SOME EXTENSIONS OF THE IDEA OF BIAS

By H. R. van der Vaart[1]

Leiden University

1. Introduction and Summary. Laplace ([13], p. 44, lines 5 and 6), in his statement concerning the "milieu de probabilité", seems to have referred to a probability distribution of the *true* value of a certain quantity ("le véritable instant du phénomène"), or, as we would say at present, to a probability distribution of a certain parameter. Thereby he differs from the attitude adopted in most of the work discussed in the present paper. Yet, one might hold that he possessed the idea of *median-unbiased* estimators. At any rate, when applying his notions to what Todhunter ([26] p. 469, art. 875) calls a case of no practical value, Laplace ([13], p. 48, lines 11 and 12 from the bottom) virtually rejected the use of arithmetic means of observations. Judging from innumerable texts, one finds that after him emphasis has long been mainly on *mean-unbiasedness* (see, however, Pitman (]20], bottom of p. 215), who mentions the existence of bias in the sense that the probability that a certain mean-unbiased estimator is less than the parameter in question is $>\frac{1}{2}$). Yet it is hard to find the requirement of mean-unbiasedness justified in print (cf., Brown ([3], lines 6–8 of Section 3): the average of independent mean-unbiased estimates is consistent; Lehmann ([14], lines 4–10 from bottom of p. 588): mean-unbiasedness flows from his general concept in the case of a quadratic loss function; Birnbaum ([2], p. 32): mean-unbiasedness is merely a technically useful property of the classical estimators in the linear estimation problem, which, at least in the case of normal errors, could equally well or preferably be justified on the basis of median-unbiasedness), much harder, in fact, than to find warnings against the hope that much is gained if an estimator be mean-unbiased (cf., Kendall ([12], Vol. 2, Section 17.9); the examples provided by Girshick, Mosteller and Savage ([9], middle of p. 20), Halmos ([10], the end of p. 43), Savage ([23], bottom of p. 244); lack of invariance under certain transformations being stressed by Halmos ([10], bottom of p. 42), Brown ([3], lines 13–16 of Section 3), Fisher ([7], p. 143, line 13 from bottom)). All the same, much interesting work has been devoted to mean-unbiased estimators, some of it investigating the conversion of biased estimators into unbiased ones (e.g., Quenouille [21], Olkin and Pratt [17]), or deriving unbiased estimators *ab initio* (e.g. Tate [25]). It is not the purpose of this paper to provide a bibliography that is at all near completeness, but it is interesting that the last two references illustrate a statement, made by Schmetterer ([24], middle of p. 215), to the effect that a close connection exists between integral equations and linear operators on the one hand, and the theory of mean-unbiased estimators on the other. This

Received September 17, 1959; revised September 12, 1960.

[1] Research sponsored by the Office of Ordnance Research under Contract No. DA-36-034-ORD-1517 (RD) while the author held a visiting appointment at the Institute of Statistics, North Carolina State College, Raleigh, N. C.

suggests that part of the motivation for the research in this field is of a mathematical, rather than a statistical nature. This view seems to be corroborated by Fraser's statement ([8], lines 12–14 from bottom of p. 49) to the effect that median-unbiasedness does not seem to lend itself to the mathematical analysis needed to find minimum risk estimates, and hence has found little application.

The present paper seeks to extend the notion of unbiasedness (and the notion of bias) in a direction different from Lehmann [14] (who gave a definition within the framework of general decision theory), and from Brown [3] (who was primarily concerned with types of unbiasedness, among them median-unbiasedness, that are invariant "under simultaneous one-to-one transformations of the parameter and (its) estimate", or rather under simultaneous strictly monotone transformations of the parameter and its estimate), and from Peterson's [19] density-unbiasedness. It originated in work by the author [29], [30], on the estimation of the latent roots of certain matrices occurring in response surface theory. It had become clear that in this case it was of primary interest whether or not the *frequency of obtaining too small* (or too large, respectively) *estimates* would be *unduly large*. The present paper will make this notion more precise. Several types of bias (or of unbiasedness, respectively) will emerge, all of them clearly invariant in the sense of Brown. Median-unbiasedness will turn out to be a special case of this larger concept. Finally, certain seemingly unfamiliar properties of the sample median, of the product-moment correlation coefficient, and of Olkin and Pratt's function of the latter [17] will be proved and used to illustrate some of the concepts discussed.

2. Some new bias concepts. Let $\varphi(P)$ be a real valued function (a "parameter") defined on a set $\mathcal{P}$ of probability distributions P on a space $\mathcal{X}$ of points x. Let the real function $f(X)$ of the random variable X represent an estimator of $\varphi(P)$. When would one call an estimate $f(x)$ too small? A reasonable answer would be: if this estimate is smaller than a certain value (possibly depending on P) which is to be called the *comparing value*, and to be denoted by $\gamma(P)$; the selection of useful comparing values will be discussed after Definition 1. When would one call the frequency of obtaining too small estimates unduly large? A reasonable answer to this second question would be: if it is larger than it would have been if a different ("better") method of estimation would have been used, that is, if it is larger than it would have been with a different estimator, which is to called the *comparing estimator*, and to be denoted by $c(X)$. This tentative argument naturally leads to the concept described in

