The Tilted Beta-Binomial Distribution in Overdispersed Data: Maximum Likelihood and Bayesian Estimation

Eugene D. Hahn
Department of Information
and Decision Sciences
Salisbury University
Salisbury, MD 21801 USA

June 8, 2022

Abstract

Binary success/failure data is an important type of data which occurs in a wide range of fields and application domains. We examine a new overdispersed statistical distribution for this type of data which is particularly suited to 'big data' contexts. The distribution, called the tilted beta-binomial distribution, has a number of attractive properties with regard to tractability and interpretability. We apply the distribution to the analysis of population data from the 2010 U.S. Census. We find the tilted beta-binomial distribution provides a superior fit compared to the beta-binomial distribution in the current data set and provides new options for modeling success/failure data. In addition to addressing overdispersion, the distribution has implications for generating new insights by revealing underlying latent subgroups in data.
Keywords: Finite mixture models, discrete distributions, Bayesian modeling, demographic data.

1 Introduction

Binary data consisting of x successes and $n - x$ failures in n trials is an important data type occurring in a very wide range of fields and application domains. This data is often modeled with the binomial distribution and hence is often termed binomial data. However, it is common to find that the variance specified by the binomial distribution is smaller than is actually observed in the data. In these instances, the data is described as either having extra-binomial variability or as exhibiting overdispersion. For example, toxicological research using litters of mice often reveals a strong effect of litter membership (Brooks et al. 1997, Luo and Paul 2018). Mice in the same litter will tend to have similar responses to a potential toxin. The litter functions as a cluster composed of similar observations, giving rise to a larger variance than would be found if all observations were truly independent. Clusters and overdispersion are a common phenomenon in diverse areas such as microbiome research (Hu et al. 2018, Aldirawi and Yang 2022), quality assurance (Hedt-Gauthier et al. 2013), hospital bed occupancy (Gange et al. 1996), library book circulation (Gelman and Sichel 1987), seed germination (Crowder 1978, Cepeda-Cuervo and Cifuentes-Amado 2020), criminology (Haining et al. 2009), and behavioral neuroscience (Ferrari and Comelli 2016). Data that has a spatial nature may also have clusters and may therefore exhibit overdispersion (Shiyomi et al. 2000, Haining et al. 2009).

A thought experiment easily shows why overdispersion is common when data occurs in clusters or otherwise has a hierarchical component. Suppose a first data set follows a simple binomial model with probability π and sample size n. Suppose a second data set is a mixture of two equally-sized binomially-distributed groups (possibly latent) that would be better described by parameters π_1 and $n/2$ for Group 1 and π_2 and $n/2$ for Group 2 with $\pi_1 \gg \pi_2$ (we ignore the topic of estimation of mixture weights for clarity of exposition). However suppose we apply the simple binomial model to this second data set as well. Then it is well known the binomial mean, $n\pi$, can still be consistently estimated in both data sets despite overdispersion (Cox 1983). Applying the binomial variance formula, $n\pi(1-\pi)$, to both data sets will however lead to an underestimate of the variance in the second data set because the reality of $\pi_1 \gg \pi_2$ has not been reflected in the statistical model. A formula can be easily derived but it suffices to realize the expectation of the squares must be larger in the second data set. Clearly the simple binomial variance will then be too small for the second data set.

In this thought experiment we know there are exactly two (possibly latent) groups with parameters as stated but in real-world contexts this level of certainty may not exist. We may be unaware of such properties of our data. However our lack of awareness does not mean the issue can be ignored. Failure to properly account for overdispersion in binomial data leads to overly precise estimates and increases in Type I error rates. As the above thought experiment shows, failure to properly account for overdispersion can also lead to poor out-of-sample predictive performance. To reduce these problems researchers began to seek better models for overdispersed binomial data using alternative distributions such as the beta-binomial distribution (Williams 1975). This trend has been assisted in recent years by increases in computing power and the existence of larger datasets which may exhibit richer patterns of clustering.

In this paper we expand the scope of knowledge about a relatively new distribution for modeling overdispersed binomial data called the tilted beta-binomial distribution. The distribution generalizes the beta-binomial distribution and is capable of modeling greater amounts of overdispersion than the beta-binomial. As for a brief review of its literature, to date Cepeda-Cuervo and Cifuentes-Amado (2020) have made the important contributions of obtaining moment expressions and displaying its suitability for regression modeling. The fact that we can find explicit expressions for moments highlights the suitability of the distribution for use in 'big data' contexts such as business, government, and research applications. This is because strictly numerical estimation of fundamental parameters can become burdensome in big data contexts. Aside from Cepeda-Cuervo and Cifuentes-Amado (2020), it seems little else about this distribution has been published to this . We extend the literature on this distribution in this paper by providing its MGF, CF and CDF. We then describe three new equivalent alternative formulations (types) for the distribution. We examine these in depth to reveal their relative advantages for different situations. We also present modeling of the mixture weight for increasing understanding of possible latent classes and examine a large real-world data set.

This paper is organized as follows. Section 2 presents the two main approaches for modeling overdispersed binomial data. Section 3 describes the tilted beta-binomial distribution and presents foundational results. Section 4 contains modeling applications from both maximum-likelihood and Bayesian viewpoints. Finally Section 5 presents conclusions and opportunities for future research.

2 Literature

The two main families of methods for handling overdispersion are the conditional approaches such as random-effects modeling arising out of the generalized linear model tradition and the unconditional binomial mixture distribution approach (Anderson 1988, Dey et al. 1997). In terms of the first approach, Williams (1982) redefined the binomial variance function to include an overdispersion parameter and then conducted quasilikelihood modeling conditional on the overdispersion parameter and a linear predictor using the logit link function (see also McCullagh and Nelder 1989, ch. 4). Such quasilikelihood approaches do not use a full distributional specification. Instead, the approach is formulated in terms of a mean function and a variance function (McCullagh and Nelder 1989) and then the model is estimated iteratively to minimize an error function. In related work generalized estimating equations for extra-binomial variation were presented by Zeger and Liang (1986). Williams (1982) also described the use of random effects models in terms of iteratively reweighted least-squares estimation. Stiratelli et al. (1984) introduced a random-effects logistic regression model for repeated binary data where overdispersion arose from random effects that followed a multivariate normal distribution. Ochi and Prentice (1984) presented a model where data was taken to be binomial conditional on a multivariate probit link function transformation of the linear predictor. Additional examples of random-effects models for overdispersed binomial data include Chen and Ahn (1997).

As for the second approach, characterizing the data unconditionally using a beta-binomial distribution in terms of an overdispersed binomial dates back to at least Skellam (1948). The beta-binomial distribution has been extensively applied to modeling overdispersion (Williams 1975, Anderson 1988, Johnson et al. 2005). Other mixture-based approaches have been used to handle overdispersion. For example, Morel and Nagaraj (1993) discussed the use of a finite mixture of binomials to model overdispersion in binomial and multinomial data. Brooks et al. (1997) introduced a finite mixture model where the data was assumed to follow a binomial distribution with probability θ and beta-binomial with probability $(1 - \theta)$. Morton (1991) formed a ratio of composed of overdispersed Poisson variables to get an extended negative hypergeometric distribution. He described that the moments of this distribution are typically not available in a closed form, but that quasi-likelihood methods can be used to model data according to this distribution.

Extensions of the beta-binomial have also been proposed throughout the statistical literature (Wilcox 1981). More closely related to the current work, Rodríguez-Avi et al. (2007) presents an extension of the beta-binomial distribution wherein the mixing distribution is the generalized beta distribution. This distribution shares the currently proposed distribution's ability to handle excessive skewness. However, the mean, variance, and other moments of the generalized beta-binomial distribution are not available in closed form and must be found numerically (Rodríguez-Avi et al. 2007, p. 57). A recent bivariate extension by this research team also lacks explicit moment expressions (Olmo-Jiménez et al. 2011). Chen and Novick (1984) also examine the posterior distribution arising from a binomial distribution that has been given a generalized beta prior.

