A Comparison of n Estimators for the Binomial Distribution

INGRAM OLKIN, A. JOHN PETKAU, and JAMES V. ZIDEK*

Estimating the number, n, of trials, given a sequence of independent success counts obtained by replicating the n-trial experiment is less studied and a considerably harder problem than estimating p for the binomial distribution. Both the method of moments and the maximum likelihood estimators of n are considered. When p is unknown and small, while n is large, both estimators become highly unstable. Stabilized versions of these estimators are proposed, and their performance is described in terms of results obtained by simulation.

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

1. INTRODUCTION

The standard problem associated with the binomial distribution is that of estimating its probability, p, of success. A much less well studied and considerably harder problem is that of estimating the number, n, of trials. In this article we assume success counts, $s_1, \ldots, s_k$, are available from independently observable binomial random variables, $S_1, \ldots, S_k$, where n and p are fixed but unknown. Our goal is to devise and compare estimators of n.

The genesis of our interest in this problem stems from the analysis of data such as that pertaining to crime in which there are unreported counts. As a tentative model for this situation, it may be assumed that the total number, n, of incidents per time period has a Poisson distribution. Let $n_1, \ldots, n_k$ and $s_1, \ldots, s_k$ respectively denote the total number and the reported number of incidents in k successive time periods. Since the mean, say λ, of the n process equals its variance, the relative variation $\sqrt{\lambda}/\lambda$ in the n_i would be small compared with that of the s process, $\sqrt{p\lambda}/(p\lambda)$. As a first approximation, it might seem reasonable to regard the n_i as constant, $n_i \simeq \lambda$; this leads to the problem just formulated.

A particular situation in which the estimation of n is required is given in Draper and Guttman (1971). There the s_i are the numbers of a particular sort of appliance, in a given locality, which are brought in for repair during week i and n is the total number of appliances in that locality. Another situation is considered by Moran (1951).

The history of this problem is sketched in Section 2. In Section 3 we discuss two natural estimators, those obtained by maximum likelihood and the method of moments. These estimators are shown to be highly unstable in certain circumstances; most of Section 3 is devoted to finding stable alternatives. The performance of these estimators is investigated numerically in Section 4.

2. HISTORY

This problem has a long history. In the work of Student (1918), Haldane (1941, 1942), and Fisher (1941, 1942) the negative binomial, Poisson, and binomial distributions were employed in fitting probability models to count data; these were thought of as a single family indexed by n and p, where $-\infty < n < 0, p < 0$ corresponded to the negative binomial; $|n| \to \infty$, $|p| \to 0$, $np = \lambda > 0$, to the Poisson; and $\max s_i \le n < \infty$, $0 < p < 1$, to the binomial distribution. Which of these models was chosen would depend then on which values of n and p were chosen.

After pointing out qualitative differences between the binomial and negative binomial, Fisher (1941, 1942) goes on to dismiss the problem of fitting a binomial as artificial and discusses instead the corresponding problem for the negative binomial. His argument for this dismissal is that for sufficiently large k, n will be known *exactly*. Although his argument is correct, it is not compelling. For if p is small, k will have to be unrealistically large before n is known with any degree of certainty.

Binet (1954) argues in favor of the binomial against the other two contenders on the grounds that in many applications the "counts" are obviously bounded by a small positive (but unknown) integer. So a priori the other two can be dismissed since they distribute probability over *all* nonnegative integers.

Fisher (1941, 1942) considers a problem addressed by Haldane (1941, 1942), namely, should the method of moments estimators (MME's) or maximum likelihood estimators (MLE's) be used to fit n and p? Haldane (1941, 1942), while recognizing that the former is not fully ef-

* Ingram Olkin is Professor, Department of Statistics, Stanford University, Stanford, CA 94305. A. John Petkau is Associate Professor and James V. Zidek is Professor, both in the Department of Mathematics, University of British Columbia, Vancouver, Canada V6T 1Y4. The work was supported in part by the National Science Foundation under a grant at Stanford University and the Natural Sciences and Engineering Research Council of Canada under Grants A4586 and A5241. The authors wish to thank Bruce Levin and Don Rubin for their comments on an earlier draft of this article and are grateful to the editor and the referees for their helpful suggestions.

© Journal of the American Statistical Association
September 1981, Volume 76, Number 375
Theory and Methods Section

ficient, provides an example where the results are in close agreement. Fisher (1941, 1942) actually computes the asymptotic efficiency of the method of moments and so is able to state rather precise conditions on n, p, and k under which the method may be used with little loss.

