ON MULTISTAGE ESTIMATION[1]

By J. R. Blum and Judah Rosenblatt[2]

University of New Mexico

1. Introduction. Let $\mathfrak{F}$ be a family of distribution functions and let $\theta(\cdot)$ be a real-valued functional defined on $\mathfrak{F}$. In this paper we shall be concerned for the most part with the problem of finding a confidence interval of preassigned length and confidence for $\theta(F)$ based on a sample from $F \varepsilon \mathfrak{F}$. For simplicity of notation we will assume that $\mathfrak{F}$ consists of univariate distributions. It will be clear that this restriction is not necessary.

Apparently Dantzig [3] was the first to point out that many such problems cannot be solved in a single stage of estimation; i.e. it is impossible to prescribe an integer n and give a confidence interval of preassigned length and confidence based on a sample of size n. Bahadur and Savage [2] showed that if $\mathfrak{F}$ is the class of all distributions for which the mean exists, it is impossible to obtain a confidence interval of prescribed length and confidence for the mean even with a purely sequential scheme. Intuitively this follows from the fact that no matter what data have been observed, there can exist a "small spike" close to $\pm\ \infty$ which affects the mean, but is not likely to affect the data. Farrell [6] showed that a purely sequential scheme is both necessary and sufficient for estimation of the median within the class of distributions possessing a unique median. This is plausible because to pin down the median, the sample median must be closely surrounded by sufficiently many other observations. Though with probability one this will occur, the necessary sample size is not determinable if it has not yet occurred. Farrell's results are actually considerably deeper, since he obtains the order of magnitude of the minimum expected sample size as the density at the median becomes small.

The earliest result of a positive nature is the paper by Stein [9] who gave a two-stage sampling procedure for estimation of the mean of a normal distribution with unknown variance. Graybill [7] gave sufficient conditions for two-stage estimation in certain parametric cases, while Weiss [10] showed that a two-stage scheme suffices for estimation of quantiles when $\mathfrak{F}$ is the class of unimodal distributions. Birnbaum and Healy [3] considered the problem of two-stage unbiased point estimation with fixed variance. Abbott and Rosenblatt [1] gave sufficient conditions for two-stage estimability with one observation on the first stage. A number of other papers treat these and related problems, e.g. Matthes [8].

Unless otherwise specified we shall assume throughout that there is available a sequence $X_1, X_2, \cdots$ of independent random variables with common distri-

Received September 29, 1962.

[1] Research supported by the National Science Foundation, Grant NSF G-21219.

[2] Sandia Corporation. Work performed under the auspices of the United States Atomic Energy Commission.

bution function $F \varepsilon \mathfrak{F}$ and a consistent sequence $\{\theta_n^*\} = \{\theta_n(X_1, \cdots, X_n)\}$ of estimators for $\theta(F)$. (Consistent estimators are readily obtained for most statistical problems by classical methods such as maximum likelihood.) Assume also that $0 < \alpha < 1$ and $\delta > 0$ are given.

DEFINITION. An *n-stage scheme* is determined by a positive integer m_1, a positive integer-valued measurable function m_2 on R^{m_1} (Euclidean m_1-dimensional space), and for each $i = 3, 4, \cdots, n$ positive integer-valued measurable functions $m_{i,j}$ on R^j for each $j = 1, 2, \cdots$. The scheme consists of

the observation of $Y_1 = (X_1, \cdots, X_{m_1})$ on stage 1,

the observation of Y_2, the next $m_2(Y_1)$ of the random variable X_i on stage 2,

the observation of Y_3, the next $m_{3,m_1+m_2}(Y_1, Y_2)$ of the random variables X_i on stage 3, and so on.

Inductively letting $g_{j-1}(Y_1, \cdots, Y_{j-2})$ be the random variable whose value is the number of random variables X_i observed on the first $j - 1$ stages, the jth stage consists in observing Y_j, the next $m_{j,g_{j-1}(Y_1,\cdots,Y_{j-2})}(Y_1, \cdots, Y_{j-1})$ of the X_i, for $j = 3, \cdots, n$.

