Estimation of Binomial Parameters when Both n, p are Unknown

A.DasGupta Herman Rubin

Purdue University

February 8,2004

ABSTRACT

We revisit the classic problem of estimation of the binomial parameters when both parameters n, p are unknown. We start with a series of results that illustrate the fundamental difficulties in the problem. Specifically, we establish lack of unbiased estimates for essentially any functions of just n or just p. We also quantify just how badly biased the sample maximum is as an estimator of n. Then we motivate and present two new estimators of n. One is a new moment estimate and the other is a bias correction of the sample maximum. Both are easy to motivate, compute, and jackknife. The second estimate frequently beats most common estimates of n in the simulations, including the Carroll-Lombard estimate. This estimate is very promising. We end with a family of estimates for p; a specific one from the family is compared to the presently common estimate $max\{1 - \frac{s^2}{\overline{X}}, 0\}$

and the improvements in mean squared error are often very significant. In all cases, the asymptotics are derived in one domain. Some other possible estimates such as a truncated MLE and empirical Bayes methods are briefly discussed.

1 Introduction

Estimation of the Binomial parameters when n, p are both unknown has remained a problem of some notoriety over half a century. Although estimation of p when n is known is the textbook problem, estimation of the n parameter with p too unknown has generated quite some literature. Classic literature on this problem includes Haldane(1941), Feldman and Fox(1968), Olkin,Petkau and Zidek(1981), Carroll and Lombard(1985), Lindsay(1985), Raftery(1988), and Hall(1994). The software reliability literature also includes a substantial body of work that relates to the binomial n problem when p is unknown; see Basu(2003), Basu and Ebrahimi(2001) for many references and Bayesian approaches, and Zacks et.al.(1990) for sequential approaches and further references. The problem continues to be important and attractive on at least four grounds : (i) it is known to be a fundamentally difficult problem, with underestimation of n being a serious practical handicap; (ii) easily computable and easily motivated estimates are still generally lacking, although the Carroll-Lombard(CL) estimate (1985), the best estimate available to date, has fine overall performance; (iii) the problem exhibits an inherent instability with the common estimates of n being vulnerable to massive fluc-

tuations under slight perturbations of one or two sample values; and (iv) the many interesting practical applications, ranging from species diversity estimation to error counting in software codes. The estimation of p when n is unknown is also an interesting problem, and indeed has produced less literature than the corresponding problem of estimating the n parameter. We address both problems in this article.

Bias manifested in the form of severe underestimation is one of the hardest features of the binomial n problem. In section 2, we establish the lack of unbiased estimates essentially for any nontrivial function of n or of p. In particular, neither n nor p are unbiasedly estimable. The result is worth recording due to its generality and for historical completeness. It is interesting that Fisher did *not* take the binomial n problem seriously. Fisher's argument was a formally correct one, but overly optimistic. Fisher argued that the binomial n problem is not a very interesting one because with k iid sample values $X_1, X_2, \ldots, X_k$, the sample maximum $X_{(k)} \stackrel{a.s.}{\to} n$, and hence, $X_{(k)} = n(a.s.)$ for all large k. While this is a technically correct statement, in section 3 we show the nearly complete practical irrelevance of this technical fact by giving an expression for the smallest value of k required

to make $P(X_{(k)} \geq m) \geq 1 - \alpha$, with $m, n \to \infty$ and $0 < \alpha < 1$ any fixed constant. For example, the smallest k such that $P(X_{(k)} \geq \frac{n}{2}) \geq .5$ is 31,500 when the true n is 100 and the true p is .3. Of course, we never have the luxury of such large sample sizes in practical studies, such as independent aerial surveys for counting species or employing readers to count software errors. It is true that it is well known that the sample maximum $X_{(k)}$ is not a reliable estimate of n. Our result goes to show just how unreliable it is. It also gives a strong hint that *other* estimates may also have a bias problem, which in fact is true of many other currently available estimates.

In section 4, we introduce two new estimates of n. The first estimate is a new moment estimate, but uses $X_{(k)}, \overline{X}$ and s^2, as opposed to just $\overline{X}$ and s^2. It is trivial to compute, easy to jackknife, easy to motivate, and simulations show that it is a fine estimate overall, with its best performance when p is small, and n is small or moderate. We also derive its asymptotic theory in one domain. The asymptotics show an interesting phenomenon; the variance of the asymptotic distribution does not involve the nuisance parameter p. This is an unexpected, and yet positive feature; for example, if we were to use the asymptotic theory to construct a confidence interval,

we can avoid estimating the nuisance parameter p. The second estimate is a bias correction of the sample maximum $X_{(k)}$. The bias correction is done in an unusual way. Using certain bounds in van Zwet(1967) on the expected values of Beta order statistics, we obtain a bound on the average bias of $X_{(k)}$, averaged over p. We then obtain an estimate for this bound on the average bias, and subtract it off from $X_{(k)}$ to produce a bias corrected estimate of n. In all the simulations we did, this estimate performs remarkably well, and beats even the Carroll-Lombard estimate under many configurations of k, n, p. We are quite excited by the performance of this estimate. Section 4 ends with explicit expressions for the jackknifed versions of each of these two estimates of n, and some further simulations of bias reduction after jackknife.

