Published in final edited form as:

Comput Stat Data Anal. 2008 January 1; 52(3): 1636–1649. doi:10.1016/j.csda.2007.05.016.

Sums of Exchangeable Bernoulli Random Variables for Family and Litter Frequency Data

Chang Yu^{*} and

Department of Biostatistics, School of Medicine, Vanderbilt University, Nashville, TN 37232, USA

Daniel Zelterman

Department of Epidemiology and Public Health, Yale University, New Haven, CT 06520, USA

Abstract

We describe new families of discrete distributions that are used to model sums of exchangeable Bernoulli random variables. These discrete distributions can be parameterized in terms of their range, mean, variance, and shape parameters. These models are fitted to an example involving mortality rates for children in a survey of families in Brazil. The methods illustrate that mortality rates in this survey increase with family size and that the correlation of within-family mortality status also depends on the family size. These methods are also applied to a laboratory study of birth defects in mice.

Keywords

Beta-binomial distribution; clustered binary data; familial data; teratology

1 Introduction

Many statistical summaries report data in terms of marginal sums without identifiable individual covariate values. In such cases we must base our statistical inference on sums of exchangeable observations. In this manuscript we explore models for such data when the outcome measured on each individual is binary valued. Applications of these methods include a toxicology and teratology study of litter-mates in an animal-based laboratory experiment and the mortality status of children in the same family.

The data of Table 1 summarizes a survey of deaths of children in northeast Brazil as reported by Sastry (1997). This area of Brazil is poorer and less developed than the rest of the country and has a corresponding higher incidence of infant mortality. The data was extracted from a household survey as part of the Demography and Health Survey. The multi-level survey used a cluster sampling scheme that selected many of the families from a relatively small number of communities. Maternity histories were collected on women aged 15–44 over a three month period in 1986. There are 2946 children in 1051 families represented in Table 1.

As an example of how to interpret the data in Table 1, consider the row with families of size $n = 3$. There were 202 families each with three children and 143 of these experienced no

*Contact for corresponding author: Email: E-mail: chang.yu@vanderbilt.edu, Phone: (615)-322-8422, Fax: (615)343-4924.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

childhood mortality; 41 families lost one child; 15 lost two; and in 3 families all three children died. This row represents a total of 80 deaths in 606 children.

No additional information about the individual children, such as their ages or birth order, was given in this table by Sastry (1997) so the children within each family will be assumed to be exchangeable. Every row in Table 1 represents the frequencies of same sized families and the columns of this triangular shaped table counts the numbers of deaths that were experienced. Each row in the table represents a frequency sample of sums of exchangeable individuals. This manuscript develops new discrete distributions to model these frequencies in Section 4. These parametric distributions are fitted to the data of Table 1 in Section 5.1.

Models for exact tests of exchangeability in triangular frequency tables such as these are described in Yu and Zelterman (2002a) in which we also provide models that include covariates at the individual and family level. Such tables also appear as summaries of teratology and toxicology experiments conducted on entire litters of laboratory animals. Several examples of these are given in Zelterman (2004, Chapter 9). Elliot, Joffe, and Chen (2006) point out that data of the type given in the Brazilian survey may contain a large bias because children who died in large families might have survived had they been born to smaller families. We will assume that all children are exchangeable across families of the same size. In Section 5.1 we demonstrate that the mortality rate and within-family dependence both vary by family size.

There are only a few methods available to model the distribution of a sum of exchangeable Bernoulli random variables. Chief among these is the beta-binomial distribution with mass function given at (1), below. Moore, Park, and Smith (2001) use this distribution to model radiation effects on fetal death in a laboratory study of mice. Altham (1978) develops two useful models, one of which is given at (2), below. The Altham model is particularly useful for fitting models using widely available log-linear model software (Zelterman, 2004, Section 8.3). Yu and Zelterman (2002a, b) provide a number of other methods for developing new distributions to model triangular shaped family data such as given in Table 1 including permutation methods for performing exact tests of significance.

In Section 2 we introduce the notation and provide some elementary background results. In Section 3 we review a method for generating new distributions. Section 4 provides specific parametric models. The example in Table 1 is modeled in Section 5.1. A second example involving birth defects in laboratory animals is examined in Section 5.2.

2 Notation and Background

Let $Z_{1n}, Z_{2n}, \dots, Z_{nn}$ denote a set of n exchangeable Bernoulli random variables. In the following narrative we will omit the second “ n ” subscript and it will be understood in the context. For the data in Table 1, the $\{Z_1, \dots, Z_n\}$ represent the alive or dead status of the children in one family of size n . Exchangeable random variables are identically distributed but not necessarily independent. In Section 5.1 we demonstrate that the survival status of children in Table 1 within the same family do not behave as independent Bernoulli random variables.

