GSERM 2020

Regression for Publishing

June 18, 2020 (second session)

Event Counts

Things That Are Not Counts

- Ordinal scales/variables
- Grouped Binary Data
  - $\frac{N \text{ of ``successes''}}{N \text{ of ``trials''}}$
  - Binomial data
  - $=$ counts only if Pr("success") is small

Count Properties

- Discrete / integer-values
- Non-negative
- "Cumulative"

Count Data: Motivation

$$\text{Arrival Rate} = \lambda$$

$$\Pr(\text{Event})_{t,t+h} = \lambda h$$

$$\Pr(\text{No Event})_{t,t+h} = 1 - \lambda h$$

$$\begin{aligned} \Pr(Y_t = y) &= \frac{\exp(-\lambda h)\lambda h^y}{y!} \\ &= \frac{\exp(-\lambda)\lambda^y}{y!} \end{aligned}$$

Poisson Assumptions

- No Simultaneous Events
- Constant Arrival Rate
- Independent Event Arrivals

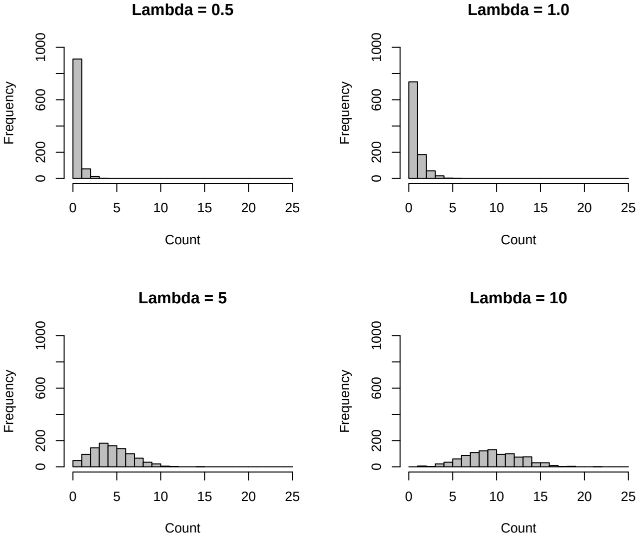
For M independent Bernoulli trials with (sufficiently small) probability of success π and where $M\pi \equiv \lambda > 0$,

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

Poisson: Characteristics

- Discrete
- $\mathrm{E}(Y) = \mathrm{Var}(Y) = \lambda$
- Is not preserved under affine transformations...
- For $X \sim \text{Poisson}(\lambda_X)$ and $Y \sim \text{Poisson}(\lambda_Y)$, $Z = X + Y \sim \text{Poisson}(\lambda_{X+Y})$ *iff* X and Y are *independent* but
- ...same is not true for differences.
- $\lambda \to \infty \iff Y \sim N$

Poissons: Examples



Suppose

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

then

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

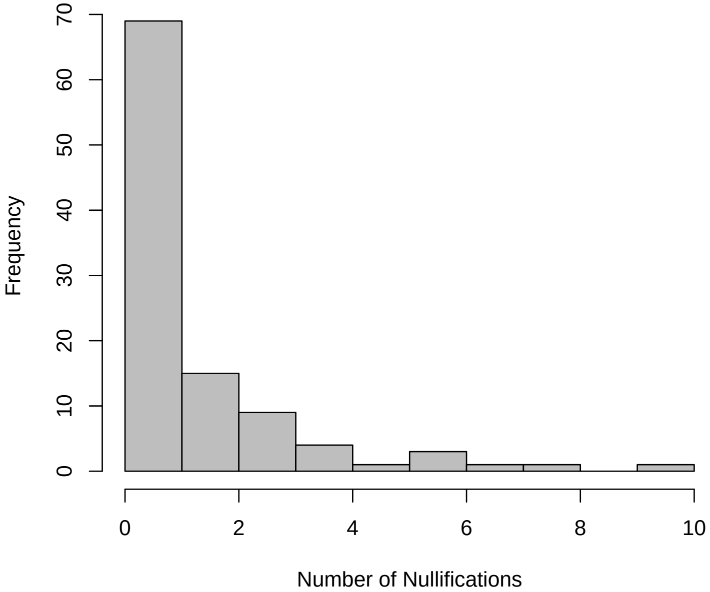
$$L = \prod_{i=1}^{N} \frac{\exp[-\exp(\mathbf{X}_i\boldsymbol{\beta})][\exp(\mathbf{X}_i\boldsymbol{\beta})]^{Y_i}}{Y_i!}$$

