Models for count data: Poisson and negative binomial

Quantitative Methods II for Political Science
Kenneth Benoit

March 4, 2009

More on maxmimum likelihood and `glm()`

- Logit model is `family=binomial(link="logit")`
- Probit model is `family=binomial(link="probit")`
- Normal model is `family=gaussian` (this is the default)
- Count model will be `family=poisson`

```
> ## classical OLS versus MLE linear-normal
> dail <- read.dta("dailcorrected.dta")
> summary(lm(votes1st ~ spend_total*incumb, data=dail))

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                  473.60391  162.45736   2.915  0.00373 **
spend_total                    0.19999    0.01166  17.147  < 2e-16 ***
incumbIncumbent             4461.74595  475.83240   9.377  < 2e-16 ***
spend_total:incumbIncumbent   -0.10328    0.02255  -4.580 5.99e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1804 on 459 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared: 0.6631, Adjusted R-squared: 0.6609
F-statistic: 301.2 on 3 and 459 DF,  p-value: < 2.2e-16

> summary(glm(votes1st ~ spend_total*incumb, data=dail), family=gaussian)

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                  473.60391  162.45736   2.915  0.00373 **
spend_total                    0.19999    0.01166  17.147  < 2e-16 ***
incumbIncumbent             4461.74595  475.83240   9.377  < 2e-16 ***
spend_total:incumbIncumbent   -0.10328    0.02255  -4.580 5.99e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 3255937)

    Null deviance: 4436367804  on 462  degrees of freedom
Residual deviance: 1494474988  on 459  degrees of freedom
  (1 observation deleted due to missingness)
AIC: 8263

Number of Fisher Scoring iterations: 2
```

Interpreting glm() output

- The "Dispersion parameter for gaussian family" is σ^2:

```
(Dispersion parameter for gaussian family taken to be 3255937)
> summary(m1.ols)$sigma^2
[1] 3255937
```

- We can replicate the OLS "F-test" using the likelihood ratio test:

```
> require(lmtest)
> lrtest(m1.mle)
Likelihood ratio test

Model 1: votes1st ~ spend_total * incumb
Model 2: votes1st ~ 1
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   5 -4126.5
2   2 -4378.4 -3 503.77  < 2.2e-16 ***
```

  The p-values are the same (and so are χ^2 and F tests, as $N \to \infty$)

- The AIC is $2k - 2\ln L$, and is a measure of the fit – smaller means better (used for comparing models)

More `glm()` interpretation

- Log-likelihoods are not reported by R as in some packages, but can extract these from `logLik()`
- AIC and deviance are closely related, both involve -2*logLik

```
> m1.mle$aic
[1] 8263.063
> -2*logLik(m1.mle)
[1] 8253.063
```

(Tricky: there are multiple ways to compute AIC...)

- The likelihood ratio test is similar to an F-test for linear models. Formula:

$$2 \ln \frac{L_L}{L_S}$$

- Likelihood ratio test in R: `lrtest(m1, m2)`

```
> summary(m2.mle <- glm(votes1st ~ spend_total*incumb + electorate + senator,
+                       data=dail))

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                 4.879e+02  4.131e+02   1.181    0.238
spend_total                 2.014e-01  1.184e-02  17.016  < 2e-16 ***
incumbIncumbent             4.462e+03  4.770e+02   9.356  < 2e-16 ***
electorate                 -2.228e-04  5.294e-03  -0.042    0.966
senator                    -5.534e+02  6.544e+02  -0.846    0.398
spend_total:incumbIncumbent -1.047e-01  2.265e-02  -4.622 4.94e-06 ***

(Dispersion parameter for gaussian family taken to be 3265077)
    Null deviance: 4436367804  on 462  degrees of freedom
Residual deviance: 1492140066  on 457  degrees of freedom
  (1 observation deleted due to missingness)
AIC: 8266.3

> require(lmtest)
> lrtest(m1.mle, m2.mle)
Likelihood ratio test

Model 1: votes1st ~ spend_total * incumb
Model 2: votes1st ~ spend_total * incumb + electorate + senator
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   5 -4126.5
2   7 -4126.2  2 0.7239     0.6963
> anova(m1.ols, m2.ols)
Analysis of Variance Table

Model 1: votes1st ~ spend_total * incumb
Model 2: votes1st ~ spend_total * incumb + electorate + senator
  Res.Df        RSS  Df  Sum of Sq      F Pr(>F)
1    459 1494474988
2    457 1492140066   2    2334922 0.3576 0.6996
```