Although the likelihood equations for the case in which both n and p are unknown are presented by Haldane (1941, 1942), the question of the uniqueness of the MLE is not discussed. It is raised by Anscombe (1950) for the negative binomial and is answered affirmatively by Levin and Reeds (1977) and again by Bonitzer (1978). The question remains unanswered for the binomial. A detailed survey of the state of this problem is given in Section 3.2.

3. THE n ESTIMATORS

Theoretical issues stimulated our interest in this problem. Since the Poisson model discussed in the introduction yields an observed success counts distribution that is Poisson with parameter $p\lambda$, estimation of λ itself is not feasible. At the same time, that model should be similar in a heuristic sense to the model that underlies the problem analyzed in this article, at least if the success probability, p, is small and the number, n, of trials is large. Yet in the latter model the parameters are identifiable, that is, have consistent estimators as $k \to \infty$. It is to be expected, therefore, that when these two models are "close," estimators of n and p will be ill behaved.

This anticipated instability of n estimators does, in fact, occur. The mean and variance of the binomial are $\mu = np$ and $\sigma^2 = \mu(1 - p)$, respectively. Thus when p is small, σ^2/μ is approximately one. In this case, the probability that $\hat\sigma^2/\hat\mu \geq 1$ is particularly large where $\hat\mu$ and $\hat\sigma^2$ denote the s-sample counterparts of μ and σ^2, respectively. When $\hat\sigma^2/\hat\mu \geq 1$, the likelihood function is maximized at $\hat n = \infty$ and $\hat p = \hat\mu/\hat n = 0$. Thus the likelihood approach leads to the rejection, in this case, of the binomial model in favor of the Poisson model obtained when $\hat n \to \infty$ and $\hat p = \hat\mu/\hat n \to 0$. Even when $\hat\sigma^2/\hat\mu < 1$, the likelihood function is almost flat if $\hat\sigma^2/\hat\mu$ is close to one. This causes the MLE (among others) to be extremely sensitive to even slight perturbations of the data and hence unreliable.

To see the instabilities described above consider the sample consisting of the success counts 16, 18, 22, 25, 27. The MLE and MME of n are, respectively, 99 and 102. Suppose the 27 appearing in this sample is a misrecorded 28. The corresponding estimators for the correct sample are 190 and 195! Incidentally, the value of n and p used to generate the above success counts were 75 and .32, respectively.

Thus the MLE, and the MME as well, fail to be robust in the presence of recording errors. Since in practice it is to be expected that the observed success counts are susceptible to these errors, this lack of robustness casts grave doubt on the value of the estimators in question. There does not seem to be any well-defined procedure to which one can appeal to overcome this lack of robustness. Although we have not attempted to formulate such a general procedure, we have exploited the opportunity presented by the particular problem presented here to deal with the question, albeit in an ad hoc and experimental fashion. Thus we shall present stabilized versions of the method of moments and maximum likelihood estimators. The simulation experiments described in Section 4 show our methods to be surprisingly effective. In particular, in the previous example the stabilized versions of the MLE and the MME give 29 and 70, respectively, when the last count is 27 and 30 and 80 when it is 28.

3.1. Method of Moments

Since $\mu = np$ and $\sigma^2 = np(1 - p)$ are the mean and variance of a given success count, S_i, the method of moments gives $\hat p = \hat\mu/\hat n$, $\hat n = \hat\mu^2/(\hat\mu - \hat\sigma^2)$, where $\hat\mu = \sum s_i/k$, $\hat\sigma^2 = \sum(s_i - \hat\mu)^2/k$ and

$$\text{MME} = \hat n. \tag{3.1}$$

The shortcomings of the MME are apparent. If $\hat\sigma^2 > \hat\mu$, then MME < 0, which is unrealistic. Even if $\hat\sigma^2 < \hat\mu$, the MME will be highly unstable when $\hat\mu$ is close to $\hat\sigma^2$.

