# PLSC 503 – Spring 2023
# Models For Event Counts

April 17, 2023

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-valued
- Non-negative
- "Cumulative"

# Count Data: Motivation

Events:

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

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

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

*Count* of events:

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

# Poisson Assumptions

Three assumptions:

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

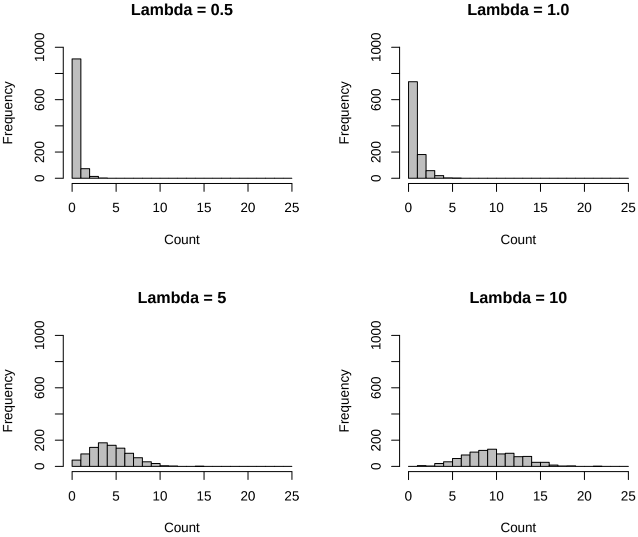
For $M$ independent Bernoulli trials with (sufficiently small) probability of success $\pi$ and where $M\pi \equiv \lambda > 0$,

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

# 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



# Poisson Regression

Suppose

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

then

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

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

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

Dahl (1957): SCOTUS overturning Acts of Congress:

- $Y_i =$ 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`) ($\bar{X} = 0.83, \mathrm{E}(\hat{\beta}) < 0$).

# Example Redux: Federal Judicial Review, 1789-2018

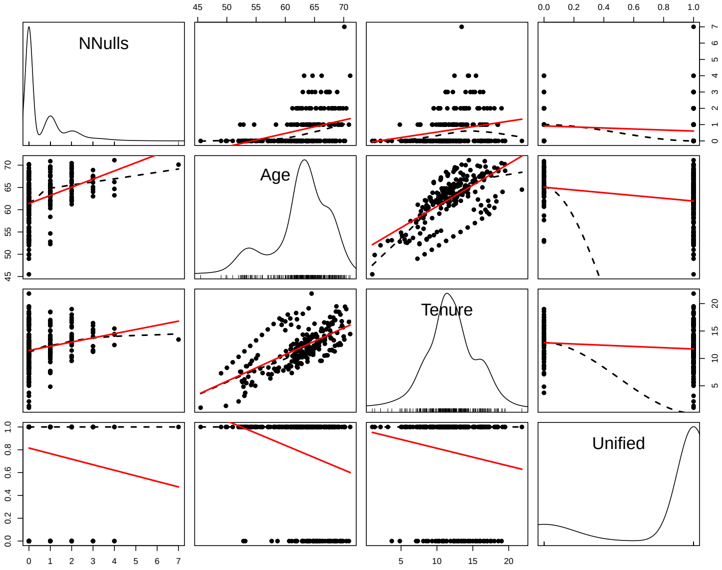
Dahl (1957):

- SCOTUS gets "out of step" with the other branches $\rightarrow$ judicial review
- Older / longer-serving justices will more likely to invalidate legislation

Data:

```
> summary(NewDahl)
      Year          NNulls            Age            Tenure          Unified
 Min.   :1789   Min.   :0.000   Min.   :45.5   Min.   : 1.0   Min.   :0.000
 1st Qu.:1846   1st Qu.:0.000   1st Qu.:60.7   1st Qu.:10.0   1st Qu.:1.000
 Median :1904   Median :0.000   Median :63.5   Median :11.8   Median :1.000
 Mean   :1904   Mean   :0.674   Mean   :62.6   Mean   :12.0   Mean   :0.783
 3rd Qu.:1961   3rd Qu.:1.000   3rd Qu.:66.0   3rd Qu.:14.1   3rd Qu.:1.000
 Max.   :2018   Max.   :7.000   Max.   :71.1   Max.   :21.8   Max.   :1.000
```