Some authors have separately applied both conditional and unconditional approaches to a given data set. For example, Dean (1992) discusses fitting binomial data with the beta-binomial distribution or modeling using a logistic link function and normal random effects. Molenberghs et al. (2012) observe that the beta-binomial distribution can be re-expressed as a binomial distribution with beta-distributed random effects. They therefore propose a modeling approach in which the random beta-distributed and normally-distributed random effects appear simultaneously (Molenberghs et al. 2012). However, Molenberghs et al. (2012) focus on a conditional modeling perspective and it does not appear that the distribution implied by their model has a closed-form expression or has closed-form moments.

One problem with the conditional approaches based on non-beta random effects is that they may become increasingly computationally demanding as the size of the data set increases. Monte-Carlo based approaches encounter particular difficulty with massive data sets, requiring either extensive run-times and memory requirements or else ingenious strategies on the part of the researcher (see, e.g., Huang and Gelman 2005). The absence of closed-form expressions for essential terms such as expectations and variances necessitates large computational effort that can become extreme with the massive data sets encountered with regularity in today's environment. The current paper shows over multiple datasets that tilted beta-binomial distribution has the desirable properties of handling large amounts of overdispersion while still retaining the tractability of closed-form expressions.

The beta-binomial distribution has attracted increased attention since the work of Skellam (1948). Johnson et al. (2005, §6.9.2) provided a literature review of early references through the 1970s. More recent references are discussed in the previous section (e.g., Brooks et al. 1997, Crowder 1978, Gange et al. 1996, Gelman and Sichel 1987, Hu et al. 2018, Luo and Paul 2018, Shiyomi et al. 2000) as well as previously in this section. As for the literature on the tilted beta-binomial, to date Cepeda-Cuervo and Cifuentes-Amado (2020) have made the important contributions of obtaining moment expressions. They have also displayed its suitability for regression modeling in a seed germination data set involving seeds on 21 plates. We extend the literature on this distribution in this paper by providing its MGF, CF and CDF. We then describe three new equivalent alternative formulations (types) for the distribution in Section 4.2.1. We examine these in depth to reveal their relative advantages for different situations. Modeling of the mixture weight is also examined in Section in the context of a large real-world data set.

3 The Tilted Beta-Binomial Distribution

3.1 Background

Hahn and López Martín (2015) proposed the tilted beta distribution. The density function of a random variable X following the tilted beta distribution with $\alpha > 0$, $\beta > 0$, $v \in [0,1]$, and $\theta \in [0,1]$ is

$$p(x|\alpha,\beta,v,\theta) = \begin{cases} (1-\theta)\big(2-2v+(4v-2)x\big) + \theta\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1} & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The tilted beta distribution is a mixture distribution where the beta distribution becomes predominant as θ increases, while the tilting distribution becomes dominant as θ decreases. As seen in (1), the tilting distribution has the following density function

$$p(x|v) = \begin{cases} 2-2v+(4v-2)x & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The tilting distribution is also a special case of the generalized Topp and Leone distribution (Kotz and van Dorp 2004) under a slightly different parameterization.

The tilted beta is capable of expressing larger variances than the beta distribution. As θ declines, the tilting distribution introduces more density at the endpoints. When $v = 1/2$ the tilting distribution is equivalent to the rectangular distribution. In this case, the rectangular component allocates density uniformly across the support. This special case is called the beta rectangular distribution (Hahn 2008) and it has seen use in project management (Herrerías-Velasco et al. 2011, García et al. 2011, Gładysz et al. 2015) as well as beta rectangular regression modeling (Bayes et al. 2012). When v is some other value, the tilting component allocates more density to one endpoint and less density to the other. This property of the v parameter allows for greater flexibility in terms of expressing skewness than does the standard beta distribution. To date the tilted beta distribution has been applied principally in project management contexts (García Pérez et al. 2016, Udoumoh et al. 2017, Salas-Morera et al. 2018) and more recently it has been used for regression modeling (Hahn 2021).

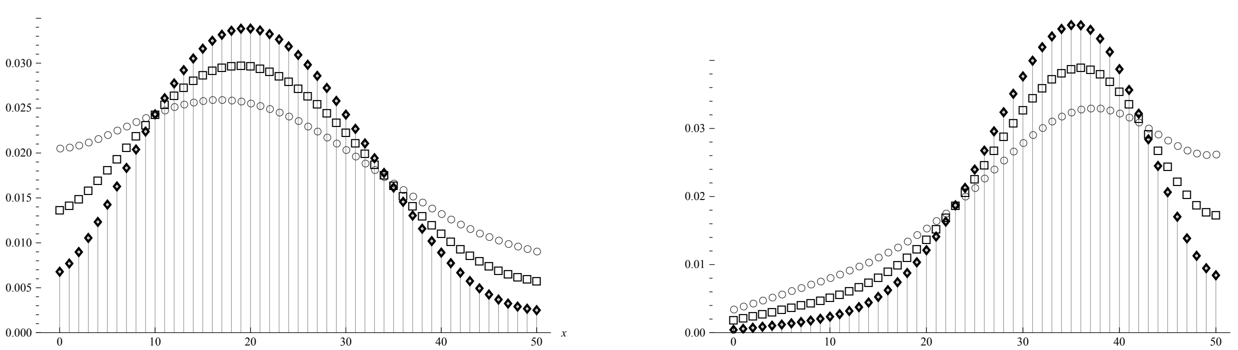
3.2 The Tilted Beta-Binomial Distribution and Previous Literature

Compounding the tilted beta in (1) with the binomial distribution leads to a distribution called the tilted beta-binomial distribution. A number of foundational aspects of this distribution have been discussed by Cepeda-Cuervo and Cifuentes-Amado (2020) under a different parameterization. We review some of these previous foundations in this subsection. The distribution's probability mass function (PMF) is

$$p(x|n,\alpha,\beta,v,\theta) = (1-\theta)\frac{2\big(x(2v-1)-nv+n+1\big)}{(n+2)(n+1)} + \theta\binom{n}{x}\left(\frac{B(x+\alpha,n-x+\beta)}{B(\alpha,\beta)}\right), \tag{3}$$

where x and n are integers such that $0 \leq x \leq n$ and $B(\cdot,\cdot)$ is the Beta function. Also we have $\alpha > 0$, $\beta > 0$, $v \in [0,1]$, and $\theta \in [0,1]$ as for the tilted beta distribution. Plots of the PMF of the tilted beta-binomial distribution with $n = 50$ appear in Figure 1.

Figure 1: Examples of Tilted Beta-Binomial Distributions



$v = 0.3, \alpha = 3, \beta = 4$ $\quad$ $v = 0.9, \alpha = 3, \beta = 4$

Both plots: circles: $\theta = 0.25$; squares: $\theta = 0.5$; diamonds: $\theta = 0.75$.

This distribution has closed-form moment expressions for the expected value and variance. Regarding the expected value, we have

$$E(x) = n\left((1-\theta)\frac{v+1}{3} + \theta\frac{\alpha}{\alpha+\beta}\right) \tag{4}$$

while for the variance we may first find

$$E(x^2) = n\left((1-\theta)\frac{2nv+n+1}{6} + \theta\frac{\alpha(n+n\alpha+\beta)}{(\alpha+\beta)(\alpha+\beta+1)}\right). \tag{5}$$

Noting that the variance is $E(x^2) - E(x)^2$ we observe the variance is expressible in closed form as

$$\mathrm{Var}(x) = n\left((1-\theta)\frac{2nv+n+1}{6} + \theta\frac{\alpha(n+n\alpha+\beta)}{(\alpha+\beta)(\alpha+\beta+1)}\right) - n^2\left((1-\theta)\frac{v+1}{3} + \theta\frac{\alpha}{\alpha+\beta}\right)^2. \quad (6)$$

Higher moments can be obtained from the moment generating function which is described in the next section.

