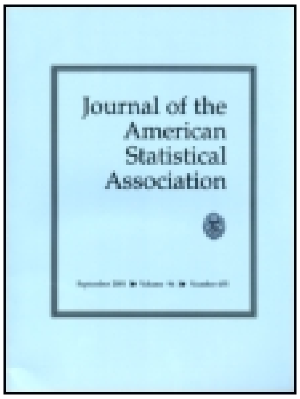

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://amstat.tandfonline.com/loi/uasa20

A Note on N Estimators for the Binomial Distribution

Raymond J. Carroll [a] & F. Lombard [b]
[a] Statistics, University of North Carolina, 315 Phillips Hall 039 A, Chapel Hill, NC, 27514, USA
[b] Statistics, University at South Africa, P.O. Box 392, Pretoria, South Africa
Published online: 12 Mar 2012.

To cite this article: Raymond J. Carroll & F. Lombard (1985): A Note on N Estimators for the Binomial Distribution, Journal of the American Statistical Association, 80:390, 423-426

To link to this article: http://dx.doi.org/10.1080/01621459.1985.10478134

A Note on N Estimators for the Binomial Distribution

RAYMOND J. CARROLL and F. LOMBARD*

Consider k success counts from a binomial distribution with unknown N and success probability p. We examine the problem of estimating N. By integrating the likelihood for N and p over a beta density for p, we obtain the beta-binomial distribution resulting in stable and reasonably efficient estimators of N, which compare favorably with and are often better than the estimates introduced by Olkin et al. (1981).

KEY WORDS: Analysis of count data; Maximum likelihood; Method of moments; Unstable estimators.

1. INTRODUCTION

Olkin, Petkau, and Zidek (OPZ; 1981) considered the problem of estimating the parameter N based on independent success counts $s_1, \ldots, s_k$ from a binomial distribution with unknown parameters N and p. They showed that the method of moments estimator (MME; see Haldane 1942) and the maximum likelihood estimator (MLE; see Fisher 1942) of N can be extremely unstable in the sense that changing an observed success count s to $s + 1$ can result in a massive change in the estimate of N. The difficulty arises when the method of moments estimates of mean and variance, $\hat{\mu}$ and $\hat{\sigma}^2$, are nearly equal, so the success probability p is apparently small.

To overcome the instability of the MME and MLE of N, OPZ introduced two estimators that they showed to be stable. The first is MLE:S, which is either the ordinary MLE or a jackknifed version of the maximum success count, depending on whether $\hat{\mu}/\hat{\sigma}^2 \geq 1 + 1/\sqrt{2}$. The stabilized MME:S also varies: if $\hat{\mu}/\hat{\sigma}^2 \geq 1 + 1/\sqrt{2}$, the usual MME is used; otherwise a ridge tracing method is employed (Hoerl and Kennard 1970).

Both MME:S and MLE:S are reasonably stable, and OPZ demonstrated in a convincing Monte Carlo Study that these estimators dominate the ordinary MME and MLE. They also showed that the ridge-stabilized MME:S is generally a better estimator of N than is the jackknife-stabilized MLE:S, except in unstable cases in which p is large. The purpose of this article is to describe a simple, stable estimator that is closely related to the MLE and seems to be competitive with and often superior to both MLE:S and MME:S in terms of mean squared error.

2. A NEW CLASS OF ESTIMATORS

The instability of the MME and MLE arises when p is apparently near zero. OPZ cited a case in which $N = 75$, $p = .32$, and the success counts are 16, 18, 22, 25, and 27. Even though p is not small, the natural estimate of it from the observed counts is $1 - \hat{\sigma}^2/\hat{\mu} = .21$. This is an example of an unstable case, since $\hat{\mu}/\hat{\sigma}^2 = 1.27$ even though $E\hat{\mu}/E\hat{\sigma}^2 = 1.84$. The extreme instability of the MLE and MME of N in this case was noted by OPZ. In the examples that motivated this research, namely counting the number of impala herds and individual waterbuck in the Kruger National Park, South Africa, it is fairly certain that p is much different from zero (see Sec. 4 for details). We reasoned from these examples that a stable procedure ought to be obtained if one smoothly builds in automatic discounting of data for which p is apparently near zero. In particular, it seemed that fairly stable procedures with good frequentist properties could be obtained by pretending that p had a beta distribution with parameters (a, b) and then looking at the likelihood obtained after integrating out p. Specifically, for $0 < p < 1$ and $N \geq s_{\max} = \max(s_1, \ldots, s_k)$, write the likelihood of the data as

$$L(N, p) = \left\{\prod_{i=1}^{k} \binom{N}{s_i}\right\} p^{\sum_1^k s_i} (1 - p)^{kN - \sum_1^k s_i}. \tag{1}$$