Recall that exchangeable means that every subset of every size has the same distribution. Formally, the random variables $Z_1, \dots, Z_n$ are said to be *exchangeable* if

$$\Pr[Z_1=z_1, \dots, Z_k=z_k] = \Pr[Z_{i_1}=z_1, \dots, Z_{i_k}=z_k]$$

for indices $\{i_1, \dots, i_k\}$, any permuted subset of $\{1, \dots, n\}$; all $k = 1, 2, \dots, n$; and for all $z_i = 0$ or 1.

Define Y_n to be the sum of the n exchangeable Bernoulli random variables:

$$Y_n = Z_1 + \cdots + Z_n.$$

In the data in Table 1, the random variable Y_n represents the number of deaths in a family of size n . The $\{Z_i\}$ are not individually observable, only their sum Y_n . One goal of this work is to develop and fit models for the distribution of Y_n for this data.

If the Z_i are mutually independent then Y_n has a binomial distribution. The well-known DeFinetti representation for the distribution of a sum of exchangeable Bernoulli random variables sampled from an infinitely large population is

$$\Pr[Y_n = i] = \binom{n}{i} \int p^i (1-p)^{n-i} f(p) dp$$

where $f(\cdot)$ is a density function on $(0, 1)$.

This representation is concerned with a sample from an infinite number of exchangeable Bernoulli's which must be positively correlated. We also want to be able to model negatively correlated indicators as well and finite numbers are of particular interest. Most studies of contagious diseases such as influenza and diseases with a genetic or environmental component such as cancer will usually demonstrate a positive correlation between siblings. Negative correlations are possible but do not appear often in practice. In studies of polio for example, exposure to the virus at a young age can have a preventative effect later in life, in which case siblings would exhibit a mutual negative correlation.

The most commonly encountered distribution for Y_n with exchangeable Z_i is the beta-binomial distribution, which can be derived as a beta distributed mixture of binomial distributions. The general form for this distribution given at (1) below, allows for negative correlations among

the Bernoulli random variables. If we define $\prod_{i=0}^{-1} = 1$ then the probability mass function of the *beta-binomial distribution* is

$$\Pr[Y_n = i] = \binom{n}{i} \prod_{r=0}^{i-1} (r\alpha + p) \prod_{s=0}^{n-i-1} (s\alpha + 1 - p) / \prod_{t=0}^{n-1} (t\alpha + 1) \quad (1)$$

for $i = 0, \dots, n$ and

$$\alpha > -\min(p, 1-p)/(n-1).$$

The p parameter in (1) is equal to $\Pr[Z_i = 1]$. The sign of α determines the sign of the correlation between any pair of Z_i . In most applications we encounter positive association between the exchangeable Bernoulli Z_i but slight negative mutual correlations are possible for finite values of n . The binomial distribution is obtained in (1) for $\alpha = 0$.

The mean and variance of the beta-binomial distribution in (1) are

$$EY_n = np$$

and

$$\text{Var } Y_n = np(1-p)\{1 + (n-1)\alpha/(\alpha+1)\}.$$

Another model for the distribution of Y_n is obtained by Altham (1978) who derives the model with probability mass satisfying

$$\Pr[Y_n = i] \propto \binom{n}{i} p^i (1-p)^{n-i} \theta^{i(n-i)} \quad (2)$$

for $i = 0, \dots, n$.

The constant of proportionality in (2) and moments of the Altham distribution do not have simple closed form expressions. The parameter $\theta > 1$ in (2) measures a negative association and $0 < \theta < 1$ indicates a positive association between the Bernoulli status of exchangeable family members $Z_1, \dots, Z_n$. Model (2) assumes that the Bernoulli distributed individuals Z_k have a pair-wise interaction that is a function of θ and there are no higher order interactions between the Z_k . The binomial distribution is obtained in (2) for $\theta = 1$.

3 A tool for new models

Other useful models for the distribution of Y_n for exchangeable Z_k are constructed by George and Bowman (1995); Bowman and George (1995); and George and Kodell (1996). These authors derive the relationship

$$\Pr[Y_n = i] = \binom{n}{i} \sum_{k=0}^{n-i} \binom{n-i}{k} (-1)^k \lambda_{i+k} \quad (3)$$

where $\lambda_0 = 1$ and

$$\lambda_k = \Pr[Y_k = k] = \Pr[Z_1 = \dots = Z_k = 1]$$

for $k = 1, 2, \dots$

Every λ_k in (3) is the probability that all individuals in a family of size k are affected. A binomial (n, p) distribution is obtained in (3) for $\lambda_k = p^k$ for $k = 0, 1, \dots$. Section 4 describes three new models that can be derived from (3). The beta-binomial distribution (1) can be obtained from (3) but the Altham model (2) is not expressible using (3). In Section 4 we demonstrate that the values for λ_k must be carefully specified in order for (3) to represent a valid distribution. Kuk (2004) and Pang and Kuk (2005) develop additional distributions based on (3).

