The Annals of Mathematical Statistics
1970, Vol. 41, No. 5, 1397–1409

STATISTICAL METHODS RELATED TO THE LAW OF THE ITERATED LOGARITHM[1]

BY HERBERT ROBBINS

Columbia University

1. Extension and applications of an inequality of Ville and Wald. Let $x_1, \cdots$ be a sequence of random variables with a specified joint probability distribution P. We shall give a method for obtaining probability inequalities and related limit theorems concerning the behavior of the entire sequence of x's.

We begin with a result of J. Ville ([16] page 100); cf. also A. Wald ([17] page 146).

Suppose that under P for each $n \geqq 1$ the random variables $x_1, \cdots, x_n$ have a probability density function $g_n(x_1, \cdots, x_n)$ with respect to a σ-finite measure μ_n on the Borel sets of n-space, and that P' is any other joint probability distribution of the sequence $x_1, \cdots$ such that $x_1, \cdots, x_n$ have a probability density function $g_n'(x_1, \cdots, x_n)$ with respect to the same μ_n. Define the likelihood ratio $z_n = g_n'/g_n$ when $g_n > 0$. Then for any $\varepsilon > 1$,

$$P(z_n \geqq \varepsilon \text{ for some } n \geqq 1) \leqq 1/\varepsilon. \tag{1}$$

To prove this, let $N =$ first $n \geqq 1$ such that $g_n' \geqq \varepsilon g_n$, with $N = \infty$ if no such n occurs; then $P(g_n = 0 \text{ for some } n \geqq 1) = 0$ and

$$\begin{aligned} P(z_n \geqq \varepsilon \text{ for some } n \geqq 1) &= P(N < \infty) = \textstyle\sum_1^\infty \int_{(N=n)} g_n \, d\mu_n \\ &\leqq 1/\varepsilon \textstyle\sum_1^\infty \int_{(N=n)} g_n' \, d\mu_n \\ &= 1/\varepsilon \cdot P'(N < \infty) \leqq 1/\varepsilon. \end{aligned} \tag{2}$$

As a first example, cf. ([16] page 52), suppose that the x's are i.i.d. Bernoulli random variables such that $P(x_i = 1) = p$, $P(x_i = 0) = 1-p$, $0 < p < 1$. If μ_n is counting measure on the space of vectors $(x_1, \cdots, x_n)$ then $g_n(x_1, \cdots, x_n) = P(x_1, \cdots, x_n) = p^{S_n}(1-p)^{n-S_n}$, where $S_n = x_1 + \cdots + x_n$. Take P' to be the uniform mixture of Bernoulli distributions with parameter $0 < \theta < 1$; then

$$\begin{aligned} g_n'(x_1, \cdots, x_n) = P'(x_1, \cdots, x_n) &= \textstyle\int_0^1 \theta^{S_n}(1-\theta)^{n-S_n} \, d\theta \\ &= \frac{S_n!(n-S_n)!}{(n+1)!}. \end{aligned}$$

Using the notation $b(n, p, x) = \binom{n}{x} p^x (1-p)^{n-x}$ and replacing ε by $1/\varepsilon$, (1) gives the inequality

$$P(b(n, p, S_n) \leqq \varepsilon/(n+1) \text{ for some } n \geqq 1) \leqq \varepsilon \quad (0 < \varepsilon < 1).$$

Received September 8, 1969.

[1] Invited Wald Lectures given at the annual meeting of the Institute of Mathematical Statistics New York, August 19–21, 1969; research supported by N. I. H. Grant 1-R01-GM-16895-01.

If we denote by $I_n(x, \varepsilon)$ the set of all $0 \leqq \theta \leqq 1$ such that $b(n, \theta, x) > \varepsilon/(n+1)$, then this is equivalent to the "confidence" statement

$$P(p \in I_n(S_n, \varepsilon) \text{ for every } n \geqq 1) \geqq 1-\varepsilon \qquad (0 < \varepsilon, p < 1). \tag{3}$$

As a second example, which we shall study in more detail, of the use of (1), let P_θ, $-\infty < \theta < \infty$, denote the probability under which the x's are i.i.d. $N(\theta, 1)$, and take $P = P_0$. If μ_n denotes Lebesgue measure in n-space and

$$\varphi(x) = (2\pi)^{-\frac{1}{2}} \exp(-x^2/2), \qquad \Phi(x) = \textstyle\int_{-\infty}^{x} \varphi(t)\,dt, \qquad S_n = x_1 + \cdots + x_n,$$

then the joint density of $x_1, \cdots, x_n$ under P_θ is

$$g_{\theta,n}(x_1, \cdots, x_n) = \textstyle\prod_1^n \varphi(x_i - \theta). \tag{4}$$

We shall take P' to be any mixture of the form $P'(\cdot) = \int_{-\infty}^{\infty} P_\theta(\cdot)\,dF(\theta)$, where F is an arbitrary probability measure on $(-\infty, \infty)$, so that

$$g_n'(x_1, \cdots, x_n) = \textstyle\int_{-\infty}^{\infty} g_{\theta,n}(x_1, \cdots, x_n)\,dF(\theta)$$
$$= \textstyle\int_{-\infty}^{\infty} \prod_1^n \varphi(x_i - \theta)\,dF(\theta),$$

$$z_n = g_n'/g_{0,n} = \textstyle\int_{-\infty}^{\infty} \exp(\theta S_n - \frac{1}{2}n\theta^2)\,dF(\theta). \tag{5}$$

