Z. Wahrscheinlichkeitstheorie verw. Gebiete 30, 117–128 (1974)

Asymptotically Efficient Non-parametric Estimators of Location and Scale Parameters. II

J. Wolfowitz*

9. Introduction

The present paper is a continuation of the paper [1] of the same name. In [1] the authors showed how to construct (asymptotically) efficient estimators of scale and location parameters and of the two jointly, when the form of the density function is unknown to the statistician (i.e., in the non-parametric case). Their estimators are functions of the "middle" $n(q-p)$ observations, where n is the total number of observations and $0<p<q<1$. The estimators are efficient modulo this fact. Since p can be chosen close to zero and q close to 1, the demands of statistical applications would probably be better served by improving this estimator rather than by eliminating the restriction to the middle $n(q-p)$ observations. However, for the purposes of statistical theory and the eventual development of a theory of non-parametric estimation, it seems of some interest to eliminate this "waste" of observations.

In the present paper we construct an estimator of the scale parameter σ which is asymptotically as efficient as the best estimator which can be constructed when the form of the density function is known to the statistician and all the observations are used. (Actually, we estimate the ratio of two σ's, because the assumptions we make are not sufficient to identify σ; see [1] and a remark in Section 15 below. If the parameter σ is identified then the method given below gives an efficient estimator of it.) It will be readily seen that the same method is applicable to estimating a location parameter μ, and μ and σ jointly. The parameter σ was chosen because a choice had to be made (it is not necessary to do both) and because it is perhaps slightly the more difficult of the two[1]. We believe that the method developed is of general interest and that it will be applicable in the development of a general theory of non-parametric estimation which has begun to emerge only recently.

In the present paper we assume familiarity with [1], whose notation and definitions are assumed herewith. Other notation will be added in Section 10, where the assumptions are stated. The numbering of the sections follows that of [1] consecutively. The assumptions will be discussed in Section 15, where the relation of this paper to work by other authors will be discussed.

---

* Research supported by the U.S. Air Force under Grant AF-AFOSR-70-1947, monitored by the Office of Scientific Research.

1 By our method, that is. It has not as yet been estimated by any other method.

After the formulae (1.3) and (1.4) (see [1]) of Bennett, Jung, and Blom were discovered, and even more after the formulae (1.7) and (1.8) of Weiss, it was trivial to conjecture that these formulae could be used to obtain non-parametric estimators of μ and σ. The difficulty was to carry out this program, since the error in estimating just one coefficient exceeds, by an order of magnitude, the error permitted the entire estimator (see [1]). In the same way, it is trivial to conjecture that full efficiency can be obtained by pushing the "cut-off" points p_n and q_n to 0 and 1, respectively, as $n \to \infty$. Carrying out this program, as is done in the papers cited in Section 15 below and in the present paper, is not at all a trivial matter, and encounters a number of difficulties.

We now give an extremely brief outline of the present paper. In Sections 11and 12 we assume that appropriate cut-off points for every n, and all the coefficients $B_j^{(n)}$, are known to the statistician, and that, between these cut-off points, the positive lower bound on the derivative g is known to the statistician. (Of course, in the nature of the problem, this is impossible, but we assume it temporarily.) In these sections we then obtain bounds (which approach zero as $p \to 0$ and $q \to 1$) on the difference between the distribution of the normalized estimator and the desired limiting normal distribution. Thus, if these (impossible) conditions were to be fulfilled, an efficient estimator would be at hand. Throughout this paper, the form of the estimator is always that of [1], but the crucial question always is what the suitable cut-off points are. The final decision is made only in Section 14, after a number of changes in different steps.

In the first half of Section 13 we prove that, if the coefficients $B_j^{(n)}$ are not known to the statistician, but estimated as in [1], then an estimator which is a function of certain of the observed variables is still efficient. In the second half of this section we estimate the cut-off points, so that, with probability approaching one, a satisfactory lower bound on g between the cut-off points, can be given.

This would seem to remove the assumptions with which the argument of Sections 11 and 12 was carried out, assumptions which involve knowledge by the statistician which he cannot possibly possess. Three obstacles still remain:

(9.1) The cut-off points determined are chance variables, not the constants which occur in the proofs of normality, like that of [2], for example.

(9.2) Since the chance cut-off points are functions of the middle observations, the latter are now not necessarily independently and identically distributed with the common truncated distribution, as required by the proofs of normality, e.g., that of [2]. Every method of moving out the cut-off points, as functions of the n observations, so that $p \to 0$ and $q \to 1$, must reckon with this difficulty.

