

ESTIMATION OF PARAMETERS IN THE BETA BINOMIAL MODEL

RAM C. TRIPATHI¹, RAMESH C. GUPTA² AND JOHN GURLAND³

¹*Division of Mathematics, Computer Science and Statistics,
The University of Texas, San Antonio, TX 78249, U.S.A.*

²*Department of Mathematics and Statistics, University of Maine,
Orono, ME 04469-5752, U.S.A.*

³*Department of Statistics, University of Wisconsin, Madison, WI 53706, U.S.A.*

(Received July 8, 1991; revised September 28, 1993)

Abstract. This paper contains some alternative methods for estimating the parameters in the beta binomial and truncated beta binomial models. These methods are compared with maximum likelihood on the basis of Asymptotic Relative Efficiency (ARE). For the beta binomial distribution a simple estimator based on moments or ratios of factorial moments has high ARE for most of the parameter space and it is an attractive and viable alternative to computing the maximum likelihood estimator. It is also simpler to compute than an estimator based on the mean and zeros, proposed by Chatfield and Goodhart (1970, *Appl. Statist.*, **19**, 240–250), and has much higher ARE for most part of the parameter space. For the truncated beta binomial, the simple estimator based on two moment relations does not behave quite as well as for the BB distribution, but a simple estimator based on two linear relations involving the first three moments and the frequency of “ones” has extremely high ARE. Some examples are provided to illustrate the procedure for the two models.

Key words and phrases: Maximum likelihood, minimum chi-square, asymptotic relative efficiency, truncated beta binomial.

1. Introduction

In studying the effect of a chemical on laboratory animals, Williams (1975) examined the data on pregnant females where the responses are recorded on each individual fetus in a litter. In each litter, the number, X_i of pups that survive the lactation period were recorded. Thus each X_i is a binomial random variable with parameter p . Recently Skurnick (1990), in planning a trial on infertility patients, considered a population of women who were not ovulating regularly. For each woman, the number of ovulatory cycles, X_i , were recorded out of six consecutive cycles, the clinically standard duration of therapy. So each X_i is a binomial random variable with parameter p . In the case of the laboratory animals, the survival

rate of the pups in a litter varies from litter to litter. Similarly in the infertility study the ovulation rates in response to therapy could be expected to vary from woman to woman. Thus the distribution of the number of successful cycles within treatment group would be overdispersed in comparison with a binomial model, where one ovulatory rate per group prevailed. In such cases, beta binomial distribution provides a flexible model for between subject variability. In this case we assume that the woman's probability of ovulating in a single cycle be distributed as the beta distribution and then her number of successful cycles is binomial conditional on that. Removing the conditioning gives rise to the beta binomial (BB) distribution.

In studying the impact of litter effect on dose-response modeling in teratology, Kupper *et al.* (1986) concluded that the failure to take litter effects into account can lead to underestimation of variances associated with the parameter estimates, so that the use of a binomial likelihood to model teratology data does not seem advisable. They advocate the use of BB distribution to introduce the varying degree of intralitter correlation. Paul (1982), in analyzing the proportions of affected fetuses in teratological experiments, observed that the BB model is superior to the multiplicative binomial model. Pack has found the BB to be superior to alternative models such as the correlated binomial of Kupper and Haseman (1978), see Pack (1986). Tarone (1982) observed that for many tumor types, the historical control rates are more variable than would be expected if they followed a binomial distribution, see also Tarone *et al.* (1981), and he fitted a BB distribution to the historical tumor rates.

Various techniques for estimating parameters in the BB or TBB (truncated beta binomial) distributions have appeared in the literature. Skellam (1948) used maximum likelihood estimates (MLE) obtainable with the aid of tables of the digamma function. Griffiths (1973) developed a method of obtaining MLE's which obviates the need of values of the digamma function but involves an iterative process for solving nonlinear equations. Nissen-Meyer (1964) proposed an iterative graphical procedure for obtaining MLE's. Williams (1982) has described a simple method of modifying standard logistic-linear analyses, in particular using the GLIM computer program to accommodate extra binomial variation. Williams (1988) has also studied the bias related to the MLE's. Brooks (1984) has suggested a modification of the GLIM method so as to provide a simple way of obtaining approximate likelihood ratio test statistic assuming a BB model. In addition to the MLE's, Chatfield and Goodhart (1970) applied an iterative technique based on equating the sample mean and the proportion of zeros to their population counterparts. Shenton (1950) showed that the asymptotic relative efficiency (ARE) of the method of moments usually exceeds 70%.

In the present paper we provide other alternative methods for the estimation of parameters in the BB model. A numerical technique, according to which partial derivatives with respect to any of the arguments of the hypergeometric function can be evaluated with high precision, has been used, see Tripathi (1975). Section 2 contains the estimation of BB parameters based on (i) mean and zeros (ii) first two sample moments (iii) the mean and the ratio of ones to zeros, and the comparison of ARE's of these estimators. Similar study is done, in Section 3, for the truncated

beta binomial (TBB) model. It turns out that the estimators based on moments or ratios of factorial moments behave surprisingly well. The method of minimum chi-squares is briefly described in Section 4 to obtain the estimators for the parameters of the BB and TBB models. Tables 1 and 2 are provided to summarize the investigation of the ARE's for the two models. Numerical examples are included in Section 5 to illustrate the procedure for the BB and TBB models, see also Tables 3 and 4.

