Modeling under or over-dispersed binomial count data by using extended Altham distribution families

Senay Asma

McMaster University, 1280 Main St W, Hamilton, Ontario, Canada

Abstract

While aiming particularly at handling under-dispersion, we explore a type of models constructed conservatively using the minimum information of first two moments for the fitting of binomial count data, which could have under, equal or over-dispersion. The extended Altham distribution (EAD) families were presented in this study. The extended Altham families are very close to the binomial distribution under equal dispersion setting, implying that they are alternative models of the binomial distribution. The feature that extended Altham families can reach the full range of dispersion outperforms some commonly used models such as extended beta-binomial and quasi-binomial which have restricted ranges of dispersion. Moreover, the extended Altham family can have double peaks at two boundaries, indicating they are feasible for fitting the double tail inflation phenomenon. This study illustrated the modeling using extended Altham families for both under-dispersed and over-dispersed binomial data resulted from disease cases within the same family.

1. Introduction

Binomial count data, a type of count data with bounded supports, arise from many disciplines such as toxicological study, medical research, ecology, agriculture, logistics management, linguistics, electronic engineering, political science, and so on. This type of data are often associated with an important quantity called proportion which is the study purpose. For the binomial count data, the most commonly used model is binomial distribution. The binomial random variable (rv) is the sum of independent and identically distributed (iid) Bernoulli rv's which have a fixed success probability for value 1. This success probability is the interested population proportion.

However, the above binomial setting is too ideal and simple. In reality, there could exist more complicated situations. For example, the success probability may be a rv instead of a fixed constant, or the Bernoulli rv's may positively or negative correlated (corresponding to attraction or repulsion). The data could even result from an aggregation of subsets

with varying upper bounds. Thus, observations could appear to be over-dispersion or under-dispersion relative to the binomial distribution.

Handling over-dispersion has received a great deal of attention and is quite mature. The common way is to us binomial mixture. Allowing varying success probabilities in the binomial distribution can yield a mixture with over-dispersion relative to the binomial distribution. A widely used model is the beta-binomial which is the binomial mixture of the beta distribution, i.e., the success probability follows a beta distribution. Refer to Wilcox [18] for a review of beta-binomial and its extensions.

However, in reality, the under-dispersion can occur, especially in the repulsion situation. Bailey [4] reported the repulsion examples of function word counts, which are under-dispersed relative to binomial due to the nature that a function word can not follow itself in general. Assuming negative correlation among Bernoulli rv's, the sum of them will result in a distribution of under-dispersion relative to the binomial. See Theorem 7.1 in Joe [12]. Viveros-Aguilera, Balasubramanian and Balakrishnan [17] constructed a concrete example using the homogeneous Markov chain for binary response. In addition, quasi-binomial and its variations prescribe non-homogeneous dependence mechanisms for successive trials by Chakraborty and Das [5]. We show another possibility leading to under-dispersion in Section 2, which is a mixture of varying upper bounds of supports.

Prentice [15] extended the beta-binomial to allow limited under-dispersion. Consul [8] proposed the quasi-binomial (type I) using an urn model, in which the success probability of the i-th trial has an additional part proportional to i ($i > 1$). This additional part in the success probability can be negative or positive, resulting in under-dispersion or over-dispersion, but both under- and over-dispersions are bounded. Some extensions of quasi-binomial can be found in Mishra, Tiwary and Singh [14], Dobson, Carreras and Newman [9], Chakraborty and Das [5] and some advanced studies in Altham [2,3]. Other models using particular mechanisms like Bailey [4] were practised in the literature too. Although there are many attempts to handle the under-dispersion case, none of them becomes a mature tool for a general case.

The descriptive statistics are not always as easy as might be expected, particularly when data exhibit skewness and/or outliers. A relevant example is given by Chatfield [6] which involves the number of issues of a particular monthly magazine read by 20 people in a year. In this example, the data has bimodal U-shape which is even more difficult to summarize than a skewed distribution. Therefore, the sample mean and standard deviation are potentially very misleading. The proportion of regular readers is a useful statistic, but it may be sensible to describe the data in words rather than with summary statistics.

Since binomial count data can arise from complex situations, none of existing models provides a unified way to handle them. Thus, there is a need to develop a unified model capable of handling various dispersion situations. To this end, we construct models with specified mean and variance using the entropy method. The resulted two-parameter models can reach the full range of dispersion, providing a unified way for modelling binomial count data with different dispersion case.Also, numerical comparison shows that the proposed models are quite close to the binomial distribution in the equal-dispersion setting. Hence, they are alternative to the binomial model in the equal-dispersion setting.

In summary, the proposed two-parameter models have the ability to better fit various binomial count data in a unified way. Based on our proposed models, we have found that the Altham distribution [1] is a special case by reparametrization. Thus, this finding uncovers the feature of full dispersion of the Altham distribution. To credit Altham, the models we proposed are named as the extended Altham distribution families (EAD).

The remainder of this paper is organized as follows. We define new exponential families in Section 2, with the computational algorithm for the probability mass function (pmf).

MLEs are derived in Section 3. We conduct simulation study and illustrate data examples in Section 4. A brief discussion is given in Section 5.

2. Model construction

In this section, we shall present the construction of extended Altham distribution families by using Kullback-Leibler (KL) divergence measure. KL is non-symmetric measure defined by

$$KL(p_i||q_i) = \sum_{i=0}^{M} p_i \log(\frac{p_i}{q_i}) \tag{2.1}$$

and it gives the distance between two probability distributions, P and Q, where Q is given distribution and P is unknown probability distribution. The distribution Q is known as a priori distribution.

For example, if a priori distribution Q is considered as a discrete uniform distribution assigns equal probability $1/(M+1)$ to every point in the support, then the closest distribution in sense of KL measure will be the distribution that has the maximum uncertainty in the support, leading to the maximum entropy. For a discrete distribution, denote the probability mass function (pmf) as $\Pr[X = i] = p_i$, $(i = 0, 1, \ldots, M)$, the mean as μ and variance as σ^2. Given the information of a priori distribution Q, mean and variance, KL optimization defines the distribution which obtain the probability distribution which satisfy minimum KL distance. Encouraging probability assignment in the support as even as possible, thus, taking advantage of given information in a minimum and conservative sense. That is

$$\min \left\{ \sum_{i=0}^{M} p_i \log(\frac{p_i}{q_i}) \right\}, \tag{2.2}$$

subject to three constrains

$$\sum_{i=0}^{M} p_i = 1, \qquad \sum_{i=0}^{M} ip_i = \mu, \qquad \sum_{i=0}^{M} i^2 p_i = \sigma^2 + \mu^2. \tag{2.3}$$

There is no explicit form of pmf in terms of parameters μ and σ^2, however, there is an explicit form in terms of Lagrangian multipliers β's:

$$p_i = q_i C(\beta_1, \beta_2) e^{i\beta_1 + i^2\beta_2}, i = 0, 1, \ldots, M, \tag{2.4}$$

where $C(\beta_1, \beta_2)$ is the normalizing constant.

Note that if Q is considered as a binomial distribution, the pmf will be Altham distribution [1]. Thus, we call this family as extended Altham distribution family. In the following, we give a formal definition.

Definition 2.1. (extended Altham distribution family): A rv X is said to be from the extended Altham distribution family, denoted as extended Altham(M, h, β_1, β_2) where $-\infty < \beta_1, \beta_2 < \infty$, if its probability mass function (pmf) is of form:

$$p_i \propto h_i \exp(\beta_1 i + \beta_2 i^2), \qquad i = 0, 1, \ldots, M, \tag{2.5}$$

where h_i is an arbitrary function with positive values and β_1 and β_2 are real parameters and satisfy

$$\sum_{i=0}^{M} \Pr[X = i] = \sum_{i=0}^{M} h_i C(\beta_1, \beta_2) \exp(i\beta_1 + i^2\beta_2) = 1, \tag{2.6}$$

$$E[X] = \sum_{i=0}^{M} i \Pr[X=i] = \sum_{i=0}^{M} i h_i C(\beta_1, \beta_2) \exp(i\beta_1 + i^2\beta_2) = \mu, \tag{2.7}$$

$$E[X^2] = \sum_{i=0}^{M} i^2 \Pr[X=i] = \sum_{i=0}^{M} i^2 h_i C(\beta_1, \beta_2) \exp(i\beta_1 + i^2\beta_2) = \sigma^2 + \mu^2, \tag{2.8}$$

where $C(\beta_1, \beta_2)$ is the normalizing constant.

