

# 1 Background

Count data describe the number of times an event occurs and take on non-negative integer values. These data cannot be described by continuous distributions, such as the normal distribution; instead, discrete distributions must be used. This section describes various discrete distributions that can be used to build a count data version of the stochastic frontier model. First, unbounded distributions like the Poisson and negative binomial distributions are described and applications to generalized linear models are explored. Then, bounded distributions that can be used to model one-sided errors, such as the binomial and beta-binomial distributions, are detailed.

As a matter of notation, let  $y$  be a  $N \times 1$  vector of count data outcomes,  $X$  be a  $N \times k$  matrix of explanatory variables, and let  $\beta$  be a  $k \times 1$  vector of slope parameters, where  $N$  is the number of observations in the data and  $k$  is the number of explanatory variables. Further, let  $\mathcal{L}$  denote the likelihood function, and let  $i$  subscripts denote the  $i$ th observation/row of a variable.

## 1.1 Traditional Stochastic Frontier Models

A typical stochastic frontier model has the form

$$y_i = X_i\beta - \delta_i + \varepsilon_i, \quad (1)$$

where  $\varepsilon_i$  is a two-sided measurement error and  $\delta_i$  is a one-sided error representing inefficiency. We will make common assumptions that  $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$  and  $\delta_i \sim N^+(\mu_\delta, \sigma_\delta^2)$ . Two methods can be employed to compute the likelihood. First, the law of total probability can be used to integrate directly over  $\delta$ :

$$f(y_i|\beta, \sigma_\varepsilon, \mu_\delta, \sigma_\delta) = \int_{\delta_i} f(y_i|\beta, \sigma_\varepsilon, \delta_i) f(\delta_i|\mu_\delta, \sigma_\delta) d\delta_i. \quad (2)$$

## 1.2 Poisson and Negative Binomial Count Models

The Poisson distribution is commonly used to model count data outcomes. The Poisson distribution is characterized by the likelihood

$$\mathcal{L}(y_i|\lambda) = \frac{\lambda^{y_i} e^{-\lambda}}{y_i!}, \quad (3)$$

where  $\lambda > 0$  is a parameter. The outcome  $y_i$  can take on any non-negative integer value, and the Poisson distribution is therefore unbounded. A key feature of the Poisson distribution is that  $E[y_i] = \text{var}(y_i) = \lambda$ , leading to two consequences. First, a generalized linear model can be constructed by setting  $\lambda = \exp(X_i\beta)$ , and therefore  $E[y_i] = \exp(X_i\beta)$ . Second, since a single parameter controls both mean and variance, data can be over-dispersed (common) or under-dispersed (less common) relative to the model with no method of correction.

The negative binomial distribution provides one method of correcting for over-dispersion relative to the Poisson distribution. There are several parameterizations of the negative binomial distribution; in this document we will use the likelihood of the form

$$\mathcal{L}(y_i|\mu, \phi) = \binom{y_i + \phi - 1}{y_i} \left( \frac{\mu}{\mu + \phi} \right)^{y_i} \left( \frac{\phi}{\mu + \phi} \right)^\phi, \quad (4)$$

where  $\mu > 0$  and  $\phi > 0$  are parameters of the distribution. Under this formulation,  $E[y_i] = \mu$  and  $\text{var}(y_i) = \mu + \mu^2/\phi$ . Notice that  $\text{var}(y_i) > \mu$ , thereby accommodating data with higher variance than the Poisson distribution. As with the Poisson distribution, the negative binomial distribution can be made into a generalized linear model by setting  $\mu = \exp(X_i\beta)$ . The parameters  $\beta$  and  $\phi$  are then free to be estimated.

A comparison of Poisson and negative binomial mass functions is shown in fig. 1. Each of the distributions shown has the same mean, and the  $\phi$  parameter is adjusted in the negative binomial distribution to increase variance.

# Poisson and Negative Binomial Distributions for $\lambda = \mu = 10$

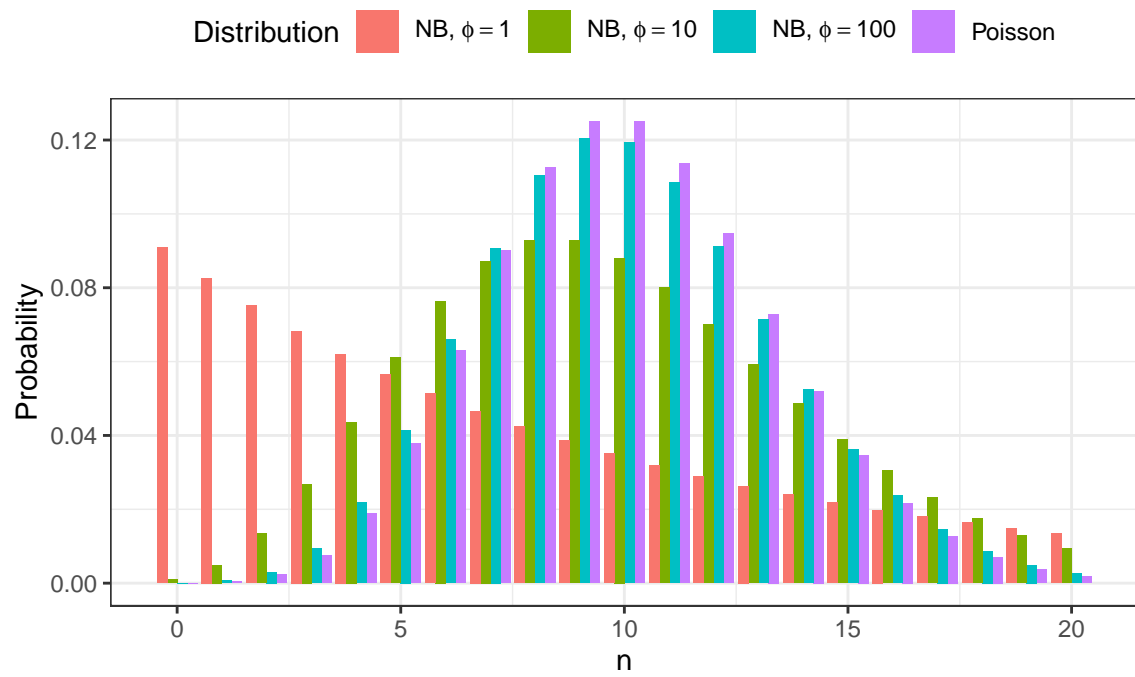


Figure 1: Comparison of Poisson and Negative Binomial Distributions