Some General Remarks on the Definition of the Concept of Biased Estimators

Prepared Under Contract No. DA-36-034-ORD-1517 (RD)

(Experimental Designs for Industrial Research)

by

H. R. van der Vaart

Institute of Statistics

Mimeo Series No. 178

July, 1957

SOME GENERAL REMARKS ON THE DEFINITION OF THE CONCEPT OF BIASED ESTIMATORS

by

H. R. van der Vaart
Institute of Statistics, University of North Carolina
Raleigh, North Carolina, U. S. A.
and
Leiden University, Netherlands

1. Statement of the problem

Suppose a parameter μ is estimated by a statistic (a random variable) $\underline{m}$, both of them real. To avoid non-essential complications $\underline{m}$ will be supposed throughout to be distributed continuously. As is well known, $\underline{m}$ is called a

$$\left\{\begin{array}{c}\text{positively biased}\\ \text{unbiased}\\ \text{negatively biased}\end{array}\right\} \quad \text{estimator of } \mu \text{ if} \quad \left\{\begin{array}{c}E(\underline{m}) > \mu\\ E(\underline{m}) = \mu\\ E(\underline{m}) < \mu\end{array}\right\}.$$

These various cases may be verbally described as follows (cf. for instance van der Waerden, 1957, p. 29):

$$\left\{\begin{array}{l}\text{a positively biased}\\ \text{an unbiased}\\ \text{a negatively biased}\end{array}\right\} \quad \text{estimator of } \mu \text{ yields an estimated value}$$

$$\text{which is} \quad \left\{\begin{array}{l}\underline{\text{on the average}} \text{ larger than}\\ \underline{\text{on the average}} \text{ equal to}\\ \underline{\text{on the average}} \text{ less than}\end{array}\right\} \quad \text{the true value.}$$

However, while this may frequently be an adequate description of what an experimenter or a statistician has in mind when he resorts to the terms biased or unbiased, it will not always meet his requirements. For instance, one may want to know not so much whether a certain estimator will be right *on the average*, as whether, say, the *frequency of obtaining too small estimates* will be *unduly large* (an example of this situation is afforded by the estimation of the latent roots of a determinant in connection with the problem of estimating the type of response surfaces). In this connection the remark made by Snedecor (1948, p. 8) with regard to the term unbiased: "In sampling from certain symmetric populations, it may be said that estimates made from unbiased samples are as likely to be in excess of the population value as in defect", is interesting as he apparently felt the need of adding something to the usual "true on the average".

This paper will explore a few alternative approaches to the general concept of bias together with the interrelations between various definitions given. For convenience' sake, only negative bias will be considered in most cases, the treatment of positive bias being then self-evident.

2. Expectation-bias and median-bias

Definition 1: The random variable $\underline{m}$ will be said to be a negatively expectation-biased estimator of μ if

$$E(\underline{m}) < \mu \qquad (2,1)$$

This is the usual definition of bias.

Definition 2: The random variable $\underline{m}$ will be said to be a negatively median-biased estimator of μ if it satisfies one of the equivalent conditions:

$$\text{Med}(\underline{m}) < \mu; \quad P(\underline{m} \leq \mu) > \frac{1}{2}; \quad P(\underline{m} \leq \mu) > P(\underline{m} \geq \mu) \qquad (2,2)$$

Replacing the three inequality signs $<$, $>$, $>$ in (2,2) by =, =, = or by $>$, $<$, $<$, one obtains the definition of median-unbiased, or positively median-biased estimators, respectively.*

* After this paper had been written, the author read the recent book: "Nonparametric methods in statistics" (Wiley, 1957) by D. A. S. Fraser, which on p. 49 defines the concept of median-unbiased estimators without, however, giving any further developments.