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

Example: Judicial Review

- $Y_i =$ number of Acts of Congress overturned by the Supreme Court in each Congress,
- 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`) ($\bar{X} = 0.83, \mathrm{E}(\hat{\beta}) < 0$).

Supreme Court Nullifications, 1789-1996



Estimation

```
> nulls.poisson<-glm(nulls~tenure+unified,family="poisson",data=Nulls)
> summary(nulls.poisson)

Call:
glm(formula = nulls ~ tenure + unified, family = "poisson", data = Nulls)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-2.367  -1.503  -0.623   0.561   4.153

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.8776     0.3713   -2.36  0.01809 *
tenure        0.0959     0.0256    3.74  0.00018 ***
unified       0.1435     0.2327    0.62  0.53747
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 251.80  on 103  degrees of freedom
Residual deviance: 237.52  on 101  degrees of freedom
AIC: 385.1

Number of Fisher Scoring iterations: 6
```

Interpretation: Incidence Rate Ratios

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

- Like ORs
- `unified`: IRR=$\exp(0.143) = 1.15$

Incidence Rate Ratios, continued

$$\text{IRR}_{X_k, X_k+\delta} = \exp(\delta \hat{\beta}_k)$$

So, a ten-year difference in `tenure`:

$$\begin{aligned}\text{IRR} &= \exp(10 \times 0.096) \\ &= \exp(0.96) \\ &= 2.61\end{aligned}$$

Incidence Rate Ratios

```
> library(mfx)
> nulls.poisson.IRR<-poissonirr(nulls~tenure+unified,
                    data=Nulls)
> nulls.poisson.IRR

Call:
poissonirr(formula = nulls ~ tenure + unified, data = Nulls)

Incidence-Rate Ratio:
           IRR Std. Err.    z   P>|z|
tenure  1.1006    0.0282 3.74 0.00018 ***
unified 1.1543    0.2686 0.62 0.53747
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Predicted Values ($\hat{Y}$s)

Mean predicted Y:

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

In-Sample

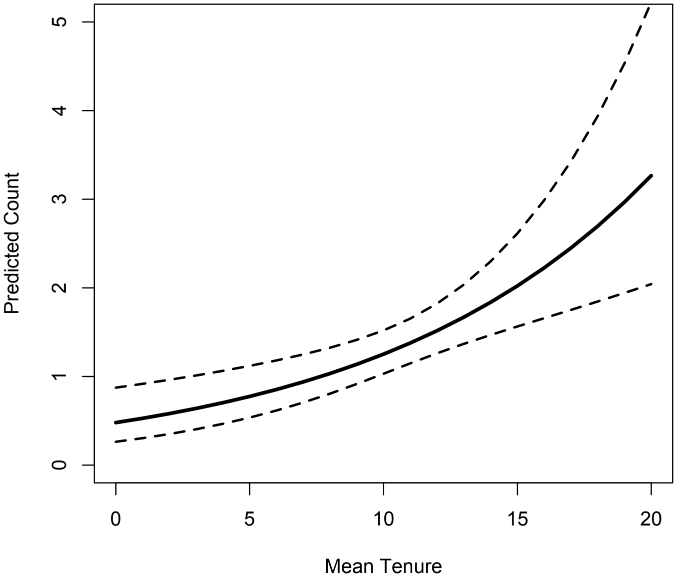
- R : in `$fitted.values`
- Stata : use `predict`

Out-of-Sample: use `predict`

Example: Out-Of-Sample Predicted Values