3.3 The Beta Rectangular Binomial Distribution and Previous Literature

As mentioned, the beta rectangular distribution is a special case of the tilted beta distribution when $v = 1/2$. Here the tilting component is flat giving a rectangular or uniform distribution. Accordingly the results for beta rectangular binomial distribution arise by making the substitution $v = 1/2$ in the relevant formulas above. In particular, the PMF is

$$\begin{aligned} p(x|n,\alpha,\beta,\theta) =& (1-\theta)\frac{\big((1-v)(n+x) - vx - 1\big)}{(n+2)(n+1)/2} \\ &+ \theta\binom{n}{x}\left(\frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(n+\alpha+\beta)}\right). \end{aligned} \quad (7)$$

For completeness the mean and variance are, as in Cepeda-Cuervo and Cifuentes-Amado (2020),

$$E(x) = n\left(\frac{1-\theta}{2} + \theta\frac{\alpha}{\alpha+\beta}\right) \quad (8)$$

and

$$\mathrm{Var}(x) = n\left((1-\theta)\frac{2n+1}{6} + \theta\frac{\alpha(n+n\alpha+\beta)}{(\alpha+\beta)(\alpha+\beta+1)}\right) - n^2\left(\frac{1-\theta}{2} + \theta\frac{\alpha}{\alpha+\beta}\right)^2. \quad (9)$$

Further theoretical topics related to the definition of tilted beta-binomial regression models such as possible choices of Bayesian priors, possible choices of link functions, and possible model comparisons using DIC (Spiegelhalter et al. 2002) can be found in Cepeda-Cuervo and Cifuentes-Amado (2020).

3.4 The Tilted Binomial Distribution

Here we present additional results for the tilted binomial distribution. This is the first component of the mixture in (3), namely

$$p(x|n,v) = \frac{2\big(x(2v-1) - nv + n + 1\big)}{(n+2)(n+1)}. \quad (10)$$

The expected value and variance can be found with (10) or by setting $\theta = 0$ in (4) and (6). In particular they are

$$E(x) = \frac{n(v+1)}{3} \quad (11)$$

and

$$\mathrm{Var}(x) = \frac{n\,(2nv(1-v) + n + 3)}{18}. \quad (12)$$

For clarity we mention that the below foundational results of this section have not appeared in the literature on the tilted beta-binomial distribution to date. We find the tilted binomial distribution's moment generating function (MGF) is

$$M_X^{TBN}(t) = \frac{e^t(n+2)(v-1) - (n+2)ve^{(n+1)t} + e^{(n+2)t}(nv+1) - nv + n + 1}{(n+1)(n+2)\left(e^t - 1\right)^2}, \tag{13}$$

where TBN is used to indicate tilted binomial. By independence of the two components of the mixture, the MGF of the tilted beta-binomial distribution is then

$$M_X^{TBB}(t) = (1-\theta)M_X^{TBN}(t) + \theta M_X^{BB}(t), \tag{14}$$

where $M_X^{BB}(t)$ is the MGF of the beta-binomial distribution. Generating functions for the beta-binomial can be found in reference material (Johnson et al. 2005, ch. 6). The characteristic function (CF) of the tilted binomial is

$$C_X^{TBN}(t) = \frac{e^{\imath t}(n+2)(v-1) - (n+2)ve^{(n+1)\imath t} + e^{(n+2)\imath t}(nv+1) - nv + n + 1}{(n+1)(n+2)\left(e^{\imath t} - 1\right)^2}, \tag{15}$$

where $\imath$ is the unit imaginary number. The CF for the tilted beta-binomial follows by independence of the two components of the mixture and the use of the CF for the beta-binomial.

We may also wish to use the CDF of the tilted binomial in certain contexts. The CDF is

$$P(X \leq x) = \frac{\big(x(2v-1) + 2n(1-v) + 2\big)(x+1)}{(n+2)(n+1)}. \tag{16}$$

Monte Carlo simulation from the tilted binomial distribution is another topic that does not seem to have been examined previously. It can be accomplished by noting the compound nature of the distribution. First, one simulates from the tilting distribution for a given value of v to obtain a simulated $\tilde{p}$. Then one draws from the binomial with parameters n and $\tilde{p}$. The tilting distribution itself can be specified as a mixture of a Beta(1,2) distribution and a Beta(2,1) distribution with mixing weights $(1-v)$ and v respectively. Equivalently, the tilting distribution can be decomposed into a mixture of two triangular distributions: one with a mode at 0 (left-triangular) and the other at the mode of 1 (right-triangular). Stein and Keblis (2009) provide methods for simulating various triangular distributions with support $(0,1)$ which can be applied to the tilting distribution with support $(0,1)$ and known v. Five random draws from the uniform distribution, $u_1, \ldots, u_4$, will be required. The quantity $t_1 = \min(u_1, u_2)$ will be a random draw from the left triangular distribution. Similarly the quantity $t_2 = \max(u_3, u_4)$ will be a random draw from the right triangular distribution. Lastly a random draw from the tilting distribution, $\tilde{p}$, is created by taking a draw from t_2 with probability v and a draw from t_1 otherwise.

4 Applications

4.1 Maximum Likelihood Estimation

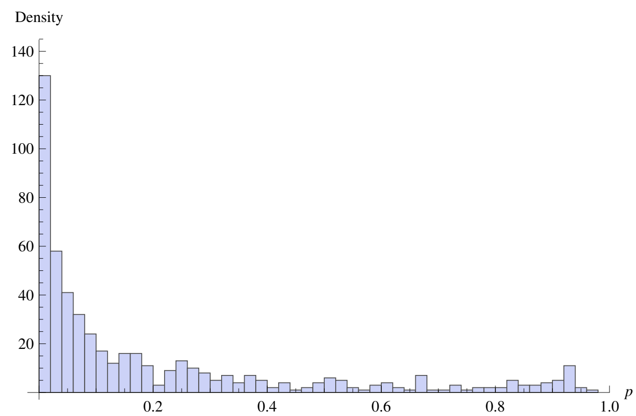
To examine the utility of the tilted beta-binomial distribution, we apply it to data from the 2010 U.S. Census. In this data set, we extract data for the state of Maryland at the Census place level. There are 518 Census places in Maryland. We denote x as the number of people indicating they were both African-American and of one race in a Maryland Census place. We denote n as the

Table 1: Maryland Data Set: Distribution Parameter Estimates and Log Likelihoods

Distribution	α	β	θ	v	Log likelihood
Beta binomial	0.373	1.332	—	—	-3497.37
Tilted beta-binomial	0.534	5.125	0.787	0.792	-3475.71

number of people indicating they were of one race in a Maryland Census place. The total of x for this extracted data set is over 1.5 million and the total of n is over 4.6 million. We fit the beta-binomial distribution and the tilted beta-binomial distribution to this data set. Maximum-likelihood estimation of distribution parameters was performed in Mathematica 8.0. Parameter estimates appear in Table 1.