β_1 and β_2 seem to govern the increasing or decreasing speed of pmf, but no direct connection with the mean and variance. The parametrization in terms of μ and σ^2 has clear explanation, however, no analytical pmf available. But this can be compensated by numerical solution.

Since constrain (2.7) implies

$$C^{-1}(\beta_1, \beta_2) = \sum_{i=0}^{M} h_i e^{i\beta_1 + i^2\beta_2} \tag{2.9}$$

hence, there are only two independent parameters: β_1 and β_2. For any discrete distribution on the support $\{0, 1, \ldots, M\}$, since

$$\mu = \sum_{i=1}^{M} i p_i = E[1 \times X] \le E[X^2] \le E[M \times X] = M \sum_{i=1}^{M} i p_i = M\mu, \tag{2.10}$$

the natural ranges of μ and σ^2 are

$$0 \le \mu \le M, \qquad \max(0, \mu - \mu^2) \le \sigma^2 = E[X^2] - \mu^2 \le M\mu - \mu^2. \tag{2.11}$$

There is no restriction for parameters μ and σ^2, thus, these two parameters can vary in their full ranges shown in (2.14). However, the ranges of β_1 and β_2 can not be determined in explicit forms.

When $M = 1$, the rv X degenerates to the Bernoulli case, and only one parameter is needed. Thus, we exclude this extreme case for the upper bound of the support, and only consider $M \ge 2$.

When $M = i$, the pmf can be expressed in terms of h_i and $\beta = (\beta_1, \beta_2)$:

$$p_i = log(h_i) + \beta[(i+1)^\alpha - i^\alpha], \tag{2.12}$$

where $\alpha > 0$ and h_i are arbitrary positive valued function.

The extended Altham distribution has only two independent parameters: β_1 and β_2.

For any discrete distribution on the support $\{0, 1, \ldots, M\}$, since

$$\mu = \sum_{i=1}^{M} i p_i = E[1 \times X] \le E[X^2] \le E[M \times X] = M \sum_{i=1}^{M} i p_i = M\mu, \tag{2.13}$$

the natural ranges of μ and σ^2 are

$$0 \le \mu \le M, \qquad \max(0, \mu - \mu^2) \le \sigma^2 = E[X^2] - \mu^2 \le M\mu - \mu^2. \tag{2.14}$$

There is no restriction for parameters μ and σ^2, thus, these two parameters can vary in their full ranges shown in (2.14). However, the ranges of β_1 and β_2 can not be determined in explicit forms.

The binomial distribution is usually referred as the equally-dispersed distribution. Assume $Y \sim binomial(M, p)$ which has pmf

$$p_i = \binom{M}{i} p^i (1-p)^{M-i}, \qquad 0 \le p \le 1, \quad i = 0, 1, \ldots, M. \tag{2.15}$$

Then $E[Y] = Mp$ and $Var[Y] = Mp(1-p)$. The ratio of variance to mean is $\frac{Var[Y]}{E[Y]} = 1 - p = 1 - \frac{E[Y]}{M}$. A discrete distribution on the same support is said to be under-dispersed or over-dispersed if its ratio is smaller or bigger than that of the binomial distribution of the same mean. That is, the comparison is regarded to the binomial distribution of the same mean.

For convenience, we define the dispersion index for discrete distribution on the support $\{0, 1, \ldots, M\}$ as follows

$$D = \frac{Var[Y]}{E[Y]\,(1 - E[Y]/M)}. \tag{2.16}$$

Then, a discrete distribution on the support $\{0, 1, \ldots, M\}$ is said to be under-dispersed, equally-dispersed or over-dispersed if its dispersion index defined in (2.16) is smaller than, equal to or bigger than 1 respectively. Obviously, the binomial distribution is equally-dispersed. However, other distributions can be equally-dispersed too.

According to (2.14), the full range of dispersion is

$$\max\left(0, \frac{1-\mu}{1-\mu/M}\right) = \frac{\max(0, \mu(1-\mu))}{\mu(1-\mu/M)} \leq D \leq \frac{M\mu - \mu^2}{\mu(1-\mu/M)} = M. \tag{2.17}$$

Note that the lower bound is $\frac{1-\mu}{1-\mu/M} > 0$ when $0 \leq \mu < 1$, and 0 otherwise. When M is large, the interval $(0, 1)$ for under-dispersion is very narrow comparing with the interval $(1, M)$ for over-dispersion, one might uses $\log(D)$ as the dispersion index. But to keep consistent with the convention, we use (2.16).

The over-dispersion is usually explained by a mixture of binomial, say the beta-binomial. We have found that under-dispersion could be caused by a mixture too, but of varying upper bounds of supports. Here we illustrate using a simple example of two-component binomial mixture.

Let $X_1 \sim \text{binomial}(M_1, p_1)$ and $X_2 \sim \text{binomial}(M_2, p_2)$, where $M_1 < M_2$. Assume $E[X_1] = E[X_2] = \mu < M_1$. Denote $I \sim \text{Bernoulli}(p)$, and define Y conditional on I as follows

$$[Y|I = 1] \sim \text{binomial}(M_1, p_1), \qquad [Y|I = 0] \sim \text{binomial}(M_2, p_2). \tag{2.18}$$

Note that the support of Y is $\{0, 1, \ldots, M_2\}$. Then

$$E[Y] = E\{E[Y|I]\} = pE[X_1] + (1-p)E[X_2] = \mu, \tag{2.19}$$

$$\begin{aligned} Var[Y] &= E[(Y-\mu)^2] = E\{E[(Y-\mu)^2|I]\} \tag{2.20} \\ &= pVar[X_1] + (1-p)Var[X_2] \\ &= p\mu(1-\mu/M_1) + (1-p)\mu(1-\mu/M_1) \\ &= \mu\{1 - [p\mu/M_1 + (1-p)\mu/M_2]\} \\ &< \mu(1-\mu/M_2). \end{aligned}$$

Thus

$$D = \frac{\mu\{1 - [p\mu/M_1 + (1-p)\mu/M_2]\}}{\mu(1-\mu/M_2)} < \frac{\mu(1-\mu/M_2)}{\mu(1-\mu/M_2)} = 1, \tag{2.21}$$

implying that Y is under-dispersed.

In order to illustrate the extended Altham distribution family, we considered the following models with different h_i functions:

$$\text{Model 1.}\quad h_i = 1; \quad \textit{flat (Discrete Uniform)} \tag{2.22}$$

$$\text{Model 2.}\quad h_i = log(M - X + 1) + 1; \quad \textit{decreasing} \tag{2.23}$$

$$\text{Model 3.}\quad h_i = \frac{M!}{(M - X)!(X)!}; \quad \textit{convex (Weighted Binomial)} \tag{2.24}$$

$$\text{Model 4.}\quad h_i = \frac{(M - X)!(X)!}{M!}; \quad \textit{concave} \tag{2.25}$$

$$\text{Model 5.}\quad h_i = X + 1; \quad \textit{increasing} \tag{2.26}$$

$$\text{Model 6.}\quad h_i = \frac{1}{(X + 1)}; \quad \textit{decreasing} \tag{2.27}$$

$$\text{Model 7.}\quad h_i = M - X + 1; \quad \textit{decreasing} \tag{2.28}$$

$$\text{Model 8.}\quad h_i = \frac{1}{(M - X + 1)}; \quad \textit{increasing} \tag{2.29}$$

$$\text{Model 9.}\quad h_i = X(M - X) + 1; \quad \textit{convex} \tag{2.30}$$

$$\text{Model 10.}\quad h_i = \frac{1}{X(M - X) + 1}; \quad \textit{concave} \tag{2.31}$$

$$\text{Model 11.}\quad h_i = log(X + 1) + 1; \quad \textit{increasing} \tag{2.32}$$

$$\text{Model 12.}\quad h_i = \frac{1}{log(X + 1) + 1}; \quad \textit{decreasing.} \tag{2.33}$$

The dispersion index for extended Altham(μ, σ^2) is $D = \frac{\sigma^2}{\mu(1-\mu/M)}$, which can reach the full range of dispersion because of no restriction on parameters μ and σ^2. Since σ^2 is independent of μ, D could be smaller than, equal to or bigger than 1. Therefore, the extended Altham family covers all dispersion situations. The extended Altham distribution family given by 2.5 includes Binomial distribution when the function $h_i = 1$,

Weighted Binomial distributions Zelterman [19] when the function h_i is the binomial coefficient and so Altham distribution [1] because it is known to be an example of a weighted binomial model.

