SEQUENTIAL METHODS IN STATISTICS
BANACH CENTER PUBLICATIONS, VOLUME 16
PWN-POLISH SCIENTIFIC PUBLISHERS
WARSAW 1985

# A SURVEY OF SEQUENTIAL STATISTICAL ANALYSIS(¹)

Z. GOVINDARAJULU

*Department of Statistics, University of Kentucky,*
*Lexington, KY, U.S.A.*

In this survey we provide the highlights of sequential analysis, tracing the new developments in testing hypotheses and point and interval estimation.

## 1. Introduction

**1.1. Brief historical background.** Sequential statistics grew out of necessity during World War II and the sequential methods were simultaneously developed in England by George Barnard and in the United States by the Princeton Research Group. Darling (1976) and Wallis (1980) trace the origins of sequential analysis, the latter providing a well-documented description of its origin. From Wallis' account it was Captain G.L. Schyler who suggested to Wallis in 1943 that one should be able to achieve some *economy in sampling by applying the single-sample test sequentially*. W. A. Wallis and M. Friedman had several discussions about this sequential setup, both with each other and with J. Wolfowitz and E. Paulson. H. Hotelling pointed out the close analogy between the problem of stopping in sequential sampling and the heat flow in the presence of an absorbing barrier and contributed the term "sequential" to describe the method of analysis. After realizing that the sequential method might require sophisticated mathematical tools, Wallis and Friedman brought the problem to the attention of A. Wald who put the finishing touches to the sequential test known to us as the *sequential probability ratio test* (SPRT). W. E. Deming remarked that sequential analysis is similar to the work of Lord Rayleigh on the problem of random walk in physics.

**1.2. Sampling inspection.** The earliest sequential procedure is the double sampling plan proposed by Dodge and Romig (1929) for sampling

(¹) Formed the basis of four lectures delivered at the Banach International Mathematical Center (Warsaw, Poland) during the semester on Sequential Methods in Statistics (September 7–December 11, 1981).

inspection. The rule is given as follows: Sample one item at a time, reject the lot as soon as the number of defectives in the sample is $\geqslant c$, and accept the lot as soon as the number of effectives in the sample is $\geqslant n-c+1$, where $c$ is a prescribed positive integer. The required sample size is random, and is at least $c$ (if all the items are defective) and is at most $n$. The sampling plan is called *curtailed inspection.*

**1.3. Stein's two-stage procedure.** Let $X$ be normally distributed having unknown mean $\theta$ and unknown variance $\sigma^2$. Testing $H_0$: $\theta = \theta_0$ against the alternative $H_1$: $\theta > \theta_0$ is known as the Student's hypothesis-testing problem. It is of interest to devise a test of Student's hypothesis whose power function does not depend on $\sigma$. However, Dantzig (1940) has shown the nonexistence of meaningful fixed-sample test procedures for this problem. Stein (1945) proposed a two-sample test with the above property, the size of the second sample depending on the result of the first.

Moshman (1958) investigated the proper choice of the initial sample size $m$ in Stein's two-stage procedure. However, the optimum choice of $m$ that maximizes a given function involves an arbitrary parameter which needs to be specified by the experimenter from nonstatistical considerations. Jurečkova (1981) also investigated the optimum choice of $m$. The study of Blumenthal and Govindarajulu (1977) indicates that Stein's two-stage procedure is quite robust against mixtures of normal populations differing in location parameters.

**1.4. Wald's sequential probability ratio test** (SPRT). Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed (i.i.d.) random variables having probability density $f(x)$. We wish to test $H_0$: $f = f_0$ versus $H_1$: $f = f_1$. Let

$$\Lambda_n = \prod_{i=1}^{n} [f_1(X_i)/f_0(X_i)].$$

Then Wald's SPRT can be described as follows: Choose two constants $A$ and $B$ such that $0 < B < A < \infty$. If experimentation has proceeded to stage $n$, accept $H_0$ if $\Lambda_n \leqslant B$; reject $H_0$ if $\Lambda_n \geqslant A$; continue sampling if $B < \Lambda_n < A$ ($n = 1, 2, \ldots$). The following theorem will ensure the finite sure termination of the SPRT.

THEOREM 1.4.1 (Wald (1947), Stein (1946)). *Let* $Z = \ln\{f_1(X)/f_0(X)\}$. *Then Wald's SPRT terminates finitely with probability one provided* $P(Z = 0) < 1$.

Now, we shall explore whether it is possible to express the bounds $A$ and $B$ explicitly in terms of $\alpha$ and $\beta$ (the type I and type II error probabilities) where

$$\alpha = \sum_{i=1}^{\infty} P_0(B < \Lambda_j < A, j = 1, \ldots, i-1 \text{ and } \Lambda_i \geqslant A) \tag{1.4.1}$$

and

$$\beta = \sum_{i=1}^{\infty} P_1(B < \Lambda_j < A, j = 1, \ldots, i-1 \text{ and } \Lambda_i \leqslant B) \tag{1.4.2}$$

However, these expressions are not easy to calculate and hence, in general, there is no hope of solving for $A$ and $B$ in terms of $\alpha$ and $\beta$. However, Wald (1947) has obtained very simple bounds for $A$ and $B$ in terms of the error probabilities.

THEOREM 1.4.2. *For Wald's* SPRT

$$A \leqslant (1-\beta)/\alpha \qquad \textit{and} \qquad B \geqslant \beta/(1-\alpha). \tag{1.4.3}$$

In obtaining the above bounds for $A$ and $B$, it is assumed that the SPRT terminates finitely with probability one. However, according to Hall, Wijsman and Ghosh (1965, p. 586) it can be shown that the inequalities given by (1.4.3) constitute approximate bounds irrespective of whether termination is certain or not, provided the error probabilities are small. Also $B \doteq \beta$ and $A \doteq 1/\alpha$ can serve as reasonable bounds.

The inequalities in Theorem 1.4.2 are almost equalities since $\Lambda_n$ does not usually obtain either a value far above $A$ or value far below $B$. So, suppose we take $A \doteq A' \equiv (1-\beta)/\alpha$ and $B \doteq B' \equiv \beta/(1-\alpha)$. When approximations $A'$ and $B'$ are used in the place of $A$ and $B$ we may not have the same error probabilities $\alpha$ and $\beta$. Let the effective error probabilities be $\alpha'$ and $\beta'$. Then one can easily show that

$$\alpha/(1-\beta) \geqslant \alpha'/(1-\beta') \qquad \text{and} \qquad \beta/(1-\alpha) \geqslant \beta'/(1-\alpha'), \tag{1.4.4}$$

from which it follows that

$$\alpha+\beta \geqslant \alpha'+\beta'. \tag{1.4.5}$$

That is, at most one of the error probabilities could be larger than the nominal error probability. Further, one can show that when $\alpha$ and $\beta$ are small, any increase in $\alpha\,[\beta]$ is not beyond a factor of $1+\beta\,[1+\alpha]$. Both $\alpha' < \alpha$ and $\beta' < \beta$, imply that the test $(B', A')$ requires substantially more observations than the test associated with $(B, A)$. Since $B \geqslant \beta/(1-\alpha)$ and $A \leqslant (1-\beta)/\alpha$, we have increased the continue sampling region. Hence, there will be an increase in the number of observations required to come to a terminal decision. However, Wald (1947, pp. 46-48) points out that the increase in the sample size caused by the approximation will be only slight.

A nice feature of the SPRT is that the approximations to $A$ and $B$ are only functions of $\alpha$ and $\beta$ and can be computed once and for all free of the underlying density. Also note that when $B = \beta/(1-\alpha)$ and $A = (1-\beta)/\alpha$ it is trivial to show that $B < 1 < A$ (since $\alpha+\beta$ cannot exceed unity).

**1.5. The operating characteristic function (OC function).** Wald (1947) devised an ingenious method of obtaining the operating characteristic (probability of accepting $H_0$) of an SPRT. Consider an SPRT of $H_0$: $f = f(x;\theta_0)$ versus $H_1$: $f = f(x;\theta_1)$. If $\theta_0$ and $\theta_1$ are the only two possible states of nature, then there is no point in considering the OC function. However, if the above hypothesis-testing problem is a simplification of, for example, $H_0$: $\theta \leqslant \theta^*$ versus $H_1$: $\theta > \theta^*$, then one would be interested in the OC function for all possible values of $\theta$.

Let $\theta$ be fixed and determine as function of that $\theta$ a value of $h$ (other than zero) for which

$$E_\theta[\{f(X;\theta_1)/f(X;\theta_0)\}^h] = 1. \tag{1.5.1}$$

This expectation is 1 when $h = 0$ but there is one other value of $h$ for which it is also 1 (see Govindarajulu (1981, pp. 604–605) or Wald (1947, pp. 158–159)). Note that $h = 1$ when $\theta = \theta_0$ and $h = -1$ if $\theta = \theta_1$. Let

$$f^*(x;\theta) = [f(x;\theta_1)/f(x;\theta_0)]^h f(x;\theta).$$

Considering the auxiliary problem of testing

$$H\colon f = f(x;\theta) \qquad \text{versus} \qquad H^*\colon f = f^*(x;\theta),$$

one can obtain that

$$P_\theta(\text{accept } H_0) = P_\theta(\text{accept } H) = P_H(\text{accept } H) = 1-\alpha^*$$

where $\alpha^*$ denotes the type I error probability for the auxiliary problem. However, solving the approximate equations

$$B^h = \beta^*/(1-\alpha^*), \qquad A^h = (1-\beta^*)/\alpha^*$$

we find that $\alpha^* \doteq (1-B^h)/(A^h-B^h)$, and hence

$$\mathrm{OC}(\theta) \doteq (A^h-1)/(A^h-B^h) \tag{1.5.2}$$

Since $h$ is a function of $\theta$, $\mathrm{OC}(\theta)$ is an implicit function of $\theta$. Each value of $h$ determines a $\theta$ and a value of $P_\theta(\text{accept } H_0)$, a point on the OC curve. (For exponential models, one obtains an explicit expression for $\theta$ in terms of $h$.) Eq. (1.5.2) does not provide a well-defined value of $\theta$ when $h = 0$ since the relation is satisfied by all $\theta$. However, using l'Hospital's rule,

$$\lim_{h\to 0} \mathrm{OC}(\theta) \doteq \ln A/(\ln A - \ln B).$$

Since, $\mathrm{OC}(\theta_0) = 1-\alpha$, $\mathrm{OC}(\theta_1) = \beta$, $\lim_{h\to\infty} \mathrm{OC}(\theta) = 1$ and $\lim_{h\to-\infty} \mathrm{OC}(\theta) = 0$, five points on the OC curve are known. The corresponding $\theta$-values can be determined from the specific problem.

**1.6. Average sample number (ASN).** The sample size needed to reach a decision in a sequential plan is a random variable which is denoted by $N$. Typically, the distribution of $N$ is too complicated to be determined. However, we have the following result pertaining to the moment-generating function of $N$.

THEOREM 1.6.1. *Let* $Z = \ln\{f_1(X)/f_0(X)\}$. *If* $P(Z=0) < 1$ *then*

$$M_N(t) = E\{\exp(tN)\} < \infty \qquad \text{for all} \quad t \in [-\theta, \infty]$$

*for some* $\theta \geqslant 0$.

Even though all the moments of $N$ are finite it is just not easy to evaluate $E(N^i)$. However, since $\ln \Lambda_N$ is a sum of a random number of i.i.d. random variables one can show (for instance, see Wolfowitz (1947b) and Johnson (1959)) that

$$E(\ln \Lambda_N) = E(N)E(Z), \qquad Z = \ln\{f(X;\theta_1)/f(X;\theta_0)\} \tag{1.6.1}$$

because the event $(N \geqslant i)$ is independent of $Z_i$, $Z_{i+1}, \ldots$ If a decision is reached at the $n$th stage, $\Lambda_n$ is approximately distributed as a Bernoulli variable with values $B$ and $A$ and

$$E(\ln \Lambda_N) \doteq (\ln B)\, P(\text{accept } H_0) + (\ln A)\, P(\text{reject } H_0). \tag{1.6.2}$$

In particular, from (1.6.1) and (1.6.2) we have

$$E_{\theta_0}(N) \doteq [(\ln B)(1-\alpha) + (\ln A)\alpha]/E_{\theta_0}(Z) \tag{1.6.3}$$

and

$$E_{\theta_1}(N) = [(\ln B)\beta + (\ln A)(1-\beta)]/E_{\theta_1}(Z). \tag{1.6.4}$$

Furthermore, if $EZ = 0$, then one can show that

$$EN = E[(\ln \Lambda_N)^2]/E(Z^2). \tag{1.6.5}$$

The identities in Equations (1.6.1) and (1.6.5) are usually referred to as Wald's first and second equations.

**1.7. Wald's fundamental identity.** Wald's (1947) identity plays a fundamental role in derving the moments of the sample size required to terminate the SPRT.

THEOREM 1.7.1 (Wald (1947)). *Let* $Z = \ln\{f(X;\theta_1)/f(X;\theta_0)\}$, $S_N = \sum_1^N Z_i$ and $C(t) = E(\exp tZ)$. *Assume that* $P(Z=0) < 1$. *Then we have*

$$E\{\exp(S_N t)[C(t)]^{-N}\} = 1 \tag{1.7.1}$$

*for every $t$ in $D$ where $D$ denotes the set of points in the complex plane such that $C(t)$ is finite and $|C(t)| \geqslant 1$.*

Differentiating (1.7.1) w.r.t. $t$ and setting $t = 0$, we obtain Wald's first equation (see (1.6.1)). If $EZ = C'(0) = 0$, then differentiating (1.7.1) twice and setting $t = 0$ we obtain Wald's second equation (1.6.5). Bahadur (1958) points out that Wald's identity can be regarded as a special case of a formula for the probability that sampling terminates finitely. Blom (1949) has extended Wald's identity to the case of independent summands and remarks that in the case of i.i.d. summands, the identity can be differentiated any number of times for any real $t$ provided $P(Z = 0) < 1$.

**1.8. Further results on OC and ASN functions.** Wald (1947) considered a Bernoulli class of random variables for which one can obtain exact expressions for the OC and ASN functions. Other families of distributions of the $Z_i$ for which $P(\Lambda_N \geqslant A)$ and $EN$ are capable of being expressed by exact formulas have been described by Kemperman (1961, pp. 70–71) and Ferguson (1967, p. 378). Let

$$\ln\{f(x;\theta_1)/f(x;\theta_0)\} = u(\theta_0, \theta_1)x + v(\theta_0, \theta)$$

for all $\theta_0$, $\theta_1$ and $u/v$ be a rational number. Then Eger (1980) reduces the exact computation of the OC function and the moments of ASN of the SPRT for $\theta_0$ versus $\theta_1$ to that of solving a system of simultaneous linear equations which differ only in their right hand side quantities.

M. N. Ghosh (1960) gave a bound on the error in the approximation to $E(N)$ given by (1.6.2) which is wider than the bound of Wald (1947, Eqs. (A.77) and (A.78)); especially, when the population is normal. He also constructed an example showing that the error in Wald's approximation to $E(N)$ is not small even though the mean and variance of $Z$ are small, so that no general statement about the validity of the approximation can be made. Hoeffding (1953, 1960) derived lower bounds for the ASN required by an arbitrary sequential test for which $\alpha+\beta < 1$. Chanda (1971) has derived results pertaining to the convergence in probability and asymptotic normality of $N$ when suitably standarized and when the underlying population of the observations belong to the exponential family.

Page (1954) suggested an improvement to Wald's approximations to the OC and ASN functions and Kemp (1958) provided even better approximations. Tallis and Vagholkar (1965) also obtain improvements to the OC and ASN approximations which are comparable to those of Page (1954) and Kemp (1958). For further details see Govindarajulu (1981, Section 2.7).

**1.9. Truncated SPRT.** Although SPRT's enjoy the property of eventually terminating with probability one, often due to limitations of cost or available number of experimental units we need to set a definite upper limit say $n_0$ on the number of observations. This may be achieved by truncating the SPRT. Thus Wald's (1947) truncated SPRT behaves like an ordinary

SPRT for $n \leqslant n_0$. If the SPRT does not lead to a terminal decision for $n \leqslant n_0$,

$$\text{reject } H_0 \text{ if } \quad 0 < \sum_{i=1}^{n_0} Z_i < \ln A,$$

and

$$\text{accept } H_0 \text{ if } \quad \ln B < \sum_{i=1}^{n_0} Z_i < 0$$

where $Z_i = \ln\{f_1(X_i)/f_0(X_i)\}$, $i = 1, 2, \ldots, H_j$: $f = f_j$, $j = 0, 1$.

When the SPRT is truncated, the error probabilities are changed from their nominal values. Govindarajulu (1981, Theorem 2.8.1.) provides sharper bounds for the true error probabilities than Wald's (1947).

Aroian (1968) proposes a direct method for evaluating the OC and ASN functions of any truncated sequential test procedure, once the acceptance, rejection and continuation regions are specified at each stage. His method involves repeated convolutions and numerical integration.

**1.10. Optimal properties of the** SPRT. The Optimal Property (OP) of the SPRT for testing a simple $H_0$ against a simple alternative $H_1$ was first proved by Wald and Wolfowitz (1948) (see Wolfowitz (1966) for additional details). Another proof which is due to LeCam appears in Lehmann (1959). Matthes (1963) has given a proof which relies on a mapping theorem.

THEOREM 1.10.1. *Among all tests (fixed-sample or sequential) for which* $P(\text{reject } H_0 \mid \theta_0) \leqslant \alpha$, $P(\text{accept } H_0 \mid \theta_1) \leqslant \beta$ *and for which* $E(N \mid \theta_i) < \infty$ $(i = 0, 1)$, the SPRT *with error probabilities* $\alpha$ *and* $\beta$ *minimizes both* $E(N \mid \theta_0)$ and $E(N \mid \theta_1)$.