We consider two-stage estimation in Section 2 and n-stage estimation in Section 3. The remainder of the paper is devoted to applications and examples. No attempt is made here to determine schemes which are optimal in any sense. Rather the emphasis is on constructive existence theorems. We feel that some of the schemes suggested here may not be too inefficient at those F which would normally demand a sample size far exceeding the first sample size (e.g., for which $m(F, \alpha, \delta)$ exceeds four times the first sample size) if one defines efficiency at a given F to be

$$m(F, \alpha, \delta)/E_F(N)$$

where N is the sample size and $m(F, \alpha, \delta)$ is, in a sense to be defined below, the optimal sample size for the problem at F.

2. Two-stage estimation. For each positive integer k we shall denote by F_k the product distribution function on Euclidean k-space induced by F, and by P_{F_k} the corresponding probability measure. P_{F_∞} is similarly defined.

Since $\{\theta_n(X_1, \cdots, X_n)\}$ is a consistent sequence it follows that for each $F \varepsilon \mathfrak{F}$, $\delta > 0$, and $\gamma \varepsilon (0, 1)$, there exists a smallest positive integer $m(F, \gamma, \delta)$ such that for all $n \geqq m(F, \gamma, \delta)$ we have

$$P_{F_n}\{|\theta_n(X_1, \cdots, X_n) - \theta(F)| \leqq \delta\} \geqq 1 - \gamma.$$

(We note that essentially if for each γ, δ, $\sup_{F\varepsilon\mathfrak{F}} m(F, \gamma, \delta) = \infty$, [i.e. if the consistency of $\{\theta_n(X_1, \cdots, X_n)\}$ is not uniform in F] then one cannot expect fixed-precision estimation in one stage of sampling.)

Let $\mathfrak{F}_{m,\gamma,\delta} = \{F \varepsilon \mathfrak{F}: m(F, \gamma, \delta) \geqq m\}$. Note that $\mathfrak{F}_{m,\gamma,\delta} \subset \mathfrak{F}_{n,\gamma,\delta}$ for $m \geqq n$ and $\mathfrak{F}_{0,\gamma,\delta} = \mathfrak{F}$.

THEOREM 1. *Suppose there exists a decreasing sequence $\{N_j\}$ of Borel subsets of R^k such that*

(i) $N_0 = R^k$ *and* $\lim_{j\to\infty} P_{F_k}\{N_j\} = 0$ *for every* $F \varepsilon \mathfrak{F}$, *and*

(ii) *there exists* $\gamma \varepsilon (0, \alpha)$ *and for each integer* j *a positive integer* n_j *such that*

$$\inf_{F\varepsilon\mathfrak{F}_{n_j,\gamma,\delta}} P_{F_k}\{N_j\} > (1-\alpha)/(1-\gamma).$$

(Note that we may assume n_j *to be strictly increasing, and shall do so.) Then there exists a two-stage procedure with* k *observations in the first stage for constructing a confidence interval for* $\theta(F)$, *of length* 2δ *and confidence* $1-\alpha$.

PROOF. We shall prove the theorem by actually constructing a confidence interval with the desired properties. Let $X^{(1)} = (X_1, \cdots, X_k)$ be observed in the first stage of sampling. From (i) it follows that with probability one there exists a positive integer-valued random variable J such that $X^{(1)} \varepsilon N_i$ for $i \leqq J-1$ and $X^{(1)} \not\varepsilon N_J$. Referring to (ii) let the second sample size be n_j and let the confidence interval be

$$[\theta_{n_J}(X_{k+1}, \cdots, X_{k+n_J}) - \delta,\ \theta_{n_J}(X_{k+1}, \cdots, X_{k+n_J}) + \delta].$$

Note that this interval has length 2δ. Also clearly n_J is a random variable (i.e. measurable) and hence the set

$$\{\theta_{n_J}(X_{k+1}, \cdots, X_{k+n_J}) \leqq a\} = \bigcup_{n=1}^{\infty}\{\theta_n(X_{k+1}, \cdots, X_{k+n}) \leqq a, n_J = n\}$$

is measurable for each real number a. Thus $\theta_{n_J}(X_{k+1}, \cdots, X_{k+n_J})$ is a random variable. It remains to be shown that the interval has the proper confidence.