2. The beta binomial family of distributions

For a binomial distribution with parameter p , the probability of success, and N , the number of independent trials, the probability generating function (p.g.f.) is $[1 + p(z - 1)]^N$. If p is regarded as a beta random variable with probability density function (p.d.f.)

$$(2.1) \quad \frac{1}{B(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1} \quad \alpha > 0, \quad \beta > 0, \quad 0 < p < 1,$$

$$\text{where } B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)},$$

the p.g.f. $G(z)$ of the resulting compound distribution, designated here as the BB family, becomes

$$(2.2) \quad G(z) = \frac{1}{B(\alpha, \beta)} \int_0^1 \{1 + p(z - 1)\}^N p^{\alpha-1} (1-p)^{\beta-1} dp \\ = {}_2F_1(-N, \alpha; \alpha + \beta; 1 - z).$$

The hypergeometric function ${}_2F_1$ has a power series representation given

$${}_2F_1(u, v; w; x) = \sum_{j=0}^{\infty} \frac{(u)_j (v)_j}{(w)_j} \frac{x^j}{j!}$$

where $(u)_j = u(u + 1)(u + 2) \cdots (u + j - 1)$, and the series converges for $|x| < 1$. From the p.g.f. (2.2), or otherwise, it can be seen that P_j , the probability of the random variable assuming the value j , is

$$(2.3) \quad P_j = \binom{N}{j} \frac{B(\alpha + j, N + \beta - j)}{B(\alpha, \beta)}, \quad j = 0, 1, \dots, N$$

and $\mu_{(j)}$, the j -th descending factorial moment, is

$$(2.4) \quad \text{and } \mu_{(j)} = \frac{(-N)_j (\alpha)_j}{(\alpha + \beta)_j} (-1)^j, \quad j = 1, 2, \dots$$

In particular the mean and variance of the beta binomial model are, respectively, given by $n\pi$ and $n\pi(1-\pi)(n\phi+1)(1+\phi)^{-1}$ where $\pi = \alpha(\alpha+\beta)^{-1}$ and $\phi = (\alpha+\beta)^{-1}$.

The binomial model is a special case of the BB model with $\phi = 0$. Obviously ϕ is the over dispersion parameter. It follows that

$$(2.5) \quad \frac{P_{j+1}}{P_j} = \frac{(N - j)(\alpha + j)}{(j + 1)(N + \beta - j - 1)}, \quad j = 0, 1, \dots, N - 1$$

and

$$(2.6) \quad \frac{\mu_{(j+1)}}{\mu_{(j)}} = \frac{(\alpha + j)(N - j)}{\alpha + \beta + j}, \quad j = 0, 1, \dots$$

(We adopt the convention here that $\mu_{(0)} = 1$.) For convenience in developing estimators of α and β and hence of π and ϕ through the techniques described below, we introduce the following notation:

$$\xi_j = \frac{\mu_{(j+1)}}{\mu_{(j)}}, \quad j = 0, 1, 2, \dots \quad \text{and}$$

$$\eta_j = \frac{P_{j+1}}{P_j}, \quad j = 0, 1, \dots, N - 1.$$

The relations (2.5) and (2.6) can now be expressed equivalently in terms of ξ_j and η_j as follows:

$$(2.7) \quad \alpha(j - N) + \beta(j + 1)\eta_j = j(N - j) - (j + 1)(N - j - 1)\eta_j, \quad j = 0, 1, \dots, N - 1,$$

$$(2.8) \quad \alpha(\xi_j - N + j) + \beta\xi_j = j(N - j - \xi_j), \quad j = 0, 1, 2, \dots$$

These relations will be used below.

2.1 Estimation of the parameters in the BB distribution

(1) Estimators based on mean and zero (mean-zeros)

Since the Negative Binomial (NB) is a limiting case of the BB distribution, and as Anscombe (1950) showed that the method of mean and zeros is more efficient than the method of moments when fitting the NB to reverse J-shaped distributions, Chatfield and Goodhart (1970) conjectured that the method of mean and zeros would have high efficiency when fitting the BB to reverse J-shaped distributions. Let \hat{P}_0 denote the sample proportion of observed zeros and $\hat{\mu}$ the sample mean. Then the estimators of α and β based on \hat{P}_0 and $\hat{\mu}$ are obtained by solving the equations

$$(2.9) \quad \frac{B(\alpha, N + \beta)}{B(\alpha, \beta)} = \hat{P}_0,$$

$$\frac{N\alpha}{\alpha + \beta} = \hat{\mu}.$$

Chatfield and Goodhart (op.cit.) present an iterative solution.

(2) *Estimators based on first two sample moments (2-moments)*

Let $\hat{\mu}_{(j)}$ denote the j -th sample factorial moment and $\hat{\xi}_j = \frac{\hat{\mu}_{(j+1)}}{\hat{\mu}_{(j)}}$. Setting $j = 0, 1$ in (2.8) and solving the two equations yields the estimators

$$(2.10) \quad \hat{\alpha} = \frac{\hat{\xi}_0(N - 1 - \hat{\xi}_1)}{\hat{\xi}_0 + N(\hat{\xi}_1 - \hat{\xi}_0)}, \quad \hat{\beta} = \frac{(N - \hat{\xi}_0)(N - 1 - \hat{\xi}_1)}{\hat{\xi}_0 + N(\hat{\xi}_1 - \hat{\xi}_0)}$$

based on the first two sample moments.