If we define the function

$$f(x, t) = \textstyle\int_{-\infty}^{\infty} \exp(xy - \frac{1}{2}y^2 t)\,dF(y) \tag{6}$$

and for reasons which will become apparent in the next section replace $F(\theta)$ in (5) by $F(\theta m^{\frac{1}{2}})$, where m is an arbitrary positive constant, then (5) becomes

$$z_n = \textstyle\int_{-\infty}^{\infty} \exp(\theta S_n - \frac{1}{2}n\theta^2)\,dF(\theta m^{\frac{1}{2}}) = \int_{-\infty}^{\infty} \exp(yS_n/m^{\frac{1}{2}} - \frac{1}{2}ny^2/m)\,dF(y) \tag{7}$$
$$= f(S_n/m^{\frac{1}{2}}, n/m),$$

and hence by (1) we have hat for i.i.d. $N(0, 1)$ x's,

$$P(f(S_n/m^{\frac{1}{2}}, n/m) \geqq \varepsilon \text{ for some } n \geqq 1) \leqq 1/\varepsilon \qquad (m > 0, \varepsilon > 1). \tag{8}$$

In order to see the meaning of (8) more clearly, suppose now that the probability measure F is confined to $(0, \infty)$, so that $f(x, t)$ defined by (6) is an *increasing* function of x. If we define for $t > 0$ the function $A(t, \varepsilon)$ = the (positive) solution x of the equation $f(x, t) = \varepsilon$, then

$$f(x, t \;\geqq \varepsilon \text{ if and only if } x \geqq A(t, \varepsilon), \tag{9}$$

and (8) can be written as

$$P(S_n \geqq m^{\frac{1}{2}} A(n/m, \varepsilon) \text{ for some } n \geqq 1) \leqq 1/\varepsilon \qquad (m > 0, \varepsilon > 1). \tag{10}$$

We note that (10) remains valid if instead of being i.i.d. $N(0, 1)$ the x's are any i.i.d. random variables having a moment generating function Ψ such that

$$\Psi(\theta) = E(e^{\theta x_1}) \leqq e^{\frac{1}{2}\theta^2} \qquad (0 < \theta < \infty). \tag{11}$$

To see this, let $g(x)$ be the density of x_1 with respect to some measure μ on $(-\infty, \infty)$ and let $g_\theta(x) = e^{\theta x} g(x)\Psi^{-1}(\theta)$; the argument which led to (10) now carries over *a fortiori* with $\varphi(x-\theta)$ replaced by $g_\theta(x)$. An example other than the $N(0, 1)$ case of a "subnormal" distribution satisfying (11) is the symmetric random walk distribution $P(x_1 = 1) = P(x_1 = -1) = \frac{1}{2}$, for which

$$\Psi(\theta) = (e^\theta + e^{-\theta})/2 = \textstyle\sum_0^\infty \theta^{2i}/(2i)! \leqq e^{\frac{1}{2}\theta^2} \qquad (-\infty < \theta < \infty).$$

Thus S_n for the symmetric random walk also satisfies (10), and for many distributions—Bernoulli, Poisson, uniform, normal square, etc.—an appropriate linear function with expectation 0 and variance 1 will provide an x_1 satisfying (11) and hence (10) in the case where F is confined to $(0, \infty)$.

We shall now illustrate the significance of (10) by considering a few particular choices of F. We write P for any distribution under which the x's are i.i.d. and (11) holds.

EXAMPLE 1. Let F be the degenerate measure which assigns mass 1 to the point $2a > 0$. Then $f(x, t) = \exp(2ax - 2a^2 t) \geqq \varepsilon$ if and only if $x \geqq at + (\log\varepsilon)/2a$. Hence (10) yields the inequality (with $d = (\log\varepsilon)/2a$)

$$P(S_n \geqq an/m^{\frac{1}{2}} + dm^{\frac{1}{2}} \text{ for some } n \geqq 1) \leqq e^{-2ad} \qquad (a, d, m \text{ all } > 0).$$

EXAMPLE 2. Let F be defined by

$$dF(y) = (2/\pi)^{\frac{1}{2}} e^{-\frac{1}{2}y^2} dy \quad \text{for} \quad 0 < y < \infty; = 0 \quad \text{elsewhere.}$$

Then for $t > -1$

$$f(x, t) = (2/\pi)^{\frac{1}{2}} \textstyle\int_0^\infty \exp(xy - y^2(t+1)/2\, dy = \dfrac{2e^{\frac{1}{2}x^2/(t+1)}}{(t+1)^{\frac{1}{2}}} \Phi\left(\dfrac{x}{(t+1)^{\frac{1}{2}}}\right).$$