The method for stabilizing the MME that we adopt is suggested by the ridge tracing method of Hoerl and Kennard (1970). Suppose a constant $\epsilon > 0$ is added to each observed success count, s_i, so that $\hat\mu \to \hat\mu + \epsilon = x$, say, while $\hat\sigma^2 \to \hat\sigma^2$. Then the MME $\to x^2/(x - \hat\sigma^2) = n(x)$, say. Thus $n(\hat\mu) =$ MME. For ϵ sufficiently large, $n(x) > 0$ even if MME < 0. Furthermore, $n'(x) = 1 - [\hat\sigma^2/(x - \hat\sigma^2)]^2$, from which it is clear that $n(x)$ reaches a minimum at $x = 2\hat\sigma^2$ and then increases to ∞. Finally, even though $n(x)$ is rapidly changing for x near $\hat\sigma^2$, it stabilizes as ϵ, and hence x, increases.

To achieve a stable version of the MME, that is, one that is insensitive to small perturbations in the s_i, we choose $\epsilon > 0$ so that $x > \hat\sigma^2$ and, somewhat arbitrarily, $n'(x) \geq -1$. Finally, for consistency we require $n(x) \geq s_{\max} + \epsilon$, where $s_{\max} = \max\{s_1, \ldots, s_k\}$; this is because $n(x)$ represents the number of trials, whereas $s_{\max} + \epsilon$ represents the maximum number of successes in the perturbed data set.

If $n'(\hat\mu) \geq -1$, that is, if

$$\hat\mu/\hat\sigma^2 \geq 1 + 1/\sqrt{2} \cong 1.71, \tag{3.2}$$

we say the MME is stable. Condition (3.2) may be thought of as a preliminary test of significance. If this test rejects stability, then ϵ is chosen so that $n'(x) = -1$; that is, $\epsilon = \epsilon_0$ where

$$\epsilon_0 = (1 + 1/\sqrt{2})\hat\sigma^2 - \hat\mu, \tag{3.3}$$

provided that $s_{\max} \leq n(\hat\mu + \epsilon_0) - \epsilon_0$; that is,

$$z/\hat\sigma \leq 1 + \sqrt{2}, \tag{3.4}$$

where $z = (s_{\max} - \hat\mu)/\hat\sigma$. At this value, $n(x) = (2 + 3/\sqrt{2})\hat\sigma^2 \doteqdot 4\hat\sigma^2$. Otherwise, ϵ_0 is chosen as large as possible subject to condition (3.4); that is, $\epsilon = \epsilon_0$ where $\epsilon_0 = [s_{\max}$

$(\hat{\mu} - \hat{\sigma}^2) - \hat{\mu}^2]/[\hat{\mu} + \hat{\sigma}^2 - s_{\max}]$. At this value, $n(x) = \hat{\sigma}^2(z/\hat{\sigma})^2/(z/\hat{\sigma} - 1)$. Observe that we may summarize these calculations by asserting that in the unstable case, $n(x) = \hat{\sigma}^2 w^2/(w - 1)$ where $w = \max[1 + \sqrt{2}, z/\hat{\sigma}]$.

Our stabilized version of MME is therefore

$$\text{MME:S} = \max\{\hat{\sigma}^2\phi^2/(\phi - 1), s_{\max}\} \qquad (3.5)$$

where

$$\phi = \begin{cases} \hat{\mu}/\hat{\sigma}^2, & \text{if } \hat{\mu}/\hat{\sigma}^2 \geq 1 + 1/\sqrt{2}, \\ \max[z/\hat{\sigma}, 1 + \sqrt{2}], & \text{if } \hat{\mu}/\hat{\sigma}^2 < 1 + 1/\sqrt{2}; \end{cases}$$

here, as before, $z = (s_{\max} - \hat{\mu})/\hat{\sigma}$.

3.2. Method of Maximum Likelihood

The likelihood function, $L(n, p)$, is given by

$$L(n, p) = \prod_{i=1}^{k} [n(n-1)\cdots(n - s_i + 1)/s_i!]p^{s_i}(1-p)^{n-s_i}$$

for $0 \leq p \leq 1$ and $s_{\max} \leq n < \infty$. This likelihood may well be maximized as $n \to \infty$, $p \to 0$, and $np = \hat{\mu}$. This limiting case corresponds to that obtained for the Poisson law. For expository convenience we will denote this limiting likelihood by $L(\infty, 0)$.

The proof of the following theorem was provided to us by Bruce Levin; the argument is an adaptation of that of Levin and Reeds (1977).

Theorem 1. An MLE of n exists. If $\hat{\mu}/\hat{\sigma}^2 \leq 1$, $L(n, p) < L(\infty, 0)$, while if $\hat{\mu}/\hat{\sigma}^2 > 1$, $L(n, p)$ is maximized by at least one pair, (n, p), with $s_{\max} \leq n < \infty$.

