The Annals of Statistics
1976, Vol. 4, No. 2, 265–280

ON CONFIDENCE SEQUENCES[1]

BY TZE LEUNG LAI

Columbia University

This paper is concerned with confidence sequences, i.e., sequences of confidence regions which contain the true parameter for every sample size simultaneously at a prescribed level of confidence. By making use of generalized likelihood ratio martingales, confidence sequences are constructed for the unknown parameters of the binomial, Poisson, uniform, gamma and other distributions. It is proved that for the exponential family of distributions, the method of using generalized likelihood ratio martingales leads to a sequence of intervals which have the desirable property of eventually shrinking to the population parameter. The problem of nuisance parameters is considered, and in this connection, boundary crossing probabilities are obtained for the sequence of Student's t-statistics, and a limit theorem relating to the boundary crossing probabilities for the Wiener process is proved.

1. Introduction. Let $X_1, X_2, \cdots$ be a sequence of independent random variables having a common distribution function $F_{\theta,\sigma}$, $(\theta, \sigma) \in \Theta$, and taking values in $\mathscr{X}$. A family

$$\{\Gamma_n(x_1, \cdots, x_n) : n \geqq 1, x_i \in \mathscr{X} \text{ for } i = 1, \cdots, n\}$$

of the subsets of the space of θ is said to constitute a $(1 - \alpha)$-*level sequence of confidence sets if*

$$P_{\theta,\sigma}[\theta \in \Gamma_n(X_1, \cdots, X_n) \text{ for all } n \geqq 1] \geqq 1 - \alpha \qquad \text{for all} \quad (\theta, \sigma) \in \Theta .$$

If the θ-space is a topological space, the confidence sequence is said to be *consistent* if

$$P_{\theta,\sigma}[\exists \hat{\theta}_n \in \Gamma_n(X_1, \cdots, X_n) \text{ for each } n \geqq 1 \text{ such that } \lim_{n\to\infty} \hat{\theta}_n = \theta] = 1 \qquad \text{for all} \quad (\theta, \sigma) \in \Theta .$$

If the θ-space is a metric space and $\rho(\Gamma)$ denotes the diameter of a subset Γ of the θ-space, then the confidence sequence is said to be *degenerate in the limit* if

$$P_{\theta,\sigma}[\lim_{n\to\infty} \rho(\Gamma_n(X_1, \cdots, X_n)) = 0] = 1 \qquad \text{for all} \quad (\theta, \sigma) \in \Theta .$$

Clearly if $\{\Gamma_n(x_1, \cdots, x_n) : n \geqq 1, x_i \in \mathscr{X}\}$ is a $(1 - \alpha)$-level confidence sequence, then so is $\{\Gamma_n^*(x_1, \cdots, x_n) : n \geqq 1, x_i \in \mathscr{X}\}$, where $\Gamma_n^*(x_1, \cdots, x_n) = \bigcap_{i=1}^n \Gamma_i(x_1, \cdots, x_i)$, and the sequence $\{\Gamma_n^*\}$ has the following nice property:

Received April 1975; revised July 1975.

[1] This paper is based on part of the author's doctoral dissertation at Columbia University. Research supported by PHS Grant No. 5-R01-GM-16895-02.

AMS 1970 *subject classifications.* Primary 62F25, 62L10.

Key words and phrases. Confidence sequences, likelihood ratio martingales, nuisance parameters, invariance, Student's t-statistics, boundary crossing probabilities.

$\Gamma_1^* \supset \Gamma_2^* \supset \cdots$. Such shrinking confidence sequences have applications in selection and ranking procedures (see [10] and [13]).

In 1967, Darling and Robbins [2], [4] obtained a $(1-\alpha)$-level consistent sequence of confidence intervals for the mean of a normal distribution when the variance is known and also when the variance is unknown, and these confidence sequences are degenerate in the limit. They also constructed a consistent $(1-\alpha)$-level sequence of upper confidence bounds for the variance of a normal distribution with unknown mean, and a $(1-\alpha)$-level confidence sequence for the median of a distribution of unknown form. Earlier in 1964, Paulson [11] constructed a $(1-\alpha)$-level sequence of confidence intervals for the mean of a normal distribution, but the sequence is not degenerate in the limit.

When θ is a real parameter, there is a well-known connection between $(1-\alpha)$-level confidence intervals for θ and tests of hypotheses of the form $H: \theta = \theta_0$ versus $K: \theta \neq \theta_0$. In the case of a $(1-\alpha)$-level consistent sequence, degenerate in the limit, of confidence sets $\{\Gamma_n(x_1, \cdots, x_n) : n \geqq 1, x_i \in \mathscr{X}\}$ for θ, the sequential test which stops sampling with $N = \inf\{n \geqq 1 : \theta_0 \notin \Gamma_n(X_1, \cdots, X_n)\}$ and rejects H is a level α test for H versus K, and this sequential test has a zero Type II error probability. When each $\Gamma_n(x_1, \cdots, x_n)$ is an interval, the confidence sequence can also be used to test $H_0' : \theta < \theta_0$ versus $H_1' : \theta > \theta_0$. Our stopping rule is N as before and our terminal decision rule is to reject H_0' if the left endpoint of the interval $\Gamma_N(X_1, \cdots, X_N)$ is larger than θ_0 and to accept H_0' if otherwise. This is a test with error probabilities uniformly $\leqq \alpha$. Such tests and similar tests for one-sided hypotheses $H_0: \theta \leqq \theta_0$ versus $H_1: \theta > \theta_0$ have been considered by Fabian [6], Farrell [7], Darling and Robbins [2], [3], [4], [5], Robbins and Siegmund [13] and Robbins [12]. Performance characteristics of such tests and the applications of confidence sequences to some detection problems are studied in [10].

In Section 3, by making use of generalized likelihood ratio martingales introduced in Section 2, we construct confidence sequences for the unknown parameters of certain important distributions. Theorem 1 of Section 4 shows that for the exponential family, the confidence sequences constructed by our method are always consistent and degenerate in the limit. In Section 5, we consider invariant confidence sequences, largely in the context of t-statistics, and show how nuisance parameters can be handled in certain situations by using the principle of invariance. Section 6 deals with boundary crossing probabilities and a limit theorem which arise in connection with Section 5. Throughout this paper, we shall denote the nth sample sum by S_n, i.e., $S_n = X_1 + \cdots + X_n$, where $X_1, X_2, \cdots$ are i.i.d. observations.

2. Generalized likelihood ratio martingales. Let μ be some σ-finite measure on the real line and μ_n be the product measure induced by μ on $R^n (n = 1, 2, \cdots)$, i.e., $d\mu_n(x_1, \cdots, x_n) = d\mu(x_1) \cdots d\mu(x_n)$. Let $X_1, X_2, \cdots$ be random variables such that for each $n \geqq 1$, $X_1, \cdots, X_n$ have a joint density function p_n with respect to μ_n.

For each $n \geqq 1$, let q_n be an extended real-valued Borel function on R^n satisfying:

(a) $q_n \geqq 0$,
(b) $\int_{-\infty}^{\infty} q_{n+1}(x_1, \cdots, x_{n+1})\, d\mu(x_{n+1}) = q_n(x_1, \cdots, x_n)$,
(c) $q_n(x_1, \cdots, x_n) = 0$ whenever $p_n(x_1, \cdots, x_n) = 0$.

Define

$$Z_n = q_n(X_1, \cdots, X_n)/p_n(X_1, \cdots, X_n) \qquad \text{if } \; p_n(X_1, \cdots, X_n) \neq 0\,,$$
$$= 0 \qquad \text{otherwise.}$$

Then Z_n, $n \geqq 1$, is a martingale (possibly with infinite expectation) and will hereafter be referred to as a *generalized likelihood ratio martingale*. If in addition to (a), (b) and (c), the sequence q_n also satisfies

(d) $\int_{R^n} q_n(x_1, \cdots, x_n)\, d\mu(x_1) \cdots d\mu(x_n) = 1$,

then Z_n, $n \geqq 1$, is called a *proper likelihood ratio martingale*.

More generally, for each $n \geqq 1$, let q_n be an extended real-valued Borel function on R^n satisfying (a) and

(e) $\int_{-\infty}^{\infty} q_{n+1}(x_1, \cdots, x_{n+1})\, d\mu(x_{n+1}) \leqq q_n(x_1, \cdots, x_n)$.

Define Z_n as above. Then Z_n, $n \geqq 1$, is a nonnegative supermartingale, and will be referred to as a *pseudo likelihood ratio supermartingale*. Our construction of confidence sequences in this paper depends on the following inequality for Z_n (cf. [12], page 1400): For any $\varepsilon > 0$,

$$P[Z_n \geqq \varepsilon \;\text{ for some }\; n \geqq m] \leqq P[Z_m \geqq \varepsilon] + \varepsilon^{-1} \int_{[Z_m < \varepsilon]} Z_m\, dP \tag{1}$$
$$\leqq \varepsilon^{-1} E Z_m\,.$$

Note that in the case where q_n satisfies (d), we have $EZ_n \leqq 1$.