For comparison purposes, we need reparametrization so that we can fix (μ, σ^2). Figure 1 and 2 displays the pmf profiles of the extended Altham distributions with h_i functions given by 2.22-2.33, mean $\mu = 5$ and various dispersions using the developed numerical algorithm. Comparing with Binomial distribution (red line), the under-dispersed extended Altham distributions (green lines) seem to have larger probability masses around the mean, while the over-dispersed extended Altham distributions (blue lines) attempt to have more masses at two boundaries. When the dispersion large enough, the pmf shows U-shape, like that of the beta-binomial distributions.

Since the extended Altham distribution can have equal dispersion, it is natural to compare it with the binomial distributions under the same means.

Figure 3 and 4 demonstrate some of them on the support $\{0, 1, \ldots, 40\}$. We see that both pmf's are very close when the mean is not close to the two boundaries. When the mean close to two boundaries, there are slight differences among two distributions, and the extended Altham distribution assigns more masses at 0 or M. For many values of M, we check the maximum absolute difference of pmf of two distributions under the same mean, and find that this maximum is no more than 3% when the mean close to boundaries, and becomes smaller when the mean close to the center of the support. The larger the M, the smaller the maximum of probability difference. From the viewpoint of distribution theory, this suggests that the binomial distribution can be approximated by the extended Altham distribution. On the other hand, for the distribution constructed using the minimum information of mean and equal-dispersion, the binomial distribution is
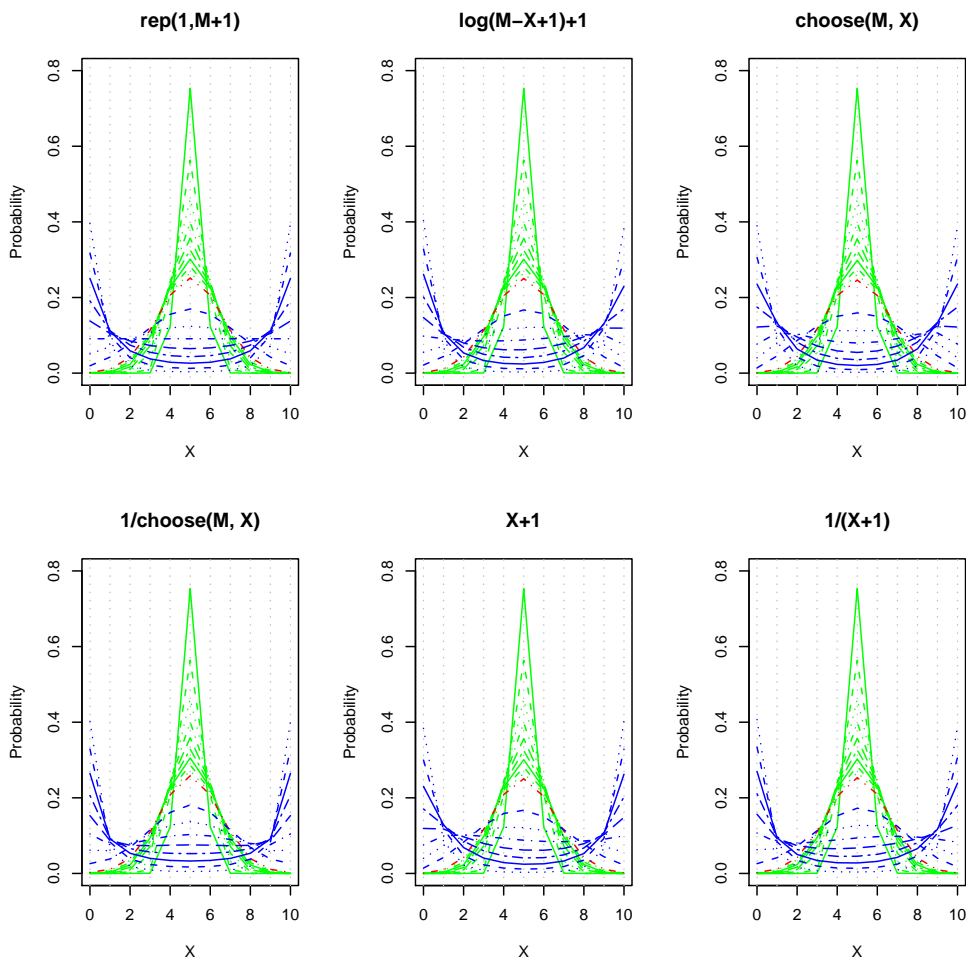
Figure 1. Probability profiles of the extended Altham distributions of mean $\mu = 5$ and various dispersions regarding to h_i given by 2.22 and 2.27 The red line indicates the equal dispersion. The blue lines correspond to over dispersions of $2, 3, \ldots, 9$, while the green lines shows under dispersions of $0.1, 0.2, \ldots, 0.9$. The most centered extended Altham distributions with the largest mass at 5 has dispersion 0.1, and the most spread extended Altham distributions with largest masses at two boundaries has dispersion 9.

very close to it. Thus, from the aspect of modelling, such a fact implies that the extended Altham distribution could be an alternative of the binomial distribution if the mean is not extremely small or large.

Note that the extended beta-binomial and quasi-binomial can handle both under-dispersion and over-dispersion too. The beta-binomial distribution is constructed using mixture. Assume the success probability in binomial distribution $p \sim \text{beta}(a, b)$ $(a > 0, b > 0)$, the pmf of beta-binomial(M, a, b) is

$$p_i = \binom{M}{i} \frac{B(a+i, b+M-i)}{B(a, b)}, \qquad i = 0, 1, \ldots, M, \tag{2.34}$$

where $B(x, y)$ is the complete beta function. See Hasemann and Kupper [11].
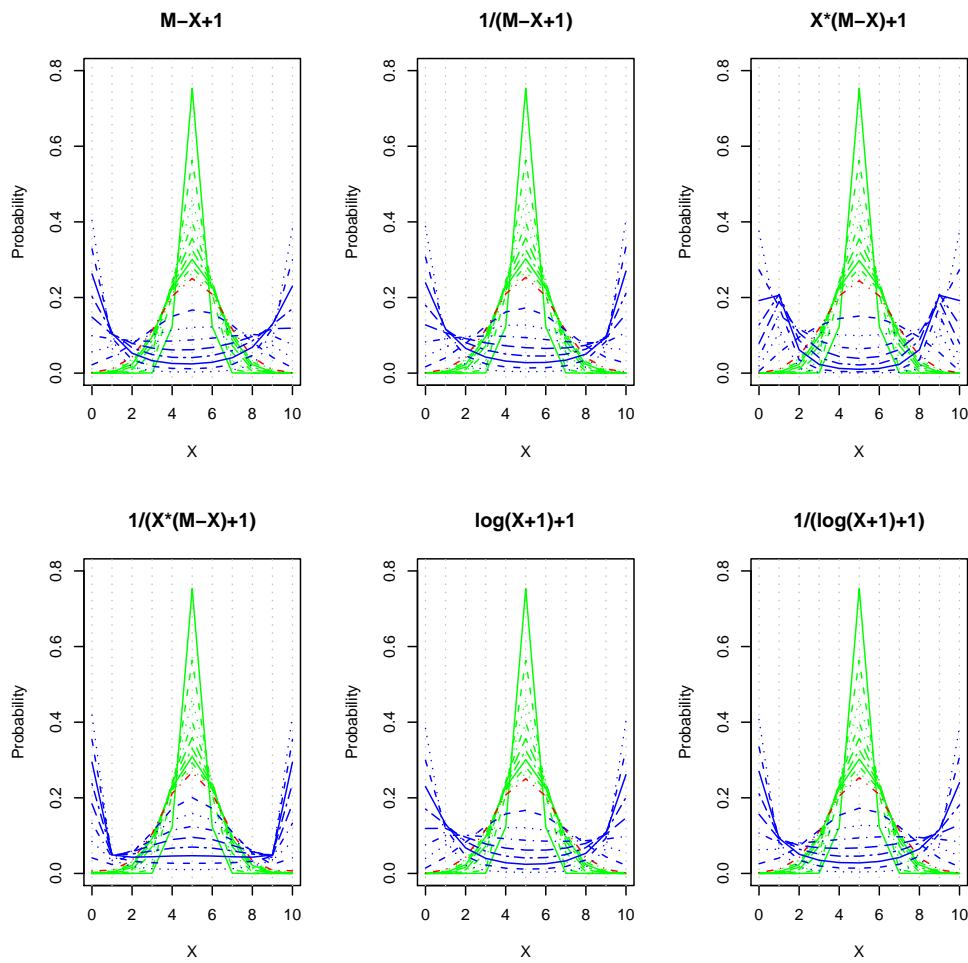
Figure 2. Probability profiles of the extended Altham distributions of mean $\mu = 5$ and various dispersions regarding to h_i given by 2.28 and 2.33. The red line indicates the equal dispersion. The blue lines correspond to over dispersions of $2, 3, \ldots, 9$, while the green lines shows under dispersions of $0.1, 0.2, \ldots, 0.9$. The most centered extended Altham distributions with the largest mass at 5 has dispersion 0.1, and the most spread extended Altham distributions with largest masses at two boundaries has dispersion 9.