To solve the equation $f(x, t) = \varepsilon$ for x we introduce the function

$$h(x) = x^2 + 2\log\Phi(x) \qquad (-\infty < x < \infty),$$

which increases from $-\infty$ to ∞ as x does, and is such that $h(x) \sim x^2$, $h^{-1}(x) \sim x^{\frac{1}{2}}$ as $x \to \infty$. Then $f(x, t) = \varepsilon$ for $x = A(t, \varepsilon) = (t+1)^{\frac{1}{2}} \cdot h^{-1}(2\log\frac{1}{2}\varepsilon + \log(t+1))$, so (10) yields (with $\varepsilon = 2e^{\frac{1}{2}a^2}\Phi(a)$) the inequality

(12) $$P(S_n \geqq (n+m)^{\frac{1}{2}} \cdot h^{-1}(h(a) + \log(n/m+1)) \text{ for some } n \geqq 1) \leqq \tfrac{1}{2}e^{-\frac{1}{2}a^2}/\Phi(a) \qquad (a \text{ and } m > 0).$$

EXAMPLE 3. For any $\delta > 0$ let F be defined by

$$dF(y) = \delta \cdot \frac{dy}{y \log 1/y (\log_2 1/y)^{1+\delta}} \quad \text{for} \quad 0 < y < e^{-e};$$
$$= 0 \quad \text{elsewhere,}$$

where we write $\log(\log y) = \log_2 y$, etc. It is impossible to evaluate $f(x, t)$ explicitly, but it can be shown by some analysis [14] that as $t \to \infty$

(13) $$A(t, \varepsilon) = \left\{2t\left(\log_2 t + \left(\frac{3}{2} + \delta\right)\log_3 t + \log\frac{\varepsilon}{2\delta\pi^{\frac{1}{2}}} + o(1)\right)\right\}^{\frac{1}{2}}.$$

The inequality (10) states for i.i.d. x's satisfying (11) that $P(S_n \geqq c_n$ for some $n \geqq 1) \leqq 1/\varepsilon$, where by proper choice of the probability measure F on $(0, \infty)$ we have seen that as $n \to \infty$

$$c_n = m^{\frac{1}{2}} A(n/m, \varepsilon) = an/m^{\frac{1}{2}} + dm^{\frac{1}{2}} \sim an/m^{\frac{1}{2}} \qquad \text{(Example 1)}$$

$$= (n+m)^{\frac{1}{2}} h^{-1}(h(a) + \log(n/m+1)) \sim (n \log n)^{\frac{1}{2}} \qquad \text{(Example 2)}$$

$$\sim (2n \log_2 n)^{\frac{1}{2}} \qquad \text{(Example 3).}$$

Other choices of F give sequences $c_n \sim n^{\beta}$ $(\frac{1}{2} < \beta < 1)$, etc. By the law of the iterated logarithm,

$$P\left(\limsup_{n\to\infty} \frac{S_n}{(2n \log_2 n)^{\frac{1}{2}}} = 1\right) = 1 \tag{14}$$

whenever the x's are i.i.d. with mean 0 and variance 1. Thus the sequences c_n of Example 3 increase about as slowly as is possible to have for i.i.d. x's with mean 0 and variance 1, $P(S_n \geqq c_n$ for some $n \geqq 1) < 1$.

Useful extensions of (10) are provided by the following remarks. The reader interested primarily in applications to statistics may proceed directly to Section 3.

(I) Returning to the general inequality (1), suppose we put

$$g_n'(x_1, \cdots, x_n) = \int_a^b g_{\theta,n}(x_1, \cdots, x_n)\, dF(\theta),$$

$$z_n = g_n'/g_n \quad \text{when} \quad g_n > 0,$$

where $P'(\cdot) = \int_a^b P_\theta(\cdot)\, dF(\theta)$ and $\{P_\theta; a < \theta < b\}$ is any family of joint probability distributions for the sequence $x_1, \cdots$ such that for each $n \geqq 1$ the random variables $x_1, \cdots, x_n$ have a probability density function $g_{\theta,n}(x_1, \cdots, x_n)$ (with respect to the same μ_n) which for fixed $x_1, \cdots, x_n$ is a Borel measurable function of θ. Then for any $\varepsilon > 0$ and $j = 1, 2, \cdots$ we shall show that

$$P(z_n \geqq \varepsilon \text{ for some } n \geqq j) \leqq P(z_j \geqq \varepsilon) + 1/\varepsilon \int_{(z_j < \varepsilon)} z_j\, dP \leqq F(a, b)/\varepsilon, \tag{15}$$

where F may now be any *σ-finite* measure on the parameter interval (a, b). (When $j = 1$ and F is a *probability* measure on (a, b) the extreme terms of (15) reduce to (1).)