Now note that there exists a unique s such that $F \varepsilon \mathfrak{F}_{s,\gamma,\delta}$ and $F \not\varepsilon \mathfrak{F}_{s+1,\gamma,\delta}$, and a unique r such that $n_{r-1} \leqq s < n_r$. Then $F \varepsilon \mathfrak{F}_{n_{r-1},\gamma,\delta}$ and hence $P_{F_k}\{N_{r-1}\} \geqq (1-\alpha)/(1-\gamma)$. It thus follows from the definition of n_J that $P_{F_k}\{n_J \geqq n_r\} \geqq (1-\alpha)/(1-\gamma)$. Consequently

$$\begin{aligned}
&P_{F_\infty}\{|\theta_{n_J}(X_{k+1}, \cdots, X_{k+n_J}) - \theta(F)| \leqq \delta\} \\
&\quad= \sum_{n=0}^{\infty} P_{F_{n+k}}\{|\theta_n(X_{k+1}, \cdots, X_{k+n}) - \theta(F)| \leqq \delta \mid n_J = n\}P_{F_k}\{n_J = n\} \\
&\quad\geqq \sum_{n=n_r}^{\infty} P_{F_{n+k}}\{|\theta_n(X_{k+1}, \cdots, X_{k+n}) - \theta(F)| \leqq \delta \mid n_J = n\}P_{F_k}\{n_J = n\} \\
&\quad= \sum_{n=n_r}^{\infty} P_{F_n}\{|\theta_n(X_{k+1}, \cdots, X_{k+n}) - \theta(F)| \leqq \delta\}P_{F_k}\{n_J = n\} \\
&\quad\geqq (1-\gamma)\sum_{n=n_r}^{\infty} P_{F_k}\{n_J = n\} \geqq (1-\gamma)(1-\alpha)/(1-\gamma) = 1-\alpha.
\end{aligned}$$

The second equality follows from the independence of $(X_{k+1}, \cdots, X_{k+n})$ and n_J, and the second inequality from the fact that $F \not\varepsilon \mathfrak{F}_{n_r,\gamma,\delta}$. The theorem is proved.

A word or two about the intuitive idea behind the theorem might be in order. It must be true, if a two-stage scheme is possible, that the necessity for a large sample size be "reflected" in the behavior of the first sample—i.e. in the distribution of the first sample. Such a property is most easily phrased in terms of k dimensional sets. We can then use the first sample to "pin down" an integer s such that we are reasonably certain that $F \varepsilon \mathfrak{F}_{s+1,\gamma,\delta}$. Once we know this we are reasonably certain that $\theta(F)$ can be precisely estimated with an additional s observations. In practice the problem is thus reduced to characterising the $\mathfrak{F}_{m,\gamma,\delta}$ and finding a sequence $\{N_j\}$ of sets which satisfy the hypotheses of the theorem. In many parametric families, e.g. the normal family, and in some larger classes, e.g. the class of unimodal distributions, such sequences are quite apparent.

We finally mention that the converse of the reasoning above led to some of the results of Section 3.

3. n-stage estimation. In this section we give sufficient conditions so that an estimation problem can be solved in n stages, and show that in some situations n stages may be required. The latter result was also proved by Kiefer and Weiss in unpublished work using a different technique.

DEFINITION. $\mathfrak{F}$ is said to be an *n-stage* family if for each $\delta > 0$, $\alpha \varepsilon (0, 1)$ there exists an n stage scheme for estimating $\theta(F)$ with an interval of length 2δ and confidence $1 - \alpha$. If $\mathfrak{F}$ is an n-stage family but not an $(n - 1)$-stage family we shall refer to it as a *true n-stage family.*

THEOREM 2. *Suppose* $\mathfrak{F} = \bigcup_m \mathfrak{F}_m$ *where*

(i) $\mathfrak{F}_m$ *is an* $(n - 1)$*-stage family for each* m *such that the final confidence interval for* $\theta(F)$ *is based on the same consistent sequence* $\{\theta_n(X_1, \cdots, X_n)\}$.

(ii) *There exists a decreasing sequence* $\{N_j\}$ *of Borel subsets of* R^k *such that* $N_0 = R^k$ *and* $\lim_{j\to\infty} P_{F_k}\{N_j\} = 0$ *for each* $F \varepsilon \mathfrak{F}$.