The mean and variance are
$$E[X] = \frac{Ma}{a+b}, \qquad Var[X] = \frac{Mab(a+b+M)}{(a+b)^2(a+b+1)}, \tag{2.35}$$
and it is over-dispersed. Prentice [15] extended the beta-binomial, denoted as $\text{EBB}(M; p, \delta)$, using the following reparametrized pmf form
$$p_i = \binom{M}{i} \prod_{j=0}^{i-1} (p + \gamma j) \prod_{j=0}^{M-i-1} (1 - p + \gamma j) \bigg/ \prod_{j=0}^{M-1} (1 + \gamma j), \qquad i = 0, 1, \ldots, M, \tag{2.36}$$
where $0 \leq p \leq 1$, $\gamma = \frac{\delta}{1-\delta}$ and
$$\delta = \gamma(1+\gamma)^{-1} \geq \max\left(\frac{-p}{M-p-1}, \ \frac{-q}{M-q-1}\right), \quad q = 1 - p. \tag{2.37}$$

Figure 3. Comparison of probability profiles between extended Altham distributions with h_i given by 2.22 - 2.27 and binomial distributions under the same means. The blue lines indicate the extended Altham distributions, while red lines correspond to binomial distributions. Any close pair of the extended Altham and binomial distributions has the same mean.

The mean and variance are

$$E[X] = Mp, \qquad Var[X] = Mp(1-p)[1 + (M-1)\delta]. \qquad (2.38)$$

The extended beta-binomial allows under-dispersion, but bounded when δ reaches it lower bound. For example, if $M = 10$ and $p = 0.5$, then the lower bound of δ is $-1/17$, and the lower bound of dispersion is approximately $D = 0.4706$.

Consul [8] proposed the quasi-binomial distribution, later termed as type I QBD$(M; p, \phi)$, with pmf

$$p_i = \binom{M}{i} p(p + i\phi)^{i-1}(1 - p - i\phi)^{M-i}, \qquad i = 0, 1, \dots, M, \qquad (2.39)$$

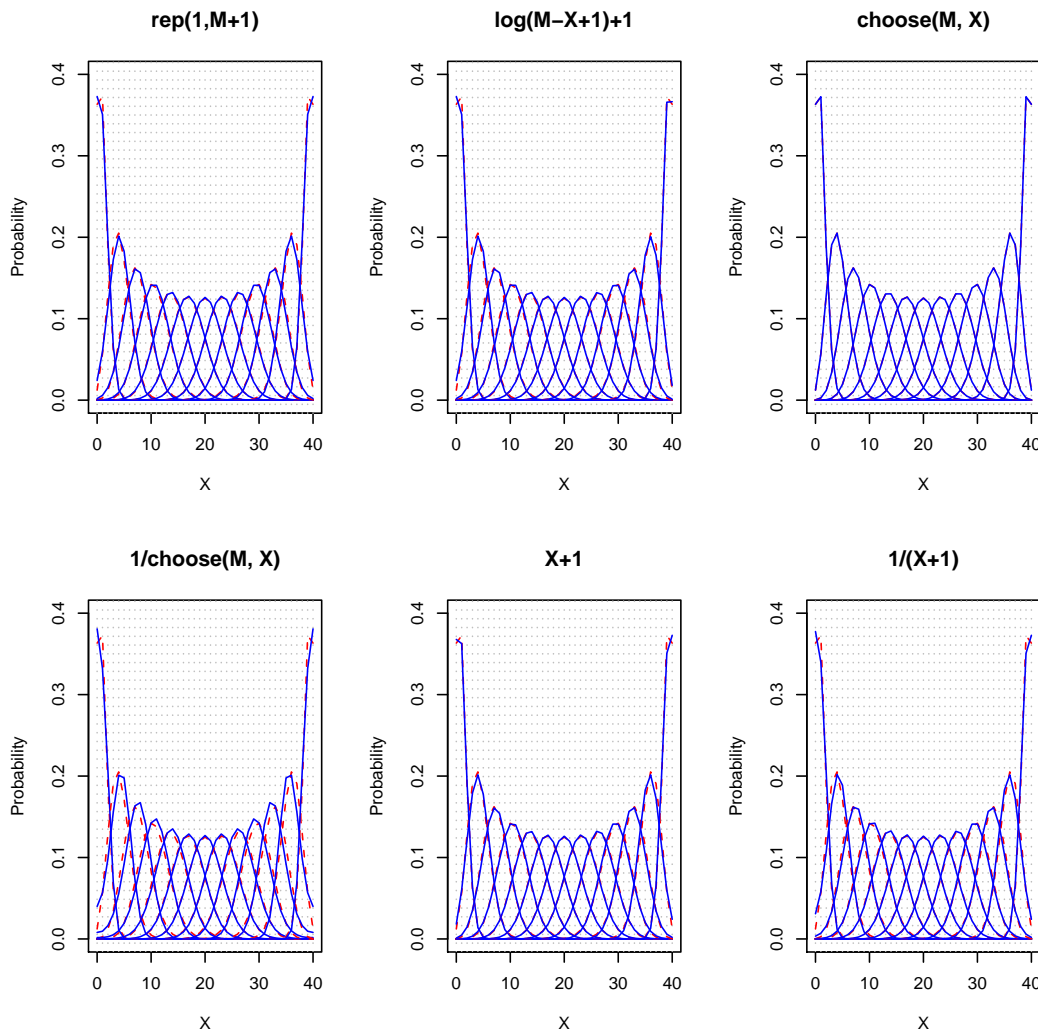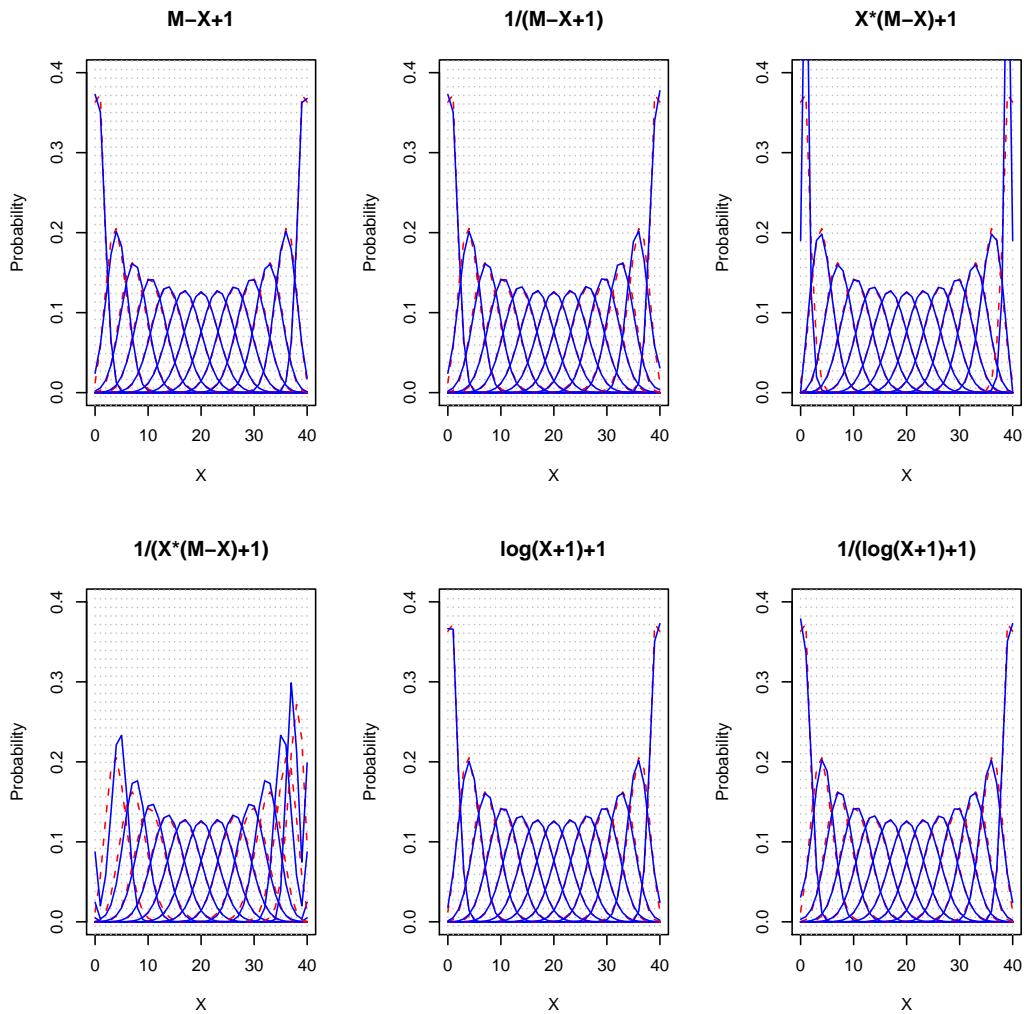where $0 \le p \le 1$ and $-p/M < \phi < (1-p)/M$.