In section 5, we address estimation of p. It is obvious that estimation of n and estimation of p are linked together, in that if we had a good estimate of n, we could hope to find good estimates for p and vice versa. Thus, by using the new moment estimate of n presented in section 3, we give a two-parameter family of estimates for p in section 5. Again, these estimates all use $X_{(k)}, \overline{X}$ and s^2. We derive the asymptotics for these estimates as well. The variance function in the limiting distribution involves

the two free parameters in the definition of the estimates. The section ends with a discussion on choosing the two free parameters and some simulations to compare a specific estimate from the family with the popular estimate $max\{1 - \frac{s^2}{\overline{X}}, 0\}$.

To summarize, we quantify the severity of the bias problem by a collection of results. We present two new estimates of n. Both show promise, and in particular the estimate obtained as a bias-correction of $X_{(k)}$ seems to be a really promising one. We also give a new two-parameter family of estimates of p. And in all cases, we derive the asymptotics of the estimates in one domain, and examine their actual performance for various configurations of k, n, p by simulations. These are the main contributions made in this article.

2 Nonexistence of Unbiased Estimates

Although exact unbiasedness is not generally considered to be important in modern statistics, we provide two results on nonexistence of unbiased estimates to illustrate the difficulty of obtaining good estimates in this problem. In fact, as far as estimation of n is considered, severe underestimation is one of the most crippling phenomena in

the problem. Thus, the results on nonexistence of unbiased estimates highlight that aspect of the problem.

Theorem 1 Let $X_1, X_2, \ldots, X_k$ be iid observations from a $Bin(n, p)$ distribution, with n, p being both unknown, $n \geq 1, 0 < p < 1$.

a) If $g(n)$ is any nonconstant function of n, there does not exist an unbiased estimate for $g(n)$;

b) If $g(p)$ is any function of p such that $g'(p)$ exists in a neighborhood of $p = 0$ and such that $c = \lim_{p \to 0} g'(p)$ exists, and is finite and nonzero, then there does not exist an unbiased estimate for $g(p)$.

Proof a) Suppose an unbiased estimate $h(X_1, X_2, \ldots, X_k)$ exists. Then, $\forall n, p, E_{n,p} h(X_1, X_2, \ldots, X_k) = n$. We may assume without loss of generality that h is a permutation invariant function. Then, using $n = 1$,

$\forall p, E_{n=1,p} h(X_1, X_2, \ldots, X_k) = 1$

$\Rightarrow (1-p)^k h(0, 0, \ldots, 0) + kp(1-p)^{k-1} h(1, 0, \ldots, 0) + \ldots + p^k h(1, 1, \ldots, 1) = 1 \forall p$;

Taking a limit as $p \to 0$, one obtains $h(0, 0, \ldots, 0) = 1$. (1)

Next, using $n = 2$,

$\forall p, E_{n=2,p} h(X_1, X_2, \ldots, X_k) = 2$

$\Rightarrow \Sigma_{i=0}^{k} \Sigma_{j=0}^{k-i} \frac{k!}{i!j!(k-i-j)!}(1-p)^{2i}(2p(1-p))^j p^{2(k-i-j)} = 2 \forall p; \quad (2)$

Again, taking a limit as $p \to 0$, one obtains $h(0, 0, \ldots, 0) = 2$, thus giving a contradiction to (1).

b) We assume as in part a) that a permutation invariant unbiased estimate $h(X_1, X_2, \ldots, X_k)$ exists. Then, using $n = 1$,

$\forall p, h(0, 0, \ldots, 0)(1-p)^k + kh(0, 0, \ldots, 0, 1)p(1-p)^{k-1} + \ldots + h(1, 1, \ldots, 1)p^k = g(p)$;differentiating with respect to p and taking a limit as $p \to 0$, one obtains :

$$-kh(0, 0, \ldots, 0) + kh(0, 0, \ldots, 0, 1) = c. \qquad (3)$$

Next, taking $n = 2$, and differentiating the equation

$\Sigma_{i=0}^{k} \Sigma_{j=0}^{k-i} \frac{k!}{i!j!(k-i-j)!}(1-p)^{2i}(2p(1-p))^j p^{2(k-i-j)} = g(p)$ with respect to p, and finally taking a limit as $p \to 0$, one obtains :

-2kh(0,0,. . . ,0) + 2kh(0,0,. . . ,0,1) = c $\qquad$ (4),

which contradicts (3), as c was assumed to be nonzero and finite.