It should be noted that the OP has been established under two conditons: (i) $B \leqslant 1 \leqslant A$, and (ii) $E(N \mid H_i) < \infty$ for $i = 0, 1$. Burkholder and Wijsman (1963) have shown that (ii) is not necessary whereas (i) is necessary. If (i) does not hold, then the SPRT is not admissible in some sense. However, if only tests which take at least one observation are considered, then there are no restrictions on $A$ and $B$ (other than $B \leqslant A$) for an SPRT to have OP.

Wijsman (1960) introduced monotonicity property (Property M) of an SPRT and showed that it implies uniqueness and the restricted optimum property (ROP), (namely that every SPRT has the optimum property among all SPRTs).

DEFINITION 1.10.1. An SPRT is said to *have the property M* if at least one of the error probabilities decreases, when the upper stopping bound of the SPRT is increased and the lower stopping bound is decreased, unless the new test is equivalent to the old one, in which case the error probabilities remain the same.

THEOREM 1.10.2 (Wijsman (1960)). *The restricted optimum property* (ROP) *implies and is implied by the monotonicity property.*

Simons (1976) shows that Wald's SPRT is optimal in a larger class of tests having error probabilities $(\alpha', \beta')$ such that

$$\alpha'/(1-\beta') \leqslant \alpha/(1-\beta) \quad \text{and} \quad \beta'/(1-\alpha') \leqslant \beta/(1-\alpha), \quad \alpha' \geqslant 0 \text{ and } \beta' \geqslant 0.$$

Krylov and Miroshnichenko (1980) showed that Wald's SPRT is optimal in the class of tests having error probabilities $(\alpha', \beta')$ for which $\alpha' \geqslant 0$, $\beta' \geqslant 0$, $W(\alpha', \beta') \geqslant W(\alpha, \beta)$ and $W(\beta', \alpha') \geqslant W(\beta, \alpha)$ where $W(x, y) = (1-x)\ln\{(1-x)/y\} + x\ln\{x/(1-y)\}$.

The uniqueness of an SPRT has been considered by Weiss (1956), Anderson and Friedman (1960) and Wijsman (1960). Wijsman (1963, Theorem 2) has considered the existence of a SPRT which in general, need not be unique. Regarding the monotonicity of its power function, we have

THEOREM 1.10.4. (Lehmann (1959)). *Let $X_1, X_2, \ldots,$ be i.i.d. random variables having probability density $f(x; \theta)$ which has monotone likelihood ratio in $T(x)$. Then, any* SPRT *for testing $H_0$: $\theta = \theta_0$ versus $H_1$: $\theta = \theta_1$ $(\theta_0 < \theta_1)$ has a nondecreasing power function.*

**1.11. The generalized** SPRT (GSPRT). Weiss (1953) introduced the generalized SPRTs which differ from Wald's SPRT's in the sense that at $i$th stage, limits $A_i$, $B_i$ are used where these numbers are predetermined constants $(i = 1, 2, \ldots)$.

Let $f_i(x)$ denote the probability (density) function of $X$ under $H_i (i = 0, 1)$. Assume that $f_0(x)$ and $f_1(x)$ are bounded everywhere and have at most a finite number of discontinuities. Let $X_1, X_2, \ldots$ denote an i.i.d. sequence. Assume that for every $n$ and any finite nonzero $c$, $P_i(c \leqslant \Lambda_n \leqslant c+\Delta c) \to 0$ as $\Delta c \to 0$ for $i = 0, 1$.

Let $(\bar{A}_1, \bar{A}_2, \ldots)$ and $(\bar{B}_1, \bar{B}_2, \ldots)$ be two sequences of predetermined constants such that $\bar{A}_i > \bar{B}_i$ for all $i$. The value $\infty$ is not excluded. Sampling is continued as long as

$$\bar{B}_n < \prod_{j=1}^{n} \{f_i(X_j)/f_0(X_j)\} < \bar{A}_n. \tag{1.11.1}$$

The first time this is violated, we accept $H_1$ if the upper bound is violated and accept $H_0$ if the lower bound is violated. If $\bar{A}_m = \bar{B}_m$ while for all $i < m$ we have $\bar{A}_i > \bar{B}_i$, the test is truncated at the $m$th step.

Weiss (1953) shows the existence of a GSPRT or a sequence of GSPRT's which is better than a given test procedure for a specified hypothesis-testing problem. Kiefer and Weiss (1957) have studied certain properties of GSPRT's. They show that under certain regularity assumptions, the distribution of the sample size under the hypotheses uniquely determine a GSPRT. Many GSPRT's are shown to be inadmissible (admissibility being defined in terms of the error probabilities and the distribution of the sample size required to come to a decision). They also provide some fine characterizations of the GSPRT's.

Aivazian (1965) obtained some generalized sequential tests (GST's) which fall into the category of GSPRT's based on independent sequences of observations for distinguishing between densities that are indexed by a vector-valued parameter. For a special case of i.i.d. sequence with a one-dimensional parameter see Govindarajulu (1981, pp. 103–109).

Even though the SPRT enjoys the optimum property, in general, its expected sample size is large for values of the parameter lying between the values specified by the null and alternative hypotheses (that is, in indifference zones, a large number of observations is expected). Armitage (1957) proposed certain restricted SPRT's, leading to closed boundaries in testing hypotheses about the mean of a normal population. Donnelly (1957) proposed straight line boundaries that meet, converted to Wiener process, and obtained certain results. Anderson (1960) considered a modified SPRT in testing the mean of a normal population with known variance and derived approximation to the OC and ASN functions. Anderson's procedure is similar to Armitage's (1957) and Donnelly's (1957) procedures and we shall describe it for the special case of equal error probabilities.

Let $X$ be distributed normally having unknown mean $\theta$ and known variance $\sigma^2$. We wish to test $H_0$: $\theta = \theta_0$ versus $H_1$: $\theta = \theta_1$ ($\theta_1 > \theta_0$) with a sequential procedure which minimizes $E_\theta(N)$ at $\theta = (\theta_0+\theta_1)/2$ or (alternatively) minimizes the maximum of $E_\theta(N)$. Replacing the observation $X$ by the transformed observation $[X-(\theta_0+\theta_1)/2]/\sigma$ and calling $\theta^* = (\theta_1-\theta_0)/2$, the hypotheses become $H_0$: $\theta = -\theta^*$ and $H_1$: $\theta = \theta^*(\theta^* > 0)$ when sampling from the normal population having mean $\theta$ and variance 1. Then the procedure is: Let $c$, $d$ be positive constants. Take (transformed) observations $X_1, X_2, \ldots$ sequentially. At $n$th stage

$$\text{reject } H_0 \text{ if } \sum_1^n X_i \geqslant c - dn,$$

$$\text{accept } H_0 \text{ if } \sum_1^n X_i \leqslant c + dn,$$

and take one more observation if neither of the above occurs. If $n_0$ observations are drawn, stop sampling and reject $H_0$ (accept $H_0$) if $\sum_1^{n_0} X_i \geqslant 0$ ($< 0$).

Anderson (1960) using Wiener's process approximation, derived expressions for the probability of rejecting $H_0$ as a function of $\theta$, and the expected length of time as infinite series of terms involving Mill's ratio. Lawing and David (1966) and Fabian (1974) derived a simple and explicit expression for the probability of accepting $H_0$ using Wiener process as an approximation.

Savage and Savage (1965) consider finite sure termination of GSPRT's based on dependent and non-identically distributed random variables. Eisenberg, Ghosh and Simons (1976) consider GSPRT's based on non-necessarily i.i.d. sequence of observations and study their properties.

**1.12. Asymptotic properties of** SPRT's. In this section we briefly mention the asymptotic behavior of the error rates and ASN function of Wald's SPRT. Let $X, X_1, \ldots$ be i.i.d. with common density $f_i(x)$ under $H_i$, $i = 0, 1$. Wald's SPRT of $H_0$ versus $H_1$ has stopping time

$$N = \inf\{n\colon\ S_n \notin (b, a)\} \tag{1.12.1}$$

where $S_n = \sum_1^n Z_j$, $Z_j = \ln [f_1(X_j)/f_0(X_j)]$ and $b, a$ are two specified constants. We also assume that $Z$ is finite with probability one (wpl). Let

$$\alpha = P_0(S_N \geqslant a) \qquad \text{and} \qquad \beta = P_1(S_N \leqslant b). \tag{1.12.2}$$

We, throughout this section, assume that $P(Z = 0) < 1$. Let $c = \min(-b, a)$ and write $\lim_c$ for $\lim_{c\to\infty}$. Then we have the following theorem.

THEOREM 1.12.1 (Berk (1973) and Govindarajulu (1968)). *Let $X$, $X_1, \ldots$ be i.i.d and $\mu = EZ$ exists. Then if $\mu > 0$* wpl

(i) $\lim_c I_{(S_N \geqslant a)} = \lim_c P(S_N \geqslant a) = 1$ *and*

(ii) $\lim_c N/a = \lim_c EN/a = 1/\mu$.

*If $\mu < 0$,* wpl

(iii) $\lim_c I_{(S_N \leqslant b)} = \lim_c P(S_N \leqslant b) = 1$, *and*

(iv) $\lim_c \{N/b\} = \lim_c \{EN/b\} = 1/\mu$.

If follows from Theorem 1.12.1 that Wald's approximation to the ASN, namely

$$EN \doteq [bP(S_N \leqslant b) + aP(S_N \geqslant a)]/\mu \tag{1.12.3}$$

is asymptotically exact. If the $X$'s are i.i.d. and $E_i|Z| < \infty$ for $i = 0, 1$, then Berk (1973) obtains

$$\lim_c a^{-1} \ln \alpha^{-1} = 1 = \lim_c (-b)^{-1} \ln(1/\beta). \tag{1.12.4}$$

On the other hand if $X$'s are i.i.d., $EZ = 0$ and $\sigma^2 = EZ^2$, then under certain regularity conditions on the variable $Z$, Berk (1973) shows that

$$\lim \sigma^2 EN/(-ab) = 1. \tag{1.12.5}$$

For futher details see Govindarajulu (1981, Section 2.13).

## 2. Sequential tests for composite hypotheses

In this section we have considered SPRT's for testing a simple hypothesis versus a simple alternative. However, in practical situations the simple null or alternative hypothesis is only a representative of a set of simple hypotheses. The compositeness of a hypothesis can arise from two situations: (i) the composite hypotheses are concerned about the parameters of interest, and (ii) the hypothesis may be simple or composite; however, there are nuisance parameters.

Wald (1974) proposed the method of weight functions to test sequentially composite hypotheses. Let $f(x;\theta)$ denote the probability (or density) function of $X$, indexed by the same unknown parameter $\theta$ (which may be vector-valued). We wish to test $H_0$: $\theta \in \omega_0$ versus $H_1$: $\theta \in \omega_1$ where the $\omega_0$ and $\omega_1$ have at least one element in them.

*Wald's method of weight functions.* Assume that there exist two functions $g_0(\theta)$ and $g_1(\theta)$ such that

$$\int_{\omega_0} g_0(\theta)\,d\theta = 1, \qquad \int_S g_1(\theta)\,dS = 1 \tag{2.1.1}$$

where $S$ denotes the boundary of $\omega_1$. Then the SPRT is based on the ratio

$$f_{1,n}/f_{0,n} = \int_S g_1(\theta) \prod_{i=1}^{n} f(x_i;\theta)\,dS \Big/ \int_{\omega_0} g_0(\theta) \prod_{i=1}^{n} f(x_i;\theta)\,d\theta \tag{2.1.2}$$

satisfying the conditions:

(i) the type I error probability, $\alpha(\theta)$ is constant on $\omega_0$;

(ii) the type II error probability $\beta(\theta)$ is constant over $S$;

(iii) for any point $\theta$ in the interior of $\omega_1$ the value of $\beta(\theta)$ does not exceed its constant value on $S$.

Wald (1947, Section A.9) claims that the weight functions $g_i(\theta)$ ($i = 0, 1$) are optimal in certain sense. However, Brown and Cohen (1981) show that tests based on weight functions $g_i$ such that $\theta_0$ (the boundary point between the two hypotheses) belongs to the support of $g_i (i = 0, 1)$ are inadmissible because any test whose continuation region has unbounded width is inadmissible (see Brown and Cohen (1981, Theorem 3.1, p. 1246). See also Berk, Brown and Cohen (1981). Using the method of weight functions, sequential binomial, chi-square, $t$, $t^2$ and $F$ tests have been derived and their properties investigated. Let us illustrate the method for the sequential $t$ and $t^2$ case.

Let $X$ be normally distributed having mean $\theta$ and unknown variance $\sigma^2$. We wish to test $H_0$: $\theta = \theta_0$ versus $H_1$: $|\theta - \theta_0| \geqslant \delta\sigma$ where $\delta > 0$. The above hypothesis-testing can arise from several practical situations.

Define the weight functions $g_0(\theta, \sigma) = 1/c$ if $0 \leqslant \sigma \leqslant c$, $\theta = \theta_0$ (and zero elsewhere) and $g_1(\theta, \sigma) = c/2$ if $0 \leqslant \sigma \leqslant c$ and $\theta = \theta_0 \pm \delta\sigma$ (and zero elsewhere), one can easily obtain

$$\psi(T; \delta, n) = \lim_{c \to \infty} \{f_{1,n}/f_{0,n}\} = \frac{\exp(-n\delta^2/2)}{\Gamma\{(n-1)/2\}} \int_0^\infty v^{-(n-3)/2} e^{-v} \cosh\{(n\delta T(2v)^{1/2}\}\, dv$$

where

$$T = (\bar{X}_n - \theta_0)/S^{1/2}, \qquad S = \sum_1^n (X_i - \theta_0)^2.$$

Since $\psi$ is a function of $|T|$, one can easily establish that (i) $\alpha(\theta, c)$ is constant on $\omega_0$ and (ii) $\beta(\theta, \sigma)$ is a function of $|\theta - \theta_0|/\sigma$. Analogously for the sequential $t$-test by taking $g_1(\theta, \sigma) = 1/c$ for $0 \leqslant \sigma \leqslant c$ and $\theta = \theta_0 + \delta\sigma$ (and zero elsewhere) we obtain the limit of the modified likelihood ratio to be

$$\psi_1(T; \delta, n) = \frac{\exp(-n\delta^2/2)}{\Gamma[(n-1)/2]} \int_0^\infty v^{(n-3)/2} \exp[-v + n\delta T(2v)^{1/2}]\, dv.$$

Thus, the sequential procedures can be based on $t_n$ where $t_n = \sqrt{n}(\bar{X} - \theta_0)/s_n$, $\bar{X}$ denoting the sample mean and $s_n^2$ the sample variance that are based on $n$ obsevations. Thus the sequential $t$(or $t^2$) test can be described as follows: If the experiment has proceeded until $n$th stage, the sampling continuation region is given by

$$B_n < t_n < A_n \quad (B'_n < |t_n| < A'_n), \qquad n = 2, 3, \ldots$$

where the bounds $B_n$ and $A_n$ ($B'_n$ and $A'_n$) are obtained by inverting the inequality: $B < \psi_1(T; \delta, n) < A\,(B < \psi(T; \delta, n) < A)$ in terms of $t_n(|t_n|)$. Rushton (1950) obtained an asymptotic expansion for $\psi_1(T)$. One can also obtain an asymptotic expression for $\psi(T)$. One can easily establish that $t$ and $t^2$ tests terminate finitely with probability one. Also one can show the power function of the sequential $t$ test is a nondecreasing function of $(\theta - \theta_0)/\sigma$. This monotonicity property may not be enjoyed by the $t^2$-test. J. K. Ghosh (1960) has investigated the double minimax property of $t$ and $t^2$. For further details see Govindarajulu (1981, Sections 3.1, 3.2).

In the context of sequential $t$, it is worthwhile to mention that Hall

(1962) proposed two sequential analogues of Stein's two-stage test procedure for testing hypotheses about the mean of a normal population with specified bounds on error probabilities when the variance is unknown. These constitute alternatives to the sequential $t$-test. Moreover, unlike the $t$-test, these procedures do not require that the alternative hypothesis be in standard deviation units.

**2.2. Invariant sequential procedures.** Where there are nuisance parameters, it is not always possible to construct weight functions so as to satisfy the specified conditions. One way to eliminate the nuisance parameter is to consider invariant procedures. First we need the following definitions.

DEFINITION 2.2.1. A function $T(x)$ on $\mathscr{X}$ is said to be a *maximal invariant* (MI) with respect to (w.r.t.) $G$ if it is invariant w.r.t. $G$ and if it takes on distinct values on distinct orbits. In other words, if $T[g(x)] = T(x)$ for all $x \in \mathscr{X}$ and $g \in G$ and $T(x_1) = T(x_2)$ implies that $x_2 = gx_1$ for some $g \in G$.

Let $\bar{G}$ denote the one-to-one group of transformations induced by $G$ on the parameter space $\Theta$. Further, $\bar{G}$ induces a partition of $\Theta$ into equivalent classes or orbits. Let $\lambda(\theta)$ denote the maximal invariant on $\Theta$ w.r.t. $\bar{G}$.

The principle of invariance gives a reduction of the data and one can test a hypothesis about $\lambda(\theta)$, the maximal invariant on $\Theta$, via the maximal invariant $U(x)$ on $\mathscr{X}$. One can further reduce the data by use of an invariantly sufficient statistic. Assume that the sufficient statistic $S$ induces a group $G_S$ of transformations on the sample space $\mathscr{S}$ of $S$.

DEFINITION 2.2.2. A function $V$ on $\mathscr{X}$ is *invariantly sufficient* for the probabilities model $Y = (\mathscr{X}, \mathscr{A}, P_\theta)$ under $G$ if

(i) $V$ is invariant under $G$,

(ii) the conditional probability of any invariant set $E$ given $V = v$ is free of $\theta$, for all $\theta \in \Theta$.

Notice that (i) implies that $V$ is a function of the maximal invariant $U$ on $\mathscr{X}$ under $G$ and (ii) implies (provided the densities exist) that

$$f_U(u; \lambda) = f_V(v; \lambda)\, h(u) \tag{2.2.1}$$

where $f_U(u; \lambda)(f_V(v; \lambda))$ denotes the probability density or function of $U$ $(V)$. Loosely speaking (2.2.1) means that $V$ contains all the information about $\lambda$ that is available in any invariant function.

