

# **Event Count Analysis**

**Tse-min Lin**  
**< tml@austin.utexas.edu >**

**Department of Government**  
**The University of Texas at Austin**

## **Types of Event Count**

### Finite Event Count

- Example: the number of positive responses to n survey items
- Model: binomial regression, (extended) beta-binomial regression
- Method: MLE

### (Potentially) Infinite Event Count

- Example: outbreaks of war in a given year; annual deaths from domestic conflict in a given country
- Model: Poisson regression, negative binomial regression
- Method: MLE

### **Problems of Using a Linear Model for Event Count Data**

- Wrong DGP: The dependent variable follows a non-normal, special distribution.
- Violation of classical assumptions (normality, homoscedasticity)
- OLS estimators are not BLUE.

**In this lecture, I will ...**

- Use a running example to walk you through the process of picking a particular distribution for event count analysis.
- Start from the Bernoulli distribution, and then progress to the Binomial distribution, Extended Beta-Binomial Distribution, Poisson distribution, the Negative Binomial Distribution, and their corresponding regression models.
- Show how event count models can be estimated with MLE.
- Compare (finite) event count models and IRT.

## The Running Example

- The problem will be determining the number of positive responses a respondent gives to  $n$  binary items in an opinion survey administered to  $m$  respondents.
- The response of a typical individual  $i$  to a typical binary item  $j$  is represented by the random variable  $X_{ij}$  which takes values on 0 or 1.
- The probability of an event occurring for individual  $i$  on item  $j$  is denoted by  $\pi_{ij}$ , i.e.,  $\Pr(X_{ij} = 1) = \pi_{ij}$ .
- The total number of positive responses of a typical individual  $i$  is represented by the random variable  $Y_i = \sum_{j=1}^n X_{ij}$ .

**Case 1: n=1 and  $\pi_{ij}=\pi$  for all i, j**

$$\Pr(Y_i = 1) = \Pr(X_{i1} = 1) = \pi_{i1} = \pi$$

$$\Pr(Y_i = 0) = \Pr(X_{i1} = 0) = 1 - \pi_{i1} = 1 - \pi$$

This is the Bernoulli distribution with the parameter  $\pi$ :

$$\Pr(Y_i = y_i) = \pi^{y_i} (1 - \pi)^{1-y_i} \quad y_i = 0,1$$

$$E(Y_i) = \pi$$

$$\text{Var}(Y_i) = \pi(1 - \pi)$$

Assumptions

- 1) There is only one item in the survey.
- 2)  $\pi$  is constant from individual to individual.

**Case 2: n is finite and  $\pi_{ij}=\pi$  for all i, j**

$$\Pr(Y_i = y_i | n, \pi) = \binom{n}{y_i} \pi^{y_i} (1 - \pi)^{n-y_i} \quad y_i = 0 \dots n$$

$$E(Y_i) = n\pi$$

$$\text{Var}(Y_i) = n\pi(1 - \pi)$$

This is the binomial distribution with parameters n and  $\pi$ .

For example, if there are three items in the survey (i.e., n=3), then

$$\Pr(Y_i = y_i | n, \pi) = \binom{3}{y_i} \pi^{y_i} (1 - \pi)^{3-y_i} \quad y_i = 0 \dots 3$$

**Assumptions**

- 1) There is a finite and discrete number of items.
- 2)  $\pi$  is constant from item to item.
- 3) Having a “1” on an item is independent of having a “1” on any other item for the same individual.
- 4)  $\pi$  is constant from individual to individual.

Note: For individual i, Assumptions (2) and (3) imply that the Bernoulli random variables  $X_{i1}, X_{i2}, \dots, X_{in}$  are identically and independently distributed (i.i.d.).

**Case 3: n is finite and  $\pi_{ij}=\pi_i$**

$Y_i$  follows the Binomial distribution for individual i:

$$\Pr(Y_i = y_i | n, \pi_i) = \binom{n}{y_i} \pi_i^{y_i} (1 - \pi_i)^{n-y_i} \quad y_i = 0 \dots n$$

If  $\pi_i$  is a random variable following the beta distribution (Gary King's version; see Appendix 1) with parameters  $\pi$  and  $\gamma$ :

$$f_\beta(\pi_i | \pi, \gamma) = \frac{\Gamma(\pi\gamma^{-1} + (1 - \pi)\gamma^{-1})}{\Gamma(\pi\gamma^{-1})\Gamma[(1 - \pi)\gamma^{-1}]} \pi_i^{\pi\gamma^{-1}-1} (1 - \pi_i)^{(1-\pi)\gamma^{-1}-1} \quad \text{with } E(\pi_i) = \pi$$

then the joint density of  $y_i$  and  $\pi_i$  are

$$f(y_i, \pi_i | n, \pi, \gamma) = \Pr(Y_i = y_i | n, \pi_i) f_\beta(\pi_i | \pi, \gamma)$$