PROOF. Let $N =$ first $n \geqq j$ such that $g_n' \geqq \varepsilon g_n$, with $N = \infty$ if no such n occurs; then

$$\begin{aligned} &P(z_n \geqq \varepsilon \text{ for some } n \geqq j) \\ &\quad = P(N < \infty) = P(N = j) + \textstyle\sum_{j+1}^{\infty} \int_{(N=n)} g_n\, d\mu_n \\ &\quad \leqq P(N = j) + 1/\varepsilon \textstyle\sum_{j+1}^{\infty} \int_{(N=n)} g_n'\, d\mu_n = P(N = j) + 1/\varepsilon P'(j < N < \infty) \\ &\quad \leqq P(N = j) + 1/\varepsilon P'(N > j) = P(z_j \geqq \varepsilon) + 1/\varepsilon \int_{(g_j' < \varepsilon g_j)} g_j'\, d\mu_j \\ &\quad = P(z_j \geqq \varepsilon) + 1/\varepsilon \int_{(z_j < \varepsilon)} z_j\, dP \leqq 1/\varepsilon \int_{(z_j \geqq \varepsilon)} z_j\, dP + 1/\varepsilon \int_{(z_j < \varepsilon)} z_j\, dP \leqq 1/\varepsilon \int z_j\, dP \\ &\quad = F(a, b)/\varepsilon. \end{aligned}$$

We remark that the first inequality of (15) holds under the assumption that $\{z_n, \mathscr{F}_n; n \geqq 1\}$ is any positive supermartingale on a probability space $(\Omega, \mathscr{F}, P)$, although the proof for that case is slightly different.

Applied to the i.i.d. $N(0, 1)$ case where F is confined to $(0, \infty)$ we obtain from (15) as an extension of (10) the result that for $\varepsilon > 0, m > 0, j = 1, 2, \cdots$ and $\tau = j/m$

(16) $$P(S_n \geqq m^{\frac{1}{2}}A(n/m, \varepsilon) \text{ for some } n \geqq j)$$
$$\leqq 1 - \Phi(A(\tau, \varepsilon)/\tau^{-\frac{1}{2}} + 1/\varepsilon \textstyle\int_0^\infty \Phi(A(\tau, \varepsilon)/\tau^{-\frac{1}{2}} y \tau^{-\frac{1}{2}}\, dF(y).$$

For example, if $dF(y) = (2/\pi)^{\frac{1}{2}}dy$ for $0 < y < \infty$ then (cf. Example 2 above)

$$f(x, t) = 2e^{\frac{1}{2}x^2/t}/t^{-\frac{1}{2}}\Phi(x/t^{\frac{1}{2}}) \quad . \qquad (t > 0),$$

and from (16) for $j = m$, $\tau = 1$ we obtain (cf. (12)) for i.i.d. $N(0, 1)$ x's

$$P(S_n \geqq n^{\frac{1}{2}}h^{-1}(h(a) + \log n/m) \text{ for some } n \geqq m) \leqq 1 - \Phi(a) + \varphi(a)(a + \varphi(a)/\Phi(a)).$$

(II) If F is a *symmetric* probability measure on $(-\infty, \infty)$ then instead of (9) we see that $f(x, t) \geqq \varepsilon$ if and only if $|x| \geqq A(t, \varepsilon)$, and (10) continues to hold with S_n replaced by $|S_n|$ when the x's are i.i.d. $N(0, 1)$, or more generally when (11) holds for all $-\infty < \theta < \infty$. For example, if $F = \Phi$ we obtain when (11) holds for all $-\infty < \theta < \infty$ that

(17) $$P(|S_n| \geqq [(n+m)(a^2 + \log(n/m + 1))]^{\frac{1}{2}} \text{ for some } n \geqq 1) \leqq e^{-\frac{1}{2}a^2}.$$

Since $(a^2 + \log t)^{\frac{1}{2}} > h^{-1}(h(a) + \log t)$ for all $t > 1$, it follows from (12) that as a one-sided version of (17) we have

(18) $$P(S_n \geqq [(n+m)(a^2 + \log(n/m + 1))]^{\frac{1}{2}} \text{ for some } n \geqq 1) \leqq \tfrac{1}{2}e^{-\frac{1}{2}a^2}/\Phi(a) \qquad (a > 0).$$

Likewise, if F is any symmetric measure on $(-\infty, \infty)$ and the x's are i.i.d. $N(0, 1)$ then we obtain from (15) as an analogue of (16) that with $\tau = j/m$

(19) $$P(|S_n| \geqq m^{\frac{1}{2}}A(n/m, \varepsilon) \text{ for some } n \geqq j)$$
$$\leqq 2(1 - \Phi(A(\tau, \varepsilon)/\tau^{\frac{1}{2}})) + 1/\varepsilon \textstyle\int_{-\infty}^{\infty} \{\Phi(A(\tau, \varepsilon)/\tau^{\frac{1}{2}} - y\tau^{\frac{1}{2}})$$
$$- \Phi(-A(\tau, \varepsilon)/\tau^{\frac{1}{2}} - y\tau^{\frac{1}{2}})\}\, dF(y).$$

From (19) with $dF(y) = (2\pi)^{-\frac{1}{2}}dy$ for $-\infty < y < \infty$ and $j = m$, $\tau = 1$ we obtain for i.i.d. $N(0, 1)$ x's that

(20) $$P(|S_n| \geqq [n(a^2 + \log n/m)]^{\frac{1}{2}} \text{ for some } n \geqq m) \leqq 2(1 - \Phi(a) + a\varphi(a)) \qquad (a > 0).$$

(III) Suppose the x's are i.i.d. with an absolutely continuous density function $g(x - \theta)$ under P_θ, and that $P = P_0$. Define for $n \geqq 1$ and $-\infty < \theta < \infty$,