# Federal Judicial Review, 1789-2018



```
> nulls.poisson<-glm(NNulls~Age+Tenure+Unified,family="poisson",
+                    data=NewDahl)
> summary(nulls.poisson)

Call:
glm(formula = NNulls ~ Age + Tenure + Unified, family = "poisson",
    data = NewDahl)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.933  -1.033  -0.544   0.393   2.954

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -12.0321     1.7258   -6.97  3.1e-12 ***
Age           0.1897     0.0296    6.42  1.4e-10 ***
Tenure       -0.0421     0.0350   -1.20     0.23
Unified      -0.0336     0.1808   -0.19     0.85
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 331.95  on 229  degrees of freedom
Residual deviance: 261.39  on 226  degrees of freedom
AIC: 484.1

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
- `Age`: IRR$=$exp$(0.19) = 1.21$

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

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

$$\begin{aligned}\text{IRR} &= \exp(10 \times 0.190) \\ &= \exp(1.90) \\ &= 6.69\end{aligned}$$

# Incidence Rate Ratios

```
> library(mfx)
> nulls.poisson.IRR<-poissonirr(NNulls~Age+Tenure+Unified,
+                               data=NewDahl)
> nulls.poisson.IRR
Call:
poissonirr(formula = NNulls ~ Age + Tenure + Unified, data = NewDahl)

Incidence-Rate Ratio:
           IRR Std. Err.     z          P>|z|
Age     1.2089    0.0357  6.42 0.00000000014 ***
Tenure  0.9588    0.0335 -1.20          0.23
Unified 0.9670    0.1749 -0.19          0.85
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Mean predicted $Y$:

$$\mathrm{E}(Y|\bar{\mathbf{X}}_i) = \exp[\bar{\mathbf{X}}_i\hat{\beta}]$$

In-Sample:

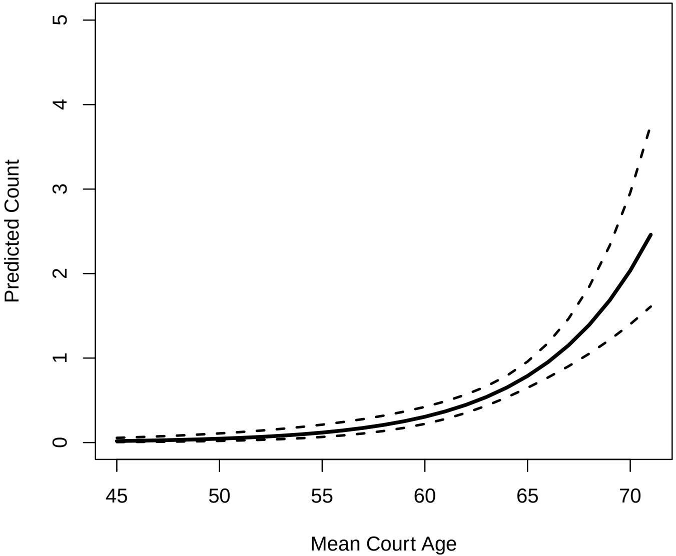
- R : in `$fitted.values`
- Stata : use `predict`

Out-of-Sample: use `predict`

# Example: Out-Of-Sample Predicted Values

```
> simdata<-data.frame(Age=seq(from=45,to=71,by=1),
+                     Tenure=mean(NewDahl$Tenure,na.rm=TRUE),
+                     Unified=1)
> 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$Age,nullhats$Yhat,t="l",lwd=3,ylim=c(0,5),ylab=
+        "Predicted Count", xlab="Mean Court Age")
> lines(simdata$Age,nullhats$UB,lwd=2,lty=2)
> lines(simdata$Age,nullhats$LB,lwd=2,lty=2)
```

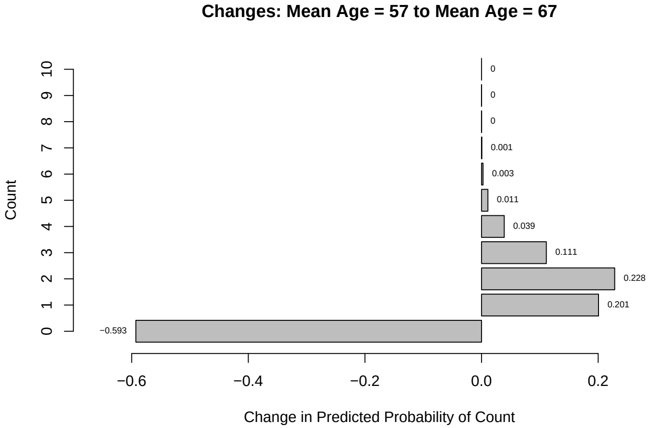
# Plotting Out-Of-Sample Predicted Values