(3) *Estimator based on the mean and the ratio of ones to zeros (1 moment-1 probability)*

Since the estimators in (2.9) involve a nonlinear equation for the zeros, it is tempting to replace it by a linear equation involving the ratio of ones to zeros obtained from equation (2.7) with $j = 0$.

Let $\hat{\eta}_0 = \hat{P}_1/\hat{P}_0$, the ratio of the proportion of observed ones to the proportion of observed zeros. Then a simple estimator can be obtained by solving the two linear equations

$$(2.11) \quad \begin{aligned} \alpha N - \beta \hat{\eta}_0 &= (N - 1)\hat{\eta}_0, \\ \alpha(N - \hat{\mu}) - \beta \hat{\mu} &= 0 \end{aligned}$$

where $\hat{\mu}$ is the sample mean. This yields the estimators

$$\begin{aligned} \hat{\alpha} &= \frac{(N - 1)\hat{\eta}_0\hat{\mu}}{N\hat{\mu} - (N - \hat{\mu})\hat{\eta}_0}, \\ \hat{\beta} &= \frac{(N - 1)(N - \hat{\mu})\hat{\eta}_0}{N\hat{\mu} - (N - \hat{\mu})\hat{\eta}_0}. \end{aligned}$$

In Subsection 2.2 we shall compare the asymptotic relative efficiencies of the estimators of π and ϕ by the above three methods.

2.2 *Comparison of asymptotic relative efficiencies*

The asymptotic relative efficiency is defined by

$$(2.12) \quad \text{ARE} = \frac{|I^{-1}|}{|V|}$$

where $I = E(\frac{\partial}{\partial \pi} \ln L)(\frac{\partial}{\partial \phi} \ln L)$ and V is the asymptotic covariance matrix of the estimators of π and ϕ . Since $\alpha = \pi/\phi$ and $\beta = (1 - \pi)/\phi$, V can be written as $V = J\Sigma J'$ where J is the jacobian of the transformation and Σ is the asymptotic covariance matrix of the estimators of α and β . Similarly $I^{-1} = JI^{*-1}J'$ where I^{*-1} is the asymptotic covariance matrix of the MLE's of α and β . Hence the ARE of the estimators of π and ϕ is the same as the ARE of the corresponding estimators of α and β . Thus we shall compute the ARE of the estimators of α and β .

Computation of the asymptotic generalized variance is straightforward, but computation of the information determinant $|E(\frac{\partial}{\partial \alpha} \ln L)(\frac{\partial}{\partial \beta} \ln L)|$, where L is the

Table 1(a). ARE of some estimators for the parameters in the beta family.

$N = 5$		ϕ				
π		.1	.3	.5	2.5	5.5
.1	(i)	.95	.95	.93	.84	.82
	(ii)	.95	.92	.92	.93	.94
	(iii)	.78	.70	.63	.40	.35
	(iv)	1.00	.99	.98	.97	.96
.3	(i)	.75	.80	.81	.80	.80
	(ii)	.99	.97	.96	.94	.94
	(iii)	.45	.45	.43	.34	.32
	(iv)	1.00	.99	.99	.97	.96
.5	(i)	.49	.60	.66	.75	.77
	(ii)	1.00	.99	.98	.94	.94
	(iii)	.21	.25	.27	.29	.29
	(iv)	1.00	.99	.99	.97	.96
.7	(i)	.25	.39	.48	.70	.75
	(ii)	.99	.97	.96	.94	.94
	(iii)	.07	.11	.15	.24	.27
	(iv)	1.00	.99	.99	.97	.96
.9	(i)	.06	.20	.31	.64	.72
	(ii)	.95	.92	.92	.93	.94
	(iii)	.01	.04	.07	.20	.25
	(iv)	1.00	.99	.98	.97	.96

(i) Mean-zeros, (ii) 2-moment, (iii) 1 moment -1 probability, (iv) minimum chi-square based on 3 moment relations.

Table 1(b). ARE of some estimators for the parameters in the beta binomial family.

$N = 10$		ϕ				
π		.1	.3	.5	2.5	5.5
.1	(i)	.89	.91	.89	.76	.74
	(ii)	.89	.83	.82	.82	.83
	(iii)	.56	.50	.44	.26	.23
	(iv)	.98	.94	.92	.88	.87
.3	(i)	.50	.66	.70	.70	.70
	(ii)	.98	.93	.90	.83	.83
	(iii)	.16	.22	.24	.21	.20
	(iv)	1.00	.97	.97	.89	.87
.5	(i)	.18	.38	.48	.63	.67
	(ii)	1.00	.97	.93	.83	.83
	(iii)	.03	.08	.11	.17	.18
	(iv)	1.00	.98	.96	.89	.87
.7	(i)	.04	.17	.28	.56	.64
	(ii)	.98	.93	.90	.83	.83
	(iii)	.003	.02	.04	.13	.16
	(iv)	1.00	.97	.95	.89	.87
.9	(i)	.002	.05	.13	.49	.60
	(ii)	.89	.83	.82	.82	.83
	(iii)	$.9 \times 10^{-4}$.003	.01	.10	.14
	(iv)	.98	.94	.92	.88	.87

(i) Mean and zeros, (ii) 2 moment, (iii) 1 moment and -1 probability, (iv) minimum chi-square based on 3 moment relations.

likelihood function, involves derivatives of the hypergeometric function with respect to the parameters appearing in it. Differentiation based on extrapolation to the limit was used for the computations in all the tables of ARE values in this article. The method mentioned above is capable of obtaining from two lower order approximations a higher order approximation and is, thus, computationally more accurate and more efficient. This method has been described in detail by Conte and deBoor ((1980), p. 333).

