Distribuição de Probabilidade de Poisson

Ernesto F. L. Amaral
Magna M. Inácio

07 de outubro de 2010
Tópicos Especiais em Teoria e Análise Política:
Problema de Desenho e Análise Empírica (DCP 859B4)

DISTRIBUIÇÃO DE PROBABILIDADE DE POISSON

– A distribuição de probabilidade de Poisson é importante porque é usada para se descrever o comportamento de eventos raros (com pequenas probabilidades).

– Esta é uma distribuição de probabilidade discreta que se aplica a ocorrências de eventos ao longo de intervalos especificados.

– A variável aleatória *x* é o número de ocorrências do evento no intervalo.

– O intervalo pode ser de tempo, distância, área, volume ou alguma unidade similar.

– Probabilidade de ocorrência do evento *x* em um intervalo é:

$$P(x) = \frac{\mu^{x} * e^{-\mu}}{x!}, onde\ e \approx 2{,}71828$$

REQUISITOS PARA DISTRIBUIÇÃO DE POISSON

– A variável aleatória *x* é o número de ocorrências de um evento ao longo de algum intervalo.

– As ocorrências devem ser aleatórias.

– As ocorrências devem ser independentes umas das outras.

– As ocorrências devem ser uniformemente distribuídas sobre o intervalo em uso.

PARÂMETROS DA DISTRIBUIÇÃO DE POISSON

– Probabilidade de ocorrência do evento *x*:

$$P(x) = \frac{\mu^{x} * e^{-\mu}}{x!}, onde\ e \approx 2{,}71828$$

– Média na distribuição de Poisson:

$$\mu$$

– Desvio padrão na distribuição de Poisson:

$$\sigma = \sqrt{\mu}$$

POISSON ≠ BINOMIAL

– Uma distribuição de Poisson difere de uma distribuição binomial em alguns aspectos fundamentais.

– A distribuição binomial é afetada pelo tamanho n da amostra e pela probabilidade p, enquanto a distribuição de Poisson é afetada apenas pela média μ.

– Na distribuição binomial, os valores possíveis da variável aleatória x são 0, 1, ..., n. Porém, numa distribuição de Poisson, os valores possíveis de x são 0, 1, 2, ..., sem qualquer limite superior.

DEBILIDADE DO MODELO DE POISSON

– A regressão de Poisson leva em consideração a heterogeneidade observada (isto é, diferenças observadas entre os membros da amostra), ao especificar a taxa média (μ_i) como uma função de x_k's observados.

– Na prática, o modelo de Poisson raramente possui bom ajuste, devido à grande dispersão (*overdispersion*) dos dados.

– O modelo subestima a quantidade de dispersão na variável dependente.

– Com três variáveis independentes, o modelo de Poisson é:

$$\mu_i = exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})$$

MODELO BINOMIAL NEGATIVO

– O modelo de regressão binomial negativo trata desta debilidade do modelo de Poisson, ao adicionar um parâmetro *α* que reflete a heterogeneidade não-observada entre as observações.

– O modelo binomial negativo adiciona um erro (*ε*) que é assumido como não correlacionado com os *x*'s:

$$\tilde{\mu}_i = exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i)$$

$$\tilde{\mu}_i = exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})exp(\varepsilon_i)$$

$$\tilde{\mu}_i = exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3})\delta_i$$

– O modelo assume que $E(\delta)=1$, o que é similar a $E(\varepsilon)=0$, no modelo de mínimos quadrados ordinários. Temos então:

$$E(\tilde{\mu}) = \mu E(\delta) = \mu$$

VALORES DE ALFA NO MODELO BINOMIAL NEGATIVO

– Na distribuição binomial negativa, o parâmetro α determina o grau de dispersão das predições.