Figure 4. Comparison of probability profiles between extended Altham distributions with h_i given by 2.28 - 2.33 and binomial distributions under the same means. The blue lines indicate the extended Altham distributions, while red lines correspond to binomial distributions. Any close pair of the extended Altham and binomial distributions has the same mean.

As pointed by Mishra, Tiwary and Singh [14], the most unfortunate result of this distribution (and other types QBD) is that the moments are series which are not possible to be summed. When $\phi \neq 0$, the probability of success in the i-th trial becomes $p + i\phi$. Positive or negative ϕ indicates attraction or repulsion of a trial to previous trials. This quasi-binomial distribution has lower bound for the under-dispersion and upper bound for the over-dispersion when ϕ reaches its lower and upper bounds respectively. For example, let $M = 10$ and $p = 0.5$. The lower and upper bounds of ϕ will be -0.05 and 0.05 respectively, and the lower and upper bounds of dispersion D will be approximately 0.4518 and 3.1847 respectively.

The range of dispersion for both extended beta-binomial and quasi-binomial distributions can be numerically displayed. However, both can not cover the full range of dispersion like the extended Altham. Since the extended beta-binomial distribution can be reparametrized in terms of mean and variance analytically, we make numerical comparison of pmf under the same mean and dispersion between this distribution and the extended

Altham distribution, and find that they are different, matching the fact that they are constructed from different angles.

3. Comparison and statistical inference

The pmf (2.5) is explicit in (β_1, β_2) and is implicit in (μ, σ^2). So, for MLE, we can solve it either by parametrization (β_1, β_2) or (μ, σ^2). Since extended Altham distribution is a member of general exponential family, the MLEs for (β_1, β_2) can be obtained by using the form given by

$$p(x|\theta) = h(x)c(\theta)e^{\sum_{i=1}^{k} w_i(\theta)t_i(x)}. \tag{3.1}$$

Then, the Log-likelihood function is,

$$L(\theta) = \sum_{j=1}^{N} log[h(x_j)c(\theta)e^{\sum_{i=1}^{k} w_i(\theta)t_i(x_j)}] \tag{3.2}$$

and the corresponding derivative is

$$\frac{\partial L(\theta)}{\partial \theta} = N\frac{c'(\theta)}{c(\theta)} + \sum_{j}^{N}\sum_{i=1}^{k} w_i(\theta)t_i(x_j). \tag{3.3}$$

Since $p(x|\theta)$ is a probability distribution, we can write

$$\int p(x|\theta) = \int h(x)c(\theta)e^{\sum_{i=1}^{k} w_i(\theta)t_i(x)}dx = 1 \tag{3.4}$$

and we can get

$$c(\theta) = \frac{1}{\int h(x)e^{\sum_{i=1}^{k} w_i(\theta)t_i(x)}dx} \tag{3.5}$$

$$c'(\theta) = -c(\theta)E[\sum_{i}^{k} \frac{\partial w_i(\theta)}{\partial \theta}t_i(x)] \tag{3.6}$$

If $c'(\theta)$ is replaced in the derivative of the log-likelihood function,

$$-NE[\sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial \theta}t_i(x)] + \sum_{j}^{N}\sum_{i=1}^{k} w_i(\theta)t_i(x_j) = 0. \tag{3.7}$$

Finally, maximum likelihood estimator of extended Altham distribution family is found as

$$E[\sum_{i=1}^{k} \frac{\partial w_i(\theta)}{\partial \theta}t_i(x)] = \frac{\sum_{j}^{N}\sum_{i=1}^{k} w_i(\theta)t_i(x_j)}{N}, \tag{3.8}$$

which means the MLE of extended Altham distribution family coincide the moment estimator.

On the other hand, we need the reparametrization of extended Altham distribution with respect to μ and σ^2 in order to be able to make appropriate comparision. First, we derive the MLE of parameter vector by employing the maximum likelihood method. $\beta = (\beta_1, \beta_2)^T$, and its asymptotic normality. Then we obtain the MLE of $\boldsymbol{\theta} = (\mu, \sigma^2)^T$ and its asymptotic normality according to (2.7) and (2.8). Note that the normalizing constant is the function of β_1 and β_2. We establish the following key results for MLEs and their asymptotic covariance matrix. Denote the moment $m_j = E[X^j]$ for $j = 1, 2, 3, 4$.

Lemma 3.1.

$$\frac{\partial C(\beta_1, \beta_2)}{\partial \beta_1} = -m_1, \qquad \frac{\partial C(\beta_1, \beta_2)}{\partial \beta_2} = -m_2,$$

$$\frac{\partial^2 C(\beta_1, \beta_2)}{\partial \beta_1^2} = m_2 - m_1^2, \quad \frac{\partial^2 C(\beta_1, \beta_2)}{\partial \beta_2^2} = m_4 - m_2^2, \quad \frac{\partial^2 C(\beta_1, \beta_2)}{\partial \beta_1 \partial \beta_2} = m_3 - m_2 m_1.$$

Proof. Taking the first and second order partial derivatives with respect to β_1 and β_2 respectively for both sides of $C(\beta_1, \beta_2)$, and then simplifying the equations will yield the results. For instance,

$$e^{C(\beta_1,\beta_2)} \times \frac{\partial C(\beta_1,\beta_2)}{\partial \beta_1} = \frac{\partial}{\partial \beta_1}\left(\sum_{i=0}^{M} h_i e^{i\beta_1 + i^2\beta_2}\right) = -\sum_{i=0}^{M} i h_i e^{i\beta_1 + i^2\beta_2}, \qquad (3.9)$$

$$e^{C(\beta_1,\beta_2)} \times \frac{\partial C(\beta_1,\beta_2)}{\partial \beta_2} = \frac{\partial}{\partial \beta_2}\left(\sum_{i=0}^{M} h_i e^{i\beta_1 + i^2\beta_2}\right) = -\sum_{i=0}^{M} i^2 h_i e^{i\beta_1 + i^2\beta_2}, \quad (3.10)$$

$$e^{C(\beta_1,\beta_2)} \times \frac{\partial^2 C(\beta_1,\beta_2)}{\partial \beta_1 \partial \beta_2} + e^{C(\beta_1,\beta_2)} \times \frac{\partial C(\beta_1,\beta_2)}{\partial \beta_2} \times \frac{\partial C(\beta_1,\beta_2)}{\partial \beta_1} \qquad (3.11)$$

