NOTES

A GENERAL CONCEPT OF UNBIASEDNESS

By E. L. Lehmann

University of California, Berkeley, and Princeton University

The term unbiasedness was introduced by Neyman and Pearson [1] in connection with hypothesis testing. A test of the hypothesis $\theta \in \omega$ against the alternatives $\theta \in \Omega - \omega$ is said to be unbiased at level α if its power function β satisfies

$$\begin{aligned} &\beta(\theta) \leq \alpha \text{ for } \theta \in \omega, \\ &\beta(\theta) \geq \alpha \text{ for } \theta \in \Omega - \omega. \end{aligned} \tag{1}$$

In 1937 Neyman [2] developed a theory of estimation by confidence sets. He established a duality with the theory of hypothesis testing, so that to each notion of one theory corresponds an analogous one in the other. In particular, he defined a family of confidence sets $A(x)$ to be unbiased if

$$P_\theta(A(X) \supset \theta') \leq P_\theta(A(X) \supset \theta) \text{ for all } \theta \text{ and } \theta'. \tag{2}$$

While the above two definitions are closely related, a third use of the term unbiasedness was made in a rather different context. In presenting their version of the Gauss-Markov theorem on least squares David and Neyman [3] defined a point estimate $\delta(X)$ of $g(\theta)$ to be unbiased if its expectation coincides with the estimated value, that is, if

$$E_\theta \delta(X) \equiv g(\theta). \tag{3}$$

It was pointed out later by Brown [4] that one obtains other analogous definitions by postulating that some central value of the distribution of $\delta(X)$ other than the mean coincides with the estimated value. Using the median as an example he defined $\delta(X)$ to be median-unbiased if

$$P_\theta(\delta(X) > g(\theta)) = P_\theta(\delta(X) < g(\theta)) \text{ for all } \theta. \tag{4}$$

In view of Wald's theory of decision functions [5] it seems tempting to try to give a definition of unbiasedness at the level of generality of this theory. Suppose we are concerned with a decision problem where the loss resulting from a decision $\delta(X)$ is $W(\theta, \delta(X))$ when the true parameter value is θ. In analogy with (2) we shall say that a decision procedure $\delta(X)$ is unbiased if for each θ

$$E_\theta W(\theta', \delta(X)) = \min \text{ when } \theta' = \theta. \tag{5}$$

This clearly reduces to Neyman's definition for confidence sets if one uses for loss function,

$$W(\theta, \delta(x)) = \begin{cases} 0 \text{ if the confidence set } \delta(x) \text{ covers } \theta, \\ 1 \text{ otherwise.} \end{cases} \tag{6}$$

In order to obtain an interpretation of condition (5), let us consider the case that for each parameter value θ there exists a unique "correct" decision d and that each d is correct for at least some θ. This is the case for example in hypothesis testing and in point estimation. Here a correct decision may be defined by the condition $W(\theta, d) = 0$. Let us say that two parameter values θ, θ' are equivalent, $\theta \sim \theta'$, if the correct decision is the same for both of them, and let us suppose that for any decision d'

$$W(\theta_1, d') = W(\theta_2, d') \text{ whenever } \theta_1 \sim \theta_2. \tag{7}$$

Then the loss $W(\theta, d')$ depends only on the actual decision taken, say d', and the decision d that would have been correct, and we may write for it $W(d, d')$. The loss $W(d, d')$ is a measure of how far the two decisions d and d' are apart, and (5) states that a decision function $\delta(X)$ is unbiased if on the average it comes closer to the correct decision than to any incorrect decision.

Let us now apply this notion to some particular examples. Let the decision to accept and reject the hypothesis $H: \theta \in \omega$ be denoted by d_0 and d_1, respectively. Since in the Neyman-Pearson theory of hypothesis testing one is concerned only with the probabilities of the two types of error, the natural associated loss function is of the form

$$\begin{aligned} W(\theta, d_0) &= \begin{cases} a \text{ if } \theta \in \Omega - \omega, \\ 0 \text{ if } \theta \in \omega; \end{cases} \\ W(\theta, d_1) &= \begin{cases} b \text{ if } \theta \in \omega, \\ 0 \text{ if } \theta \in \Omega - \omega. \end{cases} \end{aligned} \tag{8}$$

It is easy to see that in this case (5) becomes

$$\begin{aligned} P_\theta(d_1) &\leq \frac{a}{a+b} \text{ for } \theta \in \omega, \\ P_\theta(d_1) &\geq \frac{a}{a+b} \text{ for } \theta \in \Omega - \omega, \end{aligned} \tag{9}$$

where $P_\theta(d)$ denotes the probability that $\delta(X) = d$ when θ is the true parameter value. This is exactly the usual definition (1) with $\alpha = a/(a+b)$.