3 Estimation of n : Severity of Bias

The most profound difficulty in estimating the binomial n parameter when p is unknown is the severe underestimation of n, especially if either the true n is large, or the value of p is small. The underestimation is generally so drastic as to make many of the common estimates essentially useless. We illustrate the underestimation problem with the sample maximum $X_{(k)}$ as the estimate of n. Although it is generally known that $X_{(k)}$ underestimates n seriously, the results below provide some precise quantification of just how bad the bias is. It also suggests that other estimates may have a bias problem as well, if $X_{(k)}$ has such a severe problem.

First, we give a numerical example.

Example 1 This numerical example illustrates the serious difficulties in avoiding drastic underestimation of n for small or moderate values of p, and strikingly so when the true n is large.

Table 1: $E(X_{(k)}); k = 25$

	p		
n	.1	.2	.3
25	5.78	9.17	12.15
50	9.52	15.82	21.53
100	16.26	28.14	39.18
200	28.72	51.40	72.92

Thus, with p in the range of .1 to .3, with as many as 25 independent searches, the sample maximum typically estimates n as only about 20% to 40% of its true value.

The next Table gives the smallest value of k for which $P(X_{(k)} \geq \frac{n}{2}) \geq .5$.

Table 2: Smallest k such that $P(X_{(k)} \geq \frac{n}{2}) \geq .5$

	p		
n	.1	.2	.3
25	4.25×10^6	1880	40
50	6.85×10^{11}	331,000	293
100	—	3.25×10^{10}	31,500

The numbers in Table 2 are staggering and in no practical estimation problem in real life, so many independent samples would be available. Note also that the demand in Table 2 is very minimal; we only ask that the estimate is at least as large as 50% of the true value with only a 50% probability.

The following theorem quantifies the smallest k required to have $P(X_{(k)} \geq m) \geq 1 - \alpha$, for given α.

Theorem 2 Let $m, n \to \infty$ such that $\frac{n}{m} = t$ is an integer, and $\delta = \frac{1}{t} > p$. Let $0 < \alpha < 1$. Then the smallest k such that $P(X_{(k)} \geq m) \geq 1 - \alpha$ satisfies

$$k \geq \frac{(-\log \alpha)(\delta - p)}{p\sqrt{1-\delta}} \left(\frac{pq^{\frac{1}{\delta}-1}}{\delta(1-\delta)^{\frac{1}{\delta}-1}}\right)^{-m} \sqrt{2m\pi}(1 + o(1)) \quad (5).$$

For the proof of the theorem, we need the following lemma.

Lemma 1 Let $Y \sim Beta(n-m, m+1)$. Then, $P(Y \geq 1-p) \leq \frac{p(1-p)}{m-np} f_{n-m,m+1}(1-p)$, where $f_{u,v}$ denotes the density function of the $Beta(u,v)$ distribution.

The proof of this lemma can be seen in DasGupta(2000).

Proof Obviously, $P(X_{(k)} > m) \geq 1-\alpha \Leftrightarrow P(X_{(k)} \leq m) \leq \alpha \Leftrightarrow P(X_1 \leq m)^k \leq \alpha$, where $X_1 \sim Bin(n,p)$.

By a well known identity, $P(X_1 \leq m) = P(Y \leq 1-p) \geq 1 - \frac{p(1-p)}{m-np} f_{n-m,m+1}(1-p)$, where $Y \sim Beta(n-m, m+1)$, the last inequality a consequence of Lemma 1.

Now, $P(X_{(k)} \geq m) \geq 1-\alpha \Leftrightarrow k \geq \frac{-\log\alpha}{-\log P(X_1 \leq m)} \geq \frac{-\log\alpha}{-\log(1-\frac{p(1-p)}{m-np} f_{n-m,m+1}(1-p))} = \frac{-\log\alpha}{\frac{p(1-p)}{m-np} f_{n-m,m+1}(1-p)(1+o(1))}$. (6)

Using now the formula for the Beta density $f_{n-m,m+1}(1-p) = \frac{n! p^m (1-p)^{n-m-1}}{m!(n-m-1)!}$, Stirling's approximation, and writing q for $1-p$, we have from (6), after a page of algebra which we omit for brevity of space,

$\frac{p(1-p)}{m-np} f_{n-m,m+1}(1-p) = e^{-n} n^{n+\frac{1}{2}} \frac{n-m}{m-np} q^n (\frac{p}{q})^m \frac{1}{e^{-m} m^{m+\frac{1}{2}} e^{m-n} (n-m)^{n-m+\frac{1}{2}} \sqrt{2}}$

$o(1))$

$$=p\sqrt{1-\delta}q^n(\tfrac{p}{q})^m(\tfrac{1}{\delta}-1)^m\frac{1}{(\delta-p)\sqrt{2m\pi}(1-\delta)^m(1-\delta)^{m(\frac{1}{\delta}-1)}}(1+$$
$o(1))$,

, on using that $\frac{m}{n}=\delta$. (7)

¿From (7), the stated result in the theorem follows on some more algebra.