Because of the fact that for many probability distributions the expected value and the median are unequal, these two definitions are not equivalent in general: an expectation-unbiased estimator may be median-biased (a), a median-unbiased estimator may be expectation-biased (b), or even a negatively expectation-biased estimator may be positively median-biased (c). For example, $s^2 = (n - 1)^{-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$, the well-known estimator of σ^2 from samples of size n from normal distributions $N(\mu, \sigma^2)$, belongs to class (a), the median of a sample of size 2m + 1 from certain skew continuous distributions belongs to class (b) when used as an estimator of the median of the corresponding distribution, and r, the well-known product-moment correlation coefficient, belongs to class (c) when used as an estimator of the correlation coefficient ρ of the underlying (supposedly bivariate normal) distribution in case $\rho > 0$ (see Appendix). The last example is of particular interest as it shows that the contention made by Tschuprow (1939, p. 116) to the effect that the estimator r systematically underrates ρ, is dubious in that this may be taken to mean that r more frequently than not underrates ρ -- which is not true.

3. Bias concepts based on comparing estimators

Looking attentively at definitions 1 and 2 one sees they have one thing in common: all that is involved in using them is the distribution of the estimator itself and the true value of the parameter. This may seem natural enough, yet remember the motivation of the present search for alternative definitions of bias: one may be interested in the question whether the frequency of obtaining too small estimates will be unduly large and want to call an estimator with this property negatively biased.

Now both the terms too small and unduly large imply some type of comparison to be made. When would one call an estimate too small? A reasonable answer to this question seems to be: if the estimate is smaller than a certain value, to be denoted as comparing value. The choice of this comparing value is, generally speaking, arbitrary. Furthermore, when would one call a frequency (of obtaining too small estimates) unduly large? A reasonable answer to this second question seems to be: if it is larger than it would have been if a different (presumably "better") method of estimation would have been used; that is, if it is larger than with a different estimator, to be denoted as comparing estimator. Again the choice of this comparing estimator is to a large extent arbitrary.

It is evident that a definition of bias ensuing from this approach generally involves more than just the true value μ and the distribution of the estimator itself. The next three sections will give three examples of bias concepts based on this "comparing approach".

3.1. Median-bias viewed as a special case of the "comparing approach"

Take for the comparing value the true value μ of the parameter. Take for the comparing estimator any estimator m_c with $\mathrm{Med}(m_c) = \mu$. An estimator m would then be called negatively biased in the sense of section 3 if $P(m \leq \mu) > P(m_c \leq \mu) =$ [illegible]

Hence this definition is identical with the definition of median-bias; cf. the second inequality in (2,2), definition 2. If the distribution of m is symmetric, then of course median-bias and expectation-bias are equivalent and it is here that Snedecor's remark (see section 1 of this paper) seems to fit in the scheme.

The following argument is well in place here. One might want to look at the idea of an unduly large frequency of obtaining too small estimates in another way which might appear not to involve comparing estimators: one might endeavour to define the frequency of obtaining too small estimates (i.e., estimates smaller than the comparing value) as being unduly large if the probability of obtaining estimates smaller than the comparing value would be large as compared with the probability of obtaining estimates larger than the comparing value. One might feel that with μ for a comparing value this approach would lead automatically to the definition

$$P(\underline{m} \leq \mu) > P(\underline{m} \geq \mu) \tag{3.1,1a}$$

(cf. the third inequality in (2,2), definition 2), but it does not. In fact, why not use the definition

$$P(\underline{m} \leq \mu) > k.P(\underline{m} \geq \mu) \tag{3.1,1b}$$

say, with $k \neq 1$? The simple fact that one is apt indeed to prefer the inequality (3.1,1a) to (3.1,1b) shows that in applying this approach one would be using implicitly a symmetrically distributed estimator (or at least an estimator with equal probabilities of values being smaller and being larger than the true value) for a comparing estimator. An analogous remark holds true if any other comparing value (than μ) would be used; always one would have to decide about questions like: how large is k in (3.1,1b) to be chosen? So this approach is essentially identical with the approach based on comparing estimators.

3.2 Distribution-bias

Take for comparing values all conceivable values ξ of the parameter μ (the set of values ξ may be an interval, both finite and infinite).

As for the choice of the comparing estimator $\underline{m}_c$, there are often two or more estimators with different, yet (almost) equally desirable properties. In such cases there may be no point in selecting just one of these estimators for comparison so that one may want to use all of them (in most practical cases the number of estimators competing for the rôle of comparing estimator will not be too large). The concept of bias will then have to be modified into bias with respect to (a particular comparing estimator) $\underline{m}_c$.