Let us next consider point estimation where the loss is taken as the square of the error. If the function to be estimated is $g(\theta)$, condition (5) becomes

$$E_\theta[\delta(X) - g(\theta')]^2 \geq E_\theta[\delta(X) - g(\theta)]^2 \text{ for all } \theta, \theta'. \tag{10}$$

Let $E_\theta\delta(X) = h(\theta)$. In the usual case that $h(\theta)$ is one of the possible values of the function g, the left-hand side of (10) is minimized for $g(\theta') = h(\theta)$. Thus the inequality holds for all θ' if and only if $g(\theta) = h(\theta)$, which is equivalent to (3). So again (5) reduces just to the usual definition.

Even if $h(\theta)$ is not one of the possible values of g, it is easily seen that (10) is equivalent to

$$| h(\theta) - g(\theta) | = \min_{\theta'} | h(\theta) - g(\theta') | .$$

Then, if for example Ω is a real interval and g is continuous and strictly monotone, there can exist at most two values of θ for which $g(\theta) \neq h(\theta)$. If further, as is usually the case, $h(\theta)$ is continuous for all estimates δ, we must have $h(\theta) \equiv g(\theta)$.

Quite analogously one sees that if $W(\theta, \delta(x)) = |\delta(x) - g(\theta)|$, definition (5) reduces to Brown's notion of median-unbiasedness.

While the definition given here seems satisfactory in that it does reduce under reasonable assumptions to the usual concepts, it is somewhat more restrictive than appears at first sight. If for example there exists for each θ a unique correct decision d and if the loss function is of the form

$$W(\theta, d') = f(\theta)V(d, d'),$$

then, with the trivial exception of procedures for which $E_\theta V(d, \delta(x)) = 0$ for some d and some value of θ in ω_d, no unbiased procedure can exist unless $f(\theta)$ is constant on each ω_d. For let $\theta, \theta' \varepsilon \omega_d$. On substituting in (5) we see that unbiasedness implies $f(\theta') \geq f(\theta)$ and hence by symmetry $f(\theta') = f(\theta)$. In hypothesis testing for example if the loss is zero for a correct decision, it follows, again with trivial exceptions, that unbiased tests can exist only if the loss function is given by (8).

It is perhaps worth pointing out certain connections between the principle of unbiasedness and that of invariance. Consider for example the problem of estimating θ from a sample $X_1, \cdots, X_n$ where the X's are uniformly distributed on $(0, \theta)$. If one takes as loss function

$$W(\theta, \delta(x_1, \cdots, x_n)) = [\delta(x_1, \cdots, x_n) - \theta]^2/\theta^2 \tag{11}$$

the problem transforms in an obvious manner under a change of scale, and one may wish to consider only estimates having the invariance property

$$\delta(cX_1, \cdots, cX_n) = c\delta(X_1, \cdots, X_n) \text{ for all } c > 0. \tag{12}$$

If $Y = \max(X_1, \cdots, X_n)$, it is easily seen that among all invariant estimates the one that uniformly minimizes the expected loss is

$$\frac{n+2}{n+1} Y. \tag{13}$$

This estimate does not have the usual unbiasedness property since

$$E_\theta\left[\frac{n+2}{n+1} Y\right] = \frac{n(n+2)}{(n+1)^2}\theta.$$

However a simple computation shows that (13) is unbiased in the sense of (5) with respect to the invariant loss function (11).