Let us present three brief examples of the use of (3) before we proceed. Values of $\lambda_1 = p$;

$$\lambda_2 = p(p+\alpha)/(1+\alpha);$$

and for $k=3, 4, \dots$

$$\lambda_k = \frac{p(p+\alpha) \cdots \{p+(k-1)\alpha\}}{(1+\alpha) \cdots \{1+(k-1)\alpha\}} \quad (4)$$

in (3) gives us the beta-binomial distribution with mass function (1).

The hypergeometric distribution with integer-valued parameters satisfying $0 < n < N$ and $0 < m < N$ and mass function

$$\Pr[Y_n=i] = \binom{n}{i} \binom{N-n}{m-i} / \binom{N}{m}$$

for $\max\{0, n+m-N\} \leq i \leq \min\{n, m\}$ is obtained using

$$\lambda_k = \prod_{i=0}^{k-1} (m-i)/(N-i)$$

with $k=1, \dots, m$ and $\lambda_k=0$ if $k > m$.

As a third example, if $\lambda_k = \theta p^k$ for $k=1, 2, \dots$ then we have the zero-inflated (or deflated) binomial distribution with mass function

$$\Pr[Y_n=i] = \theta \binom{n}{i} p^i (1-p)^{n-i}$$

for $i=1, 2, \dots, n$ and

$$\Pr[Y_n=0] = 1 - \theta \{1 - (1-p)^n\}$$

for probability parameter p ($0 < p < 1$) and $0 < \theta < \{1 - (1-p)^n\}^{-1}$. This distribution inflates or deflates the frequency of $Y_n=0$ relative to the binomial distribution depending on whether $0 < \theta < 1$ or $\theta > 1$ respectively.

Let us return to elementary properties of the random variable Y_n in (3). Begin by noting that

$$\lambda_1 = \Pr[Z_i=1] = E[Z_i]$$

for $i=1, 2, \dots, n$ determines the marginal expectation of any one child's death risk so that

$$EY_n = n\lambda_1$$

in model (3).

The covariance of any two Z_i 's can be expressed in terms of the λ 's:

$$\text{Cov}(Z_1; Z_2) = \lambda_2 - \lambda_1^2$$

giving us

$$\text{Var } Y_n = n\lambda_1(1 - \lambda_1) + n(n-1)(\lambda_2 - \lambda_1^2). \quad (5)$$

Mutual negative correlations among the exchangeable Bernoulli random variables Z_i are possible but can only be slight when n is large because (5) must be non-negative.

An odds ratio measure between a pair of Bernoulli indicators can be expressed as

$$OR = \frac{\{E(Z_1 Z_2)\} \{E(1 - Z_1)(1 - Z_2)\}}{\{EZ_1(1 - Z_2)\} \{E(1 - Z_1)Z_2\}}.$$

We can use the exchangeability property and express this measure of dependence in terms of λ_i giving

$$OR = \lambda_2(1 - 2\lambda_1 + \lambda_2) / (\lambda_1 - \lambda_2)^2. \quad (6)$$

The OR measure is used in Section 5.1 to illustrate the dependence in mortality status of individuals within the same family among the Brazilian children listed in Table 1.

General factorial moments of Y_n have a simple expression. For $r = 1, 2, \dots$ define the factorial polynomial

$$Y^{(r)} = Y(Y-1) \cdots (Y-r+1)$$

and define $Y^{(0)} = 1$.

In the Appendix we show that $EY_2^{(r)} = n^{(r)} \lambda_r$. This expression coincides with the corresponding factorial moments of the binomial distribution when $\lambda_r = p^r$. As an example, from this expression we can calculate the third central moment

$$E(Y_n - n\lambda_1)^3 = n^{(3)}\lambda_3 + 3n^{(2)}\lambda_2(1 - \mu) + \mu(1 - \mu)(1 - 2\mu)$$

where $\mu = E(Y_n) = n\lambda_1$.

Additional properties and expressions for general central moments of Y_n satisfying (3) are given by George and Bowman (1995).

4 Specific model parameterizations

Consider the tail measure τ_n defined as

$$\tau_n = \Pr[Y_n = n | Y_{n-1} = n - 1] = \lambda_n / \lambda_{n-1}.$$

The value of τ_n describes the conditional probability of $Z_n = 1$ after having already observed $Z_1 = \dots = Z_{n-1} = 1$. If τ_n is close to one then the Bernoulli distributed indicators Z_i will ultimately become highly dependent when n is large. Loosely speaking, the joint distribution of $\{Z_1, \dots, Z_n\}$ is of dimension less than n as n becomes large. On the other hand, if τ_n has a limit that is less than 1 then, again, roughly speaking, all of the Bernoulli distributed Z_i will continue to retain a random character even when we condition on a large number of their outcomes. The word “tail” is used in this section to describe the asymptotic behavior of τ_n and does not refer to the extremes in the distribution of Y_n .

For the binomial model, $\tau_n = p$ for all n .