To come back at the comparing values ξ, note that if the set I of values ξ coincides with the union of the set of ξ with $0 < P(\underline{m} < \xi) < 1$

and the set of ξ with $0 < P(\underline{m}_c < \xi) < 1$,

then the condition

$$P(\underline{m} \leq \xi) > P(\underline{m}_c \leq \xi) \text{ for } \xi \varepsilon I$$

means exactly that $\underline{m}$ is stochastically smaller than $\underline{m}_c$ (this term was introduced by Mann and Whitney, 1947, p. 50). However, in the definition to follow, I will be allowed not to contain all these ξ-values.

The preceding considerations lead to

Definition 3. Let I be the set of all values which the parameter μ might conceivably have (depending on the specific practical problem). Then the estimator m of μ will be called negatively distribution-biased with respect to a comparing estimator $\underline{m}_c$ if

$$P(\underline{m} \leq \xi) \geq P(\underline{m}_c \leq \xi) \quad \text{for } \xi \varepsilon I, \tag{3.2,1}$$

the equality sign not holding true for at least one $\xi \varepsilon I$.

Note that whereas the definition of positive distribution-bias is self-evident after definition 3, the definition of distribution-unbiasedness with respect to $\underline{m}_c$ presents difficulties. If the $\geq$ sign in (3.2,1) would be simply replaced by an equality sign, then it would be easy to find estimators which are neither distribution-unbiased nor distribution-biased. Alternative definitions of distribution-unbiasedness will easily appear to be unsatisfactory in other respects, and will not be attempted here.

3.3. ξ_c-bias

In certain problems one may be particularly interested in one special comparing value, not the true value of the parameter μ (for example, in response surface theory the value zero plays an important rôle in connection with the estimation of the canonical regression coefficients). Call this special comparing value ξ_c, so that ξ_c is a fixed and known constant as opposed to ξ occurring in section 3.2. Handling the question of comparing estimators in the same way as in section 3.2 one arrives at

Definition 4. The estimator m of μ will be called negatively ξ_c-biased with respect to the estimator m_c if

$$P(m \leq \xi_c) > P(m_c \leq \xi_c) \tag{3.3,1}$$

The definition of ξ_c-unbiasedness presents no difficulties.

Note that estimators which are both expectation-unbiased and median-unbiased may be ξ_c-biased with respect to some estimator m_c. In fact, due to the asymmetry of ξ_c with respect to μ (ξ_c cannot be both larger and smaller than μ), even an estimator m that is distributed symmetrically around μ may be ξ_c-biased with respect to another estimator m_c that is also distributed symmetrically around μ. For example, let m be distributed $N(\mu, 2\sigma^2)$ and m_c be distributed $N(\mu, \sigma^2)$, and take $\xi_c = \mu - 2\sigma$, then m is negatively ξ_c-biased with respect to m_c. Hence, from the standpoint of ξ_c-bias an estimator which is both expectation-unbiased and median-unbiased may be worse than even an estimator which is negatively expectation-biased and negatively median-biased, but which has a smaller variance. Here the concepts of bias and of efficiency merge into each other. Another possible reason (also a mixture of bias and efficiency considerations) for preferring certain biased estimators to certain unbiased estimators was brought forward by Olekiewicz, 1950.

At this point the author remembers with pleasure a discussion with D. Hurst, Department of Experimental Statistics, N.C. State College, Raleigh, who emphasized that there is little practical value in proving that a certain estimator possesses just one out of the whole long list of properties which estimators may be desired to have, and that the main interest ought to be in estimators with several useful properties.

4. A lemma which is useful in proving certain types of bias

An almost trivial lemma which, however, sometimes provides an easy proof of an estimator being (negatively) median-biased, or distribution-biased, or ξ_c-biased is

Lemma 1.