Proof. For each n, $L(n, p)$ is maximized by $p = \hat{p} = \hat{\mu}/n$. Set $p = \hat{\mu}/n$ and differentiate, with respect to n, the log-likelihood. Then on setting the result equal to zero, obtain the following equation (c.f. Haldane 1941, 1942):

$$\sum_{i=1}^{k}{}' \sum_{j=0}^{s_i - 1} (n - j)^{-1} + k\log(1 - \hat{\mu}/n) = 0, \qquad (3.6)$$

where the accent on $\sum$ indicates summation over all i for which $s_i \geq 1$.

Suppose $\hat{\mu}/\hat{\sigma}^2 \leq 1$. It is easily seen that the left-hand side of (3.6) may be rewritten as $\int (N + u)^{-1} d[G(u) - F(u)] = d(N)$, say, where $N = n - s_{\max} \geq 0$, $G = \sum G_i$ and G_i places unit mass at the points $\{s_{\max} - s_i + 1, \ldots, s_{\max}\}$; F is $k \times$ Lebesque measure on $\{s_{\max} - \hat{\mu}, s_{\max}\}$. Let $M = G - F$. Since $(N + u)^{-1} = \int_0^\infty \exp[-(N + u)y]dy$,

$$d(N) = \int \exp(-Ny)\mathcal{L}_{dM}(y)\, dy, \qquad (3.7)$$

where $\mathcal{L}_{dM}(y) = \int \exp(-yu)\, dM(u)$. As is well known, d has no more sign changes than $\mathcal{L}_{dM}$. Let $I(u) = u$ or 0 according to whether $u \geq 0$ or $u < 0$, and let $*$ denote convolution. Then $\mathcal{L}_{I*M}(y) = \mathcal{L}_I(y)\mathcal{L}_{dM}(y) = y^{-2}\mathcal{L}_{dM}(y)$. So d has no more sign changes than $\mathcal{L}_{I*M}(y)$, which is $\int_0^\infty \exp(-uy)\, H(u)\, du$ where

$$H(u) = \int_0^u (u - x)\, dM(x) = \int_0^u [G(x) - F(x)]\, dx.$$

Thus d has no more sign changes than H. We now show that $H(u) > 0$.

It is straightforward to show that if $u > s_{\max}$, $H(u) = \frac{1}{2}k(\hat{\sigma}^2 - \hat{\mu}) > 0$. Furthermore, if $u \leq s_{\max}$, $H(u) = R(s_{\max} - u) - R(s_{\max})$ where

$$R(w) = \sum_{i=1}^{k}{}' \sum_{j=0}^{s_i - 1} (w - j)I\{0 \leq w - j\} - \tfrac{1}{2}kw^2 + \tfrac{1}{2}k(w - \hat{\mu})^2 I\{0 < w - \hat{\mu}\} \qquad (3.8)$$

and $I\{A\}$ denotes the indicator function of the set A, whatever A might be. It is easy to show that $R'(w) \leq 0$ for $w \geq \hat{\mu}$. Levin and Reeds (1977) show for $w < \hat{\mu}$ that $R(w) < 0$ implies $R'(w) < 0$. Furthermore, $R(s_{\max}) = \frac{1}{2}k(\hat{\mu} - \hat{\sigma}^2) \leq 0$ in the present case. Combining these facts leads to the conclusion that $H > 0$ and so $d > 0$. Thus the MLE, $\hat{n}$, equals infinity (the Poisson case with $\hat{n}\hat{p} = \hat{\mu}$).

Next suppose $\hat{\mu}/\hat{\sigma}^2 > 1$. Since $H(u) = \frac{1}{2}k(\hat{\sigma}^2 - \hat{\mu}) < 0$ for $u \geq s_{\max}$, $H(0) = 0$, and $H'(u) = -R'(w) \geq 0$ for $0 < u < s_{\max} - \hat{\mu}$ (or $\hat{\mu} < w < s_{\max}$), it follows that $\mathcal{L}_{I*M}(y) = \int_0^\infty \exp(-yu)\, H(u)\, du$ is < 0 or > 0 according as y approaches 0 or ∞. Hence $\mathcal{L}_{I*M}(y)$ has at least one sign change, and in fact, approaches $-\infty$ as $y \to 0$. The same comment applies therefore to $\mathcal{L}_{dM}(y)$. Finally, it follows that $d(N) < 0$ for large values of N. De Riggi (1978) gives a somewhat different argument for the case $\hat{\mu}/\hat{\sigma}^2 > 1$ than the one given here.