We then calculate the marginal density of  $y_i$  by collapsing this joint distribution over  $\pi_i$ . The result is the extended beta-binomial distribution:

$$f_{ebb}(y_i | n, \pi, \gamma) = \int_{-\infty}^{\infty} f(y_i, \pi_i | n, \pi, \gamma) d\pi_i$$

The distribution of  $Y_i$  is then

$$\begin{aligned} \Pr(Y_i = y_i | n, \pi, \gamma) &= f_{ebb}(y_i | n, \pi, \gamma) \\ &= \frac{n!}{y_i!(n - y_i)!} \prod_{j=0}^{y_i-1} (\pi + \gamma) \prod_{j=0}^{n-y_i-1} (1 - \pi + \gamma) / \prod_{j=0}^{n-1} (1 + \gamma) \end{aligned}$$

with

$$\begin{aligned} E(Y_i) &= n\pi & 0 \leq \pi \leq 1 \\ \text{Var}(Y_i) &= n\pi(1 - \pi) \left[ 1 + (n - 1)\gamma(1 + \gamma)^{-1} \right] & \gamma \geq \max \{ -\pi(n - 1)^{-1}, -(1 - \pi)(n - 1)^{-1} \} \end{aligned}$$

Assumptions

- 1) There is a finite and discrete number of items in the survey.
- 2)  $\pi$  is constant from item to item for the same individual whose responses to the items are independent – i.e., the Bernoulli random variables are i.i.d. for the same individual – but changes from individual to individual randomly according to the beta distribution.



**Case 4: n is infinite and  $\pi_i=\pi$**

$$\lim_{n \rightarrow \infty} \Pr(Y_i = y_i | n, \pi) = \lim_{n \rightarrow \infty} \binom{n}{y_i} \pi^{y_i} (1 - \pi)^{n-y_i} = \frac{e^{-\lambda} \lambda^{y_i}}{y_i!}$$

where  $\lambda = n\pi$  is a constant. That is, as  $n \rightarrow \infty$ ,  $\pi \rightarrow 0$  such that  $n\pi$  remains a constant.

This is the Poisson distribution with parameter  $\lambda$  :

$$\Pr(Y_i = y_i | \lambda) = \frac{e^{-\lambda} \lambda^{y_i}}{y_i!} \quad y_i = 0 \dots \infty$$

$$E(Y_i) = \lambda$$

$$Var(Y_i) = \lambda$$

Assumptions

- 1) Only one event per item (i.e., “point” in a continuum).
- 2) n is infinite. There are many, many items.
- 3)  $\pi$  is constant from item to item for the same individual whose responses to the items are independent – i.e., the Bernoulli random variables are i.i.d. for the same individual – and constant from individual to individual.
- 4)  $\pi$  is infinitesimally small such that  $\lambda = n\pi$  is finite.

Example

Lewis F. Richardson, in his article “The Distribution of Wars in Time” (*Journal of Royal Statistical Society*, 107:242-250) compiled the following data on outbreaks of war from 1500 to 1931 A.D.

Number of outbreaks in the year	0	1	2	3	4	>4	Total
Number of such years	223	142	48	15	4	0	432
Binomial model	216	150	52	12	2	0	432
Poisson model	216	150	52	12	2	0	432

- Suppose every year has 365 days and each day can have at most one outbreak of war. Suppose further that whether a war breaks out on each day of the year is an iid Bernoulli process. Under these assumptions, the binomial model with  $n=365$  and  $\pi=0.00189625$  will be a good model to predict the number of outbreaks of war in a year.
- Richardson fitted a Poisson model with  $\lambda=0.69213$ .

### **Case 5: n is infinite and $\pi_{ij}=\pi_i$**

Let  $\lambda_i = n\pi_i$ . Then  $Y_i \sim \text{Poisson}(\lambda_i)$ .

