

## Introduction

---

The negative binomial is traditionally derived from a Poisson–gamma mixture model. However, the negative binomial may also be thought of as a member of the single parameter exponential family of distributions. This family of distributions admits a characterization known as generalized linear models (GLMs), which summarizes each member of the family. Most importantly, the characterization is applicable to the negative binomial. Such interpretation allows statisticians to apply to the negative binomial model the various goodness-of-fit tests and residual analyses that have been developed for GLMs.

Poisson regression is the standard method used to model count response data. However, the Poisson distribution assumes the equality of its mean and variance – a property that is rarely found in real data. Data that have greater variance than the mean are termed *Poisson overdispersed*, but are more commonly designated as simply *overdispersed*. Negative binomial regression is a standard method used to model overdispersed Poisson data.

When the negative binomial is used to model overdispersed Poisson count data, the distribution can be thought of as an extension to the Poisson model. Certainly, when the negative binomial is derived as a Poisson–gamma mixture, thinking of it in this way makes perfect sense. The original derivation of the negative binomial regression model stems from this manner of understanding it, and has continued to characterize the model to the present time.

As mentioned above, the negative binomial has recently been thought of as having an origin other than as a Poisson–gamma mixture. It may be derived as a generalized linear model, but only if its ancillary or heterogeneity parameter is entered into the distribution as a constant. The straightforward derivation of the model from the negative binomial probability distribution function (PDF) does not, however, equate with the Poisson–gamma mixture-based version of the negative binomial. Rather, one must convert the canonical link and inverse canonical link to log form. So doing produces a GLM-based negative binomial

that yields identical parameter estimates to those calculated by the mixture-based model. As a non-canonical linked model, however, the standard errors will differ slightly from the mixture model, which is typically estimated using a full maximum likelihood procedure. The latter uses by default the observed information matrix to produce standard errors. The standard GLM algorithm uses Fisher scoring to produce standard errors based on the expected information matrix – hence the difference in standard errors between the two versions of negative binomial. The GLM negative binomial algorithm may be amended though to allow production of standard errors based on observed information. When this is done, the amended GLM-based negative binomial produces identical estimates and standard errors to that of the mixture-based negative binomial. This form of negative binomial was called the *log-negative binomial* by Hilbe (1993a), and was the basis of a well-used SAS negative binomial macro (Hilbe, 1994b). It is also the form of the negative binomial found in Stata's *glm* command as well as in the SAS/STAT GENMOD procedure in SPSS's GLZ command, and in GENSTAT's GLM program.

Regardless of the manner in which the negative binomial is estimated, it is nevertheless nearly always used to model Poisson overdispersion. The advantage of the GLM approach rests in its ability to utilize the specialized GLM fit and residual statistics that come with the majority of GLM software. This gives the analyst the means to quantitatively test different modeling strategies with tools built into the GLM algorithm. This capability is rarely available with models estimated using full maximum likelihood or full quasi-likelihood methods.

In this book we shall discuss in greater depth the two methods of estimating negative binomial data that have been outlined above. The complete derivation of both methods will be given, together with discussion of how the algorithms may be altered to deal with count data that should not be modeled using simple Poisson or standard negative binomial methods. In fact, we shall devote considerable space to describing the base Poisson regression model, and the manner in which its assumptions may be violated. In addition, we shall find that just as Poisson models can be overdispersed, negative binomial models can as well. Following an examination of estimating methods and overviews of both the Poisson and negative binomial models, the remainder of the book is devoted to a discussion of how to understand and deal with various enhancements to both the Poisson and traditional negative binomial models.