# Predicted Probabilities

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

# Changes in Predicted Probabilities



## "Exposure" and "Offsets"

If we relax the assumption of equal "exposure," we get:

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

i.e., the expected number of events is proportional to *exposure* $M_i$.

Note that now, instead of:

$$\ln[E(Y_i)] = \mathbf{X}_i\boldsymbol{\beta}$$

we have:

$$\ln\left[E\left(\frac{Y_i}{M_i}\right)\right] = \mathbf{X}_i\boldsymbol{\beta}$$

which is the same as:

$$\ln[E(Y_i)] = \ln(M_i) + \mathbf{X}_i\boldsymbol{\beta}$$

that is, 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     1.0000 -0.2962   -0.140 -0.398   0.0274   -0.654
Ndyads   -0.2962  1.0000    0.863  0.920  -0.0751    0.699
disputes -0.1399  0.8626    1.000  0.825  -0.1682    0.634
allies   -0.3983  0.9200    0.825  1.000  -0.1255    0.700
openness  0.0274 -0.0751   -0.168 -0.125   1.0000   -0.143
exposure -0.6544  0.6988    0.634  0.700  -0.1433    1.000
```

```
> 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)

Call:
glm(formula = disputes ~ allies + openness + log(Ndyads), family = "poisson",
    data = IR)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-2.838  -1.390  -0.758   0.605   4.731