DEFINITION 1. *The estimator $f(X)$ of $\varphi(P)$ will be called negatively $\gamma(P)$-biased relative to the estimator $c(X)$ if*

$$(2.1) \qquad P[f(X) \leqq \gamma(P)] > P[c(X) \leqq \gamma(P)] \quad \textit{for each} \quad P \in \mathcal{P}.$$

By replacing the two $\leqq$-signs in (2.1) by $\geqq$-signs one obtains the definition of *positive $\gamma(P)$-bias*. Definition 1 has left the function $\gamma(P)$ unspecified: so this type of bias comprises as many varieties as there are choices of comparing values $\gamma(P)$. Whether, in a given estimation problem, a comparing value $\gamma(P)$ and a comparing estimator $c(X)$ can be chosen so as to provide a useful variety of relative $\gamma(P)$-bias, will depend on the nature of the problem. We shall indicate two examples in the next two paragraphs and refer the reader to point e and to the last sentence of point k in Section 5 for an additional one.

As a first example, consider the problem of estimating the coefficients of the canonical form of the second degree part of the equation for a quadratic response surface. It is well known that the signs of these coefficients are important since they determine the type (hyperbolic, ellipsoidal, etc.) of the surface. So here $\gamma(P) = 0$ suggests itself as a comparing value, and if for all possible quadratic response surfaces with positive true values of the canonical coefficients a certain estimator $f(X)$ of the smallest coefficient is negatively zero-biased relative to a comparing estimator $c(X)$, it is quite clear that in this respect the comparing estimator is better than $f(X)$ (although it may be worse in some other respect; cf., for instance point d of Section 5).

The second example is connected with the concept of median-bias. Suppose that an estimator $g(X)$ of the "parameter" $\varphi(P)$ exists which satisfies the condition

$$\text{(2.2)} \qquad \text{Med}_P\, g(X) = \varphi(P) \quad \text{for each} \quad P \,\varepsilon\, \mathcal{P};$$

if more than one function $g(X)$ satisfies (2.2), just choose one of them; $\text{Med}_P\, g(X)$ denotes (one of) the median(s) of $g(X)$ under the probability distribution P on the space $\mathfrak{X}$; Lehmann [15, p. 80–83], pointing out a simple connexion between median-unbiased estimators and confidence intervals, gives a condition on $\mathcal{P}$, which guarantees that one and only one estimator $g(X)$ will satisfy (2.2). Now in (2.1) choose $\gamma(P) = \varphi(P)$, $c(X) = g(X)$, then because of (2.2) condition (2.1) becomes

$$\text{(2.3)} \quad P[f(X) \leqq \varphi(P)] > P[g(X) \leqq \varphi(P)] = P[g(X) \leqq \text{Med}_P\, g(X)] \geqq \tfrac{1}{2}.$$

Now, on one hand (2.3) means that, as an estimator of $\varphi(P)$, $f(X)$ is negatively $\varphi(P)$-biased relative to the estimator $g(X)$, on the other hand, under a certain condition, (2.3) entails the inequality $\text{Med}_P\, f(x) < \varphi(P)$, which means that, as an estimator of $\varphi(P)$, $f(X)$ is negatively median-biased (the above-mentioned condition being that not only $P[f(X) \leqq \varphi(P)] - \frac{1}{2} > 0$, which follows from (2.3), but also $P[f(X) \leqq \varphi(P)] - \frac{1}{2} > P[f(X) = \varphi(P)]$, which is certainly true if for each $P \,\varepsilon\, \mathcal{P}$ the distribution of $f(X)$ is continuous). Thus the concept of *relative* $\gamma(P)$*-bias* described in our Definition 1 is seen to *generalize* the concept of *median-bias*.

In certain contexts it is useful to admit as comparing values all values $\varphi(Q) \,\varepsilon\, \varphi(\mathcal{P})$, i.e., all possible values of the parameter. This leads to the concept described in

Definition 2. *The estimator* $f(X)$ *of* $\varphi(P)$ *will be called negatively distribution-*

biased relative to the estimator $c(X)$ if

$$(2.4)\quad P[f(X) \leqq \varphi(Q)] \geqq P[c(X) \leqq \varphi(Q)] \quad \textit{for each} \quad P \varepsilon \mathcal{P},\ Q \varepsilon \mathcal{P},$$

and the inequality is strict for at least one pair (P, Q).

By replacing the two $\leqq$-signs in (2.4) by $\geqq$-signs one obtains the definition of *positive distribution-bias*. A close connexion clearly exists between the condition for the estimator $f(X)$ being negatively distribution-biased with respect to the estimator $c(X)$, and the condition for the random variable $f(X)$ being stochastically smaller than the random variable $c(X)$; as to the latter concept see Mann and Whitney ([16], line 3 of Section 2).

