The Annals of Mathematical Statistics
1970, Vol. 41, No. 3, 783–801

DEFICIENCY

BY J. L. HODGES, JR.[1] AND E. L. LEHMANN[2]

University of California, Berkeley

1. Introduction and summary. Consider a statistical procedure (Method A) which is based on n observations, and a less effective procedure (Method B) which requires a larger number k_n of observations to give equally good performance. Comparison of the two methods involves the comparison of k_n with n, and this can be carried out in various ways. Perhaps the most natural quantity to examine is the difference $k_n - n$, the number of additional observations required by the less effective method. Such difference comparisons have been performed from time to time. (See, for example, Fisher (1925), Walsh (1949) and Pearson (1950).) Historically, however, comparisons have been based mainly on the ratio k_n/n. Thus, Fisher (1920), in comparing the mean absolute deviation with the mean squared deviation as estimates of a normal scale, found this ratio to be 1/1.14. Similarly in 1925 he found a large-sample ratio of $2/\pi$ for median compared with mean for estimating normal location, and the same value was found by Cochran (1937) for the sign test relative to the t-test in the normal case.

The reason for using the ratio rather than the difference in these cases is of course that the ratio is stable in large samples, so that a single limit value, say $e = \lim_{n\to\infty} n/k_n$, known as the *asymptotic relative efficiency* or ARE of B with respect to A, conveys a great deal of useful information in a compact form. When $e < 1$, the difference $k_n - n$, is not a useful measure in large samples because it tends to infinity with n.

The situation is however different, and in a sense even the reverse, in the many important statistical problems in which $e = 1$. It is then possible also for the difference to be stable, and the main purpose of this paper is to point out a number of problems in which this is the case. For the additional number $k_n - n$ of observations needed by Method B we suggest the term *deficiency*. If it exists, the limit value $d = \lim_{n\to\infty}(k_n - n)$, will be called the *asymptotic deficiency*. The number d summarizes the comparison much more revealingly in these cases than does the fact that e equals 1. (If it is not known a priori, the latter does not even tell us which of the two procedures is better.) Of course, d is less easy to compute than e since it in effect requires computing an additional term in the asymptotic expansions.

2. General properties of deficiency. We shall now introduce several conceptual and theoretical aspects of deficiency, using as the motivating topic the comparison of two methods of point estimation in terms of expected squared error. Suppose

Received February 17, 1969.

[1] Part of this research was done while this author was a Professor of the Adolph C. and Mary Sprague Miller Institute for Basic Research in Science, University of California, Berkeley.

[2] The research of this author was partially supported by the Office of Naval Research, Contract NONR N00014-67-A-0114-0004.

that A and B are methods of point estimation and let their expected squared errors be denoted by V_n for Method A and by V_n' for Method B, when each estimator is based on n observations. Then for each sample size n for A consider the equivalent sample size $k = k_n$ for B, such that V_k' is equal to V_n or as nearly equal as possible. In the problems with which we shall be concerned, V_n and V_n' will typically be of the form

$$V_n = \frac{c}{n^r} + \frac{a}{n^{r+1}} + o\left(\frac{1}{n^{r+1}}\right) \quad \text{and} \tag{2.1}$$

$$V_n' = \frac{c}{n^r} + \frac{b}{n^{r+1}} + o\left(\frac{1}{n^{r+1}}\right) \tag{2.2}$$

with $r > 0$. Here the positive coefficients of $1/n^r$ are assumed to be the same since otherwise the asymptotic efficiency of one procedure relative to the other would be different from 1. It occasionally happens that the second terms are of order $1/n^{r+s}$ with $s \neq 1$, and we shall see below how to modify the results for this case.

Let us assume, as will typically be the case, that both $\{V_n\}$ and $\{V_n'\}$ are strictly decreasing sequences. Since V_n and V_n' are inherently nonnegative, it then follows that V_n and V_n' are positive for all n.

In spite of the integral nature of sample size, let us for a moment continualize the problem by defining V_k' for nonintegral k in such a way that it is a continuous, strictly-decreasing function of k. (There are clearly many such functions; we shall discuss a particular choice below.) Given such a function, let k_n be the (unique) solution of the equation $V'_{k_n} = V_n$. As $n \to \infty$, it follows from (2.1) and (2.2) that $V_n \to 0$, hence $V'_{k_n} \to 0$, and thus $k_n \to \infty$. Equations (2.1) and (2.2) show further that

$$\frac{1}{n^r}\left[1 + \frac{a + o(1)}{cn}\right] = \frac{1}{k_n^r}\left[1 + \frac{b + o(1)}{ck_n}\right]$$

and hence that

$$k_n/n \to 1. \tag{2.3}$$

Defining d_n by

$$k_n = n + d_n, \tag{2.4}$$

the equation preceding (2.3) may be rewritten as

$$1 + \frac{d_n}{n} = \left[1 + \frac{b + o(1)}{cn}\right]^{1/r}\left[1 + \frac{a + o(1)}{ck_n}\right]^{-1/r} = 1 + \frac{b}{rcn} - \frac{a}{rck_n} + o\left(\frac{1}{n}\right).$$

By (2.3), this shows that

$$d_n \to \frac{b-a}{cr} \quad \text{as} \quad n \to \infty. \tag{2.5}$$

The same argument also shows that when the second term of (2.1) and (2.2) is respectively a/n^{r+s} and b/n^{r+s}, then $d_n \to \infty$ or $d \to 0$ as $s < 1$ or $s > 1$.

Formula (2.5), which is independent of the particular continuous function which is interpolated between V_k for integral k, suggests that the difference between the numbers of observations required by the two procedures tends to a limit as the sample sizes tend to infinity. However, it unjustifiably treats k as a continuous variable and consequently is difficult to interpret. The difficulty can be avoided by the method of stochastic interpolation introduced in [5], according to which sample size k or $k+1$ is chosen with probability $1-\pi$ and π respectively. This yields a continuous expected sample size $k+\pi$. Let us suppose that the performance of the resulting procedure is measured by $V'_{k+\pi} = (1-\pi)V'_k + \pi V'_{k+1}$, as will be the case when V' is a probability, an expectation or a variance. This defines V' as a continuous strictly decreasing function of its subscript as required, and it follows that

$$d_n = k_n + \pi - n \tag{2.6}$$

tends to

$$d = (b-a)/rc, \tag{2.7}$$

which we shall call the *asymptotic expected deficiency* (AED) of procedure B relative to A, or simply the deficiency of B relative to A when no confusion is likely.

Formula (2.7) shows that if V_n and V_n' are given by (2.1) and (2.2), then deficiency has the following two properties, which are analogous to the corresponding properties for efficiency.

(i) *Reflexivity.* If d is the AED of B with respect to A, then the AED of A with respect to B is $-d$.