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
```

# Contagion, Heterogeneity, and Dispersion


Lisco
26
Oshkosh
27
Lewellen
26


Seattle
WASHINGTON
MONTANA
NORTH DAKOTA
MINNESOTA
SOUTH DAKOTA
WISCONSIN
MICHIGAN
Ottawa
Toronto
NB
MAINE
VT
NH
NEW YORK
MA
Boston
CT
RI
OREGON
IDAHO
WYOMING
NEBRASKA
IOWA
Chicago
ILLINOIS
OHIO
PENN
New York
INDIANA
United States
NEVADA
UTAH
COLORADO
KANSAS
MISSOURI
San Francisco
MD
DE
NJ
WEST VIRGINIA
VIRGINIA
KENTUCKY
CALIFORNIA
Las Vegas
OKLAHOMA
TENNESSEE
NORTH CAROLINA
Los Angeles
ARIZONA
ARKANSAS
NEW MEXICO
MISSISSIPPI
SOUTH CAROLINA
ALABAMA
San Diego
TEXAS
GEORGIA
LOUISIANA
Houston
Gulf of California
FLORIDA
Gulf of Mexico
Mexico

Cats (daily values):

$$
\begin{aligned}
Y_{cats} &= \{0, 1, 1, 0, 2, 0, 1, 0, 3, 1, 2, 1, 0, 2\} \\
\bar{Y}_{cats} &= 1.0, \\
\sigma_{cats} &= 0.92.
\end{aligned}
$$

# Heterogeneity, Contagion, and Dispersion

$$\mathrm{E}(Y_{cats}) = \lambda_{cats}$$

Assumes:

- $Y = 0$ at $t = 0$,
- Exclusive events
- $t_j = t_k \,\forall\, j \neq k$
- Constant, independent Pr(Event) over $t$

Daily values:

$$
\begin{aligned}
Y_{antelope} &= \{0,0,0,0,0,0,0,0,0,0,0,0,7,7\} \\
\bar{Y}_{antelope} &= 1.0, \\
\sigma_{antelope} &= 6.46.
\end{aligned}
$$

*Positive contagion* $\rightarrow$ *overdispersion*.

Daily values:

$$
\begin{aligned}
Y_{foxes} &= \{1, 0, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1\} \\
\bar{Y}_{foxes} &= 1.0, \\
\sigma_{foxes} &= 0.15.
\end{aligned}
$$

*Negative contagion* $\rightarrow$ *underdispersion*.

# Aggregation & Cross-Period Effects

Aggregated two-day measures:

$$
\begin{aligned}
Y_{cats} &= \{1, 1, 2, 1, 4, 3, 2\} \\
Y_{antelope} &= \{0, 0, 0, 0, 0, 0, 14\} \\
Y_{foxes} &= \{1, 2, 2, 3, 2, 2, 2\}
\end{aligned}
$$

- Correct specification
- Correct distribution for $\epsilon$
- Constant $\mathrm{E}(Y|\mathbf{X}, \boldsymbol{\beta})$

$$\lambda_i \equiv \mathrm{E}(Y_i) = f[\mathbf{X}_i\boldsymbol{\beta} + Z_i\theta]$$

Examine:

$$\hat{u}_i = \delta\hat{\lambda}_i + \epsilon_i$$

where

$$\hat{u}_i = \frac{(Y_i - \hat{\lambda}_i)^2 - Y_i}{\hat{\lambda}_i\sqrt{2}}$$

- Estimate a Poisson regression of $Y_i$ on $\mathbf{X}_i$, and generate predicted counts $\hat{\lambda}_i$.
- Calculate $\hat{u}_i$ according to the equation above.
- Estimate $\delta$ using OLS, and test $H_0 : \hat{\delta} = 0$.

$$\begin{aligned} E(Y_i) \equiv \lambda_i &= \exp(\mathbf{X}_i\boldsymbol{\beta} + u_i) \\ &= \exp(\mathbf{X}_i\boldsymbol{\beta})\exp(u_i) \\ &= \lambda_i\nu_i \end{aligned}$$

$$\nu_i \sim \text{gamma}\left(1, \frac{1}{\alpha}\right)$$

$$\Pr(Y_i = y|\lambda_i, \alpha) = \left(\frac{\Gamma(\alpha^{-1} + Y_i)}{\Gamma(\alpha^{-1})\Gamma(Y_i + 1)}\right)\left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i}\right)^{\alpha^{-1}}\left(\frac{\lambda_i}{\lambda_i + \alpha^{-1}}\right)^{Y_i}$$

where

$$\Gamma(a) = \int_0^\infty \exp(-t)t^{a-1}dt$$

Basis:

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

Model has

$$\mathrm{E}(Y) = \lambda$$

$$\mathrm{Var}(Y) = \lambda(1 + \alpha\lambda),\ \alpha > 0$$

# Negative Binomial (log-)Likelihood

$$\begin{aligned}\ln L_{NB} &= \sum_{i=1}^{N}\Bigg\{\left(\sum_{j=0}^{Y_i-1}\ln(j+\alpha^{-1})\right) - \ln Y_i! - \\ &\quad (Y_i-\alpha^{-1})\ln[1+\alpha\exp(\mathbf{X}_i\boldsymbol{\beta})] + Y_i\ln\alpha + Y_i\mathbf{X}_i\boldsymbol{\beta}\Bigg\}\end{aligned}$$

So:

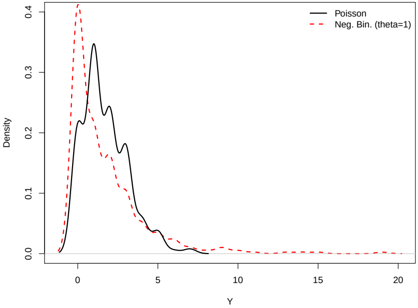
- $\alpha = 0 \iff \mathrm{E}(Y) = \mathrm{Var}(Y)$
- LR test for overdispersion:

$$-2\times(\widehat{\ln L_{Poisson}} - \widehat{\ln L_{NB}}) \sim \chi^2_1$$

- $\widehat{E(Y_i)} \equiv \hat{\lambda}_i = \exp(\mathbf{X}_i\hat{\boldsymbol{\beta}})$

# What Difference Does It Make?

```
> N<-400
> set.seed(7222009)
> X <- runif(N,min=0,max=1)
> YPois <- rpois(N,exp(0+1*X))           # Poisson
> YNB <- rnbinom(N,size=1,mu=exp(0+1*X)) # NB with theta=1.0
>
> describe(cbind(YPois,YNB))
      vars   n mean   sd median trimmed  mad min max range skew kurtosis   se
YPois    1 400 1.72 1.41      1    1.56 1.48   0   7     7 0.92     0.84 0.07
YNB      2 400 1.71 2.44      1    1.22 1.48   0  19    19 2.76    11.15 0.12
```



# What Difference Does It Make (cont'd)?

