The Annals of Statistics
1984, Vol. 12, No. 3, 1109–1116

RESULTS ON DOUBLE SAMPLE ESTIMATION FOR THE BINOMIAL DISTRIBUTION[1,2]

BY ARTHUR COHEN AND HAROLD B. SACKROWITZ

Rutgers University

A complete class result is obtained for double sample estimation of a binomial parameter. The complete class leads to some inadmissibility results for some procedures that are likely to be used in practice. It is also shown that the sample proportion as a true double estimator is not unbiased.

1. Introduction and summary. In a companion paper, Cohen and Sackrowitz (1984a), Bayes double sample estimation procedures are obtained for the binomial distribution. The model in this paper is the same as the model in Section 3 of that paper: namely, X_i, $i = 1, 2, \cdots$ are i.i.d. Bernoulli variables with parameter p, $0 \le p \le 1$. Results of this paper are as follows:

(1.1) A complete class result is obtained for double sample estimation and sequential estimation. The result is that the Bayes terminal decision (estimator) is unique. This sharply contrasts with the fixed sample estimation problem where Bayes procedures are not always unique.

(1.2) The complete class result can be applied to prove that if the loss is a linear combination of squared error and cost of sampling then any estimation procedure (with its sampling rule) where the sample proportion is the terminal decision is inadmissible except for the case where the total sample is a single observation. Thus the inadmissibility result is true for any fixed sample procedure in the sequential setting as long as the sample size is at least two. This is a very surprising finding which contrasts sharply with the fixed sample situation.

(1.3) If the terminal decision loss function is squared error divided by $p(1 - p)$ and the overall loss is a linear combination, then any double sample procedure ($n_1 > 0$, n_2 not always constant) in which the estimate is the sample proportion is inadmissible. These last two results demonstrate the inadmissibility of the Miller-Freund (MF) procedure for various linear combination loss functions. See Cohen and Sackrowitz (1984a) Section 3 for the description of the Miller-Freund procedure.

(1.4) The sample proportion as a true double sample estimate is not unbiased.

Received March 1984.

[1] Research supported by NSF-MCS-81-18564-02.

[2] This paper was delivered at a Statistical Research Conference dedicated to the memory of Jack Kiefer and Jacob Wolfowitz. The conference was held at Cornell University, July 6–9, 1983, and was a special meeting of the Institute of Mathematical Statistics, cosponsored by the American Statistical Association. Financial support was provided by the National Science Foundation, the office of Naval Research, the Army Research Office–Durham, and units of Cornell University.

AMS 1980 *subject classifications*. Primary 62L12; secondary 62C05.

Key words and phrases. Double sample, linear combination loss function, vector loss function, complete class, admissibility, unbiasedness.

Section 2 contains complete class and admissibility results. Section 3 contains results on the Miller-Freund procedure and unbiasedness.

2. Complete class theorem and admisibility results for binomial case.

In this section we prove an interesting complete class result that is true for double samples and is also true for sequential estimation. We'll treat this latter case when the loss function is

$$L(p, a) = W(|\tau - p|) + c(n), \tag{2.1}$$

with $W(0) = 0$, $W(\cdot)$ strictly convex, $C(0) = 0$, $C(n)$ strictly increasing, $\lim_{n\to\infty} C(n) = \infty$. Since the parameter space is compact, the Bayes procedures form a complete class. For the fixed sample problem of estimating p with loss function $W(|\tau - p|)$, Bayes estimators with priors putting all mass on $p = 0$ and $p = 1$ are not unique. Brown (1981) gives conditions under which such Bayes estimators are admissible. In the sequential problem here, no matter what the prior distribution is, the Bayes estimator will always be an admissible terminal decision. Bayes estimators (terminal decision) are unique for any specified sample size dictated by the Bayes stopping rule.