(9.3) One must take into account how many of the $n(q-p)$ observations lie in the prescribed interval. For fixed p and q, as in [1], this was of no consequence for the limiting distribution.

In Section 14 these difficulties are resolved and the final estimator is given. This estimator has in the limit, after normalization, the same distribution as the most efficient estimator which can be constructed as a function of all the observations and with full knowledge of the form of the distribution function.

10. Assumptions

Before giving the assumptions we add some more notations, necessary or useful facts.

Define

$$V(p,q)=\left\{\int_{G^{-1}(p)}^{G^{-1}(q)}\left(y\frac{g'(y)}{g(y)}\right)^2 g(y)\,dy-(q-p)\right.$$
$$+\frac{(G^{-1}(p)\,g(G^{-1}(p)))^2}{p}+\frac{(G^{-1}(q)\,g(G^{-1}(q)))^2}{1-q}$$
$$\left.+2G^{-1}(q)\,g(G^{-1}(q))-2G^{-1}(p)\,g(G^{-1}(p))\right\}^{-1},$$

$$V^*=\left\{\int_{-\infty}^{\infty}\left(y\frac{g'(y)}{g(y)}\right)^2 g(y)\,dy-1\right\}^{-1},$$

$$V'(p,q)=\left\{\int_{G^{-1}(p)}^{G^{-1}(q)}\left(\frac{g'(y)}{g(y)}\right)^2 g(y)\,dy+\frac{g^2(G^{-1}(p))}{p}+\frac{g^2(G^{-1}(q))}{1-q}\right\}^{-1},$$

$$V^{**}=\left\{\int_{-\infty}^{\infty}\left(\frac{g'(y)}{g(y)}\right)^2 g(y)\,dy\right\}^{-1}.$$

The significance of these quantities is as follows: $\sigma^2 V(p,q)$ is the variance of the normal distribution with mean zero which is the limit of the distribution of

$$\sqrt{n}\,(\hat{\sigma}_n(Y)-\sigma),$$

where $\hat{\sigma}_n(Y)$ is the estimator of (1.8). $V'(p,q)$ is the variance of the normal distribution with mean zero which is the limit of the distribution of

$$\sqrt{n}(\hat{\mu}_n(X)-\mu),$$

where $\hat{\mu}_n(X)$ is the estimator of (1.7). Both of these results are derived in [2], and, in a simpler way, in [5]. When g is symmetric about zero, the value of $V'(p,p)$ is supposedly given at the top of p. 150 of [1], but the expression given there involves an algebraic error. The correct expression for $V'(p,q)$ in the general case is given above in this section. V^* and V^{**}, respectively, are of course the variance of the limiting normal distributions of the maximum likelihood estimators of σ and μ, respectively, when the statistician knows the form of g and the latter satisfies the conditions of the "regular" case. These are then the best variances (of the limiting distribution) which the statistician can achieve.

(We take this opportunity to correct a few minor errors in [1]. In Section 8, $V(n)$ should be replaced by V throughout. That V is now called $V'(p,p)$ for the symmetric case, because we now need a finer differentiation. In (7.12) there should be absolute values about $\Delta(n,j)$. In the denominator of (7.10) the factor $(\lambda_{np}+\lambda_{nq})$ should be deleted. A non-trivial error made in [1] was to omit the requirement that g'' satisfy a Lipschitz condition.)

(The error made in $V(n)$ of [1] is carried over into [12], (3.9), (3.13), (3.16), and Lemma 8.)

In the present paper we make the following assumptions:

Assumption 1. $0 < V^* < \infty$. Let T_n' be any estimator of σ which is such that, for every σ, $\sqrt{n}(T_n' - \sigma)$ is asymptotically normally distributed with mean zero and variance $V'(\sigma)$. Then, for (Lebesgue) almost every σ, $V'(\sigma) \geqq \sigma^2 V^*$.

Assumption 2. The set where g is positive is the entire line. The derivatives g', g'', and g''' exist and are bounded above in absolute value.

Assumption 3. As $p \to 0$ and $q \to 1$, $V(p, q) \to V^*$.

Assumption 4. $E|V_i|^3 < \infty$. (For the definition of V_i see the first paragraph of [1].)