$y_n = x_n/|x_1|$ and

$$z_{\theta,n}(x_1, \cdots, x_n) = \frac{\int_0^\infty \prod_1^n \left\{\frac{1}{\sigma} g\left(\frac{x_i}{\sigma} - \theta\right)\right\} \frac{d\sigma}{\sigma}}{\int_0^\infty \prod_1^n \left\{\frac{1}{\sigma} g\left(\frac{x_i}{\sigma}\right)\right\} \frac{d\sigma}{\sigma}}.$$

Then $z_{\theta,n}(x_1, \cdots, x_n) = z_{\theta,n}(y_1, \cdots, y_n)$, and if the joint density of $y_1, \cdots, y_n$ under P_θ is denoted by $p_\theta(y_1, \cdots, y_n)$, then it is easy to see that

$$z_{\theta,n}(y_1, \cdots, y_n) = \frac{p_\theta(y_1, \cdots, y_n)}{p_0(y_1, \cdots, y_n)}.$$

Hence if F is any measure on $(-\infty, \infty)$ and if we put

$$z_n(x_1, \cdots, x_n) = \int_{-\infty}^\infty z_{\theta,n}\, dF(\theta),$$

then (15) holds. For example, if $g = \varphi$ and we put $dF(\theta) = (m/2\pi)^{\frac{1}{2}}\, d\theta$, $-\infty < \theta < \infty$, we obtain after some computation the following result: if the x's are i.i.d. $N(\mu, \sigma^2)$ and

$$\bar{x}_n = n^{-1}\textstyle\sum_1^n x_i, \qquad v_n^2 = n^{-1}\sum_1^n (x_i - \bar{x}_n)^2 \qquad \text{then}$$

(21) $$P(|\bar{x}_n - \mu| \geqq v_n[(tn)^{n^{-1}} - 1]^{\frac{1}{2}} \text{ for some } n \geqq m)$$
$$\leqq 2(1 - F_{m-1}(a) + a f_{m-1}(a)) \qquad (a \geqq 0,\ m = 1, 2, \cdots)$$

where $t = m^{-1}(1 + a^2/(m-1))^m$ and f_m, F_m denote the Student t density and distribution function with m degrees of freedom. Note that for large n, m

$$(tn)^{n^{-1}} - 1 \cong e^{n^{-1}\log(tn)} - 1 \cong n^{-1}\log(tn)$$
$$\cong n^{-1}\log(nm^{-1}(1 + a^2/(m-1))^m) \cong n^{-1}(a^2 + \log n/m)$$

and compare (21) with (20). The use of the scale-invariant measure $d\sigma/\sigma$ was suggested to the author by Robert Berk; cf. ([10] page 250).

2. The Wiener process and a limit theorem. Let $w(t)$ denote a standard Wiener process for $t \geqq 0$ with $w(0) = 0$. If $x_1, \cdots$ are independent $N(0, 1)$, the two sequences

(22) $$(S_1/m^{\frac{1}{2}}, S_2/m^{\frac{1}{2}}, \cdots)$$
$$(w(1/m), w(2/m), \cdots)$$

have the same joint distribution for any $m > 0$. This suggests that (10), (16), and (19) should become equalities for the Wiener process; e.g., in the case of (10) that if F is a probability measure on $(0, \infty)$ and if $f(x, 0) < \infty$ for all x, then

(23) $$P(w(t) \geqq A(t, \varepsilon) \text{ for some } t \geqq 0) = 1/\varepsilon \qquad (\varepsilon > 1).$$

To see heuristically why this should be so we remark that in (2) the only strict inequality was the replacement of $1/z_n$ by $1/\varepsilon$ on the set $(N = n)$. If m is large, the "overshoot" will be stochastically small, and the behavior of the first sequence of

(22) will be about the same as that of the continuous process $w(t)$. However, instead of trying to prove that (23) holds by letting $m \to \infty$ in (10), it is more convenient to prove (23) directly from the fact that

$$z(t) = \int_0^\infty \exp(w(t)y - \tfrac{1}{2}ty^2)\, dF(y) \qquad (t \geqq 0) \tag{24}$$

is a positive martingale with continuous sample paths for which $z(0) = 1$. It is easy to show that $z(t) \to 0$ in probability as $t \to \infty$, and from this it follows almost immediately by stopping $z(t)$ the first time it is $\geqq \varepsilon$ that

$$P(z(t) \geqq \varepsilon \text{ for some } t \geqq 0) = 1/\varepsilon \qquad (\varepsilon > 1),$$

which is equivalent to (23). The examples of F which we have already considered thus give exact boundary crossing probabilities for the Wiener process:

$$P(w(t) \geqq at + d \text{ for some } t \geqq 0) = e^{-2ad} \qquad (a > 0, d > 0),$$

$$P(w(t) \geqq (t+1)^{\frac{1}{2}} \cdot h^{-1}(h(a) + \log(t+1)) \text{ for some } t \geqq 0) = \tfrac{1}{2}e^{-\frac{1}{2}a^2}/\Phi(a) \qquad (a > 0)$$

$$P(w(t) \geqq A(t, \varepsilon) \text{ for some } t \geqq 0) = 1/\varepsilon \qquad (\varepsilon > 1),$$

for the function $A(t, \varepsilon)$ of Example 3 for which the asymptotic expression (13) holds,

$$P(|w(t)| \geqq [(t+1)(a^2 + \log(t+1))]^{\frac{1}{2}} \text{ for some } t \geqq 0) = e^{-\frac{1}{2}a^2} \qquad (a > 0),$$

etc. Only in the case of a linear boundary have such formulas been available up to now.

