A NOTE ON N ESTIMATORS FOR THE BINOMIAL DISTRIBUTION

R.J. Carroll
University of North Carolina at Chapel Hill

F. Lombard
University of South Africa, Pretoria

Keywords: Analysis of count data, maximum likelihood, method of moments, unstable estimators.

Abstract

Consider k success counts from a binomial distribution with unknown N and success probability p. We consider 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 which results in stable and reasonably efficient estimators of N, which compare favorably with and are often better than the estimates introduced by Olkin, Petkau and Zidek.

1. Introduction

Olkin, Petkau and Zidek (OPZ, 1981) consider 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 show that the method of moments estimator (MME, see Haldane (1941, 1942)) and the maximum likelihood estimator (MLE, see Fisher (1941, 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 sample mean $\hat{\mu}$ and sample variance $\hat{\sigma}^2$ of the success counts are nearly equal, so that the success probability p is apparently small.

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

Both MME:S and MLE:S are reasonably stable, and OPZ demonstrate in a convincing Monte-Carlo study that these estimators dominate the ordinary MME and MLE. They also show that the ridge-stabilized MME:S is generally a better estimator of N than the jackknife-stabilized MLE:S, except in unstable cases when p is large. The purpose of this note is to describe a simple, stable estimator which is closely related to the MLE and which 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 the MLE arises when p is apparently near zero. OPZ cite an case in which N = 75, p = 0.32 and the success counts are

16, 18, 22, 25, 27. Even though p is not small, the natural estimate of it from the observed counts is $1 - \hat{\sigma}^2/\hat{\mu} = .014$. This is an example of an unstable case since $\hat{\mu}/\hat{\sigma} = 1.01$ even though $E\hat{\mu}/E\hat{\sigma}^2 = 1.47$. The extreme instability of the MLE and MME of N in this case is noted by OPZ. In an example with which we are familiar, namely counting the number of impala herds in the Kruger National Park, it is fairly certain that p is much different from zero. We reasoned that a stable procedure ought to be obtained if one smoothly built in an automatic discounting of data for which p is apparently near zero. In particular, it seemed to us 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

$$(1) \qquad L(N,p) = \prod_{i=1}^{k} \binom{N}{s_i} p^{\Sigma_i^k s_i} (1-p)^{kN-\Sigma_i^k s_i} .$$

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

$$(2) \qquad f(p) \propto p^a (1-p)^b ,$$

where (a,b) are integers. Multiply (1) and (2) and integrate over p to obtain an integrated likelihood for N:

$$(3) \qquad \mathcal{L}(N) = \prod_{i=1}^{k} \binom{N}{s_i} \left[(kN+a+b+1)\binom{kN+a+b}{a+\Sigma_1^k s_i}\right]^{-1} \quad \text{for } N \geq s_{max} .$$

The estimate MB(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.

For every $a \geq 0$, $b \geq 0$, the integrated likelihood (3) is maximized at some finite N. This follows since $\mathcal{L}(N) \to 0$ 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. However, in our calculations, we always found that (3) was either decreasing or first increasing and then decreasing in N, suggesting that (3) does have a unique maximum.

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)-(3) gives

$$L(N{:}MLE) = \sum_{i=1}^{k} \binom{N}{s_i} \hat{p}(N)^{\Sigma_1^k s_i} (1-\hat{p}(N))^{kN-\Sigma_1^k s_i}.$$

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

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 $\Sigma_1^k s_i$.

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 multiply (1) and (2) together and then maximize this product jointly in N and p. They do not give any guidelines on how to choose (a,b) or the stability of the result.

3. Numerical Work

Our estimates are reasonably stable. In Table 1 we reproduce Table 2 of OPZ, who compute MME, MME:S, MLE, MLE:S for some particularly difficult cases. In addition, we provide the estimator MB(0,0) obtained from (3) with a=b=0 (the uniform distribution) and the estimator MB(1,1) with a=b=1. It is clear from these examples that MME and MLE are highly unstable. Also, MME:S, MLE:S, MB(0,0) and MB(1,1) are clearly stable, with MME:S, MB(0,0) and MB(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 MB(0,0) and MB(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.