Whether the random variables $\underline{m}$ and $\underline{z}$ are independent or dependent, if

$$P(\underline{z} \geqq \zeta) = 1, \qquad (c)$$

then

$$P(\underline{m} \leqq \xi) \geqq P(\underline{m}_c \leqq \xi + \zeta), \qquad (4,1)$$

where $\underline{m}_c = \underline{m} + \underline{z}$. A necessary and sufficient condition for the equality sign to hold in (4,1) is

$$P[(\underline{m} + \underline{z} > \xi + \zeta) \cap (\underline{m} \leqq \xi)] = 0 \qquad (4,2)$$

Proof

Drawing a picture will bring home the triviality of the proof.

$$P(\underline{m} \leqq \xi) = P\{[(\underline{z} \leqq \xi + \zeta - \underline{m}) \cap (\underline{m} \leqq \xi)] \cup [(\underline{z} > \xi + \zeta - \underline{m}) \cap (\underline{m} \leqq \xi)]\}$$

$$= \text{(because of condition (c) in the lemma)} =$$

$$= P[\underline{m} + \underline{z} \leqq \xi + \zeta] + P[(\underline{m} + \underline{z} > \xi + \zeta) \cap (\underline{m} \leqq \xi)] \geqq$$

$$\geqq P[\underline{m}_c \leqq \xi + \zeta],$$

which completes the proof both of (4,1) and (4,2).

It should be remarked that the equality (4,2) will represent a rather peculiar property of the joint distribution of $\underline{m}$ and $\underline{z}$, yet it is clear that (4,2) cannot be proved or disproved from the conditions of the lemma alone. Three extra conditions each of them sufficient for (4,2) not to hold, are

α) $\underline{m}$ and $\underline{z}$ are independently distributed and $P(\underline{z} = \zeta) < 1$,

β) each (measurable) set in the half-plane $\underline{z} > \zeta$ in $(\underline{m}, \underline{z})$-space has positive probability,

γ) some set $[(\underline{z} > \zeta_0) \cap (\underline{m} > \xi_0)]$ with $\zeta_0 + \xi_0 > \zeta + \xi$ has positive probability (implied by, hence weaker than β).

Remark. Lemma 1 is useful in connection with certain cases of negative bias. The analogue for positive bias is:

Lemma 1'.

Whether the random variables $\underline{m}$ and $\underline{z}$ are independent or dependent, if

$$P(\underline{z} \leq \zeta) = 1, \qquad (c')$$

then

$$P(\underline{m} \geq \xi) \geq P(\underline{m}_c \geq \xi + \zeta) \qquad (4,1')$$

where $\underline{m}_c = \underline{m} + \underline{z}$. A necessary and sufficient condition for the equality sign to hold in (4,1') is

$$P[(\underline{m} + \underline{z} < \xi + \zeta) \cap (\underline{m} \geq \xi)] = 0 \qquad (4,2')$$

When applying Lemma 1 or 1' in order to prove that $\underline{m}$ is a biased estimator of μ in one of the senses discussed in this paper, take $\zeta = 0$ and

for median-bias take $\xi = \mu$ and try to choose the random variable $\underline{z}$ in such a way that $\underline{m}_c = \underline{m} + \underline{z}$ has μ for its median,

for distribution-bias with respect to $(\underline{m} + \underline{z})$ take ξ arbitrarily from the set I of conceivable values of μ and see if $\underline{z}$ can be chosen in such a way that $(\underline{m} + \underline{z})$ is an interesting comparing estimator,

for ξ_c-bias with respect to $(\underline{m} + \underline{z})$ take $\xi = \xi_c$ and see again if $\underline{z}$ can be chosen in such a way that $(\underline{m} + \underline{z})$ is an interesting comparing estimator.

In all three cases $\underline{m}$ and $\underline{z}$ may be mutually independent or dependent.

APPENDIX

At the end of section 2 certain statements were made concerning expectation- and/or median-bias of certain estimators. In this appendix these statements will be proved and illustrated.