Note that, whereas the definitions of $\gamma(P)$*-unbiasedness* and *median-unbiasedness* are self-evident (in (2.1) and (2.3) replace $>$ by $=$), the definition of distribution-unbiasedness presents difficulties. On one hand, it seems impracticable to define distribution-unbiasedness of $f(X)$ relative to $c(X)$ otherwise than as $f(X)$ and $c(X)$ having the same distribution function. On the other hand, to call $f(X)$ distribution-unbiased relative to $c(X)$ only if $f(X)$ and $c(X)$ have the same distribution function for each $P \varepsilon \mathcal{P}$ (a condition obtained if in (2.4) the $\geqq$-sign is replaced by $=$) is unsatisfactory, because estimators will exist which are neither biased nor unbiased in this sense. Hence we will not attempt a definition of distribution-unbiasedness.

One more point has to be mentioned. One might think that it should be possible to make the rather vague notion of negative bias as an unduly large frequency of obtaining too small estimates more precise without introducing the concept of comparing estimators: one might endeavour to define the frequency of obtaining too small estimates (i.e., estimates smaller than the comparing value $\gamma(P)$) as being unduly large if the probability of obtaining estimates $\leqq \gamma(P)$ would be large as compared with the probability of obtaining estimates $\geqq \gamma(P)$; that is to say, if the ratio $P[f(X) \leqq \gamma(P)]/P[f(X) \geqq \gamma(P)]$ would be large, $>k$, say. Thus, in order to make this approach work, we should have to decide upon the value of k. Decisions of this kind would be to a large extent arbitrary. Upon a moment's reflection it turns out that about the only natural way to find "plausible" values of k consists in considering the value of the above-mentioned ratio of probabilities when another estimator, $c(X)$ say, is substituted for $f(X)$. Therewith our comparing estimator has proved indispensable.

3. A remark on terminology. To avoid confusion we note that for median-(un)-bias(edness) and mean-(un)bias(edness) (cf., Brown [3], p. 583) other terms may be substituted. For example, Eisenhart and Martin [6] use the term "downward bias in the probability sense" instead of "negative median-bias", and in a personal communication to the author (June, 1958) Eisenhart uses "probability-wise unbiasedness" instead of "median-unbiasedness".

As is well known, $f(X)$ is a mean-unbiased estimator of $\varphi(P)$ if

$$(3.1)\qquad \mathcal{E}_P f(X) = \varphi(P) \quad \text{for each} \quad P \varepsilon \mathcal{P}.$$

Instead of "mean-bias" Eisenhart and Martin [6] use "bias in the mean-value sense", in the above-mentioned communication to the author Eisenhart uses "on-the-average bias", and the present author personally prefers *expectation-bias* (similarly expectation-unbiasedness), since bias and unbiasedness of an estimator are properties of its theoretical distribution, and statistical usage tends to substitute "expectation" for "mean" in connexion with theoretical distributions.

4. A lemma. A very simple lemma, which nevertheless is a useful tool in proving that certain estimators are biased in the sense discussed in Section 2, is

LEMMA 1. *Whether the random variables* $T = t(X)$ *and* $U = u(X)$ *are independent or dependent, if*

$$P[U \geqq v] = 1, \tag{4.1}$$

then

$$P[T \leqq \tau] \geqq P[T + U \leqq \tau + v]. \tag{4.2}$$

A necessary and sufficient condition for the equality sign to hold in (4.2) *is*

$$P[T + U > \tau + v) \cap (T \leqq \tau)] = 0. \tag{4.3}$$

PROOF. Let $U^* = U - v$ and $T^* = T - \tau$; the proof is then immediate from a sketch in the (T^*, U^*)-plane.

Although joint distributions of T and U satisfying (4.3) may be rather uncommon, it is evident that (4.3) cannot be proved or disproved from the conditions of the lemma alone. Two extra conditions, each of them sufficient for (4.3) not to hold, are

(α) each (measurable) set in the half plane $U > v$ in (T, U)-space has positive probability,

(β) (implied by α:) Some set $[(U > v_0) \cap (\tau \geqq T > \tau_0)]$ with $\tau_0 + v_0 > \tau + v$, $\tau_0 < \tau$, has positive probability; this condition is satisfied for instance if U and T are independent and $P[\tau \geqq T > \tau_0] > 0$, $P[U > v_0] > 0$.

While Lemma 1 may serve to prove negative bias of the types discussed in Section 2, positive bias may be derived from Lemma 1′, which is obtained from Lemma 1 by reversing all six inequality-signs in Lemma 1, except the second inequality-sign in (4.2).

5. Supplementary remarks and examples.

(a). From Definition 1 it follows easily that relative $\gamma(P)$-bias is *transitive*: if $g_1(X)$, $g_2(X)$, and $g_3(X)$ are estimators of $\varphi(P)$, and if $g_1(X)$ is negatively $\gamma(P)$-biased relative to $g_2(X)$, and $g_2(X)$ is negatively $\gamma(P)$-biased relative to $g_3(X)$, then $g_1(X)$ is negatively $\gamma(P)$-biased relative to $g_3(X)$. Hence it is possible to arrange any number of estimators according to degree of negative $\gamma(P)$-bias; in the above case we would have: $g_1(X) > g_2(X) > g_3(X)$ (where $>$ would mean: "has more negative $\gamma(P)$-bias than"; the value of the difference $P[g_1(X) \leqq \gamma(P)] - P[g_2(X) \leqq \gamma(P)]$ would be a useful measure of how much more $\gamma(P)$-biased $g_1(X)$ is than $g_2(X)$). Although this arrangement would not without further consideration permit the conclusion that $g_3(X)$ is a better estimator than $g_2(X)$, and $g_2(X)$ a better estimator than $g_1(X)$, it is clear that in general, if $\gamma(P) < \varphi(P)$ for each $P \varepsilon \mathcal{P}$, one would tend to consider estimators to be worse as they are more biased in the sense of negative $\gamma(P)$-bias. In the same vein, if $\gamma(P) > \varphi(P)$ for each $P \varepsilon \mathcal{P}$, one would tend to consider estimators to be worse as they are more biased in the sense of positive $\gamma(P)$-bias.