```
> tenure<-seq(0,20,1)
> unified<-1
> simdata<-as.data.frame(cbind(tenure,unified))
> nullhats<-predict(nulls.poisson,newdata=simdata,se.fit=TRUE)
>
> # NOTE: These are XBs, not predicted counts.
> # Transforming:
> nullhats$Yhat<-exp(nullhats$fit)
> nullhats$UB<-exp(nullhats$fit + 1.96*(nullhats$se.fit))
> nullhats$LB<-exp(nullhats$fit - 1.96*(nullhats$se.fit))
> plot(simdata$tenure,nullhats$Yhat,t="l",lwd=3,ylim=c(0,5),ylab=
+            "Predicted Count", xlab="Mean Tenure")
> lines(simdata$tenure,nullhats$UB,lwd=2,lty=2)
> lines(simdata$tenure,nullhats$LB,lwd=2,lty=2)
>
> plot(simdata$tenure,nullhats$Yhat,t="l",lwd=3,ylim=c(0,5),ylab=
+            "Predicted Count", xlab="Mean Tenure")
> lines(simdata$tenure,nullhats$UB,lwd=2,lty=2)
> lines(simdata$tenure,nullhats$LB,lwd=2,lty=2)
```

Plotting Out-Of-Sample Predicted Values



Predicted probability is:

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

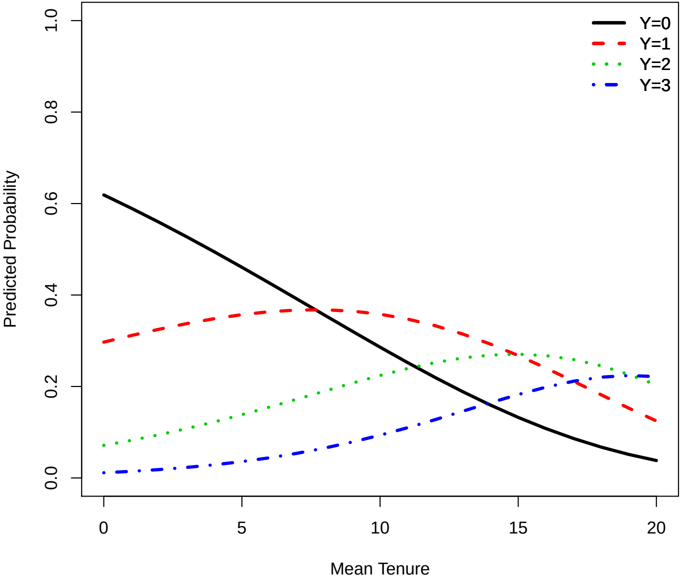
So for (e.g.) $\lambda = 1.25$:

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

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

etc.

Plotting (OOS) Predicted Probabilities



“Exposure” and “Offsets”

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

Same as including $\ln(M_i)$ in $\mathbf{X}$ with $\beta_{\ln M} = 1$.

- Example: Data on numbers of interstate disputes by country, 1950-1985
- $N = 102$, but
- `Ndyads` = number of dyad-years which were aggregated to create each observation, ranging from five to 3249
- `disputes` = number of (interstate) dispute-years that country experienced during 1950-1985
- `allies` = number of (dyadic) ally-years each country had during 1950-1985
- `openness` $= \frac{1}{36}\left(\frac{\text{Imports}_t + \text{Exports}_t}{\text{GDP}_t}\right)$ across all 36 years in the data.

“Exposure” and “Offsets”: Data

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

> summary(IR)
     ccode           Ndyads          disputes           allies            openness           exposure
 Min.   :  2    Min.    :   5    Min.    : 0.00    Min.    :    0.0    Min.    :0.032    Min.    :1.61
 1st Qu.:214    1st Qu.:  44    1st Qu.: 0.00    1st Qu.:    0.0    1st Qu.:0.185    1st Qu.:3.79
 Median :436    Median :  92    Median : 1.00    Median :   26.0    Median :0.296    Median :4.52
 Mean   :418    Mean    : 179    Mean    : 3.55    Mean    :   63.9    Mean    :0.392    Mean    :4.42
 3rd Qu.:598    3rd Qu.: 146    3rd Qu.: 4.00    3rd Qu.:   81.0    3rd Qu.:0.535    3rd Qu.:4.98
 Max.   :900    Max.    :3249    Max.    :52.00    Max.    :1283.0    Max.    :1.659    Max.    :8.09
                                                                      NA's    :12

> cor(IR,use="complete.obs")
            ccode   Ndyads disputes  allies openness exposure
ccode     1.00000 -0.29623  -0.1399 -0.3983  0.02744  -0.6544
Ndyads   -0.29623  1.00000   0.8626  0.9200 -0.07511   0.6988
disputes -0.13989  0.86257   1.0000  0.8255 -0.16819   0.6335
allies   -0.39826  0.92004   0.8255  1.0000 -0.12548   0.7003
openness  0.02744 -0.07511  -0.1682 -0.1255  1.00000  -0.1433
exposure -0.65442  0.69878   0.6335  0.7003 -0.14325   1.0000
```