Remark Theorem 2 shows the exponential growth of the value of k required for $X_{(k)}$ to exceed m with a prescribed probability as $m \to \infty$ and gives a theoretical explanation for why the numbers in Table 2 are so large.

4 Two New Estimates of n

In this section, we propose two new estimates of n when p too is unknown. One of them is a new moment estimate and the other is an estimate obtained by bias correction of the sample maximum $X_{(k)}$. The moment estimate is easier to compute, has good overall performance, and is the better of the two when p is small, less than .1 or so. The second estimate is a bit more cumbersome to compute, requiring quantiles of Beta distributions for computation, but still much easier to compute than the Carroll-Lombard estimate, and has excellent performance for larger p.

4.1 A New Moment Estimate

First we motivate the estimate. Consider the identity $n = \frac{n^{\alpha+1}(npq)^{\alpha}}{(np)^{\alpha}(nq)^{\alpha}}$. Substituting the sample maximum $X_{(k)}$ for n, the sample variance s^2 for npq, the sample mean $\overline{X}$ for np, and $X_{(k)} - \overline{X}$ for $nq = n - np$, one has the estimate :

$$\hat{n}_1 = \frac{X_{(k)}^{\alpha+1}(s^2)^{\alpha}}{\overline{X}^{\alpha}(X_{(k)}-\overline{X})^{\alpha}}. \qquad (8)$$

This is the first of the two estimates, and is a moment estimate. We first derive its asymptotic distribution. It is interesting that $\hat{n}_1$ has an asymptotic distribution free of the nuisance parameter p, a surprising and positive property.

Theorem 3 For fixed n, as $k \to \infty$, $\sqrt{k}(\hat{n}_1 - n) \stackrel{\mathcal{L}}{\Rightarrow} N(0, 2\alpha^2 n(n-1))$.

Proof Since $X_{(k)} \to n$ in probability exponentially, by Slutsky's theorem, $\sqrt{k}(\hat{n}_1 - n)$ and $\sqrt{k}(T_k - n)$, where $T_k = \frac{n^{\alpha+1}(s^2)^{\alpha}}{\overline{X}^{\alpha}(n-\overline{X})^{\alpha}}$ have the same limiting distribution. We find the limiting distribution of $\sqrt{k}(T_k - n)$ by an application of the *delta theorem*. We mention below only the main steps.

The following follow from straightforward calculations

:

(a) $\frac{\partial T_k}{\partial \overline{X}} = -\alpha n^{\alpha+1}\overline{X}^{-\alpha-1}(s^2)^{\alpha}(n-2\overline{X})(n-\overline{X})^{-\alpha}$;

(b) $\frac{\partial T_k}{\partial s^2} = \alpha n^{\alpha+1}\overline{X}^{-\alpha}(n-\overline{X})^{-\alpha}(s^2)^{\alpha-1}$;

(c) with $\mu = np, \sigma^2 = npq$, $\frac{\partial T_k}{\partial \overline{X}}|_{\mu,\sigma^2} = -\frac{\alpha(1-2p)q^{\alpha}(1-p)^{-\alpha-1}}{p}$;

(d) $\frac{\partial T_k}{\partial s^2}|_{\mu,\sigma^2} = \frac{\alpha}{pq}$;

(e) $\mu_3 = E(X_1-\mu)^3 = np(2p^2-3p+1)$, and $\mu_4 = E(X_1-\mu)^4 = npq(1+3(n-2)p-3(n-2)p^2)$.

Since $\sqrt{k}[(\overline{X}, s^2)-(\mu,\sigma^2)] \stackrel{\mathcal{L}}{\Rightarrow} N_2(0,0,\sigma^2,\mu_4-\sigma^4,\mu_3)$, it follows from the expressions in (a)-(e) above by an application of the delta theorem that $\sqrt{k}(T_k-n) \stackrel{\mathcal{L}}{\Rightarrow} N(0,2\alpha^2 n(n-1))$, after some more algebra. We omit the algebra.

Remark The asymptotic variance is therefore a decreasing function of α. However, in fixed samples, the variance is not minimized by using $\alpha = 0$, and typically, the choice $\alpha = 1$ is a very good one in the range of (n,p) values we tried in the mean squared error simulations.