For the beta-binomial distribution (1) with λ_n given at (4) we have

$$\tau_n = \{p + (n - 1)\alpha\} / \{1 + (n - 1)\alpha\}.$$

In the beta-binomial distribution, these values of τ_n are close to 1 for large values of n . Specifically

$$\tau_n = 1 - (1 - p) / \{(n - 1)\alpha + 1\} = 1 - O(1/n)$$

for large values of n in the beta-binomial distribution.

A limit of 1 for τ_n appears in the beta-binomial distribution and we generalize that model in Section 4.1. In Section 4.2 we generalize the binomial distribution by considering a choice of the λ_n for which the corresponding τ_n has a limit that is less than 1. In Section 4.3 we describe a second generalization of the binomial distribution in which τ_n approaches its limit at an exponential rate.

4.1 The beta-binomial tail model

Define the *beta-binomial tail model* obtained using $\lambda_1 = p$;

$$\lambda_2 = p(p + \alpha) / (1 + \alpha)$$

and more generally,

$$\lambda_n = \frac{p(p+1^\theta\alpha)(p+2^\theta\alpha)\cdots\{p+(n-1)^\theta\alpha\}}{(1+1^\theta\alpha)(1+2^\theta\alpha)\cdots\{1+(n-1)^\theta\alpha\}} \quad (7)$$

for $n = 3, 4, \dots$ in (3).

The beta-binomial tail model has probability parameter p and positive parameters θ and α . This distribution generalizes the beta-binomial distribution. Values of $\theta = 1$ in (7) gives the beta-binomial model (1) and $\alpha = 0$ yields the binomial distribution.

In the beta-binomial tail model (7) we have

$$\tau_n = 1 - (1-p) / \{(n-1)^\theta\alpha + 1\} = 1 - O(1/n^\theta)$$

allowing us to control the high dimensional dependence among the Bernoulli distributed Z_i .

The beta-binomial tail distribution has the same parametric form of the mean and variance as the beta-binomial distribution (1). Specifically,

$$EY_n = n\lambda_1 = np$$

and

$$\text{Var } Y_n = np(1-p)\{1 + (n-1)\alpha/(1+\alpha)\}$$

using expression (5).

These two moments are independent of the choice of θ and are the same as those of the beta-binomial distribution with parameters (n, p, α) .

Fig. 1 displays beta-binomial tail models using (7) in (3) and compares these with a beta-binomial distribution with parameters $p = .35$ and $\alpha = .15$. Values of θ range from .8 to 1.4 by .1 in this figure. Values of $\theta > 1$ shorten the lower tail and lengthen the upper tail of this distribution relative to those of the beta-binomial. The opposite is true for $0 < \theta < 1$. All distributions in this figure have expected value 3.5 and variance 4.95. The tail parameter θ only appears in $\lambda_3, \lambda_4, \dots$ so that the mean and variance of Y_n are independent of the choice of θ . This results in a family of discrete distributions parameterized by their range, mean, variance, and shape parameters.

Extreme values of θ can result in bimodal distributions with local or global modes at 0 or n . Local modes at $Y_n = 0$ and $Y_n = n$ are common because the exchangeability property results in a higher probability that $Z_1, \dots, Z_n$ all have the same value. Extreme values of θ can result in negative values of (3). It is difficult to specify the range of valid parameter values in this and other distributions in this section. Similar problems were noted in most other published works that make use of (3).

4.2 The binomial tail model

If τ_n has a limit less than 1 then the resulting joint distribution of $\{Z_n\}$ will have properties similar to those of the binomial distribution. That is, the limiting distribution of Y_n will not become degenerate when we condition on the event $\{Y_{n-1} = n - 1\}$ for large values of n .

Define the *binomial tail model* obtained using $\lambda_1 = p$;

$$\lambda_2 = p(p + \alpha) / (1 + \alpha)$$

and

$$\lambda_n = \frac{p \{p + \theta^0 \alpha\} \{p + \theta^1 \alpha\} \cdots \{p + \theta^{n-2} \alpha\}}{\{1 + \theta^0 \alpha\} \{1 + \theta^1 \alpha\} \cdots \{1 + \theta^{n-2} \alpha\}} \quad (8)$$

for $n = 3, 4, \dots$ in (3) for parameters $0 < p < 1$ and $\theta > 0$.

Most applications of (8) will have $\alpha \geq 0$ but slight negative values of α are possible when n is not too large. Examples of this distribution appear in Fig. 2 and include one model with a negative value of α . The binomial (n, p) distribution is obtained for $\alpha = 0$. The beta-binomial distribution is not a special case of model (8). In (8) we have

$$\tau_n = 1 - (1 - p) / (1 + \theta^{n-2} \alpha)$$

so that τ_n tends to 1 for $\theta > 1$ and τ_n tends to p if $0 < \theta < 1$ for large values of n .

If $\theta = 1$ then τ_n is between p and 1 depending on the value of α . The discrete distributions obtained using (8) in (3) have the same mean and variance as those of the beta-binomial distribution with parameters (n, p, α) .