Odds ratios and logit interpretation

- Logit results may sometimes be reported as “odds ratios”
- Advantage: These perform the same function as “standardized” coefficients in that they generalize the unit of change
- Disadvantage: Not meaningful for all variable types (esp. continuous variables)
- These can be obtained by exponentiating the coefficients: $e^{\tilde{\beta}_k}$ for the kth variable instance
- Interpretation: For a one-unit increase of X_k, the odds of $Y = 1$ increases by $e^{\tilde{\beta}_k}$
- Note: Odds ratios for intercept cannot be interpreted
- Standard error for odds ratios is $s^{\phi} = \phi s$, where s is $\tilde{\beta}_{se}$ estimated by logit and $\phi = e^{\tilde{\beta}}$ (the odds ratio of the logit coefficient) – but better to use confidence intervals

Odds ratios for coefficients: example in Stata

```
. logit wonseat incumb spend_total spendXinc electorate, or nolog

Logistic regression                               Number of obs   =        463
                                                  LR chi2(4)      =     239.99
                                                  Prob > chi2     =     0.0000
Log likelihood = -181.55667                       Pseudo R2       =     0.3979

------------------------------------------------------------------------------
     wonseat | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      incumb |   23.91585   20.15613     3.77   0.000     4.584567    124.7594
 spend_total |   1.000163   .0000235     6.96   0.000     1.000117    1.000209
   spendXinc |   .9999358    .000043    -1.49   0.135     .9998516     1.00002
  electorate |   .9999912   8.33e-06    -1.06   0.290     .9999749    1.000008
------------------------------------------------------------------------------
```

Odds ratios for coefficients: example in R

```
> require(foreign)
> dail <- read.dta("dailcorrected.dta")
> won.logit <- glm(wonseat ~ incumb*spend_total + electorate,
+                  family=binomial(link="logit"), data=dail)
> summary(won.logit)$coefficients
                                Estimate   Std. Error   z value     Pr(>|z|)
(Intercept)                -3.299165e+00 7.054731e-01 -4.676528 2.917722e-06
incumbIncumbent             3.174541e+00 8.427698e-01  3.766795 1.653564e-04
spend_total                 1.631863e-04 2.344526e-05  6.960312 3.395191e-12
electorate                 -8.813909e-06 8.328096e-06 -1.058334 2.899031e-01
incumbIncumbent:spend_total -6.421803e-05 4.297991e-05 -1.494141 1.351388e-01
> exp(summary(won.logit)$coefficients)
                              Estimate Std. Error      z value Pr(>|z|)
(Intercept)                 0.03691398   2.024804 9.311283e-03 1.000003
incumbIncumbent            23.91584817   2.322792 4.324127e+01 1.000165
spend_total                 1.00016320   1.000023 1.053963e+03 1.000000
electorate                  0.99999119   1.000008 3.470334e-01 1.336298
incumbIncumbent:spend_total  0.99993578   1.000043 2.244414e-01 1.144696
> exp(confint(won.logit))
Waiting for profiling to be done...
                                  2.5 %      97.5 %
(Intercept)                 0.008838667   0.1415823
incumbIncumbent             4.423489663 122.4305278
spend_total                 1.000119830   1.0002122
electorate                  0.999974731   1.0000075
incumbIncumbent:spend_total 0.999854251   1.0000237
```

When dependent variables are counts

- Many dependent variables of interest in political science may be in the form of counts of discrete events– examples:
  - international wars or conflict events
  - presidential appointments to the US Supreme Court
  - the number of coups d'état
- Characteristics: these Y are bounded between $(0, \infty)$ and take on only discrete values $0, 1, 2, \ldots, \infty$
- Imagine a social system that produces events randomly during a fixed period, and at the end of this period only the total count is observed. For N periods, we have $y_1, y_2, \ldots, y_N$ observed counts
- As with the binary dependent variable case, we need to transform both the error assumption (away from normality) and the functional form (away from linearity)