4.2 The Second Estimate

The results in Section 3 illustrate the uselessness of $X_{(k)}$ as an estimate of n due to its severe bias. We now propose an estimate of n obtained after a bias correction of $X_{(k)}$. The estimate performs surprisingly well, and beats the Carroll-Lombard estimate in our simulations for many combinations of (k, n, p). The estimate is motivated below.

$E(X_{(k)}) = \Sigma_{i=0}^{n-1} P(X_{(k)} > i) = \Sigma_{i=0}^{n-1} P_{i+1,n-i}(Y_{(1)} \leq p)$, where $Y_{(1)}$ denotes the minimum of an iid sample of size k from a $Beta(i+1, n-i)$ distribution; this relation between the binomial and the Beta distribution is well known.

Therefore,$E(X_{(k)}) = n - \Sigma_{i=0}^{n-1} P_{i+1,n-i}(Y_{(1)} > p)$, and hence,

$$\int_0^1 (E(X_{(k)}))dp = n - \int_0^1 P_{i+1,n-i}(Y_{(1)} > p)dp$$

$$= n - \Sigma_{i=0}^{n-1} E_{i+1,n-i}(Y_{(1)}). \qquad (10)$$

Now, by a result in van Zwet(1967), for $0 \leq i \leq n-2$, $E_{i+1,n-i}(Y_{(1)}) \leq F^{-1}_{i+1,n-i}(\frac{1}{k})$, and for $i = n-1, E_{i+1,n-i}(Y_{(1)}) \leq F^{-1}_{n,1}(\frac{1}{k+1}) = (\frac{1}{k+1})^{\frac{1}{n}}$, where $F^{-1}_{r,s}$ denotes the quantile function of the $Beta(r, s)$ distribution.

Substituting these bounds in (10), one obtains the following bound on the average bias of $X_{(k)}$:

$$\int_0^1 (E(X_{(k)}) - n)dp \geq -\Sigma_{i=0}^{n-2} F^{-1}_{i+1,n-i}(\tfrac{1}{k}) - (\tfrac{1}{k+1})^{\frac{1}{n}}. \quad (11)$$

Treating, as a heurism, the lower bound on the average bias as an equality, and ignoring the small number $(\frac{1}{k+1})^{\frac{1}{n}}$, a bias corrected estimate for n is obtained :

$$\hat{n}_2 = X_{(k)} + \Sigma_{i=0}^{\hat{n}-2} F^{-1}_{i+1,\hat{n}-i}(\tfrac{1}{k}),$$

where $\hat{n}$ is some suitable (preliminary) estimate of n. We use the integer part of our previously introduced estimate $\hat{n}_1$ as this preliminary estimate.

Thus, our second estimate of n is :

$$\hat{n}_2 = X_{(k)} + \Sigma_{i=0}^{[\hat{n}_1]-2} F^{-1}_{i+1,[\hat{n}_1]-i}(\tfrac{1}{k}), \quad (12)$$

where $[\hat{n}_1]$ is the integer part of $\hat{n}_1$.

4.3 Asymptotic Distribution of $\hat{n}_2$

Theorem 4 For fixed n, as $k \to \infty$, $(nk)^{\frac{1}{n-1}}(\hat{n}_2 - n) \stackrel{\mathcal{L}}{\Rightarrow} \delta_1$, where δ_1 denotes a point mass at 1.

Proof Let $A_k = \{[\hat{n}_1] = n\}, k \geq 1$. Then, on the set

A_k, $\hat{n}_2 = X_{(k)} + \Sigma_{i=0}^{n-2} F_{i+1,n-i}^{-1}(\frac{1}{k})$. Next, note that, for any $0 < x < 1$, and $i = 0, 1, \ldots, n-1$, by a simple calculation, $F_{i+1,n-i}(x) = \frac{n!}{i!(n-i-1)!} \Sigma_{j=0}^{n-i-1} (-1)^j \frac{(n-i-1)!}{j!(n-i-j-1)!} \frac{x^{i+j+1}}{i+j+1}$.

(13)

Hence, for $i = 0, 1, \ldots, n-1$, $F_{i+1,n-i}(x) = \frac{n!}{i!(n-i-1)!} \frac{x^{i+1}}{i+1}(1 + o(x))$, as $x \to 0$. This implies that for $i = 0, 1, \ldots, n-1$, $F_{i+1,n-i}^{-1}(\frac{1}{k}) = (\frac{n!}{(i+1)!(n-i-1)!} k)^{-\frac{1}{i+1}}(1+o(1))$, as $k \to \infty$. Summing over i, $(nk)^{\frac{1}{n-1}} \Sigma_{i=0}^{n-2} F_{i+1,n-i}^{-1}(\frac{1}{k}) = 1 + o(1)$, as $k \to \infty$. (14)

Thus, $P(|(nk)^{\frac{1}{n-1}}(\hat{n}_2 - n) - 1| > \epsilon)$

$\leq P(|(nk)^{\frac{1}{n-1}}(\hat{n}_2 - n) - 1| > \epsilon \cap A_k \cap \{X_{(k)} = n\}) + P(A_k^c) + P(X_{(k)} \neq n)$, and since $P(A_k^c) \to 0$ as $\hat{n}_1$ is a consistent estimate, and since evidently $P(X_{(k)} \neq n) \to 0$, the statement of the theorem follows from here by using (14).