(b). Through the concept of $\gamma(P)$-bias the notions of *bias* and of *inefficiency* merge into each other: if the estimator $f_1(X)$ of $\varphi(P)$ is distributed $N(\varphi(P), 2\sigma^2)$ and the estimator $f_2(X)$ is distributed $N(\varphi(P), \sigma^2)$, then $f_1(X)$ is both negatively $(\varphi(P) - k\sigma)$-biased $(k > 0)$ relative to $f_2(X)$ and less efficient than $f_2(X)$ (Note that in the examples b, c, d the parameter σ is assumed to be known).

(c). It should not be surprising (though it is worth while noting) that *different criteria* of (un)*bias*(edness) may be incompatible (see also point f below). Thus, even an expectation-biased and median-biased estimator like $f_3(X)$, distributed $N(\varphi(P) - \frac{1}{2}\sigma, \sigma^2)$, would from the point of view of negative $(\varphi(P) - 2\sigma)$-bias, say, be better (i.e., less biased) than the expectation-unbiased and median-unbiased estimator $f_1(X)$, distributed $N(\varphi(P), 2\sigma^2)$. This pair of estimators is interesting for yet another reason: $f_3(X)$ has also a smaller mean square error than has $f_1(X)$: $\mathcal{E}[f_3(X) - \varphi(P)]^2 < \mathcal{E}[f_1(X) - \varphi(P)]^2$. Thus the idea of $\gamma(P)$-bias, in cases where a comparing value $\gamma(P)$ naturally suggests itself, may help to bridge the gap which often exists between the requirements of least bias and least mean square error.

(d). However, the requirement of least $\gamma(P)$-bias will not always agree so well with other criteria for "good" estimators. For instance, the above-mentioned estimator $f_2(X)$ of $\varphi(P)$, distributed $N(\varphi(P), \sigma^2)$, is negatively $(\varphi(P) - k\sigma)$-biased $(k > 0)$ relative to the estimator $f_4(X)$, distributed $N(\varphi(P) + \sigma, \sigma^2)$. So, according to point a above, one would tend to consider $f_4(X)$ a better estimator. Yet, $f_4(X)$ has a number of undesirable features: for instance, it is not only positively expectation-biased and median-biased, it has also greater mean square error than $f_2(X)$. On closer inspection another thing turns out to be wrong with $f_4(X)$ as an estimator of $\varphi(P)$: for any $k > 0$ it is positively $(\varphi(P) + k\sigma)$-biased relative to $f_2(X)$; so, according to the last sentence of point a above, $f_4(X)$ is worse than $f_2(X)$ in this respect. Thus, here is a simple example where the probability of obtaining too small estimates has been corrected at the expense of enlarging the probability of obtaining too large estimates. In certain contexts this may be all right, in other contexts it may be undesirable: if negative $\gamma^l(P)$-bias $(\gamma^l(P) < \varphi(P))$ and positive $\gamma^r(P)$-bias $(\gamma^r(P) > \varphi(P))$ are about equally undesirable features, both have to be kept as small as possible in some sense. This remark points out the relationship between our concept of (relative) $\gamma(P)$-*bias* and two *other criteria* for "good" estimators:

(1) criterion 3 of L. J. Savage [23, p. 224], according to which an estimator $g_1(X)$ is called better than an estimator $g_2(X)$ if $P[g_1(X) < \gamma_1] + P[g_1(X) > \gamma_2] \leqq P[g_2(X) < \gamma_1] + P[g_2(X) > \gamma_2]$ for every $\gamma_1 \leqq \varphi(P)$, $\gamma_2 \geqq \varphi(P)$, $P \varepsilon \mathcal{P}$ (with strict inequality for some γ_1, γ_2, and some P);

(2) the approach of A. Birnbaum[2] ([2], pp. 113 seq.), who uses as a criterion the behavior of the function $a(\gamma, P; g)$ for $P \varepsilon \mathcal{P}, \gamma \varepsilon \varphi(\mathcal{P})$: $a(\gamma, P; g) = P[g(X) \leqq \gamma]$ if $\gamma < \varphi(P)$ and $a(\gamma, P; g) = P[g(X) \geqq \gamma]$ if $\gamma > \varphi(P)$.