Let $X_1, X_2, \cdots$ be i.i.d. random variables such that $\varphi(\theta) = E \exp(\theta X_1) < \infty$ for all $\theta > 0$, and let F be a measure on $(0, \infty)$. It is well known that $\int_0^\infty (\varphi(\theta))^{-n} \exp(\theta S_n)\, dF(\theta)$, $n \geqq 1$, is a martingale, and we shall call it a *moment generating function martingale*. This martingale can in fact be regarded as a generalized likelihood ratio martingale, where we define $p_\theta(x) = e^{\theta x}\psi(x)/\varphi(\theta)$ (ψ being the density function of X_1 with respect to some measure μ), $p_{\theta,n}(x_1, \cdots, x_n) = \prod_{i=1}^n p_\theta(x_i)$, and $q_n(x_1, \cdots, x_n) = \int_0^\infty p_{\theta,n}(x_1, \cdots, x_n)\, dF(\theta)$. Moment generating function martingales are used in [2], [3], [4], [5], [8], [12], [13] to obtain confidence sequences based on sample sums.

3. Confidence sequences for some important distributions. In this section, we shall construct consistent confidence sequences which are degenerate in the limit for the unknown parameters of the Bernoulli, negative binomial, uniform, exponential, Poisson and gamma distributions,

(A) *Bernoulli distribution*. Suppose $X_1, X_2, \cdots$ are i.i.d. such that $P_p[X_1 = 1] = p$, $P_p[X_1 = 0] = 1 - p = q$, $p \in (0, 1)$. Then writing $b(n, p, x) = \binom{n}{x} p^x q^{n-x}$, Robbins

[12] has shown that

$$(2)\qquad P_p[b(n,p,S_n) \leqq \alpha/(n+1) \text{ for some } n \geqq 1] \leqq \alpha \qquad \text{for all } p \in (0,1)\,,\ 0<\alpha<1\,.$$

Consider the solutions $p = p(x)$ of the equation $(n+1)b(n,p,x) = \alpha$. For fixed $x = 0, 1, \cdots, n$, the function $(n+1)b(n,p,x)$ has a maximum at $p = x/n$ and the maximum value is $(n+1)\binom{n}{x}(x/n)^x(1-x/n)^{n-x}$, which is larger than 1. Hence the equation $(n+1)b(n,p,x) = \alpha$ has two distinct roots $p = f_n(x)$ and $p = g_n(x)$ with $f_n(x) > g_n(x)$. Therefore from (2), we have

$$P_p[g_n(S_n) < p < f_n(S_n) \text{ for all } n \geqq 1] \geqq 1-\alpha \qquad \text{for all } p \in (0,1)\,.$$

We now prove that $P_p[\lim_{n\to\infty} f_n(S_n) = p = \lim_{n\to\infty} g_n(S_n)] = 1$ for all $p \in (0,1)$. We need only show that if x_n is a sequence of natural numbers such that $0 \leqq x_n \leqq n$ and $\lim_{n\to\infty} x_n/n = p_0 \in (0,1)$, then $\lim_{n\to\infty} f_n(x_n) = p_0 = \lim_{n\to\infty} g_n(x_n)$. Consider the two roots $p = f_n(x_n)$ and $p = g_n(x_n)$ of the equation

$$p^{x_n/n}(1-p)^{1-x_n/n} = \alpha^{1/n}(x_n!(n-x_n)!/(n+1)!)^{1/n}\,.$$

Since $\lim_{n\to\infty} x_n/n = p_0$ and the right-hand side of the above equation converges to $p_0^{\,p_0}(1-p_0)^{1-p_0}$, it is clear that

$$\lim_{n\to\infty} f_n(x_n) = p_0 = \lim_{n\to\infty} g_n(x_n)\,.$$

For each $n \geqq 1$, let $I_n = (g_n(S_n), f_n(S_n))$. We know that for all $p \in (0,1)$, $P_p[p \notin I_n] \leqq \alpha$, and that $P_p[p \notin I_n] \to 0$ as $n \to \infty$. It is interesting to ask how fast $P_p[p \notin I_n]$ converges to 0. The answer is given by:

$$(3)\qquad P_p[p \notin I_n] \sim \alpha[2p(1-p)]^{\frac12}(n\log n)^{-\frac12} \qquad \text{as } n \to \infty\,.$$

To prove (3), note that $P_p[p \notin I_n] = P_p[b(n,p,S_n) \leqq \alpha/(n+1)]$. Let $Y_n = (S_n - np)/(npq)^{\frac12}$. Given any $\delta \in (0,1)$, we can choose n_0 such that for all $n \geqq n_0$ and all k with $|k-np| \leqq n^{\frac1{12}}(npq)^{\frac12}$,

$$(2\pi npq)^{-\frac12}(1-\delta)\exp(-(k-np)^2/(2npq)) < b(n,p,k) < (2\pi npq)^{-\frac12}(1+\delta)\exp(-(k-np)^2/(2npq))\,.$$

Therefore for $n \geqq n_0$,

$$\begin{aligned} P_p[b(n,p,S_n) &\leqq \alpha/(n+1)] \\ &\geqq P_p[n^{\frac1{12}} \geqq |Y_n| \geqq \{2\log[(2\pi npq)^{-\frac12}(1+\delta)(n+1)/\alpha]\}^{\frac12}] \\ &\sim (2pq)^{\frac12}(n\log n)^{-\frac12}\alpha/(1+\delta)\,. \end{aligned}$$

On the other hand,

$$\begin{aligned} P_p[b(n,p,S_n) \leqq \alpha/(n+1)] &\leqq P_p[|Y_n| \geqq \{2\log[(2\pi npq)^{-\frac12}(1-\delta)(n+1)/\alpha]\}^{\frac12}] \\ &\sim (2pq)^{\frac12}(n\log n)^{-\frac12}\alpha/(1-\delta)\,. \end{aligned}$$

Since δ is arbitrary, (3) follows.

(B) *Negative binomial distribution.* Suppose $X_1, X_2, \cdots$ are i.i.d. with $P_p[X_1 = x] = \binom{x-1}{r-1}p^r(1-p)^{x-r}$, $x = r, r+1, \cdots$, where $0 < p < 1$ and r is a known positive integer. Define

$$\begin{aligned} Z_n &= \textstyle\int_0^1 \theta^{nr}(1-\theta)^{S_n - nr}\, d\theta/(p^{nr}(1-p)^{S_n-nr}) \\ &= 1/((S_n+1)b(S_n, p, nr)) . \end{aligned}$$

Then Z_n, $n \geqq 1$, is a proper likelihood ratio martingale and by (1),

$$P_p[b(S_n, p, nr) \leqq \alpha/(S_n+1) \text{ for some } n \geqq 1] \leqq \alpha \qquad \text{for all } p \in (0, 1), \quad 0 < \alpha < 1 .$$

As in (A), it can be proved that for fixed $x = nr, nr+1, \cdots$, the equation $(x+1)b(x, p, nr) = \alpha$ has two distinct roots $p = f_n(x)$ and $p = g_n(x)$ with $f_n(x) > g_n(x)$, and so for all $p \in (0, 1)$,

$$P_p[g_n(S_n) < p < f_n(S_n) \text{ for all } n \geqq 1] \geqq 1 - \alpha .$$

Now $E_p X_1 = r + r(1-p)/p$. Given $p_0 \in (0, 1)$ and any sequence of positive integers x_n with $x_n \geqq nr$ and $\lim_{n\to\infty} x_n/n = r + r(1-p_0)/p_0$, we have

$$\lim_{n\to\infty} \alpha^{1/n}((nr)!(x-nr)!/(x_n+1)!)^{1/n} = p_0{}^r(1-p_0)^{r(1-p_0)/p_0} .$$

From this it follows that $\lim_{n\to\infty} f_n(x_n) = p_0 = \lim_{n\to\infty} g_n(x_n)$. Therefore

$$P_p[\lim_{n\to\infty} f_n(S_n) = p = \lim_{n\to\infty} g_n(S_n)] = 1 .$$

(C) *Uniform distribution with range parameter.* Suppose $X_1, X_2, \cdots$ are i.i.d. with uniform density

$$\begin{aligned} p_\xi(x) &= 1/\xi \qquad && \text{if } x \in (a, a+\xi) , \\ &= 0 && \text{elsewhere,} \end{aligned}$$

where $\xi > 0$ is an unknown parameter and a is a known real number. Let $X^{(n)} = \max(X_1, \cdots, X_n)$, $X_{(n)} = \min(X_1, \cdots, X_n)$. Define

$$\begin{aligned} p_{\theta,n} &= \theta^{-n} I_{(X^{(n)}-a,\infty)}(\theta) \cdot I_{(-\infty, X_{(n)})}(a) , \\ Z_n &= \textstyle\int_{0+}^\infty p_{\theta,n}\theta^{-2} \exp(-1/\theta)\, d\theta/p_{\xi,n} . \end{aligned}$$

Then $Z_n = I_{(-\infty, X_{(n)})}(a) \cdot \gamma_{n+1}(1/(X^{(n)} - a))/p_{\xi,n}$, where γ_{n+1} is the incomplete gamma function, i.e., $\gamma_{n+1}(x) = \int_0^x t^n e^{-t}\, dt$. It follows from (1) that for any $0 < \alpha < 1$,

$$P_\xi[X^{(n)} - a < \xi < (\alpha\gamma_{n+1}(1/(X^{(n)} - a)))^{-1/n} \text{ for all } n \geqq 1] \geqq 1 - \alpha .$$

Obviously, $P_\xi[\lim_{n\to\infty} X^{(n)} = a + \xi] = 1$. It is also easy to show that $P_\xi[\lim_{n\to\infty} (\alpha\gamma_{n+1}(1/(X^{(n)} - a)))^{-1/n} = \xi] = 1$.