Let ξ be a prior probability distribution, $\delta_\xi = (\tau_\xi, \phi_\xi)$ a Bayes procedure with respect to ξ where τ_ξ is the Bayes terminal decision and ϕ_ξ the Bayes stopping rule. We prove

THEOREM 2.1. *For every ξ, there exists a ϕ_ξ such that (τ_ξ, ϕ_ξ) is admissible.*

PROOF. If ξ puts any mass on the open interval (0, 1), clearly τ_ξ is unique and the conclusion is obvious. If ξ puts all its mass at 0 the Bayes rule is to take 0 observations and estimate p by 0. If ξ puts all its mass at 1, the Bayes rule is to take 0 observations and estimate p by 1. Now suppose ξ puts mass π at 0 and mass $(1 - \pi)$ at 1. The expected risk with respect to such a prior when no observations are taken is

$$\{\pi W(|\tau^*|) + (1 - \pi) W(|\tau^* - 1|)\} \tag{2.2}$$

where τ^* is the unique value of a which minimizes the expected risk when no observations are taken. The expected risk when one observation is taken is $C(1)$. To see this, note that the Bayes terminal decision is to estimate p by 0 if the single observation is 0 and estimate p by 1 if the single observation is 1. Thus the risk is

$$W(|p|)(1 - p) + W(|1 - p|)p + C(1) \tag{2.3}$$

and hence the expected risk is $C(1)$. Since $C(n)$ is strictly increasing, the Bayes procedure cannot take more than 1 observation. In fact the Bayes procedure is to take 0 observations if (2.2) $< C(1)$, and take 1 observation if (2.2) $> C(1)$. If (2.2) $= C(1)$ then one can randomize. Hence the Bayes procedure is unique for these priors except if (2.2) $= C(1)$. In this latter case, the procedure with no observations is easily seen to be admissible since its risk is zero at $p = \tau^*$.

REMARK 2.2. In the fixed sample problem, priors which put mass only at $\{0\}$ and $\{1\}$ yield many inadmissible Bayes procedures. In the sequential or k stage problem, this does not happen.

An interesting application of Theorem 2.1 can be made for the case where (2.1) becomes

$$L(p, a) = (\tau - p)^2 + C(n). \tag{2.4}$$

Let $y_n = \sum_{i=1}^n X_i$.

THEOREM 2.3. *Any two stage procedure with either $n_1 > 0$, n_2 not constant or $n_1 > 1$, and estimator $\hat{p}_n = y_n/n$ is inadmissible.*

PROOF. It is known that the only proper prior for which $\hat{p}_n$ can be a Bayes terminal decision is the prior which puts mass only on $\{0\}$ and $\{1\}$. In the proof of Theorem 2.1 we saw that the Bayes procedure with respect to such a prior takes at most one observation. Hence the procedure with terminal decision $\hat{p}$ can only be Bayes if $n_1 = 1$ and $n_2 = 0$. Since the Bayes procedures are a complete class, the theorem follows. □

REMARK 2.4. Theorem 2.3 applies to any sequential procedure where the sample size exceeds 1 with positive probability. This includes the somewhat surprising result that any fixed sample size procedure with $n > 1$ and estimator $\hat{p}_n$ is inadmissible! For a hypothesis testing problem, a comparable type of result was obtained by Brown, Cohen, and Strawderman (1980). It is of further interest to consider the fixed sample size procedure with $n = 2$, estimator $\hat{p}_2$, and cost function $C(n) = cn$. If $c > 1/8$ the procedure taking one observation and estimating $\hat{p}_1$ is better. If $1/64 \le c \le 1/8$, the procedure is better that takes one observation and estimates by $k\sqrt{c}$ if $X_1 = 0$ and estimates by $1 - k\sqrt{c}$ if $X_1 = 1$, for any k such that $(1/8\sqrt{c}) \le k \le 1$. However, if $c < 1/64$ there is no procedure which takes exactly one observation that is better than the procedure with $n = 2$ and estimator $\hat{p}_2$. Nevertheless this latter procedure is inadmissible.