$$= \frac{\partial}{\partial \beta_2}\left(-\sum_{i=0}^{M} i h_i e^{i\beta_1 + i^2\beta_2}\right) = \sum_{i=0}^{M} i^3 h_i e^{i\beta_1 + i^2\beta_2}, \qquad (3.12)$$

thus

$$\frac{\partial C(\beta_1,\beta_2)}{\partial \beta_1} = -\sum_{i=0}^{M} i h_i e^{i\beta_1 + i^2\beta_2} e^{-C(\beta_1,\beta_2)} = -E[X] = -m_1, \qquad (3.13)$$

$$\frac{\partial^2 C(\beta_1,\beta_2)}{\partial \beta_1 \partial \beta_2} = \sum_{i=0}^{M} i^3 h_i e^{i\beta_1 + i^2\beta_2} e^{C(\beta_1,\beta_2)} - \frac{\partial C(\beta_1,\beta_2)}{\partial \beta_2} \times \frac{\partial C(\beta_1,\beta_2)}{\partial \beta_1} \qquad (3.14)$$

$$= E[X^3] - E[X^2]E[X] = m_3 - m_2 m_1. \qquad (3.15)$$

$\square$

Suppose the observations are $x_1, x_2, \ldots, x_n$. The log-likelihood is

$$\log L(\beta \mid x_1, \ldots, x_n) = \sum_{k=1}^{n} \log(\Pr[X_k = x_k]) \qquad (3.16)$$

$$= -nC(\beta_1, \beta_2) - \beta_1 \sum_{k=1}^{n} x_k - \beta_2 \sum_{k=1}^{n} x_k^2.$$

The score functions are

$$\frac{\partial \log L}{\partial \beta_1} = -n\frac{\partial C(\beta_1,\beta_2)}{\partial \beta_1} - \sum_{k=1}^{n} x_k, \qquad \frac{\partial \log L}{\partial \beta_2} = -n\frac{\partial C(\beta_1,\beta_2)}{\partial \beta_2} - \sum_{k=1}^{n} x_k^2, \qquad (3.17)$$

leading to estimating equations

$$\sum_{i=0}^{M} i h_i e^{i\beta_1 + i^2\beta_2} \Big/ \sum_{i=0}^{M} h_i e^{i\beta_1 + i^2\beta_2} = \frac{1}{n}\sum_{k=1}^{n} x_k = \bar{X}, \qquad (3.18)$$

$$\sum_{i=0}^{M} i^2 h_i e^{i\beta_1 + i^2\beta_2} \Big/ \sum_{i=0}^{M} h_i e^{i\beta_1 + i^2\beta_2} = \frac{1}{n}\sum_{k=1}^{n} x_k^2. \qquad (3.19)$$

Applying the quasi-Newton method used before, we can obtain the MLE $\widehat{\beta}$ numerically. Under regularity conditions, for β in the interior of the parameter space, the asymptotic normality holds as follows:

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \to N\left(\mathbf{0}, \boldsymbol{\Sigma}^{-1}\right), \qquad \text{as } n \to \infty, \qquad (3.20)$$

where the Hessian matrix is

$$\boldsymbol{\Sigma} = \begin{pmatrix} -E\left[\frac{\partial^2 \log L}{\partial \beta_1^2}\right] & -E\left[\frac{\partial^2 \log L}{\partial \beta_1 \partial \beta_2}\right] \\ -E\left[\frac{\partial^2 \log L}{\partial \beta_1 \partial \beta_2}\right] & -E\left[\frac{\partial^2 \log L}{\partial \beta_2^2}\right] \end{pmatrix} = n\begin{pmatrix} m_2 - m_1^2 & m_3 - m_2 m_1 \\ m_3 - m_2 m_1 & m_4 - m_2^2 \end{pmatrix}. \qquad (3.21)$$

Although $\widehat{\boldsymbol{\beta}}$ does not have an explicit form, the MLE of $\boldsymbol{\theta} = (\mu, \sigma^2)^T$ has an explicit form. From score functions (3.17), we also obtain estimating equations for μ and σ^2:

$$\mu = \bar{X}, \qquad \sigma^2 + \mu^2 = \frac{1}{n}\sum_{k=1}^{n} x_k^2, \qquad (3.22)$$

leading to the MLEs

$$\hat{\mu} = \bar{X}, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n} x_k^2 - \bar{X}^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{X})^2. \qquad (3.23)$$

Constrains (2.7) and (2.8) imply that μ and σ^2 are functions of β_1 and β_2 respectively. Denote

$$\mathbf{A} = \begin{pmatrix} \frac{\partial \mu}{\partial \beta_1} & \frac{\partial \mu}{\partial \beta_2} \\ \\ \frac{\partial \sigma^2}{\partial \beta_1} & \frac{\partial \sigma^2}{\partial \beta_2} \end{pmatrix}, \qquad (3.24)$$

where

$$\frac{\partial \mu}{\partial \beta_1} = \frac{\partial}{\partial \beta_1}\left(\sum_{i=0}^{M} i h_i e^{C(\beta_1,\beta_2)+i\beta_1+i^2\beta_2}\right) = -\frac{\partial C(\beta_1,\beta_2)}{\partial \beta_1} \times E[X] - E[X^2] = m_1^2 - m_2,$$

$$\frac{\partial \mu}{\partial \beta_2} = \frac{\partial}{\partial \beta_2}\left(\sum_{i=0}^{M} i h_i e^{C(\beta_1,\beta_2)+i\beta_1+i^2\beta_2}\right) = -\frac{\partial log(C(\beta_1,\beta_2))}{\partial \beta_2} \times E[X] - E[X^3]$$

$$= m_1 m_2 - m_3,$$

and

$$\frac{\partial \sigma^2}{\partial \beta_1} = \frac{\partial}{\partial \beta_1}\left(\sum_{i=0}^{M} i^2 h_i e^{C(\beta_1,\beta_2)+i\beta_1+i^2\beta_2}\right) - 2\mu\frac{\partial \mu}{\partial \beta_1}$$

$$= -\frac{\partial C(\beta_1,\beta_2)}{\partial \beta_1} \times E[X^2] - E[X^3] - 2m_1(m_1^2 - m_2) = 3m_1 m_2 - 2m_1^3 - m_3,$$

$$\frac{\partial \sigma^2}{\partial \beta_2} = \frac{\partial}{\partial \beta_2}\left(\sum_{i=0}^{M} i^2 h_i e^{C(\beta_1,\beta_2)+i\beta_1+i^2\beta_2}\right) - 2\mu\frac{\partial \mu}{\partial \beta_2}$$

$$= -\frac{\partial C(\beta_1,\beta_2)}{\partial \beta_2} \times E[X^2] - E[X^4] - 2m_1(m_1^2 - m_2)$$

$$= m_2^2 - m_4 - 2m_1^2 m_2 + 2m_1 m_3.$$

Then,

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \rightarrow N\left(\mathbf{0}, \mathbf{A}\boldsymbol{\Sigma}^{-1}\mathbf{A}^{\mathbf{T}}\right), \qquad \text{as } n \to \infty. \qquad (3.25)$$

Matrix $\mathbf{A}$ and $\boldsymbol{\Sigma}$ can be estimated by replacing m_j's as their estimates $\hat{m}_j$'s. Standard errors of $\hat{\mu}$ and $\hat{\sigma}^2$ can be obtained as the square root of diagonal elements of the estimated covariance matrix. There are two approaches to estimate m_j:

(1) using the sample only, $\hat{m}_j = \frac{1}{n}\sum_{k=1}^{n} x_k^j$, or
(2) using the MLEs $\widehat{\beta}$, $\hat{m}_j = \sum_{i=0}^{M} i^j p_i(\widehat{\beta})$.

The former has large variation when the sample size is not large. Thus, for small sample size, the latter is recommended.

The closed form MLEs of parameters μ and σ^2 simplifies the model fitting using the extended Altham distribution.