In Table 2, we expand the Monte-Carlo study of OPZ, comparing MME:S and MLE:S with MB(0,0) and MB(1,1). All random numbers were generated using the IMSL generators GGBTR and GGBN. The basic study was as in OPZ, so that k was randomly chosen such that $3 \le k \le 22$, p was randomly chosen such that $0<p<1$ and $1 \le N \le 100$ was randomly chosen. There were 2000 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 $0.2 \le p \le 0.8$, $0<p<(\sqrt{2}-1)$ and $(\sqrt{2}-1) \le p<1$.

Readers of an earlier version of this manuscript pointed out that our study seems biased in favor of our estimators, since p was uniformly distributed on $0<p<1$. This is not really the case, but to avoid the criticism we completely redid the study, generating Beta (A,B) random variables (see (2)) with the asymmetric choices (A,B) = (0,1), (1,2), (1,3), (1,0), (2,1), (3,1). Since in each of these studies our estimators performed as well or better than for the case (A,B) = (0,0) reported in Table 2, we do not report them.

For the general case in which $0<p<1$, the actual relative mean square error efficiencies and the percentage of stable cases were quite similar to the results of OPZ. For the stable cases, all four estimators had similar performance. For the unstable cases, MLE:S was the clear loser, with the other three estimators being similar in performance.

When we considered the special cases $.2<p<.8$, interesting results emerge. For the unstable cases, MME:S still beats MLE:S, but our estimators MB(0,0) and MB(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 considered the cases of "small p" ($0<p<\sqrt{2}-1$) and "large p" ($\sqrt{2}-1\le p<1$). 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 and, for example, MB(0,0) is only 4% less efficient than MME:S.

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

4. Discussion

By considering beta-binomial distributions, we have obtained stable estimates which 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 note that their definition of unstable ought perhaps to also depend on k. We think their work is an excellent step towards better understanding of this difficult problem. Our estimators are differently motivated than theirs, and we hope they will provide some additional insight. The advantages of our method include the flexibility of choosing a and b and the fact that we modify the likelihood by smoothly handling the nuisance parameter p. We believe that further progress is inevitable and that even better estimators can be found. For example, one might suppose a natural joint distribution for (N,p) which downweights small p and large N; an estimator with good frequentist properties might emerge from such a Bayesian analysis.

Acknowledgement

The work of R.J. Carroll was supported by the Air Force Office of Scientific Research Contract F49620 82 C 0009.

TABLE 1. N Estimates for Selected and Perturbed Samples[a]

Parameters				Estimators					
Sample #	N	p	K	MME	MME:S	MLE	MEL:S	MB(0,0)	MB(1,1)
1	75	.32	5	102	70	19	29	51	49
				195	80	190	30	57	52
2	34	.57	4	507	77	504	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	201	67	135	125
				259	162	237	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

[a]The exact samples are given in T able 2 of OPZ. For each sample number, the first entries are the N estimates for the original sample, while the second entries are the N estimates for perturbed samples obtained by adding one to the largest success count.

TABLE 2. Relative Mean Square Error Efficiencies of the N-estimates Relative to MME:S

Stable Cases

Range of p	# of Cases	MME:S	MLE:S	MB(0,0)	MB(1,1)
$0<p<1$	1367	1.00	0.99	0.99	1.03
$0.2<p<0.8$	863	1.00	0.98	0.99	1.08
$p<\sqrt{2}-1.0$	281	1.00	0.99	0.96	0.99
$p>\sqrt{2}-1.0$	1086	1.00	0.99	1.08	1.20

Unstable Cases

Range of p	# of Cases	MME:S	MLE:S	MB(0,0)	MB(1,1)
$0<p<1$	633	1.00	0.86	1.16	1.18
$0.2<p<0.8$	336	1.00	1.18	1.96	2.79
$p<\sqrt{2}-1.0$	519	1.00	0.66	0.96	0.92
$p\geq\sqrt{2}-1.0$	114	1.00	5.70	2.90	5.16

REFERENCES

Blumenthal, S. and Dahiya, R.C. (1981). Estimating the binomial parameter n. Journal of the American Statistical Association 76, 903-909.

Fisher, R.A. (1941, 1942). The negative binomial distribution. Annals of Eugenics 11, 181-187.

Haldane, J.B.S. (1941, 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.