The Canadian Journal of Statistics
Vol. 20, No. 4, 1992, Pages 353–358
La Revue Canadienne de Statistique

On the admissibility of the maximum-likelihood estimator of the binomial variance

Lawrence D. BROWN, Mosuk CHOW and Duncan K.H. FONG

Cornell University, Northeastern University and *The Pennsylvania State University*

Key words and phrases: Admissibility, binomial distributions, maximum-likelihood estimator, squared-error loss, variance.
AMS 1985 subject classifications: Primary 62C15; secondary 62F10, 62F11.

ABSTRACT

This paper addresses the admissibility of the maximum-likelihood estimator (MLE) of the variance of a binomial distribution with parameters n and p under squared-error loss. We show that the MLE is admissible for $n \leq 5$ and inadmissible for $n \geq 6$.

RÉSUMÉ

Cet article concerne l'admissibilité, sous la fonction de perte quadratique, de l'estimateur à vraisemblance maximale de la variance d'une loi binomiale de paramètres n et p. Il y est montré que l'estimateur en question n'est admissible que si $n \leq 5$.

1. INTRODUCTION

Let X be a binomial random variable, $b(n,p)$, with n fixed and $p \in [0,1]$. It is customary to estimate the mean and the variance of the binomial random variable by the maximum-likelihood estimators $\delta_0(X) = X$ and $\delta^*(X) = X(n-X)/n$, respectively. While it is well known that δ_0 is admissible under squared-error loss, the admissibility property of δ^* is unknown. [A formal proof of the admissibility of δ_0 can be deduced from the results in Karlin (1958); Skibinsky and Rukhin (1989) have lately derived an admissibility criterion for estimators of p.]

Johnson (1971) has given a necessary and sufficient condition characterizing admissible estimators of $f(p)$ relative to squared-error loss for any continuous real-valued function f on [0,1]. Although it is true that the binomial variance is a function of p, it is very difficult to apply Johnson's result directly to prove or disprove the admissibility of δ^*. Here we develop a necessary and sufficient condition for δ^* to be admissible, which is then used to tackle the unsolved problem. The approach we have adopted is similar to that considered in Kozek (1982).

The problem of estimating the normal variance has been well studied in the decision-theory literature. An excellent review, together with a historical account of the development of the problem, is given in Maatta and Casella (1990). However, there is little work on the decision-theoretic approach to variance estimation for binomial distributions. There is an exercise in Lehmann (1983) which asks for comparing the expected squared error of the uniformly minimum-variance unbiased estimator with that of the Bayes estimator for the Jeffreys prior. The UMVUE can be shown to be inadmissible using results in Berger (1990), whereas the Bayes estimator is evidently admissible. Also, the estimator $X(n-X)/(n+1)$ has been shown to be admissible by Johnson (1971). A major contribution of our paper is to point out the pitfalls of using the MLE in estimating the binomial variance. We show that, under squared-error loss, the MLE is admissible for $n \leq 5$ and inadmissible for $n \geq 6$.

The paper is divided into four sections. Section 2 gives a necessary and sufficient condition for the maximum-likelihood estimator (MLE) of the binomial variance to be admissible. Section 3 establishes the (in)admissibility of the MLE. Section 4 presents some concluding remarks.

2. A NECESSARY AND SUFFICIENT CONDITION

From Johnson (1971) it can be shown that estimators with the following representation form the class of admissible estimators of $np(1-p)$ relative to squared-error loss:

$$\delta(x) = \begin{cases} 0 & \text{for} \quad x \leq r \text{ or } x \geq s, \\ \dfrac{\displaystyle\int_0^1 np^{x-r}(1-p)^{s-x}\pi(dp)}{\displaystyle\int_0^1 p^{x-r-1}(1-p)^{s-x-1}\pi(dp)} & \text{for} \quad r+1 \leq x \leq s-1, \end{cases} \tag{2.1}$$

where r and s are integers, $-1 \leq r < s \leq n+1$, and π is a probability measure with $\pi(\{0\} \cup \{1\}) < 1$. [Note that this representation can also be derived by the stepwise algorithm proposed in Brown (1981), which is applicable to mush more general settings.] However, it is hard to show the (non)existence of a prior corresponding to the MLE δ^* by applying the above result directly. Thus an alternative version of (2.1) is needed. We establish below a necessary and sufficient condition which will be used to prove the (in)admissibility of δ^*.