Discussion For reasonably large n, the norming constant $(nk)^{\frac{1}{n-1}}$ would be approximately 1. So Theorem 4 says that $\hat{n_2}$ would vary slightly around $n + 1$. Recall, however, that the domain of asymptotics is when $k \to \infty$, with n fixed. The speed of convergence is slow. Thus, in reality, the behavior of $\hat{n_2}$ is not reflected in Theorem 4 unless $\frac{k}{n}$ is about 10, in our simulations that we did not

report here.

4.4 Some Other Possible Approaches

Of course, other approaches to estimation of the n parameter are possible, and we suspect that other reasonable estimates can be found. For brevity of space and purposes of focus, we only briefly mention a few other possible approaches, but do not investigate them here.

The MLE is always worth an investigation. Indeed, in the binomial n problem, the properties of the MLE have been studied quite extensively. It has some serious drawbacks; the drawbacks form a primary motivation for the many other estimates researchers have offered. It is very unstable, and need not be finite with probability 1. One possibility is to truncate the MLE at $X_{(k)} + c$ for some suitable positive constant c. When n is small, and p in the range of .3 or above, this type of an estimate is very good. It is not good for large n, in the range of 200, or if p is small. Another possibility, specifically directed towards the case of large n and small p, is to treat the problem as a Poisson one, and assign (perhaps independent) priors to $\lambda = np$ and p and estimate n treating it as $n = \frac{\lambda}{p}$. Empirical Bayes with such a formulation is apparently promising, but the only investigation seems

to be DasGupta, Haff and Strawderman(1998). It would be worth further investigation. Yet another possibility is to use appropriate *noninformative priors* for n, p and use posterior means or medians. A conceptual difficulty is that the n parameter is discrete, and so, for example, Jeffrey priors are not defined. We have some ideas on this, but do not report them here.

4.5 Comparative Performance of the Estimates

We present simulated mean squared errors and the bias of four estimates for some combinations of (k, n, p). The estimates are $X_{(k)}, \hat{n}_1, \hat{n}_2$, and the Carroll-Lombard(CL) estimate. We take values of p to be small to moderate; we do not simulate for larger values of p because the problem of estimating n is not that difficult unless p is small or moderate.

Table 3: Expected values of various estimates

n	p	k	$X_{(k)}$	$\hat{n}_1$	$\hat{n}_2$	C-L
20	.1	5	3.6	8.0	6.2	13.6
50	.1	10	8.47	18.93	16.58	13.85
50	.3	10	20.0	53.6	43.6	45.0
100	.1	15	15.44	39.58	31.23	27.9
100	.3	15	38.1	125.5	93.4	96.6
200	.1	25	28.8	85.49	65.84	55.5
200	.3	25	72.8	291.9	161.1	189.6

Discussion

For small p, the moment estimate $\hat{n_1}$ looks the best unless n is small, in which case the Carroll-Lombard estimate seems to have the least bias. Notice in particular how much better $\hat{n_1}$ does than the Carroll-Lombard estimate when n is large and p is small, arguably the most interesting case in practical applications. For larger p, the

Carroll-Lombard estimate has the least bias, although $\hat{n_2}$ is competitive. However, bias is only part of the assessment and we need to look at the full mean squared error. It is considered next.

Table 4: Mean Squared Error of Various Estimates

n	p	k	$X_{(k)}$	$\hat{n}_1$	$\hat{n}_2$	C-L
20	.1	5	269	176	201	47
50	.1	10	1727	1050	1145	1333
50	.3	10	903	348	161	216
100	.1	15	7154	3831	4785	5245
100	.3	15	3837	1941	386	878
200	.1	25	29313	13566	18122	20957.6
200	.3	25	16179	12534	1958	4233

Discussion

Quite interestingly, the mean squared errors present the same qualitative comparison as does the bias. For

small p, the moment estimate $\hat{n_1}$ is the best unless n is small, in which case the Carroll-Lombard estimate is the best. When p is not small, the bias-corrected estimate $\hat{n_2}$ is clearly the best; particularly impressive is by how much in these simulations $\hat{n_2}$ beats the Carroll-Lombard estimate for $n = 50, 100, 200$ and especially for the largest value $n = 200$.