Suppose for the moment that the density of p is proportional to

$$p^a(1 - p)^b, \tag{2}$$

where a and b are integers. To eliminate the nuisance parameter, multiply (1) and (2) and integrate over p to obtain an integrated likelihood for N:

$$\mathcal{L}(N) = \left\{\prod_{i=1}^{k} \binom{N}{s_i}\right\} \times \left[(kN + a + b + 1)\binom{kN + a + b}{a + \sum_1^k s_i}\right]^{-1} \quad \text{for} \quad N \geq s_{\max}. \tag{3}$$

The estimate Mbeta (a, b) of N is obtained by maximizing (3) as a function of $N \geq \max(s_1, \ldots, s_k)$. Of course, in the standard terminology, (3) is the beta-binomial likelihood.

The idea of maximizing (3) as a function of N was justified in a Bayesian context by Draper and Guttman [1971; our Eq. (3) is equivalent to their (2.8)]. A non-Bayesian justification for eliminating nuisance parameters by integrating them out is given by Barnard et al. (1962; see pp. 348–350 in particular).

For every $a \geq 0$ and $b \geq 0$, the integrated likelihood (3) is maximized at some finite N. This follows because $\mathcal{L}(N) \to 0$ as $N \to \infty$, using Stirling's formula. We do not know if (3) always has a unique maximum when considered as a continuous function of N. In our calculations, however, we always found that (3) was either decreasing or first increasing and then decreasing in N, suggesting that (3) does have a unique maximum. DeRiggi (1983) showed that the likelihood function (1) evaluated at $p = \sum s_i/kN$ is unimodal; we have been unable to prove a similar result for (3).

* Raymond J. Carroll is Professor of Statistics, University of North Carolina, 315 Phillips Hall 039 A, Chapel Hill, NC 27514; and F. Lombard is Professor of Statistics, University at South Africa, P.O. Box 392, Pretoria, South Africa. Carroll's work was supported by Air Force Office of Scientific Research Contract F 49620 82 C 0009. Lombard's work was supported by research grants from the University of South Africa and the Council for Scientific and Industrial Research. The authors thank the associate editor for detailed and helpful comments and the referees for their suggestions.

Journal of the American Statistical Association
June 1985, Vol. 80, No. 390, Theory and Methods

Of course, one need not be restricted to having the distribution of p given N be beta (a, b). Indeed, different but fairly unnatural choices of distributions for p given N lead to some familiar, unstable estimators. For example, the MLE is formed by supposing that given N, p has a point mass distribution at $\hat{p}(N) = \Sigma_1^k s_i/N$. Following the prescription that led from (1) to (3) gives

$$\mathcal{L}(N : \text{MLE}) = \left\{\prod_{i=1}^{k} \binom{N}{s_i}\right\} \hat{p}(N)^{\Sigma_1^k s_i}(1 - \hat{p}(N))^{kN - \Sigma_1^k s_i}.$$

Part of the instability of the MLE may be due to the fact that as $N \to \infty$, it does not follow that $\mathcal{L}(N : \text{MLE}) \to 0$.

A second, rather strange choice is to pretend that the density of p is proportional to $1/p$ $(0 < p < 1)$. This gives extreme weight to small values of p and, as might be expected, leads to a very unstable estimator. It turns out that this unstable estimator is equivalent to the one obtained by maximizing the conditional likelihood for N given $\Sigma_1^k s_i$; the conditional likelihood for N differs from (3) here by a multiplier not dependent on N.

In contrast to the unrestricted and conditional MLE, our method uses a proper and natural distribution for p. We have chosen to fix the choice of (a, b) in line with our experience, but one could reasonably attempt to use the data to estimate (a, b). A Bayesian might also wish to construct a proper prior distribution for the parameter (N, p), the result of which might be an estimator with good frequentist properties.