Event count model basic assumptions

First principles:

1. The probability that two events occur at precisely the same time is zero
2. During each period i, the event rate occurence λ_i remains constant and is independent of all previous events during the period
   - note that this implies no *contagion* effects
   - also known as *Markov independence*
3. Zero events are recorded at the start of the period
4. All observation intervals are equal over i

If these assumptions hold, we can model the counts as generated by a Poisson distribution:

$$f_{Poisson}(y_i|\lambda) = \begin{cases} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!} & \forall\ \lambda > 0 \text{ and } y_i = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

The Poisson distribution

$$
\begin{aligned}
f_{Poisson}(y_i|\lambda) &= \begin{cases} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!} & \forall\ \lambda > 0 \text{ and } y_i = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases} \\
Pr(Y|\lambda) &= \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!} \\
\lambda &= e^{X_i\beta} \\
\mathrm{E}(y_i) &= \lambda \\
\mathrm{Var}(y_i) &= \lambda
\end{aligned}
$$

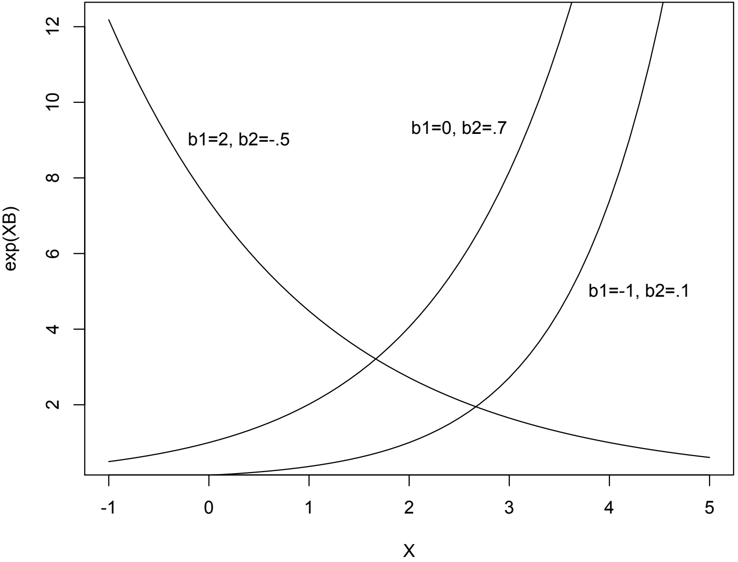
Systematic component

- $\lambda_i > 0$ is only bounded from below (unlike π_i)
- This implies that the effect cannot be linear
- Hence for the functional form we will use an exponential transformation

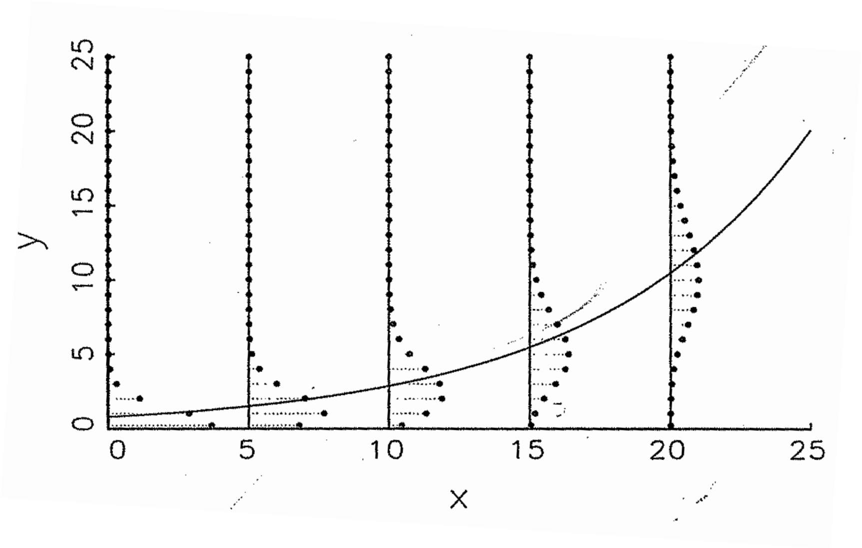
$$E(Y_i) = \lambda_i = e^{X_i\beta}$$

- Other possibilities exist, but this is by far the most common – indeed almost universally used – functional form for event count models