In Table 1(a) for $N = 5$, and Table 1(b) for $N = 10$, the ARE values for the estimators of π and ϕ are given for the following four estimators:

- i) Based on mean and zeros.
- ii) Based on the ratios of factorial moments.
- iii) Based on the mean and the ratios of ones to zeros.
- iv) Minimum chi-square estimators based on three moments.

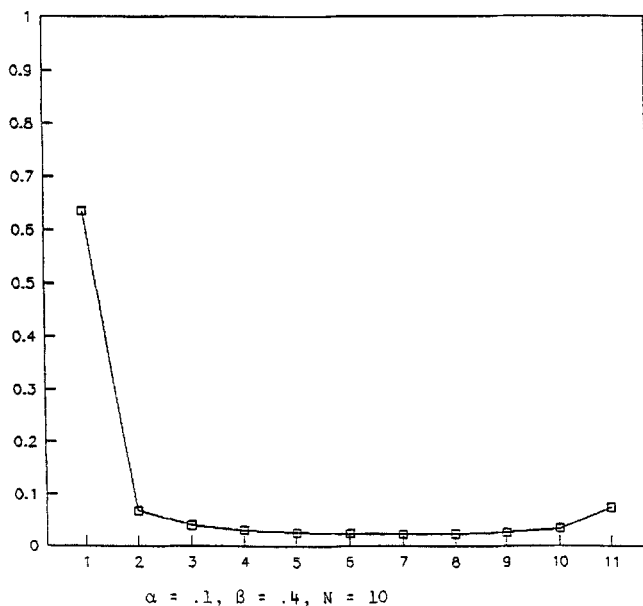
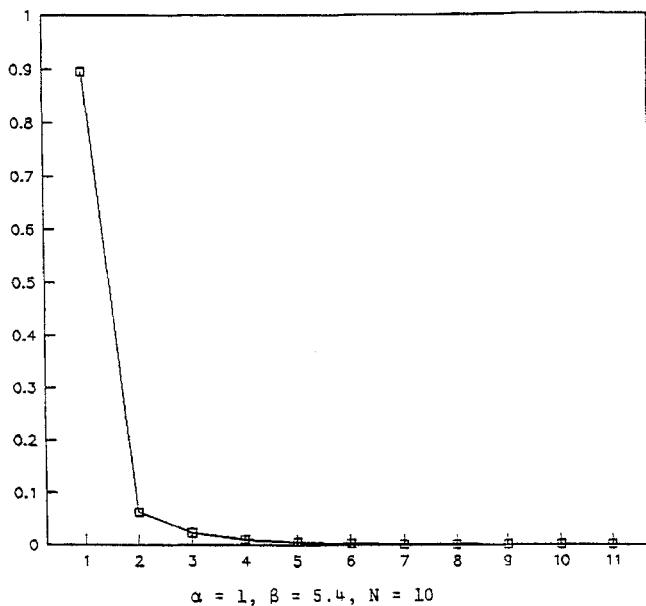


Fig. 1. Reverse J-shaped plots of beta-binomial.

The first three correspond respectively to (1), (2), (3) in Subsection 2.1. The fourth will be described in Section 4 below.

The beta binomial distribution is reverse J-shaped for small values of α , or π , see Fig. 1.

For the parameter values in the first row of Tables 1(a) and 1(b), the estimator

(i) based on the mean and zeros does not have high efficiencies especially for low values of π and high values of ϕ . The ARE of estimator (ii), based on the first two moments, is also very high in the first row of the tables. In fact, the ARE values of (ii) over all of Table 1(a) are at least 92% with some values indistinguishable from 100%. As for Table 1(b), the ARE values of estimator (ii) are well above 90% for $\phi < 2.5$ and moderate values of π that are not too small or too large. In both Tables 1(a) and 1(b), the ARE values of estimator (i) drop rapidly from its values in the first row, with some values close to or indistinguishable from zero in Table 1(b). The third estimator based on the mean and ratio of ones to zeros has low ARE values throughout the tables and, therefore, is not to be recommended.

For a fourth estimator, obtained through minimum chi-square utilizing the first three moments, designated here as (iv), the ARE values are also given in Tables 1(a) and 1(b). The technique for obtaining this estimator will be described in Section 4 below. At this point it is worthwhile noting that all its ARE values are well above 87% and surpass or are equal to the values for all the other estimators considered. In the light of the excellent performance of estimator (ii), however, which is very easily computed from the ratios of factorial sample moments, it may hardly be worth the extra effort to compute estimator (iv), except perhaps for larger values of N , e.g. $N = 10$, and for larger values of ϕ .