Extensions to the respective Poisson and negative binomials are made depending on the type of underlying problem that is being addressed. Extended models include, among others, those for handling excessive response zeros – zero-inflated Poisson, zero-inflated negative binomial, and hurdle models; for

handling responses having no possibility of zero counts – zero-truncated Poisson and zero-truncated negative binomial; having responses with structurally absent values – truncated and censored Poisson and negative binomial; and having longitudinal or clustered data – fixed, random, and mixed effects negative binomial as well as negative binomial GEE. Models may also have to be devised for situations when the data can be split into two or more distributional subsets. In fact, both Poisson and negative binomial models have been extended to account for a great many count response modeling situations. We shall attempt to give an overview of each of the major varieties mentioned here, which should provide the researcher with a map or guideline of how to handle a wide variety of count modeling situations.

Typically, extensions to the Poisson model precede analogous extensions to the negative binomial. For example, statisticians have recently created random parameter and random intercept count models to deal with certain types of correlated data. The first implementations were based on the Poisson distribution. Nearly all literature dealing with random parameter count models relates to the Poisson. Negative binomial versions have only surfaced within the past couple of years, primarily as a result of the work of William Greene. The only software available for modeling negative binomial random parameter and intercept models is LIMDEP, and even at that, it has not yet been made part of its menu system procedures.

Of the two general count regression models, the negative binomial has greater generality. In fact, as will be discussed at greater length later in the text, the Poisson can be considered as a negative binomial with an ancillary or heterogeneity parameter value of zero. It seems clear that having an understanding of the various negative binomial models, basic as well as complex, is essential for anyone considering serious research dealing with count models.

It is important to realize that the negative binomial has been derived and presented with different parameterizations. Some authors employ a variance function that clearly reflects a Poisson–gamma mixture. With the Poisson variance defined as  $\mu$  and the gamma as  $\mu^2/\alpha$ , the negative binomial variance is then characterized as  $\mu + \mu^2/\alpha$ . The Poisson–gamma mixture is clear. This parameterization is the same as that originally derived by Greenwood and Yule (1920). An inverse relationship between  $\mu$  and  $\alpha$  was also used to define the negative binomial variance in McCullagh and Nelder (1989), to which some authors refer when continuing this manner of representation.

However, shortly after the publication of that text, Nelder developed his KK system (1992), a user-defined negative binomial macro written for use with Genstat software. In this system he favored the direct relationship between  $\alpha$  and  $\mu^2$  – resulting in a negative binomial variance function of  $\mu + \alpha\mu^2$ .

Nelder has continued to prefer the direct relationship in his subsequent writings (1994). Still, relying on the 1989 work, a few authors have continued to use the originally defined relationship, even as recently as Faraway (2006).

The direct parameterization of the negative binomial variance function was favored by Breslow (1984) and Lawless (1987) in their highly influential seminal articles on the negative binomial. In the decade of the nineties, the direct relationship was used in the major software implementations of the negative binomial: Hilbe (1993b, 1994a) – XploRe and Stata, Greene (2006) – LIMDEP, and Johnston (1997) – SAS. The direct parameterization was also specified in Hilbe (1994a), Long (1997), Cameron and Trivedi (1998), and most articles and books dealing with the subject. Recently Long and Freese (2003, 2006), Hardin and Hilbe (2001, 2007), and a number of other recent authors have employed the direct relationship as the preferred variance function. It is rare now to find current applications using the older inverse parameterization.

The reason for preferring the direct relationship stems from the use of the negative binomial in modeling overdispersed Poisson count data. Considered in this manner,  $\alpha$  is directly related to the amount of overdispersion in the data. If the data are not overdispersed, i.e. the data are Poisson, then  $\alpha = 0$ . Increasing values of  $\alpha$  indicate increasing amounts of overdispersion. Values for data seen in practice typically range from 0 to about 4.

Interestingly, two books have been recently published, Hoffmann (2004) and Faraway (2006), asserting that the negative binomial is not a true generalized linear model. However, the GLM status of the negative binomial depends on whether it is a member of the single-parameter exponential family of distributions. If we assume that the overdispersion parameter,  $\alpha$ , is known and is ancillary, resulting in what has been called a LIMQL (limited information maximum quasi-likelihood) model (see Greene, 2003), then the negative binomial is a GLM. On the other hand, if  $\alpha$  is considered to be a parameter to be estimated, then the model may be estimated as FIMQL (full information maximum quasi-likelihood), but it is not a GLM.