A.1. The estimator s^2 of the variance σ^2

Be $x_1, x_2, \ldots, x_n$ an n-fold sample from a normal distribution $N(\mu, \sigma^2)$. It is well known that

$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$$

is an expectation-unbiased estimator of σ^2. It will be shown that $\underline{s}^2$ is, however, negatively median-biased, i.e., that

$$P(\underline{s}^2 \leq \sigma^2) > \frac{1}{2} \quad . \qquad (A.1.1)$$

$$\text{Now } P(\underline{s}^2 \leq \sigma^2) = \frac{1}{\Gamma(\frac{n-1}{2}) \cdot 2^{\frac{n-1}{2}}} \int_0^{n-1} e^{-\frac{y}{2}} y^{\frac{n-1}{2} - 1} dy =$$

$$= \frac{1}{\Gamma(\frac{n-1}{2})} \int_0^{\frac{1}{2}(n-1)} e^{-t} t^{\frac{n-1}{2} - 1} dt =$$

$$= 1 - Q(n-1 \mid n-1) = \gamma(\frac{n-1}{2}, \frac{n-1}{2}) \Big/ \Gamma(\frac{n-1}{2}) = I\left(\sqrt{\frac{n-1}{2}}, \frac{n-1}{2} - 1\right),$$

where the functions Q, γ and I are defined by

$$Q(\chi^2 \mid \nu) = [\Gamma(\frac{\nu}{2})]^{-1} \cdot 2^{-\frac{\nu}{2}} \cdot \int_{\chi^2}^{\infty} e^{-\frac{y}{2}} y^{\frac{\nu}{2} - 1} dy,$$

cf. Pearson and Hartley (1954), p. 122;

$$\gamma(\alpha, x) = \int_0^x e^{-t} t^{\alpha - 1} dt,$$

cf. Higher Transcendental Functions (Vol. 2, 1953), p. 133;

$$I(u,p) = I\left(\frac{m}{\sqrt{p+1}}, p\right) = \gamma(p+1, m)\Big/\Gamma(p+1),$$

cf. K. Pearson (1946) or Jordan (1950), p. 56.

Jordan (1950), p. 57 mentions that "it can be shown by the tables of the function $I(u,p)$ that ... $I(\sqrt{p+1}, p) > \frac{1}{2}$". However, this can be proved analytically. By means of the asymptotic expression for $\gamma(\alpha + 1, \alpha + \sqrt{2\alpha}\, y)$ given by Tricomi (1950) one finds that

$$\gamma(\alpha,\alpha)/\Gamma(\alpha) = \frac{1}{2} + \frac{1}{3\sqrt{2\pi\alpha}} + O(\alpha^{-1}), \qquad (A.1,2)$$

a formula which for real α can easily be proved by Laplace's method. This formula implies that for α large enough $\gamma(\alpha,\alpha)/\Gamma(\alpha)$ exceeds $\frac{1}{2}$ and that

$$\gamma(\alpha,\alpha)/\Gamma(\alpha) \longrightarrow \frac{1}{2} \text{ if } \alpha \longrightarrow \infty$$

Therefore, the inequality (A.1,1) will evidently be proved (even for small values of n) if it can be shown that

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

Now this inequality is equivalent to

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

or

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

or

$$[e^{-t}\, t^{\alpha}]_0^{\alpha} + \int_0^{\alpha} e^{-t}\, t^{\alpha} dt > \int_0^{\alpha+1} e^{-t} t^{\alpha} dt$$

or

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

As the maximum of the integrand $e^{-t}\, t^{\alpha}$ is reached for $t = \alpha$, the last inequality is clearly correct, whereby the proof is complete.

R e m a r k. Line 6 from the bottom of p. 141 of the above-cited Volume 2 of Higher Transcendental Functions contains an error: it is stated that $\Gamma(\alpha, x)/\Gamma(\alpha)$ is a monotonic decreasing function of α for $\alpha > 0$, $x > 0$, whereas in reality $\gamma(\alpha, x)/\Gamma(\alpha)$ is a decreasing function of α. (Note that $\Gamma(\alpha, x) + \gamma(\alpha, x) = \Gamma(\alpha)$; cf. Jordan (1950), p. 57, equation (3)).

The following table is computed from Pearson and Hartley (1954), p. 122 seqq., at some places checked by K. Pearson's table of the incomplete gamma function. It is interesting to note that the asymptotic expression (A.1,2) yields results which are accurate to 3 significant decimal places for $n - 1 \geq 4$.

$n - 1$	$P(s^2 \leq \sigma^2)$
1	0,683
2	0,632
3	0,608
4	0,594
5	0,584
6	0,577
7	0,571
8	0,567
9	0,563
10	0,560
20	0,542
30	0,534
40	0,530
50	0,527
60	0,524
70	0,522
98	0,519

A.2. The sample median as an estimator of the median of the distribution.