3. Properties of Miller-Freund procedure.

In this section we study MF procedures regarding the optimality properties of admissibility and unbiasedness. Theorem 2.3 implies that the MF procedure is inadmissible for linear combination loss functions with squared error loss for the terminal decision. Theorem 4.2 below will imply inadmissibility for the linear combination loss function with $(\tau - p)^2$ replaced by $(\tau - p)^2/p(1 - p)$. This is of interest since in the fixed sample case the sample proportion is proper Bayes for such a loss function. Note that in this case the Bayes terminal decision with respect to prior ξ is

$$\tau_\xi(y_1, y_2) = \frac{\int_0^1 p^y(1-p)^{n-y-1}\, d_\xi(p)}{\int_0^1 p^{y-1}(1-p)^{n-y-1}\, d_\xi(p)} = \frac{\sum_{i=y}^{n-1} (-1)^i \binom{n-y-1}{i-y} M_i(\xi)}{\sum_{i=y}^{n-1} (-1)\binom{n-y-1}{i-y} M_{i-1}(\xi)} \tag{3.1}$$

whenever $y = y_1 + y_2$ satisfies $0 < y < n$, $n = n_1 + n_2$, and $M_i(\xi)$ denotes the ith moment of ξ. For $y = 0$ or n, if ξ puts mass in a neighborhood of 0 and 1,

$\tau_\xi(y_1, y_2) = 0$ if $y = 0$ and $\tau_\xi(y_1, y_2) = 1$ if $y = n$. If ξ does not put mass in a neighborhood of 0 and/or 1, then τ_ξ in (3.1) is correct even if $y = 0$ and/or n.

The integer m will be called a strong number for the double sampling procedure δ if the probability that the total sample size (using this procedure) equals m is greater than zero. Let S_δ denote the set of strong numbers for the procedure δ.

It is clear from (3.1) that if N_ξ is the largest strong number for the Bayes procedure versus the prior distribution ξ, then any prior distribution having the same first $N_\xi - 1$ moments as ξ will yield the same terminal decision rule. The converse is also true.

LEMMA 3.1. *Let T_ξ be the Bayes terminal decision rule for the prior distribution ξ. If n is a strong number, then T_ξ determines the first $(n-1)$ moments of ξ.*

PROOF. The method of proof is to show that the system of equations in (3.1) has at most one solution in $M_1(\xi), \cdots, M_{n-1}(\xi)$ for any fixed $n_1, n_2(0), n_2(1), \cdots, n_2(n_1)$. To this end we define

$$B_j^1 = M_j(\xi) \qquad j = 0, 1, \cdots$$
$$B_j^k = B_j^{k-1} - B_{j+1}^{k-1}, \quad j = 0, 1, \cdots, \quad k = 2, 3, \cdots. \tag{3.2}$$

Note that

$$B_j^k = \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} M_{j+i}(\xi) \tag{3.3}$$

and so (3.1) may be rewritten as

$$T_\xi(y_1, y_2) = \frac{\sum_{j=y_1+y_2}^{y_1+n_2} (-1)^j \binom{n_2-y_2}{j-y} B_j^{n_1-y_1}}{\sum_{j=y_1+y_2}^{y_1-n_1} (-1)^j \binom{n_2-y_2}{j-y} B_{j-1}^{n_1-y_1}}, \quad y_1 = 0, 1, \cdots, n_1 \tag{3.4}$$

and

$$T_\xi(n_1, y_2) = \frac{\sum_{j=y_1+y_2}^{n_1+n_2-1} (-1)^j \binom{n_2-y_2-1}{j-y} B_j^1}{\sum_{j=y_1+y_2}^{n_1+n_2-1} (-1)^j \binom{n_2-y_2-1}{j-y} B_{j-1}^1} \quad \text{for} \quad y_1 = n_1.$$

For each fixed $y_1 = 0, 1, \cdots, n_1$, call the set of equations generated by (3.4) for $y_2 = 0, 1, \cdots, n_2$ the y_1th set of equations. (Recall from (3.1) that we exclude the equations for $y_1 = y_2 = 0$ and $y_1 = n$, $y_2 = n_2$). The y_1th set of equations permits us to obtain each $B_j^{n_1-y_1}$, $j = y_1, \cdots, y_1 + n_2$, in terms (in fact as a multiple) of $B_{y_1-1}^{n_1-y_1}$. Thus we will have a system of equations of the form

$$B_j^k = a_{jr} B_r^k, \quad j = r+1, \cdots, r+n_2$$

for $r = 0, \cdots, n_1 - 1$, $k = n_1 - 1 - r$. This with the system of equations in (3.2) guarantees the uniqueness of $M_1(\xi), \cdots, M_{n_1+n_2-1}(\xi)$.