Figure 2: Proportion of African-Americans in 2010 Census Places in Maryland



We examine model fit by the likelihood ratio test. The likelihood ratio test gives a value of $\chi^2 = 43.42$ on 2 degrees of freedom ($p < 0.0001$), indicating that the tilted beta-binomial distribution provides a better fit than the beta-binomial distribution. However, as can be seen in (3), when θ is at the boundary value of 1 the tilted beta-binomial distribution collapses to the beta-binomial distribution. Since theory justifying the distribution of the likelihood ratio test involves parameters being on the interior of the possible range as opposed to the boundary, the standard p value is not correct. Self and Liang (1987) studied the large-sample behavior of the likelihood ratio test under various kinds of boundary problems. Our case here corresponds to their Case 6 in that we have one additional parameter of interest (θ) with possible true value on the boundary and one additional parameter of interest (v) which has possible true value on the interior of the range. The limiting distribution of the likelihood ratio test statistic is in this case a 50:50 mixture distribution of a χ^2 distribution with one degree of freedom and a χ^2 distribution with two degrees of freedom (Self and Liang 1987, p. 608). Since this mixture would produce an even smaller p-value than a χ^2 distribution with two degrees of freedom, the p-value reported above is conservative.

4.2 Bayesian Estimation

We may also use Bayesian methods and Markov chain Monte Carlo (MCMC) estimation to fit the tilted beta-binomial model. It is well known that Bayesian methods may offer some flexibility and this flexibility can be used in estimation of tilted beta-binomial models. We provide four such types in Section 4.2.1 below.

For MCMC it is preferable to work with a common re-parameterization of the beta distribution. One transforms α and β to μ and ϕ by the relations

$$\begin{aligned} \mu &= \frac{\alpha}{\alpha+\beta}, \\ \phi &= \alpha+\beta. \end{aligned} \tag{17}$$

Here μ is the expected value of the beta distribution and ϕ can be thought of as a term in the precision or the 'sample size' of the beta distribution. In MCMC μ and ϕ will often be less correlated in the posterior than will be α and β, plus they are more interpretable.

4.2.1 Types of Tilted Beta-Binomial Models

Four types (or parameterizations) of tilted beta-binomial models are illustrated in Table 2, only one of which (Type A) has been discussed previously in the literature. The two vertical columns express two different ways that the likelihood can be expressed. In the column entitled Marginalized Likelihood, we have used the likelihood where the underlying distribution has been explicitly integrated out. For example, if x follows a Binomial(n, p) distribution, and p itself has a tilted beta distribution as in (1), then the marginal distribution of x is the tilted beta-binomial distribution (3) which has been abbreviated to f_{TBB}. To make the Table more compact, we omit reference to the parameters of the tilted beta-binomial distribution in the Table and instead refer the reader to (3). We may then specify priors for the parameters in (3) for a Bayesian model (reference to the priors' parameters will be discussed later).

However, it is not necessary that this integration be performed explicitly. Bayesian hierarchical models capitalize on the fact that a distribution may be placed on a parameter in the likelihood, and then at another level another distribution may be placed on parameter on the distribution just mentioned. It is common to refer to these higher level distributions as hyperpriors (e.g., Clayton 1996). Here we slightly different words for the purposes of clarity in discussing the different types of models in Table 2. We use the word 'priors' in the Table to refer to the distributions at the top level of the hierarchy. These are the distributions for which the Bayesian must provide a certain amount of his or her subjective input, such as his or her decision to use certain values to represent prior information or non-informativeness. For the purposes of Table 2, subjective decisions about 'priors' in the Table are the same as the commonly discussed subjective decisions about Bayesian prior specification. We use the word 'hyperlikelihood' as a compromise word to indicate that the distribution is neither a likelihood or a 'prior' as defined above. Instead it is a distribution that is one level above the likelihood in terms of a hierarchical model. Also, it is a distribution of focus in the paper because the differences in the four types of models involve the likelihoods (and the hyperlikelihoods if present). Readers who prefer a more hierarchical Bayesian way of thinking may prefer to substitute the word prior for our hyperlikelihood and hyperprior for our prior. These readers may also want to substitute the words hierarchical model for Conditional Likelihood in the Table's last column. We can now discuss the Type B model listed in the Table. Here the likelihood of x is binomial. Next, p follows the tilted beta distribution as in (1). We abbreviate the tilted beta distribution as f_{TB}. Finally, specification of priors completes the Bayesian model.

We also observe that the tilted beta-binomial distribution is a finite mixture model. This type of model can be thought of in terms of having latent classes. Dempster et al. (1977) introduced the EM algorithm and showed how it can be applied to many statistical contexts. In addition

Table 2: Four Types of Tilted Beta-Binomial Models

	Marginalized Likelihood	Conditional Likelihood
Continuous Mixture	*Type A* Likelihood: $x \sim f_{TBB}$ Priors	*Type B* Likelihood: $x \sim \text{Binomial}(n, p)$ HyperLikelihood: $p \sim f_{TB}$ Priors
Latent Class/'EM'	*Type C* Likelihood: $x \sim f_{TBN}^{(1-z)} f_{BB}^{z}$ $z \sim \text{Bernoulli}(\theta)$ Priors	*Type D* Likelihood: $x \sim \text{Binomial}(n, p)$ HyperLikelihood: $p \sim f_{TB}^{(1-z)} f_{B}^{z}$ $z \sim \text{Bernoulli}(\theta)$ Priors

they described a method for estimating finite mixture models that involves a certain change to the likelihood function. Suppose we have two components to the mixture and the two components have densities $f_1(\boldsymbol{\omega}_1)$ and $f_2(\boldsymbol{\omega}_2)$, where $\boldsymbol{\omega}_j$ is a parameter vector for jth density. Then Dempster et al. (1977, pp. 15–16) write the likelihood function for the ith observation as

$$f_1(\boldsymbol{\omega}_{1i})^{1-z_i} f_2(\boldsymbol{\omega}_{2i})^{z_i} \tag{18}$$

where z_i is an unobserved or latent Bernoulli variable. Upon taking logarithms this simplifies to either $\log(f_1(\boldsymbol{\omega}_{1i}))$ if $z_1 = 0$ or $\log(f_2(\boldsymbol{\omega}_{2i}))$ otherwise. Using the notation of this paper, we can specify that the latent z_is depend on a mixing Bernoulli parameter θ. In Bayesian modeling we can use the Gibbs sampler to estimate the latent z vector (Gelfand and Smith 1990, McLachlan and Krishnan 1997) conditional on θ. This strategy is known as data augmentation (Tanner and Wong 1987). Zaslavsky (2003) describes data augmentation as the converse to marginalization: the increase in the number of the parameters enhances tractability and creates simplifications in the model. Here we have the Type C model of Table 2 based on the latent class conceptualization. For completeness we can provide the model labeled Type D in the Table. This model has a binomial likelihood but the hyperlikelihood has a Dempster et al. (1977) latent class formulation. It would be possible to create additional variations of the above models. For example, in Section 3.4 we discussed how the tilting distribution can be thought of as a mixture of triangular distributions or as a mixture of beta distributions. We leave such variations to future research.

We estimate the four types of models from Table 2 using the Maryland data set above. We parameterize the beta component of the models in terms of μ and ϕ as described at the beginning of this section. For completeness we also provide results for α and β by estimating these parameters during the MCMC run. Uniform priors on (0,1) were used for μ, θ and v. A uniform (0,100) prior was used for ϕ. For Type C and D models, the latent z parameters were given a Bernoulli(θ) prior. After a 5000 iteration burn-in, posterior estimates were based on 200,000 iterations of the Markov chain. Table 3 displays the results. For clarity we point out that Type B, C, and D models have not appeared in the literature. We see that the posterior means are all very similar, with some slight differences apparent in the estimates of ϕ. These differences in ϕ translate to slight differences in the estimates of β.