If  $\lambda_i$  is a random variable following the gamma distribution (Gary King's version; see Appendix 2) with parameters  $\lambda$  and  $\sigma^2$ :

$$f(\lambda_i | \lambda, \sigma^2) = \frac{\lambda_i^{\lambda(\sigma^2-1)^{-1}-1} e^{-\lambda_i(\sigma^2-1)^{-1}}}{\Gamma[\lambda(\sigma^2-1)^{-1}](\sigma^2-1)^{\lambda(\sigma^2-1)^{-1}}} \quad \text{with } E(\lambda_i) = \lambda$$

then the joint density of  $y_i$  and  $\lambda_i$  are

$$f(y_i, \lambda_i | \lambda, \sigma^2) = f_p(y_i | \lambda_i) f_\gamma(\lambda_i | \lambda, \sigma^2)$$

We then calculate the marginal density of  $y_i$  by collapsing this joint distribution over  $\lambda_i$ . The result is the negative binomial distribution:

$$f_{nb}(y_i | \lambda, \sigma^2) = \int_0^{\infty} f_j(y_i, \lambda_i | \lambda, \sigma^2) d\lambda_i$$

The distribution of  $Y_i$  is then

$$\Pr(Y_i = y_i | \lambda, \sigma^2) = f_{nb}(y_i | \lambda, \sigma^2)$$

$$= \frac{\Gamma\left(\frac{\lambda}{\sigma^2-1} + y_i\right)}{y_i! \Gamma\left(\frac{\lambda}{\sigma^2-1}\right)} \left(\frac{\sigma^2-1}{\sigma^2}\right)^{y_i} (\sigma^2)^{\frac{-\lambda}{\sigma^2-1}}, \quad \lambda > 0; \sigma^2 > 1$$

$$E(Y_i) = \lambda$$

$$\text{Var}(Y_i) = \lambda\sigma^2$$

Assumptions

- 1) There are an infinite number of items.
- 2)  $\lambda$  is not constant but a random variable following the gamma distribution.

## Event Count Regression Models

In the above cases, if the expected values (e.g.,  $\pi$  of the Bernoulli, binomial, and extended beta binomial, and  $\lambda$  of the Poisson and negative binomial) are parameterized as varying *from individual to individual as a function of the characteristics of the individuals*, the resulting model will be probit/logit, binomial regression, extended beta binomial regression, Poisson regression, and negative binomial regression.

### Case 1a: Probit/Logit Model

$$\text{Probit: } \pi_i = \text{normal\_cdf}(x_i' \beta) = \Phi(x_i' \beta)$$

$$\text{Logit: } \pi_i = \text{logit\_cdf}(x_i' \beta) = \Lambda(x_i' \beta) = \frac{1}{1 + \exp(-x_i' \beta)} = \frac{\exp(x_i' \beta)}{\exp(x_i' \beta) + 1}$$

### Case 2a: Binomial Regression

$$\pi_i = \Phi(x_i' \beta) \text{ or } \pi_i = \Lambda(x_i' \beta)$$

### Case 3a: Extended Beta Binomial Regression

$$\pi_i = \Phi(x_i' \beta) \text{ or } \pi_i = \Lambda(x_i' \beta)$$

### Case 4a. Poisson Regression

$$\lambda_i = \exp(x_i' \beta)$$

### Case 5a. Negative Binomial Regression

$$\lambda_i = \exp(x_i' \beta)$$

## Extended Beta-Binomial Regression and Negative Binomial Regression

### Extended Beta-Binomial Regression

	Item 1	Item 2	Item 3	.....	Item n (Discrete Dimension)
Individual 1	$\square$	$\square$	$\square$		$\square$
	$\pi_1$	$\pi_1$	$\pi_1$		$\pi_1$
Individual 2	$\square$	$\square$	$\square$		$\square$
	$\pi_2$	$\pi_2$	$\pi_2$		$\pi_2$
.....					
Individual m	$\square$	$\square$	$\square$		$\square$
	$\pi_m$	$\pi_m$	$\pi_m$		$\pi_m$

$$\begin{aligned} \pi_i &\sim \text{beta}(\pi, \gamma) \\ E(\pi_i) &= f(x_i' \beta) \\ \text{Var}(\pi_i) &= f(x_i' \beta) (1 - f(x_i' \beta)) \gamma / (\gamma + 1) \end{aligned}$$

### Negative Binomial Regression

	Continuous Dimension (0 to $\infty$ )			
Individual 1	-----			
	$\lambda_1$	$\lambda_1$	$\lambda_1$	$\lambda_1$
Individual 2	-----			
	$\lambda_2$	$\lambda_2$	$\lambda_2$	$\lambda_2$
.....				
Individual m	-----			
	$\lambda_m$	$\lambda_m$	$\lambda_m$	$\lambda_m$

$$\begin{aligned} \lambda_i &\sim \text{gamma}(\lambda_i | \lambda, \sigma) \\ E(\lambda_i) &= g(x_i' \beta) \\ \text{Var}(\lambda_i) &= g(x_i' \beta) (\sigma^2 - 1) \end{aligned}$$