(D) *Exponential distribution with location parameter.* Suppose $X_1, X_2, \cdots$ are i.i.d. with density function

$$\begin{aligned} p_\xi(x) &= \exp(-(x-\xi)) , \qquad && x \geqq \xi , \\ &= 0 , && \text{elsewhere,} \end{aligned}$$

where ξ is an unknown parameter. Define $X_{(n)} = \min(X_1, \cdots, X_n)$, $p_{\theta,n} = \exp(-S_n + n\theta)I_{(-\infty, X_{(n)}]}(\theta)$, and let

$$Z_n = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} p_{\theta,n} \exp(-\theta^2/2)\, d\theta / p_{\xi,n}$$
$$= \exp(n^2/2 - n\xi)\Phi(X_{(n)} - n)/I_{(-\infty, X_{(n)}]}(\xi)\,,$$

where $\Phi(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} \exp(-t^2/2)\, dt$.
From (1), it follows that for any $\alpha \in (0, 1)$,

$$P_\xi[n/2 + (1/n)\log \Phi(X_{(n)} - n) + (\log \alpha)/n < \xi \leqq X_{(n)} \text{ for all } n \geqq 1]$$
$$\geqq 1 - \alpha\,.$$

Obviously $P_\xi[\lim_{n\to\infty} X_{(n)} = \xi] = 1$. Take any real number x. It is not hard to show that for $n \geqq 2$,

$$n/2 + (1/n)\log \Phi(x - n) + (\log \alpha)/n < x\,. \tag{4}$$

From (4), we have

$$n/2 + (1/n)\log \Phi(X_{(n)} - n) + (\log \alpha)/n < X_{(n)} \to \xi \qquad \text{a.s. } [P_\xi]\,.$$

On the other hand, for all large n,

$$n/2 + (1/n)\log \Phi(X_{(n)} - n) + (\log \alpha)/n$$
$$\geqq n/2 + (1/n)\log \{(2\pi)^{-\frac{1}{2}}(n - X_{(n)})$$
$$\times \exp(-(n - X_{(n)})^2/2)/(1 + (n - X_{(n)})^2)\} + (\log \alpha)/n \to \xi$$
$$\text{a.s. } [P_\xi]\,.$$

Hence $P_\xi[\lim_{n\to\infty} (n/2 + (1/n)\log \Phi(X_{(n)} - n) + (\log \alpha)/n) = \xi] = 1$.

(E) *Poisson distribution*. Suppose $X_1, X_2, \cdots$ are i.i.d. Poisson random variables with parameter λ. Let

$$Z_n = \int_0^\infty e^{-n\theta}\theta^{S_n}e^{-\theta}\, d\theta / (e^{-n\lambda}\lambda^{S_n})$$
$$= e^{n\lambda}(S_n!)\lambda^{-S_n}(n + 1)^{-(S_n+1)}\,.$$

Since Z_n, $n \geqq 1$, is a proper likelihood ratio martingale, it follows from (1) that if $0 < \alpha < 1$, then for all $\lambda > 0$,

$$P_\lambda[\text{For all } n \geqq 1, \text{ either } S_n = 0 \text{ and } e^{n\lambda} < (n + 1)/\alpha, \text{ or}$$
$$S_n > 0 \text{ and } \exp(n\lambda/S_n) < V_n n\lambda/S_n] \geqq 1 - \alpha\,, \tag{5}$$

where $V_n = (n + 1)^{(S_n+1)/S_n}(\alpha(S_n!))^{-1/S_n}S_n/n$. Let a be any positive number and consider solutions $x = x(a)$ of the equation $e^x = ax$, $x > 0$. If $a > e$, this equation has two distinct roots $x_1(a) < x_2(a)$. If $a = e$, this equation has one root, while if $a < e$, there are no roots. For $a \leqq e$, let us define $x_1(a) = 0$, $x_2(a) = \infty$. If $S_n > 0$, define $f_n(S_n) = x_2(V_n)S_n/n$, $g_n(S_n) = x_1(V_n)S_n/n$. If $S_n = 0$, let $f_n(S_n) = (\log(n + 1) - \log \alpha)/n$, and let $g_n(S_n) = 0$. Then it follows from (5) that for all $\lambda > 0$, $P_\lambda[g_n(S_n) < \lambda < f_n(S_n) \text{ for all } n \geqq 1] \geqq 1 - \alpha$. It is not hard to see that $P_\lambda[\lim_{n\to\infty} g_n(S_n) = \lambda = \lim_{n\to\infty} f_n(S_n)] = 1$.

(F) *Gamma distribution with scale parameter*. Suppose $X_1, X_2, \cdots$ are i.i.d.

with the gamma density $f_\theta(x)$, i.e., for $x > 0$, $f_\theta(x) = \theta^{-\beta}x^{\beta-1}\exp(-x/\theta)/\Gamma(\beta)$, where β is known. Let

$$\begin{aligned} Z_n &= \int_0^\infty \theta^{-n\beta-2}\exp(-S_n/\theta - 1/\theta)\,d\theta/(\lambda^{-n\beta}e^{-S_n/\lambda}) \\ &= \lambda^{n\beta}e^{S_n/\lambda}(S_n + 1)^{-(n\beta+1)}\Gamma(n\beta + 1)\,. \end{aligned}$$

Since Z_n, $n \geqq 1$, is a proper likelihood ratio martingale, we can apply (1) and obtain that if $0 < \alpha < 1$, then for all $\lambda > 0$,

$$P_\lambda[(n\beta\lambda/S_n)\exp(S_n/(n\beta\lambda)) < U_n \text{ for all } n \geqq 1] \geqq 1 - \alpha\,, \tag{6}$$

where $U_n = (n\beta/S_n)(S_n + 1)^{(n\beta+1)/n\beta}(\alpha\Gamma(n\beta + 1))^{-1/n\beta}$.

Let a be any positive number and consider solutions $x = x(a)$ of the equation $xe^{1/x} = a(x > 0)$. If $a > e$, this equation has two distinct roots $x_1(a) < x_2(a)$. If $a = e$, this equation has one root at $x = 1$, while if $a < e$, there are no roots. Define $x_1(a) = 0$ and $x_2(a) = \infty$ in the latter two cases and set $f_n(S_n) = (S_n/n\beta)x_2(U_n)$, $g_n(S_n) = (S_n/n\beta)x_1(U_n)$. It follows from (6) that

$$P_\lambda[g_n(S_n) < \lambda < f_n(S_n) \text{ for all } n \geqq 1] \geqq 1 - \alpha\,.$$

It is easy to see that

$$P_\lambda[\lim_{n\to\infty} g_n(S_n) = \lambda = \lim_{n\to\infty} f_n(S_n)] = 1\,.$$

4. A general theorem for the exponential family. In the preceding section, we have proved directly that the confidence sequences constructed in (A), (B), (E), (F) all shrink to the population parameter. Actually there is the following more general result which holds for the exponential family of distributions.

THEOREM 1. *Let $X_1, X_2, \cdots$ be* i.i.d. *having a common density $p_\theta(x) = h(\theta)e^{\theta x}$, $\theta \in \Theta$, with respect to some nondegenerate measure ν on the real line, where Θ is an interval. Let F be a measure on Θ such that $F(I) > 0$ for any open interval I contained in Θ. Then given $\varepsilon > 0$, the set*

$$G_n(S_n) = \{\lambda \in \Theta : \textstyle\int_\Theta (h(\theta)/h(\lambda))^n \exp((\theta - \lambda)S_n)\,dF(\theta) < \varepsilon\}$$

is either empty or an interval. Let $\lambda_n^{(1)} \leqq \lambda_n^{(2)}$ be the end-points of $G_n(S_n)$ if it is nonempty. Suppose $P_\lambda[G_n(S_n) \neq \emptyset$ for all large $n] = 1$. Then $P_\lambda[\lim_{n\to\infty}\lambda_n^{(1)} = \lim_{n\to\infty}\lambda_n^{(2)} = \lambda] = 1$.

PROOF. Let $\mu(\theta) = E_\theta X_1 = \int_{-\infty}^\infty h(\theta)xe^{\theta x}\,d\nu(x)$. Then

$$\mu(\theta) = -h'(\theta)/h(\theta)\,, \qquad \frac{d\mu(\theta)}{d\theta} = \operatorname{Var}_\theta X_1 > 0\,.$$

Setting $g(\lambda, \theta) = (h(\theta)/h(\lambda))^n \exp((\theta - \lambda)S_n)$, we see that

$$\partial g/\partial\lambda = n(\mu(\lambda) - S_n/n)g(\lambda, \theta)\,, \qquad \partial^2 g/\partial\lambda^2 > 0\,.$$

Hence for a fixed θ, $g(\lambda, \theta)$ is a strictly convex function of $\lambda \in \Theta$, and is decreasing for $\mu(\lambda) \leqq S_n/n$ and increasing for $\mu(\lambda) \geqq S_n/n$. From this, it is clear that either $G_n(S_n)$ is empty, or $G_n(S_n)$ is an interval with end-points $\lambda_n^{(1)} \leqq \lambda_n^{(2)}$.