(ii) *Transitivity.* If d_1 is the AED of B with respect to A, and d_2 the AED of C with respect to B (where the accuracy of C is of the same form as those of A and B), then the AED of C with respect to A is d_1+d_2.

If interpolation is not acceptable, and only integral sample sizes are permitted, the equivalent sample size k_n may be defined as the integer(s) k for which V_k' is closest to V_n. For fixed n and $d = k-n$ we find

$$V'_{n+d} - V_n = c\left[\frac{1}{(n+d)^r} - \frac{1}{n^r}\right] - \frac{a}{(n+d)^{r+1}} - \frac{b}{n^{r+1}} + o\left(\frac{1}{n^{r+1}}\right)$$

$$= [crd - ab + o(1)]\frac{1}{n^r} = \left[d - \frac{ab}{cr} + o(1)\right]\frac{cr}{n^r}.$$

For all sufficiently large n, V'_{n+d} and V_n are therefore made most nearly equal by giving d the integral value closest to ab/cr (or one of the two nearest integers if $ab/cr = 1/2, 3/2, 5/2, \cdots$). We shall call this value of d the *asymptotic integral deficiency* of B with respect to A. While for simplicity we have presented the discussion of deficiency in the context of comparing the squared errors of point estimates, much of it applies to other measures of performance. Examples will be given in Sections 4 and 5.

3. Further examples of point estimation. In the present section, we shall illustrate the use of deficiency on some additional problems of point estimation. As measure of accuracy of an estimator, we shall continue to take the expected squared error. Of course in the case of unbiased estimators, this is the variance.

EXAMPLE 3.1. *Estimation of variance; price of not knowing the mean.* Observations $X_1, X_2, \cdots$ are drawn from a distribution F with expectation ξ and variance σ^2. If ξ is given, the customary estimator for σ^2 based on n observations is

$$M_n = \frac{1}{n}\sum_{i=1}^{n}(X_i-\xi)^2. \tag{3.1}$$

This estimator is unbiased and consistent, and its variance is

$$V_n = \gamma\sigma^4/n \tag{3.2}$$

where $\gamma+1 = \mu_4/\sigma^4$ is the standardized fourth central moment of F.

The estimator M_n is of course valid only if the value given for ξ happens to be correct. For this reason, even when a value for ξ is available from theory or past experience, one may often prefer an estimator which is robust against errors in ξ. The conventional robust estimator is

$$M_n' = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 \quad \text{where} \quad \bar{X} = \frac{1}{n}\sum X_i. \tag{3.3}$$

This estimator is also unbiased and its variance (see Cramér (1946) formula (27.4.2)) may be written as

$$V_n' \doteq \sigma^4\frac{\gamma(n-1)+2}{n(n-1)}. \tag{3.4}$$

In making a choice between these estimators a highly relevant factor is the price one must pay for the robustness. The moments of any distribution F satisfy $\mu_4 \geqq \sigma^4$ and hence $\gamma \geqq 0$. In fact, $\gamma > 0$ except for one special case which we shall discuss separately. Assuming γ to be positive, the variances V_n and V_n' satisfy (2.1) and (2.2) with $r = 1$, $c = \gamma$, $a = 0$ and $b = 2$. The asymptotic relative efficiency of M' relative to M is therefore 1, and the asymptotic expected deficiency is $d = 2/\gamma$.

When F is normal, $\gamma = 2$ and hence $d = 1$. In this case the deficiency of M' with respect to M is equal to 1 in a stronger sense than that we have been discussing, since for normal samples M_n and M'_{n+1} have exactly the same distribution for all sample sizes rather than merely matching the variances in large samples.

While in the normal case it costs only one observation to protect against an erroneous value given for ξ, the deficiency of M' relative to M can be arbitrarily large for certain nonnormal populations since γ can be arbitrarily near zero. It is easy to see that the limiting value $\gamma = 0$ arises only when F puts probability 1/2 at each of the points $\xi-\sigma$ and $\xi+\sigma$. In this case $\text{Var}(M_n) = 0$ for each value of n, corresponding to the fact that $M_n = \sigma^2$ with probability 1. On the other hand, $\text{Var}(M_k') = 2\sigma^2/k(k-1)$ is always positive, approaching 0 as $k \to \infty$. One might thus say that $k_n = \infty$ for every n, so that the ARE is 0 while the AED is ∞. This agrees in a limiting sense with the conclusion that the deficiency is $2/\gamma$.

In statistical practice, special interest attaches to populations that are generally similar to the normal but depart from it by having somewhat heavier tails—corresponding to the presence of an occasional gross error. In such populations one would expect to find γ greater than its value 2 in the normal case, and hence the integral deficiency would be 0 or 1. It is reassuring that the large-sample cost of the protection afforded by use of the robust estimator M' will not exceed one observation in these frequently-encountered heavy-tailed populations.

EXAMPLE 3.2. *Estimation of variance; price of unbiasedness.* In the preceding example, suppose that ξ is unknown but that instead of (3.3) we are willing to consider any estimator of the form

$$M_c = \sum(X_i - \bar{X})^2/(n+c). \tag{3.5}$$

If $c \neq -1$, this will not be unbiased but may have a smaller expected squared error than the unbiased estimator. The deficiency of M_{-1} relative to M_c will then indicate how many observations one loses by insisting on unbiasedness, and thereby provides a basis for deciding whether or not the price is too high.

From (3.4), one easily finds

$$E(M_c - \sigma^2)^2 = \frac{\sigma^4}{n(n+c)^2}\{(n-1)[\gamma(n-1)+2]+n(c+1)^2\} \tag{3.6}$$

and hence

$$E(M_c - \sigma^2)^2 = \sigma^4\left[\frac{\gamma}{n} + \frac{(c+1)^2 - 2\gamma + 2 - 2c\gamma}{n^2} + o\left(\frac{1}{n^2}\right)\right]. \tag{3.7}$$

As an example, consider the classical estimator M_0. The deficiency of M_{-1} relative to M_0 is given by (2.7) with $r = 1$, $c = \gamma$, $a = 3-2\gamma$ and $b = 2$, and hence is equal to

$$d = (2\gamma - 1)/\gamma. \tag{3.8}$$

The classical M_0 is thus better than M_{-1} when $\gamma > \frac{1}{2}$, with the situation reversed when $\gamma < \frac{1}{2}$. When F is normal, in particular, $\gamma = 2$ and $d = -3/2$. One can therefore save an expected 1.5 observations by using the biased estimator M_0. The best value of c in the normal case is $c = 1$ for which $d = 2$ and which therefore provides an additional saving of .5 observations.