## Maximum Likelihood Estimation of Event Count Models

### Bernoulli Distribution

Let  $Y_1, Y_2, \dots, Y_n$  be a random sample of Bernoulli variables with

$$Y_i \sim \text{Bernoulli}(\pi)$$

$$f_{Y_i}(y_i | \pi) = \pi^{y_i} (1 - \pi)^{1-y_i} \quad y_i = 0, 1$$

The ML Estimator:

$$L_i(\pi) \propto \pi^{y_i} (1 - \pi)^{1-y_i}$$

$$\ln L_i(\pi) = y_i \ln(\pi) + (1 - y_i) \ln(1 - \pi)$$

$$\ln L(\pi) = \sum_{i=1}^n \ln L_i(\pi) = \sum_{i=1}^n y_i \ln(\pi) + \sum_{i=1}^n (1 - y_i) \ln(1 - \pi) = \sum_{i=1}^n y_i \ln(\pi) + (n - \sum_{i=1}^n y_i) \ln(1 - \pi)$$

$$\frac{d \ln(L(\pi))}{d\pi} = \frac{\sum_{i=1}^n y_i}{\pi} - \frac{n - \sum_{i=1}^n y_i}{1 - \pi} = \frac{\sum_{i=1}^n y_i - n\pi}{\pi(1 - \pi)} = 0$$

$$\text{Thus, } \sum_{i=1}^n y_i - n\pi = 0 \text{ and } \hat{\pi} = \frac{\sum_{i=1}^n y_i}{n} = \bar{y}.$$

### Stata Routine for Bernoulli Regression (Logit)

```
program define mylogit
  args lnf theta1
  tempvar pi y
  quietly gen double `pi' = 1/(1+exp(-`theta1'))
  quietly gen double `y' = $ML_y1
  quietly replace `lnf' = `y'*ln(`pi') + (1-`y')*ln(1-`pi')
end
ml model lf mylogit (y = x2 x3 x4)
ml maximize

logit y x2 x3 x4
```

## Binomial Distribution

Let  $Y_1, Y_2, \dots, Y_N$  be a random sample of binomial random variables with

$$Y_i \sim \text{Binomial}(y_i | n, \pi)$$

$$f_{Y_i}(y_i | n, \pi) = \binom{n}{y_i} \pi^{y_i} (1 - \pi)^{n - y_i} \quad y_i = 0, 1, 2, \dots, n$$

The ML estimator for  $\pi$  is  $\hat{\pi} = \frac{\bar{y}}{n}$ . That is,  $n\hat{\pi} = \bar{y}$ .

[Proof]

$$L_i(\pi) \propto \pi^{y_i} (1 - \pi)^{n - y_i}$$

$$\ln L_i(\pi) = y_i \ln(\pi) + (n - y_i) \ln(1 - \pi)$$

$$\ln L(\pi) = \sum_{i=1}^N L_i(\pi) = \left( \sum_{i=1}^N y_i \right) \ln(\pi) + (nN - \sum_{i=1}^N y_i) \ln(1 - \pi)$$

$$\frac{d \ln L(\pi)}{d\pi} = \left( \sum_{i=1}^N y_i \right) \left( \frac{1}{\pi} \right) + (nN - \sum_{i=1}^N y_i) \left( \frac{-1}{1 - \pi} \right) = \frac{\left( \sum_{i=1}^N y_i \right) (1 - \pi) - (nN - \sum_{i=1}^N y_i) \pi}{\pi(1 - \pi)} = 0$$

$$\left( \sum_{i=1}^N y_i \right) (1 - \pi) - (nN - \sum_{i=1}^N y_i) \pi = 0$$

$$\left( \sum_{i=1}^N y_i \right) - \left( \sum_{i=1}^N y_i \right) \pi - nN\pi + \left( \sum_{i=1}^N y_i \right) \pi = 0$$

$$\left( \sum_{i=1}^N y_i \right) - nN\pi = 0$$

$$nN\pi = \sum_{i=1}^N y_i$$

$$\text{Thus, } \hat{\pi} = \frac{\sum_{i=1}^N y_i}{nN} = \frac{y_1 + y_2 + \dots + y_N}{nN} = \frac{\bar{y}}{n}, \text{ which implies that } n\hat{\pi} = \bar{y}$$