More generally, let $\mathcal{G}$ be a group of measurable 1:1 transformations on the sample space. Let gX be the random variable that takes on the value gx when $X = x$, and suppose that when X has a distribution p_θ, $\theta \varepsilon \Omega$, then gX has a distribution $p_{\theta'}$, $\theta' \varepsilon \Omega$. Denote this θ' by $\bar{g}\theta$ and suppose that $\bar{g}\theta$ defines a 1:1 transformation on Ω. Let $\bar{\mathcal{G}}$ be the group of transformations $\bar{g}$ and assume that

there exists a group $\mathcal{G}^*$ of 1:1 transformations on the decision space D such that $\mathcal{G}^*$ is homomorphic to $\bar{\mathcal{G}}$ and

$$W(\bar{g}\theta, g^*d) = W(\theta, d) \text{ for all } \theta \in \Omega, d \in D. \tag{14}$$

Then a decision function δ is said to be invariant if

$$\delta(gX) = g^*\delta(X). \tag{15}$$

This is a natural generalization of the definition of invariance given by Hunt and Stein [6, 7], and is essentially the definition used by Peisakoff [8]. Further, δ is said to be almost invariant if (15) holds except on a set N_g of measure 0.

Whenever among all unbiased procedures there exists a unique[1] one that uniformly minimizes the risk, then it is almost invariant. This follows easily from the fact that if $\delta(X)$ is unbiased $g^*\delta(g^{-1}X)$ is also unbiased. It is not in general true that conversely an optimum invariant test is necessarily unbiased. However, this result does hold under certain restrictions.[2] If

(i) $\mathcal{G}$ is transitive, i.e., given any θ, θ' there exists $\bar{g}$ such that $\theta = \bar{g}\theta'$,

(ii) $\mathcal{G}^*$ is commutative,

and if among all invariant (or almost invariant) procedures there exists one that uniformly minimizes the risk, then it is unbiased.

To see this, let δ be invariant and such that for any other invariant procedure δ'

$$E_\theta W(\theta, \delta'(X)) \geq E_\theta W(\theta, \delta(X)).$$

Let $\theta' \neq \theta$, $\theta = \bar{g}\theta'$, say. Then

$$E_\theta W(\theta', \delta(X)) = E_\theta W(\theta, g^*\delta(X)) \geq E_\theta W(\theta, \delta(X)).$$

Here the inequality follows since by (ii) the invariance of $\delta(X)$ implies that $g^*\delta(X)$ is also invariant.

While assumptions (i) and (ii) are satisfied in many estimation problems, (i) will in general not hold in a problem of hypothesis testing because of the asymmetry of d_0 and d_1. Here the result in question follows when the loss function is given by (8) from the fact that if a test is unbiased so is any test that is uniformly better, together with the unbiasedness and invariance of the test $\varphi(x) \equiv a/(a + b)$ (i.e., the test that rejects the hypothesis with probability $a/(a + b)$ regardless of the observations).

That the result is not true in general if we drop either one of the two conditions (i) or (ii) can be seen from the following example. For estimating the mean ξ of a normal variable with unknown variance σ^2 when the loss function is $[(\delta(x) - \xi)/\sigma]^2$, the best invariant estimate is X both with respect to the group

$$\mathcal{G}_1: gx = x + b, \qquad -\infty < b < \infty$$

[1] Throughout, this is understood to mean unique up to a set of measure zero.

[2] I am grateful to the referee for pointing out an error in my original statement of this result.

and with respect to

$$\mathcal{G}_2: gx = ax + b, \qquad 0 < a < \infty, -\infty < b < \infty.$$

For this problem an unbiased estimate in the sense of (5) does not exist, and it is seen that $\mathcal{G}_1$ satisfies (ii) but not (i) while $\mathcal{G}_2$ satisfies (i) but not (ii).

The notion of unbiasedness in many cases leads to reasonable decision procedures and this seems to be in general the value of such concepts. On the other hand there is no guarantee that an optimum unbiased procedure is necessarily satisfactory. As an example (for another example see [9]) consider a Poisson variable X which is observed only if $X \neq 0$, so that the distribution of X is given by

$$P(X = K) = \frac{\lambda^K}{K!} e^{-\lambda}(1 - e^{-\lambda})^{-1}, \qquad K = 1, 2, \cdots. \tag{15}$$

It is desired to estimate the probability $e^{-\lambda}$ of X being zero, and the loss function is squared error. The condition of unbiasedness gives

$$\sum_{K=1}^{\infty} \delta(K) \frac{\lambda^K}{K!} \equiv 1 - e^{-\lambda} \equiv \sum_{K=1}^{\infty} (-1)^{K+1} \frac{\lambda^K}{K!}, \tag{16}$$

so that $\delta(K) = (-1)^{K+1}$. Thus the estimate takes on only impossible values and instead of decreasing with K as one would expect, it does not depend on the order of magnitude of K at all.