Be $F(x)$ a continuous distribution function with $dF(x)/dx$ different from zero in one (finite or infinite) x-interval, and $x_1, x_2, \ldots, x_{2m+1}$ a sample of odd size from the corresponding distribution. Rearrange the values in the sample, and use the notation $x^{(1)} \leq x^{(2)} \leq \ldots \leq x^{(2m+1)}$. As $F(x)$ is assumed to be a continuous distribution function, the occurrence of equality signs has probability zero; $x^{(m+1)}$ is the sample median. It is well known (cf. for instance Wilks, 1948, equation (16) on p. 16) that

$$P(x^{(m+1)} \leq p) = \frac{\Gamma(2m+2)}{\Gamma(m+1)\Gamma(m+1)} \cdot \int_{-\infty}^{p} [F(y)]^m [1-F(y)]^m dF(y). \qquad (A.2,1)$$

Denote by G the inverse (defined for $0 < F < 1$) of the function F so that $G(\frac{1}{2})$ is the median of the distribution considered. Substitute $F(y)$ by F in the integrand of (A.2,1) and use the equality $F\left\{G(\frac{1}{2})\right\} = \frac{1}{2}$, then

$$P\left\{x^{(m+1)} \leq G(\tfrac{1}{2})\right\} = \frac{\Gamma(2m+2)}{\Gamma(m+1)\Gamma(m+1)} \cdot \int_{0}^{1/2} F^m(1-F)^m \, dF. \qquad (A.2,2)$$

Because of $\int_{0}^{1/2} F^m(1-F)^m \, dF = \int_{1/2}^{1} H^m(1-H)^m \, dH$

(as appears from substituting $F = 1-H$), one obtains

$$\int_{0}^{1/2} F^m(1-F)^m \, dF = \frac{1}{2} \int_{0}^{1} F^m(1-F)^m \, dF = \frac{1}{2} B(m+1,m+1) . \qquad (A.2,3)$$

The equations (A.2,2) and (A.2,3) prove that $x^{(m+1)}$ is a median-unbiased estimator of the median G(1/2) whether $F(x)$ represents a skew distribution or a symmetric one.

However, the expectation of $x^{(m+1)}$ equals

$$\frac{\Gamma(2m+2)}{\Gamma(m+1)\Gamma(m+1)} \int_{-\infty}^{+\infty} y \, [F(y)]^m \, [1-F(y)]^m \, dF(y) =$$

$$= \frac{1}{B(m+1,m+1)} \int_{0}^{1} G(F) \, . \, F^m(1-F)^m \, dF =$$

$$= G(\tfrac{1}{2}) + \frac{1}{B(m+1,m+1)} \int_0^1 [G(F) - G(\tfrac{1}{2})] \;.\; F^m(1-F)^m \, dF \qquad (A.2,4)$$

Now $\int_0^1 [G(F) - G(\tfrac{1}{2})] \;.\; F^m(1-F^m \, dF = \int_{-1/2}^{+1/2} [G(\tfrac{1}{2}+h) - G(\tfrac{1}{2})] \;.\; (\tfrac{1}{4}-h^2)^m \, dh =$

$$= \int_0^{1/2} \left\{ [G(\tfrac{1}{2}+h) - G(\tfrac{1}{2})] - [G(\tfrac{1}{2}) - G(\tfrac{1}{2}-h)] \right\} (\tfrac{1}{4}-h^2)^m \, dh. \qquad (A.2,5)$$

Hence the second term in the third member of (A.2,4) equals zero for distributions which are symmetric with respect to the median, since for these distributions $G(\frac{1}{2}+h) - G(\frac{1}{2})$ is an odd function of h. Further, this same second term in (A.2,4) is $\left\{ \begin{matrix} \text{positive} \\ \text{negative} \end{matrix} \right\}$ for those distributions for which

$$G(\tfrac{1}{2}+h) - G(\tfrac{1}{2}) \gtrless G(\tfrac{1}{2}) - G(\tfrac{1}{2}-h) \qquad (0 < h < \tfrac{1}{2}) \qquad (A.2,6)$$

Hence for such skew distributions as satisfy (A.2,6) the statistic $\underline{x}^{(m+1)}$ is an expectation-biased estimator of the median $G(\frac{1}{2})$ (though not only for this type of skew distribution, of course.)