(iii) *For each* $\epsilon > 0$ *and each integer* j *there exists an integer* $m_j(\epsilon)$ *such that*

$$\inf_{\substack{F \varepsilon \bigcup \mathfrak{F}_i \\ i \leqq m_j(\epsilon)}} P_{F_k}\{N_j\} > 1 - \epsilon.$$

Then $\mathfrak{F}$ *is an n-stage family requiring at most k observations on the first stage.*

The proof of Theorem 2 differs only in minor details from that of Theorem 1 and will be omitted.

For each $(n - 1)$-stage family $\mathfrak{F}_m$ of Theorem 2 and each $\gamma \varepsilon (0, 1)$ let $k(m, \gamma)$ be the minimal number of observations required on the first stage for the $(n - 1)$-scheme. Then we have

THEOREM 3. *Suppose in addition to the hypotheses of Theorem* 2, *each* $\mathfrak{F}_m$ *is a true* $(n - 1)$*-stage family and that* $\lim_{m\to\infty} k(m, \gamma) = \infty$ *for each* $\gamma \varepsilon (0, 1)$. *Then* $\mathfrak{F}$ *is a true n-stage family.*

PROOF. For if k is any first stage size and $\gamma \varepsilon (0, 1)$, there exists m such that $k(m, \gamma) > k$. Thus even if we know that $F \varepsilon \mathfrak{F}_m$, we would need an additional $n - 1$ stages for m sufficiently large.

Theorem 3 enables us to construct true n-stage families. We give one such example:

Let F_0 be the uniform distribution on [0, 1] and let $\mathfrak{R}$ be the family of uniform distributions on $[\theta, \theta + 1]$ with $\theta > 0$. Let $\{p_n, n = 1, 2, 3, \cdots\}$ be a sequence of numbers with $0 < p_n < p_{n+1} < 1$ and $\lim_n p_n = 1$. Define

$$\mathfrak{F}_m^{(2)} = \{F: F(x) = p_m F_0(x - m) + (1 - p_m) R(x - m), R \varepsilon \mathfrak{R}\}.$$

Let $\theta(F)$ be the mean of F. Clearly each $\mathfrak{F}_m^{(2)}$ is a true one-stage family using the sample mean as an estimator. Also since for fixed m, when θ is large a significant contribution to the mean is due to $R(x - m)$, we must make certain that one or more observations in the sample come from $R(x - m)$. To assure this it is necessary that the sample size increase unboundedly with m. Now let $\mathfrak{F}^{(2)} = \bigcup_m \mathfrak{F}_m^{(2)}$, and let $N_j = [j, \infty)$. Then it is easily verified that Theorem 3 applies and $\mathfrak{F}^{(2)}$ is a true two-stage family. Now define

$$\mathfrak{F}^{(n)} = \bigcup_m \mathfrak{F}_m^{(n)}$$

where

$$\mathfrak{F}_m^{(n)} = \{F: F(x) = p_m F_0(x - m) + (1 - p_m) G(x - m), G \varepsilon \mathfrak{F}^{(n-1)}\}$$

where $\mathfrak{F}^{(n-1)}$ has been inductively defined as a true $(n - 1)$-stage family. Repeating the above arguments we see that $\mathfrak{F}^{(n)}$ is a true n-stage family (requiring but one observation on the first stage).

4. Quantiles of unimodal populations. Let $\mathfrak{F}$ be the class of unimodal distributions—(i.e., P_F is unimodal if there exists a given point x such that for each two intervals I and I' to the right [left] of x having the same length, if I is to the right of [left of] I' then $P_F(I) \leqq P_F(I')$). In this section we obtain fixed length confidence intervals for quantiles in two stages.