### Stata Routine for Binomial Regression

```
program define mybinomial
  args lnf theta1
  tempvar y pi
  quietly gen double `y'=$ML_y1
  quietly gen double `pi' = exp(`theta1')/(1+exp(`theta1'))
  quietly replace `lnf' = ln(comb(n,`y'))+`y'*ln(`pi') + (n-`y')*ln(1-`pi')
end
ml model lf mybinomial (y = x2 x3 x4)
ml maximize
```

### Stata Routine for Beta-Binomial Regression

```
program define mybetabinomial
  args lnf theta1 theta2
  tempvar y pi gamma
  quietly gen double `y'=$ML_y1
  quietly gen double `pi' = exp(`theta1')/(1+exp(`theta1'))
  quietly gen double `gamma' = exp(`theta2')
  quietly replace `lnf' = ln(comb(n,`y'))
  local j=0
  while `j'<=(n-1) {
    quietly replace `lnf' = `lnf'+ln(`pi'+`gamma'*`j') if `j'<=(`y'-1)
    quietly replace `lnf' = `lnf'+ln(1-`pi'+`gamma'*`j') if (`j'<=n-`y'-1)
    quietly replace `lnf' = `lnf'-ln(1+`gamma'*`j')
    local j=`j'+1
  }
end
ml model lf mybetabinomial (y = x2 x3 x4) ()
ml maximize
```

## Poisson Distribution

Let  $Y_1, Y_2, \dots, Y_n$  be a random sample of Poisson variables with

$Y_i \sim \text{Poisson}(\lambda)$

$$f_{Y_i}(y_i | \lambda) = \frac{e^{-\lambda} \lambda^{y_i}}{y_i!} \quad y = 0, 1, 2, \dots \quad \lambda > 0$$

Note that  $E(y_i) = \text{Var}(y_i) = \lambda$ .

The ML estimator for  $\lambda$  is  $\hat{\lambda} = \bar{y}$ .

[Proof]

$$L_i(\lambda) \propto e^{-\lambda} \lambda^{y_i}$$

$$\ln(L_i(\lambda)) = -\lambda + y_i \ln(\lambda)$$

$$\ln(L(\pi)) = \sum_{i=1}^n L_i(\lambda) = -n\lambda + \ln(\lambda) \sum_{i=1}^n y_i$$

$$\frac{d \ln(L(\lambda))}{d\lambda} = -n + \frac{\sum_{i=1}^n y_i}{\lambda} = 0$$

$$\text{Thus, } -n\lambda + \sum_{i=1}^n y_i = 0 \text{ and } \hat{\lambda}_{MLE} = \frac{\sum_{i=1}^n y_i}{n} = \bar{y}.$$

## Stata Routine for Poisson Regression

```
program define mypoissonreg
  args lnf theta1
  tempvar lambda
  quietly gen double `lambda' = exp(`theta1')
  quietly replace `lnf' = -`lambda'+$ML_y1*ln(`lambda')-lnfact($ML_y1)
end
ml model lf mypoissonreg (y= x2 x3 x4)
ml maximize

poisson y x2 x3 x4
```



## Stata's Implementation of Negative Binomial Regression

In Stata, the negative binomial regression is derived by reparameterizing the Poisson parameter  $\lambda$  as

$$\lambda_i = \mu_i v_i$$

with  $\mu_i$  in turn reparameterized as

$$\mu_i = \exp(x_i' \beta)$$

and  $v_i$  following the gamma distribution with  $E(v_i) = 1$  and  $Var(v_i) = \alpha$ .

Equivalently,

$$\ln \lambda_i = \ln \mu_i v_i = \ln \mu_i + \ln v_i = x_i' \beta + \varepsilon_i$$

where  $v_i = \exp(\varepsilon_i)$  follows the gamma distribution with  $E(v_i) = 1$  and  $Var(v_i) = \alpha$ .

The resulting negative binomial distribution is

$$\Pr(Y_i = y_i) = \frac{\Gamma(\frac{1}{\alpha} + y_i)}{y_i! \Gamma(\frac{1}{\alpha})} \left( \frac{\alpha \mu_i}{1 + \alpha \mu_i} \right)^{y_i} \left( \frac{1}{1 + \alpha \mu_i} \right)^{\frac{1}{\alpha}}$$

with  $E(Y_i) = \mu_i = \exp(x_i' \beta)$  and  $Var(Y_i) = \mu_i(1 + \alpha \mu_i)$ .