Assumption 5. There exists an s, $0 < s < 1$, such that

a) for $x < 0$,

$$G\left(\frac{x}{\sigma}\right) < \left[\frac{1}{\sigma} g\left(\frac{x}{\sigma}\right)\right]^s,$$

and

b) for $x > 0$,

$$1 - G\left(\frac{x}{\sigma}\right) < \left[\frac{1}{\sigma} g\left(\frac{x}{\sigma}\right)\right]^s.$$

The quantity s may depend on σ.

The assumptions will be discussed in Section 15. We will carry out the proof in the next four sections under the following additional assumptions, for the sake of a little simplicity. These additional assumptions will be eliminated in an almost trivial manner in Section 14.

Assumption 2'. The statistician knows K_1, the largest of the bounds in Assumption 2.

Assumption 4'. The statistician knows an upper bound on

$$\frac{E|V_i|^3}{[EV_i^2 - (EV_i)^2]^{\frac{3}{2}}}.$$

Assumption 5'. The statistician knows the number s.

11. Heuristic Introduction to the Proof

In this section we describe some of the basic ideas of the proof in a non-rigorous manner for the sake of easier understanding. These ideas are carried out rigorously in Section 12. The proof is then completed in Sections 13 and 14, which depend on ideas not discussed in this section. Only a superficial familiarity with [1] is needed for Sections 11 and 12.

Let α_i, $i = 1, 2, \ldots$ be a descending sequence of positive numbers such that $\alpha_1 < 1$ and $\alpha_i \to 0$. Assume now that the statistician knows sequences $\{p_i\}$ and $\{q_i\}$ such that

(11.1) always $0 < p_i < q_i < 1$, and $(q_i - p_i) \to 1$, and

(11.2) whenever $\sigma G^{-1}(p_i)<x<\sigma G^{-1}(q_i)$,

$$\alpha_i<\frac{1}{\sigma}g\left(\frac{x}{\sigma}\right)=d(x), \quad \text{say.}$$

Suppose also that the statistician knows the coefficients $B_j^{(n)}$ in the estimator (1.8) of [1], i.e.,

$$(11.3) \quad \hat{\sigma}_n^{(i)}(Y)=\frac{\sum_{j=np_i+1}^{nq_i-1}\frac{1}{n}B_j^{(n)}Y_j^{(n)}+B_{np_i}'^{(n)}Y_{np_i}^{(n)}+B_{nq_i}'^{(n)}Y_{nq_i}^{(n)}}{\sum_{j=np_i+1}^{nq_i-1}\frac{1}{n}B_j^{(n)}G^{-1}\left(\frac{j}{n}\right)+B_{np_i}'^{(n)}G^{-1}(p_i)+B_{nq_i}'^{(n)}G^{-1}(q_i)}.$$

(Of course, the problem is such that the statistician cannot possibly know these things. In Sections 13 and 14 we will remove these assumptions.)

We will then find an integer N_i such that, when $n \geqq N_i$, the distribution of

$$(11.4) \qquad \sqrt{n}\,(\hat{\sigma}_n^{(i)}(Y)-\sigma)$$

differs from the normal distribution with mean zero and variance $\sigma^2 V(p_i, q_i)$, by at most α_i, uniformly in the argument of the distribution function and all G which satisfy out assumptions.

From the determination of N_i it will follow that:

(11.5) The above conclusion about the uniform approach of the distribution of (11.4) to its limit is a fortiori true if, in $\hat{\sigma}_n^{(i)}(Y)$, we replace p_i and q_i, respectively, by p and q such that $p_i<p<q<q_i$.

(11.6) We may choose the N_i strictly increasing. Then $N_i \uparrow \infty$.

For any n let $i(n)$ be the largest i such that $N_i \leqq n$. It follows from the above that the distribution of

$$(11.7) \qquad \sqrt{n}(\hat{\sigma}_n^{(i(n))}(Y)-\sigma)$$

approaches the normal distribution with mean zero and variance $\sigma^2 V^*$. Hence $\hat{\sigma}_n^{i(n)}(Y)$ is an (asymptotically) efficient estimator of σ.

The coefficients $B_j^{(n)}$ are, of course, unknown to the statistician. We will estimate them as in [1]. We will have to prove, and will do so in Section 13, that

(11.8) The ratio of the estimators of σ_1 and σ_2, with the coefficients estimated as in [1], is an asymptotically efficient estimator of $\frac{\sigma_1}{\sigma_2}$. (This was also proved in [1], but there α_i did not approach zero.) The distribution function of this ratio differs from the distribution function of the ratio of estimators with coefficients known, by at most a multiple of α_i.