3. The truncated beta binomial (TBB) family of distributions

We designated here the TBB family as the BB distribution with zeros excluded, and its p.g.f. as $G^*(z)$. Thus

$$G^*(z) = \frac{{}_2F_1(-N, \alpha; \alpha + \beta; 1 - z) - P_0}{1 - P_0}$$

where ${}_2F_1(-N, \alpha; \alpha + \beta; 1 - z)$ and P_0 are respectively the p.g.f. and the probability of the zero count for the BB family. Let P_j^* and $\mu_{(j)}^*$ denote the probability of the j -th count and the j -th factorial moment respectively, of the TBB family. Further, let

$$\begin{aligned} \xi_j^* &= \frac{\mu_{(j+1)}^*}{\mu_{(j)}^*}, & j &= 1, 2, \dots, \\ \eta_j^* &= \frac{P_{j+1}^*}{P_j^*}, & j &= 1, 2, \dots, N - 1, \\ \xi_0^* &= \mu_{(1)}^* = \frac{N\alpha}{(\alpha + \beta)(1 - P_0)} & \text{where } P_0 &= {}_2F_1(-N, \alpha; \alpha + \beta; 1). \end{aligned}$$

Then the η_j^* and ξ_j^* satisfy the same equations as do the η_j 's and ξ_j 's in (2.7) and (2.8), but with the range of the j subscript starting at 1.

A further relation can be obtained by summing the modified version of (2.5) corresponding to the TBB distribution with the P_j replaced by P_j^* and the summation ranging from 1 to $N - 1$. This yields

$$(3.1) \quad \alpha(N - \xi_0^*) + \beta(P_1^* - \xi_0^*) = P_1^*(1 - N).$$

This relation, along with (2.7) and (2.8) suitably modified, will be used in forming estimators of the parameters as described below.

3.1 *Estimation of the parameters in the TBB distribution*

(1) *Estimator based on first two moment relations*

As in Section 3, let $\hat{\mu}_{(j)}^*$ denote the j -th sample factorial moment and

$$\hat{\xi}_j^* = \frac{\hat{\mu}_{(j+1)}^*}{\hat{\mu}_{(j)}^*}.$$

Setting $j = 1, 2$ in (2.8) and solving the two resulting equations, yields the estimator

$$\hat{\alpha} = \frac{\hat{\xi}_2^*(N - 1 - \hat{\xi}_1^*) - 2\hat{\xi}_1^*(N - 2 - \hat{\xi}_2^*)}{\hat{\xi}_2^*(\hat{\xi}_1^* - N + 1) - \hat{\xi}_1^*(\hat{\xi}_2^* - N + 2)},$$

$$\hat{\beta} = \frac{2(N - 2 - \hat{\xi}_2^*)(\hat{\xi}_1^* - N + 1) - (N - 1 - \hat{\xi}_1^*)(\hat{\xi}_2^* - N + 2)}{\hat{\xi}_2^*(\hat{\xi}_1^* - N + 1) - \hat{\xi}_1^*(\hat{\xi}_2^* - N + 2)}$$

based on the first two moment relations.

(2) *Estimator based on first moment relation and one involving $\mu_{(1)}^*$ and P_1^* (relation (3.1))*

Setting $j = 1$ in (2.8) for the first equation, and including (3.1) as the second equation yields the following estimator

$$\hat{\alpha} = \frac{(N - 1 - \hat{\xi}_1^*)(\hat{P}_1^* - \hat{\mu}_{(1)}^*) - \hat{P}_1^*\hat{\xi}_1^*(1 - N)}{(\hat{P}_1^* - \hat{\mu}_{(1)}^*)(\hat{\xi}_1^* - N + 1) - \hat{\xi}_1^*(N - \hat{\mu}_{(1)}^*)},$$

$$\hat{\beta} = \frac{(\hat{\xi}_1^* + 1 - N)(N - \hat{\mu}_{(1)}^*) + \hat{P}_1^*(1 - N)(\hat{\xi}_1^* - N + 1)}{(\hat{P}_1^* - \hat{\mu}_{(1)}^*)(\hat{\xi}_1^* - N + 1) - \hat{\xi}_1^*(N - \hat{\mu}_{(1)}^*)}$$

where $\hat{\mu}_{(1)}^*$, \hat{P}_1^* and $\hat{\xi}_1^*$ are the sample counterparts of $\mu_{(1)}^*$, P_1^* and ξ_1^* respectively.

3.2 *Comparison of the asymptotic relative efficiency of estimators*

In Table 2(a), for $N = 5$, and Table 2(b) for $N = 10$, appear the ARE values for four estimators as follows:

- i) Based on first two moment relations.
- ii) Based on first moment relation and one involving $\mu_{(1)}^*$ and P_1^* .
- iii) Minimum chi-square estimator based on first three moment relations.
- iv) Minimum chi-square estimator based on first two moment relations and one involving $\mu_{(1)}^*$ and P_1^* .

The first two estimators correspond respectively to (1), (2) in Subsection 3.1. The minimum chi-square technique required for obtaining estimators (iii) and (iv) will be described in Section 4.

The entries of Tables 2(a) and 2(b) suggest that the ARE values of estimator (ii) are higher when $N = 5$ than those of estimator (i) over an upper triangular

Table 2(a). ARE of some estimators for the parameters in the TBB family.