EXAMPLE 3.3. *Median versus quasi-median.* Consider the problem of estimating the center θ of a symmetric distribution. Let $X_1, X_2, \cdots$ be a sample from the distribution, and denote the ordered sample by $Y_1 \leqq Y_2 \leqq \cdots$. Suppose that the sample size is odd, and consider as two unbiased estimators of θ: the median $M' = Y_{h+1}$ based on $k = 2h+1$ observations, and the quasi-median $M = (Y_m + Y_{m+2})/2$ based on $n = 2m+1$ observations. It was pointed out in [7] on the basis of a heuristic argument that up to terms of order n^{-2}, $\operatorname{Var} M = \operatorname{Var} M'$ provided $k = n+2$, for all F satisfying suitable regularity conditions. For these F, the AED of M' with respect to M is therefore equal to two.

EXAMPLE 3.4. *Comparison of the minimax and unbiased estimator of binomial p.* Let X denote the number of successes in n binomial trials with success probability p. Then the unbiased estimator of p with uniformly minimum variance is $M_n = X/n$ while the minimax estimator is

$$M_n' = \frac{\sqrt{n}}{1+\sqrt{n}}\frac{X}{n}+\tfrac{1}{2}\cdot\frac{1}{1+\sqrt{n}},$$

and these estimators have expected squared errors

$$V_n = \frac{pq}{n} \quad \text{and} \quad V_n' = \frac{1}{4(1+\sqrt{n})^2}$$

respectively. If $p \neq \frac{1}{2}$, the ARE of V_n' to V_n is $pq/4 < 1$ so that for large n, M_n is superior to M_n'. However, for $p = \frac{1}{2}$, the situation is reversed. The ARE is 1, and $V_n' < V_n$. To determine the deficiency in this case, we must compare $V_n = 1/4n$ with

$$V_n' = \frac{1}{4n}\left[1-\frac{2}{\sqrt{n}}+o\left(\frac{1}{\sqrt{n}}\right)\right].$$

Here the remark applies which was made following (2.5). We have $r = 1$ and $s = \frac{1}{2}$ and the asymptotic deficiency of V_n relative to V_n' is infinite. In fact, if we equate V_{kn} with V_n', we find $k_n = n+2n^{\frac{1}{2}}+1$ so that $k_n-n = 2n^{\frac{1}{2}}+1$. The deficiency computation thus shows that the classical estimator requires a much larger sample size than the minimax estimates if they are to have the same expected squared error at $p = \frac{1}{2}$, in spite of the fact that the corresponding asymptotic efficiency is 1.

4. Confidence sets for normal means. Let $X_{ij}, j = 1, \cdots, n; i = 1, \cdots, p$, be independent observations from p normal distributions with expectations $\xi_1, \cdots, \xi_p$ and common variance σ^2, and suppose we wish to obtain a confidence set for the vector $(\xi_1, \cdots, \xi_p)$. It is sometimes possible to treat σ as known, the value being taken from past experience or theoretical considerations. The best confidence sets then are spheres centered at $(X_1., \cdots, X_p.)$ where $X_i. = \sum X_{ij}/n$ and with a fixed radius computed from the χ^2-distribution. The confidence coefficient of this procedure could of course be seriously invalid if the given value of σ should happen to be incorrect. For this reason, it is often preferred to ignore the given information and instead estimate σ from the data. The resulting confidence sets, which are based on the F-distribution, will again be spheres centered on $(X_1., \cdots, X_p.)$ but the radius will now be a random variable. Estimation of σ involves some loss of effectiveness; the resulting confidence sets will tend to be larger. The choice between the two procedures will depend mainly on two considerations: the degree of reliance placed on the given σ, and the cost of ignoring it. As a contribution to the second aspect we shall in the present section determine the deficiency of the F- with respect to the χ^2-confidence sets, where we shall take as our measure of performance the volume or expected volume of the confidence spheres.

For known σ, the confidence sets at confidence level $1-\alpha$ are given by

$$n\sum(\xi_i - X_{i\cdot})^2/\sigma^2 \leqq v_p. \tag{4.1}$$

Here v_p is determined by the equation

$$\Psi(v_p) = 1-\alpha \tag{4.2}$$

where Ψ denotes the cumulative distribution function of the χ^2-distribution with p degrees of freedom. For simplicity of notation, let us choose the scale so that $\sigma = 1$.

If σ is unknown, the confidence sets are

$$n\sum(\xi_i - X_{i\cdot})^2 \leqq CT \tag{4.3}$$

where $T = \sum\sum(X_{ij}-X_{i\cdot})^2/p(n-1)$. If we put $t = p(n-1)$, then tT has the χ^2-distribution with t degrees of freedom. The constant C (which of course depends on t) is determined by the equation

$$E\Psi(CT) = 1-\alpha. \tag{4.4}$$

As $t\to\infty$, $T\to\sigma^2 = 1$ in probability and hence $C \to v_p$. For a comparison of the two families of confidence sets, it is necessary to find the difference between C and v_p to terms of order $1/t$. Since T is close to 1, this can be accomplished by expanding Ψ about C, and for this purpose we note that the density ψ of Ψ is of the form

$$\psi(y) = Ky^{\frac{1}{2}p-1}e^{-\frac{1}{2}y} \tag{4.5}$$

and hence that its derivative is

$$\psi'(y) = [(\tfrac{1}{2}p-1)y^{-1}-\tfrac{1}{2}]\psi(y). \tag{4.6}$$

Thus

$$\begin{aligned}\Psi(CT) &= \Psi[C+C(T-1)]\\ &= \Psi(C)+C(T-1)\psi(C)+\tfrac{1}{2}C^2(T-1)^2\psi'(C)+\tfrac{1}{6}\psi''(U)C^3(T-1)^3\end{aligned}$$

where U depends on T.

Since $E(T-1) = 0$, $E(T-1)^2 = 2/t$ and $E|T-1|^3 = O(t^{-2})$, it can be shown that

$$\begin{aligned}1-\alpha &= \Psi(C)+\tfrac{1}{2}C^2[(\tfrac{1}{2}p-1)C^{-1}-\tfrac{1}{2}]\psi(C)+o(t^{-1})\\ &= \Psi[C+Ct^{-1}\{(\tfrac{1}{2}p-1)-\tfrac{1}{2}C\}+o(t^{-1})].\end{aligned}$$

