads)

Poisson regression                                Number of obs   =        102
                                                  LR chi2(2)      =      42.40
                                                  Prob > chi2     =     0.0000
Log likelihood = -233.55463                       Pseudo R2       =     0.0832
------------------------------------------------------------------------------
    disputes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      allies |  -.0006058   .0001333    -4.54   0.000    -.0008671   -.0003445
    openness |  -1.604059   .3167503    -5.06   0.000    -2.224878   -.9832395
       _cons |  -3.290605   .1194627   -27.55   0.000    -3.524748   -3.056463
      Ndyads | (exposure)
------------------------------------------------------------------------------
```

Once we "control" for exposure, the `allies` variable has a negative sign (which is what we would expect), though the strength of the association is small. We can accomplish something very similar by generating a variable for the log of exposure and including it among the covariates:

```
. gen exposure = ln(Ndyads)

. poisson disputes allies openness exposure

Poisson regression                                Number of obs   =        102
                                                  LR chi2(3)      =     462.09
                                                  Prob > chi2     =     0.0000
Log likelihood = -229.95478                       Pseudo R2       =     0.5012
------------------------------------------------------------------------------
    disputes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      allies |  -9.48e-06   .0002569    -0.04   0.971    -.0005129     .000494
    openness |  -1.444624   .3119605    -4.63   0.000    -2.056056    -.833193
    exposure |   .8109774   .0709538    11.43   0.000     .6719105    .9500444
       _cons |  -2.426567   .3434595    -7.07   0.000    -3.099735   -1.753398
------------------------------------------------------------------------------

. test exposure=1

 ( 1)  [disputes]exposure = 1

           chi2(  1) =    7.10
         Prob > chi2 =    0.0077
```

These latter results indicate that:

- The effect relationship between `disputes` and `openness` remains more or less unchanged throughout these models – an unsurprising fact, given that `openness` is not at all strongly related to exposure.
- `allies`' relationship to `disputes` is effectively zero, once we address each nation's degree of exposure, and
- The increase in `disputes` as a function of `exposure` is not quite proportional / one-to-one.

Finally, note that as a practical matter, if (a) all of the observations have the same (or very similar) exposures, and/or (b) for the most part, $Y_i$ is significantly less than $M_i$ (that is, no observation "comes close" to experiencing its maximum possible number of events) then the issue of exposure is not a big deal, and you can probably safely ignore it.