In the analysis of count data, the negative binomial is inevitably a competitor to the binomial; only the domain of the likelihood and not the function itself differs for these two models. For the former model, the argument corresponding to the one just given has already been carried out by Levin and Reeds (1977) and again by Bonitzer (1978). They show that $L(n, \hat{\mu}/n) > L(\infty, 0)$ for exactly one n in the range $-\infty < n < 0$ if $\hat{\mu}/\hat{\sigma}^2 < 1$, and further that if $\hat{\mu}/\hat{\sigma}^2 \geq 1$, then $L(n, \hat{\mu}/n) < L(\infty, 0)$ for all n in that range.

We may now combine all of these results to obtain a rather complete picture of the likelihood derived from either of these two models or the Poisson. Over the range $-\infty \leq n \leq \infty$, the behavior of $L(n, \hat{\mu}/n)$ is summarized as follows: if $\hat{\mu}/\hat{\sigma}^2 < 1$, then L is maximized at a unique, negative (finite) value of n; if $\hat{\mu}/\hat{\sigma}^2 = 1$, then L is maximized at $|n| = \infty$; if $\hat{\mu}/\hat{\sigma}^2 > 1$, then L is maximized at one or more positive (finite) values of n. Thus, in the analysis of count data, maximum likelihood prescribes the negative binomial, Poisson, or binomial models according to whether $\hat{\mu}/\hat{\sigma}^2$ is $<$, $=$, or $>$ 1.

All further discussion is restricted to the case of the binomial in which $s_{\max} \leq n < \infty$. In this situation the MLE, like the MME, may be highly unstable; small perturbations in the s_i may cause gross changes in the MLE. Although conditions guaranteeing the stability of the MLE cannot be specified precisely, instability increases as $\hat{\mu}/\hat{\sigma}^2 \to 1$ from above. Somewhat arbitrarily, and to achieve concurrence with the MME, we will call the MLE stable if $\hat{\mu}/\hat{\sigma}^2 \geq 1 + 1/\sqrt{2}$.

One approach to the stabilization of the MLE that mimics that of Section 3.1 would consist of perturbing the data, $s_i \rightarrow s_i + \epsilon$, with ϵ chosen large enough to make the corresponding value of $d(\epsilon) \leq 0$. This alternative is not practical in view of the complexity of the function d and because, unlike the MME, it cannot be defined for nonintegral data. The numerical process of determining a sufficiently large integral ϵ, if such an ϵ exists, would be extremely difficult. This would in turn prove costly when carried out repeatedly, as would have to be done in evaluating the performance of the resulting estimator in a long sequence of simulated experiments like that carried out in Section 4. Finally, it would not be possible to define the resulting estimator explicitly, as is done in (3.5) for MME:S.

An alternative approach follows: $s_{\max} = \hat{n}_k$, say, is adopted as an initial (stable) estimate of n. Obviously $\hat{n}_k$ is biased; it underestimates n. To reduce this bias we then apply the jackknife technique (see, e.g., Cox and Hinkley 1974, p. 261).

In general terms, the jackknife procedure may be described as follows. Suppose given a rule, $\hat{\theta}_n$, which from any given sample of size n estimates θ. Delete successively the ith observation, $i = 1, \ldots, n$, and from the resulting sample of size $n - 1$, apply the rule to obtain the estimate, say $\hat{\theta}^i_{n-1}$. Next obtain the pseudovalues, $\hat{\theta}_i = n\hat{\theta}_n - (n - 1)\hat{\theta}^i_{n-1}$, $i = 1, 2, \ldots, n$. Then the jackknife estimate is given by $\hat{\theta} = \sum_1^n \hat{\theta}_i/n$.