This result is obvious for values of p for which $\psi''(y)$ is bounded i.e. for $p \geqq 6$, and easily follows by truncation for the remaining values of p. Comparing this equation with (4.2), we find $v_p = C+Ct^{-1}[(\frac{1}{2}p-1)-\frac{1}{2}C]+o(t^{-1})$ and hence

$$C = v_p - v_p t^{-1}[(\tfrac{1}{2}p-1)-\tfrac{1}{2}v]+o(t^{-1}). \tag{4.7}$$

Since the radius of the confidence spheres (4.1) is $r = (v_p/n)^{\frac{1}{2}}$ when $\sigma = 1$, their volume is given by

$$V_n = K_p r^p = K_p(pv_p/N)^{\frac{1}{2}p} \tag{4.8}$$

where $N = pn$ is the total sample size and K_p is a constant. Similarly, the radius of the confidence spheres (4.3) is $r' = (CT/n)^{\frac{1}{2}}$, so that their expected volume is

$$E(V_n') = K_p(C/n)^{\frac{1}{2}p}E(T^{\frac{1}{2}p}). \tag{4.9}$$

Now

$$E(T^{\frac{1}{2}p}) = 1 + A_p t^{-1} + o(t^{-1}) \tag{4.10}$$

where we shall determine A_p below. Hence, using the fact that $t = p(n-1)$, we have

$$E(V_n') = K_p\left(\frac{pv_p}{N}\right)^{p/2}\left\{1 - \frac{p}{2N}\left[\left(\frac{p}{2}-1\right)-\frac{v_p}{2}\right]+\frac{A_p}{N}+o\left(\frac{1}{N}\right)\right\}. \tag{4.11}$$

It follows that V_n and $E(V_n')$ satisfy (2.1) and (2.2) with $r = \frac{1}{2}p$, $c = (pv_p)^{\frac{1}{2}p}K_p$, $a = 0$ and $b = (pv_p)^{\frac{1}{2}p}K_p\{A_p - \frac{1}{2}p[(\frac{1}{2}p-1)-\frac{1}{2}v_p]\}$. By (2.7), the AED is therefore

$$d_p = 2p^{-1}A_p - (\tfrac{1}{2}p-1) + \tfrac{1}{2}v_p\,. \tag{4.12}$$

To complete the evaluation of d, it remains to find A_p. Now the moments of T are given by (see for example [6])

$$E(T^{\frac{1}{2}p}) = \frac{(t+2a-2)(t+2a-4)\cdots t}{t^a} = 1 + \frac{a(a-1)}{t} + o\left(\frac{1}{t}\right) \quad \text{if} \quad p = 2a$$

and

$$E(T^{\frac{1}{2}p}) = \frac{(t+2a-1)(t+2a-3)\cdots(t+1)}{t^a}\left[1 - \frac{1}{4t}, + o\left(\frac{1}{t}\right)\right]$$

$$= 1 + (a^2 - \tfrac{1}{4})\frac{1}{t} + o\left(\frac{1}{t}\right) \quad \text{if} \quad p = 2a+1.$$

Substituting the values $A_{2a} = a(a-1)$ and $A_{2a+1} = a^2 - \frac{1}{4}$ into (4.12) shows that for all p the deficiency of the F- to the χ^2-confidence sets is

$$d_p = v_p/2. \tag{4.13}$$

The deficiency thus increases with $1-\alpha$ and with p. The values of d corresponding to a number of values of α and p are shown in the table below. It is interesting to note how large some of these values are.

p	α				
	.1	.05	.025	.01	.005
1	1.353	1.921	2.512	3.317	3.940
2	2.303	2.996	3.689	4.605	5.298
5	4.618	5.535	6.416	7.543	8.375
10	7.994	9.154	10.242	11.605	12.594

The result extends without any difficulty to the case of unequal sample sizes. If the number of observations in the ith sample is n_i, the total number of observations $N = \sum n_i$ and $\rho_i = n_i/N$, the confidence sets for known σ are

$$N \sum \rho_i(\xi_i - X_{i\cdot})^2/\sigma^2 \leqq v_p. \tag{4.14}$$

These are ellipsoids whose volume is $K_p(v_p/N)^p/\rho_1 \cdots \rho_p$ instead of (4.8). Similarly, the expected volume of the ellipsoids for unknown σ is $K_p(C/N)^p E(T^p)/\rho_1 \cdots \rho_p$ and (4.13) now follows exactly as before.

5. Student's t-test versus the $\bar{X}$-test. In the preceding sections we were concerned with the comparison of two procedures the performance of which was measured by a single real number: A variance, or volume of a confidence set, which happened to be independent of the unknown parameters, or which we considered only at specified values of the parameters. This is however atypical; more usually, performance is measured by a function of the parameters and an equivalent sample size will have to be defined in terms of an overall matching of two such functions.

In the present section we shall consider a testing problem, where the performance is described by the power functions of the tests in question but where through special circumstances the difficulty alluded to need be confronted only partially. The problem is the normal one-sample location problem. The variables $X_1, X_2, \cdots$ are assumed to be independently normally distributed with expectation ξ and variance σ^2, and the hypothesis $H: \xi = 0$ is to be tested against the alternatives $\xi > 0$. As in the preceding section, we shall compare the test that is appropriate when a given value of σ can be relied upon, with the t-test, which places no reliance on such an assumed value of σ, in terms of deficiency. The determination of this deficiency enables us to throw some fresh light on the interesting solution to the cost problem given by John Walsh (1949). (We are indebted to Prof. J. Hemelrijk for pointing out to us the connection of our results with the work of Walsh.)

When σ is known, the hypothesis H is accepted at level α on the basis of n observations $X_1, \cdots, X_n$ when

$$n^{\frac{1}{2}}\bar{X}_n/\sigma \leqq u_\alpha \tag{5.1}$$

where

$$u_\alpha = \Phi^{-1}(1-\alpha). \tag{5.2}$$

For simplicity of notation, we shall suppress the subscript α of u in the following discussion and suppose furthermore that $\sigma = 1$. The acceptance probability of the test (5.1) is then

$$\beta(\xi, n) = \Phi(u - n^{\frac{1}{2}}\xi) = \Phi(u - \theta) \tag{5.3}$$

where $\theta = n^{\frac{1}{2}}\xi$.

Consider now the t-test based on k observations, with acceptance region

$$k^{\frac{1}{2}}\bar{X}_k/S_k \leqq c_k, \tag{5.4}$$

where $S_k^2 = \sum(X_i - \bar{X}_k)^2/(k-1)$. The acceptance probability of this second test can be written as

$$\gamma(\xi, k) = E\Phi(c_k S_k - k^{\frac{1}{2}}\xi). \tag{5.5}$$

Let us put $\eta = k^{\frac{1}{2}}\xi$, add and subtract c_k within the parenthesis, and expand the right-hand side about $c_k - \theta$. Suppressing the subscript k in c and S, formula (5.5) then becomes

$$\gamma(\xi, k) = E\{\Phi(c-\eta) + \varphi(c-\eta)c(S-1) + \tfrac{1}{2}\varphi'(c-\eta)c^2(S-1)^2 + \tfrac{1}{6}\varphi''(T)c^3(S-1)^3\},$$

where T depends on S and η. From the facts that

$$E(S_k) = 1 + \frac{a}{k} + O\left(\frac{1}{k^2}\right), \qquad \operatorname{Var}(S_k) = \frac{1}{2k} + O\left(\frac{1}{k^2}\right)$$

and $E|S_k - E(S_k)|^3 \leqq M/k^2$, it then follows that

$$\gamma(\xi, k) = \Phi(c-\eta) + \frac{ac}{k}\varphi(c-\eta) + \frac{c^2}{4k}\varphi'(c-\eta) + O\left(\frac{1}{k^2}\right).$$

