A MEAN-SQUARE-ERROR CHARACTERIZATION OF BINOMIAL-TYPE DISTRIBUTIONS

BY M. C. K. TWEEDIE

University of Liverpool

1. Hodges and Lehmann (1950), with an acknowledgement of priority to Rubin, have shown that, if θ^* is an estimator of the parameter θ based on a single observation on the binomial distribution

$$\text{prob}\,\{X = x\} = \binom{N}{x}\theta^x(1-\theta)^{N-x}, \qquad \text{for } x = 0, 1, \cdots, N, \tag{1}$$

with N fixed and known and a loss function $W(\theta, \theta^*) = (\theta - \theta^*)^2$, the minimax estimator is

$$\theta^M = (X + \tfrac{1}{2}N^{\frac{1}{2}})/(N + N^{\frac{1}{2}}). \tag{2}$$

This is also the Bayesian estimator when θ has a Bayesian prior probability element of the form

$$d\lambda(\theta) = (\theta(1-\theta))^{\frac{1}{2}N^{\frac{1}{2}}-1}\,d\theta/B(\tfrac{1}{2}N^{\frac{1}{2}}, \tfrac{1}{2}N^{\frac{1}{2}}). \tag{3}$$

This distribution of θ is therefore a least favourable one for the estimator θ^M. The proofs are simplified by the facts that

(i) θ^M is linear in the random variable X,

(ii) the mean square error (mse) or risk function of θ^M is independent of θ.

2. Since the binomial family of distributions is one in which the sample mean of a constant number of independent observations is both a sufficient statistic and a minimum variance unbiased estimator (mvue) of their common expectation, it is of interest to investigate whether the two properties, (i) and (ii), apply to any other families of distributions which possess minimum variance unbiased estimators.

One may, therefore, seek to replace the binomial by other "exponential-type" distributions. In the notation of Lehmann's book (1959), page 50, (1) might be replaced by the probability element

$$C(\theta)e^{Q(\theta)T(x)}h(x)\,d\mu(x), \tag{4}$$

in which μ is a measure function; for then the sample mean of $T(X)$ is the mvue of $ET(X)$. It is clear, however, that little is gained in the present problem by using (4) instead of the simpler formulation,

$$e^{\alpha x - F(\alpha)}\,d\mu(x). \tag{5}$$

This formulation, or, rather, a closely related one, was found convenient in a paper (1947) in which I called it a Laplacian distribution, and it has also (cf.

Received 16 August 1965; revised 7 August 1966.

Lehmann (1959), page 115) been shown to be a Pólya-type case of the general exponential form (4). A set of probability distributions which share a common function F may be called a family of distributions in the Laplacian class.

3. Using (5), the cumulant-generating function (cgf) of X can be written (cf. Fisher (1934)) as

$$\text{c.g.f.} = \ln E(e^{tX}) = F(\alpha + t) - F(\alpha). \tag{6}$$

Then

$$EX = \kappa_1 = F'(\alpha), \qquad VX = \kappa_2 = F''(\alpha) = d\kappa_1/d\alpha. \tag{7}$$

If X' has the binomial distribution (1), and if we redefine $X = x_0 + X'\Delta x$, where x_0 and Δx are constants, independent of X and θ, so that the possible values of X are $x_0, x_0 + \Delta x, \cdots, x_0 + N\Delta x$, then

$$\text{cgf of } X = \ln Ee^{tX} = x_0 t + N \ln (1 - \theta + \theta e^{t\Delta x}). \tag{8}$$

The distribution of X may be said to be of the binomial type. Conversely, if a cgf is of the form (8), then the distribution of X is of the binomial type.