Thus, if  $\alpha = 0$ , then  $Var(Y_i) = E(Y_i)$ , and negative binomial distribution collapses to Poisson distribution. If  $\alpha > 0$ , then  $Var(Y_i) > E(Y_i)$ , and negative binomial distribution is justified. This is called "over-dispersion." Therefore, testing  $\alpha$  against zero can be used to determine whether there is over-dispersion.

## Stata Command for Negative Binomial Regression

nbreg y x2 x3 x4

## Poisson/Negative Binomial Regression as Log-Linear Model

Let  $Y_i \sim \text{Poisson}(\lambda_i)$ .

$$\Pr(Y_i = y_i | \lambda_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \quad y_i = 0 \dots \infty \quad E(Y) = \lambda_i \quad \text{Var}(Y) = \lambda_i$$

Parameterizing  $\lambda_i$  as an exponential function of a linear combination of some covariates leads to Poisson regression:

$$\lambda_i = \exp(x_i' \beta)$$

$$\ln(\lambda_i) = \ln(E(Y_i | x_i)) = x_i' \beta$$

Adding a random component to the log-linear model leads to negative binomial regression:

$$\ln(\lambda_i) = \ln(E(Y_i | x_i)) = x_i' \beta + \varepsilon_i$$

where  $v_i = \exp(\varepsilon_i)$  follows the gamma distribution with  $E(v_i) = 1$  and  $\text{Var}(v_i) = \alpha$ .

### Event Rate and Offset

If each unit has a (known) different *exposure*  $T_i$ , e.g., different length of time period of different area of space in which events take place, then

$$Z_i = \frac{Y_i}{T_i}$$

is an “event rate” variable. To model this variable, note that

$$E(Z_i | x_i) = \frac{E(Y_i | x_i)}{T_i}$$

$$\ln(E(Z_i | x_i)) = \ln\left(\frac{E(Y_i | x_i)}{T_i}\right) = \ln(E(Y_i | x_i)) - \ln(T_i) = x_i' \beta$$

$$\ln(E(Y_i | x_i)) = \ln(T_i) + x_i' \beta$$

Thus, the term  $\ln(T_i)$  adjusts for exposure. It is called *offset*. Note that the coefficient of the term is constrained to 1. This can be done for the negative binomial regression model as well.

## Appendix 1. Beta Distribution

### Conventional Form

$$f(y | a, b) = \frac{\Gamma(a+b)y^{a-1}(1-y)^{b-1}}{\Gamma(a)\Gamma(b)} \quad 0 < y < 1 \quad a, b > 0$$

(Note:  $\Gamma$  is a gamma function:  $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$ )

$$\text{with } E(y) = \frac{a}{a+b}, \text{Var}(y) = \frac{ab}{(a+b)^2(a+b+1)}.$$

### Gary King's Parameterization

$$f_\beta(y | \rho, \gamma) = \frac{\Gamma(\rho\gamma^{-1} + (1-\rho)\gamma^{-1})}{\Gamma(\rho\gamma^{-1})\Gamma[(1-\rho)\gamma^{-1}]} y^{\rho\gamma^{-1}-1}(1-y)^{(1-\rho)\gamma^{-1}-1} \quad 0 < y < 1 \quad \rho, \gamma > 0$$

Comparing King's expression with the conventional form gives

$$a = \frac{\rho}{\gamma}, b = \frac{1-\rho}{\gamma}$$

Solving for  $\rho$  and  $\gamma$  gets

$$\rho = \frac{a}{a+b}, \gamma = \frac{1}{a+b}$$

Thus, for King's distribution,

$$E(y) = \frac{a}{a+b} = \rho, \text{Var}(y) = \frac{ab}{(a+b)^2(a+b+1)} = \frac{\rho(1-\rho)\gamma}{\gamma+1}$$

King adopts the particular form of the beta distribution so that one of the parameters is the expected value.

## Appendix 2. Gamma Distribution

### Conventional Form

$$f(y | r, \lambda) = \frac{\lambda^r y^{r-1} e^{-\lambda y}}{\Gamma(r)} \quad y > 0 \quad r, \lambda > 0$$

$$\text{with } E(y) = \frac{r}{\lambda}, \text{ Var}(y) = \frac{r}{\lambda^2}$$

### Gary King's Parameterization

$$f(y | \phi, \sigma^2) = \frac{y^{\phi(\sigma^2-1)^{-1}-1} e^{-y(\sigma^2-1)^{-1}}}{\Gamma[\phi(\sigma^2-1)^{-1}](\sigma^2-1)^{\phi(\sigma^2-1)^{-1}}} \quad y > 0 \quad \phi > 0 \quad \sigma > 1$$