(e). Summarizing, it seems fitting to note that, if it is of primary importance to avoid an unduly large frequency of obtaining too small estimates, then the concept of (relative) negative $\gamma(P)$-bias, $\gamma(P) < \varphi(P)$, leads to a useful criterion for good estimators. Similarly, positive $\gamma(P)$-bias is a useful concept if it is important to avoid too large estimates. Examples of this situation are the estimation of latent roots (discussed at the end of Section 1, and in the first example after Definition 1 in Section 2), and the estimation of the correlation coefficient (to be discussed in points i and k below): in both these cases the comparing value of interest is zero. However, if it is important to avoid errors of under-estimation and over-estimation at the same time, then criterion 3 of Savage [23, p. 224] and the approach by Birnbaum, although the latter leads only to a partial ordering of estimators, provide more natural criteria than does the concept of distribution-bias (Definition 2 in Section 2)–though this term is useful in that it permits a succinct statement of certain results.

(f). Next, we will give examples of an expectation-unbiased estimator which is median-biased (variance), of a median-unbiased estimator which is expectation-biased (median), of a negatively expectation-biased estimator which is positively median-biased (correlation-coefficient), of median-bias becoming less when an estimator is corrected for expectation-bias (variance), and of median-bias becoming worse when an estimator is corrected for expectation-bias (correlation-coefficient).

(g). Let $X_1, X_2, \cdots, X_n$ be an n-fold sample from a normal distribution $N(\mu, \sigma^2)$. Then

$$S^2 = (n - 1)^{-1} \cdot \sum_{i=1}^{n} (X_i - \bar{X})^2$$

is an expectation-unbiased estimator of the *variance* σ^2. However, S^2 is negatively median-biased: we will show that

$$P[S^2 \leqq \sigma^2] > \tfrac{1}{2}. \tag{5.1}$$

Note that

$$P[S^2 \leqq \sigma^2] = 1 - Q(n - 1 \mid n - 1) = \gamma\{\tfrac{1}{2}(n - 1), \tfrac{1}{2}(n - 1)\} \,/\, \Gamma\{\tfrac{1}{2}(n - 1)\}, \tag{5.2}$$

where the function $Q(\chi^2 \mid \nu)$ is defined by Pearson and Hartley [18, p. 122] and $\gamma(\alpha, x) = \int_0^x e^{-t}t^{\alpha-1}\, dt$, (cf., [1], Vol. 2, p. 133). Now, as

$$e^{-\alpha}\alpha^\alpha > \int_\alpha^{\alpha+1} e^{-t}t^\alpha\, dt,$$

[2] I want to thank the referee for drawing my attention to Mr. Birnbaum's paper and to the connection between his approach and mine.

we have

$$\alpha \int_0^\alpha e^{-t} t^{\alpha-1}\, dt = e^{-\alpha}\alpha^\alpha + \int_0^\alpha e^{-t} t^\alpha\, dt > \int_0^{\alpha+1} e^{-t} t^\alpha\, dt,$$

whence

$$\alpha\gamma(\alpha, \alpha) > \gamma(\alpha + 1, \alpha + 1),$$

so

$$\gamma(\alpha, \alpha)/\Gamma(\alpha) > \gamma(\alpha + 1, \alpha + 1)/\Gamma(\alpha + 1) \quad \text{for any} \quad \alpha > 0.$$

Hence, for any β, $0 < \beta \leqq 1$, $\lim_{n\to\infty} \gamma(\beta + n, \beta + n)/\Gamma(\beta + n) = L(\beta) \geqq 0$ exists and

$$(5.3) \qquad \gamma(B + n, \beta + n)/\Gamma(\beta + n) > L(\beta)$$

for any integer $n > 0$. From the asymptotic expression for $\gamma(\alpha + 1, \alpha + (2\alpha)^{\frac{1}{2}}y)$, given by Tricomi [27, p. 144, eq. (27)], one can derive that

$$(5.4) \qquad \gamma(\alpha, \alpha)/\Gamma(\alpha) = \tfrac{1}{2} + [3(2\pi\alpha)^{\frac{1}{2}}]^{-1} + O(\alpha^{-1}),$$

which shows that in (5.3) $L(\beta) = \frac{1}{2}$, independent of β. Therefore,

$$(5.5) \qquad \gamma(\alpha, \alpha)/\Gamma(\alpha) > \tfrac{1}{2} \quad \text{for any} \quad \alpha > 0,$$

which, together with (5.2), proves (5.1).

Equation (5.2) permits the calculation of $P[S^2 \leqq \sigma^2]$ from the table of χ^2 by Pearson and Hartley [18, p. 122]. For $n - 1 = 1, 2, 3$ one finds $P[S^2 \leqq \sigma^2] =$ 0.683, 0.632, 0.608. For $n - 1 \geqq 4$ the asymptotic expression (5.4) turns out to yield results which are accurate to 3 significant decimal places!

The median-bias of S^2, hence of S, is of interest in quality control, cf., Eisenhart [5]. The present author wants to thank Mr. Eisenhart for his kind letter (of June, 1958), in which he mentioned this interesting article as well as the abstract [6], where six different estimators of σ are investigated as to their median-bias, and the report [4], where among other things a table for $P[s \leqq \sigma]$ is given.