(11.9) We will determine $\{p_i\}$ and $\{q_i\}$ so that (11.1) and (11.2) are satisfied, not literally, but in a probabilistic sense. This will cause some difficulties which will have to be eliminated by additional arguments. This will be done in the second half of Section 13 and in Section 14.

12. Proof of the Theorem when the Statistician Knows $\{p_i\}$, $\{q_i\}$, and the Coefficients $B_j^{(n)}$

Our proof in this case leans heavily on the clever proof of normality in [2], p. 4–6, and is actually a refinement of this proof for purposes not present in [2]. For this reason and to avoid needless repetition we assume familiarity with these pages of [2] and will indicate where modifications are to be made. Our problem, not present in [2], is to obtain an N_i large enough so that the distribution of (11.7) is, uniformly in the argument of the distribution function, within α_i of the normal distribution with mean zero and variance $\sigma^2 V(p_i, q_i)$.

Let p and q of [2] now be p_i and q_i, respectively. We now concern ourselves with the conditional distribution of S_n of [2], given that $V_n = v$ and $W_n = w$, with $K\left(\frac{j}{n}\right)$ of [2] equal our $B_j^{(n)}$. From the definition of V_n and W_n of [2] it follows that there exists an absolute (i.e., independent of G) constant H_i and an N_i such that, for $n \geqq N_i$,

$$P\{|V_n| < H_i, |W_n| < H_i\} \geqq 1 - \frac{\alpha_i}{6}. \tag{12.1}$$

The inequality in (12.1) provides the bounded region in the (v, w) plane discussed in [2], p. 6. The expression $Q''(\theta_j, n)$ of [2] comes from the derivative of $B_j^{(n)}$ of [1]. The latter (derivative) is a sum of terms, each of which consists of a product of non-negative powers of g', g'', g''', and G^{-1}, divided by a positive power of g. Now g', g'', g''' are, by Assumption 2, bounded above in absolute value. $\left|G^{-1}\left(\frac{j}{n}\right)\right|$ is bounded above by $|G^{-1}(\frac{1}{2})| + \alpha_i^{-1}\sigma^{-1}$ which, for i sufficiently large, is less than $i\alpha_i^{-1}$. Consequently, in the interval (p_i, q_i), $Q''(\theta_j, n)$ of [2] is bounded above in absolute value by a known multiple of a negative power of α_i. Increasing N_i, if necessary, this bound can be made less than N_i^{γ} for a $\gamma < \frac{1}{2}$.

It follows from [3], Lemma 2, that, for i sufficiently large,

$$P\left\{\max_{np_i \leqq j \leqq nq_i} \left[Y_j - G^{-1}\left(\frac{j}{n}\right)\right]^2 < (n\alpha_i^2)^{-1}\right\} > 1 - \frac{\alpha_i}{6}. \tag{12.2}$$

In the notation of [2], therefore, we have then

$$|\delta_n(v, w)| < \frac{n \cdot n^{\gamma} \cdot \max\limits_{np_i \leqq j \leqq nq_i} \left[Y_j - G^{-1}\left(\frac{j}{n}\right)\right]^2}{2n^{\frac{1}{2}}}. \tag{12.3}$$

Hence, increasing N_i if necessary, we have

$$P\left\{|\delta_n(v, w)| < \frac{\alpha_i}{6}\right\} > 1 - \frac{\alpha_i}{6}. \tag{12.4}$$

We can now directly consider the conditional distribution of S_n of [2], given that $V_n = v$ and $W_n = w$. From the argument of [2], especially p. 6, line 6, we see that everything now depends on the distribution of T_n of [2]. The function Q ([2], p. 3, line 9) is still at our disposal, subject to the condition given in [2]. We define $Q(z)$ as

(12.5)
$$\int_{G^{-1}(\frac{1}{2})}^{z} K(G(y))\,dy.$$

The argument by which we proved Q'' bounded by N_i^γ applies a fortiori to Q', and we obtain that

(12.6)
$$|Q(z)| < |z - G^{-1}(\tfrac{1}{2})|\, N_i^\gamma.$$

From (12.6), the definition of the chance variables Z_i of [2], the definition of T_n, Assumption 4′, and the Berry-Esseen theorem, we conclude that, for N_i sufficiently large, the distribution of T_n of [2] differs from the normal distribution with mean zero and variance one by less than $\frac{\alpha_i}{6}$, uniformly for v and w in the set $|v| < H_i$, $|w| < H_i$, for i large.