Our method differs from that of Blumenthal and Dahiya (1981), who multiplied (1) and (2) together and then maximized this product jointly in N and p. They did not give any guidelines on how to choose (a, b) or on the stability of the result. From our limited calculations, it is clear that their choice of (a, b) must operate entirely differently from ours. In fact, our integration over p seems to induce stability for much smaller values of (a, b) than is the case with the Blumenthal and Dahiya method.

3. NUMERICAL WORK

The estimate of N obtained from maximizing (3) is reasonably stable. In Table 1 we analyze the examples listed in table 2 of OPZ, who computed MME, MME:S, MLE, and MLE:S for some particularly difficult cases; the MLE differs slightly from that of OPZ in cases 2 and 6 because of the extreme flatness of the likelihood in these cases. In addition, we provide the estimator Mbeta (0, 0) obtained from (3) with $a = b = 0$ (the uniform distribution) and the estimator Mbeta (1, 1) with $a = b = 1$. It is clear from these examples that MME and MLE are highly unstable. In addition, MME:S, MLE:S, Mbeta (0, 0), and Mbeta (1, 1) are clearly stable, with MME:S, Mbeta (0, 0), and Mbeta (1, 1) giving rather similar results. Cases 6 and 8 are particularly illustrative. Case 6 is an unstable case with large p, and here MLE:S dominates MME:S, with our Mbeta (0, 0) and Mbeta (1, 1) falling somewhere in between. Case 8 is unstable with small p, and MLE:S is now much worse than MME:S; again our estimators fall between the two, although they are more efficient in this case. The different behavior in unstable cases is reflected in the Monte Carlo study we now describe.

Table 1. N Estimates for Selected and Perturbed Samples

Sample	Parameters: N	p	K	Estimators: MME	MME:S	MLE	MLE:S	Mbeta (0, 0)	Mbeta (1, 1)
1	75	.32	5	102	70	19	29	51	49
				195	80	190	30	57	52
2	34	.57	4	507	77	514	31	52	47
				<0	91	∞	32	59	52
3	37	.17	20	65	25	66	11	26	23
				154	27	159	13	29	25
4	48	.06	15	18	10	15	7	9	8
				135	12	125	9	12	10
5	40	.17	12	32	26	40	21	27	25
				61	32	79	22	33	29
6	74	.68	12	210	153	213	67	135	125
				259	162	266	69	144	131
7	55	.48	20	71	69	71	43	64	63
				79	74	81	45	70	67
8	60	.24	15	67	49	67	24	45	41
				88	53	90	26	49	45

NOTE: The exact samples are given in table 2 of OPZ. For each sample number, the first entries are the N estimates for the original sample, and the second entries are the N estimates for perturbed samples obtained by adding one to the largest success count.

In Table 2 we expand the Monte Carlo study of OPZ, comparing MME:S and MLE:S with Mbeta (0, 0) and Mbeta (1, 1). All random numbers were generated by using the IMSL generators GGBTR and GGBN. The basic study was as in OPZ, so k was randomly chosen such that $3 \le k \le 22$, p was uniformly chosen such that $0 < p < 1$, and $1 \le N \le 100$ was uniformly chosen. There were 2,000 randomly generated cases. A case was called stable if $\hat{\mu} \ge (1 + 1/\sqrt{2})\hat{\sigma}^2$ and unstable otherwise. We also considered subcases in which $.2 \le p \le .8$, $0 < p \sqrt{2} - 1$ and $\sqrt{2} - 1 \le p < 1$.