THEOREM 1. *δ^* is admissible if and only if there exists a measure m' on $[0,\infty)$ and a nonnegative constant c such that*

$$b_k = \int_0^\infty t^k m'(dt) < \infty \qquad \textit{for} \quad k = 0, \ldots, n-1, \tag{2.2}$$

$$b_n = \int_0^\infty t^n m'(dt) + c < \infty, \tag{2.3}$$

$$b_i = b_{n-i}, \qquad i = 0, \ldots, n, \tag{2.4}$$

$$b_{x+1} = \left(\frac{n}{\delta^*(x)} - 2\right) b_x - b_{x-1}, \qquad x = 1, \ldots, n-1. \tag{2.5}$$

Proof. Necessity: If δ^* is admissible, there exists a probability measure π for which (2.1) holds (with $r=0$ and $s=n$). Moreover, since $\delta^*(x) = \delta^*(n-x)$, we can always find such π which is symmetric about $p = \frac{1}{2}$. Let $\pi = \pi' + \pi''$, where the measures π' and π'' are orthogonal and π'' is concentrated on $\{1\}$. Then

$$\int_0^1 p^k(1-p)^{n-k}\pi(dp) = \int_0^1 \left(\frac{p}{1-p}\right)^k d\pi^*(p), \qquad 0 \leq k \leq n-1,$$

and

$$\int_0^1 p^n \pi(dp) = \int_0^1 \left(\frac{p}{1-p}\right)^n d\pi^*(p) + \pi''(1),$$

where $d\pi^* = (1-p)^n\, d\pi'$. Denote by m' the measure on $[0,\infty)$ induced by the mapping $p \longrightarrow p/(1-p)$ and π^*, and set $c = \pi''(1)$; then the above integrals are the requred b_k and (2.1) becomes (2.5). Finally, from the symmetric property of π, it is easy to check that $b_i = b_{n-i}$ for $0 \le i \le n$.

Sufficiency: Using the transformation $p = t/(1+t)$, we obtain

$$b_k = \int_0^1 \left(\frac{p}{1-p}\right)^k d\pi^*(p), \qquad 0 \le k \le n-1$$

$$b_n = \int_0^1 \left(\frac{p}{1-p}\right)^n d\pi^*(p) + c,$$

where π^* is a proper measure on [0,1] with $\pi^*(1) = 0$. Define

$$d\pi' = \frac{d\pi^*/(1-p)^n}{\displaystyle\int_0^1 d\pi^*(p)/(1-p)^n + c},$$

$$\pi''(p) = \begin{cases} \dfrac{c}{\displaystyle\int_0^1 d\pi^*(p)/(1-p)^n + c} & \text{if} \quad p = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Noting that $\int_0^1 d\pi^*(p)/(1-p)^n$ is finite since b_n is finite, we see that $\pi = \pi' + \pi''$ is a probability measure on [0,1] with $\pi(\{0\} \cup \{1\}) < 1$. Now we express b_k as integrals with respect to the probability measure π, and (2.5) becomes (2.1). Thus δ^* is admissible. Q.E.D.

3. THE (IN)ADMISSIBLITY OF δ^*

We first show that δ^* is admissible for $n \le 5$ and then establish the inadmissibility result when $n \ge 6$.

THEOREM 2. *δ^* is admissible for $n \le 5$ under squared-error loss.*

Proof. The case $n = 1$ is trivial. When $2 \le n \le 5$, the conditions (2.4) and (2.5) in Theorem 1 became

(1) $b_0 = b_1 = b_2 > 0$ for $n = 2$,
(2) $b_1 = b_2 = 2b_0/3$ and $b_3 = b_0 > 0$ for $n = 3$,
(3) $b_1 = b_2 = b_3 = 3b_0/7$ and $b_4 = b_0 > 0$ for $n = 4$,
(4) $b_1 = b_4 = 28b_0/95$, $b_2 = b_3 = 24b_0/95$, and $b_5 = b_0 > 0$ for $n = 5$.

It can be shown that, with the b_i related as given above, the matrices

$$\|b_{i+j}\|_{i,j=0}^{m}, \ \|b_{i+j+1}\|_{i,j=0}^{m-1} \qquad \text{for} \quad n = 2m$$

and

$$\|b_{i+j}\|_{i,j=0}^{m-1}, \ \|b_{i+j+1}\|_{i,j=0}^{m-1} \qquad \text{for} \quad n = 2m-1$$

are positive semidefinite for the above four cases. By Theorem V.10.1 in Karlin and Studden (1966), we are assured of the existence of a measure m' such that (2.2) and (2.3) hold. Thus, δ^* is admissible. Q.E.D.

THEOREM 3. *δ^* is inadmissible for $n \geq 6$ under squared-error loss.*

Proof. We prove the result by contradiction.

Case 1. n is even. Let $n = 2m$. (2.1) gives

$$m/2 = \delta^*(m)$$
$$= \frac{\int_0^1 np(1-p)p^{m-1}(1-p)^{m-1}\pi(dp)}{\int_0^1 p^{m-1}(1-p)^{m-1}\pi(dp)}. \tag{3.1}$$

Since $np(1-p) \leq m/2$, it follows that the probability measure

$$\frac{p^{m-1}(1-p)^{m-1}\pi(dp)}{\int_0^1 p^{m-1}(1-p)^{m-1}\pi(dp)}$$

is concentrated on the one point $p = \frac{1}{2}$, where $np(1-p) = m/2$. Consequently, π is concentrated on $p = 0, \frac{1}{2}, 1$. Hence (2.1) then implies $\delta(x) = m/2$ for $2 \leq x \leq n-2$. If $n \geq 6$, this means that the estimators δ^* cannot be represented in the form (2.1), and so must be inadmissible.