Let $c^{-1/2}$ be an integer. The following theorem indicates the implications the set of strong numbers has on the terminal decision rule y/n.

THEOREM 3.2. *Let δ be a procedure whose terminal decision rule is y/n and for which S_δ contains more than one non-zero element. Then δ is not a Bayes procedure.*

PROOF. Let N be an element of S_δ and say δ is Bayes with respect to ξ^*. Since equation (3.1) is satisfied for $\tau(y_1, y_2) = y/n$ and ξ is taken to be uniform on [0, 1], it follows from Lemma 3.1 that ξ^* and the uniform distribution must agree in at least the first $(N - 1)$ moments. Thus $M_i(\xi^*) = (i + 1)^{-1}$, $i = 1, 2, \cdots, N$.

We now consider the posterior terminal decision risk, $R_{\xi^*}(y_1)$, given $X_1 = y_1$ for such an estimator. That is

$$\begin{aligned} R_{\xi^*}(y_1) &= E[(y/n - p)^2 p^{-1}(1 - p)^{-1} \mid y_1] \\ &= E[E[(y/n - p)^2 p^{-1}(1 - p)^{-1} \mid y_1, y_2] \mid y_1] \\ &= \sum_{y_2=0}^{n_2} \frac{\int_0^1 (y/n - p)^2 p^{y-1}(1 - p)^{n-y-1}\, d\xi^*(p)}{\int_0^1 p^y (1 - p)^{n-y}\, d\xi^*(p)} \\ &\qquad \cdot \frac{\int_0^1 \binom{n_2}{y_2} p^y (1 - p)^{n-y}\, d\xi^*(p)}{\int_0^1 p^{y_1}(1 - p)^{n_1-y_1}\, d\xi^*(p)} \\ &= \frac{\sum_{y_2=0}^{n_2} \binom{n_2}{y_2} \int_0^1 (y/n - p)^2 p^{y-1}(1 - p)^{n-y-1}\, d\xi^*(p)}{\int_0^1 p^{y_1}(1 - p)^{n_1-y_1}\, d\xi^*(p)} . \end{aligned} \tag{3.5}$$

For any $n \leq N$ and letting ξ_0 denote the uniform distribution, it follows from (3.5) and equality of the first $(n - 1)$ moments that

$$\begin{aligned} &R_{\xi^*}(y_1) \int_0^1 p^{y_1}(1 - p)^{n_1-y_1}\, d\xi^*(p) \\ &= R_{\xi^*}(y_1) \int_0^1 p^{y_1}(1 - p)^{n_1-y_1}\, d\xi_0(p) \\ &= \sum_{y_2=0}^{n_2} \binom{n_2}{y_2} \left\{ \left[\int_0^1 \left(\frac{y}{n} - p\right)^2 p^{y-1}(1 - p)^{n-y-1}\, d\xi^*(p) - (-1)^{n-y-1} \right.\right. \\ &\qquad \left.\left. \cdot \int_0^1 p^n\, d\xi^*(p) \right] + (-1)^{n-y-1} \int_0^1 p^n\, d\xi^*(p) \right\} \\ &= \sum_{y_2=0}^{n_2} \binom{n_2}{y_2} \left\{ \int_0^1 \left(\frac{y}{n} - p\right)^2 p^{y-1}(1 - p)^{n-y-1}\, d\xi_0(p) \right. \\ &\qquad \left. + (-1)^{n-y-1} \left(\int_0^1 p^n\, d\xi^*(p) - \int_0^1 p^n\, d\xi_0(p) \right) \right\} \end{aligned} \tag{3.6}$$

as the expression in square brackets depends only on the first $n - 1$ moments of ξ^*. Thus

$$\begin{aligned} R_{\xi^*}(y_1) &= R_{\xi_0}(y_1) + (-1)^{n_1-y_1-1} \frac{M_n(\xi^*) - M_n(\xi_0)}{\int_0^1 p^{y_1}(1 - p)^{n_1-y_1}\, d\xi_0(p)} \sum_{y_2=0}^{n_2} \binom{n_2}{y_2} (-1)^{n_2-y_2} \\ &= R_{\xi_0}(y_1) = \frac{1}{n} . \end{aligned}$$