Readers of an earlier version of this article pointed out that our study seemed biased in favor of our estimators, since p was uniformly distributed on $0 < p < 1$. To avoid this criticism we redid the study completely, generating beta (A, B) random variables [see (2)] with the asymmetric choices $(A, B) = (0, 1)$, (1, 2), (1, 3), (1, 0), (2, 1), and (3, 1). Since in each of these studies, our estimators performed as well as or better than

Table 2. Relative Mean Square Error Efficiencies of the N Estimates Relative to MME:S

Range	Stable Cases: No.	MME:S	MLE:S	Mbeta (0, 0)	Mbeta (1, 1)	Unstable Cases: No.	MME:S	MLE:S	Mbeta (0, 0)	Mbeta (1, 1)
$0 < p < 1$	1,367	1.00	.99	.99	1.03	633	1.00	.86	1.16	1.18
$.2 < p < .8$	863	1.00	.98	.99	1.08	336	1.00	1.18	1.96	2.79
$p < \sqrt{2} - 1.0$	281	1.00	.99	.96	.99	519	1.00	.66	.96	.92
$p > \sqrt{2} - 1.0$	1,086	1.00	.99	1.08	1.20	114	1.00	5.70	2.90	5.16

in the case $(A, B) = (0, 0)$ reported in Table 2, we do not report them.

For the unrestricted case in which $0 < p < 1$, the actual relative mean squared error efficiencies and the percentages of stable cases were similar to the results of OPZ. For the stable cases, the four estimators performed equally well. For the unstable cases, MLE:S was the clear loser, with the other three estimators being similar in performance.

When we consider the special cases $.2 < p < .8$, interesting results emerge. For the unstable cases, MME:S still beats MLE:S, but our estimators Mbeta (0, 0) and Mbeta (1, 1) are vastly superior to the other two. An intuitive reason for this may be that our estimators downweight the possibility that p is near zero.

Following OPZ, we also consider the cases of "small" p ($0 < p < \sqrt{2} - 1$) and "large" p ($\sqrt{2} - 1 \leq p < 1$). The former case is, as expected, least favorable to our estimators, which discount the possibility that p is small. However, our estimators still perform well; for example, Mbeta (0, 0) is only 4% less efficient in terms of relative mean squared error than MME:S, and Mbeta (1, 1) loses only 1% efficiency.

More striking results emerge when p is large ($\sqrt{2} - 1 \leq p < 1$). For the stable cases (90%) in this subset, our estimators have a definite advantage over MME:S and MLE:S, especially Mbeta (1, 1). For the few unstable cases, MLE:S is much better than MME:S (a fact noted by OPZ); even in these cases, our estimators perform competitively, and overall Mbeta (1, 1) emerges as the clear winner.

We found that the stabilized MLE:S was much more negatively biased than the three other stabilized estimates. Though all were negatively biased in general, the stabilized MLE:S had a bias in the unstable cases of almost 60% of the true value of N, versus 20% for the moments estimate MME:S and 30%–35% for our suggestions Mbeta (0, 0) and Mbeta (1, 1). Interestingly, for the 82% of unstable cases with true probability less than $\sqrt{2} - 1$, our suggestions were negatively biased, and for the other 18% of unstable cases, the bias was positive.

4. EXAMPLES

This research was motivated by the following two examples, the second of which is especially difficult. The counts of impala herds and individual waterbucks were obtained on five successive cloudless days in a small area of the Kruger Park. Counting was done from a light aircraft by five highly trained and experienced wildlife officials. The assumption of independent binomial counts with approximately equal success probabilities seems reasonable in this example, but the assumption is of course not absolutely indisputable.

Example 1. The observed number of herds of at least 25 impala were given as 15, 20, 21, 23, and 26. This is an unstable case, since $\hat{\mu}/\hat{\sigma}^2 = 1.59$. The various estimators are MME = 57, MLE:S = 28, MLE = 53, Mbeta (0, 0) = 42, MME:S = 54, and Mbeta (1, 1) = 42. When we changed the largest count from 26 to 27, the estimators MME:S, MLE:S, Mbeta (0, 0), and Mbeta (1, 1) exhibited little change. The moments estimator MME, however, changed from 57 to 77 and the MLE changed from 53 to 74. Note how the stabilized MLE:S is the smallest here, which is in line with the extreme negative-bias results found in the Monte Carlo study in the previous section. The conditional maximum likelihood estimator Mbeta (−1, 0) was fairly unstable here—95 for the original data but 215 for the perturbed data.