Here the boundedness of φ'' insures that the error term is of order $1/k^2$ uniformly in η. Using the fact that $\varphi'(x) = -x\varphi(x)$ and once more the expansion of Φ, γ can be now written as

$$\gamma(\xi, k) = \Phi\left[\left(c + \frac{ac}{k} - \frac{c^3}{4k}\right) - \eta\left(1 - \frac{c^2}{4k}\right)\right] + O\left(\frac{1}{k^2}\right).$$

Putting $\eta = 0$ in this equation yields

$$1 - \alpha = \Phi(u) = \Phi\left(c + \frac{ac}{k} - \frac{c^3}{4k}\right) + O\left(\frac{1}{k^2}\right)$$

and hence

$$u = c + \frac{ac}{k} - \frac{c^3}{4k} + O\left(\frac{1}{k^2}\right),$$

and thus in particular also $u = c + O(1/k)$. Substitution in the last expression for γ and the equation $\eta = (k/n)^{\frac{1}{2}}\theta$ finally gives the approximation

$$\gamma(\xi, k) = \Phi\left[u - \left(\frac{k}{n}\right)^{\frac{1}{2}}\theta\left(1 - \frac{u^2}{4k}\right)\right] + O\left(\frac{1}{k^2}\right), \tag{5.6}$$

where the error term is of order $1/k^2$ uniformly in θ.

If we now equate expressions (5.3) for $\beta(\xi, n)$ and (5.6) for $\gamma(\xi, k)$ to find the value $k = k_n$ for which the two tests give the same power (up to terms of order $1/n$) for a fixed value of θ, we find in analogy to (2.5) that

$$d_n = k_n - n \to u^2/2. \tag{5.7}$$

This result is independent of θ and hence of ξ, and we see in fact that the maximum difference $\sup|\beta(\xi, n) - \gamma(\xi, k_n)|$ is of order $1/n^2$ if and only if k_n satisfies (5.7).

As did (2.5), relation (5.7) suffers from the difficulty that k is not a continuous variable. This can be avoided as before by stochastic interpolation. Instead of a fixed sample size k, we take either k or $k+1$ observations with probability $1-\pi$ and π respectively, the acceptance probability of the resulting test being

$$(1-\pi)\gamma(\xi, k)+\pi\gamma(\xi, k+1).$$

If this is equated with $\beta(\xi, n)$ we find as in Section 2 that the difference between the expected sample size $k_n+\pi$ for the t-test and the sample size n for the normal test tends to

$$d = \tfrac{1}{2}u_\alpha^{\ 2}, \tag{5.8}$$

which is therefore the AED in the present case. The following are the values of d corresponding to a number of values of α.

α	.10	.05	.025	.01	.005
d	.821	1.353	1.921	2.706	3.3175

In large samples, those values may be interpreted as the price, measured by the number of additional observations, required to protect oneself against possible error in the given value of σ. In our judgment, the values of d are small enough so that it would rarely be reasonable to rely on the given σ.

Since the value of d is asymptotic, the question arises how good the approximation is for a given finite value of n or k. Some indications regarding the accuracy of the approximation for small samples will be given at the end of the section.

In the foregoing, we were concerned with testing the hypothesis $\xi = 0$ against the alternatives $\xi > 0$. By symmetry, (5.8) is clearly also the deficiency when the problem is that of testing $\xi = 0$ against the alternatives $\xi < 0$. Consider now the problem of testing $\xi = 0$ against the two-sided alternatives $\xi \neq 0$ at significance level α. The power function of the symmetric two-sided normal or t-test is just the sum of the power functions of the corresponding two one-sided tests at level $\alpha/2$. Since an expected addition to n of $d = \frac{1}{2}u_{\alpha/2}^2$ observations will bring both of the one-sided t-power functions into coincidence with the corresponding normal power functions, it is seen that the deficiency in the two-sided case is just

$$d = \tfrac{1}{2}u_{\alpha/2}^2. \tag{5.9}$$

This two-sided testing problem is closely related to the problem of obtaining confidence intervals for ξ, which in turn is the special case $p = 1$ of the problem treated in Section 4. The deficiency found there was $\frac{1}{2}v$, with v defined by $\Psi(v) = 1-\alpha$ where Ψ is the cumulative distribution function of a χ^2 variable with 1 degree of freedom. Clearly $v^{\frac{1}{2}} = u_{\alpha/2}$, and hence the deficiency (5.9) agrees with the deficiency (4.13) for $p = 1$.

Procedures intermediate between the one-sided and symmetric two-sided tests are the asymmetric two-sided tests. The corresponding comparisons in these cases are quite different from those treated above, because it is not possible to match the power curves to the same order as in the one-sided case.

Let us return now to the one-sided case and, as in Section 2, consider the alternative approach of integral deficiency, in which the equivalent sample size is defined as the integer k_n for which $\gamma(\xi, k)$ is closest to $\beta(\xi, n)$. To make this precise, it is necessary to specify what is meant by the functions $\gamma(\xi, k)$ and $\beta(\xi, n)$ being as close as possible. Since (5.3) and (5.6) express β and γ conveniently in terms of θ, we could specify a value of θ and ask that the acceptance probabilities be as close as possible for this particular value of θ. However, such a choice would typically have to be rather arbitrary (although in the present case the large-sample result would be independent of θ) and we shall instead define the equivalent sample size to be the integer k_n which minimizes

$$\sup_\xi |\gamma(\xi, k)-\beta(\xi, n)|. \tag{5.10}$$

Here the supremum is taken over $\xi > 0$ since the tests are one-sided.

To examine the behavior of $\gamma(\xi, k)$ for k near n, let us write $k = n+d$ where d is now an integer. Then

$$\left(\frac{k}{n}\right)^{\frac{1}{2}} = \left(\frac{n+d}{n}\right)^{\frac{1}{2}} = 1+\frac{d}{2n}+O\left(\frac{1}{n^2}\right)$$

so that by (5.6)

$$\begin{aligned}\gamma(\xi, n+d) &= \Phi\left[u-\theta\left(1+\frac{d}{2n}-\frac{u^2}{4n}\right)\right]+O\left(\frac{1}{n^2}\right) \\ &= \Phi(u-\theta)+\frac{1}{n}(\tfrac{1}{2}u^2-d)\cdot\tfrac{1}{2}\theta\varphi(u-\theta)+O\left(\frac{1}{n^2}\right),\end{aligned} \tag{5.11}$$

and hence

$$\gamma(\xi, n+d)-\beta(\xi, n) = (\tfrac{1}{2}u^2-d)\tfrac{1}{2}\theta\varphi(u-\theta)\cdot\frac{1}{n}+O\left(\frac{1}{n^2}\right).$$

Let us now find the value of d for which $\gamma(\xi, n+d)$ fits most closely to $\beta(\xi, n)$ in the sense of minimizing (5.10).