(h). Let $X_1, X_2, \cdots, X_{2m+1}$ be an *odd-sized* sample from a univariate distribution with continuous distribution function $F(x)$, for which $dF(x)/dx$ is positive in one (finite or infinite) x-interval. Rearranging the $2m + 1$ values in the sample, use the notation $X^{(1)} \leqq X^{(2)} \leqq \cdots \leqq X^{(2m+1)}$. Under the conditions stated the occurrence of equality-signs has probability zero; $X^{(m+1)}$ is the sample *median*. Let G denote the inverse (defined for $0 < F < 1$) of the function F, so that $G(\frac{1}{2})$ is the median of the distribution considered.

From the well-known formula

$$P[X^{(m+1)} \leqq \gamma] = [B(m + 1, m + 1)]^{-1} \cdot \int_{-\infty}^{\gamma} [F(y)]^m \cdot [1 - F(y)]^m\, dF(y)$$

it follows immediately that

$$(5.6) \quad P[X^{(m+1)} \leqq G(\tfrac{1}{2})] = [B(m + 1, m + 1)]^{-1} \cdot \int_0^{\frac{1}{2}} F^m(1 - F)^m\, dF = \tfrac{1}{2}.$$

Hence the sample median $X^{(m+1)}$ is a *median-unbiased* estimator of the median $G(\frac{1}{2})$, whether $F(x)$ represents a skew distribution or a symmetric one.

On the other hand, it is very simple to define classes of continuous distribution functions such that the sample median $X^{(m+1)}$ is an expectation-biased estimator of the median $G(\frac{1}{2})$. For we have

$$
\begin{aligned}
B(m+1, m+1)\cdot[\mathcal{E}X^{(m+1)} - G(\tfrac{1}{2})] &= \int_{-\infty}^{+\infty} [y - G(\tfrac{1}{2})] \\
\cdot[F(y)]^m[1 - F(y)]^m \, dF(y) &= \int_0^1 [G(F) - G(\tfrac{1}{2})]\cdot F^m(1-F)^m \, dF \\
&= \int_{-\frac{1}{2}}^{+\frac{1}{2}} [G(\tfrac{1}{2} + h) - G(\tfrac{1}{2})]\cdot(\tfrac{1}{4} - h^2)^m \, dh \\
&= \int_0^{\frac{1}{2}} \{[G(\tfrac{1}{2} + h) - G(\tfrac{1}{2})] - [G(\tfrac{1}{2}) - G(\tfrac{1}{2} - h)]\}\cdot(\tfrac{1}{4} - h^2)^m \, dh.
\end{aligned} \tag{5.7}
$$

Hence a sufficient condition for the sample median being a positively expectation-biased estimator of the median is that $G(\frac{1}{2} + h) - G(\frac{1}{2}) > G(\frac{1}{2}) - G(\frac{1}{2} - h)$, $0 < h < \frac{1}{2}$, which describes a certain type of skewness of the distribution function $F(x)$.

(i). Let $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$ be an n-fold sample from a bivariate normal distribution with *correlation coefficient* ρ. Define the sample correlation coefficient R[3] in the usual way by

$$R = [\textstyle\sum (X_i - \bar{X})(Y_i - \bar{Y})]/[\sum (X_i - \bar{X})^2 \sum (Y_j - \bar{Y})^2]^{\frac{1}{2}}.$$

It is well known (cf., Kendall ([12], Vol. 1, p. 344, eq. (14.55)), Romanovsky ([22], p. 42, eq. (128)), and reference ([1], Vol. 1, p. 59, eq. (10), and p. 114, eq. (1))) that

$$
\begin{aligned}
\mathcal{E}R = \rho\cdot g(n, \rho^2) &= \rho\cdot\{\Gamma(\tfrac{1}{2}n)\}^2\cdot\{\Gamma[\tfrac{1}{2}(n-1)]\cdot\Gamma[\tfrac{1}{2}(n+1)]\}^{-1} \\
&\qquad\cdot F(\tfrac{1}{2}, \tfrac{1}{2}; \tfrac{1}{2}(n+1); \rho^2) \\
&= \rho\cdot\Gamma(\tfrac{1}{2}n)\cdot\{\Gamma(\tfrac{1}{2})\cdot\Gamma[\tfrac{1}{2}(n-1)]\}^{-1}\cdot\int_0^1 t^{\frac{1}{2}-1}(1-t)^{\frac{1}{2}n-1}(1-t\rho^2)^{-\frac{1}{2}} \, dt,
\end{aligned} \tag{5.8}
$$

where Euler's integral representation for the hypergeometric function has been applied. From $(1 - t\rho^2)^{-\frac{1}{2}} < (1 - t)^{-\frac{1}{2}}$ for any $\rho^2 \neq 1$, $t \neq 0$, it follows that for any $\rho^2 \neq 1$ the integral in the last member of (5.8) is less than

$$\int_0^1 t^{\frac{1}{2}-1}(1-t)^{\frac{1}{2}(n-1)-1} \, dt = B(\tfrac{1}{2}, \tfrac{1}{2}(n-1)),$$

whence in (5.8) $g(n, \rho^2) < 1$ for any $\rho^2 \neq 1$, so that $|\mathcal{E}R| < |\rho|$: R is a *negatively expectation-biased* estimator of ρ *if* $\rho > 0$, a *positively expectation-biased* estimator *if* $\rho < 0$.