Case 2. n is odd. Let $n = 2m - 1 > 6$, the conditions (2.4) and (2.5) imply

$$b_{m-1} = b_m = \alpha > 0,$$

$$b_{m-2} = b_{m+1} = \left(1 + \frac{1}{m(m-1)}\right)\alpha,$$

$$b_{m-3} = b_{m+2} = \left\{\left(1 + \frac{1}{m(m-1)}\right)\left(2 + \frac{9}{(m+1)(m-2)}\right) - 1\right\}\alpha,$$

$$b_{m-4} = b_{m+3} = \left[\left\{\left(1 + \frac{1}{m(m-1)}\right)\left(2 + \frac{9}{(m+1)(m-2)}\right) - 1\right\}\right.$$
$$\left.\times\left\{2 + \frac{25}{(m+2)(m-3)}\right\} - \left(1 + \frac{1}{m(m-1)}\right)\right]\alpha.$$

Define $dm'' = t^{m-4}dm'$; then (2.2) gives

$$b_{m-4+k} = \int_0^\infty t^k m''(dt), \qquad k = 0, \ldots, 6.$$

By Theorem V.10.1 in Karlin and Studden (1966), the matrix

$$A = \|b_{m-4+i+j}\|_{i,j=0}^3$$

must be positive semidefinite. If we write

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

where A_{ij} are 2×2 matrices, the determinant of A can be computed as

$$\det A = (\det A_{11})(\det A_{22}) \det(I - A_{22}^{-1}A_{21}A_{11}^{-1}A_{12})$$
$$= \frac{-4(2m-1)^4(m^2-m+1)(8m^4-16m^3-63m^2+71m-52)}{m^4(m-3)(m-2)^3(m-1)^4(m+1)^3(m+2)}\alpha^4.$$

It can be checked that $\det A < 0$ for $m \geq 4$ and hence A is not positive semidefinite. This contradiction leads to the conclusion that δ^* is inadmissible for any odd number $n \geq 7$. Q.E.D.

4. CONCLUDING REMARKS

In this paper we have proved that the MLE of the variance of a binomial distribution is admissible for $n \leq 5$ and inadmissible for $n \geq 6$. When $n \leq 5$, it can be shown that the MLE is a stepwise Bayes estimator with respect to a prior (of p) which depends on n. Since

$$\int_0^1 p^a(1-p)^b\pi(dp) = \sum_{i=o}^{b} \binom{b}{i} (-1)^i M(a+i),$$

where $M(k) = \int_0^1 p^k\pi(dp)$ is the kth moment of π, the condition (2.1) relates the MLE with the moments of the prior π. Using a result (Theorem IV.1.1) in Karlin and Studden (1966), we obtain

$$M(2) = M(1) - \tfrac{1}{4} \text{ for } n = 2,$$

$$M(2) = M(1) - \tfrac{2}{9} \text{ for } n = 3,$$

$$M(2) = M(1) - \tfrac{3}{14} \text{ for } n = 4,$$

$$M(2) = M(1) - \tfrac{4}{19} \text{ for } n = 5.$$

It is clear that the moments are different in the four cases. Thus the prior is dependent on n in spite of the fact that the MLE is admissible for $2 \leq n \leq 5$.

ACKNOWLEDGEMENT

Duncan K.H. Fong's research was supported by NSF grant DMS-9003926. The authors thank a referee for some helpful comments.

REFERENCES

Berger, J.O. (1990). On the inadmissibility of unbiased estimators. *Statist. Probab. Lett.*, 9, 381–384.

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

Johnson, B. McK. (1971). On admissible estimators for certain fixed sample binomial problems. *Ann. Math. Statist.*, 42, 1579–1587.

Karlin, S. (1958). Admissibility for estimation with quadratic loss. *Ann. Math. Statist.*, 29, 406–436.

Karlin, S., and Studden, W.J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics*. Wiley, New York.

Kozek, A. (1982). Towards a calculus for admissibility. *Ann. Statist.*, 10, 825–837.

Lehmann, E.L. (1983). *Theory of Point Estimation*. Wiley, New York.

Maatta, J.M., and Casella, G. (1990). Developments in decision-theoretic variance estimation (with discussion). *Statist. Sci.*, 5, 90–120.

Skibinsky, M., and Rukhin, A.L. (1989). Admissible estimators of binomial probability and the inverse bayes rule map. *Ann. Inst. Statist. Math.*, 41, 699–716.

Received 24 September 1990
Revised 6 January 1992
Accepted 15 January 1992

Department of Mathematics
White Hall
Cornell University
Ithaca, NY 14853
U.S.A.

Department of Management Science
340 Beam BAB
Pennsylvania State University
University Park, PA 16802
U.S.A.