The bias-corrected estimate $\hat{n_2}$ appears to be extremely promising and would be worth further investiagtion at other configurations of k, n, p.

4.6 Further Bias Reduction by Jackknife

Because the estimates $\hat{n}_1, \hat{n}_2$ both have explicit formulae, each estimate can be jackknifed with minimal computation in order to achieve some further reduction in bias. In fact, for both $\hat{n}_1, \hat{n}_2$, an explicit formula for the jackknifed version can be produced, which can be directly used for further use, e.g., simulation of bias and mean squared error. The expression for the jackknifed version is stated below. The derivation of this expression involves some straightforward but tedious calculation, and we omit it.

Proposition 1 The jackknifed version of $\hat{n}_1$ equals $\hat{n}_{1,J} = \frac{1}{k} \Sigma_{i=1}^{k} \hat{n}_{1,-i}$, where

$$\hat{n}_{1,-i} = \frac{\frac{k-1}{k-2}X_{(k)}^2[(k-1)^2s^2 - k(X_{(i)}-\overline{X})^2]}{(k\overline{X}-X_{(i)})[(k-1)X_{(k)}-k\overline{X}+X_{(i)}]}, \text{ for } i < k,$$

(15) and

$$\hat{n}_{1,-k} = \frac{\frac{k-1}{k-2}X_{(k-1)}^2[(k-1)^2s^2 - k(X_{(k)}-\overline{X})^2]}{(k\overline{X}-X_{(k)})[(k-1)X_{(k-1)}-k\overline{X}+X_{(k)}]}. \quad (16)$$

An explicit formula for the jackknifed version of $\hat{n}_2$ follows from the expression in Proposition 1 above.

Proposition 2 The jackknifed version of $\hat{n}_1$ equals $\hat{n}_{2,J} = \frac{1}{k}\Sigma_{i=1}^{k}\hat{n}_{2,-i}$, where

$$\hat{n}_{2,-i} = X_{(k)} + \Sigma_{j=0}^{\hat{n}_{1,-i}-2} F^{-1}_{j+1,\hat{n}_{1,-i}-j}(\tfrac{1}{k-1}), \text{ for } i < k, \quad (17) \text{ and}$$

$$\hat{n}_{2,-k} = X_{(k-1)} + \Sigma_{j=0}^{\hat{n}_{1,-k}-2} F^{-1}_{j+1,\hat{n}_{1,-k}-j}(\tfrac{1}{k-1}). \quad (18)$$

Some simulations on the amount of bias reduction due to jackknife is presented in the next table. The bias reduction due to jackknife in $\hat{n_1}$ is quite impressive. We noticed in our simulations that jackknife did not help for the other estimate $\hat{n_2}$; so we do not report the results on jackknifing for $\hat{n_2}$.

Table 5: Bias Before and After Jackknife

n	p	k	$\hat{n}_1$	$\hat{n}_{1,J}$	$\hat{n}_2$
200	.3	25	91.9	52.7	38.9
100	.3	15	24.5	10.4	6.6
50	.3	10	3.6	-1.6	6.4

5 Estimation of p

Using the one parameter family of estimates $\hat{n}_1$ for n, we propose a two parameter family of estimates for p. Again, first we motivate the estimates. For $0 \leq \lambda \leq 1$, write p as $p = \frac{(np)^{2\lambda-1}(np-npq)^{1-\lambda}}{n^{\lambda}}$. Substituting $\overline{X}$ for np, s^2 for npq, and $\hat{n}_1$ for n, on a little algebra, the following two parameter family of estimates for p is obtained :

$$\hat{p} = \frac{\overline{X}^{(\alpha+2)\lambda-1}(\overline{X}-s^2)^{1-\lambda}(X_{(k)}-\overline{X})^{\alpha\lambda}}{X_{(k)}^{(\alpha+1)\lambda}(s^2)^{\alpha\lambda}} \quad (19)$$

We present the asymptotic distribution of $\hat{p}$ below.

Theorem 5 For fixed n, as $k \to \infty$, $\sqrt{k}(\hat{p} - p) \stackrel{\mathcal{L}}{\Rightarrow} N(0, \tau^2(\alpha, \lambda))$, where

$\tau^2(\alpha,\lambda) = n^{2(\alpha\lambda-1)}p^{2(\alpha\lambda-\frac{1}{2})}[q(2-2nq-3p)+4(n-1)q\lambda((\alpha+1)p-1)+2(n-1)\lambda^2((\alpha+1)p-1)^2]$

(20).