Remark. If the sample size is even, 2m say, the statistic $\frac{1}{2}\left(\underline{x}^{(m)} + \underline{x}^{(m+1)}\right)$ is customarily used as an estimator of the sample median. For certain skew distributions this is a median-biased estimator of the median. If $\phi\left(x^{(1)}, x^{(2)}, \ldots, x^{(2m)}\right)$ is to be a median-unbiased estimator of the median, then the function ϕ will not be independent of the distribution function F. So for even sample sizes there is no exact analogue to the sample median in samples of odd size.

A.3. The estimator r of the correlation coefficient ρ

Be (x_1,y_1), (x_2,y_2) ..., (x_n,y_n) an n-fold sample from a bivariate normal distribution with correlation coefficient ρ. Define the sample correlation coefficient in the usual way by

$$r = \left\{ \sum_{i=1}^{n} (x_i-\bar{x})(y_i-\bar{y}) \right\} \Big/ \left\{ \sum_{i=1}^{n} (x_i-\bar{x})^2 \sum_{j=1}^{n} (y_j-\bar{y})^2 \right\}^{\frac{1}{2}}$$

It is well known (cf. Kendall, 1947, p. 344, eq. (14.55), and Romanowsky, 1925, p. 42, eq. (128)) that

$$E(\underline{r}) = \rho . g(n, \rho^2) = \rho . \frac{\Gamma^2(\frac{n}{2})}{\Gamma(\frac{n-1}{2})\Gamma(\frac{n+1}{2})} . F(\tfrac{1}{2}, \tfrac{1}{2}; \tfrac{n+1}{2}; \rho^2) =$$

$$= \rho . \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n-1}{2})} \int_0^1 t^{\frac{1}{2}-1} (1-t)^{\frac{n}{2}-1} (1-t\rho^2)^{-\frac{1}{2}} dt. \qquad (A.3,1)$$

The last equality follows from Euler's integral representation for the hypergeometric function (used in this context as early as 1925 by Romanowsky, l.c.):

$$F(a,b;c;z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1} (1-t)^{c-b-1} (1-tz)^{-a} dt,$$

see for instance Higher Transcendental Functions (vol. 1, 1953), p. 59, eq. (10), and p. 114, eq. (1).

From

$$(1-t\rho^2)^{-\frac{1}{2}} < (1-t)^{-\frac{1}{2}} \text{ for any } \rho^2 \neq 1, \ t \neq 0$$

it follows that

$$\int_0^1 t^{\frac{1}{2}-1} (1-t)^{\frac{n}{2}-1} (1-t\rho^2)^{-\frac{1}{2}} dt < \int_0^1 t^{\frac{1}{2}-1} (1-t)^{\frac{n-1}{2}-1} dt =$$

$$= B(\tfrac{1}{2}, \tfrac{n-1}{2}) = \frac{\Gamma(\frac{1}{2})\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})} \text{ for any } \rho^2 \neq 1,$$

hence by the fourth and second members of (A.3,1) that

$$g(n, \rho^2) < 1 \text{ for any } \rho^2 \neq 1.$$

Therefore r is a negatively expectation-biased estimator of ρ if ρ is positive, as was stated at the end of section 2.

R e m a r k. The well-known asymptotic series for E(r) may be derived readily from (A.3,1).

In order to investigate whether r is a median-biased estimator of ρ, we use formula (25) of Hotelling (1953)*, p. 200 (note that Hotelling's n denotes the number which is one less than the sample size), which entails

$$P(r \geq \rho) = \frac{n-2}{\sqrt{2\pi}} \frac{\Gamma(n-1)}{\Gamma(n-\frac{1}{2})} (1-\rho^2)^{\frac{n-1}{2}} \int_\rho^1 \frac{(1-r^2)^{\frac{n-4}{2}}}{(1-\rho r)^{n-\frac{3}{2}}} F(\tfrac{1}{2}, \tfrac{1}{2}; n-\tfrac{1}{2}; \tfrac{1+\rho r}{2}) dr \qquad (A.3,2)$$