Let $\lim_{n\to\infty} S_n(\omega)/n = E_\lambda X_1 = \mu(\lambda)$ and $G_n(S_n(\omega)) \neq \emptyset$ for all large n. We assert that $\lim_{n\to\infty} \lambda_n^{(1)}(\omega) = \lambda = \lim_{n\to\infty} \lambda_n^{(2)}(\omega)$. Since $\lambda_n^{(1)}(\omega) \leqq \lambda_n^{(2)}(\omega)$, it suffices to prove that $\liminf_{n\to\infty} \lambda_n^{(1)}(\omega) \geqq \lambda \geqq \limsup_{n\to\infty} \lambda_n^{(2)}(\omega)$. First suppose that λ is an interior point of Θ. Then for all large n, $S_n(\omega)/n$ is an interior point of $\mu[\Theta]$, and so $\mu^{-1}(S_n(\omega)/n)$ can be defined and $\lambda_n^{(1)}(\omega) \leqq \mu^{-1}(S_n(\omega)/n) \leqq \lambda_n^{(2)}(\omega)$. Assume that $\liminf_{n\to\infty} \lambda_n^{(1)}(\omega) < \lambda$. Then we can choose an interior point λ_1 of Θ and a subsequence n_j such that $\lambda_{n_j}^{(1)}(\omega) < \lambda_1 < \mu^{-1}(S_{n_j}(\omega)/n_j)$ and $\lambda_1 < \lambda$. We note that since $\lambda_1 \in (\lambda_{n_j}^{(1)}(\omega), \lambda_{n_j}^{(2)}(\omega))$,

$$\begin{aligned} \varepsilon &\geqq \liminf_{j\to\infty} \int \exp\{(\theta - \lambda_1)S_{n_j}(\omega) + n_j \log(h(\theta)/h(\lambda_1))\}\, dF(\theta) \\ &\geqq \int_{(\lambda_2,\lambda_3)} \liminf_{j\to\infty} \exp\{n_j[(\theta - \lambda_1)S_{n_j}(\omega)/n_j + \log(h(\theta)/h(\lambda_1))]\}\, dF(\theta)\,, \end{aligned}$$

where λ_2, λ_3 are chosen as follows. Consider the continuous function $\psi(\xi, \theta) = \xi(\theta - \lambda_1) + \log(h(\theta)/h(\lambda_1))$. Now $\psi(\mu(\lambda), \theta) = 0$ at $\theta = \lambda_1$ and $(\partial/\partial\theta)\psi(\mu(\lambda), \theta) = \mu(\lambda) - \mu(\theta) \gtrless 0$ according as $\lambda \gtrless \theta$. Hence $\psi(\mu(\lambda), \theta_0) > 0$ for any $\theta_0 \in (\lambda_1, \lambda)$. By the continuity of ψ, we can therefore choose $(\lambda_2, \lambda_3) \subset \Theta$ and a neighborhood U of $\mu(\lambda)$ such that $\psi(\xi, \theta) > 0$ for all $(\xi, \theta) \in U \times (\lambda_2, \lambda_3)$. Therefore if j is large enough, $(\theta - \lambda_1)S_{n_j}(\omega)/n_j + \log(h(\theta)/h(\lambda_1)) > 0$ for all $\theta \in (\lambda_2, \lambda_3)$, and so using the fact that $F(\lambda_2, \lambda_3) > 0$, we have

$$\varepsilon \geqq \int_{(\lambda_2,\lambda_3)} \lim_{j\to\infty} \exp\{n_j[(\theta - \lambda_1)S_{n_j}(\omega)/n_j + \log(h(\theta)/h(\lambda_1))]\}\, dF(\theta) = \infty\,.$$

Hence we obtain a contradiction and we must have $\liminf_{n\to\infty} \lambda_n^{(1)}(\omega) \geqq \lambda$. A similar argument proves that $\limsup_{n\to\infty} \lambda_n^{(2)}(\omega) \leqq \lambda$. In the case where λ is a boundary point of Θ, we need only define $\mu^{-1}(S_n(\omega)/n) = \lambda$ if $S_n(\omega)/n \notin \Theta$ and observe that $\lambda_n^{(1)}(\omega) \leqq \mu^{-1}(S_n(\omega)/n) \leqq \lambda_n^{(2)}(\omega)$ still holds for all large n. □

5. Nuisance parameters and invariant confidence sequences. Suppose X_1, $X_2, \cdots$ are i.i.d. random variables with a common distribution function $F_{\theta,\sigma}$, $(\theta, \sigma) \in \Theta$. Let X_1 take values in $\mathscr{X}$ and let G be a group of transformations on $\mathscr{X}$ leaving the family $\{F_{\theta,\sigma}, (\theta, \sigma) \in \Theta\}$ invariant. Let $\bar{G}$ be the induced group on Θ. We assume that if $\bar{g}(\theta, \sigma) = (\theta', \sigma')$, then θ' depends only on $\bar{g}$ and θ and not on σ, so that $\bar{g}$ induces a transformation on the space of θ. Let $\{I_n(x_1, \cdots, x_n): n \geqq 1, x_i \in \mathscr{X}\}$ constitute a $(1 - \alpha)$-level sequence of confidence sets in the θ-space, i.e.,

$$P_{\theta,\sigma}[\theta \in I_n(X_1, \cdots, X_n) \text{ for all } n \geqq 1] \geqq 1 - \alpha \qquad \text{for all } (\theta, \sigma) \in \Theta\,.$$

For each transformation $g \in G$, denote by g^* the transformation acting on subsets I of the θ-space and defined by $g^*I = \{\bar{g}\theta : \theta \in I\}$. We say that $\{I_n(x_1, \cdots, x_n): n \geqq 1, x_i \in \mathscr{X}\}$ is invariant under G if $g^*I_n(x_1, \cdots, x_n) = I_n(gx_1, \cdots, gx_n)$ for all $n \geqq 1$, $x_i \in \mathscr{X}$ and $g \in G$.

In the rest of this section, we shall study the problem of finding confidence sequences for a location parameter in the presence of a nuisance scale parameter. Suppose $X_1, X_2, \cdots$ are i.i.d. random variables with probability density function $\sigma^{-1}g((x - \theta)/\sigma)$, $\sigma > 0$, $-\infty < \theta < \infty$. Let G be the group of scale changes on the real line. For each n, this group induces a group of transformations on

$R^n - (\{0\} \times R^{2n-1})$ of the form $(x_1, \cdots, x_n) \to (cx_1, \cdots, cx_n)$, $c > 0$, and a maximal invariant is $(x_1^*, \cdots, x_n^*)$, where $x_i^* = x_i/|x_1|$. (Note that $P_{\theta,\sigma}[X_1 = 0] = 0$.) The joint density of the maximal invariant $(X_1^*, \cdots, X_n^*)$ is $\int_0^\infty \prod_1^n \{(1/\tau)g(x_i^*/\tau - \theta/\sigma)\}\, d\tau/\tau$ $(x_1^* = \pm 1)$, which depends only on the maximal invariant θ/σ under the group $\bar{G}$. Define

$$(7) \qquad h_{\theta,n}(x_1, \cdots, x_n) = \frac{\int_0^\infty \prod_1^n \{(1/\tau)g(x_i/\tau - \theta)\}\, d\tau/\tau}{\int_0^\infty \prod_1^n \{(1/\tau)g(x_i/\tau)\}\, d\tau/\tau} = \frac{\int_0^\infty \prod_1^n \{(1/\tau)g(x_i^*/\tau - \theta)\}\, d\tau/\tau}{\int_0^\infty \prod_1^n \{(1/\tau)g(x_i^*/\tau)\}\, d\tau/\tau}\,.$$

Then for any θ, $\{h_{\theta,n}(X_1, \cdots, X_n), \mathscr{F}_n, n \geqq 1\}$ is a proper likelihood ratio martingale under $P_{0,\sigma}$ for any $\sigma > 0$, where $\mathscr{F}_n$ is the σ-field generated by $X_1^*, \cdots, X_n^*$.