The chance variable (11.7) is a linear function of T_n, V_n, and W_n of [2]. Following the proof of [6], p. 367–370, it is not difficult to prove that, when N_i is sufficiently large, because of Assumptions 1 and 2, the distribution of the normalized chance variable V_n differs from its limiting normal distribution by less than $\frac{\alpha_i}{6}$, and the conditional distribution of the normalized chance variable W_n, given that $|V_n| < H_i$, differs from its limiting normal distribution by less than $\frac{\alpha_i}{6}$. We have already proved the corresponding result for the conditional distribution of T_n, given the event in (12.1). Now, it is easy to see that, if the distribution of a chance variable Z' differs from a distribution G', say, by less than β, then, for any non-zero constant c, the distribution of cZ' differs from the distribution G'_0 $\left(G'_0(x) \equiv G'\left(\frac{x}{c}\right)\right.$ when $c > 0$, with a corresponding definition when $\left.c < 0\right)$ by less than β. From these facts it is not difficult to see that, when N_i is sufficiently large, the distribution of (11.4) differs from the normal distribution with mean zero and variance $\sigma^2\, V(p_i, q_i)$ by at most α_i. Since $V(p_i, q_i) \to V$ it follows that $\hat{\sigma}_n^{i(n)}(Y)$ is an asymptotically efficient estimator of σ.

If G, $\{p_i\}$, and $\{q_i\}$ were known to the statistician our work would now be finished. Of course, they are not known, and, in the nature of the problem, cannot be known.

13. Proof of (11.8). Determination of p_i and q_i

We begin by proving (11.8). It follows from Lemma 2 of [3] that, for large i, a "belt" of constant half-thickness $(\alpha_i n^{\frac{1}{2}})^{-1}$, about the graph of the distribution function $G\left(\frac{x}{\sigma}\right)$, will include the graph of the empiric distribution function of $Y_1, \ldots, Y_n$ with probability $\geqq 1 - \alpha_i$. We now retrace the argument of [1], Section 5, conditioned upon the event in the last sentence. In the interval (p_i, q_i), $G^{-1}\left(\frac{j}{n}\right)$

is bounded above in absolute value by $|G^{-1}(\frac{1}{2})|+\alpha_i^{-1}\sigma^{-1}$, which is less than $i\alpha_i^{-1}$ for i sufficiently large. We will now estimate the errors incurred in estimating

$$\sigma G^{-1}\left(\frac{x}{\sigma}\right), \quad \frac{1}{\sigma} g\left(\frac{x}{\sigma}\right), \quad \frac{1}{\sigma^2} g'\left(\frac{x}{\sigma}\right), \quad \text{and} \quad \frac{1}{\sigma^3} g''\left(\frac{x}{\sigma}\right)$$

as in [1], Section 5. This will give us simultaneous bounds on the errors of all the $C(n, i, j)$ of Section 3.

The error in estimating $c(n, i, j)$ of [1] is, from the construction of the "belt", not greater than $(\alpha_i^2 n^{\frac{1}{2}})^{-1}$.

Proceeding as in (5.5), the first of the errors in estimating $d(n, i, j)$ of [1] is not greater than $2(\alpha_i^2 n^{\frac{1}{4}})^{-1}$. The second of the errors is not greater than $K_1(\alpha_i n^{\frac{1}{4}})^{-1}$, where K_1 is the largest of the bounds on $|g'|$, $|g''|$, $|g'''|$. For large enough i, the larger of these two bounds is $2(\alpha_i^2 n^{\frac{1}{4}})^{-1}$.

We continue in this manner as in Section 5. The details are not important, because it is obvious what the conclusion will be, and the actual computations are tedious. The conclusion is that, for large i, the error in estimating $C(n, i, j)$ of [1] is, with probability $\geqq 1-\alpha_i$, simultaneously for all j such that Y_j lies in the interval $(\sigma G^{-1}(p_i), \sigma G^{-1}(q_i))$, bounded in absolute value by a multiple of $n^{-\frac{1}{16}}$ multiplied by a negative power of α_i. (In [1] there was no need to take into account this negative power of α_i, because α_i did not approach zero.) Increasing N_i, if necessary, we can make this bound less than $n^{-\frac{1}{32}}$.