In this text, the negative binomial is estimated as both a GLM and as a full maximum (quasi-)likelihood model. As a GLM, the model has associated fit and residual statistics, which can be of substantial use during the modeling process. However, in order to obtain a value of  $\alpha$ , i.e. to make  $\alpha$  known, it must be estimated. The traditional, and most reasonable, method of estimating  $\alpha$  is by a non-GLM maximum likelihood algorithm. Extensions to the negative binomial model, e.g. zero-inflated, zero-truncated, and censored models, are nearly all based on FIMQL methods. I shall be using both methods of estimation when modeling basic Poisson and negative binomial data. How these

methods are used together will become apparent as we progress through the text.

The first chapter provides a brief overview of count response regression models. Incorporated in this discussion is an outline of the variety of negative binomial models that have been constructed from its basic parameterization. Each extension from the base model is considered as a response to a violation of model assumptions. We list seven types of violation to the standard negative binomial model. Enhanced negative binomial models are identified as solutions to the respective violations.

Chapter 2 examines the two major methods of parameter estimation relevant to modeling Poisson and negative binomial data. We begin by illustrating the construction of distribution-based statistical models. That is, starting from a probability distribution, we follow the logic of establishing the estimating equations that serve as the focus of the fitting algorithms. Given that the Poisson and traditional negative binomial, also referred to as NB-2, are members of the exponential family of distributions, we define the exponential family and its constituent terms. In so doing we derive the iteratively reweighted least squares (IRLS) algorithm and the form of the algorithm required to estimate the model parameters. Secondly, we define maximum likelihood estimation and show how the modified Newton–Raphson algorithm works in comparison to IRLS. We shall discuss the reason for differences in output between the two estimation methods, and explain when and why differences occur.

Chapter 3 is devoted to the derivation of the Poisson log-likelihood and estimating equations. The Poisson traditionally serves as the basis for deriving the negative binomial – at least for one variety of negative binomial. Regardless, Poisson regression is the fundamental method used to model counts. We identify how overdispersion is indicated from Poisson model output, and some of the methods used to deal with it. We also discuss the rate parameterization of the count models. We find that rates can be thought of in a somewhat analogous manner to the denominators in binomial models. There are important differences though – which we discuss. The subject relates to the topic of offsets.

Chapter 4 details the difference in real versus apparent overdispersion. Criteria are specified which can be used to distinguish real from apparent overdispersion. Simulated examples are constructed that show how apparent overdispersion can be eliminated. We show how overdispersion affects otherwise equi-dispersed data. Finally, scaling of standard errors, application of robust variance estimators, jackknifing, and bootstrapping of standard errors are

all evaluated in terms of their effect on inference. An additional section related to negative binomial overdispersion is provided, showing that overdispersion is a problem for all count models, not simply for Poisson models. This chapter is vital to the development of the negative binomial model.

In Chapter 5 we define the negative binomial probability distribution function (PDF) and proceed to derive the various statistics required to model the canonical and traditional form of the distribution. Additionally, we derive the Poisson–gamma mixture parameterization that is used in maximum likelihood algorithms. In this chapter it becomes clear that the negative binomial is a full member of the exponential family of generalized linear models. We discuss the nature of the canonical form, and the problems that have been claimed to emanate when applying it to real data. We then re-parameterize the canonical form of the model to derive the traditional log-linked form (NB-2).

In Chapter 6 we discuss the development and interpretation of the NB-2 model. Examples are provided that demonstrate how the negative binomial is used to accommodate overdispersed Poisson data. Goodness-of-fit statistics are examined, in particular methods used to determine whether the negative binomial fit is statistically different from a Poisson. Residuals appropriate to evaluation of a negative binomial analysis are derived and explained.