Now, we are ready to state Stein's theorem.

THEOREM 2.2.1 (Stein's Theorem). *Under certain assumptions, if $S$ is sufficient for the probability model $(\mathscr{X}, \mathscr{A}, P_\theta)$ and $U_S$ is a maximal invariant on $\mathscr{S}$ under $G_S$, then $V(x) = U\ (S(x))$ is invariantly sufficient for the class of probability models $Y_{\boldsymbol{\theta}} = \{Y_\theta, \theta \in \Theta\}$.*

Next let us define *transitivity*.

DEFINITION 2.2.3. For each $n$, let $T_n$ be a function of $X_{(n)} = (X_1, \ldots, X_n)$. If, for each $n$ and all $\theta$, the conditional distribution of $T_{n+1}$ for given $X_{(n)}$ is equal to the conditional distribution of $T_{n+1}$ for given $T_n$, then $T = (T_1, T_2, \ldots)$ is said to be a *transitive sequence* for $Y_\theta$. (That is, all the information about $T_{n+1}$ contained in $X_{(n)}$ is carried by $T_n(X_{(n)})$).

*Invariant* SPRT's. Cox (1952a) proposed that sequential tests of simple hypotheses about $\lambda(\theta)$, constituting composite hypotheses about $\theta$, may be obtained by applying a SPRT or a GSPRT based on a sequence of statistics the distribution of which depends only on $\lambda$. An invariantly sufficient and transitive sequence is a natural candidate for such a sequence. Since $V_n$ is sufficient for the distribution of any invariant function of which $V_n$ is a function, $V_n$ is sufficient for the distribution of $V_{(n)} = (V_1, \ldots, V_n)$. Consequently, the ratio of densities of $V_{(n)}$ at fixed values of $\lambda$, namely $\lambda_1$ and $\lambda_0$, (on which any GSPRT is based) becomes the ratio of densities of $V_n$ at $\lambda_1$ and $\lambda_0$. Hence at any stage $n$, the GSPRT based on $V$ depends only on $V_n$ and not on $V_{(n)}$ and is thus a $V$-rule. This factorization is the essence of Cox's theorem.

*Properties of V-rules.* Wald's bounds provide approximate upper bounds on the error probabilities. If $V_n$ has a monotone likelihood ratio (MLR), the SPRT has been shown to terminately surely. (See Wirjosudirdjo (1961) and Ifram (1965).) If $V_n$ has MLR in $\lambda$, its OC function is monotone in $\lambda$ (which occurs in exponential families of densities). However, approximations to OC functions are not available in general. Very little is known about the ASN functions except those based on empirical studies.

Berk (1970) considers the sure termination of SPRT's based on exchangeable likelihood ratio and gives unified results for many of the SPRT's. Wijsman (1971) considers the exponential boundedness of the stopping time of an SPRT (that is, for some $0 < c < \infty$, $0 < \varrho < 1$, we have $P(N > n) < c\varrho^n$, $n = 1, 2, \ldots$ where $N$ denotes the stopping time). A distribution function is said to be obstructive if the stopping time of the SPRT based on the i.i.d. sequence is not exponentially bounded. Berk (1973) provides asymptotic results for invariant SPRT's that are analogous to those for SPRT's. Further, he shows that Bhate's (1955) approximation is asymptotically valid.

Pollak and Siegmund (1975) obtain approximations to the expected sample size of some sequential tests. Let $X_1, X_2, \ldots$ be an i.i.d. sequence of random variables having the exponential density function

$$f(x; \theta) = \exp(\theta x - \psi(\theta)), \qquad \theta \in \Theta$$

with respect to some $\sigma$-finite measure $\nu$. Let $S_n = \sum_1^n X_i \, (n = 0, 1, \ldots)$ let $G$ be a probability distribution on $\Theta$. Then for some $\theta_0 \in \Theta$

$$h(x, t) = \int_{\Theta} \exp[(y-\theta_0)x - t(\psi(y)-\psi(\theta_0))]\, dG(y)$$

and

$$T = \inf\{n:\ h(S_n, n) \geqslant a\}, \qquad (a > 1).$$

Statistical applications of the stopping rule $T$ have been given by Robbins (1970) where using simple arguments it was shown that

$$P_{\theta_0}(T < \infty) \leqslant 1/a \qquad (a > 1).$$

Letting

$$I(\theta) = (\theta-\theta_0)\psi'(\theta) - (\psi(\theta)-\psi(\theta_0))$$

and assuming that for $\theta \neq \theta_0$, $G'$ exists in a neighborhood of $\theta$ and is positive and continuous at $\theta$, Pollak and Siegmund (1975) obtain (as $a \to \infty$)

$$E_\theta(T) \simeq \left[2\ln a + \ln((\ln a)/I(\theta)) - \ln\{2\pi(G'(\theta))^2/\psi''(\theta)\} - 1\right]/2I(\theta) + o(1).$$

They also obtain some variations of the above result. Lai and Siegmund (1977, 1979) derive some further refinements to the expected sample size, analysing the excess over the boundary and apply their results to the sequential $t$, $\chi^2$ and $F$ test procedures. They develop renewal theory for nonlinear functions of a random walk $S_n$ by expanding the function and applying classical renewal theory to the dominant linear term. On the other hand, Woodroofe (1976a, 1977) considers the first passage of a random walk $S_n$ to a nonlinear boundary which he expands around an appropriate point and analyses it. Woodroofe's method seems to involve heavy computations.

**2.3. Likelihood ratio test procedures.** If it is not feasible to apply the method of weight functions, and if invariant considerations do not apply, then one should look for a procedure in which sample estimates replace the true values of the nuisance parameters. Estimates like BAN (best asymptotically normal) estimates will suffice. In particular we confine ourselves to maximum likelihood estimates, which have some well-known desirable properties. Maximum likelihood SPRT's have been proposed by Bartlett (1946) and D. R. Cox (1963). First we shall give Cox's procedure which is a slight modification of that of Bartlett. Let $X$ have density or probability function $f(x; \theta, \eta)$ where $\eta$ is considered to be the nuisance parameter and we are interested in testing $H_0$: $\theta = \theta_0$ versus $H_1$: $\theta = \theta_1$. Then consider the sequential procedure based on the likelihood ratio

$$Z_n = l(\theta_1, \hat{\eta}_n) - l(\theta_0, \hat{\eta}_n) \tag{2.3.1}$$

where

$$l(\theta, \eta) = \sum_{i=1}^{n} \ln f(X_i; \theta, \eta)$$

and $\hat{\eta}_n$ denotes the maximum likelihood estimate (mle) of $\eta$. Then if $|\theta_i - \theta|$ is sufficiently small ($i = 0, 1$) Cox (1963) has shown that for large $n$, one can carry out Wald's SPRT with bounds

$$\ln B \doteq C_n \ln \{\beta/(1-\alpha)\} \quad \text{and} \quad \ln A \doteq C_n \ln \{(1-\beta)/\alpha\} \tag{2.3.2}$$

where

$$C_n = \{1 - I^2_{\hat{\theta},\hat{\eta}} / I_{\hat{\theta},\hat{\theta}} I_{\hat{\eta},\hat{\eta}}\},$$

$$\hat{\theta} = \hat{\theta}_n \text{ is the mle of } \theta, \quad \hat{\eta} = \hat{\eta}_n,$$

$$I_{\theta\theta} = \operatorname{var}\{\partial \ln f(X; \theta, \eta)/\partial\theta\} \text{ etc.}$$

There might be some situations where it is easier to calculate the mle's of $\eta$ when $\theta = \theta_0$ and $\theta = \theta_1$ than to compute the joint mle's $\hat{\theta}_n$ and $\hat{\eta}_n$. Bartlett (1946) proposed that the sequential likelihood ratio procedure be based on

$$Z'_n = l(\theta, \hat{\eta}_1) - l(\theta_0, \hat{\eta}_0) \tag{2.3.3}$$

where $\hat{\eta}_1 (\hat{\eta}_0)$ denotes the mle of $\eta$ when $\theta = \theta_1 (\theta_0)$. For large $n$, Bartlett (1946) represents $Z'_n$ as a sum of i.i.d. random variables and hence Wald's approximations to the boundary values, OC function and ASN are applicable. It should be noted that asymptotically, Cox's and Bartlett's procedures are equivalent.

Let the $X$'s belong to a one parameter exponential family having the density

$$f(x; \theta) = \exp\{\theta x - b(\theta)\} \quad \text{for some} \quad \theta \in (\underline{\theta}, \bar{\theta}) \tag{2.3.4}$$

with respect to some nondegenerate sigma finite measure. Robbins and his associates have proposed "open ended tests" which like the one-sided SPRT, continue sampling indefinitely (with prescribed probability) when the null hypothesis is true and stop only if the alternative is to be accepted. Lorden (1973) has investigated the generalized likelihood ratio approach to the problem of open ended tests of $H_0$: $\theta = \theta_0$ versus $H_1$: $\theta > \theta_0$ for the family of densities given by (2.3.4).

**2.4. Tests for three hypotheses about the normal mean.** There are some practical situations which require a choice among three or more courses of action for example, the tolerance limits set to a machine may be too high, too low or acceptable. One can formulate this problem in terms of sequential testing among three hypotheses. If the three hypotheses can be ordered in terms of an unknown parameter, a sequential test may be devised by performing two SPRT's simultaneously, one between each pair of neighboring hypotheses (which are ordered). Armitage (1947, 1950) and Sobel and Wald (1949) obtained sequential tests satisfying certain conditions on the error probabilities.

*Armitage–Sobel–Wald Test.* Let $X$ be normally distributed having mean $\theta$ and known variance which without loss of generality, can be set to be unity. We are interested in accepting one of

$$H_0\colon \theta = \theta_0, \qquad H_1\colon \theta = \theta_1, \qquad H_2\colon \theta = \theta_2 \qquad (\theta_0 < \theta_1 < \theta_2) \quad (2.4.1)$$

on the basis of an i.i.d. sequence, $\{X_n\}$ $(n = 1, 2, \ldots)$. The above formulation was considered by Armitage (1947) whereas Sobel and Wald (1949) considered $H_0$: $\theta = \theta_0$, $H_1$: $\theta_1 \leqslant \theta \leqslant \theta_2$, $H_2$: $\theta = \theta_3$. Thus, we are considering the special case of $\theta_2 = \theta_1$. Since $T_n = X_1 + \ldots + X_n$ is sufficient for $\theta$, the fixed sample size procedure would be:

$$\text{accept } H_0 \text{ if } T_n \leqslant t_0, \qquad \text{accept } H_1 \text{ if } t_0 < T_n < t_1, \qquad \text{accept } H_2 \text{ if } T_n > t_1, \quad (2.4.2)$$

where $t_0$ and $t_1$ are chosen subject to

$$P(\text{reject } H_0 \mid H_0) \leqslant \gamma_0, \qquad P(\text{reject } H_1 \mid H_1) \leqslant \gamma_1, \qquad P(\text{reject } H_2 \mid H_2) \leqslant \gamma_2. \quad (2.4.3)$$

The sequential procedure is given by:

Let $R_1$ denote the SPRT for $H_0$ versus $H_1$ and $R_2$ be the SPRT for $H_1$ versus $H_2$. Then both $R_1$ and $R_2$ are carried out at each stage until either: one of the procedures leads to a decision to stop before the other. Then the former is stopped and the latter is continued until it leads to a decision to stop or: both $R_1$ and $R_2$ lead to a decision to stop at the same stage in which case no further experimentation is carried out. The final decison rule is:

$$\begin{aligned} &\text{accept } H_0 \text{ if } R_1 \text{ accepts } H_0 \text{ and } R_2 \text{ accepts } H_1, \\ &\text{accept } H_1 \text{ if } R_1 \text{ accepts } H_1 \text{ and } R_2 \text{ accepts } H_1, \\ &\text{accept } H_2 \text{ if } R_1 \text{ accepts } H_1 \text{ and } R_2 \text{ accepts } H_2. \end{aligned}$$

Sobel and Wald (1949) provide a sufficient condition for the impossiblity of accepting both $H_0$ and $H_2$. The above rule terminates finitely with probability one. The bounds for the two SPRT's can be determined by specifying another constant $\eta$ besides $\gamma_0$, $\gamma_1$ and $\gamma_2$ and one can obtain lower and upper bounds for $\eta$. One can derive the OC function and bounds for the ASN.

*Remark* 2.4.1. Although the Sobel–Wald procedure is not an optimum procedure in the sense that the terminal decision is not in every case, a function of only the sufficient statistic, Sobel and Wald (1949) claim that their procedure is not far from being optimum. All the above considerations can be extended to $k$-decision problems. Armitage (1950) independently of Sobel and Wald (1949), has proposed a sequential procedure for $k$-hypotheses testing problem.

Paulson (1963) provided a sequential solution to the following problem: Let $\Theta$ denote the unknown mean of a normal population. Let $\{I_j\}$ ($j = 1, \ldots, k$) denote $k$ nonoverlapping intervals whose union is the real line. Based on a sample $\{X_i\}$ it is of much practical interest to decide to which of the $k$ intervals $\theta$ belongs. Solution is available both when $\sigma$ is known and unknown.

Armitage (1947), Rushton (1952) and Billard and Vagholkar (1969) propose sequential procedures for testing a two-sided hypothesis about the normal mean. Also, Armitage's procedure is applicable to testing two-sided alternatives about parameters indexing arbitrary distributions. These procedures are adaptations of carrying out two SPRT's simultaneously. For further details see Govindarajulu (1981, Section 3.10).

**2.5. Efficiency of the** SPRT. Let $\{X_i\}$ be an i.i.d. sequence having density function $f(x;\theta)$ where $\theta$ is real and the parameters space $\Theta$ is a part of the real line. Suppose we are testing $H_0$: $\theta = \theta_0$ versus $H_1$: $\theta = \theta_1$ where $\theta_0 \neq \theta_1$, subject to the prescribed error probabilities $\alpha$ and $\beta$. If the optimum fixed-sample size is $n(\alpha, \beta)$ the relative efficiency of the SPRT at $\theta \in \Theta$ is defined by

$$R_e(\theta) = n(\alpha, \beta)/E_\theta(N) \tag{2.5.1}$$

where $N$ is the stopping time of the SPRT. In particular, if $X$ is normal with mean $\theta$ and known variance $\sigma^2$ and if $\beta = \alpha$ then

$$R_e(\theta_1) = R_e(\theta_0) = 2z_{1-\alpha}^2\left[(1-2\alpha)\ln\left(\frac{1-\alpha}{\alpha}\right)\right]^{-1} \tag{2.5.2}$$

where $z_{1-\alpha}$ denotes the $100(1-\alpha)$th percentile of the standard normal distribution. One can easily verify (at least numerically) that the right-side expression is increasing in $\alpha$ and the percentage saving by the SPRT is 46.4 when $\alpha = .10$. Paulson (1947) has shown that when $\theta_1$ is close to $\theta_0$, the efficiency of the SPRT relative to the optimum fixed-sample size procedure is free of the form of $f(x;\theta)$ and the particular values of $\theta_0$ and $\theta_1$. If $X$ is normal with mean $\theta$ and known variance $\sigma^2$, then one can obtain

$$\inf_\theta R_e(\theta) = \frac{n(\alpha, \alpha)}{\sup_\theta E_\theta(N)} = \left[\frac{2z_{1-\alpha}}{\ln\{(1-\alpha)/\alpha\}}\right]^2 = \psi(\alpha), \qquad \text{for} \quad 0 < \alpha < \tfrac{1}{2}. \tag{2.5.3}$$

One can also show that $\lim_{\alpha\to 0}\psi(\alpha) = 0$, $\lim_{\alpha\to 1/2}\psi(\alpha) = \pi/2$ and that $\psi(\alpha)$ is monotonically increasing in $\alpha$.

Bechhofer (1960) has studied the limiting relative efficiency of the SPRT for the normal case when the error probabilities are related to each other and they tend to zero. He brings out some of the surprises that are never anticipated. Sakaguchi (1967) has extended Bechhofer's (1960) results to the exponential family using Chernoff's (1952) information number discriminating between the two densities. Berk (1973) obtained elegant results for the asymptotic efficiency of Wald's SPRT relative to the best nonsequential test having the same error probabilities ($\alpha$, $\beta$).

**2.6. Bayes sequential procedures.** For the binomial problem, Vagholkar and Wetherill (1960) gave a most economical sequential sampling method which is based on the basic theory developed by Barnard (1954). Lindley and Bartlett (1965) gave an optimal Bayes squential procedure that can be solved numerically by the backward induction techniques of dynamic programming. Chernoff (1968) summarizes the results pertaining to the Bayes sequential testing procedure for the mean of a normal population. Schwarz (1962) provided an elegant characterization of the asymptotic (as $c$ the cost per observation tends to zero) shape of the Bayes continuation region, when the density function $f(x;\theta)$, $\theta \in \Theta$, belongs to the one parameter exponential family and the hypothesized parameter sets are separated by an indifference zone. Assuming no indifference zone, Kiefer and Sacks (1963) have generalized and extended the results of Schwarz (1962), for testing two composite hypotheses. Woodroofe (1976) studied in detail Schwarz's (1962) approximation to the Bayesian test when the underlying population is normal. Using the method of truncation and backward induction, Arrow, Blackwell and Girshick (1949) obtain the best (Bayes) sequential procedures; in particular, they characterize the best acceptance regions in the space of prior probability distributions for the $k$-decision problems.