In our case the rule consists of taking the largest order statistic to estimate n. Thus for the sample of k values of s_i, we obtain $\hat{n}_k = s_{[k]} = s_{\max}$. Then $\hat{n}^i_{k-1}$ are determined for $i = 1, \ldots, k$. Of these k estimates based on samples of size $k - 1$, $k - 1$ estimates are equal to $s_{[k]}$ and the remaining one is $ks_{[k]} - (k - 1)s_{[k-1]}$. It follows that the jackknife estimator of n, JK, is given by

$$JK = s_{\max} + [(k - 1)/k]\,[s_{\max} - s_{[k-1]}]. \tag{3.9}$$

The latter estimate eliminates bias of the order $1/k$. Additional bias of order $1/k^2$ may be removed by applying the two-stage jackknife (see, e.g., Gray and Schucany 1972). The simulation experiments described in Section 4, however, indicate that the latter procedure leads to no additional reduction in the mean squared error of estimation and thus MLE:S is defined as

$$\text{MLE:S} = \begin{cases} \text{MLE}, & \text{if } \hat{\mu}/\hat{\sigma}^2 \geq 1 + 1/\sqrt{2}, \\ \text{JK}, & \text{if } \hat{\mu}/\hat{\sigma}^2 < 1 + 1/\sqrt{2}, \end{cases} \tag{3.10}$$

where JK is as defined in (3.9). Since we have not established uniqueness of the MLE in the case $\hat{\mu}/\hat{\sigma}^2 > 1$, there is still some ambiguity in (3.10). We adopt the following as our approximation to the MLE in the case $\hat{\mu}/\hat{\sigma}^2 \geq 1 + 1/\sqrt{2}$. If $d(0) \leq 0$, MLE $= s_{\max}$; if $d(0) > 0$, MLE is taken as the solution of $d(\text{MLE} - s_{\max}) = 0$, which is obtained by Newton's method with $s_{\max}$ as an initial guess.

4. NUMERICAL RESULTS

To evaluate the efficacy of the n estimators derived in Section 3, a series of simulated experiments were carried out. One possible experimental program would have consisted of generating for fixed n, p, and k a sequence of success count vectors, $(s_1, \ldots, s_k)$, and then evaluating the performance of the n estimators in some overall average sense for this sequence. This procedure would then be repeated for other choices of n, p, and k. This approach was rejected in favor of an alternative of the sort proposed by Dempster, Schatzoff, and Wermuth (1977). Thus at each step, values of n, p, and k are generated at random, and in turn, a single success count vector. After a sequence of such steps is carried out, the overall performance of the estimators is investigated. This method seems to us to better indicate how well the estimators would have performed had they been used in a wide variety of individual problems.

In the absence of knowledge about the particular problems to which the estimators might be applied, and to insure that they are adequately tested in a wide variety of problems, n, p, and k were generated by means of a

Table 1. Comparison of MME:S and MLE:S

				Wins		Relative Error	
	Run	Cases	Ties	MME:S	MLE:S	MME:S	MLE:S
Stable cases	1	711 71%	337 47%	190 27%	184 26%	.349	.349
	2	694 69%	314 45%	209 30%	171 25%	.352	.351
	1 + 2	1405 70%	651 46%	399 28%	355 25%	.350	.351
Unstable cases	1	289 29%	21 7%	208 72%	60 21%	.551	.620
	2	306 31%	26 8%	224 73%	56 18%	.557	.649
	1 + 2	595 30%	47 8%	432 73%	116 19%	.554	.635

uniform distribution. The ranges, $0 < p < 1$, $1 \leq n \leq 100$, and $3 \leq k \leq 22$ seemed realistic and were employed.

Two distinct runs of 1,000 cases were carried out. The n estimators, MME:S and MLE:S, were in each case rounded to the nearest integer and then compared by two criteria. The first was simply a count of the number of times each estimator won, that is, was closer to the true value of n; in many cases these estimators tied. The second was a measure of the relative error of estimation, namely, the square root of the average value of $(\hat{n} - n^2)/n^2$ for the cases under consideration.

Each of the 1,000 cases, in each of the two runs, was classified as stable if and only if $\hat{\mu}/\hat{\sigma}^2 \geq 1 + 1/\sqrt{2}$. A summary of the performance of MME:S and MLE:S in terms of these criteria is presented in Table 1; the similarity of the results obtained in the two runs indicates that at least for the restricted ranges of the parameters under consideration, the combined set of 2,000 cases should provide a reasonably reliable basis for comparison of MME:S and MLE:S.

Examination of Table 1 reveals that the conventional n estimators obtained by maximum likelihood and the method of moments appear to work satisfactorily and equally well in the stable case (sample mean/sample variance $\geq 1 + 1/\sqrt{2}$). In particular, MME > 0 and MLE $< \infty$. In terms of their proximity to n, MME wins in 28 percent of the cases and MLE in 25 percent of the cases; a tie occurs in the remaining cases. The measure of relative error of estimation is .35 for both methods. Incidentally, both the one-stage and two-stage jackknife procedures do somewhat worse than the MLE and MME in these stable cases (the measure of relative error is .39 for the former two).