We may also calculate the Monte Carlo (MC) standard errors for the parameters for the different types of models. These are displayed in Table 4. As we might expect, the Type A model has

Table 3: Maryland Data Set: Bayesian Parameter Estimates by Model Type

Para-	Posterior Means				Posterior Standard Deviations			
meter	*Type A*	*Type B*	*Type C*	*Type D*	*Type A*	*Type B*	*Type C*	*Type D*
α	0.544	0.544	0.544	0.544	0.045	0.045	0.045	0.045
β	5.911	5.925	5.915	5.906	1.336	1.372	1.352	1.348
μ	0.087	0.087	0.087	0.087	0.013	0.013	0.013	0.013
ϕ	6.454	6.469	6.458	6.450	1.365	1.402	1.381	1.378
v	0.648	0.647	0.647	0.648	0.191	0.192	0.192	0.192
θ	0.746	0.746	0.746	0.746	0.057	0.058	0.057	0.057

the lowest MC error due to the fact that the mixing distribution has been explicitly integrated out. Model Type B has somewhat greater MC error. Separate calculations on greater precision numbers indicate that Type B has about 11% greater MC error than does Type A averaging the bottom four parameters in the Table. Recall that the first two parameters in the Table, α and β, are transformations of estimated parameters μ and ϕ. Hence these two parameters have been excluded from the computation of the average MC error to eliminate redundancy. For Type C, the average MC error is approximately 50% greater than that of Type A on average when including only the bottom four parameters of the Table. Model Type C has 518 latent Bernoulli variables that are also estimated along with the parameters in the Table. The uncertainty of these parameters propagates to the parameters shown in the Table. Finally Type D has the highest average MC error. However, the average MC error for Type D is only about 4.5% larger than that of Type C. In summary the conditional likelihood approaches of Table 2 have slightly larger MC errors than their marginalized likelihood counterparts. By comparison, moving to a latent class formulation with unobserved z variables can increase MC error noticeably when number of trials is substantial as in the current data set.

It is customary in Bayesian modeling to present model fit indices such as DIC (Spiegelhalter et al. 2002) and the effective number of parameters for DIC, p_D, or WAIC (Watanabe 2010) and p_{WAIC}, when presenting results such as in Table 3. However, all four Types of model displayed are conceptually equivalent to one another so the issue of which one fits "best" is not easily answerable. Indeed, differences in model structuring are known to lead to differences in the values of DIC. This issue causes concern for some authors (Smith 2002) while others describe this as versatility (DeIorio and Robert 2002). As observed by Smith (2002), comparison of DIC for our different Types of models is unlikely to be meaningful as they correspond to separate types of prediction problems. Here the effective number of parameters will include contributions from z_i for Types C and D. These will increase with the sample size, potentially making the effective number of parameters considerably greater for these model Types. Given the above, we have omitted these measures from Table 3. Nonetheless it is possible to see whether the tilted beta-binomial model fits better than a beta-binomial model. For the Type A tilted beta-binomial model, we found DIC to be 6958.15 ($p_D = 3.150$) and WAIC to be 6958.47 ($p_{\text{WAIC}} = 3.447$). A marginalized (Type A) beta-binomial model was created with the same Uniform (0,1) prior for μ and Uniform (0,100) prior for ϕ. We used 5000 iterations for burn-in and 200,000 iterations for estimation. We found DIC to be 6998.72 ($p_D = 1.988$) and WAIC to be 6998.81 ($p_{\text{WAIC}} = 2.060$) for the beta-binomial model. The large differences in DIC and WAIC suggest the tilted beta-binomial model fits better than the beta-binomial model in this dataset.

In summary Type A appears to be the most attractive for reducing MC error in the current dataset. We may therefore wonder if Type B is needed since Type B appears to be less competitive

Table 4: Maryland Data Set: Monte Carlo Error Estimates by Model Type

	Posterior Means			
Parameter	*Type A*	*Type B*	*Type C*	*Type D*
α	0.0002	0.0003	0.0003	0.0004
β	0.0138	0.0158	0.0195	0.0212
μ	0.0001	0.0002	0.0002	0.0002
ϕ	0.0140	0.0161	0.0198	0.0215
v	0.0018	0.0019	0.0028	0.0028
θ	0.0006	0.0006	0.0009	0.0009

than Type A. However, if it is necessary to calculate the CDF of either observed values of x or of predicted values of x, then Type B may be more attractive than Type A especially as the values of x increase. The CDF of the binomial distribution is readily computed by most statistical packages and can be easily programmed elsewhere using the incomplete and the complete beta functions. The CDF of the tilted beta-binomial distribution is less attractive to compute. We have found an exact expression for its CDF, but it is cumbersome and involves regularized generalized hypergeometric functions in the beta-binomial component. The Type A CDF could instead be found by resorting to brute force evaluation and summation of probabilities but again the simplicity of the Type B CDF may be more attractive. The CDF will be needed in certain situations such as the existence of truncated data or the calculation of exceedances. Type C and Type D have even greater MC errors. Still these models can be attractive when we wish to model θ_i as a function of other explanatory values. We now turn to an example of this.

4.2.2 Bayesian Modeling of θ_i

We may expand a Type C or a Type D model by allowing θ to vary over observations, θ_i, and adding the following predictive specification.

$$\begin{aligned} z_i &\sim \text{Bernoulli}(\theta_i), \\ \text{logit}(\theta_i) &= a_0 + a_1 w_{1i} + \ldots + a_k w_{ki}, \\ a_0, \ldots, a_k &\sim \text{Normal}(\cdot). \end{aligned} \tag{19}$$

Here $w_1, \ldots, w_k$ are observed predictor variables, $a_0, \ldots, a_k$ are coefficients to be estimated and Normal$(\cdot)$ indicates a normal distribution with the user's choice of prior mean and prior precision. Possible variations of (19) include using an alternative link function such as the probit link, different priors for $a_0, \ldots, a_k$, and other changes. We re-estimate the Type D model of Table 3 using (19). There are many variables we might use to predict θ_i including the relative counts of another demographic variable or else geographical aspects of the Census place. For simplicity we suppose that more populated places might have a different probability of being in the beta-binomial component versus the tilted binomial component of the mixture. We therefore use the logarithm of n_i as our predictor variable w_1. This variable was centered around its mean. We used normal distributions with means of zero and precisions of 0.04 as priors for a_0 and a_1. All other priors were as before. Estimates were based on 200,000 iterations of the Markov chain after 5000 iterations of burn-in. Table 5 contains the parameters estimates for the two models.

The Table shows that v has changed noticeably after the introduction of the predictor for θ_i. Without the predictor, v is larger with a posterior mean of 0.648. This value indicates a positively

Table 5: Parameter Estimates in Two Type D Models

	Type D without predictor			*Type D with predictor*		
Parameter	Mean	SD	95% CI	Mean	SD	95% CI
α	0.544	0.045	(0.462, 0.640)	0.581	0.066	(0.462, 0.722)
β	5.906	1.348	(3.958, 9.165)	14.150	2.792	(9.460, 20.390)
μ	0.087	0.013	(0.060, 0.112)	0.040	0.005	(0.031, 0.051)
ϕ	6.450	1.378	(4.443, 9.768)	14.730	2.841	(9.954, 21.070)
v	0.648	0.192	(0.303, 0.978)	0.176	0.046	(0.093, 0.273)
θ	0.746	0.057	(0.623, 0.839)			
a_0				0.079	0.222	(-0.361, 0.512)
a_1				-1.151	0.145	(-1.459, -0.891)