**2.7. Tests of power one.** Let $X_1, X_2, \ldots$ be i.i.d. normal ($\theta$, 1) and we wish to test $H_0$: $\theta \leqslant 0$ versus $H_1$: $\theta > 0$. For the stopping time, Robbins (1970) proposes

$$N = \text{first } n \geqslant 1 \text{ such that } S_n \geqslant c_n, \quad = \infty \text{ if no such } n \text{ occurs}, \tag{2.7.1}$$

and when $N < \infty$, continue sampling indefinitely and do not reject $H_0$, where

$$S_n = X_1 + \ldots + X_n \quad \text{and} \quad c_n = (n \ln n)^{1/2}. \tag{2.7.2}$$

Note that $P_\theta(\text{not reject } H_0) = P(S_n < c_n \text{ for all } n \geqslant 1) = 0$ for all $\theta > 0$, because $c_n/n \to 0$ and $S_n/n \to \theta > 0$ as $n \to \infty$. Thus the test has power one against $H_1$. Clearly, the proposed test will rarely terminate when $\theta < 0$. One can also obtain

$$E_\theta(N) \geqslant -2\ln P_0(N < \infty)/\theta^2, \quad \text{for every} \quad \theta > 0. \tag{2.7.3}$$

Darling and Robbins (1968a) have shown that

$$E_0(N) \leqslant \frac{c_K}{\theta} + \frac{\phi(\theta)}{\theta \Phi(\theta)} + 1 \tag{2.7.4}$$

where $K = E_\theta(N)$. Thus the proposed test has type I error probability $\leqslant \alpha$ (uniformly) and type II error probability $= 0$. By the same token while the sample size $N$ is finite with probability one when $H_1$ is true, $N$ is equal to $\infty$ with probability $\geqslant 1-\alpha$ when $H_0$ is true. Darling and Robbins (1968a) provide a practical situation where the preceding procedure is meaningful. Barnard (1969) quite independently of Darling and Robbins (1968a) has proposed tests of power one for a Bernoulli problem. For details the reader is referred to Govindarajulu (1981, Section 3.13).

**2.8. Locally most powerful (LMP) tests.** Let $f(x;\theta)$ denote a family of probability densities. Let $\Theta$, the parameter space be a subinterval of the real line. Let $X$, $X_1$, $X_2$, ... be an i.i.d. sequence of random variables having the above density. We wish to test $H_0$: $\theta = \theta^*$ versus $H_1$: $\theta > \theta^*$. Berk (1975a) has shown that one can obtain sequential tests that are LMP for testing $H_0$ versus $H_1$. Let $\alpha$ denote the level of significance and $\nu$ the expected time under $H_0$. Then among all level $\alpha$ sequential tests whose expected stopping times under $H_0$ do not exceed $\nu$, the given test is LMP. These LMP tests have for their stopping times:

$$N = \inf[n\colon\ S_n \notin (b, a)] \tag{2.8.1}$$

where $S_n = \sum_1^n r(X_i)$, $r(x) = \{\partial \ln f(x;\theta)/\partial\theta\}_{\theta=\theta^*}$. Typically $-b$ and $a$ are both positive and they are determined (in principle) by $\alpha$ and $\nu$ where

$$\alpha = P(S_n \geqslant a) \qquad \text{and} \qquad \nu = EN$$

and unless otherwise specified all probabilities and expectations are computed under $H_0$. Further, Berk (1975a) has shown that an LMP test is a Wald's SPRT of $H_0'$: $\theta = \theta_0$ versus $H_1'$: $\theta = \theta_1$ where $\theta_0 < \theta^* < \theta_1$ when $f(x;\theta) = \exp\{\theta x - b(\theta)\}$.

**2.9. Nonparametric tests.** Let $X(Y)$ have the d.f. $F(x)(G(y))$ and let $D^+(F, G) = \sup_x [F(x) - G(x)]$. Assume that $X$ and $Y$ are independent. We wish to test $H_0$: $F(x) \leqslant G(x)$ for all $x$. Darling and Robbins (1968b) propose a power one sequential test for the above hypothesis and obtain a lower bound for type I error probability and upper bound for EN. They also propose power one tests for three other hypotheses.

*Sequential sign test.* Let $X$ and $Y$ be independent and let $p = P(X < Y)$. Suppose we wish to test $H_0$: $F(x) = G(x)$, all $x$ versus $H_1$: $F(x) < G(x)$ for all $x$. This is equivalent to testing $H_0'$: $p = \frac{1}{2}$ versus $H_1'$: $p > \frac{1}{2}$. If we observe pairs $(X_1, Y_1)$, $(X_2, Y_2)$, ... reducing the data to signs of $(X_i - Y_i)$, $i = 1, 2, \ldots$ Hall, Wijsman and Ghosh (1965, Section 1.8) propose an SPRT based on the reduced data constituting a Bernoulli sequence with $P$ (positive sign) $= p$.

*Rank order* SPRT's *based on Lehmann alternatives: two sample case.* Let $X$ have d.f. $F(x)$ and $Y$ have d.f. $G(y)$. Suppose we wish to test $H_0$: $F(x) = G(x)$ for all $x$. If $H_1$: $G(x) \geqslant F(x)$ with strict inequality for some $x$, then Lehmann's (1953) alternative which is contained in $H_1$ is $H_{1\Delta}$: $G = F^{\Delta}$. Similarly if $H_1$: $F(x) \geqslant G(x)$, then an associated Lehmann alternative is $H_{1\Delta}$: $G = 1-(1-F)^{\Delta}$ for $\Delta > 0$. The general Lehmann alternative is given by $G = h(F)$ such that $h(0) = 0$, $h(1) = 1$ and $h(\cdot)$ is nondecreasing in $(\cdot)$. Let $X_1, \ldots, X_m$ denote a random sample from $F$ and $Y_1, \ldots, Y_n$ denote a random sample from $G$ where $X$'s and $Y$s are mutually independent and $F$ and $G$ are assumed to be continuous. Let $s_1 < s_2 < \ldots < s_n$ denote the ranks of $Y$'s in the combined sample of size $N = m+n$. Then $S = (s_1, \ldots, s_n)$ is called the *rank order*. One can derive an explicit expression for the probability of a rank order under Lehmann alternatives. Using this result Wilcoxon, Rhodes and Bradley (1963) developed configural rank sum test (SCR-test) and the rank sum test. For both of these tests, observations are taken in groups of $k$ $X$'s and $k$ $Y$'s with observations ranked within the groups. Bradley, Merchant and Wilcoxon (1966) provide a modified version of the configural group rank test proposed earlier, whose superiority is established via Monte Carlo studies. Based on Monte Carlo studies Bradley, Martin and Wilcoxon (1965) infer that the rank sum test derived on the basis of Lehmann alternatives is reasonable for data from normal populations differing only in locations. Savage and Sethuraman (1966) consider the case $k = 1$, (that is, only a single pair is observable at each stage) and derive the SPRT based on the ratio of $P(s_1, \ldots, s_n | H_1)/P(s_1, \ldots, s_n | H_0)$ computed under Lehmann alternatives. Savage and Sethuraman (1966) and Sethuraman (1970) consider the exponented boundedness of the stopping time. Their combined effort yields that under the assumption that a certain random variable is zero with probability less than 1, the stopping time is exponentially bounded. From this it follows that the SPRT terminates finitely with probability one and the moment-generating function of the stopping time is finite.

The asymptotic normality of the log of the test criterion is established by Govindarajulu (1968) and Sethuraman (1970) from which the finite sure termination of the SPRT follows. Berk and Savage (1968) consider sequentially testing $H_0$: $G = F$ versus the alternative $H_1 = \psi$: $F = \psi_1(K)$, and $G = \psi_2(K)$ where $\psi_1$ and $\psi_2$ are given df's on $[0, 1]$ that specify $\psi$, and $K$ ranges through all df's.

*One-sample rank order* SPRT's *for symmetry.* Let $V_1$, $V_2$, ... be i.i.d. random variables observed sequentially and having a continuous d.f. $F$. We wish to test the hypothesis

$H_0$: $F(v)+F(-v) = 1$ for all $v$ (that is, $F$ is symmetric about zero).

Let

$$H(v) = P(V \leqslant v \mid V \geqslant 0) = \{F(v)-F(0)\}/\{1-F(0)\}$$

and

$$G(v) = P(|V| \leqslant v \mid V < 0) = \{F(0)-F(-v)\}/F(0)$$

for $v \geqslant 0$ and $H$ and $G$ are zero for $v < 0$. Thus

$$F(v) = \begin{cases} F(0)\{1-G(-v)\}, & \text{for } v < 0, \\ H(v)+F(0)\{1-H(v)\}, & \text{for } v \geqslant 0. \end{cases}$$

Then we can rewrite $H_0$ as

$$H_0\colon\ H(v) = G(v) \quad \text{for all } v \quad \text{and} \quad F(0) = 1/2,$$

and take

$$H_a\colon\ H(v) \neq G(v) \quad \text{for some } v.$$

Assuming that $H_a$: $H(v) = 1-[1-G(v)]^{\Delta}$, $v \geqslant 0$ with $F(0) = \Delta/(1+\Delta)$, Weed and Bradley (1971) proposed two sequential procedures. One based on within-group configuration of signed ranks and the other based on within-group sums of positive signed ranks.

Parent (1968) defined sequential rank and signed sequential rank which makes it unnecessary to re-rank all the observations at each stage. Weed, Bradley and Govindarajulu (1974) have considered SPRT's based on ratios of probabilities of rank orders computed under Lehmann alternatives and established sure termination of the procedure under certain assumptions on the underlying d.f. $F$. Miller (1970) proposed a sequential procedure based on the sequential rank for testing symmetry and shows that asymptotically the test criterion behaves like a Brownian motion process. Lombard (1976) generalizes the results of Miller (1970). Groeneveld (1971) proposed a different sequential test for symmetry.

*c-sample rank order* SPRT's. Govindarajulu (1977) considered $c$-sample rank order SPRT's based on Lehmann alternatives and studied the exponential boundedness of the corresponding stopping times. The sufficient condition is again that a random variable takes the value zero with probability strictly less than one. The rank order SPRT's for randomness against trend are given in Govindarajulu (1981).

*Asymptotic efficiency.* Lai (1978) has extended the concept of Pitman efficiency to the sequential case. For details see Govindarajulu (1981, pp. 316–318).

## 3. Sequential estimation

In some applications, formulating a problem in terms of hypothesis-testing would be somewhat artificial; however, formulating it as an estimation problem would be more realistic. In the fixed-sample size situation there is a close connection between acceptance regions and confidence regions, whereas that analogy does not hold in the sequential case. Hence there is a need for a theory of sequential estimation. The stopping rules in sequential testing may not be meaningful in sequential estimation. However, Wijsman (1981, 1982) brings out the close analogy between a sequential $(1-\alpha)$-confidence set for an unknown parameter and the associated family of level $\alpha$ SPRT's or the GSPRT's. Siegmund (1981) also brought out similar analogy when the unknown parameter is the normal mean. See also Siegmund (1978).

**3.1. Stein's two-stage procedure.** There does not exist a fixed-sample size procedure for estimating the mean of a normal population (when the variance is unknown) with a confidence interval of fixed width and specified confidence coefficient. Stein (1945) presented a two-stage procedure, in which the size of the second sample depends upon the observations in the first sample.

Ghurye and Robbins (1954) gave a two-stage procedure for estimating the difference of two means when the variances are unknown with extension to the nonnormal populations. Richter (1960) gave a two-stage procedure for estimating the common mean of two normal populations.

**3.2. Sufficiency and completeness.** Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables having common p.d.f. $f(x;\theta)$. We wish to estimate $\theta$ by some function $\delta(X_1, \ldots, X_i)$ while using a stopping rule which is closed (that is, for every $\theta$, $P_\theta(N \leqslant n) \to 1$ as $n \to \infty$, although not necessarily uniformly in $\theta$). The sample space is $E_1 + E_2 + \ldots$, where $E_i$ is contained in $R^i$ and consists of those points $(X_1, \ldots, X_i)$ which serve as stopping points. Let $N$ denote the stopping variable. Let $T_n = T(X_1, \ldots, X_n)$ be a sufficient statistic for the joint density of $X_1, \ldots, X_n$.

DEFINITION 3.2.1. With the preceding notation, the sequence $(T_1, T_2, \ldots)$ is called a *sufficient sequence* for the sequential model. Then one can easily show that $(T_N, N)$ is a sufficient statistic for $\theta$ in the sequential case.

In Section 2.2 we have defined transitivity of a sequence of statistics. Transitivity of $\{T_n\}$ implies that all the information concerning $T_{n+1}$ contained in $\underline{X}_n = (X_1, \ldots, X_n)$ is also contained in the function $T_n(\underline{X}_n)$. Bahadur (1954) showed that if $\{T_n\}$ is sufficient and transitive, then any closed sequential procedure based on $\{\underline{X}_n\}$ is equivalent to a procedure which at stage $n$ is based only on $T_n$. In the case of i.i.d. variables, $\{T_n\}$ is transitive if $T_{n+1}(\underline{X}_{n+1}) = \psi_n(T_n(\underline{X}_n), X_{n+1})$, for every $n > 1$. The exponential family enjoys this property.

Next, we consider the completeness of $(T_N, N)$. Assume that $T_n$ is complete for every fixed $n$.

DEFINITION 3.2.2. The family of distributions of $(T_N, N)$ is said to be *complete* (*boundedly complete*) if for every (bounded) $g(n, t_n)$, $E_\theta\{g(N, T_N)\} = 0$ for all $\theta$ implies that $g(n, t_n) = 0$ almost everywhere for all $n > 1$.

Lehmann and Stein (1950) found a general necessary condition for completeness of the statistic $(N, T_N)$ and examined the stopping rules for which $(N, T_N)$ is complete. In the case of normal $(\theta, 1)$, $(N, T_N)$ is complete if $N = m$ (here $T_m = \sum_1^m X_i$). Let $S_m$ be the set of values of $T_m$ for which we stop at the $m$th observation. A necessary and sufficient condition for $(N, T_N)$ to be complete is that the $S_i$'s are disjoint intervals, each lying immediately above the preceding one. Zaidman, Linnik and Romanovski (1969) consider sequential estimation of the natural parameter in (i) binomial, (ii) multinomial, (iii) Poisson and (iv) Wiener processes.

**3.3. Cramér–Rao lower bound.** Suppose that we are interested in estimating $\theta$ with loss function given by $r(\delta(x); \theta) = [\delta(x) - \theta]^2$. Assume that we have limited resources (forcing a bound on the total cost of experimentation $C(N)$). Then we seek to minimize the risk function (expectation of loss function) subject to an upperbound $n_0$ on the expectation of $C(N)$. If we restrict ourselves to unbiased estimates of $\theta$, we would be interested in lower bounds for the variance of such estimates. The Cramér–Rao inequality was extended to the sequential case by Wolfowitz (1947b) which will be given below.

THEOREM 3.3.1. *Let $T_N = \delta(X_1, \ldots, X_N)$ be an estimate of $\theta$, such that $E_\theta(T_N) = \theta + b(\theta)$. Suppose that differentiation underneath the summation and integral signs is permissible in $E_\theta(1) = 1$ and $E_\theta(T_N) = \theta + b(\theta)$ where $b'(\theta)$ exists. Then*

$$\operatorname{Var}(T_N|\theta) \geq \left\{E_\theta(N) E_\theta\left[\frac{\partial}{\partial\theta} \ln f(X; \theta)\right]^2\right\}^{-1} [1 + b'(\theta)]^2. \tag{3.3.1}$$

COROLLARY 3.3.1.1. *If we restrict ourselves to sequential unbiased estimation procedures for which $E_\theta(N) \leq n_0$, then*

$$\operatorname{Var}(T_N|\theta) \geq \left\{n_0 E_\theta\left[\frac{\partial}{\partial\theta} \ln f(X; \theta)\right]^2\right\}^{-1} \tag{3.3.2}$$

*If $T_N$ is an unbiased estimator for $h(\theta)$, then one can analogously obtain*

$$\operatorname{Var}(T_N \mid \theta) \geq [h'(\theta)]^2 \Big/ \left\{ E_\theta(N)\, E_\theta \left[ \frac{\partial}{\partial \theta} \ln f(X;\theta) \right]^2 \right\}. \tag{3.3.3}$$

Seth (1949) extended Bhattacharyya's (1946) bounds to the sequential case, which in some respect are more general than those of Wolfowitz (1947b).

**3.4. Consistency of estimators.** Given an unbiased estimator and a sufficient statistic for a certain parameter of interest, using Rao–Blackwell procedure, we can obtain another (perhaps) better estimator by taking the expectation of the given estimator conditional on the sufficient statistic. Wolfowitz (1947a) considered the consistency of such estimators for the binomial parameter in the sequential case. Loynes (1969) studied the consistency of the Rao–Blackwell type of sequential estimators for the general case under certain regularity conditions. He also provides some examples. Berk (1969) considered the strong consistency of Rao–Blackwell type of estimators.

**3.5. Certain double sampling procedures.** Suppose that we are interested in estimating an unknown parameter $\theta$ having specified accuracy, using as small a sample size as possible. The precision could be in terms of variance $a(\theta)$, some given function of $\theta$. Another problem of interest is to estimate $\theta$ by a confidence interval having a specified width and a specified confidence coefficient $\gamma$. In general it is not possible to obtain a fixed-sample size procedure meeting the specifications. Thus, one has to resort to some kind of sequential sampling. Although a general large sample theory is provided by Anscombe (1949, 1952), it is somewhat difficult to construct a sequential sampling procedure leading to an estimate having the required properties. Also, the ordinary sequential procedure has the additional drawback of requiring calculations at each stage. Cox (1952b) proposed a double sampling procedure in which one draws a preliminary sample of observations which determines how large the total sample size should be. Stein's procedure is a special case of Cox's procedure because the underlying distribution is known. The double sampling procedures are different from those used in industrial inspection, because in the latter case, the second sample is of fixed size. Although the double sampling procedures have optimum large sample properties, they are likely to be reasonable when based on small samples. Moreover, Cox's procedure is applicable even if there are nuisance parameters.

Suppose we wish to estimate a single unknown parameter $\theta$ with specified variance $a(\theta)/\lambda$ where $\lambda$ is tending to infinity. Then Cox provides an estimator having bias $O(\lambda^{-1})$ and variance $a(\theta)\lambda^{-1}[1+O(\lambda^{-1})]$. For further details see Govindarajulu (1981, pp. 381–383).