4.3 The exponential tail model

For probability parameter p , define

$$\lambda_k = \prod_{i=1}^k \tau_i$$

for $k = 1, 2, \dots$ with

$$\tau_i = 1 - (1 - p) \exp \{-\alpha(i - 1)\}.$$

In this model the limit of τ_n is 1 for $\alpha > 0$ and this limit is approached at an exponential rate. Generally $\alpha \geq 0$ in this model but small negative values of α are possible for finite n and these serve to model negative dependence among the Bernoulli indicators Z_i . The binomial

distribution is obtained when $\alpha = 0$. Examples of this distribution are plotted in Fig. 3 including one example with $\alpha < 0$.

This two-parameter distribution has expected value $E Y_n = np$ and variance

$$\text{Var } Y_n = np(1-p) + n(n-1)p(1-p)(1-e^{-\alpha}).$$

The non-negative variance constraints $\alpha \geq \log((n-1)/n)$.

5 Two Numerical Examples

In this section we fit models to the Brazilian family data of Table 1 and also examine a study involving birth defects in laboratory animals.

5.1 Childhood Mortality in Brazil

We return to the data in Table 1 discussed in the introduction. We want to show that larger families experienced higher death rates. We also model the change in dependence among the Bernoulli outcomes of their individual children as a function of family size.

In this example and the following in Section 5.2, we fit models by minimizing the deviance function

$$\Lambda(\theta) = 2 \sum_n \sum_{i=0}^n X_{ni} \log \{X_{ni}/E(X_{ni})\}$$

where X_{ni} is the number of families of size n that lost i children for $n = 1, 2, \dots$ and $i = 0, 1, \dots, n$.

The expected frequencies are

$$E(X_{ni}) = f_n \Pr[Y_n = i | \theta]$$

for fitted values of θ that minimize the deviance $\Lambda(\theta)$.

The f_n here and in Table 1 are the number of families of size n . The most basic example is to assume that mortality status is independent within families and that the mortality rate is constant across families. In this case we have the binomial model with

$$\Pr[Y_n = i] = \binom{n}{i} p^i (1-p)^{n-i}.$$

The estimate for p in this model is $\hat{p} = 430/2946$ with a deviance chi-squared of 152.10 on 35 df. Other models described here are also fitted and summarized in Table 2. None of the models fitted in this table include the effect of family size as a covariate. The lack of fit in all of these models is due to both a non-constant mortality rate and a lack of independence, as we illustrate next.

Let us address the non-constant mortality rate across different family sizes n in Table 1 by writing

$$\text{logit}(p) = \text{logit}(p_n) = \gamma_0 + \gamma_1 n + \gamma_2 n^2. \quad (9)$$

The fitted values of $(\widehat{\gamma}_0, \widehat{\gamma}_1, \widehat{\gamma}_2)$ are described in Table 3. The Wald chi-squared tests of $\gamma_1 = 0$ and $\gamma_2 = 0$ confirm that mortality rates in Table 1 increases with family size. This represents a huge improvement in model fit over the simple binomial model with one p parameter for all families. The deviance of model (9) is 56.86 on 33 df is still an inadequate fit, but represents a large improvement over a single parameter binomial model.

Changes in within-family dependence are also modeled using

$$\log(\alpha) = \log\{\alpha(n)\} = \beta_0 + \beta_1 n + \beta_2 n^2$$

and the fitted parameters are included in Table 3.

The remainder of the analysis of these data consists of modeling parameters as quadratic functions of family size in a generalized linear model fashion. In Table 3 we see that the binomial model has a poor fit but the deviances of all other models that include dependencies among the Bernoulli-distributed children have a much better fit. The fit of all of these are almost identical, in terms of their deviances.

Fig. 4 plots the estimated mortality rate $p(n)$ and the odds ratio measure OR given at (6) for the fitted exponential tail distribution. The mortality rate clearly increases with family size. The odds ratio is generally decreasing, although not monotonic, in family size.

In summary, mortality increases with family size and within-family mortality status becomes more independent with family size. Of course, it is impossible to tell from these data whether the loss of a child would lead parents to have another or if having another puts the older siblings at greater risk. Larger families also have a wider distribution of ages of the children so their mortality status is likely to be more mutually independent.

5.2 Birth defects in laboratory animals

This is data from Tyl *et al.* (1988) that is also examined by Brooks *et al.* (1997), Dominici and Parmigiani (2001), Li and Hunt (2004), and others using a variety of statistical methods. Pregnant mice were fed Di(2-ethylhexyl)phthalate (DEHP), a widely used plasticizing agent, at one of several different exposure levels in order to study developmental toxicity and birth defects among their offspring. We examined all combined malformations and fetal deaths among the litters born to these dams. Each litter was associated with a covariate vector x that included five components: an intercept; and linear and quadratic terms for both litter sizes and exposure levels. The deviances and functional forms of the fitted models are given in Table 4.