Let $0 < p < 1$. Then the p-quantile $\theta_p(F)$ is unique for each $F \varepsilon \mathfrak{F}$, and it is known, (see Cramér [4]), that $\{\theta_{p,n}^*\}$, the sequence of sample p-quantiles based on $X_1, \cdots, X_n$, is consistent for $\theta_p(F)$. Let $I_1 = [\theta_p(F) - \delta, \theta_p(F)]$ and $I_2 = [\theta_p(F), \theta_p(F) + \delta]$. Now for each $\gamma \varepsilon (0, 1)$ one can easily determine a sequence $\{c_m\}$ of positive numbers with $\lim_{m\to\infty} c_m = 0$ such that

$$\mathfrak{F}_{m,\gamma,\delta} \subset \{F \varepsilon \mathfrak{F}: \min[P_F(I_1), P_F(I_2)] < c_m\}.$$

Choose k so that $1 - \max[p^k, (1 - p)^k] > (1 - \alpha)/(1 - \gamma)$ (so that with high probability at least one of the first k observations will be in $(-\infty, \theta_p(F)]$ and one in $[\theta_p(F), \infty))$ and for each j define $N_j = \{(x_1, \cdots, x_k): \max_{i=1,\cdots,k} |x_i| > j\}$. Clearly $\lim_{j\to\infty} P_{F_k}\{N_j\} = 0$ for each $F \varepsilon \mathfrak{F}$. To apply Theorem 1 we assume $F \varepsilon \mathfrak{F}_{m,\gamma,\delta}$ and consider the following four cases:

(i) $\theta_p(F) \leqq -j$. Then $P_F\{(-\infty, -j]\} \geqq p$,

(ii) $\theta_p(F) \geqq j$. Then $P_F\{[j, \infty)\} \geqq 1 - p$,

(iii) $-j < \theta_p(F) < j$ and $P_F(I_1) < c_m$. Then from the unimodality and the fact that $F \varepsilon \mathfrak{F}_{m,\gamma,\delta}$,

$$P_F\{(-j, \theta_p(F)]\} \leqq (2j/\delta)c_m \qquad [\theta_p(F) \leqq x]$$

and hence $P_F\{(-\infty, -j]\} \geqq p - (2j/\delta)c_m$ or

$$P_F\{[\theta_p(F), j)\} \leqq (2j/\delta)c_m \qquad [\theta_p(F) \geqq x]$$

and hence $P_F\{[j, \infty)\} \geqq 1 - p - (2j/\delta)c_m$,

(iv) $-j < \theta_p(F) < j$ and $P_F(I_2) < c_m$. Then similarly to (iii) either $P_F\{(-\infty, -j]\} \geqq p - (2j/\delta)c_m$ or $P_F\{[j, \infty)\} \geqq 1 - p - (2j/\delta)c_m$.

Since these four cases are exhaustive we conclude that for $F \varepsilon \mathfrak{F}_{m,\gamma,\delta}$

$$P_{F_k}\{N_j\} \geqq 1 - \max[(p + (2\delta/\delta)c_m)^k, (1 - p + (2\delta/\delta)c_m)^k\}.$$

Since $c_m \to 0$ we see that Theorem 1 applies. This result was first obtained by Weiss [10] using different techniques.

5. Translation-scale parameter families. Let F_0 be a known nondegenerate distribution function and define

$$\mathfrak{F} = \{F:F(x) = F_0((x - \mu)/\sigma), -\infty < \mu < \infty, \sigma > 0\}.$$

We shall show that Theorem 1 applies for fixed-precision estimation of μ and σ. To do this let $\theta_{p_1}^0$ and $\theta_{p_2}^0$ be two distinct unique quantiles of F_0 and let $\theta_{p_1}^0$ and $\theta_{p_2}^0$ be the corresponding quantiles of F. Then $\sigma = (\theta_{p_2} - \theta_{p_1})/(\theta_{p_2}^0 - \theta_{p_1}^0)$ and $\mu = \theta_{p_1} - \sigma\theta_{p_1}^0$. Hence it is clear that fixed-precision confidence intervals for μ and σ can be obtained from such intervals for θ_{p_1} and θ_{p_2}, (i.e., if for each $\delta > 0$ we can find a confidence interval of length 2δ and confidence at least $1 - \alpha$ for θ_{p_1} and θ_{p_2}, then we can also do so for μ and σ).