In the unstable cases, MME:S does appreciably better than MLE:S, which is, by construction, just the one-stage jackknife procedure in these cases. By construction both these estimators are stable and, in particular, MME:S > 0 and MLE:S $< \infty$ in all cases. The measure of relative error of estimation is .55 and .64, respectively, for the MME:S and MLE:S. At the same time, the former estimator is closer to n than the latter in 73 percent of the cases (they tie in only 8 percent). At first glance, this rather substantial victory might seem at variance with the MME:S's narrower victory margin as measured by the error of estimation. The difference is explained by the fact that the MME:S makes occasional gross errors that inflate the error of estimation markedly but only add one to the MLE:S's win count; for example, in one case $n = 53$, MLE:S $= 33$, and MME:S $= 155$. Incidentally, the error of estimation of the two-stage jackknife procedure is marginally greater than that of MLE:S (the one-stage jackknife procedure) in these unstable cases; since this procedure is dominated in both the stable and unstable cases, it is eliminated from further consideration.

While the results presented in Table 1 are informative, such a coarse summary may mask many types of interactions. To investigate the nature of the dependence upon the parameters, the 2,000 cases were examined in detail by grouping the parameter ranges as follows: $0 \leq p \leq (\sqrt{2} - 1)/2 \doteq .207$, $.207 < p \leq \sqrt{2} - 1 \doteq .414$, $.414 < p \leq .7$, $.7 < p \leq 1.0$; $1 \leq n \leq 10$, $10 < n \leq 50$, $50 < n \leq 100$; $3 \leq k \leq 8$, $9 \leq k \leq 15$, $16 \leq k \leq 22$. The individual performance measures (and various transformations of these measures) were then regarded as outcomes in a $4 \times 3 \times 3$ factorial experiment and were subjected to various exploratory analyses, stable and unstable cases being considered separately. The more interesting results are reported here.

For stable cases, the fraction of the untied cases in which MME:S was closer to the true n than MLE:S exhibited a strong effect due to p (MME:S doing relatively better in those cases where $0 < p < \sqrt{2} - 1$), as well as a hint of possible $n \times k$ interactions. In the unstable cases, the effect due to p of this latter measure was similar in direction but considerably more exaggerated; there was also some evidence of an effect due to n (MME:S doing relatively better as n increased) and possible $n \times p$ interactions. Overall, MME:S appears to have the advantage over MLE:S in both the stable and unstable cases, except for those cases corresponding to the largest values of p, namely, $.7 < p < 1.0$.

The measure of the relative error of estimation was also examined in detail, separately for MME:S and MLE:S. For the stable cases, strong effects due to both k and p were evident (estimation error decreasing as k and p increased) for both MME:S and MLE:S; $k \times n$ interactions were also present in the case of MME:S. Even this detailed examination, however, did not reveal any substantial difference in performance (as measured by the relative error of estimation) of the two estimators in these stable cases. The situation in the unstable cases is somewhat more complex; strong effects due to both n and p were evident (estimation error decreasing as p increased and n decreased) for both MME:S and MLE:S. An effect due to k (estimation error being largest for the smallest values of k) was also apparent for MME:S; in addition, there was some evidence of interactions for both estimators. In these unstable cases, this measure of the relative error of estimation, in agreement with the counts of wins and ties, indicates that MME:S has a considerable advantage in those cases in which $0 \leq p \leq \sqrt{2} - 1$; on the other hand, it also indicates that MLE:S has a considerable advantage in those cases in which $\sqrt{2} - 1 < p \leq 1$.

5. CONCLUDING REMARKS

Our analysis of the stability of the conventional n estimators obtained by maximum likelihood and the method of moments reveals that, in both cases, the statistic $\hat{\mu}/\hat{\sigma}^2$ ($\equiv$ sample mean/sample variance) captures the essential nature of the data configuration in this particular problem; our criterion of stability, which was obtained by a somewhat different argument, depends only on this statistic.