**3.6. Minimax estimation.** Let the random variable $X$ have the density function $f(x;\theta)$. Suppose we desire to estimate $h(\theta)$ by $\delta(X)$. Let $r(\delta(X);\theta)$ and $R(\delta;\theta)$ respectively denote the loss incurred and risk associated with the estimator $\delta(X)$.

DEFINITION 3.6.1. An estimator $\delta^*$ is said to be *minimax* if for any other estimator $\delta(X)$

$$\sup_\theta R(\delta^*;\theta) \leqslant \sup_\theta R(\delta;\theta). \tag{3.6.1}$$

For minimizing $\sup_\theta R(\delta;\theta)$ subject to $E_\theta(M) \leqslant m$, Hodges and Lehmann (1951) obtain minimax estimates for normal mean using Cramér–Rao inequality and show that they are sample means based on a fixed-sample size. If $\theta$ is bounded, then $\bar{X}_m$ is neither admissible nor minimax for all monotone loss functions. Another way of computing minimax estimates is via Bayes estimation with respect to least favorable prior distributions. Several specific examples have appeared in the literature in which the minimax estimator turns out to be the one based on fixed sample size. Wald (1971) considered the minimax estimation of the mean of a rectangular distribution and succeeded in obtaining an honest sequential procedure.

Ibragimov and Has'minskii (1972) show that fixed-sample size procedures are asymptotically minimax in the class of sequential plans with quadratic loss functions. Ibragimov and Has'minskii (1974a) also show that in the minimax sense, sequential estimation procedures having a mean number of observations not exceeding $n$, do not yield asymptotic advantage for power loss functions. Imposing certain regularity conditions on the loss function and assuming that the underlying probability measures are locally asymptotically normal, Efroimovich (1980) shows that in the minimax sense, sequential estimation schemes having a mean number of observations not exceeding $n$, do not yield asymptotic advantage for a much broader class of loss functions which includes the power functions as a special case.

**3.7. Certain nonsequential Bayes sequential estimates.** Here we consider certain optimal Bayes estimation procedures that are nonsequential. Let $\underline{X}_n = (X_1, \ldots, X_n)'$ denote a vector of $n$ observations and let $T(\underline{X}_n)$ be designed to estimate the parameter $\theta$ that indexes the distribution of $X_i$. Also $r(T;\theta)$ denotes the loss incurred in using $T$ for estimating $\theta$. In sequential estimation, one chooses the estimation rule $T(\underline{X}_n)$ and a stopping rule to minimize $E\{CN + r(T;\theta)\}$ where $C$ denotes the constant cost per observation. However, there are several situations where the optimal stopping rule turns out to be simply $N = n_0$, a constant. Blackwell and Girshick (1954) gave a general result characterizing the situation under which the optimal Bayes sequential procedure is nonsequential. In the following we give their result.

THEOREM 3.7.1. *Let $X_1, X_2, \ldots$ be an i.i.d. sequence of random variables. If $E[r(T(\underline{X}_n); \theta) | \underline{X}_n] \leqslant K(n)$ for all $\underline{X}_n$ and $n$ where $k(n) \to 0$ as $n \to \infty$, then the optimal sequential procedure is truncated. If $E[r(T(\underline{X}_n); \theta) | \underline{X}_n] = k(n)$ (that is, it is a function of $n$ only), then the optimal sequential procedure is a fixed-sample size procedure.*

Using Theorem 3.7.1, Whittle and Lane (1967) obtained sufficient conditions under which, the optimal sequential estimate of the parameter (scalar or vector valued) of an exponential family of densities is based on a fixed-sample size.

**3.8. Large-sample theory for estimators.** Anscombe (1949) provided a large-sample theory for sequential estimators when there is only one unknown parameter in the distribution of the observations. Using an heuristic argument, he showed that an estimation formula valid for fixed sample size remained valid when the sample size was determined by a sequential stopping rule. Another proof was given by Cox (1952c) suggesting that fixed-sample size formulas might be valid generally, for sequential sampling, provided the sample size is large. Anscombe (1952) introducing the concept of "uniform continuity in probability" (UCIP) simplified his previous work and gave sufficient conditions for the asymptotic normality of a statistic indexed by a positive integer valued random variable: one of the basic conditions being UCIP. Let $\{Y_n\}$ be a sequence of statistics and $W_n$ be a linear measure of dispersion of $Y_n$ like its standard deviation. Then $\{Y_n\}$ is said to have UCIP if for given small $\varepsilon$ and $\eta$ there exists a large $\nu$ and a small positive $c$ such that for any $\eta \geqslant \nu$ we have $P[|Y_{n'} - Y_n| < \varepsilon W_n$ simultaneously for all integers $n'$ such that $|n' - n| < cn] > 1 - \eta$. It is of interest to know under what conditions we will have UCIP. The following result is proved in Govindarajulu (1981, p. 421).

THEOREM 3.8.1. *Let $Y_n$ be an estimate of $\theta$ calculated from the first $n$ observations. Let*

$$Y_n - \theta = n^{-1} \sum_{i=1}^{n} Z_i + R_n \tag{3.8.1}$$

*where the $Z_i$ are independent with $EZ_i = 0$ and* var $Z_i \leqslant b \leqslant \infty$ *and $n^{1/2} R_n = o(1)$ almost surely. Then $Y_n$ is* UCIP.

*Remark* 3.8.1.1. In many applications, like the sample quantile and the maximum likelihood estimate, the representation in (3.8.1) is satisfied.

Since in most practical applications, we will be concerned with random sums of i.i.d. random variables, the following result would supplement the main result of Anscombe (1952).

THEOREM 3.8.2 (Renyi (1957) and Wittenberg (1964)). *Let $X_1, X_2, \ldots$ be i.i.d. random variables having mean* 0 *and variance* 1, *and define $S_n = X_1 + \ldots +$*

$+X_n$. *If* $N_1, N_2, \ldots$ *is a sequence of positive integer-valued random variables (defined on the same probability space) such that* $N_n/n$ *converges in probability to a positive constant* $\xi$, *then* $S_{N_n}/(n\xi)^{1/2}$ *converges in law to a standard normal variable as* $n \to \infty$.

Bhattacharyya and Mallik (1973) employ Theorem 3.8.2 in order to establish the asymptotic normality of the stopping time of Robbins' (1959) procedure for estimating the normal mean $\mu$ when the variance is uknown, with $(\hat{\mu}-\mu)^2+cn$ as the loss function where $c$ denotes the cost per observation and $\hat{\mu} = n^{-1}\sum_1^n X_i$ and the $X_i$ are i.i.d. normal $(\mu, \sigma^2)$. In particular they show that

$$(N-c^{-1/2}\sigma)(\tfrac{1}{2}c^{-1/2}\sigma)^{1/2} \stackrel{d}{=} \text{normal}\,(0, 1) \qquad \text{as} \quad c\downarrow 0. \tag{3.8.2}$$

Siegmund (1968) proved the following result for the asymptotic normality of stopping times that commonly arise in sequential estimation.

THEOREM 3.8.3. *Let* $X_1, X_2, \ldots$ *be i.i.d. random variables with* $EX_1 = \mu > 0$, var $X_1 = \sigma^2 < \infty$ *and let* $T_n = X_1 + \ldots + X_n$. *If* $N$ *is the smallest* $n$ *for which* $T_n \geqslant c^{-1}n^{\delta}$, $0 < \delta < 1$, *then as* $c \downarrow 0$

$$\{\mu(1-\delta)\sigma^{-1}\lambda_c^{-1/2}\}(N-\lambda_c) \stackrel{d}{\approx} \text{normal}(0, 1) \tag{3.8.3}$$

*where* $\lambda_c = (c\mu)^{1/(1-\delta)}$.

Woodroofe (1977) has studied the asymptotic normality of stopping times for more general barriers, when the underlying random variables are positive.

## 4. Specific problems in estimation

In the section we shall briefly deal with some specific problems that have been considered in the literature.

**4.1. Estimation of the normal mean.** Anscombe (1953), Ray (1957) and Starr (1966a) considered the estimation of the normal mean with specified width and confidence coefficient.

Let $X_1, X_2, \ldots$ be i.i.d. normal $(\mu, \sigma^2)$. We wish to estimate $\mu$ with an interval having width $2d$ and confidence coefficient $1-\alpha(0 < \alpha < 1)$ when $\sigma$ is unknown. For any positive integer $n$ let

$$Y_n = \sum_{i=1}^{n} X_i/n. \tag{4.1.1}$$

Let $u$ denote the $(1-\alpha/2)$th fractile of the standard normal distribution. If $\sigma$ is known, the confidence interval (CI) is $(Y_n-d,\ Y_n+d)$ where $n$ is determined from

$$nd^2 = \sigma^2 u^2. \tag{4.1.2}$$

If $\sigma$ is unknown the first order asymptotic procedure is to continue taking observations and to stop as soon as

$$s^2 \leqslant d^2 n/u^2 \tag{4.1.3}$$

where $s^2 = \sum_{i=1}^{n} (X_i - Y_n)^2/(n-1)$. Let $N$ denote the stopping variable. Letting

$$U_i = (iX_{i+1} - \sum_{j=1}^{i} X_i)^2/i(i+1), \qquad i = 1, \ldots, n-1$$

and $Y_n = (X_1 + \ldots + X_n)/n$, one can rewrite the stopping rule as stop sampling as soon as

$$\sum_{i=1}^{n-1} U_i \leqslant d^2 n(n-1)/u^2. \tag{4.1.4}$$

If further observations $X_{n+1}, X_{n+2}, \ldots$ are taken, the sequence $U_i$ is increased by adding $U_n, U_{n+1}, \ldots$, the earlier elements remaining unchanged. Anscombe's (1953) procedure, when $d$ is small and $n$ is large is given by: stop sampling when

$$\sum_{i=1}^{n-1} U_i \leqslant d^2 n(n-2.676-\tfrac{1}{2}u^2)/u^2 \qquad (n \geqslant 4). \tag{4.1.5}$$

The constant 2.676 can be modified by addition of a constant or any function of $n$ of order $O(n^{1/2})$ when $n$ is large. The expected sample size obtained by Anscombe (1953) for the rule (4.1.5) is

$$E(N) = \sigma^2 u^2/d^2 + (1+u^2)/2, \tag{4.1.6}$$

where the first term denotes the number of observations required if $\sigma$ is known.

If we wish to estimate $\mu$ with specified standard error $a$, then the rule is to stop sampling when

$$\sum_{i=1}^{n-1} U_i \leqslant a^2(n-5)(n+0.824) \qquad (n \geqslant 6) \tag{4.1.7}$$

and the expected sample size is $E(N) = (\sigma^2/a^2) + 2$. An analogous procedure holds for sequentially estimating the difference of two normal populations having common unknown variance.

Ray's (1957) modified procedure is to stop as soon as

$$s^2 t_{n-1}^2/n \leqslant d^2 \tag{4.1.8}$$

where $t_{n-1}$ denotes $(1-\alpha/2)$th fractile of the $t$-distribution with $n-1$ degrees of freedom. If the procedure is slightly modified to taking three observations at the initial stage and observing pairs at each subsequent stage the rule becomes: stop sampling when

$$\sum_{i=1}^{2m} U_i \leqslant d^2(2m+1)2m/t_{2m}^2. \tag{4.1.9}$$

Ray (1957) indicates that one can write down explicit expressions for $P_{2m+1}$ which is the probability that we stop sampling at the $(2m+1)$st observation, from which one can easily compute the coverage probability and the mean and variance of $N$. Ray's (1957) approach can be used for estimating the mean with given standard error or estimating the difference of two means.

Starr (1966a) gave another modified procedure which is as follows: observe the $X$'s one at a time and stop at $X_n$ if $N$ is the first integer $n \geqslant n_0$ such that $s_n^2 \leqslant nd^2 t_{n-1}^2$, where $n_0 \geqslant 2$ is a fixed integer; then compute the interval $I_N = [Y_N - d, Y_N + d]$. Let

$$\lambda = \sigma/d, \qquad C(\lambda) = P(\mu \in I_N), \qquad D(\lambda) = EN, \tag{4.1.10}$$

$$\tau_\lambda = C(\lambda)/(1-\alpha) \qquad \text{and} \qquad \eta_\lambda = D(\lambda)/u^2\lambda^2. \tag{4.1.11}$$

Then Starr (1966a) shows that

$$P(N = n_0) \to 1, \qquad C(\lambda) \to 1 \qquad D(\lambda) \to n_0 \qquad \text{as} \quad \lambda \to 0 \tag{4.1.12}$$

and

$$\tau_\lambda \to 1 \qquad \eta_\lambda \to 1 \qquad \text{as} \quad \lambda \to \infty. \tag{4.1.13}$$

Without affecting the results in (4.1.12) and (4.1.13), one can modify Starr's (1966a) procedure to the following where computations are more tractable:

$$\begin{gathered} N^* \text{ is the first odd integer } n \geqslant n_0^* \text{ such that } s_n^2 \leqslant nd^2/t_{n-1}^2, \\ n_0^* \text{ is a fixed odd integer} \geqslant 3. \end{gathered} \tag{4.1.14}$$

Starr (1966a) computes numerical values of $\gamma^* = \inf_{0<\lambda<\infty} C^*(\lambda)$ for $n_0^* = 3$ and 5 and $\alpha = 0.005$ and 0.01 (where $C^*(\lambda)$ is defined analogously for rule (4.1.14)) which are close to the nominal values. Let us further modify rule (4.1.14) so that the experimenter takes a *fixed* number of additional observations say $j$ after having decided (nominally) to terminate sampling. If $j = 2$ or 4, then $\gamma^*$ is very close to $1-\alpha$ (numerical computations indicate this).

Starr (1966b) has also considered the general loss function in estimating $\mu$ by $Y_n = \sum_1^n X_i/n$ given by

$$r(Y_n; \mu) = A\,|Y_n - \mu|^r + n^t \tag{4.1.15}$$

and obtains analogous results. Starr and Woodroofe (1969) consider the special case when $t = 1$ and show that the regret, equal to the difference in the expected loss with optional stopping and the expected loss which would be incurred if $\sigma$ were known, is a bounded function of $\sigma$.

Simons (1968) has shown that one can obtain a fixed-width confidence interval for the mean of a normal distribution with unknown variance $\sigma^2$ using a procedure which overcomes ignorance of $\sigma$ with no more than a finite number of observations. So far we have been able to show that the difference between expected sample size and the optimal fixed sample size is a bounded constant. Woodroofe (1977) has further analyzed this bounded constant (for istance, see Govindarajulu (1981, Eq. (5.1.105)).

Paulson (1964, 1969) obtains confidence interval sequences for the normal mean, normal variance and difference of two normal means. For further details the reader is referred to Govindarajulu (1981, Section 5.2).

*Multi-variate case.* Khan (1968) considered point estimation of the mean vector of a $p$-variate normal distribution when the dispersion matrix is a diagonal matrix of unknown elements. Rohatgi and Rastogi (1973) have considered point estimation of a linear combination of the mean vector when the dispersion matrix is diagonal and unknown. Callahan (1966), Ghosh, Sinha and Mukhopadhyay (1976) have considered point estimation of the mean vector when the covariance is unknown (which need not be diagonal). Mallik (1971) considered point estimation of the common mean of a bivariate normal distribution when the ratio of the variances is unknown and the covariance is zero, using a quadratic loss function and linear cost function. (He uses the results of the one-armed bandit problem.) Robbins, Simons and Starr (1967) consider fixed-width confidence interval estimation of the difference of two normal means when the variances are unknown and the covariance is zero. Ghosh and Mukhopadhyay (1980) consider point estimation of the difference between two normal means under quadratic loss and linear cost.

**4.2. Estimation of the mean.** In this section we shall present the large-sample point and interval estimates for the population mean. First we shall present the results of Chow and Robbins (1965) which are fundamental to this problem.

LEMMA 4.2.1. *Let $\{Y_n\}$ be a sequence of random variables such that $Y_n > 0$ a.s. (almost surely) and $\lim_{n\to\infty} Y_n = 1$. a.s. Let $f(n)$ be any sequence of constants such that $f(n) > 0$, $\lim_{n\to\infty} f(n) = \infty$, $\lim_{n\to\infty} \{f(n)/f(n-1)\} = 1$, and for each $t > 0$ define*

$$N = N(t) = \text{smallest } k \geqslant 1 \text{ such that } Y_k \leqslant f(k)/t. \tag{4.2.1}$$

*Then, $N$ is well-defined and nondecreasing as a function of $t$,*

$$N \to \infty, \quad a.s., \qquad EN \to \infty, \qquad f(N)/t \to 1 \quad a.s. \qquad as \quad t \to \infty. \tag{4.2.2}$$

LEMMA 4.2.2. *If the assumptions of Lemma 4.2.1 are satisfied and if $E(\sup_n Y_n) < \infty$, then $Ef(N)/t \to 1$ as $t \to \infty$.*

Let $X_1, X_2, \dots$ be an i.i.d. sequence of observations from some population. We wish to set up a confidence interval having specified width $2d$ and coverage probability $1-\alpha$ for the population mean $\mu$. If the variance $\sigma^2$ is known and $d$ is small compared with $\sigma$, then for any $n \geqslant 1$ define

$$\bar{X}_n = n^{-1}\sum_1^n X_i, \qquad I_n = [\bar{X}_n - d, \bar{X}_n + d]$$

and let $u$ denote the $(1-\alpha/2)$th fractile of the standard normal distribution. Then for $n$ determined by

$$n = \text{smallest integer} \geqslant (u^2\sigma^2)/d^2, \tag{4.2.3}$$

$$\lim_{d\to 0} P(\mu \in I_n) = \lim_{d\to 0} P(\sqrt{n}|\bar{X}_n - \mu|/\sigma \leqslant d\sqrt{n}/\sigma) = \Phi(u) - \Phi(-u) = 1-\alpha.$$

However, in many practical situations, $\sigma^2$ is unknown and hence no fixed-sample size procedure is available. Let

$$V_n^2 = n^{-1}\sum_1^n (X_i - \bar{X}_n)^2 + n^{-1} \qquad (n \geqslant 1), \tag{4.2.4}$$

and $\{u_n\}$ be a sequence of positive constants tending to $u$. Define

$$N = \text{smallest } k \geqslant 1 \text{ such that } V_k^2 \leqslant (kd^2)/u_k^2. \tag{4.2.5}$$

Then we have the following main theorem of Chow and Robbins (1965).