Chapter 7 addresses alternative parameterizations of the negative binomial. We begin with a discussion of the geometric model, a simplification of the negative binomial where the overdispersion parameter has a value of one. When the value of the overdispersion parameter is zero, NB-2 reduces to a Poisson model. The geometric distribution is the discrete correlate of the negative exponential distribution. We then address the interpretation of the canonical link derived in Chapter 5. We thereupon derive and discuss how the linear negative binomial, or NB-1, is best interpreted. Finally, the NB-2 model is generalized in the sense that the ancillary or overdispersion parameter itself is parameterized by user-defined predictors for generalization from scalar to observation-specific interpretation. NB-2 can also be generalized to parameterize the negative binomial exponent. This model is called the NB-P model.

Chapter 8 deals with a common problem faced by researchers handling real data. In many situations the data at hand exclude a zero count. Other data situations have an excessive number of zeros – far more than defined by usual count distributions.

Zero-truncated and zero-inflated Poisson (ZIP) and negative binomial (ZINB) models, as well as hurdle models, have been developed to accommodate these two types of data situations. Hurdle models are typically used when the data have excessive zero counts, much like zero-inflated models. There are differences, however. Detailed are logit, probit, and complementary loglog

negative binomial hurdle models. Finally, we examine negative binomial models having endogenous stratification.

Chapter 9 discusses truncated and censored data and how they are modeled using appropriately adjusted Poisson and negative binomial models. Two types of parameterizations are delineated for censored count models: econometric or dataset-based censored and survival, or observation-based censored, parameterizations.

The final chapter addresses the subject of negative binomial panel models. These models are used when the data are either clustered or when they are in the form of longitudinal panels. We derive and examine unconditional and conditional fixed effects and random effects Poisson and negative binomial regression models. Population averaged panel models, also referred to as generalized estimating equations (GEE) are also examined as are random intercept and random coefficient multilevel negative binomial models.

Several appendices are associated with the text. The titles are listed in the Contents.

# 1

## Overview of count response models

Count response models are a subset of discrete response regression models. Discrete models address non-negative integer responses. Examples of discrete models include:

### RESPONSE

- Binary: binary logistic and probit regression
- Proportional: grouped logistic, grouped complementary loglog
- Ordered: ordinal logistic and ordered probit regression
- Multinomial: discrete choice logistic regression
- Count: Poisson and negative binomial regression

A count response consists of any discrete response of counts, e.g. the number of hits recorded by a Geiger counter, patient days in the hospital, and goals scored at major contests. All count models aim to explain the number of occurrences, or counts, of an event. The counts themselves are intrinsically heteroskedastic, right skewed, and have a variance that increases with the mean of the distribution.

### 1.1 Varieties of count response model

Poisson regression is the basic count model upon which a variety of other count models are based. The Poisson distribution may be characterized as

$$f_y(y; \mu) = e^{-\mu} \mu^y / y!, \quad y = 0, 1, 2, \dots; \quad \mu > 0 \quad (1.1)$$

where the random variable  $y$  is the count response and parameter  $\mu$  is the mean. Often,  $\mu$  is also called the rate or intensity parameter. Unlike most other distributions, the Poisson does not have a distinct scale parameter. Rather, the scale is assumed equal to the location parameter  $\mu$ .



The Poisson distribution may also include an exposure variable associated with  $\mu$ . The variable,  $t$ , is considered to be the length of time or exposure during which events or counts occur. If  $t = 1$ , then the Poisson probability distribution reduces to the standard form. If  $t$  is a constant, or varies between events, then the distribution can be parameterized as

$$f_y(y; \mu) = e^{-t\mu} (t\mu)^y / y! \tag{1.2}$$

When included in the data, modelers enter the natural log of  $t$  as an offset in the model estimation. Playing an important role in estimating both Poisson and negative binomial models, offsets are discussed at greater length in Chapter 3.