Let $K_i' = \alpha_i^{-\frac{1}{2}}$. Then

$$P\{\sigma G^{-1}(p_i) < Y_i < \sigma G^{-1}(q_i) \text{ when } np_i + K_i'\sqrt{np_i} < j < nq_i - K_i'\sqrt{n(1-q_i)}\} > 1 - 2\alpha_i. \tag{13.1}$$

From now on in this section we shall use the estimator (11.3) with np_i replaced by $np_i + K_i'\sqrt{np_i}$ and nq_i replaced by $nq_i - K_i'\sqrt{n(1-q_i)}$.

We now proceed almost exactly as in Section 3 of [1]. Let $\hat{\sigma}_n^{i(n)}(Y_1)$ and $\hat{\sigma}_n^{i(n)}(Y_2)$ be, respectively, estimators of σ_1 and of σ_2, computed as though G (and hence the coefficients B_j) were known. Let W_1 be the ratio of the corresponding estimators with the coefficients estimated as in [1], Section 5. As in [1], Section 3, using the bounds obtained above, we conclude that, with probability $\geqq 1 - 6\alpha_{i(n)}$,

$$\sqrt{n}\left(W_1 - \frac{\hat{\sigma}_n^{i(n)}(Y_1)}{\hat{\sigma}_n^{i(n)}(Y_2)}\right) < 4(\alpha_{i(n)}^2 n^{\frac{1}{32}})^{-1}. \tag{13.2}$$

Increasing N_i, if necessary, the bound in (13.2) can be made less than α_i. This proves (11.8).

We now turn to (11.9). Increase N_i, if necessary, so that $N_i > \alpha_i^{-3}$. Let $H_n(\cdot)$ be the empiric distribution function of $Y_1, \ldots, Y_n$. Using the method of [4] we construct a confidence belt of constant thickness with confidence coefficient $\geqq 1-\alpha_i$. Increasing N_i, if necessary, we make the half-thickness of this belt less than α_i^2. From this we obtain a maximal x-interval, say J_0, such that, with confidence coefficient $\geqq 1-\alpha_i$, for every $x \in J_0$,

$$\min[H(x), 1-H(x)] \geqq \alpha_i^s. \tag{13.3}$$

It follows from Assumption 5 that $d(x) \geqq \alpha_i$ for $x \in J_0$, with probability $\geqq 1-\alpha_i$. Let J^* be the set of indices j such that $Y_j \in J_0$. Let

$$j_1 = \min\{j | j \in J^*\}, \qquad j_2 = \max\{j | j \in J^*\} \tag{13.4}$$

and

$$p_i = \frac{j_1 + K_i' n^{\frac{1}{2}}}{n}, \qquad q_i = \frac{j_2 - K_i' n^{\frac{1}{2}}}{n}. \tag{13.5}$$

Then

$$P\{d(x) \geqq \alpha_i \text{ whenever } \sigma G^{-1}(p_i) < x < \sigma G^{-1}(q_i)\} > 1 - 3\alpha_i. \tag{13.6}$$

It follows, from the above construction of p_i and q_i, by a tedious but obvious argument (for which, in the paragraph which follows (13.2), we made the half-thickness of the confidence belt less than α_i^2), that

$$(q_i - p_i) \quad \text{converges stochastically to one}. \tag{13.7}$$

However, this is not (11.9), because the p_i and q_i we have just constructed are chance variables. Moreover, these p_i and q_i are not necessarily independent of the Y_j with $j_1 \leqq j \leqq j_2$, so that the latter Y_j cannot be used as in the theory developed in this section. Section 14 is devoted to overcoming these difficulties. When this will have been done the proof of the theorem will be complete.

14. Completion of the Proof

Let $\{N_i\}$ be the sequence hitherto obtained, and replace each N_i by $2N_i^2$. This is to be the final sequence $\{N_i\}$. Always, as before, $i(n)$ is the largest i such that $N_{i(n)} \leqq n$. Write $n' = n - \sqrt{n}$. We have, from the construction of the final sequence $\{N_i\}$, that always

$$n' p_i + K_i' \sqrt{n' p_i} < n' p_i + n' \sqrt{p_i} < 2n' \sqrt{p_i} \tag{14.1}$$

and

$$n'(1-q_i) + K_i' \sqrt{n'(1-q_i)} < n'(1-q_i) + n' \sqrt{1-q_i} < 2n' \sqrt{1-q_i}. \tag{14.2}$$

Define

$$P\{p_i \geqq \tfrac{1}{64}\} = \beta_n, \qquad P\{q_i \leqq 1 - \tfrac{1}{64}\} = \beta_n'. \tag{14.3}$$

Since p_i and q_i converge stochastically to zero and one, respectively, $\beta_n + \beta_n' \to 0$.