Therefore the posterior risk given $Y_1 = y_1$ in taking an additional $n_2 = n_2(y_1)$

observations is

$$(n_1 + n_2)^{-1} + c(n_1 + n_2)$$

independently of y_1. Thus S_δ can contain only one element $N = c^{-1/2}$. □

Theorem 3.2 implies that the Miller-Freund procedure cannot be admissible for the loss functions $(\tau - p)^2 + c(n_1 + n_2)$ even when the loss due to terminal decision is modified to $(\tau - p)^2/p(1 - p)$. The same conclusion is true if the loss function is $(\tau - p)^2 + c \log(n_1 + n_2)$.

There is one very special formulation for which the Miller-Freund procedure is admissible. Such a formulation requires the loss vector

$$((\tau - p)^2/p(1 - p), c \log(n_1 + n_2)), \tag{3.7}$$

and also must limit procedures whose first sample size is n_1, the first sample size (whatever it may be) of the MF procedure.

THEOREM 3.3. *Among all procedures whose first sample size is* n_1, *the* MF-n_1, $\Delta = c(n_1 + 1)/n_1$, *is admissible for the loss vector* (3.7).

PROOF. Cohen and Sackrowitz (1984b) show that a decision theory problem whose admissible rules are the same as those for the given vector loss problem can be formulated by introducing the dummy parameter ν as follows. Let $\nu = 1$, 2. Define

$$L((p, \nu); a) = \begin{cases} (\tau - p)^2/p(1 - p) & \text{if } \nu = 1 \\ c \log n & \text{if } \nu = 2. \end{cases} \tag{3.8}$$

In this setting it makes sense to speak of Bayes rules (with respect to priors on (p, ν)). We consider the prior distribution for which the conditional distribution of p given $\nu = 1$ is uniform on [0, 1] while the conditional distribution of p given $\nu = 2$ is Beta (α, α), where $B(\cdot, \cdot)$ denotes the Beta function. Also let $P(\nu = 2) = B(\alpha, \alpha)/(1 + B(\alpha, \alpha)) = 1 - P(\nu = 1)$. It is easy to see that for this prior the terminal decision rule for any fixed $n_1 + n_2(y_1)$ is $\hat{p} = (y_1 + y_2)/(n_1 + n_2(y_1))$. The posterior risk given $Y_1 = y_1$ is then a multiple (which depends only on y_1) of

$$\begin{aligned}(n_1 + n_1(y_1))^{-1} &\int_0^1 \theta^{y_1}(1 - \theta)^{n_1 - y_1}\, d\theta \\ &+ c \log(n_1 + n_2(y_1)) \int_0^1 \theta^{y_1 + \alpha - 1}(1 - \theta)^{n_1 - y_1 + \alpha - 1}\, d\theta.\end{aligned} \tag{3.9}$$

The value of $n_2(y_1) \geq 0$ which minimizes (3.9) is

$$n_2(y_1) = \max\left(\frac{\Gamma(y_1 + 1)\Gamma(n_1 - y_1 + 1)\Gamma(n_1 + 2\alpha)}{c\Gamma(n_1 + 2)\Gamma(y_1 + \alpha)\Gamma(n_1 - y_1 + \alpha)}, n_1\right) - n_1. \tag{3.10}$$

Taking $\alpha \to 0$, we obtain

$$n_2(y_1) = \max\left(\frac{y_1(n_1 - y_1)}{c(n_1 + 1)n_1}, n_1\right) - n_1 = \left[\left[\frac{n_1}{c(n_1 + 1)} \hat{p}_1(1 - \hat{p}_1) - n_1\right]\right]^+.$$

See Cohen and Sackrowitz (1984a), Section 2 for the definition of $[[\cdot]]^+$. Thus MF-n_1, Δ is seen to be the limit of a sequence of Bayes procedures for loss function (3.8) among the class of procedures with fixed n_1. Admissibility can be established by using a Blyth type argument. (See for example, Ferguson (1967), page 141.)