Let F be a measure on the real line and define

$$(8) \qquad h_n(x_1, \cdots, x_n) = \int_{-\infty}^\infty h_{\theta,n}(x_1, \cdots, x_n)\, dF(\theta)\,.$$

Then $\{h_n(X_1, \cdots, X_n), \mathscr{F}_n, n \geqq 1\}$ is a generalized likelihood ratio martingale under $P_{0,\sigma}$ and so it follows from (1) that for any $\varepsilon > 0$,

$$(9) \qquad \begin{aligned} &P_{0,\sigma}[h_n(X_1, \cdots, X_n) \geqq \varepsilon \text{ for some } n \geqq m] \\ &\qquad \leqq P_{0,\sigma}[h_m(X_1, \cdots, X_m) \geqq \varepsilon] \\ &\qquad\qquad + \varepsilon^{-1} \int_{[h_m(X_1,\cdots,X_m)<\varepsilon]} h_m(X_1, \cdots, X_m)\, dP_{0,\sigma}\,. \end{aligned}$$

We note that in fact, both sides of the inequality (9) do not depend on σ. Given $\alpha \in (0, 1)$, we can choose F, ε and m so that the right-hand side of (9) is equal to α. Define

$$I_n(x_1, \cdots, x_n) = \{\theta \in (-\infty, \infty) : h_n(x_1 - \theta, \cdots, x_n - \theta) < \varepsilon\}\,.$$

Then $\{I_n(x_1, \cdots, x_n) : n \geqq m, -\infty < x_i < \infty\}$ constitutes a $(1 - \alpha)$-level sequence of confidence sets for the location parameter θ and is invariant under G.

As an example, suppose $X_1, X_2, \cdots$ are i.i.d. normal random variables with unknown mean μ and unknown variance σ^2. In this particular case, the martingale $h_{\theta,n}(X_1, \cdots, X_n)$ defined by (7) reduces to the following:

$$(10) \qquad Z_{\theta,n} = (\exp(-n\theta^2/2)/\Gamma(n/2)) \int_0^\infty y^{n/2-1} \exp\{-y + \theta S_n(2y/\textstyle\sum_1^n X_i^2)^{\frac{1}{2}}\}\, dy\,.$$

Let $a > 0$ and define for $n \geqq m$ $(\geqq 2)$

$$(11) \qquad \begin{aligned} &\bar{X}_n = \textstyle\sum_1^n X_i/n\,, \qquad v_n^2 = \sum_1^n (X_i - \bar{X}_n)^2/n\,, \\ &\lambda = m^{-1}(1 + a^2/(m-1))^m\,, \qquad \xi_n = v_n((\lambda n)^{1/n} - 1)^{\frac{1}{2}}\,, \\ &I_n = (\bar{X}_n - \xi_n, \bar{X}_n + \xi_n)\,. \end{aligned}$$

Setting $dF(\theta) = (m/2\pi)^{\frac{1}{2}}\, d\theta$, $-\infty < \theta < \infty$, the inequality (9) gives rise to

$$(12) \qquad P_{\mu,\sigma}[\mu \in I_n\ \forall\, n \geqq m] \geqq 1 - 2(1 - F_{m-1}(a) + af_{m-1}(a)) \qquad \forall\, \mu, \sigma\,,$$

where f_m and F_m denote respectively the Student t density and distribution function with m degrees of freedom, a result which was obtained by Robbins [12].

In [10], by using invariance considerations, we construct simultaneous confidence sequences for the contrasts among the unknown means of k normal populations with a common unknown variance and obtain selection and ranking procedures for these populations. Here we shall consider another problem in connection with k normal populations with unknown means and a common unknown variance. For simplicity, assume $k = 2$, so that we have $X_1, X_2, \cdots$ which are i.i.d. $N(\mu, \sigma^2)$ and $Y_1, Y_2, \cdots$ which are i.i.d. $N(\tilde{\mu}, \sigma^2)$. The problem is to find a confidence sequence for μ in terms of $X_1, Y_1, X_2, Y_2, \cdots$. Note that although the observations $Y_1, Y_2, \cdots$ do not give us information about μ, they give us information about the common unknown variance σ^2.

The invariance considerations which have led to the use of the martingale (8) in the construction of invariant confidence sequences above can now be extended to the present situation. For each n, a maximal invariant with respect to the group of transformations on $R^{2n} - (\{0\} \times R^{2n-1})$ of the form $(x_1, \cdots, x_n; y_1, \cdots, y_n) \to (cx_1, \cdots, cx_n; cy_1 + b, \cdots, cy_n + b)$, $c > 0$, b real, is $(x_1^*, \cdots, x_n^*; y_1', \cdots, y_n')$, where $x_i^* = x_i/|x_1|$, $y_i' = (y_i - y_1)/|x_1|$, $(i = 1, \cdots, n)$. In general, if $X_1, \cdots, X_n$ are i.i.d. with density function $\sigma^{-1}g((x - \theta)/\sigma)$ and are independent of $Y_1, \cdots, Y_n$ which are i.i.d. with density function $\sigma^{-1}g((y - \tilde{\theta})/\sigma)$, $\sigma > 0$, $\theta, \tilde{\theta} \in (-\infty, \infty)$, then the joint density of $X_1^*, \cdots, X_n^*, Y_1', \cdots, Y_n'$ is $p_{\theta/\sigma}(x_1^*, \cdots, x_n^*; y_1', \cdots, y_n')$, where $p_\lambda(a_1, \cdots, a_n; b_1, \cdots, b_n) = \int_0^\infty \int_{-\infty}^\infty \prod_1^n \{(1/\rho)g(a_i/\rho - \lambda)\} \prod_1^n \{(1/\rho)g(b_i/\rho - \mu)\}\, d\mu\, d\rho$. $(X_1^* = \pm 1, Y_1' = 0)$. Hence for any θ and any measure F on $(-\infty, \infty)$, if we define

$$U_{\theta,n} = \frac{p_\theta(X_1^*, \cdots, X_n^*; Y_1', \cdots, Y_n')}{p_0(X_1^*, \cdots, X_n^*; Y_1', \cdots, Y_n')} = \frac{p_\theta(X_1, \cdots, X_n; Y_1, \cdots, Y_n)}{p_0(X_1, \cdots, X_n; Y_1, \cdots, Y_n)},$$

$$U_n = \int_{-\infty}^\infty U_{\theta,n}\, dF(\theta),$$

then $U_{\theta,n}$ and U_n are martingales with respect to the σ-fields $\mathscr{G}_n$ (generated by $X_1^*, \cdots, X_n^*, Y_1', \cdots, Y_n'$) under $P_{0,\tilde{\theta},\sigma}$ for any $\tilde{\theta} \in (-\infty, \infty)$ and $\sigma > 0$. In the particular case where $g(x) = (2\pi)^{-\frac{1}{2}} \exp(-x^2/2)$ and $dF(\theta) = (m/2\pi)^{\frac{1}{2}}\, d\theta$, $-\infty < \theta < \infty$, we have

$$U_n = (m/n)^{\frac{1}{2}}\{1 + n\bar{X}_n^2/(\sum_1^n (X_i - \bar{X}_n)^2 + \sum_1^n (Y_i - \bar{Y}_n)^2)\}^{(2n-1)/2},$$

where $\bar{X}_n = \sum_1^n X_i/n$ and $\bar{Y}_n = \sum_1^n Y_i/n$. Generalizing to k normal populations with the same variance, we have the following theorem:

THEOREM 2. *Suppose $X_n^{(i)}$, $i = 1, \cdots, k$; $n = 1, 2, \cdots$, are independent normal random variables such that $X_n^{(i)}$ is normally distributed with mean μ_i and variance σ^2. Let $a > 0$ and define for $n \geqq m$,*

$$\bar{X}_n^{(i)} = \sum_{r=1}^n X_r^{(i)}/n, \qquad v_n^2 = \sum_{i=1}^k \sum_{r=1}^n (X_r^{(i)} - \bar{X}_n^{(i)})^2/kn,$$

$$t = m^{-1}(1 + a^2/k(m-1))^{k(m-1)+1}, \qquad \xi_n = v_n[(tn)^{1/(k(n-1)+1)} - 1]^{\frac{1}{2}},$$

$$I_n^{(i)} = (\bar{X}_n^{(i)} - \xi_n, \bar{X}_n^{(i)} + \xi_n).$$

Then for each $i = 1, \cdots, k$,

$$P[\mu_i \notin I_n^{(i)} \text{ for some } n \geqq m] \leqq 2(1 - F_{k(m-1)}(a) + af_{k(m-1)}(a)),$$

where $f_{k(m-1)}$ and $F_{k(m-1)}$ denote respectively the Student t density and distribution function with $k(m-1)$ degrees of freedom.