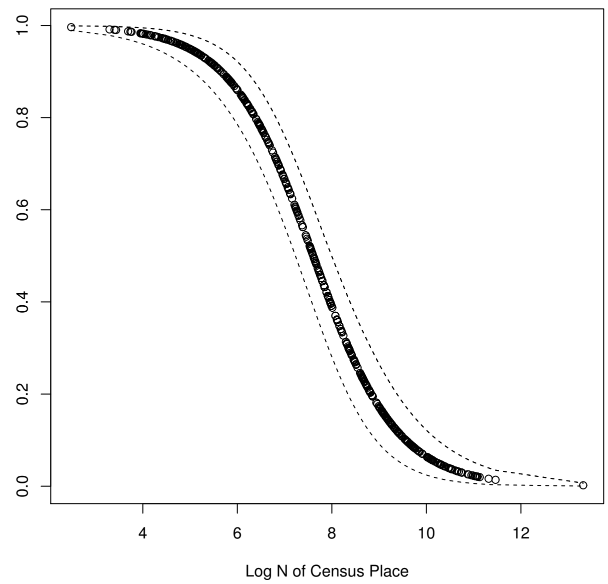
inclining tilt that favors larger values of p being more probable. In addition v has noticeable uncertainty as shown by its standard deviation of 0.192. When the predictor is added, v changes to 0.176 which indicates a declining tilt. The standard deviation of v also becomes much smaller (0.046). The 95% posterior credible intervals (CIs) for v in the two models, (0.303, 0.978) and (0.093, 0.273), do not overlap. While this observation does not constitute formal model comparison and while there are other ways of assessing the magnitude of the parameter difference such as embedding both models into a meta-model and estimating the difference directly, it does provide some preliminary evidence that the predictor for θ_i has had at least a modest impact on other parameters. A similar set of observations can be made about μ and ϕ. For example, ϕ becomes larger (i.e., the beta component becomes more precise) in the model that has the predictor for θ_i. Also at the informal level we can see the 95% CIs for ϕ do not overlap.

We can examine model fit somewhat less informally using the deviance since both models are the same type. We find that the model with the predictor in Table 5 has a somewhat smaller value of WAIC (WAIC = 4065.3, p_{WAIC} = 252.36) indicating improved model fit compared to the model without the predictor (WAIC = 4069.8, p_{WAIC} = 254.73). We also find that the model with the predictor has a somewhat smaller value of DIC (DIC = 4162.5, p_D = 445.97) indicating improved model fit compared to the model without the predictor (DIC = 4163.8, p_D = 446.46). Notice that the difference in DIC is not as large and so the evidence is weaker. However, DIC may be affected by parameter skewness. Most of the parameters in Table 5 have noticeable amounts of skewness. Given the above, the values of WAIC may be more useful since they take into account the entirety of the posterior distributions including skewness.

We now turn to a discussion of the coefficients in the predictor for θ_i. We begin by mentioning that the number of people of one race living in a Census place in Maryland during the 2010 US Census varies widely depending on the Census place. At the large end, Baltimore city had over 608,000 people in this category while the second most populous Census place was Columbia with less than 100,000 people. At the minimum, Port Tobacco Village town had less than 20 people of one race. Taking the logarithm of n_i was beneficial for accommodating this wide range of values.

We find in Table 5 that the intercept a_0 is not far from zero and its 95% credible interval includes zero. However, the a_1 coefficient is farther from zero with a posterior mean of -1.151. The 95% credible interval for this parameter excludes zero by a large margin. We may regard this as evidence that fit was improved when larger values of $\log(n_i)$ were associated with smaller values of θ_i. This means that Census places with relatively small numbers of people such as Port Tobacco Village town tend to be allocated to the beta-binomial component of the mixture. Conversely

Figure 3: Posterior Means and 95% Credible Intervals of θ_i in Type D Model



Census places with relatively large numbers of people (such as Baltimore, Columbia, and Silver Spring, MD) tend to be allocated to the tilted binomial component of the mixture. A plot of the 518 posterior means of θ_i appears in Figure 3. Figure 3 also displays splines fit to 2.5% posterior quantiles of θ_i and to the 97.5% posterior quantiles of θ_i with dashed lines for better readability. Since the median number of people in a Maryland Census place is not especially large (about 2200), we may wish to see if this asymmetry justifies the use of an asymmetric link function for the predictor such as the complementary log-log link. One benefit of the Type D (and C) model is that choices of link functions can be explored. We therefore re-estimated the Type D model using the specification of (19) but with the logit link being replaced with the complementary log-log. Interestingly, this change did cause the 95% credible interval for a_1 to exclude the value of zero. However, the model fit measures (WAIC = 4068.11, DIC = 4164.1) showed that this change decreased model fit overall.

4.2.3 Bayesian Modeling of μ_i and v_i

It is also possible to provide a regression model for other parameters of the tilted beta-binomial distribution. Possibilities include

$$\text{logit}(\mu_i) = b_0 + b_1 u_{1i} + \ldots + b_k u_{ki} \tag{20}$$

and

$$\text{logit}(v_i) = c_0 + c_1 t_{1i} + \ldots + c_k t_{ki}, \tag{21}$$

where $u_1, \ldots, u_k$ and $t_1, \ldots, t_k$ are predictor variables and $b_0, \ldots, b_k$ and $c_0, \ldots, c_k$ are coefficients. Again priors will be needed for $b_0, \ldots, b_k$ and $c_0, \ldots, c_k$. We provide an example of (20) using the current data set. We use the number of people in a Census place who were Asian and of one race as a predictor. Due to the existence of large values and also values of zero, we took the logarithm

Table 6: Parameter Estimates for Type D Model with Predictor for μ_i and θ_i

Parameter	Mean	SD	95% CI
ϕ	15.389	3.220	(10.237, 22.773)
v	0.358	0.100	(0.195, 0.588)
a_0	0.747	0.237	(0.306, 1.235)
a_1	-0.614	0.104	(-0.833, -0.424)
b_0	-2.718	0.130	(-2.975, -2.466)
b_1	0.269	0.025	(0.221, 0.321)

of 1 plus the number of people in a Census place who were Asian and of one race to be u_{1i}. This transformation closely approximates the logarithm of its argument for moderate to large values of the argument but remains defined when the argument is zero as can happen here. This variable was centered around its mean. We used normal distributions with means of zero and precisions of 0.04 as priors for b_0 and b_1. All other priors were as before (except for the prior for μ which was necessarily removed). Estimates were based on 200,000 iterations of the Markov chain after 5000 iterations of burn-in. Table 6 contains the parameters estimates for the two models.

The Table shows that the transformed independent variable u_{1i} did predict the dependent variable. The b_1 coefficient is positive and its 95% posterior credible interval excludes the value of zero. The posterior mean of b_0 is -2.718. Table 6 shows that the posterior distribution of v has again changed with the model extension. We see the posterior mean of v has increased to 0.358. We may compare the model of Table 6 to the model that has a predictor in Table 5 We find that the current model has a much better fit according to WAIC (WAIC = 4041.6 p_{WAIC} = 246.1). Similarly we find that the current model has a much better fit according to DIC (DIC = 4149.9, p_D = 448.5). We briefly consider the issue of prior sensitivity for ϕ. As an alternative prior, we refit the model using an exponential distribution with rate parameter set to 0.02. We found the posterior distribution for ϕ using this prior was similar to the posterior distribution for ϕ using the existing uniform prior.

4.2.4 Computational Issues in Practice

It may also be of interest to examine computational topics in this relatively underexplored distribution. As an example we compare the run times for the models of Table 2 in Section 4.2.1. Here we limit our observations to estimation in the WinBUGS software family. The tilted beta distribution and the tilted beta-binomial distribution do not exist natively in WinBUGS, having been described after WinBUGS' initial creation. Fortunately there are methods for incorporating new distributions in WinBUGS using special methods described in the software's manual. We could have recorded run times based only on these special methods. However the Type B and Type D models use the binomial distribution which *does* exist in WinBUGS. Since the binomial distribution has obviously been compiled into WinBUGS internal code, it is reasonable to conclude that the internal binomial distribution would be computed faster than would a special-method approach that would have to be computed at run time without prior compiling. This would have skewed the timing results. We then considered using a special-method binomial distribution for the Type B and Type D models. However, there is the issue of uncertainty about the effect of multiple special-methods in one model and also about the generalizability of WinBUGS special methods to other software environments.