[3] Note that here R is not the multiple correlation coefficient: capitals are used throughout the paper in order to denote random variables.

In order to investigate possible median-bias of R as an estimator of ρ, use formula (25) of Hotelling [11, p. 200] (note that Hotelling's n stands for our $n - 1$), by which

$$(5.9)\quad \begin{aligned} P[R \geqq \rho] &= (n-2)\cdot\Gamma(n-1)\cdot(2\pi)^{-\frac{1}{2}}\cdot\{\Gamma(n-\tfrac{1}{2})\}^{-1}\cdot(1-\rho^2)^{\frac{1}{2}(n-1)} \\ &\quad \cdot\int_\rho^1 (1-r^2)^{\frac{1}{2}(n-4)}(1-\rho r)^{-\frac{1}{2}(2n-3)}F(\tfrac{1}{2},\tfrac{1}{2};n-\tfrac{1}{2};\tfrac{1}{2}(1+\rho r))\,dr. \end{aligned}$$

By means of the substitution $(r - \rho)/(1 - \rho r) = y$ the second member of (5.9) after some patient algebra reduces to

$$(5.10)\quad \begin{aligned} &\frac{(n-2)\Gamma(n-1)}{\sqrt{2\pi}\,\Gamma(n-\frac{1}{2})}\int_0^1 (1-y^2)^{\frac{1}{2}(n-4)}(1+\rho y)^{\frac{1}{2}} \\ &\qquad\qquad \cdot F\left(\tfrac{1}{2},\tfrac{1}{2};n-\tfrac{1}{2};1-\frac{1-\rho^2}{2(1+\rho y)}\right)dy. \end{aligned}$$

As ρ increases from 0 to 1, the integrand of (5.10) increases with ρ for any $y > 0$. Since $P[R \geqq \rho] = \frac{1}{2}$ if $\rho = 0$, this means that $P[R \geqq \rho] > \frac{1}{2}$ if $\rho > 0$: R is a *positively median-biased* estimator of ρ *if* $\rho > 0$. The same argument yields the result that R is a *negatively median-biased* estimator of ρ *if* $\rho < 0$. In fact, $P[R \leqq \rho]$ equals the second member of (5.9) after integration from ρ to 1 has been replaced by integration from -1 to ρ; in the expression thus obtained one may replace ρ by $-|\rho|$ if $\rho < 0$; from the elementary substitution $r = -z$ in the resulting integral it then follows immediately that if $\rho < 0$, $P[R \leqq \rho]$ exactly equals the second member of (5.9), hence equals (5.10), if only $|\rho|$ is substituted for ρ; hence $P[R \leqq \rho] > \frac{1}{2}$ if $\rho < 0$.

This example is interesting since it shows that the contention made by Tschuprow [28, p. 116], to the effect that the estimator R *systematically underrates* ρ, is dubious in that it may be taken to mean that R more frequently than not underrates ρ—which is not true as R more frequently than not overrates $\rho(\rho > 0)$.

(j). With S^2, see under point g, compare $*S^2$ as an estimator of σ^2: $*S^2 = n^{-1}\cdot\sum (X_i - \bar{X})^2$. Evidently $*S^2 = S^2(1 - n^{-1})$, hence $\sigma^2 > \text{Med}\, S^2 > \text{Med}\, *S^2$. At the same time $\sigma^2 = \mathcal{E}S^2 > \mathcal{E}*S^2$. So when the estimator $*S^2$ is replaced by S^2, its expectation-bias is corrected, and its median-bias becomes less. Unfortunately, such a state of affairs is not universal as is shown by the next example.

(k). With R, see under point i, compare $*R$ as an estimator of ρ:

$$(5.11)\qquad *R = R\cdot F(\tfrac{1}{2},\tfrac{1}{2};\tfrac{1}{2}(n-1);1-R^2),$$

cf., Olkin and Pratt ([17], p. 202, eq. (2.3)). The second member of (5.11) is a strictly increasing function of R, cf., [17, Section 2.2]. Hence

$$(5.12)\quad \text{Med}_\rho\, *R = (\text{Med}_\rho R)\cdot F\{\tfrac{1}{2},\tfrac{1}{2};\tfrac{1}{2}(n-1);1-(\text{Med}_\rho\, R)^2\}.$$

As $|\text{Med}_\rho\, R| < 1$ if $|\rho| < 1$, the hypergeometric series in (5.12) is easily seen to be strictly larger than 1 if $|\rho| < 1$. Therefore, if $0 < \rho < 1$, $\text{Med}_\rho\, R > \rho > 0$

(see under point i) and $\text{Med}_\rho\, {}^*R > \text{Med}_\rho\, R$; if $0 > \rho > -1$, $\text{Med}_\rho\, R < \rho < 0$ (see under point i) and $\text{Med}_\rho\, {}^*R < \text{Med}_\rho\, R$.

