Models for count data: Poisson and negative binomial

ME104: Linear Regression Analysis – Kenneth Benoit

August 23, 2012

Class and homework last week

```
. mlogit valueposition us age higheduc highrelig usXhighrelig, base(1)
Multinomial logistic regression    Number of obs   =       1594
                                   LR chi2(15)     =     305.27
                                   Prob > chi2     =     0.0000
Log likelihood = -1892.5991        Pseudo R2       =     0.0746

------------------------------------------------------------
valueposit~n |      Coef.   Std. Err.      z    P>|z|
-------------+----------------------------------------------
sc_populism  |
          us |    .961763   .2110147     4.56   0.000
         age |  -.0006348   .0043677    -0.15   0.884
    higheduc |  -.6378294   .1624709    -3.93   0.000
   highrelig |   .3030148   .2512754     1.21   0.228
usXhighrelig |   1.046964   .3177656     3.29   0.001
       _cons |  -1.409909   .2428747    -5.81   0.000
-------------+----------------------------------------------
moral elit~m |
          us |  -.8423041    .210285    -4.01   0.000
         age |    .011921   .0040808     2.92   0.003
    higheduc |   .0111539   .1494356     0.07   0.941
   highrelig |   .2244204   .1820275     1.23   0.218
usXhighrelig |  -.5558658   .3501997    -1.59   0.112
       _cons |  -1.184776   .2222915    -5.33   0.000
-------------+----------------------------------------------
(valueposition==scientific elitism is the base outcome)
```

Class and homework last week

```
. mlogit valueposition us age higheduc highrelig usXhighrelig, base(1)

------------------------------------------------------------
valueposit~n |      Coef.   Std. Err.      z    P>|z|
-------------+----------------------------------------------
moral popu~m |
          us |  -.9441626    .279601    -3.38   0.001
         age |  -.0000473   .0046199    -0.01   0.992
    higheduc |  -.2209183   .1718877    -1.29   0.199
   highrelig |   .0151649   .2281207     0.07   0.947
usXhighrelig |   1.634334   .3713226     4.40   0.000
       _cons |  -1.068606   .2435958    -4.39   0.000
------------------------------------------------------------
(valueposition==scientific elitism is the base outcome)
```

Class and homework last week

- If we denote by π_{SE} the probability of Scientific Elitism, and π_{SP}, π_{ME}, and π_{MP} the probabilities of the other three value positions, the estimated model states that

$$
\begin{aligned}
\hat{L}_{\text{SP}} &= \log(\hat{\pi}_{\text{SP}}/\hat{\pi}_{\text{SE}}) \\
&= -1.41 + 0.96U - 0.006A - 0.64E + 0.30R + 1.04(UR) \\
\hat{L}_{\text{ME}} &= \log(\hat{\pi}_{\text{ME}}/\hat{\pi}_{\text{SE}}) \\
&= -1.18 - 0.84U + 0.01A + 0.01E + 0.22R - 0.56(UR) \\
\hat{L}_{\text{MP}} &= \log(\hat{\pi}_{\text{MP}}/\hat{\pi}_{\text{SE}}) \\
&= -1.07 - 0.94U - 0.00005A - 0.22E + 0.02R + 1.63(UR)
\end{aligned}
$$

  where U denotes "US", A "Age", E "Higheduc" and R "Highrelig", and $UR = U \times R$

Class and homework last week

- For example, the coefficient of High education in the model for $\log(\pi_{\mathsf{SP}}/\pi_{\mathsf{SE}})$ is -0.638, and $\exp(-0.638) = 0.53$.
- Thus the odds of selecting Scientific populism rather than Scientific elitism are 47% lower (as $1 - 0.53 = 0.47$) for a respondent with high level of education than for a respondent with a low level of education, controlling for age, region and religiosity.
- The 95% confidence interval for this odds ratio is (0.37; 0.73).