THEOREM 4.2.1. *If $0 < \sigma < \infty$, then*
(i) $(d^2N)/u^2\sigma^2 \to 1$ *a.s. (asymptotic optimality),*
(ii) $P(\mu \in I_N) \to 1-\alpha$ *(asymptotic consistency),*
(iii) $(d^2 EN)/u^2\sigma^2 \to 1$ *(asymptotic efficiency) as $d \to 0$.*

Extensive numerical computations carried out by Ray (1957) and Starr (1966a) for the normal case indicate that when $\alpha = 0.05$, the lower bound of $P(\mu \in I_N)$ for all $d > 0$ where $N$ is the smallest odd integer $k \geqslant 3$ such that $(k-1)^{-1}\sum_1^k (X_i - \bar{X}_k)^2 \leqslant (d^2k)/u_k^2$ is about 0.929 if the values of $u_k$ are taken from the $t$-distribution with $k-1$ degrees of freedom.

Nadas (1969) extended Theorem 4.2.1 so as to take care of other specified accuracies. We say that we are estimating with

(i)

absolute accuracy if $I_n = (\mu\colon |\bar{X}_n - \mu| \leqslant d) \quad (d > 0),$ (4.2.6)

(ii)

proportional accuracy if $I_n = (\mu\colon |\bar{X}_n - \mu| \leqslant p|\mu|), \quad \mu \neq 0,\ 0 < p < 1.$ (4.2.7)

The stopping rule for the proportional accuracy is to stop at $M$ where

$$M = \min_{n \geqslant 1} \{n\colon (V_n/\bar{X}_n)^2 \leqslant n(p/u)^2\}. \tag{4.2.8}$$

Nadas (1969) obtains Theorem 4.2.1 and improves it by obtaining a uniform upper bound on $EN - n(d)$. If the d.f. of $X_1$ is continuous having a finite fourth moment then $EM - m(d)$ is also bounded above where

$$n(d) = \min_{n \geqslant 1} \{n\colon \sigma^2 \leqslant n(d/u)^2\}, \qquad m(d) = \min_{n \geqslant 1} \{n\colon \varrho^2 \leqslant n(p/u)^2\}, \qquad \varrho = \sigma/|\mu|. \tag{4.2.9}$$

*Point estimates for the mean.* With the earlier notation, assuming that $0 < \sigma < \infty$, set

$$r_n = (\bar{X}_n - \mu)^2 + cn \tag{4.2.10}$$

where $c$ is a constant which is proportional to the cost per observation. If $\sigma$ is known, then the risk $R_n(c) = Er_n$ is minimized by taking

$$n = n^0 = (1/c)^{1/2}, \qquad \text{hence} \qquad R_{n^0}(c) = 2cn^0. \tag{4.2.11}$$

If $\sigma$ is unknown, use the stopping rule: stop sampling at $N$ where

$$N = \text{smallest integer } n \geqslant n_0 \text{ such that } n \geqslant (s_n + n^{-\gamma})(1/c)^{1/2}$$

where $n_0 (\geqslant 2)$ is the initial sample size, $\gamma (> 0)$ is a specified constant and $s_n^2$ denotes the sample variance (unbiased version). After we have stopped, estimate $\mu$ by $\bar{X}_N$. Since $s_n$ converges to $\sigma$ in probability, we have $P(N < \infty) = 1$. Let

$$R(c) = E(\bar{X}_N - \mu)^2 + cEN.$$

Then Ghosh and Mukhopadhyay (1979) obtained the following result.

THEOREM 4.2.2. (a) *If* $E|X_1|^{4v} < \infty$ *for some* $v > 0$ *and* $\sigma^2 > 0$, *then as* $c \to 0$

(i) $N$ *is decreasing in* $c$; $N \to \infty$ *a.s.*,

(ii) $(N/n^0) \to 1$ *a.s.*,

(iii) *for every positive* $m < v$, $E(N/n^0)^m = 1$.

(b) *If* $E|X_1|^8 < \infty$ *and* $\sigma^2 > 0$, *then with* $0 < \gamma < 1/4$, $\{R(c)/R_{n^0}(c)\} \to 1$ *as* $c \to 0$.

**4.3. Invariant estimation of the location parameter.** Let $X$ have p.d.f. $f(x-\theta)$ and an estimate $\hat{\theta}_n$ of $\theta$ based on i.i.d. $X_1, \ldots, X_n$ is said to be translation invariant or simply regular if $\hat{\theta}_n(x_1+c, \ldots, x_n+c) = c+$ $+\hat{\theta}_n(x_1, \ldots, x_n)$. For such regular estimates, $E_\theta(\hat{\theta}_n-\theta)^2$ obviously does not depend on $\theta$ and hence there always exists an optimal estimate $\hat{\theta}_n$ called the *Pitman estimate* such that

$$E_\theta(\hat{\theta}_n-\theta)^2 = \min_{\hat{\theta}\in\mathscr{R}} E_\theta(\hat{\theta}_n-\theta)^2$$

where $\mathscr{R}$ denotes the class of regular estimates. Salyt (1969) has shown that for distributions concentrating their masses in bounded intervals, invariant sequential estimates of a translation parameter are better than (optimal) Pitman estimates based on fixed-sample sizes when the loss function is quadratic. For various special distributions, $E_\theta(\theta_\tau-\theta)^2$ turns out to be considerably smaller than $E_\theta(\hat{\theta}_n-\theta)^2$ where $\tau$ denotes the stopping time of a sequential procedure and $\hat{\theta}_n$ denotes the (optimal) Pitman estimate. In particular, for the uniform density on $(-a, a)$, Salyt (1969) shows that

$$E_\theta(\hat{\theta}_n-\theta)^2/E_\theta(\hat{\theta}_\tau-\theta)^2 \to 3 \qquad \text{as} \quad n\to\infty\,.$$

Taking the loss function to be $r(t_\tau; \theta) = (t_\tau-\theta)^2+c\tau$, $c > 0$, where $t_\tau$ denotes the estimate of $\theta$ based on stopping time $\tau$, Salyt (1970) obtains the optimal sequential estimation procedure for the location parameter $\theta$. He also establishes a certain integral relation between the risk and a function of the expected sample size for the optimal estimate. He applies these results to the uniform density on $(-1/2+\theta, 1/2+\theta)$ and shows the optimality of the sequential procedure.

Linnik and Romanovski (1970a) consider sequential unbiased estimation of the location parameter $\theta$, the sequential plan having expected stopping time bounded by $n$. They show that if the Ibragimov–Has'minskii (1970) conditions on the absence of discontinuities in the information function are satisfied, sequential estimators can yield only an infinitesimal small relative gain in the mean square deviation when compared with the fixed-sample size procedure. However, the presence of discontinuities in the information can drastically change the situation and sequential procedures may be beneficial. Linnik and Romanovski (1970b) survey the results on sequential estimation procedures that are published in Soviet journals.

Under certain regularity assumptions, Klimov (1972) has shown that the optimal sequential invariant estimation procedure for $\theta$ is based on a fixed number of observations $n$ where $n$ can be determined by minimizing the risk function which does not depend on the unknown parameter $\theta$.

The standard hypernormal density is defined by

$$f(x) = C\exp\{Ae^{\gamma x}+Bx\}, \qquad x\in R^1$$

where $A$, $\gamma$ are real, $A$, $\gamma \neq 0$ and $C$ is a normalizing constant. The normal and the exponential (on $(-\infty, a)$ and on $(a, \infty)$) densities can be obtained as limiting cases of the hypernormal density function. Kagan, Linnik, Romanovski and Rukhin (1971) prove that an invariant sequential estimation procedure for the location parameter of a hypernormal distribution, with respect to a loss function, coincides with a fixed sample size procedure. If an additional restriction such as $E_\theta g(\tau) \leqslant N$ is imposed on the sequential procedure, then the optimal stopping is a randomized rule between two fixed integers. In other words, from the sequential estimation point of view the hypernormal d.f. 's are "self-controlled" in the sense that preceding observations are not needed for the decision to stop at a given moment or to continue sampling.

Ibragimov and Has'minskii (1972) propose two sequential stopping rules for invariantly estimating the location parameter. Let the density $f(x)$ have a finite but positive number of discontinuities of the second kind (jumps) located at $x_1, \ldots, x_r$, $r \geqslant 1$, and let $p_i = f(x_i+0)$ and $q_i = f(x_i-0)$. If for any $\gamma > 0$,

$$E_0|f'(x)/f(x)|^{1+\gamma} < \infty \qquad \text{and either} \qquad \sum_1^r p_i q_i \neq 0$$

or not all the differences $p_i - q_i$ are of the same sign, then Ibragimov and Has'minskii (1972) show that the sequential procedure is more efficient than the fixed sample Pitman estimate, with respect to the power loss functions. By confining to power loss functions and assuming that there exists an $\varepsilon > 0$ such that $E_0|X|^\varepsilon < \infty$ and that the density function $f(x)$ has bounded variation and a finite but positive number of jumps, Ibragimov and Has'minskii (1974b) show that the sequential procedure is more advantageous than a fixed sample size procedure. On the other hand, if $\sum p_i q_i = 0$ and the differences $p_i - q_i$ have the same sign, then asymptotically the sequential procedure is equivalent to a fixed sample size procedure. It should be noted that Ibragimov and Has'minskii (1972, 1974b) confine to sequential rules having expected stopping time bounded by $n$. If the density is uniform on $(\theta - 1/2, \theta + 1/2)$ then their stopping rule I coincides with Wald's (1971) sequential procedure (see Section 3.6) and Salyt's (1969) procedure.

**4.4. Fixed-width confidence intervals for an arbitrary parameter.** In this section, we shall be concerned with fixed-width confidence intervals for an unknown parmeter in the presence of nuisance parameters which was considered by Khan (1969). Let $p(x; \theta_1, \theta_2)$ denote the density function of a random variable $X$ indexed by real valued parameters $\theta_1$ and $\theta_2$, where $\theta_2$ is considered to be the nuisance parameter. For $\theta_1$, we wish to set up a confidence interval of width $2d$ and coverage probability $1-\alpha$. Let $I(\theta) = ((I_{ij}))$, $i, j = 1, 2$ where $\theta = (\theta_1, \theta_2)'$ and

$$I_{ij} = E\{[\partial \ln p/\partial\theta_i][\partial \ln p/\partial\theta_j]\}, \qquad 1 \leqslant i, j \leqslant 2. \tag{4.4.1}$$

Assume that $I(\theta)$ is positive definite and let $((\lambda_{ij})) = I^{-1}(\theta)$. Also let $\hat{\theta}_1(n)$ and $\hat{\theta}_2(n)$ denote the maximum likelihood estimates of $\theta_1$ and $\theta_2$ respectively. It is well known that $\hat{\theta}_1(n)$ is asymptotically normal with mean $\theta_1$ and variance $\lambda_{11}/n$. Let $\{u_n\}$ be a sequence of positive constants converging to $u$, the $(1-\alpha/2)$th fractile of the standard normal d.f. If $\theta_1$ and $\theta_2$ are known the fixed-width C.I. is

$$J_n = [\hat{\theta}_1(n)-d,\ \hat{\theta}_1(n)+d] \text{ and the optimal sample size is}$$
$$n_d = \text{smallest } n \geqslant u^2\lambda_{11}/d^2 = n^0 \text{ (say)}. \tag{4.4.2}$$

Since typically $\lambda_{11} = \lambda_{11}(\theta_1, \theta_2)$ and $\theta_1$ and $\theta_2$ are unknown, we use the following stopping rule: starting with $n \geqslant m$, stop at $N$ where

$$N = \inf\{n \geqslant m\colon\ n \geqslant u_n^2\hat{\lambda}_{11}(n)/d^2\} \qquad \text{where} \qquad \hat{\lambda}_{11}(n) = \lambda_{11}(\hat{\theta}_1(n), \hat{\theta}_2(n)). \tag{4.4.3}$$

Using the strong consistency properties of the maximum likelihood estimates, one can show that $P(N < \infty) = 1$. Khan (1969) obtained the following result.

THEOREM 4.4.1. *If the maximum likelihood estimates are strongly consistent* (see, for instance Wald (1949) for regularity assumptions, or Govindarajulu (1981, Theorem 4.11.3)) *and if*

$$E(\sup\hat{\lambda}_{11}(n)) < \infty \tag{4.4.4}$$

*then*

(i) $N/n^0 \to 1$ *a.s.*,
(ii) $P(\theta \in J_N) \to 1-\alpha$ *and*
(iii) $EN/n^0 \to 1$ *as* $d \to 0$.

For the normal case, in estimating the mean or the variance, Khan (1969) shows that (4.4.4) is satisfied by using a result of Wiener (1939).

**4.5. Fixed precision estimates.** Recently Zieliński and his colleagues have considered fixed precision sequential estimates for parameters. Let $\{\Omega, \mathscr{F}, (p_\theta\colon \theta \in \Theta)\}$ denote the statistical structure and $g$ be a mapping from $\Theta$ into a metric space $\mathscr{X}$ with $\|x-y\|$ denoting the distance between $x$ and $y$ belonging to $\mathscr{X}$. Let $\{X_t\}$, $t = 1, 2, \ldots$ be a sequence of $\mathscr{X}$ valued random elements on $(\Omega, \mathscr{F})$ which is assumed to converge to $g(\theta)$ whenever the distribution of $\{X_t\}$ is generated by $P_\theta$. Given $\varepsilon > 0$ and $\gamma \in (0, 1)$, the problem of finding fixed precision estimate for $g(\theta)$ is in finding a stopping variable $\tau\colon \Omega \to \{1, 2, \ldots\}$ such that $P_\theta\{\|X_\tau - g(\theta)\| < \varepsilon\} \geqslant 1-\gamma$ for all $\theta \in \Theta$. Assuming $k$ independent copies $\{X_t^{(i)}\}$, $i = 1, 2, \ldots, k$ of $\{X_t\}$ Zieliński (1977, 1978a) proves the existence of fixed precision estimates, provides a method of constructing such estimates and applies this method to some examples.

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables distributed in the interval $[0, \theta]$ where $\theta$ is unknown. Sierociński (1980) considered fixed precision estimation of $\theta$. As a particular case he solves the problem of estimating $\theta$, when the $X_i$ are uniform on $(0, \theta)$.

Following Zieliński's (1978b) approach and assuming that there exists $k (k \geqslant 4)$ independent copies of the sequence $\{X_i^{(j)}\}$, $j = 1, 2, \ldots, k$, (where for each fixed $j$, the $X_i^{(j)}$ may be dependent). Gołdys (1981) obtains a fixed precision sequential procedure for estimating a normal mean.

### 4.6. Estimation in linear regression

*Fixed-width confidence bounds.* Let $y_1, y_2, \ldots$ be a sequence of independent observations with

$$y_i = \beta' x^{(i)} + \varepsilon_i \tag{4.6.1}$$

where $\beta'$ is an unknown $1 \times p$ vector, $x^{(i)}$ a known $p \times 1$ column vector, and $\varepsilon_i$ a random variable having an unknown distribution $F$ with mean 0 and finite but unknown variance $\sigma^2$. We wish to find a region $W$ in $p$-dimensional Euclidean space such that $P(\beta \in W) = 1 - \alpha$ and such that the length of the interval cut-off on the $\beta_i$-axis by $W$ has width not exceeding $2d$, $i = 1, \ldots, p$. If $\sigma$ is known the usual practice is to construct the confidence region

$$\{\hat{\beta}(n) - \beta\}' (X_n X_n') \{\hat{\beta}(n) - \beta\} \leqslant d^2 \tag{4.6.2}$$

where

$$\hat{\beta}(n) = (X_n X_n')^{-1} X_n Y_n \tag{4.6.3}$$

$Y_n = (y_1, \ldots, y_n)$, $X_n = (x^{(1)}, \ldots, x^{(n)})$ a $p \times n$ $(p \leqslant n)$ matrix and where $X_n$ is assumed to be of full rank.

If $\sigma$ is unknown the least squares estimate of $\sigma^2$ is

$$\hat{\sigma}^2(n) = n^{-1} Y_n' (I_n - X_n' (X_n X_n')^{-1} X_n) Y_n \tag{4.6.4}$$

Gleser (1965) obtains the following results.

THEOREM 4.6.1. *If*

(i)
$$X_n / n^{1/2} \to 0 \qquad a.s. \tag{4.6.5}$$

and

(ii)
$$n^{-1} (X_n X_n') \to \Sigma \tag{4.6.6}$$

*where $\Sigma$ is a $p \times p$ positive definite matrix, then*

$$P\{n(\hat{\beta}(n) - \beta)' (\hat{\beta}(n) - \beta) \leqslant d^2\} = P(T(\lambda_1, \ldots, \lambda_p)) \leqslant d^2/\sigma^2\} \tag{4.6.7}$$

*where $\lambda_1, \ldots, \lambda_p$ are the characteristic roots of $\Sigma^{-1}$ and $T(\lambda_1, \ldots, \lambda_p)$ has the distributions of a weighted sum of p independent chi-square variables having one degree of freedom, the $\lambda_i$'s being the weights.*

*Sequential procedure.* Let $\{v_n\}$ be any sequence of constants converging to the number $v$ satisfying

$$P(T(\lambda_1, \ldots, \lambda_p) \leqslant v) = \alpha. \tag{4.6.8}$$

Initially observe $y_1, \ldots, y_{n_0}$ ($n_0 \geqslant p$). Then sample term by term, stopping according to the variable $N$ defined by

$$N = \text{smallest } k \geqslant n_0 \text{ such that } k^{-1}(\hat{\sigma}^2(k)+k^{-1}) \leqslant d^2/v_k. \tag{4.6.9}$$

When sampling is stopped at $N = n$, construct the region $W_n$ given by

$$W_n = \{Z\colon (Z-\hat{\beta}(n))'(Z-\hat{\beta}(n)) \leqslant d^2\}. \tag{4.6.10}$$

Gleser (1965) shows that the preceding procedure is asymptotically consistent and efficient as $d \to 0$.