4. Minimization of the average mse of a linear estimator of an expectation.

Let $\bar{X}$ be the mean of a fixed number (n) of independent observations on (5). If $\kappa_1(\theta)$, i.e., EX, which $= E\bar{X}$, is estimated by $a + b\bar{X}$, in which a and b are constants, and if θ has a prior distribution λ, then

$$\text{mse} = E\{(a + b\bar{X} - \kappa_1(\theta))^2 \mid \theta\} = b^2\kappa_2/n + (a + (b - 1)\kappa_1)^2 \tag{9}$$

$$\text{average mse} = E(\text{mse}) = b^2E\kappa_2/n + (b - 1)^2V\kappa_1 + (a + (b - 1)E\kappa_1)^2.$$

Write

$$\nu = E\kappa_2/V\kappa_1 .$$

Then the average mse is minimized by

$$a = \nu E\kappa_1/(\nu + n), \qquad b = n/(\nu + n).$$

The estimator is then

$$(\nu E\kappa_1 + n\bar{X})/(\nu + n),$$

which is, incidentally, a weighted mean of the sample mean $\bar{X}$ and the prior mean $E\kappa_1$. It has

$$\text{mse} = \{(\kappa_1 - E\kappa_1)^2\nu^2 + n\kappa_2\}/(\nu + n)^2, \tag{10}$$

of which the average value is

$$\min_{a,b} E(\text{mse}) = E\kappa_2/(\nu + n) = 1/(1/V\kappa_1 + 1/E\kappa_2).$$

5. We now prove the following preliminary theorem:

THEOREM 1. *The normal and the binomial type are the only non-degenerate* *families of univariate Laplacian distributions for which there exists a linear function of the sample mean $\bar{X}$ which has a constant* mse *as an estimator of EX.*

PROOF. Inserting (7) in (9), we require

$$(11) \qquad b^2F''(\alpha)/n + (a + (b-1)F'(\alpha))^2 = c^2, \quad \text{a constant.}$$

One solution of (11) is to take $b = 1$ and $F''(\alpha)$ to be constant. This leads to a general univariate normal distribution of X, with a degenerate solution in which, for any one value of α, X takes only one value, with unit probability.

The regular solution of (11) has $b \neq 1$ and gives

$$a\alpha + (b-1)F(\alpha) = \{b^2/(n(b-1))\} \ln (ke^{h\alpha} + e^{-h\alpha}) + \text{constant},$$

in which $h = (b-1)cn/b^2$ and k is a constant of integration. Thus, the cgf (6) of X is

$$(12) \quad F(\alpha + t) - F(\alpha)$$
$$= n^{-1}(b/(b-1))^2 \ln ((e^{-2h\alpha} + ke^{2ht})/(e^{-2h\alpha} + k)) - (a + c)t/(b-1).$$

By taking limits, under suitable conditions, one gets the special solutions already mentioned. The regular integral has $b \neq 1$ and is seen to have an imaginary period $\pi i/h$ in t and thus, subject to further conditions, to be the cgf of a lattice distribution with step equal to $2h$. The origin is arbitrary and is adjusted by the value of a. The coefficient $b^2/(b-1)^2$ of the logarithm appears to be the secondary parameter, in the terminology of my 1947 paper. In fact, (12) is of the binomial type (8), with

$$\Delta x = 2h, \qquad x_0 = -(a + c)/(b-1),$$
$$N = b^2/((b-1)^2 n), \qquad \theta = k/(e^{-2h\alpha} + k).$$

One could take $k = 1$ without loss of generality by simultaneously changing the origin of α.

Then

$$h = \tfrac{1}{2}\Delta x, \qquad b = (Nn)^{\frac{1}{2}}/((Nn)^{\frac{1}{2}} \pm 1), \qquad c = \mp\tfrac{1}{2}N\Delta x/((Nn)^{\frac{1}{2}} \pm 1),$$
$$a = \pm(x_0 + \tfrac{1}{2}N\Delta x)/((Nn)^{\frac{1}{2}} \pm 1),$$
$$k = e^{-\alpha\Delta x}\theta/(1-\theta).$$