```
> # Regressions:
>
> summary(glm(YPois~X,family="poisson")) # Poisson

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.009637   0.085337  -0.113     0.91
X            1.030573   0.131992   7.808 5.82e-15 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 516.06  on 399  degrees of freedom
Residual deviance: 453.55  on 398  degrees of freedom
AIC: 1274.4



> summary(glm.nb(YPois~X))               # NB

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.009629   0.085345  -0.113     0.91
X            1.030557   0.132007   7.807 5.86e-15 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for Negative Binomial(7837.699) family taken to be 1)

    Null deviance: 515.96  on 399  degrees of freedom
Residual deviance: 453.46  on 398  degrees of freedom
AIC: 1276.5

              Theta:  7838
          Std. Err.:  135342
Warning while fitting theta: iteration limit reached

 2 x log-likelihood:  -1270.451
```

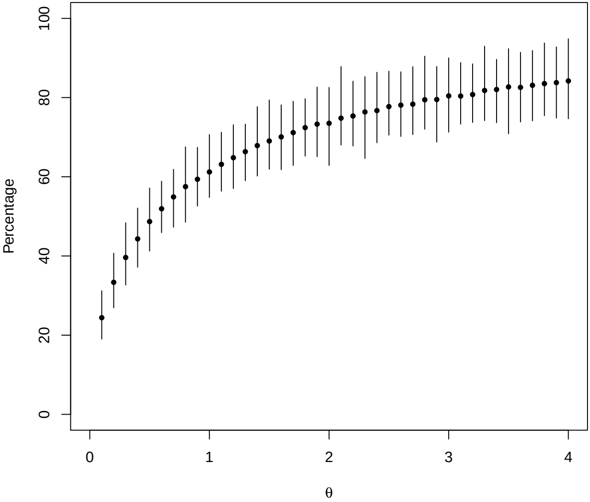
# What Difference Does It Make (cont'd)?

```
> # More regressions:
>
> summary(glm(YNB~X,family="poisson"))  # Poisson

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.03170    0.08593  -0.369    0.712
X            1.06109    0.13248   8.009 1.15e-15 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1118.0  on 399  degrees of freedom
Residual deviance: 1052.1  on 398  degrees of freedom
AIC: 1698.6



> summary(glm.nb(YNB~X))                # NB

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.03525    0.13650  -0.258    0.796
X            1.06773    0.22809   4.681 2.85e-06 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for Negative Binomial(0.8499) family taken to be 1)

    Null deviance: 436.92  on 399  degrees of freedom
Residual deviance: 414.81  on 398  degrees of freedom
AIC: 1407.4

              Theta:  0.850
          Std. Err.:  0.109

 2 x log-likelihood:  -1401.354
```

# Poisson Regression Underestimates N.B. Variances

```
Sims <- 250 # (250 sims each)
theta <- seq(0.1,4,by=0.1) # values of theta
diffs<-matrix(nrow=Sims,ncol=length(theta))