Under the extended Altham model, the MLE of dispersion index D is $\hat{D} = \frac{\hat{\sigma}^2}{\hat{\mu}(1-\hat{\mu}/M)}$.
Denote $\mathbf{B} = \left(\frac{\partial D}{\partial \mu}, \frac{\partial D}{\partial \sigma^2} \right)$, where

$$
\begin{aligned}
\frac{\partial D}{\partial \mu} &= \frac{\partial}{\partial \mu} \left[\frac{\sigma^2}{M^2} \left(\frac{1}{M-\mu} + \frac{1}{\mu} \right) \right] = \frac{\sigma^2}{M^2} \left(\frac{1}{(M-\mu)^2} - \frac{1}{\mu^2} \right) \\
&= \frac{\sigma^2(2\mu - M)}{M\mu^2(M-\mu)^2} = \frac{(m_2 - m_1^2)(2m_1 - M)}{Mm_1^2(M-m_1)^2}, \\
\frac{\partial D}{\partial \sigma^2} &= \frac{1}{\mu(1-\mu/M)} = \frac{1}{m_1(1-m_1/M)}.
\end{aligned}
$$

Then,

$$
\sqrt{n}\left(\hat{D} - D \right) \to N\left(0, \mathbf{B}\mathbf{A}\boldsymbol{\Sigma}^{-1}\mathbf{A}^T\mathbf{B}^T \right), \qquad \text{as } n \to \infty. \tag{3.26}
$$

Let s_D^2 be the estimate of asymptotic variance $\mathbf{B}\mathbf{A}\boldsymbol{\Sigma}^{-1}\mathbf{A}^T\mathbf{B}^T$. The standard error of $\hat{D}$
is s_D, and an asymptotic CI of significant level α for dispersion index D is $\hat{D} \pm z_{\alpha/2}s_D$,
where $z_{\alpha/2}$ is the $100(1-\alpha/2)\%$ quantile of the standard normal distribution.
Let P denote the extended Altham family (2.5),

$$
P = \left\{ f_\theta(x) \propto h(x)e^{\beta_1 x + \beta_2 x^2} | \theta = (\beta_1, \beta_2) : -\infty < \beta_1, \beta_2 < \infty, h = 1, 2, \ldots, 12 \right\} \tag{3.27}
$$

where $X = 0, 1, 2, \ldots, M$. Assume $f_\theta = f_{\tilde{\theta}}$, then the expression

$$
log(\frac{h(x)}{\tilde{h}(x)}) + (\beta_1 - \tilde{\beta}_1)x + (\beta_2 - \tilde{\beta}_2)x^2 = 0 \tag{3.28}
$$

is satisfied for all x only when all its coefficients are equal to zero, which is only possible
when $h = \tilde{h}$, $\beta_1 = \tilde{\beta}_1$ and $\beta_2 = \tilde{\beta}_2$. Hence, we conclude that the extended Altham family
is identifiable iff $log(\frac{h(x)}{\tilde{h}(x)}) \neq \beta_1 x + \beta_2 x^2, \beta_1, \beta_2 \neq 0$.

4. Simulation study and examination of existing examples

In the literature, some scholars tried different models. Bailey [4] proposed a particular
probabilistic model based on the Markov property to study the author's writing style
by investigation of occurrences of function word in 5-word and 10-word samples. Two
data sets from Macaulay's 'Essay on Milton' [13] and from Chesterton's essay 'About the
workers' [7] respectively were fitted. Chakraborty and Das [5] fitted QBD I and QBD II
models for four data sets from other authors, these examples were actually truncated count
data, not from true binomial experiments. The observed and expected frequencies, as well
as the values of goodness-of-fit of fitted models were reported in both papers, thus, we can
compare the fitting of the extended Altham models with theirs using the the quantity of
the goodness-of-fit under the same data grouping schemes. Dispersion investigation shows
that all examples are under-dispersed in Bailey [4], and over-dispersed in Chakraborty and
Das [5]. The comparison results are reported in Table 1 and Table 2. Table 1 gives the
fitting comparison of extended Altham models with the model proposed by Bailey [4] for
5-word and 10-word samples of function word occurrence from two authors (Macaulay's
work, Chesterton's work*). Data sets (see Appendix Table A1) and original fittings are
referred to Bailey [4].

Table 1. Fitting comparison of extended Altham models with the model proposed by Bailey [4] for 5-word and 10-word samples

Model	5-word $(0.61, 0.35, 0.66)$	10-word $(1.05, 0.64, 0.68)$	5-word* $(0.61, 0.35, 0.66)$	10-word* $(1.05, 0.64, 0.68)$
Bailey's model	8.16	6.38	2.76	4.93
1	0.0819	0.4887	0.3432	2.0268
2	0.0809	0.4869	0.3404	2.0263
3	0.1027	1.0299	0.3356	**1.6936**
4	0.0659	**0.1549**	0.3440	2.4571
5	0.0974	0.7956	0.3367	1.8095
6	0.0695	0.2599	0.3431	2.2801
7	0.0804	0.4847	0.3405	2.0307
8	0.0843	0.4925	0.3396	2.0232
9	0.2073	6.4423	**0.3199**	1.7605
10	**0.0324**	0.8428	0.3536	5.4180
11	0.1029	0.9111	0.3355	1.7519
12	0.0657	0.2019	0.3441	2.3670

Table 2. Fitting comparison of extended Altham models and the fitted QBD I and QBD II models by Chakraborty and Das [5] for four data sets

Model	Example 1 $(0.41, 0.51, 1.39)$	Example 2 $(0.68, 0.81, 1.37)$	Example 3 $(2.50, 3.37, 2.70)$	Example 4 $(0.92, 0.93, 1.23)$
QBD I	0.075	3.608	0.457	0.941
QBD II	**0.067**	3.618	0.324	0.944
1	0.8834	4.0709	0.3488	2.1207
2	0.4713	4.3125	0.4443	2.4330
3	1.7661	2.5235	0.5243	**0.7481**
4	0.3429	6.0471	0.2100	4.5916
5	2.1230	2.8038	0.4936	0.9950
6	0.1998	5.6710	0.3157	4.1748
7	0.4870	4.3679	0.4160	2.5023
8	1.4354	3.8007	0.3871	1.8306
9	6.9637	**0.5060**	1.3459	3.6989
10	0.9193	15.1880	**0.0016**	20.2011
11	2.6102	2.4775	0.5451	0.7993
12	0.0988	6.2147	0.3025	4.9262

Table 2 gives the fitting comparison of extended Altham models and the Chakraborty and Das [5] fitted QBD I and QBD II models for four data sets (see Appendix Tables A2-A5). Data sets and original fittings are referred to Chakraborty and Das [5]. The χ^2-values of goodness-of-fit are obtained under the same data grouping schemes. ($\bar{x}$, s^2, $\hat{D}$) are given for each example, where $\bar{x}$ is sample mean, s^2 is sample variance and $\hat{D}$ is sample dispersion index. In all examples in Bailey [4], the extended Altham models fits better than the model proposed by Bailey. Refering to samples from Macaulay's work, for the 5-word and the 10-word samples we get the appropriate extended Altham models (2.31) with $\chi^2 = 0.0324$ and (2.25) with $\chi^2 = 0.1549$, respectively.

Regarding to Chesterton's work, we found that the appropriate models for the 5-word* and for the 10*-word samples are extended Altham models (2.30) with $\chi^2 = 0.3199$ and (2.24) with $\chi^2 = 0.16936$, respectively.

In fact, most of the extended Altham models beat the Bailey's model. Moreover, the χ^2 testing at significant level 10% will accept the extended Altham model, but reject the Bailey's model. This might indicate that the original setting of probabilistic mechanism needs further adjustment or refinement.

In Example 2, 3 and 4 in Chakraborty and Das [5], the extended Altham model is better than QBD I and QBD II, while in first example the QBD I and QBD II are better than the extended Altham model. However, the results of acceptance or rejection from the χ^2 test at significant level 10% for all three models are the same. The above examination shows that the extended Altham model can be a safe tool in explorative analysis without special preference in model specification, and also can be an alternative model if other favoured models do not fit data well.

Now we apply the proposed extended Altham model to over-dispersed binomial data resulted from a survey of deaths of children in northest Brazil and the counts the frequencies of 430 childhood deaths in 2946 families of sizes up to eight children. Maternity histories were collected on women aged 15 to 44 over a 3-month period in 1986. The original data was published by Sastry [16] and later it was used for demonstration of different weighted binomial models by Zelterman [19]. We get the sample data regarding to families that has more than three siblings (see Appendix Table A6). From this point of view, the results of extended Altham modelling are given in Table 3.