The heuristic argument which suggested the truth of (23) suggests further, because of the central limit theorem, that the limit relations

$$\begin{aligned} &\lim_{m\to\infty} P(S_n \geqq m^{\frac{1}{2}}A(n/m, \varepsilon) \text{ for some } n \geqq 0) \\ &\quad = P(w(t) \geqq A(t, \varepsilon) \text{ for some } t \geqq 0) = 1/\varepsilon \qquad (\varepsilon > 1), \end{aligned}$$

$$\begin{aligned} &\lim_{m\to\infty} P(S_n \geqq m^{\frac{1}{2}}A(n/m, \varepsilon) \text{ for some } n \geqq \tau m) \\ &\quad = P(w(t) \geqq A(t, \varepsilon) \text{ for some } t \geqq \tau) \\ &\quad = 1 - \Phi(A(\tau, \varepsilon)/\tau^{\frac{1}{2}}) + 1/\varepsilon \int_0^\infty \Phi(A(\tau, \varepsilon)/\tau^{\frac{1}{2}} - y\tau^{\frac{1}{2}})\, dF(y) \qquad (\varepsilon > 0, \tau > 0) \end{aligned} \tag{25}$$

together with analogous relations for $|S_n|$ and $|w(t)|$ should hold *whenever the* x's *are* i.i.d. *with mean* 0 *and variance* 1, normal or not. This is true under some mild assumption about the behavior of the function $A(t, \varepsilon)$ as $t \to \infty$; it is sufficient to assume that $A(t, \varepsilon)/t^{\frac{1}{2}}$ is ultimately non-decreasing. Hence the inequalities previously obtained for every finite $m > 0$ when (11) holds now become limit theorems as $m \to \infty$ with no parametric assumptions about the x's. A full discussion of this is given in [14]. The statistical significance of limit theorems such as (25) will be discussed in the following sections.

3. Confidence sequences and tests with uniformly small error probability for the mean of a normal distribution with known variance. Let $x_1, \cdots$ be independent

$N(\theta, 1)$ where θ is an unknown parameter, $-\infty < \theta < \infty$. Define the intervals $I_n = ((S_n - c_n)/n, (S_n + c_n)/n)$ $(n \geqq 1)$, where $S_n = x_1 + \cdots + x_n$ and c_n is any sequence of positive constants such that $c_n/n \to 0$ as $n \to \infty$. Then

$$P_\theta(\theta \in I_n \text{ for every } n \geqq 1) = P_0(|S_n| < c_n \text{ for every } n \geqq 1)$$
$$= 1 - P_0(|S_n| \geqq c_n \text{ for some } n \geqq 1).$$

We saw for example in (17) that if

$$c_n = [(n+m)(a^2 + \log(n/m+1))]^{\frac{1}{2}} \qquad (m > 0) \tag{26}$$

then $P_0(|S_n| \geqq c_n \text{ for some } n \geqq 1) \leqq e^{-\frac{1}{2}a^2}$. Hence $P_\theta(\theta \in \bigcap_1^\infty I_n) \geqq 1 - e^{-\frac{1}{2}a^2}$, which can be made as near 1 as we please by choosing a sufficiently large; e.g., for

$$a^2 \cong 6, \qquad 1 - e^{-\frac{1}{2}a^2} = .95. \tag{27}$$

Thus for $a^2 = 6$ and any $m > 0$, the sequence I_n with c_n defined by (26) forms a "confidence sequence" for an unknown θ with coverage probability $\geqq .95$ (cf. [17] pages 153–156). As $m \to \infty$ the coverage probability $\to .95$ by (25).

Choosing for example $m = 1$, the half-width of I_n is

$$\frac{c_n}{n} = \left[\frac{n+1}{n^2}(6 + \log(n+1))\right]^{\frac{1}{2}} \sim \left[\frac{\log n}{n}\right]^{\frac{1}{2}} \to 0 \qquad \text{as } n \to \infty. \tag{28}$$

Of course, for any *fixed n*

$$P_\theta((S_n - 1.96n^{\frac{1}{2}})/n < \theta < (S_n + 1.96n^{\frac{1}{2}})/n) = .95, \tag{29}$$

a 95% confidence interval for θ of half-width $1.96/n^{\frac{1}{2}}$. However, by (14)

$$P_\theta((S_n - 1.96n^{\frac{1}{2}})/n < \theta < (S_n + 1.96n^{\frac{1}{2}})/n \text{ for every } n \geqq 1) = 0,$$

and this remains true if 1.96 is replaced by any constant, no matter how large.