$N = 5$		ϕ				
		π	.1	.3	.5	2.5
.1	(i)	.81	.79	.81	.76	.88
	(ii)	1.00	.99	.97	.85	.81
	(iii)	.99	.98	.97	.97	.97
	(iv)	1.00	1.00	1.00	1.00	.99
.3	(i)	.80	.81	.83	.88	.88
	(ii)	.98	.97	.94	.83	.80
	(iii)	.98	.98	.98	.97	.97
	(iv)	1.00	1.00	1.00	.99	.99
.5	(i)	.80	.83	.85	.88	.89
	(ii)	.97	.93	.90	.81	.79
	(iii)	.97	.97	.98	.97	.97
	(iv)	1.00	1.00	1.00	.99	.99
.7	(i)	.84	.87	.88	.89	.89
	(ii)	.95	.90	.86	.79	.77
	(iii)	.97	.97	.97	.97	.97
	(iv)	1.00	1.00	1.00	.99	.99
.9	(i)	.93	.92	.91	.89	.89
	(ii)	.93	.84	.81	.77	.76
	(iii)	.98	.98	.98	.97	.97
	(iv)	1.00	1.00	.99	.99	.99

(i) 2 moment relations, (ii) 1 moment relation and another relation involving $\mu_{(1)}^*$ and P_1^* , (iii) minimum chi-square based on 3 moment relations, (iv) minimum chi-square based on 2 moment relations and the relation involving $\mu_{(1)}^*$ and P_1^* .

Table 2(b). ARE of some estimators for the parameters in the TBB family.

$N = 10$		ϕ				
		π	.1	.3	.5	2.5
.1	(i)	.54	.52	.55	.63	.63
	(ii)	.99	.97	.95	.75	.68
	(iii)	.85	.79	.79	.79	.79
	(iv)	1.00	.99	.99	.95	.93
.3	(i)	.57	.58	.60	.63	.63
	(ii)	.97	.94	.90	.72	.66
	(iii)	.83	.81	.80	.80	.79
	(iv)	.99	.99	.98	.94	.92
.5	(i)	.67	.65	.65	.64	.64
	(ii)	.97	.91	.86	.69	.65
	(iii)	.86	.83	.82	.80	.79
	(iv)	.99	.98	.97	.93	.92
.7	(i)	.82	.75	.71	.65	.64
	(ii)	.96	.87	.81	.66	.63
	(iii)	.92	.86	.84	.80	.79
	(iv)	1.00	.98	.96	.92	.91
.9	(i)	.96	.84	.77	.65	.64
	(ii)	.89	.79	.73	.63	.61
	(iii)	.97	.90	.86	.80	.80
	(iv)	.98	.95	.93	.91	.90

(i) 2 moment relations, (ii) 1 moment relation and another relation involving $\mu_{(1)}^*$ and P_1^* , (iii) minimum chi-square based on 3 moment relations, (iv) minimum chi-square based on 2 moment relations and the relation involving $\mu_{(1)}^*$ and P_1^* .

region in the π - ϕ plane. The ARE values of the estimator (i) are higher than those of estimator (ii) in the remaining part of the parameter space. However, when $N = 10$, the ARE values of estimator (ii) are higher than those of estimator (i) except when π is close to 1. However, the estimators (iii) and (iv) based on minimum chi-square have quite high ARE values for both the tables except (iii) having some low ARE values when π is small and $N = 10$.

4. Minimum chi-square estimators

Consider a distribution with the probability function $p(x, \theta)$ where θ is a $k \times 1$ vector. To obtain minimum chi-square estimator of θ , we consider s ($s > k$) functions $\tau_1, \tau_2, \dots, \tau_s$ of the first few moments and/or probabilities of the distribution such that $\tau = W\theta$ where $\tau' = (\tau_1, \tau_2, \dots, \tau_s)$ and W is an $s \times k$ matrix of known constants. The vector θ may be some reparameterized version of the parameters of the distribution. Let \underline{t} be a sample counterpart of τ and $\hat{\Sigma}$ be a consistent estimate of the asymptotic covariance matrix Σ of \underline{t} . Then, the minimum chi-square estimator of θ is obtained by minimizing (Barankin and Gurland (1951))

$$Q = (\underline{t} - W\theta)' \hat{\Sigma}^{-1} (\underline{t} - W\theta)$$

and is given by

$$\hat{\theta} = (W' \hat{\Sigma}^{-1} W)^{-1} W' \hat{\Sigma}^{-1} \underline{t}.$$

The minimum value \hat{Q} of Q obtained by replacing θ by $\hat{\theta}$ gives a test statistic for testing fit of the underlying model. The asymptotic null distribution of \hat{Q} is χ^2 with $s - k$ degrees of freedom (d.f.).