Table 3. Fitting extended Altham models for the childhood death in Brazilian family data

Model	Number of siblings (n)				
	4	5	6	7	8
	$\bar{x} = 0.49$	$\bar{x} = 0.99$	$\bar{x} = 1.34$	$\bar{x} = 1.80$	$\bar{x} = 2.33$
	$s^2 = 0.52$	$s^2 = 1.16$	$s^2 = 1.78$	$s^2 = 1.48$	$s^2 = 1.72$
	$\hat{D} = 1.13$	$\hat{D} = 1.34$	$\hat{D} = 1.59$	$\hat{D} = 1.06$	$\hat{D} = 1.04$
1	0.3147	1.7954	4.6315	1.4215	0.9545
2	0.3120	1.7737	4.6929	1.4044	0.9534
3	0.5310	2.2675	**3.1773**	2.0239	**0.9450**
4	0.2573	2.0807	7.2788	1.0064	1.0830
5	0.4297	2.0184	3.6284	1.7884	0.9473
6	**0.2631**	1.8905	6.2074	1.1332	1.0152
7	0.3100	**1.7623**	4.7541	1.3936	0.9555
8	0.3195	1.8329	4.5469	1.4509	0.9561
9	3.4693	9.4540	3.8744	5.2578	1.7549
10	1.1462	7.7364	25.3762	**0.6005**	1.8558
11	0.4698	2.0462	3.4754	1.8721	0.9594
12	0.2636	2.0075	6.5180	1.0801	1.0180

According to Table 3, it is obvious that we have huge improvement over the previously examined models. Moreover, extended Altham model has the advantage of having only two parameters.

The last example that we consider is the data that was collected on the sex of the first four children carried out at the A Maxwell Evans Clinic by Elwood and Coldman [10] on 1022 newly diagnosed women with primary breast cancer who had four or fewer children and for whom the sex of each child was known. The data shows mean ages at diagnosis

by number and sex of children. Elwood and Coldman [10] made the analysis in order to observe a possible relationship between the age at diagnosis in women with breast cancer and the sex of their offspring.

Table 4. Fitting extended Altham model for diagnosis of breast cancer by number and sex of children

Model	Number of siblings (n)	
	3	4
	$\bar{x} = 1.51$	$\bar{x} = 1.93$
	$s^2 = 0.77$	$s^2 = 1.15$
	$\hat{D} = 0.82$	$\hat{D} = 1.16$
1	3.2216	4.4021
2	**2.4989**	**3.0160**
3	3.4059	4.1389
4	3.2205	4.7523
5	4.0819	5.5094
6	2.6270	3.4797
7	2.6185	3.2717
8	3.9468	5.7224
9	5.2482	3.8781
10	4.4652	6.4805
11	4.3434	5.7529
12	2.5426	3.3352

Actually, they didn't mention any models for their data. Since their data includes under-dispersed and over-dispersed cases in the same experiment, we decided to use their data (see Appendix Table A7). The number of siblings bigger than two is considered. The summary results of fitting extended Altham model is given in Table 4. In Table 4, we can see that the distribution of the number of diagnosis of breast cancer in the family that has 3 children is under-dispersed ($\hat{D} = 0.82$) and the similar distribution for the family that has 4 children is over-dispersed ($\hat{D} = 1.16$). And extended Altham model (2.23) is best fit for the both cases.

5. Discussion

The extended Altham distribution family is constructed by Kullback-Leibler divergence measure. It turns out to be a particular type of extended Altham distribution, with simple form of pmf from the parametrization of Lagrangian multipliers, which may rendered it to be overlooked previously. Since the construction is very conservative, it is relatively safer than the binomial as well other models developed based on particular probabilistic mechanisms.

The capability to reach the full range of dispersion makes the extended Altham a flexible model for binomial data of various dispersion situations. Thus, it can serve as an explorative model first to avoid wrong specification (say using the binomial model). Because of the conservative feature of the extended Altham, its fitting can be refined or improved by a better model like QBD or EBB, based on revealed dispersion information.

The closed form MLEs simplify the fitting for data, thus, facilitating the application for general end-users, although the calculation of pmf requires the numerical algorithm. The development of a regression framework is in progress.

References

[1] P.M.E. Altham, *Two Generalizations of the binomial distribution*, Appl. Statist. **27** (2), 162-167, 1978.

[2] P.M.E. Altham and R.K.S. Hankin, *Multivariate generalizations of the multiplicative binomial distribution: Introducing the MM package*, J. Stat. Softw **46** (12), 1-23, 2012.

[3] P.M.E. Altham and J.K. Lindsey, *Analysis of the human sex ratio by using overdispersion models*, Appl. Statist. **47** (1), 149-157, 1998.

[4] B.J.R. Bailey, *A model for function word counts*, Appl. Statist. **39** (1), 107-114, 1990.

[5] S. Chakraborty and K.K. Das, *On some properties of a class of weighted quasibinomial distributions*, J. Statist. Plann. Inference **136** (1), 159-182, 2006.

[6] C. Chatfield, *Problem Solving: A Statistician's Guide*, Chapman & Hall, London, 1988.

[7] G.K. Chesterton, *Selected Essays of G. K. Chesterton*, London: Methuen, 1949.

[8] P.C. Consul, *A simple urn model dependent upon predetermined strategy*, Sankhya A Series B **36** (4), 391-399, 1974.

[9] I. Dobson, B.A. Carreras and D.E. Newman, *A probabilistic loading-dependent model of cascading failure and possible implications for blackouts*, Proceedings of the 36th Annual Hawaii International Conference on System Sciences (HICSS'03), 2003.

[10] M. Elwood and A. Coldman, *Age of mothers with breast cancer and sex of their children*, Br Med J **282** (6265), 734, 1981.

[11] J.K. Haseman and L.L. Kupper, *Analysis of dichotomous response data from certain toxicological experiments*, Biometrics **35** (1), 281-293, 1979.

[12] H. Joe, *Multivariate Models and Dependence Concepts*, Chapman & Hall, London, 1997.

[13] L. Macaulay, *Literary Essays Contributed to the Edinburgh Review*, London: Oxford University Press, 1923.

[14] A. Mishra, D. Tiwary and S.K. Singh, *A class of quasi-binomial distributions*, Sankhya A Series B **54** (1), 67-76, 1992.

[15] R.L. Prentice, *Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors*, J. Amer. Statist. Assoc. **81** (394), 321-327, 1986.

[16] N. Sastry, *A nested frailty model for survival data, with an application to the study of child survival in northeast Brazil*, J. Amer. Statist. Assoc. **92** (438), 426-434, 1997.

[17] R. Viveros-Aguilera, K. Balasubramanian and N. Balakrishnan, *Binomial and negative binomial analogues under correlated Bernoulli trials*, Amer. Statist. **48** (3), 243-247, 1994.

[18] R.R. Wilcox, *A review of the beta-binomial model and its extensions*, J. Educ. Stat. **6** (1), 3-32, 1981.

[19] D. Zelterman, *Discrete Distributions: Applications in the Health Sciences*, John Wiley & Sons, West Sussex, 2004.

Appendix

Table A1. Underdispersed word counts [4]

Occurences	0	1	2	3
5-word	45	49	6	0
10-word	27	44	26	3

Table A2. Observed and expected frequencies of European Corn borer in 1296 Corn plants [5]

No. of borers per plant	0	1	2	3	≥ 4
Observed no. of plants	907	275	88	23	3

Table A3. Distribution of yeast cells per square in a haemacytometer [5]

No. of cells per square	0	1	2	3	4	5
Observed no. of squares	213	128	37	18	3	1

Table A4. Distribution of number of seeds by time of day [5]

Time	0	1	2	3	4	5
Observed no. seeds	7	4	5	5	4	7

Table A5. Distribution of number of hits per square [5]

No. of hits	0	1	2	3	4	5
No. of 1/4 km squares	229	211	93	35	7	1

Table A6. The frequency of childhood deaths in Brazilian families [19]

Number of siblings n	Number of families f_n	Number of deaths m_n	Number of affected siblings 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	2946	430								

Table A7. Diagnosis of breast cancer by number and sex of children [10]

No of Children	0	1		2			3				4				
No of boys	0	0	1	0	1	2	0	1	2	3	0	1	2	3	4
No of patients	284	93	71	65	134	83	26	71	75	26	11	21	30	28	4