Let $\theta_{p_1,n}^*$ be the corresponding sample quantile based on $X_1, \cdots, X_n$. The sequence $\{\theta_{p_1,n}^*\}$ is consistent for θ_{p_i} (see e.g. Cramér, [4]). Then it is clear that for each $\gamma \varepsilon (0, 1)$ one can easily determine a sequence $\{\sigma_m\}$ (depending on p_i and F_0) with $\lim_{m\to\infty} \sigma_m = \infty$ such that

$$\mathfrak{F}_{m,\gamma,\delta} \{F \varepsilon \mathfrak{F}: \sigma > \sigma_m\}.$$

Now let p be the maximal discrete mass of F_0 and choose k such that $1 - p^k > (1 - \alpha)/(1 - \gamma)$. As $\sigma \to \infty$ we see that at least $1 - p$ of the mass F drifts to $\pm\infty$.

As before let $N_j = \{(x_1, \cdots, x_k): \max_{i=1,\cdots,k} |x_i| > j\}$ and we see that Theorem 1 applies once again.

6. Stationary Gaussian Markov processes. Such processes depend on three parameters μ, σ, α, with $-\infty < \mu < \infty$, $\sigma > 0$ and $\alpha \varepsilon (-1, 1)$, and may be represented in the form

$$X_n = \sigma(1 - \alpha^2)^{\frac{1}{2}} \sum_{i=0}^{\infty} \alpha^i Y_{n-1} + \mu, \qquad n = 0, \pm 1, \cdots$$

where the sequence $\{Y_n, n = 0, \pm 1, \cdots\}$ consists of independent normal random variables with zero means and unit variances. We shall show that we may estimate the parameters μ, σ, α in two stages. Rather than going into details we note that it is sufficient to obtain high probability upper bounds on σ and α^2. To obtain such bounds observe that X_1 is $N(\mu, \sigma^2)$ and hence (see Abbott and Rosenblatt, [1]) for given $\gamma_1 \varepsilon (0, 1)$ there is a function f_1 (positive at nonzero values of its argument) such that $P\{\sigma^2 < f_1(X_1)\} \geqq 1 - \gamma_1$. Similarly $X_1 - X_3$ is $N(0, 2\sigma^2[1 - \alpha^2])$ and hence for given $\gamma_2 \varepsilon (0, 1)$ there is a function f_2 (positive at nonzero values of its argument) such that

$$P\{f_2(X_1 - X_3) < 2\sigma^2(1 - \alpha^2)\} \geqq 1 - \gamma_2 .$$

Consequently

$$P\{\sigma^2 \leqq f_1(X_1), \alpha^2 \leqq 1 - f_2(X_1 - X_3)/2f_1(X_1)\} \geqq 1 - \gamma_1 - \gamma_2$$

which establishes the desired result, since with probability one $X_1 \neq 0$.

REFERENCES

[1] ABBOTT, J. H., and ROSENBLATT, JUDAH (1963). Two stage estimation with one observation on first stage. *Ann. Inst. Statist. Math.*

[2] BAHADUR, R. R., and SAVAGE, L. J. (1956). The nonexistence of certain statistical procedures in nonparametric problems. *Ann. Math. Statist.* **27** 1115–1122.

[3] BIRNBAUM, ALLAN, and HEALY, WILLIAM C. JR. (1960). Estimates with prescribed variance based on two-stage sampling. *Ann. Math. Statist.* **31** 662–676.

[4] CRAMÉR, H. (1951). *Mathematical Methods of Statistics*. Princeton Univ. Press.

[5] DANTZIG, G. B. (1940). On the nonexistence of tests of "Student's" hypothesis having power functions independent of σ. *Ann. Math. Statist.* **11** 186–192.

[6] FARRELL, ROGER. Sequentially determined bounded length confidence intervals. Unpublished doctoral thesis, Univ. of Illinois.

[7] GRAYBILL, R. (1958). Determining sample size for a specified width confidence interval. *Ann. Math. Statist.* **29** 282–287.

[8] MATTHES, T. On two-stage sampling procedures. Unpublished doctoral thesis, Columbia Univ.

[9] STEIN, CHARLES (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* **16** 243–258.

[10] WEISS, LIONEL (1960). Confidence intervals of preassigned length for quantiles of unimodal populations. *Naval Research Logistics Quarterly*, **7** 251–256.