6. Boundary crossing probabilities for the sequence of Student's t-statistics. Suppose $X_1, X_2, \cdots$ are i.i.d. normal random variables with mean 0 and variance 1. In this section, we consider the sequence t_n of Student's t-statistics:

$$t_{n-1} = (n-1)^{\frac{1}{2}}\bar{X}_n/v_n$$

where $\bar{X}_n$ and v_n are defined by (11). Let F be a measure on $(0, \infty)$ such that $F(0, \infty) > 0$ and let $Z_{\theta,n}$ be defined by (10). The martingale $Z_n = \int_0^\infty Z_{\theta,n}\, dF(\theta)$ can be used to study boundary crossing probabilities for the sequence t_n. Define

$$(13) \qquad \begin{aligned} \psi_\theta(x, n) &= \int_0^\infty y^{n/2-1} \exp(\theta x(2y)^{\frac{1}{2}} - y - (n/2)\theta^2)\, dy/\Gamma(n/2)\,, \\ \psi(x, n) &= \int_0^\infty \psi_\theta(x, n)\, dF(\theta)\,. \end{aligned}$$

Let m be an integer $\geqq 2$. Suppose that $\psi(x, m) < \infty$ for all x. Then for $n \geqq m$, $\psi(x, n) < \infty$ for all x (see Lemma 1 below) and so given $\varepsilon > 0$, the equation $\psi(x, n) = \varepsilon$ has a unique solution $x = B_F(n, \varepsilon)$. We note that

$$\begin{aligned} Z_n = \psi(S_n(\textstyle\sum_1^n X_i^2)^{-\frac{1}{2}}, n) \geqq \varepsilon \\ &\Leftrightarrow S_n(\textstyle\sum_1^n X_i^2)^{-\frac{1}{2}} \geqq B_F(n, \varepsilon) \\ &\Leftrightarrow S_n/v_n \geqq n^{\frac{1}{2}} B_F(n, \varepsilon)\{1 - B_F^2(n, \varepsilon)/n\}^{-\frac{1}{2}} \\ &\Leftrightarrow t_{n-1} \geqq B_F(n, \varepsilon)\{(n-1)/n\}^{\frac{1}{2}}\{1 - B_F^2(n, \varepsilon)/n\}^{-\frac{1}{2}}\,. \end{aligned}$$

Set $B_F^*(n, \varepsilon) = B_F(n, \varepsilon)\{(n-1)/n\}^{\frac{1}{2}}\{1 - B_F^2(n, \varepsilon)/n\}^{-\frac{1}{2}}$. Then it follows from (1) that

$$(14) \qquad \begin{aligned} &P[t_{n-1} \geqq B_F^*(n, \varepsilon) \text{ for some } n \geqq m] \\ &\quad \leqq P[t_{m-1} \geqq B_F^*(m, \varepsilon)] + \varepsilon^{-1} \int_{[t_{m-1} < B_F^*(m,\varepsilon)]} Z_m\, dP\,. \end{aligned}$$

The above argument generalizes in an obvious manner to give boundary crossing probabilities for the sequence $|t_n|$ if F is a symmetric measure on $(-\infty, \infty)$ which assigns measure 0 to $\{0\}$. In particular, for $dF(\theta) = (m/2\pi)^{\frac{1}{2}}\, d\theta$, $-\infty < \theta < \infty$, we obtain that for any $a > 0$,

$$(15) \qquad \begin{aligned} &P[|t_{n-1}| \geqq (n-1)^{\frac{1}{2}}\{(n/m)^{1/n}(1 + a^2/(m-1))^{m/n} - 1\}^{\frac{1}{2}} \text{ for some } n \geqq m] \\ &\quad \leqq 2(1 - F_{m-1}(a) + af_{m-1}(a)) \end{aligned}$$

(cf. (12)). As shown by Robbins [12], the boundary in (15) is asymptotic to $(\log n)^{\frac{1}{2}}$ as $n \to \infty$.

Since $v_n \to 1$ a.s., the law of the iterated logarithm implies that

$$\limsup_{n\to\infty} t_{n-1}/(2 \log_2 n)^{\frac{1}{2}} = 1 \quad \text{a.s.}$$

where $\log_2 n$ denotes $\log\log n$, $\log_3 n$ denotes $\log(\log_2 n)$, etc. It is therefore natural to ask whether by choosing F suitably in (13), we can obtain a boundary of the order of magnitude $(2 \log_2 n)^{\frac{1}{2}}$ as $n \to \infty$. Take any $\delta > 0$ and choose the measure

$$dH(\theta) = d\theta/\{\theta(\log|\theta|)(\log_2|\theta|)^{1+\delta}\}\,, \qquad 0 < \theta < e^{-e}$$
$$= 0 \qquad \text{elsewhere.}$$

We assert that the boundary $B_H^*(n, \varepsilon)$ corresponding to this measure H has the following asymptotic expansion:

$$B_H^*(n, \varepsilon) = \left\{2\left[\log_2 n + (\tfrac{3}{2} + \delta)\log_3 n + \log\frac{\varepsilon}{2(\pi)^{\frac12}} + o(1)\right]\right\}^{\frac12}.$$

A proof of the above assertion can be found in [10].

We now proceed to prove a limit theorem involving the boundary crossing probabilities of the martingale $Z_n = \int_0^\infty Z_{\theta,n}\,dF(\theta)$ from which we have obtained boundary crossing probabilities of the sequence of t-statistics. First we note that Z_n can be written as $\varphi_n(S_n/(\sum_1^n X_i^2/n)^{\frac12}, n)$, where we define

$$\begin{aligned} &g_n(y) = n^{n/2}\int_0^\infty z^{n/2-1}\exp\left(y(2z)^{\frac12} - nz\right)dz/\Gamma(n/2)\,,\\ (16)\qquad &\varphi_n(x, t) = \int_0^\infty \exp\left(-t\theta^2/2\right)g_n(x\theta)\,dF(\theta)\,,\\ &f(x, t) = \int_0^\infty \exp\left(x\theta - \frac{t}{2}\theta^2\right)dF(\theta)\,. \end{aligned}$$

For any $m > 0$, if we replace $F(\theta)$ by $F(\theta(m)^{\frac12})$, we then obtain

$$(17)\qquad \int_0^\infty Z_{\theta,n}\,dF(\theta(m)^{\frac12}) = \varphi_n(m^{-\frac12}S_n/(\textstyle\sum_1^n X_i^2/n)^{\frac12}, n/m)\,.$$

Let $W(t)$, $t \geqq 0$, be the standard Wiener process. By Lemma 2 below, $g_n(x\theta) \to e^{x\theta}$ as $n \to \infty$. Since for large m, the two sequences $\{m^{-\frac12}S_n/((1/n)\sum_1^n X_i^2)^{\frac12}, n \geqq \tau m\}$ and $\{W(n/m), n \geqq \tau m\}$ have approximately the same joint distribution, this heuristic argument suggests the following theorem on the Wiener-process approximation for the boundary crossing probabilities of the sequence of t-statistics.

THEOREM 3. *Suppose $X_1, X_2, \cdots$ are* i.i.d. *random variables with mean* 0 *and variance* 1. *Let F be any measure on* $(0, \infty)$ *such that for some $h \geqq 0$, $k \geqq 1$ and all real x,*

$$(18)\qquad \int_0^\infty\int_0^\infty \exp\left(x\theta(y)^{\frac12} - ky - \frac{h}{2}\theta^2\right)dy\,dF(\theta) < \infty\,.$$

For $\theta > 0$, define $Z_{\theta,n}$ by (10). *Let $\varepsilon > 0$ and $\tau > h$.*

(i) *For any $\lambda > \tau$, $\lim_{m\to\infty} P[\int_0^\infty Z_{\theta,n}\,dF(\theta(m)^{\frac12}) \geqq \varepsilon$ for some $\lambda m \geqq n \geqq \tau m] = P[\int_0^\infty \exp(\theta W(t) - (t/2)\theta^2)\,dF(\theta) \geqq \varepsilon$ for some $\lambda \geqq t \geqq \tau]$.*

(ii) *Let $A_F(t, \delta) = \inf\{x : \int_0^\infty e^{\theta x - (t/2)\theta^2}\,dF(\theta) \geqq \delta\}$. If either X_1 is normal or there exists $0 < \delta < \varepsilon$ such that $\liminf_{t\to\infty}(2t\log\log t)^{-\frac12}A_F(t, \delta) > 1$, then $\lim_{m\to\infty} P[\int_0^\infty Z_{\theta,n}\,dF(\theta(m)^{\frac12}) \geqq \varepsilon$ for some $n \geqq \tau m] = P[\int_0^\infty \exp(\theta W(t) - (t/2)\theta^2)dF(\theta) \geqq \varepsilon$ for some $t \geqq \tau]$.*

LEMMA 1. *Let g_n, φ_n and f be as defined in* (16). *Assume* (18). *For any $x \in (-\infty, \infty)$ and $t > h$, $f(x, t) < \infty$ and*

$$\sup_{n>k}\varphi_n(x, t) \leqq \int_0^\infty (\sup_{n>k} g_n(x\theta))\exp(-t\theta^2/2)\,dF(\theta) < \infty\,.$$

PROOF. The fact that $f(x, t) < \infty$ for $t > h$ follows easily from (18). To prove the remaining part of the theorem, we note that the function $\exp(\theta x(2z)^{\frac{1}{2}} - (n/2)z)$ is decreasing in z for $z \geqq 2\theta^2 x^2/n^2$. In particular if $\theta x \leqq n/2$, this function is decreasing in z for $z \geqq \frac{1}{2}$. Therefore if $\theta x \leqq n$ and $n \geqq 2$, we have

$$\begin{aligned}(19)\qquad &\int_0^\infty z^{n/2-1} \exp(\theta x(2z)^{\frac{1}{2}} - nz)\, dz \\ &\quad \leqq e^{\theta x} \int_0^{\frac{1}{2}} e^{-nz} z^{n/2-1}\, dz + \int_{\frac{1}{2}}^\infty e^{-z} \exp(\theta x(2z)^{\frac{1}{2}} - (n/2)z)\, dz \\ &\quad \leqq e^{\theta x} n^{-n/2}\Gamma(n/2) + e^{\theta x - n/4}\,.\end{aligned}$$