For illustration, we present the procedure for obtaining the minimum chi-square estimators for the parameters of the BBD based on $s = 3$ relations (estimator (iv), Table 1(a)). This seems to be a reasonable choice because larger values of s will involve higher order moments and thus will introduce larger sampling fluctuations in estimating Σ . Consider two moment relations obtained from (2.8) by putting $j = 0, 1$ and one probability relation obtained from (2.7) by putting $j = 0$. This gives

$$\begin{aligned} \alpha(\xi_0 - N) + \beta\xi_0 &= 0, \\ \alpha(\xi_1 - N + 1) + \beta\xi_1 &= (N - 1 - \xi_1), \\ \alpha(\xi_2 - N + 2) + \beta\xi_2 &= 2(N - 2 - \xi_2). \end{aligned}$$

Solving these for α and β gives:

$$(4.1) \quad \begin{aligned} \tau_1 = \alpha &= \frac{\xi_0(N - 1 - \xi_1)}{\xi_0(\xi_1 - N + 1) - \xi_1(\xi_0 - N)}, \\ \tau_2 = \beta &= \frac{2(N - 2 - \xi_2)(\xi_0 - N)}{\xi_2(\xi_0 - N) - \xi_0(\xi_2 - N + 2)}, \\ \tau_3 = \alpha &= \frac{\xi_2(N - 1 - \xi_1) - 2\xi_1(N - 2 - \xi_2)}{\xi_2(\xi_1 - N + 1) - \xi_1(\xi_2 - N + 2)}. \end{aligned}$$

Then, on taking $\underline{\tau}' = (\tau_1, \tau_2, \tau_3)$, $W = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ and $\underline{\theta}' = (\alpha, \beta)$, we have $\underline{\tau} = W\underline{\theta}$. If \underline{t} be the sample counterpart of $\underline{\tau}$ and $\hat{\Sigma}$ be a consistent estimate of the asymptotic covariance matrix of \underline{t} , then the minimum chi-square estimator of τ is given by (4.1). Other minimum chi-square estimators for the BBD and the TBB distribution can be developed similarly by considering desired number of moments and/or probability relations.

5. Numerical examples

In this section we present fits to two data sets: one by the BBD and the other by the TBB. For comparison some fits to these data as obtained by other authors are also included. For obtaining minimum chi-square estimators, a consistent estimate of Σ , the asymptotic covariance matrix of \underline{t} , is needed which usually involves higher order moments. In order to avoid large random fluctuations resulting from substituting sample moments for the population moments involved in Σ , it is advisable to obtain estimates of such moments by putting consistent estimates of the parameters.

The 2-parameter BBD

Consider the data on distribution of "weeks" for 12 weeks appearing in the columns (1) and (2) of Table 3. The column (2) of this table contains the number of consumers out of 50 who purchased at least one unit of some specified consumer goods in exactly r out of 12 weeks, $r = 0, 1, 2, \dots, 12$. Denoting by p , the probability of making at least one purchase, the number of weeks out of n (12 in our case) in which the consumer makes at least one purchase will be binomially distributed with parameters n and p . However, the value of p will vary from consumer to consumer and so it is reasonable to consider p as a random variable with a beta distribution.

Column (3) shows a fit by the BBD using the method of mean and zeros (Chatfield and Goodhart (1970)). Columns (4) and (5) contain the fits to this data by the BBD using the estimator based on first two moments, and the minimum chi-square estimator based on 3 moment relations. For each of these fits the value of the Pearson's chi-square statistic along with its d.f. are also given. For the fits in columns (4) and (5) we also give the probabilities associated with the Pearson's chi-square statistic. For the fit in column (5), we give the value of the minimum chi-square statistic along with the associated d.f. and the probability. For the fits in columns (4) and (5), the values .793 and .722 of the Pearson's chi-square statistic with the associated probabilities .977 and .982 respectively show very good fits. This conclusion is also supported by the value .217 of the minimum chi-square statistic with the probability value .641 for the minimum chi-square fit in the column (5). Both of these fits are close to the fit obtained by Chatfield and Goodhart (1970) based on the method of mean and zeros.

The truncated BB distribution (TBB)

Consider the data on family epidemics of the common cold for families of size 5 appearing in columns (1) and (2) of Table 4. Column (3) of the table shows the fit

Table 3. A fit by BB family to the distribution of "weeks" (Chatfield and Goodhart (1970)).

(1)	(2)	(3)	(4)	(5)
No of weeks	Observed frequency	BB family	BB family	BB family
0	25	25.0	24.0	23.7
1	7	8.5	9.0	9.3
2	6	5.1	5.5	5.6
3	4	3.5	3.7	3.8
4	3	2.5	2.6	2.6
5	2	1.8	1.8	1.8
6	1	1.3	1.3	1.3
7	1	.9	.9	.8
8	0	.6	.6	.5
9	1	.4	.3	.3
10	0	.2	.2	.2
11	0	.1	.1	.1
12	0	.0	0.0	.0
Total	50	49.9	50	50
χ_p^2		.805	.722	.793
d.f.		5	5	5
$P_{\chi_p^2}$.982	.977
χ_{\min}^2				.217
d.f.				1
$P_{\chi_{\min}^2}$.641
$\hat{\pi}$.125	.125	.125
$\hat{\phi}$.321	.286	.269

(3) Fit obtained by BB family using the method of mean and zeros (Chatfield and Goodhart (1970)).

(4) Fit obtained by BB family using estimators based on two moment relations.

(5) Fit obtained by BB family using minimum chi-square estimator based on 3 moment relations.

by the TBBD using maximum likelihood estimates (Griffiths (1973)). Columns (4) and (5) show fits by the TBBD utilizing the estimates based on first two moments and the probability of the first count, and the minimum chi-square estimates based on 2 moment relations and the relations involving $\mu_{(1)}^*$ and P_1^* . For all the fits the value of the Pearson's chi-square statistic is given along with its d.f. For the fits in columns (4) and (5) the values of .766 for the Pearson's chi-square statistic with the associated probability of .682 show very good fits. The value .027 of the minimum chi-square statistic with the associated probability .870 for the minimum chi-square fit in column (5) also shows very good fit. The closeness of the fit in column (4) based on the simple estimators obtained from the first two moments and the probability of the first count is remarkable.