REMARK 3.4. If one does not limit the class of procedures to those whose first sample is n_1 then MF-n_1, Δ will not be a limit of the sequence of Bayes procedures given in the proof of Theorem 3.3. The sequence of priors there is eventually putting all the mass on the cost of the observations part of the problem and so a limit of such Bayes procedures would not take more than one observation.

Another result is concerned with the question of whether the sample proportion as an estimator, resulting from a true double sample procedure, is unbiased. The answer is no.

THEOREM 3.5. *Let δ be a true double sample procedure with estimator $\hat{p} = [(y_1 + y_2)/(n_1 + n_2(y_1))]$. The estimator is not unbiased.*

PROOF. Suppose $\hat{p}$ is unbiased. Then letting $\hat{p}_1 = y_1/n_1$, $\hat{p}_2 = y_2/n_2(y_1)$ we have

$$\begin{aligned} p &= \frac{E(Y_1 + Y_2)}{n_1 + n_2(Y_1)} \\ &= E\left(\frac{n_1 p_1}{n_1 + n_2(Y_1)}\right) + E\left(\frac{n_2(Y_1)p_2}{n_1 + n_2(Y_1)}\right) \qquad (3.11) \\ &= E\left(\frac{n_1 \hat{p}_1}{n_1 + n_2(Y_1)}\right) + pE\left(\frac{n_2(Y_1)}{n_1 + n_2(Y_1)}\right). \end{aligned}$$

Let $n(Y_1) = n_1 + n_2(Y_1)$. (Note below we need $n(-1)$ so define $n(-1) = 1$). Thus (3.11) becomes

$$(3.12) \quad 0 = E\left(\frac{Y_1 - n_1 p}{n(Y_1)}\right) = E\left(\frac{Y_1(1 - p) - (n_1 - Y_1)p}{n(Y_1)}\right), \quad \text{all} \quad 0 < p < 1.$$

Some algebra shows that for all $0 < p < 1$

$$E(p(n_1 - Y_1)/n(Y_1)) = E((1 - p)Y_1/n(Y_1 - 1)).$$

Thus (3.12) becomes

$$E_\xi(Y_1(n_1^{-1}(Y_1) - n_1^{-1}(Y_1 - 1))) = 0 \quad \text{all} \quad 0 < p < 1$$

which by completeness of the binomial distribution implies $n_1(y_1)$ is constant.

REMARK 3.6. Consider the loss functions $[(\tau - p)^2/p(1-p)] + c(n_1 + n_2)$, and $[(\tau - p)^2/p(1-p)] + c\log(n_1 + n_2)$. Suppose only nonrandomized procedures are permitted. Then it can be shown that the only admissible unbiased estimator resulting from a double sample procedure is a single sample procedure. This can be established using the information inequality for sequential estimators. See Wetherill (1966), page 134. The analogue of this fact will be true for estimating the expectation parameter for all one-dimensional exponential families satisfying the appropriate regularity conditions.

REFERENCES

BROWN, L. D. (1981). A complete class theorem for statistical problems with finite sample spaces. *Ann. Statist.* **9** 1289–1300.

BROWN, L. D., COHEN, A., and STRAWDERMAN, W. E. (1980). Complete classes for sequential tests of hypotheses. *Ann. Statist.* **8** 377–398.

COHEN, A. and SACKROWITZ, H. B. (1984a). Bayes double sample estimation procedures. *Ann. Statist.* **12** 1035–1049.

COHEN, A. and SACKROWITZ, H. B. (1984b). Decision theory results for vector risks with applications. *Statistics and Decisions*, to appear.

FERGUSON, T. S. (1967). *Mathematical Statistics.* Academic, New York.

MILLER, I. and FREUND, J. E. (1977). *Probability and Statistics for Engineers*, Second Ed. Prentice-Hall, Englewood Cliffs, New Jersey.

WETHERILL, G. B. (1966). *Sequential Methods in Statistics.* Methuen, London.

DEPARTMENT OF STATISTICS
HILL CENTER-BUSCH CAMPUS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903.