On the other hand, if $\theta x > n/2$ and $n \geqq 2$, we have

$$\begin{aligned}(20)\qquad &\int_0^\infty z^{n/2-1} \exp(\theta x(2z)^{\frac{1}{2}} - nz)\, dz \\ &\quad \leqq \exp(6\theta^2 x^2/n) \int_0^{18\theta^2 x^2/n^2} e^{-nz} z^{n/2-1}\, dz \\ &\qquad + \int_{18\theta^2 x^2/n^2}^\infty e^{-z} \exp(\theta x(2z)^{\frac{1}{2}} - (n/2)z)\, dz \\ &\quad \leqq n^{-n/2}\Gamma(n/2) \exp(6\theta^2 x^2/n) + e^{-3n/4}\,.\end{aligned}$$

The last inequality above makes use of the fact that $\theta x > n/2$ and the monotonicity of $\exp(\theta x(2z)^{\frac{1}{2}} - (n/2)z)$. It then follows from (19) and (20) that

$$(21)\qquad g_n(x\theta) \leqq n^{n/2}(e^{\theta x - n/4} + e^{-3n/4})/\Gamma(n/2) + e^{\theta x} + \exp(6\theta^2 x^2/n)\,.$$

The desired conclusion follows easily from (21). □

LEMMA 2. *As $n \to \infty$, $g_n(y) \to e^y$, the convergence being uniform for y belonging to any compact subset of the real line.*

PROOF. Apply Laplace's asymptotic formula (cf. [15]). □

LEMMA 3. *With the same assumption as in Theorem 3, in the case where the measure F has bounded support, we have for any $\lambda > \tau$,*

$$\max_{\lambda \geqq t \geqq \tau} \varphi_{[mt]}(m^{-\frac{1}{2}} S_{[mt]}/(\textstyle\sum_1^{[mt]} X_i^2/[mt])^{\frac{1}{2}}, [mt]/m) \to_{\mathscr{L}} \max_{\tau \leqq t \leqq \lambda} f(W(t), t)$$

as $m \to \infty$, where "$\to_{\mathscr{L}}$" denotes convergence in distribution.

PROOF. Since $n^{-1} \sum_1^n X_i^2 \to 1$ a.s., given any sequence of positive numbers $m_\nu \uparrow \infty$, we can construct processes $\{X^{(\nu)}(t), t \geqq \tau\}$, $\nu = 1, 2, \cdots$, having for each ν the same distribution as $\{m_\nu^{-\frac{1}{2}} S_{[m_\nu t]}/(\sum_1^{[m_\nu t]} X_i^2/[m_\nu t])^{\frac{1}{2}}, t \geqq \tau\}$, defined on a common probability space Ω, and a standard Wiener process $\{W(t), t \geqq \tau\}$ on the same space, such that for any subsequence ν_j increasing rapidly enough, $\max_{\tau \leqq t \leqq \lambda} |X^{(\nu_j)}(t) - W(t)| \to 0$ a.s. as $j \to \infty$ (cf. [1] page 279). Given any $\rho > 0$, we can choose j_0 such that if $\tau \leqq t \leqq \lambda'$ and $j \geqq j_0$, $|X^{(\nu_j)}(t, \omega) - W(t, \omega)| < \rho$. Since $g_n(y)$ is increasing in y, we have for all $\theta \geqq 0$,

$$\begin{aligned} g_{[m_{\nu_j} t]}(\theta W(t, \omega) - \theta\rho) &\leqq g_{[m_{\nu_j} t]}(\theta X^{(\nu_j)}(t, \omega)) \\ &\leqq g_{[m_{\nu_j} t]}(\theta W(t, \omega) + \theta\rho)\,. \end{aligned}$$

Now F has bounded support, say $F[c, \infty) = 0$. By Lemma 2, we can choose n_0 such that

$$\begin{aligned} &n \geqq n_0 \quad \text{and} \\ &\quad \min_{\tau \leqq t \leqq \lambda, 0 \leqq \theta \leqq c} (\theta W(t, \omega) - \theta\rho) \leqq y \leqq \max_{\tau \leqq t \leqq \lambda, 0 \leqq \theta \leqq c} (\theta W(t, \omega) + \theta\rho) \\ &\qquad \Rightarrow \exp(-\rho + y) \leqq g_n(y) \leqq \exp(\rho + y)\,. \end{aligned}$$

Now choose $j_1 \geqq j_0$ such that $[m_{\nu_j}\tau] > n_0$ for all $j \geqq j_1$. Hence if $\tau \leqq t \leqq \lambda$, $j \geqq j_1$, then for all $\theta \in [0, c]$,

$$\begin{aligned}\exp(\theta W(t,\omega) - (c+1)\rho) &\leqq g_{[m_{\nu_j}t]}(\theta X^{(\nu_j)}(t,\omega)) \\ &\leqq \exp(\theta W(t,\omega) + (c+1)\rho)\,.\end{aligned}$$

Therefore $g_{[m_{\nu_j}t]}(\theta X^{(\nu_j)}(t,\omega)) \to \exp(\theta W(t,\omega))$ as $j \to \infty$, the convergence being uniform for $\theta \in [0, c]$ and $t \in [\tau, \lambda]$. This implies that as $j \to \infty$,

$$\varphi_{[m_{\nu_j}t]}(X^{(\nu_j)}(t,\omega), [m_{\nu_j}t]/m_{\nu_j}) \to f(W(t,\omega), t)\,,$$

the convergence being uniform for $\tau \leqq t \leqq \lambda$. □

PROOF OF THEOREM 3. For $c > 0$, define

$$\begin{aligned}\varphi_n^c(x,t) &= \textstyle\int_{(0,c)} \exp(-t\theta^2/2) g_n(x\theta)\, dF(\theta)\,; \\ \tilde\varphi_n^c(x,t) &= \textstyle\int_{[c,\infty)} \exp(-t\theta^2/2) g_n(x\theta)\, dF(\theta)\,.\end{aligned}$$

Define g_n, φ_n and f as in (16). By Lemma 3,

$$\begin{aligned}&P[\varphi_n(m^{-\frac12}S_n/(\textstyle\sum_1^n X_i^2/n)^{\frac12}, n/m) \geqq \varepsilon \text{ for some } \lambda m \geqq n \geqq \tau m] \\ (22)\quad &\geqq P[\max_{\tau\leqq t\leqq\lambda} \varphi^c_{[mt]}(m^{-\frac12}S_{[mt]}/(\textstyle\sum_1^{[mt]} X_i^2/[mt])^{\frac12}, [mt]/m) \geqq \varepsilon] \\ &\to P\left[\max_{\tau\leqq t\leqq\lambda} \textstyle\int_{(0,c)} \exp\left(\theta W(t) - \dfrac{t}{2}\theta^2\right) dF(\theta) \geqq \varepsilon\right] \quad \text{as} \quad m\to\infty \\ &(\to P[f(W(t), t) \geqq \varepsilon \text{ for some } \tau \leqq t \leqq \lambda] \text{ as } c \to \infty).\end{aligned}$$

We shall now prove the following fact: For every $\gamma > 0$,

$$(23)\quad \begin{aligned}\lim_{c\to\infty} \limsup_{m\to\infty} P[\tilde\varphi_n^c(m^{-\frac12}S_n/(\textstyle\sum_1^n X_i^2/n)^{\frac12}, n/m) \geqq \gamma \\ \text{for some } n \geqq \tau m] = 0\,.\end{aligned}$$

Using (21), we obtain that on the event $[|S_n| < (1-h/\tau)^{\frac12}n/5$ and $\sum_1^n X_i^2/n > \frac12]$,

$$(24)\quad g_n(\theta m^{-\frac12}S_n/(\textstyle\sum_1^n X_i^2/n)^{\frac12}) \leqq 2\{\exp(2\theta m^{-\frac12}S_n) + \exp(\frac12(1-h/\tau)\theta^2 n/m)\}$$

for $n \geqq n_0$. Since $P_m = P[|S_n| \geqq (1-h/\tau)^{\frac12}n/5$ or $\sum_1^n X_i^2/n \leqq \frac12$ for some $n \geqq \tau m] \to 0$ as $m \to \infty$, it then follows from (24) that for $c \geqq c_0$,

$$\begin{aligned}&P[\tilde\varphi_n^c(m^{-\frac12}S_n/(\textstyle\sum_1^n X_i^2/n)^{\frac12}, n/m) \geqq \gamma \text{ for some } n \geqq \tau m] \\ &\quad\leqq P[\textstyle\int_{[c,\infty]} \exp(2\theta m^{-\frac12}S_n - n\theta^2/2m)\, dF(\theta) \geqq \gamma/4 \text{ for some } n \geqq \tau m] + P_m \\ &\quad\to P\left[\textstyle\int_{[c,\infty]} \exp\left(2\theta W(t) - \dfrac{t}{2}\theta^2\right) dF(\theta) \geqq \gamma/4 \text{ for some } t \geqq \tau\right] \\ &\qquad\qquad \text{as} \quad m \to \infty\,.\end{aligned}$$

The last relation above follows from Theorem 2 of [14] (see Remark (b) on page 1411), since $A_F^c(t, \gamma/4) \geqq \frac12 ct(1 + o(1))$ as $t \to \infty$ by Theorem 1 of [8], where $A_F^c(t,\gamma/4) = \inf\{x: \int_{[c,\infty)} \exp(\theta x - (t/2)\theta^2)\, dF(\theta) \geqq \gamma/4\}$. Since

$$\lim_{c\to\infty} P\left[\textstyle\int_{[c,\infty)} \exp\left(4\theta W(t) - \dfrac{t}{2}\theta^2\right) dF(\theta) \geqq \gamma/4 \text{ for some } t \geqq \tau\right] = 0\,,$$

we obtain (23). Using (22) and (23), it is easy to see (i).