We took the more labor-intensive route of using WinBUGS development tools to compile all likelihood and hyperlikelihood distributions of this paper into WinBUGS native code (WinBUGS was written in a language called Component Pascal). For increased computational comparability, we also compiled our own version of the binomial distribution into Component Pascal for use in the timing runs. These compiled codes were then made available for WinBUGS to call. To summarize, the timing tests used author-compiled versions of the binomial, tilted beta, and tilted beta-binomial distributions. All of the parameters labeled as Priors in Table 2 made use of existing distributions in WinBUGS such as the uniform and normal distributions. Hence, they were already computationally comparable and there was no need to write Component Pascal code for them.

Runs were timed for 200,000 iterations after a 5,000-iteration burn-in was completed. We found that the Type A model took about 19% more time to run than did the Type B model. We also found that the Type C model took about 17% more time to run than did the Type D model. This gives additional nuance to the decision process about model Type. At least in the WinBUGS environment, the Type B and Type D models run faster than their marginalized counterparts. This means their Monte Carlo errors could be potentially reduced by completing additional iterations in a certain amount of time. As discussed in Section 4.2.2, the latent class formulation does have some useful features mitigating against the longer run times. The slower run times of Types A and C are likely due to the fact that these Types have likelihoods that are more computationally expensive to calculate. Evaluating (3) is more computationally expensive than is evaluating the binomial likelihood alone. In this dataset there are 518 likelihood terms to be evaluated per iteration. However in Types B and D the expensive calculation of (3) only needs to occur once per iteration since (3) is a hyperlikelihood in these model Types.

Additional information comes from a review of the effective sample sizes produced by the 200,000 iteration runs of the four model Types. The effective sample size (Plummer et al. 2006) attempts to quantify the autocorrelated information content of the samples from the Markov chain as if they were produced by independent sampling. Lower values are thus less attractive since effective sample sizes decrease as chain autocorrelation increases. The results appear in the first four columns of Table 7 using our Maryland data set. Here we see the effects of increased chain autocorrelation in Types B and D respectively versus Types A and C respectively. We also see that the latent class formulations of Type C and D have lower effective sample sizes than the continuous mixture Types A and B. The rightmost four columns of Table 7 are the effective sample sizes divided by the amount of time in seconds of the respective runs. This gives the effective sample size per unit of time (in seconds) of the four Types. In the previous paragraph we described the faster run times of Types B and D. We see for example Type A μ estimation produced 2.496 effectively independent MCMC samples per second but that Type B μ estimation produced 3.112 effectively independent MCMC samples per second. We see that Type B has less attractive effective sample sizes versus A when the run size is fixed (at 200,000 iterations here). However, its inexpensive likelihood calculation makes it preferable in terms of a fixed run time since it is producing nearly 25% more effective samples per second for μ (and 20% to 35% more samples per second for the other parameters). Types C and D look considerably less attractive in terms of overall effective samples and effective samples per second. Type D is slightly preferable versus Type C in terms of effective samples per second but the gap is narrower. To summarize, Types B and D may be worth considering when the number of rows of data increases. Here we have 518 rows of data and Types B and D are more attractive than Types A and C respectively. We would expect this performance gap to increase in big data applications. However, for smaller data sets Type B and D may have less attractive performance than A and C respectively.

Table 7: Effective Sample Sizes and Effective Sample Sizes per Unit of Time by Model Type

Para-meter	Effective Sample Sizes				Effective Sample Sizes per Unit of Time			
	Type A	*Type B*	*Type C*	*Type D*	*Type A*	*Type B*	*Type C*	*Type D*
μ	9724.9	7443.0	4180.6	3899.6	2.496	3.112	0.769	0.839
ϕ	9647.2	7094.6	4605.2	4011.9	2.476	2.966	0.847	0.863
v	12193.8	10073.6	4597.1	4323.6	3.130	4.211	0.846	0.930
θ	10342.4	8257.4	4097.5	3765.4	2.655	3.452	0.754	0.810

5 Conclusions

This paper provides an extensive treatment and many new results for the tilted beta-binomial distribution, including its MGF and CF. We also discuss the newly-introduced Type B, C, and D tilted beta-binomial models that may be considered for implementation in applied research. Of course, the choice of which model type is best for a particular situation depends on which of several factors is most important to the data analyst. The data analyst may consider whether he or she is most concerned with minimizing MC error on a per-iteration basis, having a more tractable CDF, or taking advantage of a latent class formulation. Accordingly we discussed the implications of choice of model type with respect to these new considerations. Run time is another factor that may be pertinent to a data analyst, and we examined this topic in detail. Simulation from the distribution was also introduced to the literature.

The tilted beta-binomial distribution nests the beta-binomial distribution as a special case when $\theta = 1$ and the tilted binomial when $\theta = 0$. We provided new results for the latter distribution. These special cases necessarily lead to the issues of hypothesis testing and model comparison. We discussed likelihood-based hypothesis testing of the tilted beta-binomial against the beta-binomial in Section 4.1 (new to the literature), and Bayesian model comparison using DIC and WAIC in Section 4.2. In the current dataset we found that the tilted beta-binomial distribution was more appropriate than was the beta-binomial for the models considered here. We also discuss predictive modeling of θ in Section 4.2.2.

The current paper also appears to be the first application of this distribution to 'big data'. The total sample size for the analyses was over 4.6 million distributed over a sample size of 518 binomial trials. We find evidence for the existence of both a beta-binomial component and a tilted binomial component, which generates implications for decision makers seeking to better understand latent statewide demographic trends for African-Americans, Asians, and other demographic groups. It would have been possible to use the dataset to examine other demographic data such as age, education or other factors. We leave these analyses for future research. It would also be possible to extend our estimation of a single parameter ϕ to a regression model where ϕ_i varied over observations. Much of the above discussion would apply to this kind of extension. Since ϕ exists on the space of non-negative real numbers, certain straightforward adjustments would be needed. We also note that it is possible to perform likelihood-based model fit assessment of mixtures using parametric bootstrap techniques (Suesse et al. 2017). We leave this for future research. In closing, the tilted beta-binomial distribution has potential to be a useful new tool in the analysis of binomial data.