Comparing King's expression with the conventional form gives

$$r = \frac{\phi}{\sigma^2 - 1}, \lambda = \frac{1}{\sigma^2 - 1}$$

Solving for  $\phi$  and  $\sigma^2$  gets

$$\phi = \frac{r}{\lambda}, \sigma^2 = \frac{\lambda + 1}{\lambda}$$

Thus, for King's distribution,

$$E(y) = \frac{r}{\lambda} = \frac{\frac{\phi}{\sigma^2 - 1}}{\frac{1}{\sigma^2 - 1}} = \phi$$

$$\text{Var}(y) = \frac{r}{\lambda^2} = \left(\frac{r}{\lambda}\right) * \left(\frac{1}{\lambda}\right) = \phi(\sigma^2 - 1)$$

King adopts the particular form of the gamma distribution so that one of the parameters is the expected value.

### Appendix 3. Event Count Models vs. Item Response Theory (IRT)

#### IRT Models as Mixed Effect Logit Models

According to Rijmen, Tuerlinckx, De Boeck and Kuppens (2003) and De Boeck and Wilson (2004), IRT can be represented as a mixed logit model.

$$\pi_{ij} = \frac{\exp(x_{ij}'\beta_j + z_{ij}'\theta_i)}{1 + \exp(x_{ij}'\beta_j + z_{ij}'\theta_i)}$$

where  $x_{ij}$  is a p-dimensional person-by-item covariates;  $z_{ij}$  is a q-dimensional vector of person-by-item covariates;  $\beta_j$  is a p-dimensional vector of item-specific fixed effects; and  $\theta_i$  is a q-dimensional vector of individual-specific random effects. Several special cases can be derived from this specification.

#### The Basic Rasch Model

If we assume  $p = q = 1$  and  $x_{ij} = z_{ij} = 1$ , then

$$\pi_{ij} = \frac{\exp(\beta_j + \theta_i)}{1 + \exp(\beta_j + \theta_i)}$$

This is the basic Rasch model where  $\beta_j$  is the difficulty parameter of item  $j$ , and  $\theta_i$  is a random variable the latent ability of individual  $i$ . A simple one-parameter IRT model, the basic Rasch model is a mixed-effect logistic regression model containing no covariates but only constant terms in  $x_{ij}$  and  $z_{ij}$ .

#### The Latent Rasch Regression Model with Individual-Specific Covariates for Fixed Effects

If we allow individual-specific covariates  $x_i$  for fixed effects in the basic Rasch model above, we derive a more general latent Rasch regression model:

$$\pi_{ij} = \frac{\exp(x_i'\beta_j + \theta_i)}{1 + \exp(x_i'\beta_j + \theta_i)}$$

A special case of this model is when all items share the same vector  $\beta_j$  (i.e.,  $\beta_j = \beta$  for  $j = 1, 2, 3, \dots, n$ ). In this case, we can drop the  $j$  subscript from  $\pi_{ij}$  to get

$$\pi_i = \frac{\exp(x_i'\beta + \theta_i)}{1 + \exp(x_i'\beta + \theta_i)}$$

This is a Rasch regression model for homogeneous items with both fixed effects and random effects.

### The Binomial Regression Model

Based on the Rasch model for homogeneous items above, if we further remove the random effect, we get

$$\pi_i = \frac{\exp(x_i' \beta)}{1 + \exp(x_i' \beta)}$$

which happens to be the binomial regression model with

$$\Pr(Y_i = y_i) = \binom{n}{y_i} \pi_i^{y_i} (1 - \pi_i)^{n - y_i}$$

with a logit link for  $\pi_i$ . Note that the constant term of the linear form  $x_i' \beta$  is the common difficulty parameter shared by all items. This shows that the binomial regression model is a special case of the latent Rasch regression model.

### The Latent Rasch Regression Model vs. the Extended Beta-Binomial Model

Recall that the Rasch latent regression model is

$$\pi_{ij} = \frac{\exp(x_i' \beta_j + \theta_i)}{1 + \exp(x_i' \beta_j + \theta_i)}$$

Taking the logit function on both sides gets

$$\text{logit}(\pi_{ij}) = \ln\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = x_i' \beta_j + \theta_i$$

where  $\theta_i$  is the (random) ability parameter generally assumed to follow a normal distribution with zero mean and constant variance:  $\theta_i \sim N(0, \sigma^2)$ . Thus,

$$\text{logit}(\pi_{ij}) \sim N(x_i' \beta_j, \sigma^2)$$

In contrast, the extended beta-binomial regression model assumes

$$\pi_i \sim B(\rho_i, \gamma) \text{ with } \rho_i = \frac{\exp(x_i' \beta)}{1 + \exp(x_i' \beta)} = \text{logit}^{-1}(x_i' \beta), \text{ or, } \pi_i \sim B(\text{logit}^{-1}(x_i' \beta), \gamma).$$

Clearly, the two models are similar in that both make distributional assumptions about the Bernoulli parameter  $\pi_{ij}$ . They are different in the specification of the assumptions.