Exponential link function



Exponential link function



Likelihood for Poisson

$$
\begin{aligned}
\mathrm{L}(\lambda|y) &= \prod_{i=1}^{N} \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \\
\ln \mathrm{L}(\lambda|y) &= \sum_{i=1}^{N} \ln\left[\frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}\right] \\
&= \sum_{i=1}^{N} \left\{ \ln e^{-\lambda_i} + \ln(\lambda_i^{y_i}) + \ln\left(\frac{1}{y_i!}\right) \right\} \\
&= \sum_{i=1}^{N} \{ -\lambda_i + y_i \ln(\lambda_i) - \ln(y_i!) \} \\
&= \sum_{i=1}^{N} \left\{ -e^{X_i\beta} + y_i \ln e^{X_i\beta} - \ln y_i! \right\} \\
&\propto \sum_{i=1}^{N} \left\{ -e^{X_i\beta} + y_i X_i \beta - dropped \right\} \\
\ln L(\beta|y) &\propto \sum_{i=1}^{N} \left\{ X_i \beta y_i - e^{X_i\beta} \right\}
\end{aligned}
$$

The Negative Binomial model

- Generalize the Poisson model to:

$$f_{nb}(y_i|\lambda_i, \sigma^2) \text{ where :}$$

  - σ^2 is the variability (a new parameter v. Poisson)
  - λ_i is the expected number of events for i
  - λ is the average of individual λ_is
- Here we have dropped Poisson assumption that $\lambda_i = \lambda \ \forall \ i$
- New assumption: Assume that λ_i is a random variable following a *gamma* distribution (takes on only non-negative numbers)
- For the NB model, $\text{Var}(Y_i) = \lambda_i \sigma^2$ for $\lambda_i > 0$ and $\sigma^2 > 0$

The Negative Binomial model cont.

- For the NB model, $\text{Var}(Y_i) = \lambda_i \sigma^2$ for $\lambda_i > 0$ and $\sigma^2 > 0$
- How to interpret σ^2 in the negative binomial
  - when $\sigma^2 = 1.0$, negative binomial $\equiv$ Poisson
  - when $\sigma^2 > 1$, then it means there is overdispersion in Y_i caused by correlated events, or heterogenous λ_i
  - when $\sigma^2 < 1$ it means something strange is going on
- When $\sigma^2 \neq 1$, then Poisson results will be inefficient and standard errors inconsistent
- Functional form: same as Poisson

$$\text{E}(y_i) = \lambda$$

- Variance of λ is now:

$$\text{Var}(y_i) = \lambda_i \sigma^2 = e^{X_i \beta} \sigma^2$$

Poisson and negative binomial example

[switch to R for example here]

Negative binomial likelihood

$$f_{nb}(y_i \mid \lambda_i, \sigma^2) = \frac{\Gamma\left(\frac{\lambda_i}{\sigma^2-1} + y_i\right)}{y_i!\,\Gamma\left(\frac{\lambda_i}{\sigma^2-1}\right)} \left(\frac{\sigma^2-1}{\sigma^2}\right)^{y_i} \sigma^{2\,\frac{-\lambda_i}{\sigma^2-1}}$$

$\lambda_i > 0$

$\sigma^2 > 1$

$\Gamma(\cdot)$ is gamma function

$$E(Y_i) \equiv \lambda_i = e^{x_i\beta}$$

$$V(Y_i) = \lambda_i \sigma^2 = e^{x_i\beta}\sigma^2$$

as $\sigma^2 \to 1$, approximates Poisson

$$\ln L(\beta, \sigma^2 \mid y) = \sum_{i=1}^{n} \left\{ \ln\Gamma\left(\frac{\lambda_i}{\sigma^2-1} + y_i\right) - \ln\Gamma\left(\frac{\lambda_i}{\sigma^2-1}\right) + y_i \ln(\sigma^2-1) - \ln(\sigma^2)\left(y_i + \frac{\lambda_i}{\sigma^2-1}\right) \right\}$$