References

Aldirawi H, Yang J (2022) Modeling sparse data using MLE with applications to microbiome data. Journal of Statistical Theory and Practice 16(1), Article number: 13

Anderson DA (1988) Some models for overdispersed binomial data. Australian Journal of Statistics 30(2):125–148

Bayes CL, Bázan JL, García C (2012) A new robust regression model for proportions. Bayesian Analysis 7(4):841–866

Brooks SP, Morgan BJT, Ridout MS, Pack SE (1997) Finite mixture models for proportions. Biometrics 53(3):1097–1115

Cepeda-Cuervo E, Cifuentes-Amado MV (2020) Tilted beta binomial linear regression model: A Bayesian approach. Journal of Mathematics and Statistics 16(1):1–8

Chen JJ, Ahn H (1997) Marginal models with multiplicative variance components for overdispersed binomial data. Journal of Agricultural, Biological, and Environmental Statistics 2(4):440–450

Chen JJ, Novick MR (1984) Bayesian analysis for binomial models with generalized beta prior distributions. Journal of Educational Statistics 9(2):163–175

Clayton DG (1996) Generalized linear mixed models. In: Gilks WR, Richardson S, Spiegelhalter DJ (eds) Markov Chain Monte Carlo in Practice, Chapman & Hall/CRC, Boca Raton, FL, pp 275–301

Cox DR (1983) Some remarks on overdispersion. Biometrika 70(1):269–274

Crowder MJ (1978) Beta-binomial anova for proportions. Journal of the Royal Statistical Society Series C (Applied Statistics) 27(1):34–37

Dean CB (1992) Testing for overdispersion in Poisson and binomial regression models. Journal of the American Statistical Association 87(418):451–457

DeIorio M, Robert CP (2002) Discussion on the paper by Spiegelhalter, Best, Carlin, and van der Linde. Journal of the Royal Statistical Society, Series B 64(4):619–620

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society Series B (Methodological) 39(1):1–38

Dey DK, Gelfand AE, Peng F (1997) Overdispersed generalized linear models. Journal of Statistical Planning and Inference 64(1):93–107

Ferrari A, Comelli M (2016) A comparison of methods for the analysis of binomial clustered outcomes in behavioral research. Journal of Neuroscience Methods 274:131–140

Gange SJ, Munoz A, Saez M, Alonso J (1996) Use of the beta-binomial distribution to model the effect of policy changes on appropriateness of hospital stays. Journal of the Royal Statistical Society Series C (Applied Statistics) 45(3):371–382

García C, García Pérez J, van Dorp J (2011) Modeling heavy-tailed, skewed and peaked uncertainty phenomena with bounded support. Statistical Methods and Applications 20(4):463–486

García Pérez J, López Martín MdM, García García C, Sánchez Granero MA (2016) Project management under uncertainty beyond beta: The generalized bicubic distribution. Operations Research Perspectives 3:67–76

Gelfand AE, Smith AFM (1990) Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association 85(410):398–409

Gelman E, Sichel HS (1987) Library book circulation and the beta-binomial distribution. Journal of the American Society for Information Science 38(1):4–12

Gładysz B, Skorupka D, Kuchta D, Duchaczek A (2015) Project risk time management – A proposed model and a case study in the construction industry. Procedia Computer Science 64:24–31

Hahn ED (2008) Mixture densities for project management activity times: A robust approach to PERT. European Journal of Operational Management 188(2):450–459

Hahn ED (2021) Regression modeling with the tilted beta distribution: A Bayesian approach. The Canadian Journal of Statistics 49(2):262–282

Hahn ED, López Martín MdM (2015) Robust project management with the tilted beta distribution. SORT-Statistics and Operations Research Transactions 39(2):253–272

Haining R, Law J, Griffith D (2009) Modelling small area counts in the presence of overdispersion and spatial autocorrelation. Computational Statistics & Data Analysis 53(8):2923–2937

Hedt-Gauthier BL, Mitsunaga T, Hund L, Olives C, Pagano M (2013) The effect of clustering on lot quality assurance sampling: A probabilistic model to calculate sample sizes for quality assessments. Emerging Themes in Epidemiology 10(11)

Herrerías-Velasco J, Herrerías-Pleguezuelo R, van Dorp J (2011) Revisiting the PERT mean and variance. European Journal of Operational Research 210(2):448 – 451

Hu T, Gallins P, Zhou Y (2018) A zero-inflated beta-binomial model for microbiome data analysis. Stat 70:271–289

Huang Z, Gelman A (2005) Sampling for Bayesian computation with large datasets. Unpublished manuscript available at `http://www.stat.columbia.edu/~gelman/research/unpublished/comp7.pdf`, accessed Aug. 17, 2012

Johnson NL, Kemp AW, Kotz S (2005) Univariate Discrete Distributions, 3rd edn. Wiley, Hoboken, NJ

Kotz S, van Dorp JR (2004) Beyond Beta: Other Continuous Families of Distributions with Bounded Support. World Scientific, Singapore

Luo R, Paul S (2018) Estimation for zero-inflated beta-binomial regression model with missing response data. Statistics in Medicine 37(26):3789–3813

McCullagh P, Nelder JA (1989) Generalized Linear Models, 2nd edn. Chapman & Hall, London

McLachlan G, Krishnan T (1997) The EM Algorithm and Extensions. Wiley, New York

Molenberghs G, Verbeke G, Iddi S, Demétrio CGB (2012) A combined beta and normal random-effects model for repeated, overdispersed binary and binomial data. Journal of Multivariate Analysis 111:94–109

Morel JG, Nagaraj NK (1993) A finite mixture distribution for modelling multinomial extra variation. Biometrika 80(2):363–371

Morton R (1991) Analysis of extra-multinomial data derived from extra-Poisson variables conditional on their total. Biometrika 78(1):1–6

Ochi Y, Prentice RL (1984) Likelihood inference in a correlated probit regression model. Biometrika 71(3):531–543

Olmo-Jiménez MJ, Martínez-Rodríguez AM, Conde-Sánchez A, Rodríguez-Avi J (2011) A generalization of the bivariate beta-binomial distribution. Journal of Statistical Planning and Inference 141(7):2303–2311

Plummer M, Best N, Cowles K, Vines K (2006) CODA: Convergence diagnosis and output analysis for MCMC. R News 6:7–11

Rodríguez-Avi J, Conde-Sánchez A, Sáez-Castillo AJ, Olmo-Jiménez MJ (2007) A generalization of the beta-binomial distribution. Journal of the Royal Statistical Society Series C (Applied Statistics) 56(1):51–61

Salas-Morera L, Arauzo-Azofra A, García-Hernández L, Palomo-Romero J, Ayuso-Muñoz J (2018) New approach to the distribution of project completion time in pert networks. Journal of Construction Engineering and Management 144(10):04018094

Self S, Liang KY (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. Journal of the American Statistical Association 82(398):605–610

Shiyomi M, Takahashi S, Yoshimura J (2000) A measure for spatial heterogeneity of a grassland vegetation based on the beta-binomial distribution. Journal of Vegetation Science 11(5):627–632

Skellam JG (1948) A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. Journal of the Royal Statistical Society Series B (Methodological) 10(2):257–261

Smith J (2002) Discussion on the paper by Spiegelhalter, Best, Carlin, and van der Linde. Journal of the Royal Statistical Society, Series B 64(4):619–620

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) Bayesian measures of model complexity and fit (with discussion). Journal of the Royal Statistical Society, Series B 64(4):583–639

Stein WE, Keblis MF (2009) A new method to simulate the triangular distribution. Mathematical and Computer Modelling 49(5–6):1143–1147

Stiratelli R, Laird N, Ware JH (1984) Random-effects models for serial observations with binary response. Biometrics 40(4):961–971

Suesse T, Rayner JCW, Thas O (2017) Assessing the fit of finite mixture distributions. Australian & New Zealand Journal of Statistics 59(4):463–483

Tanner MA, Wong WH (1987) The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82(398):528–540

Udoumoh EF, Ebong DW, Iwok IA (2017) Simulation of project completion time with Burr XII activity distribution. Asian Research Journal of Mathematics 6(4):1–14

Watanabe S (2010) Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. Journal of Machine Learning Research 11:3571–3594

Wilcox RR (1981) A review of the beta-binomial model and its extensions. Journal of Educational Statistics 6(1):3–32

Williams DA (1975) The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. Biometrics 31(4):949–952

Williams DA (1982) Extra-binomial variation in logistic linear models. Journal of the Royal Statistical Society Series C (Applied Statistics) 31(2):144–148

Zaslavsky AM (2003) Hierarchical Bayesian modeling. In: Press SJ (ed) Subjective and Objective Bayesian Statistics: Principles, Models and Applications, Wiley, Hoboken, NJ, pp 336–358

Zeger SL, Liang KY (1986) Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42(1):121–130