Table 3. N Estimates in the Waterbuck Data

a	Mbeta (a, 0) Original Data	Perturbed Data
0	146	155
−.25	159	168
−.50	179	193
−.75	225	251
−.90	311	367
−1.00	1,545	>4,000

Example 2. The observed number of waterbucks was 53, 57, 66, 67, and 72. Since $\hat{\mu}/\hat{\sigma}^2 = 1.32$, this is a highly unstable case. For the observed data, the estimates of N are MME = 272, MLE:S = 72, MLE = 265, Mbeta (0, 0) = 146, MME:S = 199, and Mbeta (1, 1) = 140. When we changed the largest count from 72 to 73, the estimates became MME = 362, MLE:S = 78, MLE = 355, Mbeta (0, 0) = 155, MME:S = 215, and Mbeta (1, 1) = 146. Note again the apparent extreme bias of the stabilized MLE:S.

The conditional MLE was again very unstable here—1,545 for the original data and >4,000 for the perturbed data. Because of the bias observed in the Monte Carlo study, we did some experimentation with the estimator Mbeta $(a, 0)$, with $a \leq 0$. The results are displayed in Table 3. Whether a reasonable, perhaps data-based choice of the value of a in Mbeta $(a, 0)$ would improve on the estimators we have studied remains to be seen. As noted in the previous section, it is in the highly unstable cases such as these waterbuck data for which negative bias is of most concern. Casella (1984) discusses an interesting graphical device for assessing the degree of instability of a given set of data. It seems that he implicitly suggests a data-dependent choice of (a, b), something along the lines of using Mbeta (0, 0) for stable cases but smoothly adjusting to Mbeta $(a, 0)$ as the instability increases; the waterbuck data suggest that we must stay strictly away from the conditional maximum likelihood estimate Mbeta (−1, 0).

5. ASYMPTOTIC THEORY

An illuminating general asymptotic theory for this problem awaits development. The stabilized method of moments and maximum likelihood estimators of OPZ have not been fully studied. Of course, as $k \to \infty$ for fixed N, all estimators discussed in this article are consistent. We have also considered an asymptotic theory for the estimators Mbeta (a, b) in the case of fixed (a, b), $N \to \infty$, $k \to \infty$, and $\sqrt{k}/N \to 0$. The results of this asymptotic theory are not too interesting because we find that regardless of the choice of (a, b),

$$k^{1/2}(\text{Mbeta }(a, b) - N)/N \xrightarrow{\mathcal{L}} \text{normal}\left(\text{mean} = 0, \text{variance} = \frac{2(1-p)^2}{p^2}\right). \quad (4)$$

Note that (4) suggests a large effect in variance for smaller values of p.

We have obtained two theoretical results that shed light on the behavior of our procedures. Both results involve fairly intricate calculations, which will not be presented here.

The first theoretical result illustrating the role of (a, b) in (3) and in the estimators Mbeta (a, b) occurs when we fix k—the number of observers—and let $N \xrightarrow{P} \infty$. In this case, none of the estimators will be consistent in the sense that $\hat{N}/N \xrightarrow{P} 1$. If we fix $k = 2$ and let the chi-squared limit distribution of $\{2p(1 - p)\}^{-1}(s_1 - s_2)^2/N$ be denoted by χ_1^2, then $\hat{\mu}/N \rightarrow p$ and $\{4p(1 - p)\}^{-1}\hat{\sigma}^2/N \overset{\mathfrak{L}}{\Rightarrow} \chi_1^2$. We can thus expect interesting results here because there will still be a positive probability of an unstable case. In fact, we obtain the following:

Lemma 1. In (3) take $a = b$ and $k = 2$. Let the chi-squared limit distribution of $\{2p(1 - p)\}^{-1}(s_1 - s_2)^2/N$ be denoted by χ_1^2. Then as $N \rightarrow \infty$,

$$\frac{\text{Mbeta}(b, b)}{N} \overset{\mathfrak{L}}{\Rightarrow} \frac{p}{(4b + 4)} \{6b + 3 + (1 - p)\chi_1^2 + \sqrt{(6b + 3 + (1 - p)\chi_1^2)^2 - 4(b + 1)(8b + 2)}\}. \quad (5)$$

The lemma illustrates one interesting facet of our estimators Mbeta (b, b) obtained from maximizing (3). The unstable cases are those in which $(s_1 - s_2)^2$ is large relative to $(s_1 + s_2)$. This corresponds to the situations in (5) in which χ_1^2 is large. Taking the limit of the right side of (5) as $\chi_1^2 \rightarrow \infty$, we obtain the proportionality

$$(5) \propto \chi_1^2 p(1 - p)/(2b + 2). \quad (6)$$

Equation (6) shows that the effect of increasing the smoothing parameter $a = b$ in (3) is a type of shrinkage. This agrees with the intuitive notion that the effect of larger (a, b) is to discount the possibility that p is small. This simple asymptotic theory helps explain why in Table 2 the most severe unstable cases are better handled by $(a = 1, b = 1)$ than by Mbeta $(a = 0, b = 0)$.