To prove (ii), first assume that $\liminf_{t\to\infty} (2t \log\log t)^{-\frac{1}{2}} A_F(t, \delta) > 1$ for some $0 < \delta < \varepsilon$. Let $\delta \leqq \rho < (1+\eta)\rho \leqq \varepsilon$. Then there exists $0 < \zeta < 1$ such that

$$\liminf_{t\to\infty} (2t \log\log t)^{-\frac{1}{2}} \zeta A_F(t, \rho) > 1 . \tag{25}$$

Using (21), we obtain that on the event $[|S_n| < \zeta n/5, (\sum_1^n X_i^2/n)^{\frac{1}{2}} > \zeta]$,

$$g_n(\theta m^{-\frac{1}{2}} S_n/(\textstyle\sum_1^n X_i^2/n)^{\frac{1}{2}}) \leqq \left(1 + \frac{\eta}{2}\right)\{\exp(\theta m^{-\frac{1}{2}} S_n/\zeta) + \exp(\tfrac{1}{4}\theta^2 \cdot n/m)\} \tag{26}$$

for $n \geqq n_1$. From (26), it follows that for $\lambda \geqq \lambda_0$,

$$\begin{aligned} &P[\varphi_n(m^{-\frac{1}{2}} S_n/(\textstyle\sum_1^n X_i^2/n)^{\frac{1}{2}}, n/m) \geqq (1+\eta)\rho \ \text{ for some } \ n \geqq \lambda m \ \text{ and} \\ &\quad |S_n| < \zeta n/5, (\textstyle\sum_1^n X_i^2/n)^{\frac{1}{2}} > \zeta \ \text{ for all } \ n \geqq \lambda m] \\ &\qquad \leqq P[\textstyle\int_0^\infty \exp(\theta m^{-\frac{1}{2}} S_n/\zeta - n\theta^2/2m)\, dF(\theta) \geqq \rho \ \text{ for some } \ n \geqq \lambda m] \\ &\qquad = P[m^{-\frac{1}{2}} S_n \geqq \zeta A_F(n/m, \rho) \ \text{ for some } \ n \geqq \lambda m] \\ &\qquad \to P[W(t) \geqq \zeta A_F(t, \rho) \ \text{ for some } \ t \geqq \lambda] \quad \text{ as } \quad m \to \infty . \end{aligned}$$

The last relation above follows from Theorem 2 of [14] since (25) holds. Since $\lim_{\lambda\to\infty} P[W(t) \geqq \zeta A_F(t, \rho)$ for some $t \geqq \lambda] = 0$, we obtain

$$\begin{aligned} \lim_{\lambda\to\infty} \limsup_{m\to\infty} P[\varphi_n(m^{-\frac{1}{2}} S_n/(\textstyle\sum_1^n X_i^2/n)^{\frac{1}{2}}, n/m) \geqq (1+\eta)\rho \\ \text{for some } \ n \geqq \lambda m] = 0 . \end{aligned} \tag{27}$$

Noting that

$$\begin{aligned} &P[\varphi_n(m^{-\frac{1}{2}} S_n/(\textstyle\sum_1^n X_i^2)^{\frac{1}{2}}, n/m) \geqq \varepsilon \ \text{ for some } \ n \geqq \tau m] \\ &\quad \leqq P[\varphi_n(m^{-\frac{1}{2}} S_n/(\textstyle\sum_1^n X_i^2)^{\frac{1}{2}}, n/m) \geqq \varepsilon \ \text{ for some } \ \lambda m \geqq n \geqq \tau m] \\ &\qquad + P[\varphi_n(m^{-\frac{1}{2}} S_n/(\textstyle\sum_1^n X_i^2)^{\frac{1}{2}}, n/m) \geqq (1+\eta)\rho \ \text{ for some } \ n \geqq \lambda m] , \end{aligned}$$

we obtain (ii) from (i) and (27).

We now prove (ii) under the alternative assumption that X_1 is normal. In this case, $\int_0^\infty Z_{\theta,n}\, dF(\theta(m)^{\frac{1}{2}})$ is a martingale and so it follows from (1) that

$$\begin{aligned} &P[\varphi_n(m^{-\frac{1}{2}} S_n/(\textstyle\sum_1^n X_i^2/n)^{\frac{1}{2}}, n/m) \geqq \varepsilon \ \text{ for some } \ n \geqq \tau m] \\ &\quad \leqq P[\varphi_{[\tau m]}(m^{-\frac{1}{2}} S_{[\tau m]}/(\textstyle\sum_1^{[\tau m]} X_i^2/[\tau m])^{\frac{1}{2}}, [\tau m]/m) \geqq \varepsilon] \\ &\qquad + \varepsilon^{-1} \textstyle\int_{[\varphi_{[\tau m]}(\ \ )<\varepsilon]} \varphi_{[\tau m]}(\ \ )\, dP . \end{aligned} \tag{28}$$

Since $\varphi_{[\tau m]}(m^{-\frac{1}{2}} S_{[\tau m]}/(\sum_1^{[\tau m]} X_i^2/[\tau m])^{\frac{1}{2}}, [\tau m]/m) \to_{\mathscr{L}} f(W(\tau), \tau)$, (see the proof of Lemma 3), we obtain from (28) that

$$\begin{aligned} &\limsup_{n\to\infty} P[\varphi_n(m^{-\frac{1}{2}} S_n/(\textstyle\sum_1^n X_i^2/n)^{\frac{1}{2}}, n/m) \geqq \varepsilon \ \text{ for some } \ n \geqq \tau m] \\ &\quad \leqq P[f(W(\tau), \tau) \geqq \varepsilon] + \varepsilon^{-1} \textstyle\int_{[f(W(\tau),\tau)<\varepsilon]} f(W(\tau), \tau)\, dP \\ &\quad = P[f(W(t), t) \geqq \varepsilon \ \text{ for some } \ t \geqq \tau] . \end{aligned}$$

The last relation above is due to Robbins and Siegmund [14]. □

7. Acknowledgment. The author wishes to thank Professor David O. Siegmund for his guidance and encouragement. Thanks are also due to Professor Herbert E. Robbins for several valuable suggestions, and to Professor W. J. Hall and the referee for their helpful comments.

REFERENCES

[1] Breiman, L. (1968). *Probability*. Addison-Wesley, Reading, Mass.
[2] Darling, D. A. and Robbins, H. (1967). Iterated logarithm inequalities. *Proc. Nat. Acad. Sci. U.S.A.* **57** 1188-1192.
[3] Darling, D. A. and Robbins, H. (1967). Inequalities for the sequence of sample means. *Proc. Nat. Acad. Sci. U.S.A.* **57** 1577-1880.
[4] Darling, D. A. and Robbins, H. (1967). Confidence sequences for mean, variance and median. *Proc. Nat. Acad. Sci. U.S.A.* **58** 66-68.
[5] Darling, D. A. and Robbins, H. (1968). Some further remarks on inequalities for sample sums. *Proc. Nat. Acad. Sci. U.S.A.* **60** 1175-1182.
[6] Fabian, V. (1956). A decision function. *Czechoslovak Math. J.* **6** 31-41.
[7] Farrell, R. H. (1964). Asymptotic behavior of the expected sample size in certain one sided tests. *Ann. Math. Statist.* **35** 36-72.
[8] Lai, T. L. (1976). Boundary crossing probabilities for sample sums and confidence sequences. To appear in *Ann. Probability*, April 1976.
[9] Lai, T. L. (1976). Power-one tests based on sample sums. To appear.
[10] Lai, T. L. (1971). Confidence sequences and martingales. Ph.D. dissertation, Columbia Univ.
[11] Paulson, E. (1964). Sequential estimation and closed sequential decision procedures. *Ann. Math. Statist.* **35** 1048-1058.
[12] Robbins, H. (1970). Statistical methods related to the law of the iterated logarithm. *Ann. Math. Statist.* **41** 1397-1409.
[13] Robbins, H. and Siegmund, D. (1968). Iterated logarithm and related statistical procedures. *Mathematics of the Decision Sciences* **2** Lectures in Appl. Math., Amer. Math. Soc.
[14] Robbins, H. and Siegmund, D. (1970). Boundary crossing probabilities for the Wiener process and sample sums. *Ann. Math. Statist.* **41** 1410-1429.
[15] Widder, D. V. (1946). *The Laplace Transform*. Princeton Univ. Press.

Department of Mathematical Statistics
Columbia University
New York, N.Y. 10027