Table 4. Data on family epidemics of the common cold for families of size 5, i.e., $N = 5$, (Heasman and Reid (1961)) (Griffiths (1973), Table 3, Column 2).

(1)	(2)	(3)	(4)	(5)
No of cases	Observed	TBB family		
		(Griffiths (1973))	TBB family	TBB family
1	156	156.4	156.2	156.2
2	55	53.2	53.2	53.3
3	19	21.7	21.8	21.8
4	10	8.4	8.4	8.4
5	2	2.3	2.3	2.3
Total	242	242.0	232.0	242.0
χ_p^2		.8	.766	.766
d.f.		2	2	2
$P_{\chi_p^2}$.682	.682
χ_{\min}^2				.027
d.f.				1
$P_{\chi_{\min}^2}$				0.870
$\hat{\pi}$		0	.003	.002
$\hat{\phi}$.34	.342	.343

(3) Fit obtained by BB family using the method of mean and zeros (Chatfield and Goodhart (1970)).

(4) Fit obtained by BB family using estimators based on 1 moment relation and another relation involving $\mu_{(1)}^*$ and P_1^* .

(5) Fit obtained by truncated BB family using minimum chi-square estimators based on 2 moment relations and the relation involving $\mu_{(1)}^*$ and P_1^* .

6. Conclusion and summary

The paper presents some alternative methods for estimating the parameters of the beta binomial and truncated beta binomial models. Some of these methods are attractive since they yield estimators based on linear equations. The method of minimum chi-square yields estimators having high asymptotic relative efficiency in addition to being best asymptotically normal (BAN). This facilitates the construction of confidence intervals and tests regarding the parameters.

Acknowledgements

The authors are thankful to the referees for some useful suggestions which enhanced the presentation.

REFERENCES

Anscombe, F. J. (1950). Sampling theory of the negative binomial and the logarithmic series distributions, *Biometrika*, **37**, 358-382.

- Barankin, E. W. and Gurland, J. (1951). *On Asymptotically Normal Efficient Estimators*, University of California Publications in Statistics, 1, 89-129.
- Brooks, R. J. (1984). Approximate likelihood ratio tests in the analysis of Beta Binomial data, *Appl. Statist.*, **33**(3), 285-289.
- Chatfield, C. and Goodhart, G. J. (1970). The beta-binomial model for consumer purchasing behavior, *Appl. Statist.*, **19**, 240-250.
- Conte, S. D. and deBoor, C. (1980). *Elementary Numerical Analysis*, 3rd ed., McGraw Hill Book Company, New York.
- Griffiths, D. A. (1973). Maximum likelihood estimation for the beta binomial distributions and an application to the household distribution of the total number of cases of a disease, *Biometrics*, **29**, 637-648.
- Heasman, M. A. and Reid, D. D. (1961). Theory and observations in family epidemics of the common cold, *Brit. J. Prev. Soc. Med.*, **15**, 12-16.
- Kupper, L. L. and Haseman, J. K. (1978). The use of a correlated binomial model for the analysis of certain toxicological experiments, *Biometrics*, **34**, 69-76.
- Kupper, L. L., Portier, C., Hogan, M. D. and Yamamoto, E. (1986). The impact of litter effects on dose-response modeling in teratology, *Biometrics*, **42**, 85-98.
- Nissen-Meyer, S. (1964). Evaluation of screening in medical diagnosis, *Biometrics*, **20**, 730-755.
- Pack, S. E. (1986). Hypothesis testing for proportions with over dispersion, *Biometrics*, **42**, 967-972.
- Paul, S. R. (1982). Analysis of proportions of affected fetuses in teratology experiments, *Biometrics*, **38**, 361-370.
- Shenton, L. R. (1950). Maximum likelihood and the efficiency of the method moment, *Biometrika*, **37**, 111-116.
- Skellam, J. G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of a success as variable between the sets of trials, *J. Roy. Statist. Soc. Ser. B*, **10**, 257-261.
- Skurnick, J. H. (1990). Small sample power calculations based on the Beta Binomial model for comparison of probabilities, paper presented at the ENAR meeting, Baltimore, Maryland.
- Tarone, R. E. (1982). The use of historical control information in testing for a trend in proportions, *Biometrics*, **34**, 69-76.
- Tarone, R. E., Chu, K. C. and Ward, J. M. (1981). Variability in the rates of some common naturally occurring tumors in F344 rats and B6C3F, mice, *Journal of the National Cancer Institute*, **66**, 1175-1181.
- Tripathi, R. C. (1975). Families of discrete distributions with probability generating function involving hypergeometric functions, Ph.D. Thesis, Department of Statistics, University of Wisconsin, Madison.
- Williams, D. A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity, *Biometrics*, **31**, 949-952.
- Williams, D. A. (1982). Extra binomial variation in logistic linear models, *Appl. Statist.*, **31**(2), 144-148.
- Williams, D. A. (1988). Estimation bias using the beta binomial distribution in teratology, *Biometrics*, **44**, 305-309.