So substituting $^*R$ for R as an estimator of ρ corrects expectation-bias, but makes *median-bias worse*. Finally, it is evident from (5.11) that $^*R$ and R are *zero-unbiased* with respect to each other.

REFERENCES

[1] BATEMAN MANUSCRIPT PROJECT, *Higher Transcendental Functions*, Vol. 1; Vol. 2; McGraw-Hill, New York, 1953.

[2] ALLAN BIRNBAUM, "A unified theory of estimation I," *Ann. Math. Stat.*, Vol. 32, (1961), pp. 112–135.

[3] GEORGE W. BROWN, "On small-sample estimation," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 582–585.

[4] JOSEPH M. CAMERON, "Tables for computing confidence limits for σ," *SEL Note* 54-11, Statistical Engineering Laboratory, National Bureau of Standards, 1954.

[5] CHURCHILL EISENHART, "Probability center lines for standard deviation and range charts," *Industrial Quality Control*, Vol. 6 (1949), pp. 24–26.

[6] CHURCHILL EISENHART AND CELIA S. MARTIN, "The relative frequencies with which certain estimators of the standard deviation of a normal population tend to underestimate its value," *Ann. Math. Stat.*, Vol. 19 (1948), p. 600, Abstract.

[7] SIR RONALD A. FISHER, *Statistical Methods and Scientific Inference*, Hafner Pub., New York, 1956.

[8] D. A. S. FRASER, *Nonparametric Methods in Statistics*, John Wiley and Sons, New York, 1957.

[9] M. A. GIRSHICK, FREDERICK MOSTELLER, AND L. J. SAVAGE, "Unbiased estimates for certain binomial sampling problems with applications," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 13–23.

[10] PAUL R. HALMOS, "The theory of unbiased estimation," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 34–43.

[11] HAROLD HOTELLING, "New light on the correlation coefficient and its transforms," *J. Roy. Stat. Soc.*, *Ser.* B, Vol. 15 (1953), pp. 193–232.

[12] MAURICE G. KENDALL, *The Advanced Theory of Statistics*, Vol. 1, 3rd ed., 1947; Vol. 2, 2nd ed., 1948, Griffin, London.

[13] P. S. DE LAPLACE, "Mémoire sur la probabilité des causes par les événements," *Oeuvres complètes de Laplace*, Vol. 8, pp. 27–65, Gauthier-Villars, Paris, 1891; (originally published in: *Mémoires de l'Académie royale des Sciences de Paris* (Savants étrangers), Vol. 6 (1774), pp. 621 seq., par M. de la Place).

[14] E. L. LEHMANN, "A general concept of unbiasedness," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 587–592.

[15] E. L. LEHMANN, *Testing Statistical Hypotheses*, John Wiley and Sons, New York and Chapman & Hall, London, 1959.

[16] H. B. MANN AND D. R. WHITNEY, "On a test of whether one of two random variables is stochastically larger than the other," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 50–60.

[17] INGRAM OLKIN AND JOHN W. PRATT, "Unbiased estimation of certain correlation coefficients," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 201–211.

[18] E. S. PEARSON AND H. O. HARTLEY, *Biometrika Tables for Statisticians*, Vol. 1, Cambridge University Press, Cambridge, 1954.

[19] RAYMOND P. PETERSON, "Density unbiased point estimates," *Ann. Math. Stat.*, Vol. 25 (1954), p. 398–401.

[20] E. J. G. PITMAN, "The 'closest' estimates of statistical parameters," *Proc. Camb. Philos. Soc.*, Vol. 33, (1937), pp. 212–222.

[21] M. H. QUENOUILLE, "Notes on bias in estimation," *Biometrika*, Vol. 43 (1956), pp. 353–360.

[22] V. ROMANOWSKY, "On the moments of standard deviations and of correlation coefficient in samples from normal population," *Metron*, Vol. 5, no. 4 (December 31, 1925), pp. 31–46.

[23] LEONARD J. SAVAGE, *The Foundations of Statistics*, John Wiley and Sons, New York, 1954.

[24] LEOPOLD SCHMETTERER, *Einführung in die Mathematische Statistik*, Springer, Wien, 1956.

[25] R. F. TATE, "Unbiased estimation: functions of location and scale parameters," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 341–366.

[26] I. TODHUNTER, *A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace*, Chelsea, New York, 1949.

[27] F. G. TRICOMI, "Asymptotische Eigenschaften der unvollständigen Gammafunktion," *Mathematische Zeitschrift*, Vol. 53 (1950), pp. 136–148; cf., *Math. Rev.*, Vol. 13 (1952), p. 553.

[28] A. A. TSCHUPROW, *Principles of the Mathematical Theory of Correlation* (transl. by M. Kantorowitsch.), Hodge, London, 1939.

[29] H. R. VAN DER VAART, "Some results on the probability distribution of the latent roots of a symmetric matrix of continuously distributed elements, and some applications to the theory of response surface estimation," Institute of Statistics, University of North Carolina Mimeo Series, No. 189 (1958), 1 + 40 p .

[30] H. R. VAN DER VAART, "On certain types of bias in current methods of response surface estimation," *Bull. Inst. Internat. Stat.* Vol. 37 (3) (1960), pp. 191-203.