Ignoring Exposure

```
> IR.fit1<-glm(disputes~allies+openness,data=IR,family="poisson")
> summary(IR.fit1)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.1559498  0.1117581  10.343  < 2e-16 ***
allies       0.0025184  0.0001159  21.734  < 2e-16 ***
openness    -1.1144132  0.2773631  -4.018 5.87e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 732.68  on 101  degrees of freedom
Residual deviance: 392.97  on  99  degrees of freedom
  (12 observations deleted due to missingness)
AIC: 588.29

Number of Fisher Scoring iterations: 6
```

Correcting for Exposure

```
> IR.fit2<-glm(disputes~allies+openness,data=IR,family="poisson",
  offset=log(Ndyads))
> summary(IR.fit2)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.2906055  0.1194616 -27.545  < 2e-16 ***
allies      -0.0006058  0.0001333  -4.544 5.52e-06 ***
openness    -1.6040587  0.3167415  -5.064 4.10e-07 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

    Null deviance: 320.19  on 101  degrees of freedom
Residual deviance: 277.79  on  99  degrees of freedom
  (12 observations deleted due to missingness)
AIC: 473.11

Number of Fisher Scoring iterations: 5
```

Correcting for Exposure (continued)

```
> IR.fit3<-glm(disputes~allies+openness+log(Ndyads),data=IR,
+              family="poisson")
> summary(IR.fit3)

Coefficients:
               Estimate  Std. Error z value              Pr(>|z|)
(Intercept) -2.42656676  0.34345252   -7.07      0.0000000000016 ***
allies      -0.00000948  0.00025687   -0.04                 0.97
openness    -1.44462460  0.31193821   -4.63      0.0000036368547 ***
log(Ndyads)  0.81097748  0.07095243   11.43 < 0.0000000000000002 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 732.68  on 101  degrees of freedom
Residual deviance: 270.59  on  98  degrees of freedom
  (12 observations deleted due to missingness)
AIC: 467.9

Number of Fisher Scoring iterations: 5
```

Test $\beta_{\texttt{exposure}} = 1.0$

```
> # z-test:

> 2*pnorm((0.811-1)/.071)
[1] 0.007768438



> # Wald test:

> wald.test(b=coef(IR.fit3),Sigma=vcov(IR.fit3),Terms=4,H0=1)

Wald test:
----------
Chi-squared test:
X2 = 7.1, df = 1, P(> X2) = 0.0077
```

Count Models: Extensions, Variants, etc.

- Overdispersion $\rightarrow$ Negative Binomial Model
  - May be due to *event dependence* and/or *unobserved heterogeneity*
  - Fits an additional parameter for the variance of the counts
  - Rarely: Underdispersion ($\rightarrow$ continuous parameter binomial model)

- Models for "excess" zeros
  - Zero-Inflated models: mixture-based
  - Hurdle models: compound distributions

- Models for time series & panel data with event counts

- Models for sample selection and endogeneity...

GLM Interpretation, Generally

- Nonlinearity means $\frac{\partial E(Y)}{\partial X} \neq c$.
- $\rightarrow$ obtaining marginal effects requires "holding all else constant"
- Model fit is usually best thought of in predictive terms
- Deviance residuals can be used for diagnostics