Our second useful asymptotic theoretical result illustrating the role of (a, b) in our estimator Mbeta (a, b) occurs under the following specifications:

$$N \rightarrow \infty, \qquad k \rightarrow \infty, \qquad k^{1/2}/N \rightarrow 0$$
$$a = \alpha k^{1/2}, \qquad b = \beta k^{1/2}. \quad (7)$$

Lemma 2. Consider the assumptions (7) with $\alpha > 0$ and $\beta > 0$. Then

$$k^{1/2}(\text{Mbeta}(a, b) - N)/N \overset{\mathfrak{L}}{\Rightarrow} \text{normal}\left(\frac{2(1 - p)^2\{(\alpha + \beta)p - \alpha\}}{p^2}, \frac{2(1 - p)^2}{p^2}\right). \quad (8)$$

Consider what Lemma 2 says qualitatively for the case $\alpha = b$ and $\alpha = \beta$, which was examined in our Monte Carlo study. Equation (8) suggests that in our Monte Carlo we should have found more negative bias for the case in which the true success probability $p < \sqrt{2} - 1$ than for the case $p \geq \sqrt{2} - 1$. This we found to be true for stable cases taken together as well as unstable cases taken together.

6. DISCUSSION

By considering beta-binomial distributions, we have obtained stable estimates that are at least competitive with, and in some instances superior to, the stabilized MME and MLE introduced by OPZ. OPZ were primarily interested in easily computed stable estimators with good efficiency properties, and thus it is natural that they did not consider refinements of their methods. In particular, they noted that perhaps their definition of unstable also ought to depend on k. We think their work is an excellent step toward better understanding of this difficult problem. Our estimators are differently motivated than theirs, and we hope that they will provide some additional insight. The advantages of our method include the flexibility of choosing a and b and the modification of the likelihood by smooth handling of the nuisance parameter p. We believe that further progress is inevitable and that even better estimates can be found. For example, one might suppose a natural joint distribution for (N, p) that downweights small p and large N; an estimator with good frequentist properties might emerge from such as Bayesian analysis.

Finally, little is known about the shape of $\mathfrak{L}(N)$ in (3), so the question of finding a confidence interval for N remains to be addressed.

[*Received May 1983. Revised October 1984.*]

REFERENCES

Barnard, G. A., Jenkins, G. M., and Winsten, C. B. (1962), "Likelihood Inference and Time Series" (with discussion), *Journal of the Royal Statistical Society*, Ser. A, 125, 321–372.

Blumenthal, S., and Dahiya, R. C. (1981), "Estimating the Binomial Parameter N," *Journal of the American Statistical Association*, 76, 903–909.

Casella, G. (1984), "Stabilizing Binomial N Estimators," manuscript submitted for publication.

DeRiggi, D. F. (1983), "Unimodality of Likelihood Functions for the Binomial Distribution," *Journal of the American Statistical Association*, 78, 181–183.

Draper, N., and Guttman, I. (1971), "Bayesian Estimation of the Binomial Parameter," *Technometrics*, 13, 667–673.

Fisher, R. A. (1942), "The Negative Binomial Distribution," *Annals of Eugenics*, 11, 181–187.

Haldane, J. B. S. (1942), "The Fitting of Binomial Distributions," *Annals of Eugenics*, 11, 179–181.

Hoerl, A. E., and Kennard, R. W. (1970), "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics*, 12, 55–67.

Olkin, I., Petkau, A. J., and Zidek, J. V. (1981), "A Comparison of N Estimators for the Binomial Distribution," *Journal of the American Statistical Association*, 76, 637–642.