Thus there are two solutions:

$$a + b\bar{X} = (\bar{X}(Nn)^{\frac{1}{2}} \pm (x_0 + \tfrac{1}{2}N\Delta x))/((Nn)^{\frac{1}{2}} \pm 1)$$
$$= x_0 + \tfrac{1}{2}N\Delta x + (\bar{X} - x_0 - \tfrac{1}{2}N\Delta x)(Nn)^{\frac{1}{2}}/((Nn)^{\frac{1}{2}} \pm 1),$$

which ranges from $x_0 \pm \frac{1}{2}N\Delta x/((Nn)^{\frac{1}{2}} \pm 1)$ to $x_0 + N\Delta x \mp \frac{1}{2}N\Delta x/((Nn)^{\frac{1}{2}} \pm 1)$, and has mse $= (\frac{1}{2}N\Delta x)^2/((Nn)^{\frac{1}{2}} \pm 1)^2$.

The solution with the upper sign includes the minimax estimator of $N\theta$ which is obtainable from (2). It dominates the solution which has the lower sign, but the latter is not completely worthless; for its mse is less than the maximum mse of the constant 'estimator' $x_0 + \frac{1}{2}N\Delta x$, namely $(\frac{1}{2}N\Delta x)^2$, when $Nn > 4$.

6. Hodges and Lehmann (1950) show that the appropriate special case of the solution in Section 5 with the upper sign is the minimax solution for the binomial (which has $h = \frac{1}{2}$ and $a = -c$) by showing that it is equal to the conditional expectation of $\kappa_1(\theta)$ over the posterior distribution which results from combining (1) and (3), in accordance with their Theorem 2.1. In the present note, however, in which the orientation is different, a characterization of distributions of the binomial type is obtained by combining Theorem 1 with some of the results given in Section 4.

THEOREM 2. *If*

(i) *the real random variable X has a non-degenerate distribution in a family of the Laplacian class, with parameter θ;*

(ii) *attention is confined to estimators of EX which are linear in $\bar{X}$;*

(iii) *θ has a prior distribution λ in which EX has bounded positive variance; then there exists an estimator which both minimizes the average* mse *and has a constant* mse, iff

(iv) *the family of distributions of X is of the binomial type;*

(v) *λ is such that $E\theta = \frac{1}{2}$ and $V\theta = 1/(4(Nn)^{\frac{1}{2}} + 1))$*, i.e., $(E\theta - E(\theta^2))/V\theta$ *is the square root of a positive integer (namely, Nn).*

PROOF. From Theorem 1, only the normal and binomial-type families need be considered. Condition (iii), in conjunction with (10), excludes the normal.

In the binomial type, $\kappa_1 = x_0 + N\theta\Delta x$ and $\kappa_2 = N(\Delta x)^2\theta(1 - \theta)$. When these expressions are substituted in (10), the constancy of the resulting mse requires the coefficients of θ^2 and θ to vanish. This leads to $\nu = (n/N)^{\frac{1}{2}}$ and $E\theta = \frac{1}{2}$; and (v) is equivalent to these two conditions.

As a matter of interest, if $E\,\theta = \frac{1}{2}$ but $n \neq N\nu^2$,

$$\max_\theta (\text{mse}) = \max(n, N\nu^2)\cdot N(\tfrac{1}{2}\Delta x)^2/(\nu + n)^2.$$

This method of proving Theorem 2 was suggested by a referee's comments. An alternative method involves combining (10) with (7) and solving the resulting differential equation for F.

REFERENCES

FISHER, R. A. (1934). Two new properties of mathematical likelihood. *Proc. Royal Soc. Ser. A.* **144** 285–307.

HODGES, J. and LEHMANN, E. L. (1950). Some problems in minimax point estimation. *Ann. Math. Statist.* **21** 182–197.

LEHMANN, E. L. (1959). Testing Statistical Hypotheses. Wiley, New York.

TWEEDIE, M. C. K. (1947). Functions of a statistical variate with given means, with special reference to Laplacian distributions. *Proc. Cambridge Philos. Soc.* **43** 41–49.