*Remark* 4.6.1. The addition of $n^{-1}$ to $\sigma^2(n)$ in (4.6.9) is unnecessary if $F$ is continuous. Gleser's (1965) assumptions (4.6.5) and (4.6.6) are found to be strong and they are weakened by Albert (1966) and Srivastava (1971). Also the latter authors obtain spherical confidence regions for the regression parameters. For further details see Govindarajulu (1981, pp. 500–502).

*Point estimation.* Mukhopadhyay (1947) considered point estimation of regression parameters with quadratic loss and linear cost of sampling.

**4.7. Fixed-width confidence intervals for the coefficient of variation.** In order to compare the variability in several distributions described in different units, it is desirable to have a measure of relative variation. Thus the need for the coefficient of variation (C.V) which expresses the standard deviation as a percentage of the mean. Govindarajulu (1976) has given a fixed width sequential procedure for the coefficient of variation.

Let $X_1, X_2, \ldots$ be an i.i.d. sequence of normal variables having unknown mean $\mu$ and unknown variance $\sigma^2$. Suppose we wish to estimate $c = \sigma/|\mu|$ with a confidence interval having a specified length $2d$ and specified coefficient $1-\alpha$ (asymptotically as $d \to 0$). Let $\bar{X}_n = (X_1 + \ldots + X_n)/n$, $s_n^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$ and $c_n = s_n/|\bar{X}_n|$. Also let $a = (1+\mu^2/2\sigma^2)^{1/2}$ and $a_n = (1+\bar{X}_n^2/2s_n^2)^{1/2}$.

Let $\{u_k\}$ be a sequence of positive constants tending to $u$ as $k \to \infty$ where $u$ denotes the $(1-\alpha/2)$th fractile of the standard normal d.f. Then define the stopping variable $N$ as

$$N = \text{smallest } n > 2 \text{ for which } n \geqslant u_n^2\, \hat{c}_n^4\, \hat{a}_n^2/d^2 \tag{4.7.1}$$

and give the confidence interval $I_N = (\hat{c}_N - d, \hat{c}_N + d)$ with $\hat{c}_N = s_N/|\bar{X}_N|$. Then it is shown that the above procedure has asymptotic properties of optimality, consistency and efficiency (see Govindarajulu 1981, Section 5.6).

Govindarajulu (1981, Section 5.6) also considers sequential estimation with proportional closeness criterion and sequential point estimation.

*Remark* 4.7.1. Assumption of normality is not necessary provided fourth moments are finite. The machinery set up for the above problem can easily be adapted in order to set up fixed-width confidence intervals for $p = P(X < Y)$ which plays an important role in reliability theory.

**4.8. Asymptotically optimal Bayes estimation.** In Section 3.7 we have considered certain Bayes sequential procedures that are based on fixed sample sizes. Here we shall consider Bayes estimates that are asymptotically optimal (AO). Every sequential procedure $S$ prescribes a stopping rule $N$ and an estimator $d$. Also, the posterior distribution of the parameter $\theta$, given $\{N = n\}$ is independent of the stopping rule; consequently, the estimator $d$ is independent of the stopping rule $N$ for any specified prior distribution $G(\theta)$ and the loss function $r(d; \theta)$. Let $r_1(S; \theta)$ denote the loss due to erroneous estimation and $cN(S, G)$ denote the cost due to stopping according to the rule $N$ where $c$ denotes cost per observation. If $EN(S, G) < \infty$, then the component of the risk due to cost will go to zero as $c \to 0$. Thus, the stopping rule will, in general, require a very large sample as $c \to 0$. Kiefer and Sacks (1963) define the asymptotic optimality (AO) of a stopping rule and Bickel and Yahav (1967) define the asymptotic pointwise optimality (APO) of a stopping rule. Bickel and Yahav (1967, Theorem 3.1) show that under a mild condition for constant cost of observation, APO implies AO. Yahav and Bickel (1968) give some APO rules and study their properties. Bickel and Yahav (1969) show that when their result is specialized to the normal $(\mu, 1)$ case where $\mu$ has normal prior with mean $\mu_0$ and variance $\sigma^2$, the Bayes rule is to take a fixed-sample of size $N(c)$ which is the positive integer closest to $(c^{1/2}\sigma^{-1} - \sigma^{-2})$. Gleser and Kunte (1976) develop asymptotically optimal sequential Bayes interval estimation procedures when the loss is a linear combination of the length of the interval, the indicator function for noncoverage, and the sample size.

*Bayesian estimation of the binomial parameter.* Cabilio (1977), motivated by clinical trials in which $p$ being close to zero or unity will lead to a dramatic decision, considers the loss function given by

$$L(\delta_n; n, p) = \{(\delta_n - p)/pq\}^2 + nc$$

where $\delta_n$ denotes an estimate of $p$, $q = 1-p$ and $c$ is the cost per observation. Then the sequential rule is given by

$$N = \text{first } n \geqslant 1 \text{ such that } n^2(1-\hat{\delta}_n)\hat{\delta}_n \geqslant 1/c.$$

Cabilio (1977) obtains the asymptotic normality, asymptotic efficiency and a.s. convergence properties of the sequential procedure. Since for fixed $c$ the procedure performs badly when $p$ is near 0 or 1, a uniform prior on $p$ is assumed and the optimal Bayes procedure is shown to exist having a bounded sample size.

Shapiro and Waldrop (1981) consider Bayes sequential estimation for one parameter exponential family and natural conjugate prior using loss related to the Fisher information and linear cost of sampling. Tractable expressions for the Bayes estimator and the posterior expected loss are given.

**4.9. Confidence sequences for certain parameters.** Darling and Robbins (1967b, 1967c) give confidence sequences for the normal mean when variance is known and for the population median. Robbins (1970) obtains the same using elementary arguments. Darling and Robbins (1967a) obtain a confidence sequence for the normal mean when the variance is unknown. (For references on Darling and Robbins (1967a–1967c) see Robbins (1970)). The advantage of the confidence sequence $I_n$ (having confidence coefficient $1-\alpha$) over a fixed-sample size confidence interval is that it enables us to pursue the unknown parameter $\theta$ throughout the entire sequence of observations $X_1$, $X_2, \dots$ with an interval $I_n$ the width of which shrinks to zero as the sample size $n$ increases, in such a way that with probability $\geqslant 1-\alpha$ the interval $I_n$ includes the unknown parameter at every stage. Lombard (1977) derives a confidence sequence for the regression coefficient in a linear regression model. By making use of generalized likelihood ratio martingales, Lai (1976) has constructed confidence sequences for the unknown parameters of binomial, Poisson, uniform, gamma and other distributions. The problem of nuisance parameters is also considered.

**4.10. Nonparametric confidence intervals.** Farrel (1966) considers interval estimation for a quantile of the population. Geertsema (1970), using the methods of Section 4.3, constructs nonparametric fixed-width confidence intervals for certain parameters, especially the location of symmetry. In particular, he studies procedures based on the sign statistic and Wilcoxon statistic. He shows that the asymptotic efficiencies of these sequential procedures compared to the one based on the $t$-statistic coincide with the well-known Pitman efficiencies. Sen and Ghosh (1971) consider fixed-width confidence interval estimation of location parameter based on the one-sample rank order statistic. Govindarajulu (1974) considered fixed width confidence interval for $P(X < Y)$. Weiss and Wolfowitz (1972) give an optimal fixed-length confidence limit for the location parameter of an unknown distribution. For further details, see Govindarajulu (1981, Section 5.11).

## References

[1] S. A. Aivazyan, *Distinguishing neighboring hypotheses about frequency distribution in the scheme of the generalized probability ratio tests*, Theor. Probab. Appl. **10** (1965), 646–658.

[2] A. Albert, *Fixed size confidence ellipsoids for linear regression parameters*, Ann. Math. Statist. **37** (1966), 1602–1630.

[3] T. W. Anderson, *A modification of the sequential probability ratio test to reduce the sample size*, Ann. Math. Stat. **31** (1960), 165–197.

[4] T. W. Anderson and M. Friedman, *A limitation of the optimum property of the sequential probability ratio test*, in: Olkin et al (ed), *Contributions to Probability and Statistics. Essays in Honor of Harold Hotelling*, Stanford University Press, (1960), 57–59.

[5] F. J. Anscombe, *Large sample theory of sequential estimation*, Biometrika **36** (1949), 455–458.

[6] –, *Large sample theory of sequential estimation*, Proc. Cambridge Philos. Soc. **48** (1952), 600–607.

[7] –, *Sequential estimation*, J. Roy. Statist. Soc. Ser B **15**(1953), 1–29.

[8] –, *Fixed-sample size analysis of sequential observations*, Biometrics **10**(1954) 137–144. Correction ibid. **13**(1957), 543.

[9] P. Armitage, *Some sequential tests of Student's hypothesis*, J. Roy. Statist. Soc. Suppl. **9**(1947), 250–263.

[10] –, *Sequential analysis with more than two alternative hypotheses, and its relation to discriminant function analysis*, J. Roy. Statist. Soc. Ser B **12**(1950), 137–144.

[11] –, *Restricted sequential procedures*, Biometrika **44**(1957), 9–26.

[12] L. A. Aroian, *Sequential analysis-direct method*, Technometrics **10**(1968), 125–132.

[13] K. J. Arrow, D. Blackwell and M. A. Girshick, *Bayes and minimax solutions of sequential decision problems*, Econometrica **17**(1949), 213–244.

[14] R. R. Bahadur, *Sufficiency and statistical decision functions*, Ann. Math. Statist. **25**(1954), 423–462.

[15] –, *A note on the fundamental identity of sequential analysis*, Ann. Math. Statist. **29**(1958), 534–543.

[16] G. A. Barnard, *Sampling inspection and statistical decisions*, J. Roy. Statist. Soc. Ser. B **16**(1954), 151–174.

[17] M. S. Bartlett, *The large sample theory of sequential tests*, Proc. Cambridge Philos. Soc. **42**(1946), 239–244.

[18] R. Bechhofer, *A note on the limiting relative efficiency of the Wald sequential probability test*, J. Amer. Statist. Soc. Suppl. **55**(1960), 660–663.

[19] R. H. Berk *Strong consistency of certain sequential estimators*, Ann. Math. Statist. **40**(1969), 1492–1495; see also correction: Ann. Math. Statist. **42**(1971), 1135–1137.

[20] –, *Stopping time of SPRTS based on exchangeable models*, Ann. Math. Statist. **41**(1970), 979–990.

[21] –, *Some asymptotic aspects of sequential analysis*, Ann. Statist. **1**(1973), 1126–1138.

[22] –, *Locally most powerful sequential tests*, Ann. Statist. **3**(1975), 373–381.

[23] R. H. Berk and I. R. Savage, *The information in a rank-order and stopping time of some associated SPRTs*, Ann. Math. Statist. **39**(1968), 1661–1674.

[24] R. H. Berk, L. D. Brown and A. Cohen, *Bounded stopping times for a class of sequential Bayes tests*, Ann. Statist. **9**(1981), 834–845.

[25] D. H. Bhate, *Some Properties of Sequential Tests*, Doctoral Thesis, University of London, London 1955.

[26] A. Bhattacharyya, *On some analogues of the amount of information and their use in statistical estimation*, Sankhyā Ser. A **8**(1946), 1–14.

[27] P. J. Bickel and J. A. Yahav, *Asymptotically pointwise optimal procedures in sequential analysis*, Proc. Fifth Berkeley Symp. Math. Statist. Prob., University of California Press **1**(1967), 401–413.

[28] –, –, *On an A.P.O. rule in sequential estimation with quadratic loss*, Ann. Math. Statist. **40**(1969), 417–426.

[29] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, Wiley, New York 1954, 254–255; Reprinted by Dover Publications, Inc. (1980).

[30] G. Blom, *A generalization of Wald's fundamental identity*, Ann. Math. Statist. **20**(1949), 439–444.

[31] S. Blumenthal and Z. Govindarajulu, *Robustness of Stein's two-stage procedure for mixtures of normal populations*, J. Amer. Statist. Assoc. **72**(1977), 192–196.

[32] R. A. Bradley, D. C. Martin and F. Wilcoxon, *Sequential rank tests I. Monte Carlo studies of the two-sample procedure*, Technometrics 7(1965), 463–483.

[33] R. A. Bradley, S. D. Merchant and F. Wilcoxon, *Sequential rank tests II; modified two-sample procedure*, Technometrics **8**(1966), 615–623.

[34] L. D. Brown and A. Cohen, *Inadmissibility of large classes of sequential tests*, Ann. Statist. **9**(1981), 1239–1247.

[35] D. L. Burkholder and R. A. Wijsman, *Optimum properties and admissibility of sequential tests*, Ann. Math. Statist. **34**(1963), 1–18.

[36] P. Cabilio, *Sequential estimation in Bernoulli trials*, Ann. Statist. **5**(1977), 342–356.

[37] J. D. Callahan, *On sequential estimation procedures for the mean vector of a multivariate normal distribution*, Tech. Rep. No. 132, Dept. of Statist., John Hopkins University (1969).

[38] K. C. Chanda, *Asymptotic distribution of the sample size for a sequential probability ratio test*, J. Amer. Statist. Assoc. **66**(1971), 178–183.

[39] H. Chernoff, *A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations*, Ann. Math. Statist. **23**(1952), 493–507.

[40] –, *Optimal stochastic control*, Sankhyā Ser. A **30**(1968), 221–252.

[41] Y. S. Chow and H. Robbins, *On the asymptotic theory of fixed-width sequential confidence intervals for the mean*, Ann. Math. Statist. **36**(1965), 457–462.

[42] D. R. Cox. *Sequential tests for composite hypotheses*, Proc. Cambridge Philos. Soc. **48**(1952a), 290–299.

[43] –, *Estimation by double sampling*, Biometrika **39**(1952b), 217–227.

[44] –, *A note on the sequential estimation of means*, Proc. Cambridge Philos. Soc. **48**(1952c), 447–450.

[45] –, *Large sample sequential tests for composite hypotheses*, Sankhyā Ser. A **25**(1963), 5–12.

[46] G. B. Dantzig, *On the nonexistence of tests of "Students" hypothesis having power functions independent of $\sigma$*, Ann. Math. Statist. **11**(1940), 186–192.

[47] D. A. Darling, *The birth, growth and blossoming of sequential analysis*, in: D. B. Owen (ed.), *On the History of Statistics and Probability*, Marcel Dekker, New York and Basel 1976, 369–375.

[48] D. A. Darling and H. Robbins, *Some further remarks on inequalities for sample sums*, Proc. Nat. Acad. Sci. U.S.A. **60** (1968a), 1175–1182.

[49] –, –, *Some nonparametric sequential tests with power one*, Proc. Nat. Acad. Sci U.S.A. **61** (1968b), 804–809.

[50] H. F. Dodge and H. G. Romig, *A method of sampling inspection*, Bell, Syst. Tech. J. **8** (1929), 613–631.

[51] T. G. Donnelly, *A Family of Truncated Sequential Tests*, Doctoral Dissertation, University of North Carolina (1957).

[52] S. Yu. Efroimovich, *On sequential estimation under the condition of local asymptotic normality*, Theor. Probab. Appl. **25** (1980), 27–40.

[53] K. M. Eger, *A direct method of the computation of the OC and of the moments of the sample number for SPRTs in the case of discrete random variables*, Math. Operationsforsch. Statist. Ser. Statist. **11** (1980), 499–514.

[54] B. Eisenberg, B. K. Ghosh and G. Simons, *Properties of generalized sequential probability ratio tests*, Ann. Statist. **4** (1976), 237–251.

[55] V. Fabian, *Note on Anderson's sequential procedures*, Ann. Statist. **2** (1974), 170–176.

[56] R. H. Farrel, *Bounded length confidence intervals for the p-point of a distribution function, III*, Ann. Math. Statist. **37** (1966), 586–592.

[57] T. S. Ferguson, *Mathematical Statistics: A Decision Theoretic Approach*, Academic Press, New York 1967, 378–383.

[58] J. C. Geertsema, *Sequential confidence intervals based on rank tests*, Ann. Math. Statist. **41** (1970), 1016–1026.

[59] J. K. Ghosh, *On some properties of sequential t-test*, Calcutta Statist. Assoc. Bull. **9** (1960), 77–86.

[60] M. N. Ghosh, *Bounds for the expected sample size in a sequential probability ratio test*, J. Roy. Statist. Soc. Ser. B **22** (1960), 360–367.

[61] M. Ghosh and N. Mukhopadhyay, *Sequential point estimation of the mean when the distribution is unspecified*, Comm. Statist. A – Theory Methods **8** (1979), 637–652.

[62] –, –, *Sequential point estimation of the difference of two normal means*, Ann. Statist. **8** (1980), 221–225.

[63] M. Ghosh, B. K. Sinha and N. Mukhopadhyay, *Multi-variate sequential point estimation*, J. Multivariate Anal. **6** (1976), 281–294.

[64] S. G. Ghurye and H. Robbins, *Two-stage procedures for estimating the difference between means*, Biometrika **41** (1954), 146–152.

[65] L. J. Gleser, *On the asymptotic theory of fixed-size sequential confidence bounds for linear regression parameters*, Ann. Math. Statist. **36** (1965), 463–467. (Correction note Ann. Math. Statist. **37** (1966), 1053–1055).

[66] L. J. Gleser and S. Kunte, *On asymptotically optimal sequential Bayes interval estimation procedures*, Ann. Statist. **4** (1976), 685–711.