All of the fitted models in Table 4 that include dependencies among litter-mates exhibit a great improvement in the deviance over the binomial model that assumes independence. The beta-binomial model has the smallest deviance over all ten parameter models. The additional θ parameter in the beta-binomial tail model has a significantly better fit (chi-squared difference=4.13, 1 df, $p=.042$) over the beta-binomial distribution.

6 Discussion

Bowman and George (1995) motivated the use of (3) over the generalized estimating equations (GEE) approach because (3) allows for specification of interactions of all orders for the Bernoulli indicators. GEE limits these specifications to means and variances alone. This limitation may not be as restrictive as it first appears. In the models we develop here, few demonstrate a large improvement, if any, over the beta-binomial model, in terms of reduction in the deviance. This lack of widespread improvement over the beta-binomial distribution was also experienced by Brooks *et al.* (1997), Kuk (2004), and Pang and Kuk (2005). It is possible that the GEE approach may come very close to the best representation for data of the type described in Section 5.1.

A second major difficulty in the use of (3) is the careful specification of the λ_i parameters necessary to obtain all non-negative probabilities generated by this model. While (3) always sums to $\lambda_0 = 1$, not every choice of parameter values results in non-negative probabilities. There are only a few parametric examples of models that make use of (3) that result in all non-negative probabilities. One example is the completely monotone function examined by Kuk (2004). It is difficult to describe the range of parameter values of the models developed in Section 4 that result in all non-negative probabilities when using (3).

Acknowledgments

This research was supported by the U.S. National Institutes of Health with grant RR00095 awarded to the GCRC at Vanderbilt University Medical Center (Chang Yu) and with grant MH62294 awarded to the Yale Center for Interdisciplinary Research in AIDS (Daniel Zelterman).

Appendix: Factorial moments of Y_n

For $r = 1, 2, \dots$ denote

$$Y^{(r)} = Y(Y-1)\cdots(Y-r+1)$$

and $Y^{(0)} = 1$.

Equation (3) shows

$$EY_n^{(r)} = \sum_{i=0}^n i^{(r)} \binom{n}{i} \sum_{k=0}^{n-i} \binom{n-i}{k} (-1)^k \lambda_{i+k}.$$

Since $i^{(r)} = 0$ for $i < r$, write

$$EY_n^{(r)} = n^{(r)} \sum_{i=r}^n \frac{(n-r)!}{(i-r)!} \sum_{k=0}^{n-i} (-1)^k \lambda_{i+k} / k! (n-i-k) !.$$

Then use a change of variables: $s = i + k$; and $t = k$ to show

$$EY_n^{(r)} = n^{(r)} \sum_{s=r}^n \sum_{t=0}^{s-r} \lambda_s (-1)^t (n-r)! / t! (n-s)! (s-r-k)!.$$

Multiply and divide by $(s-r)!$ giving

$$EY_n^{(r)} = n^{(r)} \sum_{s=r}^n \lambda_s \binom{n-r}{s-r} \sum_{t=0}^{s-r} \binom{s-r}{t} (-1)^t.$$

The inner summation is equal to $(1-1)^{s-r}$ which is zero except when $s=r$, proving

$$EY_n^{(r)} = n^{(r)} \lambda_r.$$

References

- Altham PME. Two generalizations of the binomial distribution. *Applied Statistics* 1978;27:162–7.
- Bowman D, George EO. A saturated model for analyzing exchangeable binary data: Applications to clinical and developmental toxicity studies. *Journal of the American Statistical Association* 1995;90:871–9.
- Brooks SP, Morgan BJT, Ridout MS, Pack SE. Finite mixture models for proportions. *Biometrics* 1997;53:1097–115. [PubMed: 9333342]
- Dominici F, Parmigiani G. Bayesian semiparametric analysis of developmental toxicology data. *Biometrics* 2001;57:150–7. [PubMed: 11252591]
- Ekholm A, Smith PWF, McDonald JW. Marginal regression analysis of a multivariate binary response. *Biometrika* 1995;82:847–54.
- Elliot MR, Joffe MM, Chen Z. A potential outcomes approach to developmental toxicity analyses. *Biometrics* 2006;62:352–60. [PubMed: 16918899]
- George EO, Bowman D. A full likelihood procedure for analyzing exchangeable binary data. *Biometrics* 1995;51:512–23. [PubMed: 7662840]
- George EO, Kodell RL. Tests of independence, treatment heterogeneity, and dose-related trend with exchangeable binary data. *Journal of the American Statistical Association* 1996;91:1602–10.
- George EO, Wu H. Testing for treatment related trend with partially exchangeable clustered data. *Biometrics* 1997;53:479–87. [PubMed: 9192447]
- Kuk AYC. A litter-based approach to risk assessment in developmental toxicity studies via a power family of completely monotone functions. *Applied Statistics* 2004;53:369–86.
- Li CS, Hunt D. Regression splines for threshold selection with application to random-effects logistic dose-response model. *Computational Statistics & Data Analysis* 2004;46:1–9.
- Moore DF, Park CK, Smith W. Exploring extra-binomial variation in teratology data using continuous mixtures. *Biometrics* 2001;57:490–4. [PubMed: 11414574]
- Pang Z, Kuk AYC. A shared response model for clustered binary data in developmental toxicity studies. *Biometrics* 2005;61:1076–84. [PubMed: 16401281]
- Sastry N. A nested frailty model for survival data, with an application to the study of child survival in northeast Brazil. *Journal of the American Statistical Association* 1997;92:426–34. [PubMed: 12155403]
- Tyl RW, Price CJ, Marr MC, Kimmel CA. Developmental toxicity evaluation of dietary di(2-ethylhexyl) phthalate in Fischer 344 rats and CD-1 mice. *Fundam Appl Toxicol* 1988;10:395–412. [PubMed: 3371580]
- Yu C, Zelterman D. Statistical inference for familial disease clusters. *Biometrics* 2002a;58:481–91. [PubMed: 12229982]