It is necessary to distinguish three cases, and we first suppose that for some integer d_0,

$$d_0 < \tfrac{1}{2}u^2 < d_0+\tfrac{1}{2}. \tag{5.12}$$

The other two cases: $d_0-\frac{1}{2} < \frac{1}{2}u^2 < d_0$ and $d_0+\frac{1}{2} = \frac{1}{2}u^2$ can be handled similarly. Furthermore, we shall assume that $\alpha < \frac{1}{2}$ so that $u > 0$; again, the other case is completely analogous. The function $\frac{1}{2}\theta\varphi(u-\theta)$ then achieves a unique maximum, say h, at a point $\theta = \theta_0 > 0$. The value of h is also the maximum of $|\frac{1}{2}\theta\varphi(u-\theta)|$. If for the moment we ignore the uniform error terms of order $1/n^2$, we find that the acceptance curves of the three t-tests $\gamma(\xi, n+d_0-1)$, $\gamma(\xi, n+d_0)$ and $\gamma(\xi_t\, n+d_0+1)$ all attain their maximum distance from $\beta(\xi, n)$ at θ_0, where they are respectively: low by $(d_0+1-\frac{1}{2}u^2)h/n$; high by $(\frac{1}{2}u^2-d_0)h/n$; and high by $(\frac{1}{2}u^2-d_0+1)h/n$. Of these three curves, $\gamma(\xi, n+d_0)$ is thus closest to $\beta(\xi, n)$ and it is closer than its nearer competitor by $[2(d_0-\frac{1}{2}u^2)+1]h/n = A$. For values of k other than these three, the superiority of fit of $\gamma(\xi, n+d_0)$ is still greater: from (5.11), it is seen that for $\xi > 0$, $\gamma(\xi, n+d)$ falls monotonely as d increases, so that $\gamma(\xi, n+d_0+2)$, $\gamma(\xi, n+d_0+3), \cdots$ are still further below $\beta(\xi, n)$ than is $\gamma(\xi, n+d_0)$, while $\gamma(\xi, n+d_0-2)$, $\gamma(\xi, n+d_0-3), \cdots$ are still further above $\beta(\xi, n)$ than is $\gamma(\xi, n+d_0-1)$.

To summarize: if we ignore the error terms, the t-acceptance-curve which best fits $\beta(\xi, n)$ is $\gamma(\xi, n+d_0)$, and its superiority of fit over its nearest rival is proportional to $1/n$. However, we know that the error terms tend to zero uniformly at rate $1/n^2$. For n sufficiently large and with k_n/n near 1, the true acceptance curves will all differ from their approximations by less than $\frac{1}{2}A$. It follows that for sufficiently large n, the equivalent sample size is $k_n = n+d_0$. Dropping assumption (5.12), the equivalent sample size is then finally $k_n = n+d$ where d is the integer(s) closest to the AED namely $\frac{1}{2}u_\alpha{}^2$.

The problem of determining the cost (in terms of the number of lost observations) was treated as early as 1949 by John E. Walsh [10]. His solution was based on a formula of Johnson and Welch (1939). By a heuristic argument, in which a chi-variable is treated as if it were normal, these authors were led to

$$\Phi\left\{k^{\frac{1}{2}}\xi\left[1-\frac{u^2}{2(k-1)}\right]^{\frac{1}{2}}-u\right\} \tag{5.13}$$

as an approximation for the acceptance probability of the t-test. Walsh identified this expression with the acceptance probability (5.3) of the $\bar{X}$-test, treating n as if it were a continuous variable. This leads at once to the expression

$$k-n = \tfrac{1}{2}u^2k/(k-1) \tag{5.14}$$

which Walsh refers to as approximately the number of "sample values 'wasted' by using a t-test." The quantity $k-n$ is an approximation to what we have called deficiency.

In order to compare (5.13) with our analysis, note that it can be written as

$$\tfrac{1}{2}u^2\frac{k}{k-1} = \tfrac{1}{2}u^2+\frac{1}{2k}u^2+O\left(\frac{1}{k^2}\right). \tag{5.15}$$

Walsh notes that the first term on the right-hand side can be used to approximate the number of "wasted" observations if k is not too small. It is of course just our AED (5.8). From this point of view we have provided a more rigorous justification of Walsh's formula, which is perhaps also more easily interpreted: With stochastic interpolation (or some other equivalent interpolation method) the dominant term of Walsh's formula for the number of wasted observations is the limit value of the difference of equivalent sample sizes.

We are however not able to give a similar justification to the second term of (5.15). In fact, the correct term of order $1/k$ for the deficiency depends on the $1/k^2$-term in the power of t. Unfortunately, the normal approximation (5.13) has an error of this order. From the results of Section 6 of [6] it is easy to see that, for fixed α and fixed noncentrality parameter δ, the acceptance probability of the t-test with f degrees of freedom is given by

$$\Phi\left[u-\delta+\frac{\delta u^2}{4f}+\frac{6u^2\delta-5u^4\delta+4u^3\delta^2}{96f^2}\right]+O\left(\frac{1}{f^3}\right). \tag{5.16}$$

On the other hand, the approximation (5.13), on expansion, can be written as

$$\Phi\left[u-\delta+\frac{\delta u^2}{4f}+\frac{3u^4\delta}{96f^2}\right]+O\left(\frac{1}{f^3}\right).$$

This differs from the correct value by a term of order $1/f^2$.

One could of course use (5.16) to derive an expression for the deficiency accurate to terms of order $1/n$. The deficiency to this order depends on the precise manner in which the acceptance curves are matched. In particular, if they are matched at a fixed power π (as well as at the null hypothesis) the $1/n$ term of the expansion for k_n-n would depend on the value of π.

We conclude with a numerical illustration of the comparison of t with $\bar{X}$ in the one-sided case. The table below shows power values of the $\bar{X}$-test at level $\alpha = .05$, for $n = 4$ and $n = 8$ observations. Also shown in each case are the corresponding power values of the stochastically-interpolated t-test, as well as the power for the bracketing values, the best fitting (with minimax matching) interpolated sample sizes being $k_4 = 5.6377$ and $k_8 = 9.5149$ respectively. The small-sample expected deficiencies (1.6377 at $n = 4$ and 1.5149 at $n = 8$) are close enough to the asymptotic value (1.3528), so that the asymptotic theory provides reasonable guidance even at these very small sample sizes. The approach to the limit is also consistent with the fact that the difference between the small-sample value and the asymptotic value is of order $1/n$. Note that the matching is satisfactorily close (the maximum difference is .00670 at $n = 4$ and .00176 at $n = 8$), and is decreasing consistently with the theoretical rate $1/n^2$.