Because of the limited nature of  $\pi_{ij}$  (i.e.,  $0 \leq \pi_{ij} \leq 1$ ), while Rasch regression transforms  $\pi_{ij}$  into  $\text{logit}(\pi_{ij})$  in order to use the unlimited normal distribution, extended beta-binomial regression directly lets  $\pi_i$  follow the limited beta distribution. By making such assumptions, both models allow conditional heterogeneity across units.

They are, however, different in terms of heterogeneity across items. The Rasch model, by allowing  $\beta_j$  to vary across items, directly introduces within-unit heterogeneity. The extended beta-binomial model does not have such a mechanism, although it does have the flexibility in accommodating within-unit heterogeneity (Lin, Nakamura, Morgan, and Helfer 2014).

#### Similarities and Differences between Extended Beta-Binoimial and Rasch Regressions

	Extended Beta-Binomial	Rasch Latent Regression
Specification of $\pi_{ij}$	$\pi_i \sim B(\text{logit}^{-1}(x_i' \beta), \gamma)$	$\text{logit}(\pi_{ij}) \sim N(x_i' \beta_j, \sigma^2)$
Heterogeneity across Units	Yes	Yes
Heterogeneity across Items	Flexible	Yes
Inter-Item Correlation in Data	Yes (+/-)	Yes (+)
Subsuming Binomial Model?	Yes	Yes
Special Case to the Other?	No	No

The Monte-Carlo simulation results shown in Tables 1 and 2 indicate:

- Both beta-binomial regression and Rasch regression can rather accurately estimate parameters when applied to data generated from their own DGPs.
- Both models can produce biased and inefficient estimates when applied to data generated from each other, but the problems caused by fitting data to the wrong model is more severe for Rasch regression than for extended beta-binomial regression.
- Using IRT models for event count data is not always a good strategy.

Table 1. Monte Carlo Simulation: Beta-Binomial DGP

	10 Items		5 Items	
	Beta-Binomial Regression Estimates		Beta-Binomial Regression Estimates	
Mean	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$
n=200	-0.204	0.304	-0.200	0.301
n=600	-0.202	0.297	-0.204	0.300
n=1000	-0.199	0.302	-0.201	0.298
RMSE	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$
n=200	0.074	0.071	0.082	0.080
n=600	0.040	0.042	0.046	0.047
n=1000	0.032	0.032	0.034	0.035
	Latent Rasch Regression Estimates		Latent Rasch Regression Estimates	
Mean	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$
n=200	-0.313	0.463	-0.303	0.456
n=600	-0.308	0.453	-0.308	0.454
n=1000	-0.304	0.460	-0.303	0.451
RMSE	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$
n=200	0.147	0.177	0.149	0.179
n=600	0.114	0.154	0.116	0.156
n=1000	0.105	0.160	0.106	0.151

Note: True parameter values:  $\beta_1 = -0.2$ ,  $\beta_2 = 0.3$



Table 2. Monte Carlo Simulation: Item Response Theory DGP

	10 Items		5 Items	
	Beta-Binomial Regression Estimates		Beta-Binomial Regression Estimates	
Mean	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$
n=200	-0.147	0.221	-0.140	0.210
n=600	-0.147	0.221	-0.142	0.210
n=1000	-0.146	0.222	-0.140	0.209
RMSE	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$
n=200	0.069	0.088	0.078	0.099
n=600	0.056	0.080	0.062	0.091
n=1000	0.054	0.078	0.061	0.091
	Latent Rasch Regression Estimates		Latent Rasch Regression Estimates	
Mean	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$
n=200	-0.202	0.302	-0.200	0.299
n=600	-0.201	0.302	-0.203	0.298
n=1000	-0.200	0.302	-0.200	0.298
RMSE	$\beta_1$	$\beta_2$	$\beta_1$	$\beta_2$
n=200	0.072	0.074	0.088	0.085
n=600	0.043	0.041	0.050	0.048
n=1000	0.031	0.032	0.039	0.037

Note: True parameter values:  $\beta_1 = -0.2$ ,  $\beta_2 = 0.3$