Yu C, Zeltermann D. Sums of dependent Bernoulli random variables and disease clustering. *Statistics & Probability Letters* 2002b;57:363–73.

Zeltermann, D. *Discrete Distributions Applications in the Health Sciences*. Chichester: Wiley; 2004.

Figure 1.

Beta-binomial tail model probability distribution mass functions obtained using (7) in (3) with parameters $\lambda_1 = p = .35$; $n = 10$; $\alpha = .15$ and $\theta = .8$ to 1.4 by $.1$. The beta-binomial distribution corresponding to $\theta = 1$ is given by a dashed line. All distributions in this figure have expected value 3.5 and variance 4.95 . Mass points of these discrete distributions have been connected with lines for clarity.

Figure 2.

Examples of the binomial tail distributions obtained using (8) in (3) for parameters $n = 10$, $p = .35$, and $\theta = 1$. Values of α vary from 0 to .25 by .05. The binomial(10, .35) mass function corresponds to $\alpha = 0$ and is given in a dashed line. One model with negative $\alpha = -.0075$ is given. All distributions in this figure have expected value 3.5.

Figure 3.

Mass functions of the exponential tail distribution for parameters $n = 10$, $p = 0.4$ and $\alpha = 0$ to .25 by .05. This figure includes an example of negative association for $\alpha = -.03$. The binomial ($n = 10$, $p = .4$) distribution corresponds to $\alpha = 0$ and is highlighted with a dashed line. All distributions in this figure have expected value equal to 4.

Figure 4.

Fitted mortality probability and odds ratios for different family sizes using exponential tail model described in Section 4.3. Fitted parameters are given in Table 3. The estimated mortality rate $\hat{p}=\hat{\lambda}_1$ is given in a solid line using the scale at left and the odds ratio given at (6) is plotted with a dashed line with the scale on the right.

Table 1
The frequency of 430 deaths among 2946 children in 1051 Brazilian families as reported by Sastry (1997).

Size of family n	Number of families f_n	Number of deaths	Number of deaths i							
			0	1	2	3	4	5	6	7+
1	267	12	255	12						
2	285	48	239	44	2					
3	202	80	143	41	15	3				
4	110	54	69	30	9	2	0			
5	104	103	43	34	15	9	3	0		
6	50	67	15	18	8	5	3	0	1	
7	21	38	4	4	7	4	2	0	0	0
8	12	28	1	2	4	3	1	1	0	0
Totals		1051	430							

Table 2

Fitted parameter values and deviances for the Brazilian mortality data. Estimated standard errors are based on inverse Hessian obtained from routine nlm in R.

Model and Parameters	Estimate	Est'd SE	Null hypothesis	1 df Chi-squared	p-value	Deviance
Hypergeometric						152.54
Binomial						152.10
p	0.146	0.005				
Altham (p, θ)						144.30
p	0.123	0.007				
θ	1.105	0.029	$\theta = 1$	13.38	.0003	
Beta-binomial (p, α)						98.54
p	0.136	0.005				
α	0.130	0.018	$\alpha = 0$	50.24	<.0001	
Beta-binomial tail						98.54
p	0.136	0.005				
α	0.130	0.018	$\alpha = 0$	50.78	<.0001	
θ	0.971	0.227	$\theta = 1$	0.02	.90	
Binomial tail						98.36
p	0.136	0.005				
α	0.130	0.019	$\alpha = 0$	49.18	<.0001	
θ	2.044	0.341				
Exponential tail						98.59
p	0.136	0.005				
α	0.122	0.016	$\alpha = 0$	54.80	<.0001	

Table 3

Fitted parameter values and deviances for the Brazilian mortality data. These parameterizations are quadratic functions of the family size n covariate.