6. Bayes versus unbiased estimation of a normal mean. In a deficiency investigation in which the performance of the procedures in question depends on a parameter, this will typically also be true of the deficiency. For an overall comparison of the procedures it is then necessary to adjust the sample sizes so as to provide an optimum fit for the two performance curves, for example, by minimizing their maximum difference, This approach which was illustrated in the preceding section, arises also in the following example.

Let $X_1, \cdots, X_n$ be independently normally distributed with mean θ and common variance which without essential loss of generality we shall take to be equal to 1, and consider the problem of estimating θ, with squared error as loss. The standard estimator is $\delta = \bar{X} = \sum X_i/n$ with risk equal to

$$R_n(\theta) = n^{-1}. \tag{6.1}$$

TABLE 1
Power of $\bar{X}$-test with n observations, matched by t-test with k observations, $\alpha = .05$

$n = 4$

	t-power:			$\bar{X}$-power:	
$4^{1/2}\Delta$	$k = 5$	$k = 6$	$k = 5.6377$	$n = 4$	difference
−.8	.00797	.00606	.00675	.00725	−.00050
0	.05	.05	.05	.05	0
.8	.18549	.21366	.20345	.19910	.00435
1.2	.29964	.35457	.33467	.32821	.00646
1.4	.36606	.43494	.40998	.40328	.00670
1.6	.43638	.51779	.48830	.48211	.00619
2.0	.57976	.67695	.64174	.63876	.00298
2.4	.71191	.80806	.77322	.77491	−.00169
2.8	.81942	.89964	.87058	.87598	−.00540
3.0	.86193	.93094	.90594	.91231	−.00637
3.2	.89690	.95406	.93335	.94004	−.00669
3.4	.92485	.97046	.95394	.96038	−.00644
4.0	.97484	.99361	.98681	.99074	−.00393
4.8	.99585	.99949	.99817	.99920	−.00103

$n = 8$

	t-power:			$\bar{X}$-power:	
$8^{1/2}\Delta$	$k = 9$	$k = 10$	$k = 9.5149$	$n = 8$	difference
−.8	.00762	.00663	.00711	.00725	−.00014
0	.05	.05	.05	.05	0
.8	.19303	.20704	.20024	.19910	.00114
1.2	.31588	.34312	.32991	.32821	.00170
1.4	.38750	.42156	.40504	.40328	.00176
1.6	.46308	.50316	.48372	.48211	.00161
2.0	.61522	.66238	.63950	.63876	.00074
2.4	.75098	.79649	.77441	.77491	−.00050
2.8	.85572	.89223	.87452	.87598	−.00146
3.0	.89485	.92547	.91062	.91231	−.00169
3.2	.92563	.95022	.93829	.94004	−.00175
3.4	.94898	.96791	.95873	.96038	−.00165
4.0	.98634	.99308	.98981	.99074	−.00093
4.8	.99849	.99947	.99899	.99920	−.00021

Most Bayesian analyses presuppose that θ has a prior normal distribution, say with mean μ and variance τ^2. The Bayes estimator resulting from this assumption is

$$\delta' = (\mu + n\tau^2 \bar{X})/(n\tau^2 + 1) \tag{6.2}$$

and its risk function is

$$R_n'(\theta) = [n\tau^4 + (\theta - \mu)^2]/(n\tau^2 + 1)^2. \tag{6.3}$$

The estimator δ' arises also in a different context. Suppose there are available m earlier observations $U_1, \cdots, U_m$ from a normal distribution with unit variance and hopefully the same mean θ. Then instead of δ one might wish to use the estimator $(m\bar{U} + n\bar{X})/(m+n)$. For given $\bar{U}$, this coincides with δ' if we put $\tau^2 = m^{-1}$ and $\mu = \bar{U}$.

In trying to compare the two estimators δ and δ', we note first that for every value of θ, the asymptotic efficiency of δ' relative to δ is 1. Efficiency thus suggests that there is little to choose between the estimators; as we shall see, a deficiency analysis shows instead that the differences between the two estimators may be quite striking and that their comparison presents a rather complex problem.

In view of the equivalence, established in Section 2, of stochastic interpolation with the formally simpler treatment of sample size as a continuous variable, let us suppose that δ is based on $n+d$ observations without restricting d to integral values. Then

$$R_{n+d}(\theta) = \frac{1}{n} - \frac{d}{n^2} + o\left(\frac{1}{n^2}\right) \tag{6.4}$$

while

$$R_n'(\theta) = \frac{1}{n} + \frac{1}{n^2\tau^4}[(\theta-\mu)^2 - 2\tau^2] + o\left(\frac{1}{n^2}\right). \tag{6.5}$$

Thus, putting $D(\theta) = n^2\tau^4[R_n'(\theta) - R_{n+d}(\theta)]$, we have

$$D(\theta) = (\theta - \mu)^2 - 2\tau^2 + d\tau^4. \tag{6.6}$$

Suppose now that we wish to determine the deficiency $d = d(\theta)$ at a fixed given value of θ. This is obtained by setting $D(\theta) = 0$, and yields

$$d = \frac{2}{\tau^2} - \frac{(\theta-\mu)^2}{\tau^4}. \tag{6.7}$$

As θ tends to $+\infty$ or $-\infty$, d tends to $-\infty$, so that for large values of $|\theta|$, δ is greatly superior to δ'. The maximum value of d occurs for $\theta = \mu$ and is $2/\tau^2$; for $\theta = \mu$ and small values of τ, δ' is therefore the much better estimator. (Note however that d takes on large negative values for any value of $\theta \neq \mu$ if τ is small.)

In view of this very strong dependence of d on θ, one cannot expect to find a compromise which is generally satisfactory. Nonetheless, it is interesting to see what results are obtained by different criteria.

As a first possibility, let us suppose that θ is really a random variable with the postulated prior distribution. Then the average number of observations lost as a result of using δ instead of δ' is obtained by determining d so that $ER_{n+d}(\theta) = ER_n'(\theta)$, where the expectation is taken with respect to the assumed normal distribution of θ. Then

$$ER_n'(\theta) = \frac{1}{n} - \frac{1}{\tau^2 n^2} \quad \text{while} \quad ER_{n+d}(\theta) = \frac{1}{n} - \frac{d}{n^2}$$

and the Bayes deficiency is therefore $d = \tau^{-2}$. It is interesting to note that in the non-Bayesian model with m prior observations, $\tau^{-2} = m$ is just the number of observations lost if δ is used instead of δ'.