We are now ready to give the definition of our final estimator and to prove the desired result.

The original observed (i.i.d.) chance variables are $V_1, \ldots, V_n$. Let $Y_1, \ldots, Y_{n'}$ *now* be the chance variables $V_{\sqrt{n}+1}, \ldots, V_n$ ordered in ascending size. We determine p_i and q_i in the manner described in Section 13 as functions of $V_1, \ldots, V_{\sqrt{n}}$. From the construction of Section 13, Chebyshev's inequality, (14.1), and (14.2), it follows that the probability exceeds $1 - 4\alpha_i - \beta_n - \beta_n'$ that $p_i < \frac{1}{64}$, $1 - q_i < \frac{1}{64}$, and

$$Y_{2n'\sqrt{p_i}}, \ldots, Y_{n'(1-2\sqrt{1-q_i})} \tag{14.4}$$

lie in an interval in which $d(\cdot) > \alpha_i$. Our final estimator is defined to be the estimator of [1] in terms of the variables (14.4), with, of course, the coefficients B_j estimated as in [1]. It follows from the final definition of N_i and Section 13 that the distribution of our normalized estimator of σ, i.e., of $\sqrt{n}$(estimator $-\sigma$), when the chance variables p_i and q_i are fixed, differs from the normal distribution with mean zero and variance $\sigma^2\, V(2\sqrt{p_i}, 1-2\sqrt{1-q_i})$ by at most α_i, on a set of values of p_i and q_i whose probability exceeds $1-4\alpha_i-\beta_n-\beta_n'$ for large i. From Assumption 3 it follows that $V(2\sqrt{p_i}, 1-2\sqrt{1-q_i})$ converges stochastically to V^*. Since $\alpha_i \subset 0$, the limiting distribution of our normalized estimator of σ is normal, with mean zero and variance $\sigma^2\, V^*$, the smallest (for (Lebesgue) almost every σ) variance which can be attained by an asymptotically normal estimator even when the statistician knows G. The corresponding conclusion holds for our estimator of $\frac{\sigma_1}{\sigma_2}$. This proves the desired result.

15. Miscellaneous Remarks

First we eliminate Assumptions 2′ and 4′. To do this, compute N_i as if K_1 were i, and i were also an upper bound on the expression in Assumption 4′. For all i sufficiently large this procedure will be correct. To eliminate Assumption 5′, compute N_i as if $s=(-\log \alpha_i)^{-\frac{1}{2}}$. For all i sufficiently large this, too, will be correct. Thus we have eliminated Assumptions 2′, 4′ and 5′ in a trivial manner. Another method of doing this is to estimate the several quantities from a "wasted" subsample of size $\sqrt{n}$.

Our program for proving the (asymptotic) efficiency of our non-parametric estimator of σ is to show that its limiting distribution is the same as that of the best estimator which can be constructed even when the statistician knows g. Assumption 1 says that the maximum likelihood estimator is efficient. Sufficient conditions for the latter are known (e.g., [10, 11]), but necessary conditions are not known and, in the nature of things, are not likely to be known. It is clear, however, that this assumption implies conditions which are not necessarily independent of the conditions in the other assumptions. The first part of Assumption 2 is usually made in proofs of properties of the maximum likelihood estimator.

Assumption 3 is perhaps essential for the problem. Because of the obvious difficulties in the tails of the distribution, where g is small, it would seem that one is forced to omit the end observations and move out with p and q at the proper rate only.

Assumption 4 could be weakened by requiring only the finiteness of a $(2+\delta)$-th moment. This assumption is probably not an essential one and may be necessary only because of the particular method of proof employed.

Assumption 5 says in effect that the tails of the distribution must not approach zero too slowly. It is probable that it is needed because of the particular method of proof. An assumption about the monotonicity of g would render it unnecessary, as would other assumptions.