Table 2. n Estimates for Selected and Perturbed Samples

n	Parameters p	k	Sample	Estimators MME	MME:S	MLE	MLE:S
75	.32	5	16,18,22,25,27	102	70	99	29
			16,18,22,25,28[a]	195	80	190	30
34	.57	4	14,18,20,26	507	77	504	31
				<0	91	∞	32
37	.17	20	4,4,4,4,5,5,5,5,6,6,6,6,7,9,9,10,10,10,11,11	65	25	66	11
				154	27	159	13
48	.06	15	0,1,1,2,2,2,3,3,3,4,4,4,4,5,6	18	10	15	7
				135	12	125	9
40	.17	12	6,7,7,7,8,8,9,9,9,10,11,16	32	26	40	21
				61	32	79	23
74	.68	12	40,42,42,43,44,48,49,52,53,53,54,61	210	153	201	67
				259	162	237	69
55	.48	20	17,23,24,25,25,26,26,26,27,27,28,28,28,29,30, 30,30,31,33,38	71	69	71	43
				79	74	81	45
60	.24	15	11,11,12,12,13,13,14,16,17,17,18,18,20,20,22	67	49	67	24
				88	53	90	28

[a] This is the perturbed sample obtained by adding one to the largest success count. For simplicity, the perturbed samples are not displayed in the remaining cases.

The results in Table 1 indicate that, for the parameter ranges considered, stable cases occur about 70 percent of the time when n, p, and k are generated at random. In terms of performance, there appears to be little to choose between MME:S (see 3.5) and MLE:S (see 3.10) in these cases; since MME:S is much simpler to compute, its use is recommended in stable cases. In the unstable cases, on the other hand, MME:S is a clear winner and therefore is again the recommended estimator.

A referee has pointed out that since $\hat{\mu}/\hat{\sigma}^2 \rightarrow q^{-1}$ as $k \rightarrow \infty$, our stability criterion (3.2) will classify a case as unstable whenever k is large and $p < .414$. Indeed, for large k, we may evaluate Prob $\{\hat{\mu}/\hat{\sigma} \geq a\} \cong$ Prob $\{\hat{\mu} \geq anpq(k-1)/k\} \cong 1 - \Phi(\sqrt{npqk}(a - q^{-1}))$. Since it would seem natural that for any fixed n and p a reasonable criterion of stability should be more easily satisfied for larger values of k, this crude calculation indicates that our criterion is not entirely satisfactory. Indeed, evidence of this deficiency for moderate values of k was discovered in our empirical study, although these results are not reported here. Since in any case the stability criterion is somewhat arbitrary, this suggests the possibility of exploring an alternative that is k dependent.

In spite of these reservations, we note that our methods have resulted in n estimators that are not only consistent but also stable. In Table 2 we present results for a few specific cases; for each case the estimates of n are also presented for the perturbed sample obtained by adding one to the largest success count of the original sample. Although these cases are not representative (indeed, these cases were selected from among the more difficult cases), the results indicate that our approach has provided an improvement on the conventional n estimators.

[Received February 1980. Revised January 1981.]

REFERENCES

ANSCOMBE, F.J. (1950), "Sampling Theory of the Negative Binomial and Logarithmic Series Distributions," *Biometrika*, 37, 358–382.

BINET, F.E. (1954), "The Fitting of the Positive Binomial Distribution When Both Parameters are Estimated From the Sample," *Annals of Eugenics*, 18, 117–119.

BONITZER, J. (1978), "Unicité de la Solution de l'Équation de Vraisemblance de la Loi Binomiale Négative," *Revue de Statistique Appliquée*, 26, 55–59.

COX, D.R., and HINKLEY, D.V. (1974), *Theoretical Statistics*, London: Chapman and Hall.

DEMPSTER, A.P., SCHATZOFF, M., and WERMUTH, N. (1977), "A Simulation Study of Alternatives to Ordinary Least Squares," *Journal of the American Statistical Association*, 72, 77–91.

DE RIGGI, D.F. (1978), "The Existence of Maximum Likelihood Estimates for Binomial Distributions," unpublished manuscript.

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

FISHER, R.A. (1941, 1942), "The Negative Binomial Distribution," *Annals of Eugenics*, 11, 182–187.

GRAY, H.L., and SCHUCANY, W.R. (1972), *The Generalized Jackknife Statistic*, New York: Marcel Dekker.

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.

LEVIN, B., and REEDS, J. (1977), "Compound Multinomial Likelihood Functions are Unimodal: Proof of a Conjecture of I.J. Good," *Annals of Statistics*, 5, 79–87.

MORAN, P.A. (1951), "A Mathematical Theory of Animal Trapping," *Biometrika*, 38, 307–311.

STUDENT (1918), "An Explanation of Deviations From Poisson's Law in Practice," *Biometrika*, 12, 211–215.