A unique feature of the Poisson distribution is the relationship of its mean to the variance – they are equal. This relationship is termed equidispersion. The fact that it is rarely found in real data has driven the development of more general count models, which do not assume such a relationship.

The Poisson regression model derives from the Poisson distribution. The relationship between  $\mu$ ,  $\beta$ , and  $x$ , the fitted mean of the model, parameters, and model covariates or predictors respectively, is parameterized such that  $\mu = \exp(x\beta)$ . So doing guarantees that  $\mu$  is positive for all values of  $\eta$ , the linear predictor, and for all parameter estimates. By attaching the subscript,  $i$ , to  $\mu$ ,  $y$ , and  $x$ , the parameterization can be extended to all observations in the model. The subscript can also be used when modeling non-iid observations.

As shall be described in greater detail later in this book, the Poisson model carries with it various assumptions. Violations of Poisson assumptions usually result in overdispersion, where the variance of the model exceeds the value of the mean. Violations of equidispersion indicate correlation in the data, which affect standard errors of the parameter estimates. Model fit is also affected. Chapter 4 is devoted to this discussion.

A simple example of how distributional assumptions may be violated will likely be instructional at this point. We begin with the base count model – the Poisson. The Poisson distribution defines a probability distribution function for non-negative counts or outcomes. For example, given a Poisson distribution having a mean of 2, some 39% of the outcomes are predicted to be zero. If, in fact, we are given an otherwise Poisson distribution having a mean of 2, but with 50% zeros, it is clear that the Poisson distribution may not adequately describe the data at hand. When such a situation arises, modifications are made to the Poisson model to account for discrepancies in the goodness of fit of the underlying distribution. Models such as zero-inflated Poisson and zero-truncated Poisson directly address such problems.

The above discussion regarding distributional assumptions applies equally to the negative binomial. A traditional negative binomial distribution having

a mean of 2 and an ancillary parameter of 1.5 yields a probability of approximately 40% for an outcome of zero. When the observed number of zeros substantially differs from the theoretically imposed number of zeros, the base negative binomial model can be adjusted in a manner similar to the adjustments mentioned for the Poisson.

Early on, researchers developed enhancements to the Poisson model, which involved adjusting the standard errors in such a manner that the presumed overdispersion would be dampened. Scaling of the standard errors was the first method developed to deal with overdispersion from within the GLM framework. It is a particularly easy tactic to take when the Poisson model is estimated as a generalized linear model. We shall describe scaling in more detail later in the text. Nonetheless, most count models required more sophisticated adjustments than simple scaling.

Again, the negative binomial is normally used to model overdispersed Poisson data, which spawns our notion of the negative binomial as an extension of the Poisson. However, distributional problems affect both models, and negative binomial models themselves may be overdispersed. Both models can be extended in similar manners to accommodate any extra correlation or dispersion in the data that result in a violation of the distributional properties of each respective distribution (Table 1.1). The enhanced or advanced Poisson or negative binomial model can be regarded as a solution to a violation of the distributional assumptions of the primary model.

The following list enumerates the types of extensions that are made to both Poisson and negative binomial regression. Thereafter, we provide a bit more detail as to the nature of the assumption being violated and how it is addressed by each type of extension. Later chapters are devoted to a more detailed examination of each of these model types.

Earlier in this chapter we described violations of Poisson and negative binomial distributions as related to excessive zero counts. Each distribution has an expected numbers of counts for each value of the mean parameter; we saw how for a given mean, an excess – or deficiency – of zero counts result in overdispersion. However, it must be understood that the negative binomial has an additional ancillary or heterogeneity parameter, which, in concert with the value of the mean parameter, defines (in a probabilistic sense) specific expected values of counts. Substantial discrepancies in the number of counts, i.e. how many zeros, how many ones, how many twos, and so forth, observed in the data from the expected frequencies defined by the given mean and ancillary parameter (NB model), result in correlated data and hence overdispersion. The first two items in Table 1.1 directly address this problem.