Model and Parameters	Estimate	Est'd SE	Null hypothesis	1 df Chi-squared	p-value	Deviance
Binomial: $\text{logit}(p) = \gamma_0 + \gamma_1 n + \gamma_2 n^2$						
γ_0	-3.301	0.203				56.86
γ_1	0.492	0.095	$\gamma_1 = 0$	26.57	< .001	
γ_2	-0.024	0.010	$\gamma_2 = 0$	5.58	0.018	
Altham: $\text{logit}(p) = \gamma_0 + \gamma_1 n + \gamma_2 n^2$; $\log(\theta) = \beta_0 + \beta_1 n + \beta_2 n^2$						
γ_0	-3.834	0.257				19.28
γ_1	1.018	0.130	$\gamma_1 = 0$	60.90	< .001	
γ_2	-0.083	0.015	$\gamma_2 = 0$	29.04	< .001	
β_0	-0.220	0.233				
β_1	-0.091	0.088	$\beta_1 = 0$	1.08	0.299	
β_2	0.016	0.008	$\beta_2 = 0$	3.48	0.062	
Beta binomial: $\text{logit}(p) = \gamma_0 + \gamma_1 n + \gamma_2 n^2$; $\log(\alpha) = \beta_0 + \beta_1 n + \beta_2 n^2$						
γ_0	-3.329	0.215				19.42
γ_1	0.507	0.103	$\gamma_1 = 0$	24.07	< .001	
γ_2	-0.026	0.112	$\gamma_2 = 0$	5.34	0.021	
β_0	-4.562	1.526				
β_1	1.341	0.712	$\beta_1 = 0$	3.55	0.060	
β_2	-0.161	0.080	$\beta_2 = 0$	4.04	0.045	
Beta-binomial tail: $\text{logit}(p) = \gamma_0 + \gamma_1 n + \gamma_2 n^2$; $\log(\alpha) = \beta_0 + \beta_1 n + \beta_2 n^2$						
γ_0	-3.330	0.215				19.19
γ_1	0.507	0.103	$\gamma_1 = 0$	24.10	< .001	
γ_2	-0.026	0.011	$\gamma_2 = 0$	5.38	0.020	
β_0	-4.762	1.601				
β_1	1.446	0.755	$\beta_1 = 0$	3.57	0.055	
β_2	-0.173	0.085	$\beta_2 = 0$	4.09	0.043	
θ	1.160	0.238	$\theta = 1$	0.46	0.499	
Binomial tail: $\text{logit}(p) = \gamma_0 + \gamma_1 n + \gamma_2 n^2$; $\log(\alpha) = \beta_0 + \beta_1 n + \beta_2 n^2$						
						20.05

Model and Parameters	Estimate	Est'd SE	Null hypothesis	1 df Chi-squared	p-value	Deviance
γ_0	-3.312	0.214				
γ_1	0.495	0.103	$\gamma_1 = 0$	23.07	< .001	
γ_2	-0.024	0.011	$\gamma_2 = 0$	4.80	0.029	
β_0	-4.194	1.537				
β_1	1.141	0.729	$\theta_1 = 0$	2.45	0.117	
β_2	-0.138	0.083	$\theta_2 = 0$	2.75	0.097	
θ	2.231	0.376				19.33
Exponential tail: $\text{logit}(p) = \gamma_0 + \gamma_1 n + \gamma_2 n^2$; $\log(\alpha) = \beta_0 + \beta_1 n + \beta_2 n^2$						
γ_0	-3.329	0.215				
γ_1	0.507	0.103	$\gamma_1 = 0$	23.97	< .001	
γ_2	-0.026	0.011	$\gamma_2 = 0$	5.31	0.021	
β_0	-4.516	1.444				
β_1	1.279	0.672	$\beta_1 = 0$	3.62	0.057	
β_2	-0.153	0.076	$\beta_2 = 0$	4.09	0.043	

Table 4

Fitted models for the data from Tyl, *et al.* (1988). In litters of size n , the Y_n denotes combined outcomes of either malformations or fetal death. The covariate vector $\mathbf{x}$ contains an intercept, and linear and quadratic terms for both litter sizes and exposure doses.

Model	Parameterization	Deviance
Binomial Altham	$\text{logit}(p) = \beta' \mathbf{x}$	345.15
	$\text{logit}(p) = \beta' \mathbf{x}$	240.44
	$\log(\theta) = \gamma' \mathbf{x}$	
Beta binomial	$\text{logit}(p) = \beta' \mathbf{x}$	227.68
Beta-binomial tail	$\log(a) = \gamma' \mathbf{x}$	
	$\text{logit}(p) = \beta' \mathbf{x}$	223.55
	$\log(a) = \gamma' \mathbf{x}$	
Binomial tail	one θ	
	$\text{logit}(p) = \gamma' \mathbf{x}$	260.95
	$\log(a) = \beta' \mathbf{x}$	
Exponential tail	$\text{logit}(p) = \gamma' \mathbf{x}$	234.81
	$\log(a) = \beta' \mathbf{x}$	