The advantage of the confidence sequence I_n compared to a fixed sample size confidence interval is that it allows us to "follow" the unknown θ throughout the whole sequence $x_1, \cdots$ with an interval I_n whose length shrinks to 0 as the sample size increases, in such a way that with probability $\geqq .95$ the interval I_n contains θ at every stage. (This is also true of the smaller intervals $J_n = \bigcap_1^n I_k \subset I_n$, although for some n it might happen that $J_n = \varnothing$.) The validity of the relation $P_\theta(\theta \in I_n) \geqq .95$ is therefore unaffected by the possibility that n may be a random variable dependent on the whole sequence $x_1, \cdots$; in other words, the confidence level .95 is unaffected by any kind of optional stopping which could vitiate (29).

The disadvantage of using the sequence I_n is evident from a comparison of the numerical value of (28) with $1.96/n^{\frac{1}{2}}$.

If we wish to test $H^-: \theta < 0$ versus $H^+: \theta > 0$ ($\theta = 0$ being excluded) we can define the stopping time $N =$ first $n \geqq 1$ such that $|S_n| \geqq c_n$ and accept H^+ or H^- according as $S_N \geqq c_N$ or $S_N \leqq -c_N$. Since $S_n/n \to \theta \neq 0$ under H^- or H^+, while $c_n/n \to 0$, it follows that $P_\theta(N < \infty) = 1$ for $\theta \neq 0$, while if $\theta > 0$,

$$P_\theta(\text{accept } H^-) = P_\theta(S_n \leqq -c_n \text{ before } S_n \geqq c_n)$$
$$< P_0(S_n \leqq -c_n \text{ before } S_n \geqq c_n) = \tfrac{1}{2}P_0(|S_n| \geqq c_n \text{ for some } n \geqq 1)$$
$$\leqq \tfrac{1}{2}e^{-\frac{1}{2}a^2},$$

and similarly, P_θ (accept H^+) $\leqq \frac{1}{2}e^{-\frac{1}{2}a^2}$ for any $\theta < 0$. Thus the error probability of this test is uniformly $\leqq \frac{1}{2}e^{-\frac{1}{2}a^2}$ for all $\theta \neq 0$. (Exactly the same argument holds if the x's have the distribution $P(x_i = 1) = p = 1 - P(x_i = -1)$, with $\theta = 2p - 1$.)

Of course, $P_0(N < \infty) \leqq e^{-\frac{1}{2}a^2}$, so the test will rarely terminate when $\theta = 0$. The expected sample size $E_\theta(N)$ is, however, finite for every $\theta \neq 0$, approaching ∞ as $\theta \to 0$ and 1 as $|\theta| \to \infty$.

4. Tests with power 1. Again, let $x_1, \cdots$ be independent $N(\theta, 1)$ but now let $H_0 : \theta \leqq 0$, $H_1 : \theta > 0$ be the hypotheses which are to be tested. Put

$$N = \text{first } n \geqq 1 \text{ such that } S_n \geqq c_n$$
$$= \infty \text{ if no such } n \text{ occurs,} \tag{30}$$

and agree when $N < \infty$ to stop sampling with x_N and reject H_0 in favor of H_1; if $N = \infty$, continue sampling indefinitely and do not reject H_0.

For $\theta \leqq 0$ we have

$$P_\theta(\text{reject } H_0) = P_\theta(N < \infty) \leqq P_0(N < \infty)$$
$$= P_0(S_n \geqq c_n \text{ for some } n \geqq 1).$$

If we are using the c_n sequence (26) we have by (18)

$$P_0(S_n \geqq c_n \text{ for some } n \geqq 1) < \tfrac{1}{2}e^{-\frac{1}{2}a^2}/\Phi(a) \tag{31}$$

(if we were to use for c_n the smaller sequence of (12), (31) would still hold and with approximate equality for large m). Hence the type I error probability of the test has the upper bound $P_\theta(\text{reject } H_0) < \frac{1}{2}e^{-\frac{1}{2}a^2}/\Phi(a)$ for all $\theta \leqq 0$, while the type II error probability is

$$P_\theta(\text{not reject } H_0) = P_\theta(S_n < c_n \text{ for all } n \geqq 1) \equiv 0 \qquad \text{for all } \theta > 0,$$

since $c_n/n \to 0$ and $S_n/n \to \theta > 0$ as $n \to \infty$. Thus the test has power 1 against the alternative $\theta > 0$.

Of course the test will rarely terminate when $\theta \leqq 0$. Some people may consider this intolerable, but that is an unreasonable attitude in many practical situations.