At the other extreme, in a sense, is the minimax criterion according to which d is determined so as to minimize $\sup|R_{n+d}(\theta) - R_n'(\theta)|$. Neglecting terms of order $o(1/n^2)$, this means minimizing $\sup|D(\theta)|$, where $D(\theta)$ is given by (6.6). Unfortunately, it is seen that this maximum discrepancy is infinite for all values of d, so that the minimax criterion breaks down.

A possible approach to the problem is presented by a compromise between the Bayes and minimax principles. Prior information rarely prescribes a specific distribution. Rather it may indicate that θ lies in an interval, say $I = (\mu - A, \mu + A)$ and that values closer to μ are more likely than more distant values. If symmetry with respect to μ seems reasonable, mathematical convenience suggests a normal distribution which is centered at μ and assigns high probability to the interval I. In comparing δ with δ' we may then be willing to restrict attention to θ in I and minimize $\sup|D(\theta)|$ for $\theta \in I$.

Since both $D(\mu)$ and $D(\mu - A)$ are increasing functions of d, it is easily seen from (6.6) that the unique minimax solution is obtained by putting $D(\mu) = -D(\mu - A)$. The resulting deficiency is

$$d = \frac{2}{\tau^2} - \frac{A^2}{2\tau^4} \tag{6.8}$$

and the corresponding minimax value is

$$\sup_{\theta \in I} |D(\theta)| = A^2/2. \tag{6.9}$$

It follows from (6.5) that $d > 0$ (i.e. the Bayes estimator is "better" in the sense being considered), $d = 0$ or $d < 0$ as

$$\tau < A/2, \quad = A/2, \quad \text{or} \quad > A/2. \tag{6.10}$$

This suggests that, when fitting a prior distribution to a given A, one should choose for τ a value less than $A/2$. Condition (6.10) can be given an alternative formulation. Suppose in deciding on a prior distribution we determine τ, for a given value of A, so that $P\{|\theta - \mu| \leqq A\} = \gamma$ where γ is some fixed probability level typically close to 1. Then $A/\tau = u$ where $\Phi(u) = (1 + \gamma)/2$. Thus

$$d = \frac{2}{\tau^2}\left(1 - \frac{u^2}{4}\right). \tag{6.11}$$

The most interesting case is that of small τ, since in this case there is a substantial difference between the two estimators δ and δ' being compared. As $\tau \to 0$, $d \to +\infty$ or $-\infty$ as u is < 2 or > 2, and remains $= 0$ when $u = 2$. The condition $u = 2$ means that (approximately) $\gamma = .95$. Thus, if we fit a normal prior to the given interval I, it should be such as to assign to I a probability of at least .95.

7. Further possibilities. There are many examples similar to the ones treated in the preceding sections, which could be handled in a completely analogous manner. The approach of Examples 1 and 2 of Section 3 to the estimation of variance extends, for instance, to the estimation of covariance and of higher moments. Similarly, the result of Section 4 concerning confidence sets for the simultaneous estimation of several means generalizes to confidence sets for the mean of a multivariate normal distribution. The method for obtaining the deficiency of Student's one-sample t-test applies also to the other t-tests, for example, the two-sample t-test or the Bartlett-Scheffé test for the Behrens–Fisher problem. More generally, it applies also to substitute t-tests in which the usual denominator is replaced by some other estimator of the standard deviation, provided only this estimator is independent of the numerator. (We plan to discuss such substitute t-tests in this manner in a separate paper.) Finally, the comparison in Section 6 of the standard estimator of a normal mean with a Bayes estimator extends to certain other situations involving exponential families, for example, the estimation of binomial p.

The extensions mentioned above require no new techniques and there are other applications for which this is true, for example, to comparisons in certain design-problems. More interesting perhaps are a number of problems in which the deficiency concept appears to be useful but where its application presents certain technical difficulties stemming from the fact that the computation of deficiency requires higher-order asymptotic terms than we encounter in the usual efficiency analyses. We shall mention only three problems of this type, all of a nonparametric nature: (i) What is the deficiency of the normal scores test or of Van der Waerden's X-test with respect to the t-test? (ii) What is the deficiency of the test based on Kendall's rank correlation coefficient with respect to that based on Spearman's coefficient? (iii) If $X_1, X_2, \cdots$ is a sample from a symmetric distribution centered on θ, what is the deficiency of the Wilcoxon test of $H: \theta = \theta_0$ and of the associated estimator of θ based on all averages $(X_i + X_j)/2$ with $i \leqq j$ relative to that based only on the averages with $i < j$?

In addition to providing useful information for specific comparisons, the deficiency concept raises a number of more general questions, of which we shall again mention three. (i) In parametric situations such as that of Example (i) of the preceding paragraph, there often exist rank tests of efficiency 1. Do there also exist rank tests of deficiency zero? (ii) The asymptotic relative efficiency of two tests is typically the same as that of two estimators based on them. Does this result in any sense carry over to deficiency? (iii) In a situation involving several parameters, consider the likelihood ratio test for the hypothesis specifying the value of one of the parameters when the other parameters are either known or unknown. Under suitable regularity conditions, is it true that there is then a finite deficiency, and can one give a simple formula for it? (A simple counter example is provided by the problem of estimating the range a of the rectangular distribution $R(\theta, \theta+a)$. Here the asymptotic efficiency of the best unbiased estimator with unknown θ to the corresponding estimator with known θ is $2^{-\frac{1}{2}}$. However, in this case the estimators of both a and θ are superefficient.)

REFERENCES

[1] Cochran, W. G. (1937). The efficiencies of the binomial series tests of significance of a mean and of a correlation coefficient. *J. Roy. Statist. Soc. Ser. A* **100** 69–73.

[2] Cramér, Harold (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press.

[3] Fisher, R. A. (1920). A mathematical examination of the methods of determining the accuracy of an observation by the mean error and by the mean square error. *Monthly Notices Roy. Astron. Soc.* **80** 758–770.

[4] Fisher, R. A. (1925). Theory of statistical estimation. *Proc. Cambridge Philos. Soc.* **22** 700–725.

[5] Hodges, J. L., Jr. and Lehmann, E. L. (1956). The efficiency of some nonparametric competitors of the t-test. *Ann. Math. Statist.* **27** 324–335.

[6] Hodges, J. L., Jr. and Lehmann, E. L. (1967a). Moments of chi and power of t. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 187–201.

[7] Hodges, J. L., Jr. and Lehmann, E. L. (1967b). On medians and quasi-medians. *J. Amer. Statist. Assoc.* **62** 926–931.

[8] Johnson, N. L. and Welch, B. L. (1939). Applications of the non-central t-distribution. *Biometrika* **31** 362–389.

[9] Pearson, E. S. (1950). Some notes on the use of range. *Biometrika* **37** 88–92.

[10] Walsh, John E. (1949). On the "information" lost by using a t-test when the population variance is known. *J. Amer. Statist. Assoc.* **44** 122–125.