As a final remark we mention, without going into details, the following extension of the notion of unbiasedness. Instead of comparing $E_\theta W(\theta', \delta(X))$ only with $E_\theta W(\theta, \delta(X))$ we may ask that $E_\theta W(\theta', \delta(X))$ be a nondecreasing function of $v(\theta, \theta')$, where $v(\theta, \theta')$ in some sense measures the distance between θ and θ'. This notion is a generalization of one used by P. L. Hsu [10] in the theory of hypothesis testing. It is also closely connected with the principle of invariance. In fact if there exists a group of transformations leaving the problem invariant then with a suitable definition of $v(\theta, \theta')$ it is easy to see under weak assumptions on the loss function that Theorem 7.1 of [7] generalizes to the present case. This theorem states essentially that the totality of procedures for which $E_\theta W(\theta', \delta(X))$ depends only on $v(\theta, \theta')$ coincides with the totality of invariant procedures.

REFERENCES

[1] J. Neyman and E. S. Pearson, "Contributions to the theory of testing statistical hypotheses. I. Unbiased critical regions of type A and type A_1," *Stat. Res. Mem.*, Vol. 1 (1936), pp. 1–37.

[2] J. Neyman, "Outline of a theory of statistical estimation based on the classical theory of probability," *Phil. Trans. Roy. Soc. London*, Series A, Vol. 236 (1937), pp. 333–380.

[3] F. N. David and J. Neyman, "Extension of the Markoff theorem on least squares," *Stat. Res. Mem.*, Vol. 2 (1938), pp. 105–116.

[4] G. W. Brown, "On small sample estimation," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 582–585.

[5] A WALD, *Statistical Decision Functions*, John Wiley and Sons, 1950.
[6] G. HUNT AND C. STEIN, "Most stringent tests of statistical hypotheses," unpublished.
[7] E. L. LEHMANN, "Some principles of the theory of testing hypotheses," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 1–26.
[8] M. PEISAKOFF, "Transformation parameters," unpublished thesis, Princeton University, 1950.
[9] P. R. HALMOS, "The theory of unbiased estimation," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 34–43.
[10] P. L. HSU, "Analysis of variance from the power function standpoint," *Biometrika*, Vol. 32 (1941), pp. 62–69.

ONE-SIDED CONFIDENCE CONTOURS FOR PROBABILITY DISTRIBUTION FUNCTIONS[1]

BY Z. W. BIRNBAUM AND FRED H. TINGEY[2]

University of Washington

Summary. Let $F(x)$ be the continuous distribution function of a random variable X, and $F_n(x)$ the empirical distribution function determined by a sample $X_1, X_2, \cdots, X_n$. It is well known that the probability $P_n(\epsilon)$ of $F(x)$ being everywhere majorized by $F_n(x) + \epsilon$ is independent of $F(x)$. The present paper contains the derivation of an explicit expression for $P_n(\epsilon)$, and a tabulation of the 10%, 5%, 1%, and 0.1% points of $P_n(\epsilon)$ for $n = 5, 8, 10, 20, 40, 50$. For $n = 50$ these values agree closely with those obtained from an asymptotic expression due to N. Smirnov.

1. Introduction. Let X be a random variable with the continuous probability distribution function $F(x) = \text{Prob.}\{X \leq x\}$. An ordered sample $X_1 \leq X_2 \leq \cdots \leq X_n$ of X determines the empirical distribution function

$$F_n(x) = \begin{cases} 0 & \text{for } x < X_1, \\ \dfrac{k}{n} & \text{for } X_k \leq x < X_{k+1}, \qquad k = 1, 2, \cdots, n-1, \\ 1 & \text{for } X_n \leq x. \end{cases}$$

The function

$$F^+_{n,\epsilon}(x) = \min [F_n(x) + \epsilon, 1],$$

also determined by the sample, will be called an *upper confidence contour*. It is well known [2] that the probability

$$P_n(\epsilon) = \text{Prob.}\{F(x) \leq F^+_{n,\epsilon}(x) \text{ for all } x\}$$

of $F(x)$ being everywhere majorized by $F^+_{n,\epsilon}(x)$ is independent of the distribution $F(x)$. An expression for $P_n(\epsilon)$ in determinant form was given by A. Wald and

[1] Presented to the American Mathematical Society on April 28, 1951.
[2] Research under the sponsorship of the Office of Naval Research.