In [7–9] the authors give non-parametric estimators of μ whose limiting distribution after normalization has variance V^{**}. The first result of this kind was in [7], which is a brilliant tour de force not likely to be applicable to other problems. Nor has it yet been demonstrated that the methods of the interesting papers [8] and [9] will solve the problem of estimating σ, or of estimating μ and σ jointly, or of estimating the appropriate functions of these parameters when they are not identified. Incidentally, it is not easy to give natural conditions under which σ will be identified. The assumptions of these papers and those of the present paper are not directly comparable, since no set is uniformly stronger or weaker than another.

We devote a few words to Assumption 2. The actual estimator contains no derivatives, but only difference quotients. Preliminary work by the author suggests that one may be able to dispense with at least some of these derivatives. Assumption 1 may well have implicit consequences about derivatives. Also, in a non-parametric problem, where the statistician does not know the function g, is it likely that he would know that g is symmetric[2]? How shall we compare such an assumption with Assumption 2?

Until recently, the theory of non-parametric estimation consisted of a number of ad hoc procedures for which no optimality properties had been proved and, most probably, could not be proved because the procedures were not really optimal. With the publication of the two pioneering papers of this series and then of the subsequent papers, asymptotically optimal estimation procedures have been obtained in all papers for μ, and in one paper for μ, σ, and μ and σ jointly. Now it is obvious that no statistician dealing with a practical problem will ever employ the estimators of [7, 1, 8, 9], or the present paper. They are too complicated to compute, and the convergence rate of their distributions to their limits is too slow. Their value lies in their being existence theorems, as it were, that optimal estimators do exist and can be obtained in these different ways. It seems to the author that a next big step in the theory would be to get away from the limitation to scale and location parameters, in the direction of the general parameters treated in the parametric theory. The method of proof of the present paper may lend itself to this purpose.

When a comprehensive non-parametric theory of estimation becomes available, it will be desirable, on the one hand, to reduce the regularity assumptions, and, on the other, to make the theory accessible for practical purposes, e.g., by speeding up convergence and making compromises in the interest of computational simplicity. It may then happen that, by using estimators which are functions of the middle $n(q-p)$ observations with $(q-p)$ close to unity, one will gain much practical convenience and applicability, at the expense of a small loss of efficiency.

Acknowledgement. The author gratefully acknowledges valuable communications with Professor Lionel Weiss. The formulae for the V's given in Section 10 were furnished by him. Interesting conversations were held with Miss Abigail Sachs.

---

[2] Except in [1] (and the present paper), μ has always been estimated under the assumption that g is symmetric. This is also the case in the manuscript by C. J. Stone just received by us.

References

1. Weiss, L., Wolfowitz, J.: Asymptotically efficient non-parametric estimators of location and scale parameters. Z. Wahrscheinlichkeitstheorie verw. Gebiete **16**, 134–150 (1970)
2. Weiss, L.: On the asymptotic distribution of an estimate of a scale parameter. Naval Res. Logist. Quart. **10**, 1-9 (1963)
3. Dvoretzky, A., Kiefer, J., Wolfowitz, J.: Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. Ann. Math. Statist. **27**, 642–669 (1956)
4. Wald, A., Wolfowitz, J.: Confidence limits for continuous distribution functions. Ann. Math. Statist. **10**, 105–118 (1939)
5. Weiss, L.: Asymptotic properties of maximum likelihood estimators in some non-standard cases. J. Amer. Statist. Assoc. **66**, 345–350 (1971)
6. Cramer, H.: Mathematical methods of statistics. Princeton: Princeton University Press 1946
7. Van Eeden, C.: Efficiency-robust estimation of location. Ann. Math. Statist. **41**, 172–181 (1970)
8. Fabian, V.: Asymptotically efficient stochastic approximation; the RM case. Ann. Statist. **1**, 486–495 (1973)
9. Sacks, J.: An asymptotically efficient sequence of estimators of a location parameter. (To appear in Ann. Statist.)
10. LeCam, L.: Les proprietes asymptotiques des solutions de Bayes. Publ. Inst. Statist. Univ. Paris 7, fascicule 3–4, 17–35
11. Bahadur, R.R.: On Fisher's bound for asymptotic variances Ann. Math. Statist. **35**, 1545–1552 (1964)
12. Weiss, L., Wolfowitz, J.: Optimal, fixed length, non-parametric sequential confidence limits for a translation parameter. Z. Wahrscheinlichkeitstheorie verw. Gebiete **24**, 203–209 (1972)

J. Wolfowitz
Department of Mathematics
University of Illinois
Urbana, Ill. 61801 USA

(Received October 12, 1973)