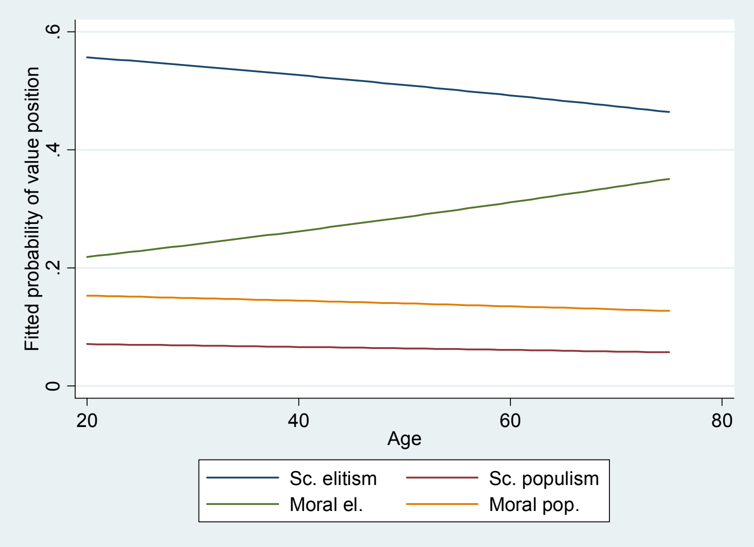
Class and homework last week

- Fitted probabilities are useful as part of the interpretation of the model:

$$
\begin{aligned}
\hat{\pi}_{SE} &= 1/[1+\exp(\hat{L}_{SP})+\exp(\hat{L}_{ME})+\exp(\hat{L}_{MP})] \\
\hat{\pi}_{SP} &= \exp(\hat{L}_{SP})/[1+\exp(\hat{L}_{SP})+\exp(\hat{L}_{ME})+\exp(\hat{L}_{MP})] \\
\hat{\pi}_{ME} &= \exp(\hat{L}_{ME})/[1+\exp(\hat{L}_{SP})+\exp(\hat{L}_{ME})+\exp(\hat{L}_{MP})] \\
\hat{\pi}_{MP} &= \exp(\hat{L}_{MP})/[1+\exp(\hat{L}_{SP})+\exp(\hat{L}_{ME})+\exp(\hat{L}_{MP})]
\end{aligned}
$$

- These can be calculated at any values of the explanatory variables you choose to use for illustration
- Examples of fitted probabilities are shown on the following pages
  - given age, for a person who lives in Europe, has a high level of education and never attends religious services.
  - given combinations of the three binary explanatory variables, for a person aged 50

Class and homework last week



Class and homework last week

		Europe				US			
Education	**Religiosity**	*Sci elit.*	*Sci pop.*	*Mor. elit.*	*Mor. pop.*	*Sci elit.*	*Sci pop.*	*Mor elit.*	*Mor. pop.*
Low	Low	0.47	0.11	0.26	0.16	0.50	0.31	0.12	0.07
	High	0.42	0.14	0.29	0.15	0.23	0.56	0.04	0.16
High	Low	0.51	0.06	0.29	0.14	0.60	0.20	0.14	0.06
	High	0.47	0.08	0.33	0.13	0.33	0.42	0.06	0.19

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:

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

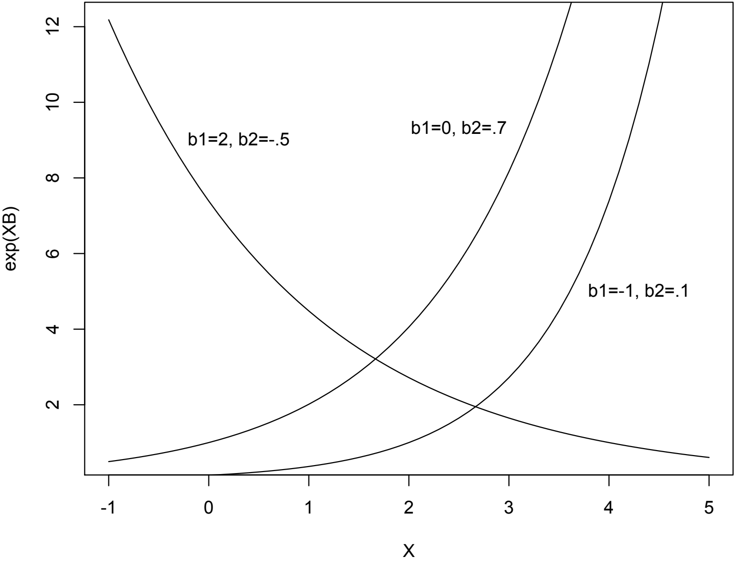
The Poisson distribution

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

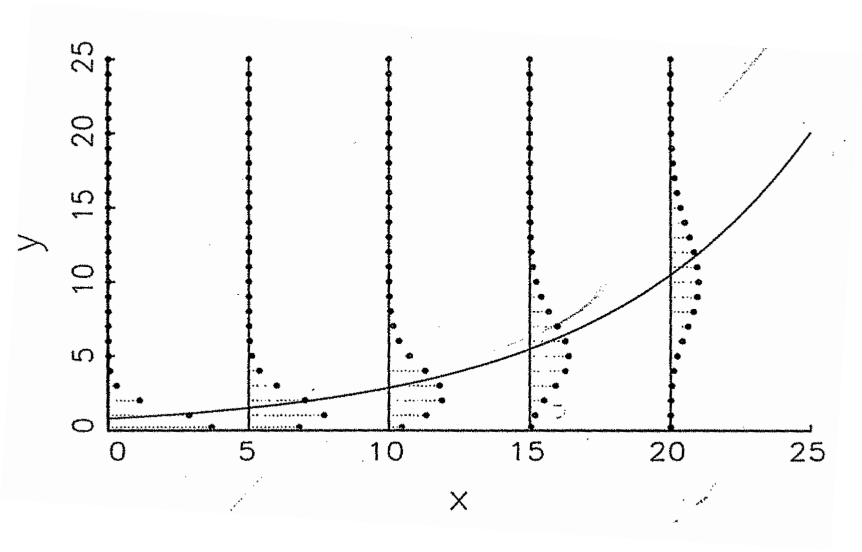
Systematic component

- $\lambda_i > 0$ 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

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

Predicted values from Benoit (1996)

```
> weede <- read.dta("weede.dta")
> z.out <- zelig(ssal6080 ~
+                 fh73+lpopln70+lmilwp70, model="poisson", data=weede)
> (x.out <- setx(z.out, fh73=2:14))
   (Intercept) fh73 lpopln70 lmilwp70
1            1    2    4.036    0.954
2            1    3    4.036    0.954
3            1    4    4.036    0.954
4            1    5    4.036    0.954
5            1    6    4.036    0.954
6            1    7    4.036    0.954
7            1    8    4.036    0.954
8            1    9    4.036    0.954
9            1   10    4.036    0.954
10           1   11    4.036    0.954
11           1   12    4.036    0.954
12           1   13    4.036    0.954
13           1   14    4.036    0.954
> s.out <- sim(z.out, x=x.out)
> summary(s.out)

  Model: poisson
  Number of simulations: 1000

Mean Values of X (n = 13)
(Intercept)        fh73     lpopln70     lmilwp70
      1.000       8.000        4.036        0.954

Pooled Expected Values: E(Y|X)
  mean     sd   2.5%  97.5%
0.3221 0.1085 0.1449 0.5697

Pooled Predicted Values: Y|X
  mean     sd   2.5%  97.5%
0.3259 0.5840 0.0000 2.0000
```

Replicate part of Table 3 from Benoit (1996)

```
> ## replicate part of Table 3 from Benoit (1996)
> z.tab2NBpoldem <- zelig(butterw ~ poldem65, model="negbin", data=weede)
> x.tab2NBpoldem <- setx(z.tab2NBpoldem, poldem65=c(0,20,55,85,100))
> s.tab2NBpoldem <- sim(z.tab2NBpoldem, x=x.tab2NBpoldem)
> cbind(apply(s.tab2NBpoldem$qi$ev, 2, mean),
+       apply(s.tab2NBpoldem$qi$ev, 2, sd))
       [,1]   [,2]
[1,] 1.7378 0.4969
[2,] 1.4819 0.3092
[3,] 1.1445 0.1644
[4,] 0.9364 0.1971
[5,] 0.8532 0.2290
> x.tab2NBfh73 <- setx(z.tab2NBfh73, fh73=c(2,4,7,12,14))
> s.tab2NBfh73 <- sim(z.tab2NBfh73, x=x.tab2NBfh73)
> cbind(apply(s.tab2NBfh73$qi$ev, 2, mean),
+       apply(s.tab2NBfh73$qi$ev, 2, sd))
       [,1]   [,2]
[1,] 1.4611 0.3421
[2,] 1.3210 0.2414
[3,] 1.1470 0.1709
[4,] 0.9308 0.2273
[5,] 0.8642 0.2674
```

TABLE 3

Fitted Values: Bivariate Negative Binomial Model

	Expected War Count			*Expected War Count*	
POLDEM 1965	*Butterworth*	*Small-Singer*	*Freedom House 1973*	*Butterworth*	*Small-Singer*
0	1.84	0.79	2	1.55	0.66
20	1.53	0.62	4	1.36	0.55
55	1.10	0.42	7	1.12	0.42
85	0.84	0.30	12	0.81	0.27
100	0.73	0.25	14	0.71	0.23
Mean *SE*	(0.27)	(0.14)		(0.23)	(0.11)

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 Stata 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

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