Proof The proof is similar to that of Theorem 3. $\sqrt{k}(\hat{p}-p)$ and $\sqrt{k}(T_k^*-p)$, where $T_k^* = \frac{\overline{X}^{(\alpha+2)\lambda-1}(\overline{X}-s^2)^{1-\lambda}(n-\overline{X})^{\alpha\lambda}}{n^{(\alpha+1)\lambda}(s^2)^{\alpha\lambda}}$ have the same limiting distribution. The limiting distribution of $\sqrt{k}(T_k^* - p)$ follows from the joint bivariate normal limiting distribution for $(\overline{X}, s^2)$ by applying the delta theorem, as in Theorem 3. The intermediate algebra is messy but straightforward, which we omit.

Remark The free parameters α, λ are to be chosen. There are clearly no uniformly best choices for all n, p. In our simulations, $\alpha, \lambda \approx 1$ gave about the best values for bias and mean squared error for various combinations of n, p, k. We report a small simulation for the mean squared error in the next table using $\alpha = \lambda = 1$; for some meaningful assessment, we compare its mean squared error and bias with that of the presently common moment estimate of p given by $\tilde{\mathrm{p}} = max\{1 - \frac{s^2}{\overline{X}}, 0\}$. A discussion will follow suit.

Table 6: Mean Squared Error and Bias of $\hat{p}$ and $\tilde{p}$

n	p	k	$E(\hat{p}$	$E(\tilde{p})$	MSE of $\hat{p}$	MSE of $\tilde{p}$
50	.1	10	.322	.226	.077	.073
50	.3	10	.311	.351	.017	.064
100	.1	15	.280	.204	.042	.054
100	.3	15	.262	.334	.008	.049
200	.1	25	.245	.170	.025	.034
200	.3	25	.219	.298	.009	.033

Discussion

It seems that the mean squared error of the new estimate $\hat{p}$ we propose is always smaller than that of the moment estimate $\tilde{p}$; the mean squared error of $\hat{p}$ is many factors of magnitude smaller for the larger true value of p, namely $p = .3$. This is indeed encouraging.

As regards bias, generally the moment estimate $\tilde{p}$ is better, with a couple of exceptions in the simulation pre-

sented above. Thus, it seems that while the moment estimate $\tilde{p}$ has a smaller bias, it suffers from a much larger variance than our new proposed estimate $\hat{p}$. But we must be cautious as the simulation is limited; however, the findings as far as these simulations are quite promising.

Acknowledgement. We would like to thank Soumen Lahiri for his gracious help with the second estimate. Nels Grevstad and Tonglin Zhang verified some of our simulations. Shelley Zacks, as always, provided a wealth of ideas and information, for which we are, as usual, most grateful.

References

Basu,S.(2003). Bayesian inference for the number of undetected errors, Statistics in Industry, Handbook of Statistics, 22, 1131-1150, North-Holland, Amsterdam.

Basu,S. and Ebrahimi,N.(2001). Bayesian capture-recapture methods for error detection and estimation of population size : heterogeneity and dependence, Biometrika, 88, 1, 269-279.

Carroll,R.J. and Lombard,F.(1985). A note on n estimators for the Binomial distribution, Jour. Amer. Statist. Assoc., 80, 390, 423-426.

DasGupta,A.(2000). Best constants in Chebyshev inequalities with various applications, Metrika, 51, 3, 185-200.

DasGupta,A.,Haff.L.R., and Strawderman,W.E.(1998). Empirical Bayes estimation of binomial parameters when both n,p are unknown, Technical Report, Purdue University.

Feldman,D. and Fox,M.(1968). Estimation of the parameter n in the binomial distribution, Jour. Amer. Statist. Assoc., 63, 150-158.

Haldane,J.B.S.(1941). The fitting of binomial distributions, Ann.Eugenics, 11, 179-181.

Hall,P.(1994). On the erratic behavior of estimators of N in the binomial(n,p) distribution, Jour. Amer. Statist. Assoc., 89, 425, 344-352.

Lindsay, B.(1985). Errors in inspection : Integer parameter maximum likelihood in finite populations, Jour. Amer. Statist. Assoc., 80, 392, 879-885.

Olkin,I., Petkau,A.J. and Zidek,J.V.(1981). A comparison of n estimators for the binomial distribution, Jour. Amer. Statist. Assoc., 76, 375, 637-642.

Raftery,A.(1988). Inference for the Binomial N parameter: A hierarchical Bayes approach, Biometrika, 75, 223-228.

van Zwet,W.R.(1967). An inequality for expected values of sample quantiles, Ann. Math. Statist., 38, 1817-1821.

Zacks,S., Pereira,C.A., and Leite,J.G.(1990). Bayes sequential estimation of the size of a finite population, Jour. Stat.Planning and Inf., 25, 3, 363-380.