set.seed(7222009)
for(j in 1:length(theta)) {
  for(i in 1:Sims) {
    X<-runif(N,min=0,max=1)
    Y<-rnbinom(N,size=theta[j],mu=exp(0+1*X))
    p<-glm(Y~X,family="poisson")
    nb<-glm.nb(Y~X)
    diffs[i,j]<- ((sqrt(vcov(p))[2,2]) / sqrt(vcov(nb))[2,2])*100
  }
}
```

## Percentage of True (Negative Binomial) S.E. From Fitted Poisson Model, by $\theta = \frac{1}{\alpha}$



Model fitting (in R):

- `glm.nb` (in `MASS`)
- `negbinomial` (in `VGAM`)
- `negbin` (in `aod`)
- `glmnb.fit` (in `statmod`)
- Probably others...

Model interpretation + diagnostics:

- `fitNBP` (in `statmod`) (dispersion parameter estimation)
- `negbinirr` (in `mfx`) (IRRs)
- `negbinmfx` (in `mfx`) (marginal effects)

"Continuous parameter binomial":

$$\Pr(Y_i = y|\lambda_i, \alpha) = \frac{\frac{\Gamma\left(\frac{-\lambda_i}{\alpha-1}+1\right)}{Y_i!\Gamma\left(\frac{-\lambda_i}{\alpha-1}-Y_i+1\right)}(1-\alpha)^{Y_i}(\alpha)^{\frac{-\lambda_i}{\alpha-1}-Y_i}}{D_i}$$

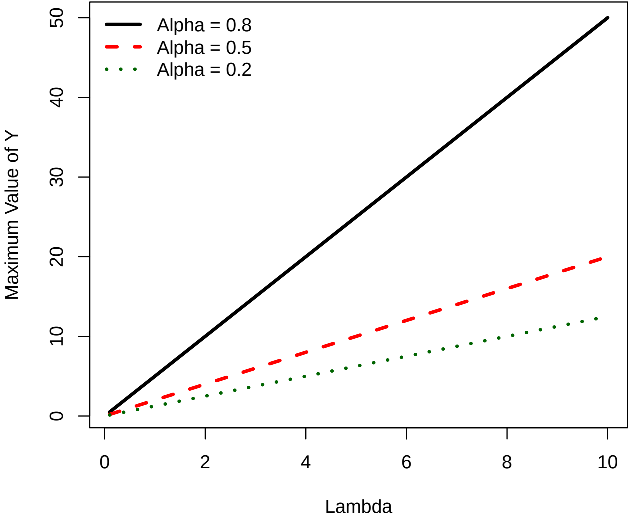
where $D_i = \sum_0^{\frac{-\lambda_i}{\alpha-1}+1}$ of the binomial distribution...

# Are You Down With The CPB?

CPB:

- ...also has $\mathrm{E}(Y_i) = \lambda_i$ [with $\mu_i = \exp(\mathbf{X}_i\boldsymbol{\beta})$]
- ...has $\mathrm{Var}(Y) = \lambda_i\alpha$ with $0 < \alpha < 1$
- ... reduces to the standard Poisson when $\alpha = 1$
- ...imposes a theoretical "upper limit" on the count variable. In particular,

$$\max(Y_i) = \frac{-\lambda_i}{\alpha - 1}.$$


Alpha = 0.8
Alpha = 0.5
Alpha = 0.2
Maximum Value of Y
Lambda
0
10
20
30
40
50
0
2
4
6
8
10

# CPB (log-)Likelihood

$$
\begin{aligned}
\ln L_{CPB} &= \sum_{i=1}^{N} \Big\{ \ln \Gamma\left(\frac{-\lambda_i}{\alpha - 1} + 1\right) - \ln \Gamma\left(\frac{-\lambda_i}{\alpha - 1} - Y_i + 1\right) \\
&\quad + Y_i \ln(1-\alpha) + \left(\frac{-\lambda_i}{\alpha - 1} - Y_i\right)\ln(\alpha) - \ln(D_i) \Big\}
\end{aligned}
$$

# Example Redux: Judicial Review

Recall:

```
> summary(nulls.poisson)

Call:
glm(formula = NNulls ~ Age + Tenure + Unified, family = "poisson",
    data = NewDahl)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.933  -1.033  -0.544   0.393   2.954

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -12.0321     1.7258   -6.97  3.1e-12 ***
Age           0.1897     0.0296    6.42  1.4e-10 ***
Tenure       -0.0421     0.0350   -1.20     0.23
Unified      -0.0336     0.1808   -0.19     0.85
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 331.95  on 229  degrees of freedom
Residual deviance: 261.39  on 226  degrees of freedom
AIC: 484.1

Number of Fisher Scoring iterations: 6
```

# Overdispersion Test: "By Hand"

```
> Phats<-fitted.values(nulls.poisson)
> Uhats<-((NewDahl$NNulls-Phats)^2 - NewDahl$NNulls) / (Phats * sqrt(2))

> summary(lm(Uhats~Phats))

Call:
lm(formula = Uhats ~ Phats)

Residuals:
   Min     1Q Median     3Q    Max
-1.431 -0.796  0.068  0.257  9.458

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.061      0.139    0.44     0.66
Phats          0.190      0.170    1.12     0.27

Residual standard error: 1.18 on 228 degrees of freedom
Multiple R-squared:  0.00543,Adjusted R-squared:  0.00107
F-statistic: 1.24 on 1 and 228 DF,  p-value: 0.266
```

$\rightarrow$ no particular evidence of overdispersion here. However...