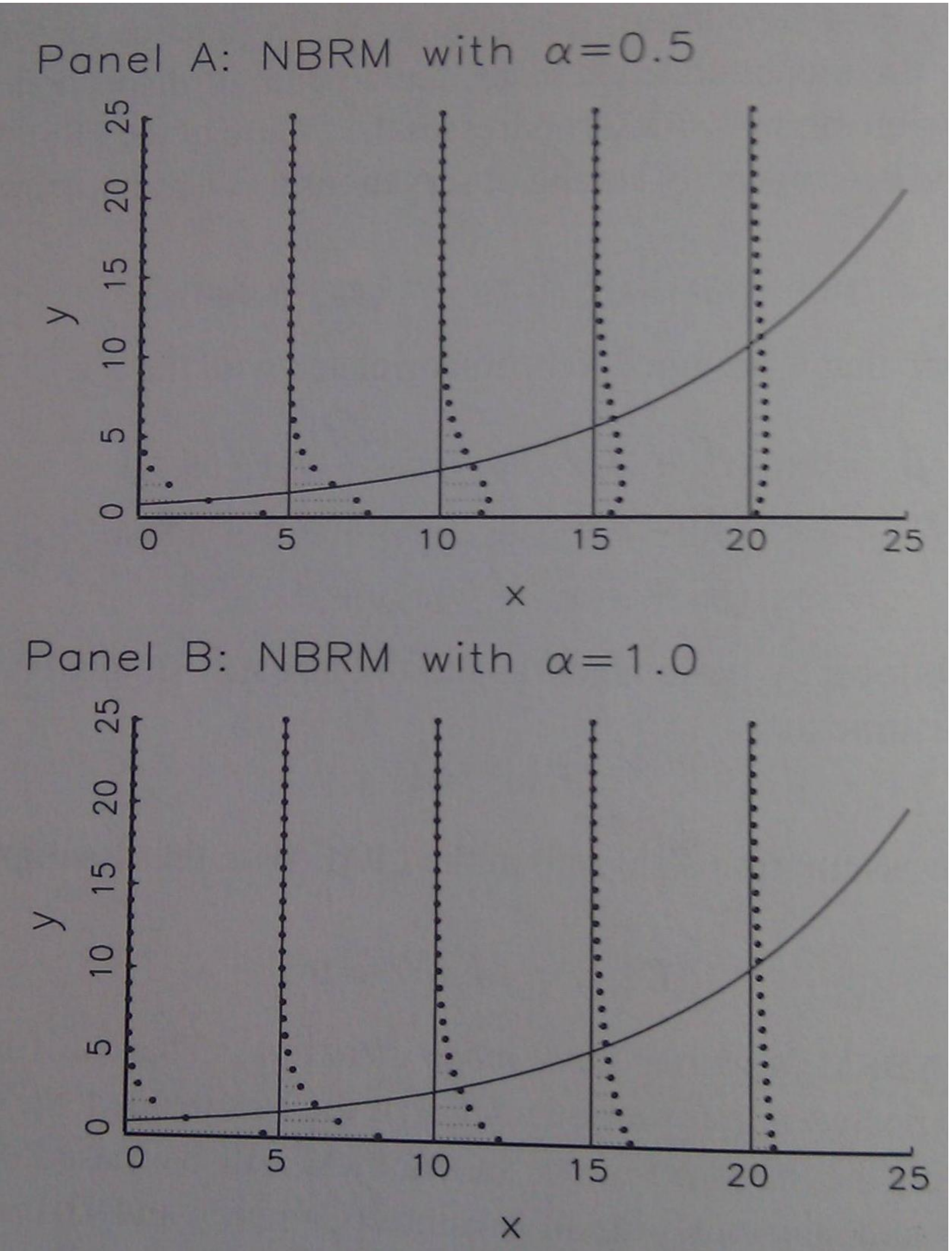
– A dispersão das contagens preditas para um determinado valor de x é maior do que no modelo de Poisson.

– Há uma maior probabilidade de contagem de zero.

– Maiores valores de α resultam em maior dispersão dos dados.

– Se α=0, o modelo binomial negativo se torna similar ao modelo de Poisson, o que acaba sendo o teste central para verificar sobre-dispersão (*overdispersion*).

EXEMPLOS DE VALORES DE ALFA



– Painel B possui maior valor de α, por isso há maior dispersão dos dados.

MODELOS DE CONTAGEM DE ZERO INFLACIONADO

– O modelo binomial negativo melhora a subestimação de zeros do modelo de Poisson, com o aumento da variância condicional (ε), sem mudar a média condicional (μ).

– Os modelos de contagem de zero inflacionado (*zero-inflated count models*) corrigem a falha do modelo de Poisson, ao levar em consideração a dispersão e excesso de zeros.

– Isto é realizado ao mudar a estrutura da média, permitindo que zeros sejam gerados em dois processos distintos.

DOIS GRUPOS NO MODELO DE ZERO INFLACIONADO

– O modelo zero inflacionado permite que um grupo de indivíduos tenha sempre probabilidade igual a um, ao aumentar a variância condicional e a probabilidade de contagem de zeros [P(0)=1].

– O modelo zero inflacionado assume que há dois grupos latentes (não-observados):

  – Um indivíduo no “grupo sempre-zero” tem um resultado zero com probabilidade igual a 1 [P(0)=1].

  – Um indivíduo no “grupo não-sempre-zero” pode ter um resultado zero, mas há uma probabilidade não-zero que haja uma contagem positiva:

$$0<P(0)<1 \quad \text{ou} \quad 0<P(>0)<1$$