[67] B. Goldys, *Fixed precision estimation of a normal mean*, Unpublished manuscript (1981).

[68] Z. Govindarajulu, *Asymptotic normality of two-sample rank order sequential probability ratio test based on Lehmann alternatives*, Unpublished manuscript (1968).

[69] –, *Fixed-width confidence intervals for $P(X < Y)$*, in: F. Proschan and R. J. Serfling (eds.) *Reliability and Biometry, Statistical Analysis of Life Length* SIAM, Philadelphia 1974, 747–757.

[70] –, *Fixed-width confidence intervals for the coefficient of variation*, Bull. Inst. Math. Acad. Sinica **4** (1976), 273–279. Correction; ibid. **7**, 231–234.

[71] –, *Stopping time of a c-sample rank order SPRT.* Trans. Seventh Prague Conference, Academia Publishing House of the Czech. Acad. of Sciences, Vol. A, 1977, 163–174.

[72] –, *The Sequential Statistical Analysis of Hypothesis Testing, Point and Interval Estimation, and Decision Theory*, American Sciences Press, Columbus, Ohio 1981.

[73] R. A. Groeneveld, *A note on the sequential sign test*, Amer. Statist. **25** (1971), 15–16.

[74] W. J. Hall, *Some sequential analogs of Stein's two-stage test*, Biometrika **49** (1962), 367–378.

[75] W. J. Hall, R. A. Wijsman and J. K. Ghosh, *The relationship between sufficiency and invariance with applications in sequential analysis*, Ann. Math. Statist. **36** (1965), 575–614.

[76] J. L. Hodges, Jr. and E. L. Lehmann, *Some applications of the Cramér-Rao inequality*. Proc. Second Berkeley Symp. Math. Statist. Prob., University of California Press **1** (1951), 13–22.

[77] W. Hoeffding, *A lower bound for the average sample number of a sequential test*, Ann. Math. Statist. **24** (1953), 127–130.

[78] –, *Lower bounds for the expected sample size and the average risk of a sequential procedure*, Ann. Math. Statist. **31** (1960), 352–368.

[79] I. A. Ibragimov and R. Z. Has'minskii, *Asymptotic behavior of generalized Bayes estimates*, Soviet Math. Dokl. **11** (1970), 1181–1185.

[80] –, –, *Information inequalities and super efficient estimates*, Dokl. Akad. Nauk SSSR **204** (1972), 1300–1302.

[81] —, —, *On sequential estimation*, Theor. Probab. Appl. **19** (1974a), 233–244.

[82] —, —, *On sequential estimation of the location parameter for families of distributions with discontinuous densities*, Theor. Probab. Appl **19** (1974b), 669–682.

[83] A. F. Ifram, *On the asymptotic behavior of densities with applications to sequential analysis*, Ann. Math. Statist. **36** (1965), 615–637.

[84] N. L. Johnson, *A proof of Wald's theorem on cumulative sums*, Ann. Math. Statist. **30** (1959), 1245–1247.

[85] J. Jurečkova, Personal Communication (1981).

[86] A. M. Kagan, Yu. V. Linnik, I. V. Romanovski and A. L. Rukhin, *Families with "self control"*, Soviet Math. Dokl. **12** (1971), 1167–1171.

[87] K. W. Kemp, *Characteristic and the average sample number of some sequential tests/AE for calculating the operating*, J. Roy. Statist. Soc. Ser. B **20** (1958), 379–386.

[88] J. H. B. Kemperman, *The Passage Problem for a Stationary Markov Chain*, University of Chicago Press, Chicago, Illinois, 1961.

[89] R. A. Khan, *Sequential estimation of a mean vector of a multivariate normal distribution*, Sankhyā Ser. A **30** (1968), 331–334.

[90] —, *A general method of determining fixed-width confidence intervals*, Ann. Math. Statist. **40** (1969), 704–709.

[91] J. Kiefer and J. Sacks, *Asymptotically optimum sequential inference and design*, Ann. Math. Statist. **34** (1963), 705–750.

[92] J. Kiefer and L. Weiss, *Some properties of generalized sequential probability ratio tests*. Ann. Math. Statist. **28** (1957), 57–74.

[93] G. P. Klimov, *Invariant decision rules in sequential analysis*, Soviet Math. Dokl. **13** (1972), 268–271.

[94] N. V. Krylov and T. P. Miroshnichenko, *On extending the optimality region of Wald's test*, Theor. Prob. Appl **25** (1980), 165–169.

[95] T. L. Lai, *On confidence sequences*, Ann. Statist. **4** (1976), 265–280.

[96] —, *Pitman efficiencies of sequential tests and uniform limit theorems in nonparametric statistics*, Ann. Statist. **6** (1978), 1027–1047.

[97] T. L. Lai and D. Siegmund, *A nonlinear renewal theory with applications to sequential analysis I*, Ann. Statist. **5** (1977), 946–954.

[98] —, —, *A nonlinear renewal theory with applications to sequential analysis II*, Ann. Statist. **7** (1979), 60–76.

[99] W. D. Lawing and H. T. David *Likelihood ratio computation of operating characteristic*, Ann. Math. Statist. **37** (1966), 1704–1716.

[100] E. L. Lehmann, *The power of rank tests*, Ann. Math. Statist. **24** (1953), 23–43.

[101] —, *Testing Statistical Hypotheses*, Wiley, New York 1959, 104–110.

[102] E. L. Lehmann and C. Stein, *Completeness in the sequential case*, Ann. Math. Statist. **21** (1950), 376–385.

[103] Yu. V. Linnik and I. V. Romanovski, *On the theory of sequential estimation*, Soviet Math. Dokl **11** (1970a) 1196–1198.

[104] —, —, *Some new results in sequential estimation theory*, in: *Proc. Sixth Berkeley Symp. Math. Statist. and Probability*, University of California Press at Berkeley and Los Angeles (1970b), 85–96.

[105] F. Lombard, *Truncated sequential tests based on one-sample rank order statistics*, South African Statist. J. **10** (1976), 177–185.

[106] —, *Sequential procedures based on Kendall's tau statistic*, South African Statist. J. **11** (1977), 79–87.

[107] G. Lorden, *Open-ended tests for Koopman-Darmois families*, Ann. Statist. **1** (1973), 633–643.

[108] R. M. Loynes, *The consistency of certain sequential estimators*. Ann. Math. Statist. **40** (1969), 568–574.

[109] A. K. Mallik, *Sequential estimation of the common mean of two normal populations*, Tech. Rep. No. 42, Dept. of Statistics, Stanford University, Stanford 1971.

[110] T. K. Matthes, *On the optimality of sequential probability ratio tests*, Ann. Math. Statist. **34** (1963), 18–21.

[111] R. G. Miller, *Sequential signed rank test*, J. Amer. Statist. Assoc. **65** (1970), 1554–1561.

[112] J Moshman, *A method for selecting the size of the initial sample in Stein's twosample procedure*, Ann. Math. Statist. **29** (1958), 667–671.

[113] N. Mukhopadhyay, *Sequential estimation of regression parameters in Gauss-Markoff set-up*. J. Ind. Statist. Assoc. **12** (1974), 39–43.

[114] E. S. Page, *An improvement to Wald's approximation for some properties of sequential tests*, J. Roy Statist. Soc. Ser. B 16 (1954), 136–139.

[115] E. A. Parent, Jr., *Sequential Ranking Procedures*, Tech. Rep. No. 80, Department of Statistics, Stanford University, Stanford 1965.

[116] E. Paulson, *A note on the efficiency of the Wald sequential test*, Ann. Math. Statist. **18** (1947), 447–450.

[117] –, *A sequential procedure for choosing one of k hypotheses concerning the unknown mean of a normal distribution*, Ann. Math. Statist. **34** (1963), 549–554.

[118] –, *Sequential estimation and closed sequential decision procedures*, Ann. Math. Statist. **35** (1964), 1048–1058.

[119] –, *Sequential interval estimation for the means of normal populations*, Ann. Math. Statist. **40** (1969), 509–516.

[120] M. Pollak and D. Siegmund, *Approximations to the expected sample size of certain sequential tests*, Ann. Statist. **3** (1975), 1267–1282.

[121] W. D. A. Ray, *Sequential confidence intervals for the mean of a normal population with unknown variance*, J. Roy. Statist. Soc. Ser. B **19** (1957), 133–143.

[122] A. Renyi, *On the asymptotic distribution of the sum of random number of random variables*, Acta. Math. Acad. Sci. Hungar. **8** (1957), 193–199.

[123] M. R. Reynolds, Jr., *A sequential signed-rank test for symmetry*, Ann. Statist **3** (1975), 382–400.

[124] D. Richter, *Two stage experiments for estimating a common mean*, Ann. Math. Statist. **31** (1960), 1164–1173.

[125] H. Robbins, *Statistical methods related to the law of the iterated logarithm*, Ann. Math. Statist. **41** (1970), 1397–1409.

[126] H. Robbins, G. Simons and N. Starr, *A sequential analogue of the Behrens-Fisher problem*. Ann. Math. Statist. **38** (1967), 1384–1391.

[127] V. K. Rohatgi and S. C. Rastogi, *On sequential estimation of a certain estimable function of the mean vector of a multivariate normal distribution*, J. Austral. Math. Soc. **15** (1973), 291–295.

[128] S. Rushton, *On a sequential t-test*, Biometrika **37** (1950), 326—333.

[129] –, *On a two-sided sequential t-test*, Biometrika **39** (1952), 302–308.

[130] M. Sakaguchi, *Relative efficiency of the Wald SPRT and the Chernoff information number*, Kodai. Math. Sem. Rep. **19** (1967), 138–145.

[131] A. I. Salyt, *Some results on sequential estimation of a translation parameter*, Soviet Math. Dokl. **10** (1969), 1365–1367.

[132] –, *On optimal sequential estimation of a translation parameter in the class of invariant procedures*. Soviet Math. Dokl. **11** (1970), 1031–1033.

[133] I. R. Savage and L. J. Savage, *Finite stopping time and finite expected time*, J. Roy. Statist. Soc. Ser. B **37** (1965), 284–289.

[134] I. R. Savage and J. Sethuraman, *Stopping time of a rank order sequential probability ratio test based on Lehmann alternatives*, Ann. Math. Statist. **37** (1966). Correction: Ann. Math. Statist. **38** (1967), 1309.

[135] G. Schwarz, *Asymptotic shapes of Bayes sequential testing regions*, Ann. Math. Statist. **33** (1963), 224–236.

[136] P. K. Sen and M. Ghosh, *On bounded length sequential confidence intervals based on one-sample rank-order statistics*, Ann. Math. Statist. **42** (1971), 189–203.

[137] G. R. Seth, *On the variance of estimates*, Ann. Math. Statist. **20** (1959), 1–27.

[138] J. Sethuraman, *Stopping time of a rank order sequential probability ratio test based on Lehmann alternatives, II*, Ann. Math. Statist. **41** (1970), 1322–1333.

[139] C. P. Shapiro and R. L. Wardrop, *Bayes sequential estimation for one parameter exponential families*, J. Amer. Statist. Assoc. **75** (1981), 984–988.

[140] D. O. Siegmund, *On the asymptotic normality of one-sided stopping rules*. Ann. Math. Statist. **39** (1968), 1493–1497.

[141] –, *Estimation following sequential tests*, Biometrika **65** (1978), 341–349.

[142] –, *Advanced Seminar on Sequential Analysis and its Applications*. Series of the lectures held at the University of Kentucky during June 15–19, 1981, and sponsored by the Office of Naval Research, (1981).

[143] A. Sierociński, *On fixed precision estimation of the maximal value of a random variable*, Demonstratio Math. **XIII** (1980), 995–1003.

[144] G. D. Simons, *An improved statement of optimality for sequential probability ratio tests*, Ann. Statist. **4** (1976), 1240–1243.

[145] M. Sobel and A. Wald, *A sequential decision procedure for choosing one of three hypotheses concerning the unknown mean of a normal distribution*, Ann. Math. Statist. **20** (1949), 502–522.

[146] M. S. Srivastava, *On fixed-width confidence bounds for regression parameters*, Ann. Math. Statist. **42** (1971), 1403–1411.

[147] N. Starr, *The performance of a sequential procedure for fixed-width interval estimate*. Ann. Math. Statist. **36** (1966a), 36–50.

[148] –, *On the asymptotic efficiency of a sequential procedure for estimating the mean*, Ann. Math. Statist. **37** (1966b), 1173–1185.

[149] N. Starr and M. B. Woodroofe, *Remarks on sequential point estimation*, Proc. Nat. Acad. Sci. U.S.A. **63** (1969), 285–288.

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

[151] –, *A note on cumulative sums*, Ann. Math. Statist. **17** (1946), 498–499.

[152] G. M. Tallis and M. K. Vagholkar, *Formulas to improve Wald's approximation for some properties of sequential test*, J. Roy. Statist. Soc. Ser B **27** (1965), 74–81.

[153] M. K. Vagholkar and G. B. Wetherill, *The most economical binomial sequential probability ratio test*, Biometrika **47** (1960), 103–109.

[154] A. Wald, *Sequential Analysis*, Wiley, New York 1947; Reprinted by Dover Publications Inc. (1973).

[155] –, *Statistical Decision Functions*, Second (Cor.) Edition, Chelsea Publishing Company, New York 1971, 164–167.

[156] A. Wald and J. Wolfowitz, *Optimum character of the sequential probability ratio test*, Ann. Math. Statist. **19** (1948), 326–339.

[157] A. W. Wallis, *The statistical research group*, J. Amer. Statist. Assoc. **75** (1980), 320–334.

[158] H. D. Weed, Jr. and R. A. Bradley, *Sequential one-sample grouped signed rank tests for symmetry*, J. Amer. Statist. Assoc. **66** (1971), 321–326.

[159] H. D. Weed, Jr., R. A. Bradley and Z. Govindarajulu, *Stopping times of two rank order sequential probability ratio tests for symmetry based on Lehmann alternatives*, Ann. Statist. **2** (1974), 1314–1322.

[160] L. Weiss, *Testing one simple hypothesis against another*, Ann. Math. Statist. **24** (1953), 273–281.

[161] –, *On the uniqueness of Wald sequential tests*, Ann. Math. Statist. **27** (1956), 1178–1181.
[162] L. Weiss and J. Wolfowitz, *Optimal, fixed-length nonparametric sequential confidence limits for a translation parameter*, Z. Wahrsch. Verw. Gebiete **25** (1972), 203–209.
[163] P. Whittle and R. O. D. Lane, *A class of situations in which a sequential estimation procedure is nonsequential*, Biometrika **54** (1967), 229–234.
[164] N. Wiener, *The ergodic theorem*, Duke Math. J. **5** (1939), 1–18.
[165] R. A. Wijsman, *A monotonicity property of the sequential probability ratio test*, Ann. Math. Statist. **31** (1960), 677–684; Correction – ibid. **3** (1975), 796.
[166] –, *Existence, uniqueness and monotonicity of sequential probability ratio tests*, Ann. Math. Statist. **34** (1963), 1541–1548.
[167] –, *Exponentially bounded stopping time of sequential probability ratio test for composite hypotheses*, Ann. Math. Statist. **42** (1971), 1859–1869.
[168] –, *Confidence sets based on sequential tests*, Comm. Statist. A – Theory Methods **10** (1981), 2137–2147.
[169] –, *Sequential confidence sets: estimation oriented versus test oriented construction*, in: S. S. Gupta and J. O. Berger (eds.), Proc. Third Purdue Symp on Statist. Dec. Theor. and Related Topics Academic Press, New York 1982. Vol. 2. 435–450.
[170] F. Wilcoxon, L. J. Rhodes, and R. A. Bradley *Two sequential two-sample grouped rank tests with applications to screening experiments*, Biometrics **19** (1963), 58–84.
[171] S. Wirjosudirdjo, *Limiting Behavior of a Sequence of Density Ratios*, Ph. D. Thesis University of Illinois, Urbana 1961; Abstract: Ann. Math. Statist. **33** (1962), 296–297.
[172] H. Wittenberg, *Limiting distributions of random sums of independent random variables*, Z. Wahrsch. Verw. Gebiete **3** (1964), 7–18.
[173] J. Wolfowitz, *Consistency of sequential binomial estimates*, Ann. Math. Statist. **18** (1947a) 131–135.
[174] –, *Efficiency of sequential estimates and Wald's equation for sequential processes*, Ann. Math. Statist. **18** (1947b), 215–230.
[175] –, *Remark on the optimum character of the sequential probability ratio test*, Ann. Math. Statist. **37** (1966), 726–727.
[176] M. Woodroofe, *A renewal theorem for curved boundaries and moments of first passage times*, Ann. Probab. **4** (1976a) 67–80.
[177] –, *Frequentist properties of Bayesian sequential test*, Biometrika **63** (1976b), 101–110.
[178] –, *Second order approximations for sequential point and interval estimation*, Ann. Statist. **5** (1977), 985–995.
[179] J. A. Yahav and P. J. Bickel, *Asymptotically optimal Bayes and minimax procedures in sequential estimation*, Ann. Math. Statist. **39** (1968), 442–456.
[180] R. A. Zaidman, Yu. V. Linnik and I. V. Romanovski, *Sequential estimation plans and Markov stopping times*, Soviet Math. Dokl. **10** (1969), 484–487.
[181] R. Zieliński, *A stopping rule for sequential estimation processes: Two existence theorems*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. **XXVI** (1978), 441–444.
[182] –, *A class of stopping rules for fixed precision sequential estimates*, Zastos. Mat. **XVII** (1982), 277–281.
[183] –, *Fixed precision estimate of mean of a Gaussian sequence with unknown covariance structure*, in *Mathematical Statistics and Probability Theory. Proc. Sixth Internat. Conference, Wisla (Poland)*, Lecture Notes in Statist. **2**, Springer – Verlag, Berlin 1980, 360–364.

*Presented to the semester*
*Sequential Methods in Statistics*
*September 7 –December 11, 1981*