```
> library(MASS)
> nulls.NB<-glm.nb(NNulls~Age+Tenure+Unified,data=NewDahl)
> summary(nulls.NB)

Coefficients:
            Estimate Std. Error z value      Pr(>|z|)
(Intercept) -12.4061     1.9829   -6.26 0.00000000039 ***
Age           0.1970     0.0342    5.75 0.00000000876 ***
Tenure       -0.0486     0.0408   -1.19          0.23
Unified      -0.0467     0.2145   -0.22          0.83
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for Negative Binomial(2.41) family taken to be 1)

    Null deviance: 261.45  on 229  degrees of freedom
Residual deviance: 204.57  on 226  degrees of freedom
AIC: 478.2

Number of Fisher Scoring iterations: 1

              Theta:  2.41
          Std. Err.:  1.14

 2 x log-likelihood:  -468.21


> # alpha:
> 1 / nulls.NB$theta
[1] 0.416
```

# Alternative NB Regression

```
> library(msme)
> nulls.nb2<-nbinomial(NNulls~Age+Tenure+Unified,data=NewDahl)
> summary(nulls.nb2)

Call:
ml_glm2(formula1 = formula1, formula2 = formula2, data = data,
    family = family, mean.link = mean.link, scale.link = scale.link,
    offset = offset, start = start, verbose = verbose)

Deviance Residuals:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -1.674  -0.986  -0.508  -0.289   0.336   2.390

Pearson Residuals:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  -1.03   -0.66   -0.37   -0.01    0.37    4.06

Coefficients (all in linear predictor):
              Estimate     SE      Z        p      LCL     UCL
(Intercept)   -12.2623 2.0203 -6.070 1.28e-09 -16.2221 -8.3026
Age             0.1944 0.0352  5.517 3.45e-08   0.1253  0.2634
Tenure         -0.0465 0.0428 -1.088    0.277  -0.1304  0.0373
Unified        -0.0490 0.2160 -0.227     0.82  -0.4725  0.3744
(Intercept)_s   0.4161 0.1987  2.094   0.0363   0.0266  0.8057

Null deviance: 261  on  228 d.f.
Residual deviance: 205  on  225 d.f.
Null Pearson: 293  on  228 d.f.
Residual Pearson: 219  on  225 d.f.
Dispersion: 0.975
AIC:  478

Number of optimizer iterations:  78
```

See here for details...

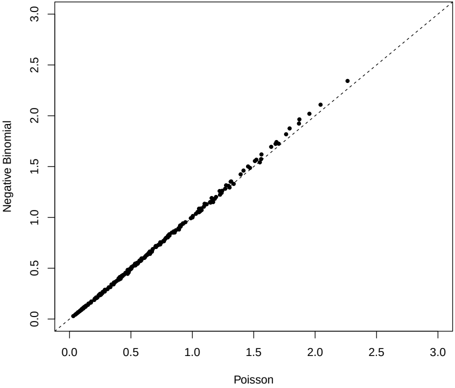
# Comparing Estimates

```
> # Coefficient estimates:
>
> cbind(nulls.poisson$coefficients,coef(nulls.NB))
                [,1]     [,2]
(Intercept) -12.0321 -12.4061
Age           0.1897   0.1970
Tenure       -0.0421  -0.0486
Unified      -0.0336  -0.0467
```

```
> # Estimated standard errors:
>
> cbind(diag(sqrt(vcov(nulls.poisson))),diag(sqrt(vcov(nulls.NB))))
              [,1]   [,2]
(Intercept) 1.7258 1.9829
Age         0.0296 0.0342
Tenure      0.0350 0.0408
Unified     0.1808 0.2145
```

# Predicted Values: Poisson and NB

```
> plot(nulls.poisson$fitted.values,nulls.NB$fitted.values,pch=20,
+       xlab="Poisson",ylab="Negative Binomial",main="",
+       xlim=c(0,3),ylim=c(0,3))
> abline(a=0,b=1,lwd=1,lty=2)
```



## More Things

- Models where Over- / Underdispersion $= f(\mathbf{Z}_i\gamma)$
- Models for Censored / Truncated Counts
- "Zero-Inflated" and "Hurdle" Count Models
- Time Series of Counts
- Panel Data Models for Event Counts
- Spatial Count Models
- etc...