By means of the substitution

$$\frac{r-\rho}{1-\rho r} = y,$$

suggested by section 5 of Hotelling (1953), the second member of equality (A.3,2) reduces after some patient algebra to

$$\frac{n-2}{\sqrt{2\pi}} \frac{\Gamma(n-1)}{\Gamma(n-\frac{1}{2})} \int_0^1 (1-y^2)^{\frac{n-4}{2}} (1+\rho y)^{\frac{1}{2}} F(\tfrac{1}{2}, \tfrac{1}{2}; n-\tfrac{1}{2}; 1 - \frac{1-\rho^2}{2(1+\rho y)}) dy \qquad (A.3,2a)$$

As ρ increases from 0 to 1, the function $(1+\rho y)^{\frac{1}{2}}$ increases for any $y > 0$,

the function $\frac{1-\rho^2}{1+\rho y}$ decreases for any $y > 0$,

hence $1 - \frac{1-\rho^2}{2(1+\rho y)}$ increases for any $y > 0$,

---

* The author wants to thank Dr. R. J. Hader, Institute of Statistics, Raleigh, for drawing his attention to this paper.

hence $F(\frac{1}{2}, \frac{1}{2}; n-\frac{1}{2}; 1 - \frac{1-\rho^2}{2(1+\rho y)})$ increases for any $y > 0$,

hence the integrand of (A.3,2a) increases with ρ for any $y > 0$.

As $P(\underline{r} \geq \rho) = \frac{1}{2}$ if $\rho = 0$ (the direct verification of this elementary fact from (A.3,2a) is straightforward, though rather intricate), this means that $P(\underline{r} \geq \rho) > \frac{1}{2}$ if $\rho > 0$. Therefore r is a positively median-biased estimator of ρ if ρ is positive, which completes the proof of all statements made at the end of section 2.

LITERATURE CITED

Higher Transcendental Functions, (Bateman Manuscript Project), Vol. 1, 1953, xxvi + 302 pp., Vol. 2, 1953, xvii + 396 pp; McGraw Hill.

Hotelling, H., New light on the correlation coefficient and its transforms; Journal of the Royal Statistical Society, Ser. B (Methodological) 15, 1953, 193-232.

Jordan, Ch. Calculus of finite differences; 2nd ed., New York, Chelsea, 1950, xxi + 652 pp.

Kendall, M. G., The advanced theory of statistics, Vol. 1; London, Griffin, 1947, xii + 457 pp.

Mann, H. B. and Whitney, D. R., On a test of whether one of two random variables is stochastically larger than the other; Ann. Math. Stat. 18, 1947, 50-60.

Olekiewicz, M., On the efficiency of biased estimates; Ann. Univ. Mariae Curie-Skłodowska, Sect. A., 3, 1950, 103-140.

Pearson, E. S. and Hartley, H. O., Biometrika Tables for Statisticians, Vol. 1, Cambr. Univ. Pr., 1954, xiv + 238 pp.

Pearson, K., Tables of the incomplete Γ-functions; Cambr. Univ. Pr. Re-issue 1946, xxxi + 164 pp.

Romanowsky, V., On the moments of standard deviations and of correlation coefficient in samples from normal population: Metron, Vol. 5, N. 4; December 31, 1925; 31-46.

Snedecor, G. W. Statistical methods, applied to experiments in agriculture and biology; 4th ed. (3rd print). Iowa State College Press, 1948, xvi + 485 pp.

Tricomi, F. G. Asymptotische Eigenschaften der unvollständigen Gammafunktion; Mathematische Zeitschrift 53, 1950, 136-148; cf. also Math. Rev. 13, 1952, 553.

Tschuprow, A. A., Principles of the mathematical theory of correlation (translated by M. Kantorowitsch); London, W. Hodge and Co., 1939, x + 194 pp.

Waerden, B. L. van der, Mathematische Statistik (Grundlehren der Mathematischen Wissenschaften Bd. 87) Berlin, Springer, 1957, ix + 360 pp.

Wilks, S. S., Order statistics; Bulletin of the American Mathematical Society 54, 1948, 6-50.