Concerning the expected sample size $E_\theta(N)$ when $\theta > 0$, it can be shown that for *any* stopping rule N of the sequence $x_1, \cdots$ the inequality

$$E_\theta(N) \geqq -2\log P_0(N < \infty)/\theta^2 \tag{32}$$

must hold for every $\theta > 0$. Thus, if we are willing to tolerate an N for which $P_0(N < \infty) = .05$, then necessarily $E_\theta(N) \geqq 6/\theta^2$ for every $\theta > 0$; however, no such N will minimize $E_\theta(N)$ uniformly for all $\theta > 0$. For the N given by (30), if like (26) the function c_t is concave for $t \geqq 1$ it can be shown [5] that

$$E_\theta(N) \leqq \frac{c_{E_\theta(N)}}{\theta} + \frac{\varphi(\theta)}{\theta\Phi(\theta)} + 1, \tag{33}$$

which gives an implicit upper bound for $E_\theta(N)$ as a function of $\theta > 0$. For (26) with $m = 1$ and $a^2 \doteq 9$, for example, we obtain from (32) and (33) the bounds

$$\begin{aligned} 1040 &< E_{.1}(N) < 1800 \\ 10.4 &< E_1(N) < 15 \\ 2.6 &< E_2(N) < 5 \\ &\vdots \end{aligned}$$

More precise estimates of $E_\theta(N)$ could be obtained from Monte Carlo methods which, for obvious reasons, are not directly applicable to estimating the Type I error, for which we have the upper bound (31) of .0056 for $a^2 = 9$.

Other examples of the methods described above in testing, selection, and ranking procedures are indicated in the references. In the next two sections we shall discuss some non-parametric "open-ended" procedures.

5. Confidence sequences for the median. Let $z_1, \cdots$ be i.i.d. with $P(z_i \leqq M) = \frac{1}{2}$ and let $z_1^{(n)} \leqq z_2^{(n)} \leqq \cdots$ denote the ordered values $z_1, \cdots, z_n$. The usual confidence interval for M for a single value of n is based on the normal approximation to the binomial distribution which gives the relation

$$P(z_{a_1}^{(n)} \leqq M \leqq z_{a_2}^{(n)}) \cong 2\Phi(a) - 1 \qquad \text{for large } n$$

where

$$\begin{aligned} a_1 &= a_1(n) = \text{largest integer} \leqq \tfrac{1}{2}(n - an^{\frac{1}{2}}), \\ a_2 &= a_2(n) = \text{smallest integer} \geqq \tfrac{1}{2}(n + an^{\frac{1}{2}}). \end{aligned}$$

To construct a confidence sequence for M, let

$$\begin{aligned} x_i &= 1 \quad \text{if} \quad z_i \leqq M, \\ &= -1 \quad \text{if} \quad z_i > M, \qquad S_n = x_1 + \cdots + x_n, \end{aligned}$$

and let c_n be some sequence of positive constants. Define

$$\begin{aligned} b_1 &= b_1(n) = \text{largest integer} \leqq \tfrac{1}{2}(n - c_n), \\ b_2 &= b_2(n) = \text{smallest integer} \geqq \tfrac{1}{2}(n + c_n). \end{aligned}$$

Then

$$\begin{aligned} &P(z_{b_1}^{(n)} \leqq M \leqq z_{b_2}^{(n)} \text{ for every } n \geqq m) \\ &\qquad \geqq 1 - P(|S_n| \geqq c_n \text{ for some } n \geqq m). \end{aligned}$$

Using for example the sequence $c_n = [n(a^2 + \log n/m)]^{\frac{1}{2}}$ for which (20) gives the approximation for large m

$$P(|S_n| > c_n \text{ for some } n \geqq m) \cong 2(1-\Phi(a)+a\varphi(a)),$$

we have for large m

$$P(z_{b_1}^{(n)} \leqq M \leqq z_{b_2}^{(n)} \text{ for every } n \geqq m) \cong 2\Phi(a)-1-2a\varphi(a)$$
$$\equiv H(a), \text{ say,}$$

where

a	$2\Phi(a)-1$	$H(a)$
2	.9546	.7386
2.5	.9876	.9001
2.8	.9948	.9506
3	.9974	.9710
3.5	.9996	.9933

and $b_1(m) = a_1(m)$, $b_2(m) = a_2(m)$. Remember that by the law of the iterated logarithm

$$P(z_{a_1}^{(n)} \leqq M \leqq z_{a_2}^{(n)} \text{ for every } n \geqq m) = 0 \qquad (m = 1, 2, \cdots).$$

6. Kolmogorov–Smirnov tests with power 1. For any two distribution functions G, H write

$$D^+(G, H) = \sup_{-\infty<t<\infty}(G(t)-H(t)), \quad D(G, H) = \sup_{-\infty<t<\infty}|G(t)-H(t)|.$$

Let $x_1, \cdots$ be i.i.d. with df $F_x(t) = P(x_i \leqq t)$ and let $y_1, \cdots$ be i.i.d. with df $F_y(t)$, the x's and y's being independent. Denote by $F_x^n(t)$, $F_y^n(t)$ the sample df's of $x_1, \cdots, x_n$ and $y_1, \cdots, y_n$.

Consider the hypothesis

$$H_0: F_x(t) \leqq F_y(t) \qquad \text{for every } -\infty < t < \infty.$$

To test H_0 define

$$N = \text{first integer } n \geqq m \text{ such that } D^+(F_x^n, F_y^n) \geqq f(n)/n$$
$$= \infty \quad \text{if no such } n \text{ occurs,}$$

where $f(n)$ is some positive sequence such that $f(n)/n \to 0$ as $n \to \infty$. If H_0 is false, and $D^+(F_x, F_y) = d > 0$, then by the Glivenko–Cantelli theorem, as $n \to \infty$ $D^+(F_x^n, F_y^n) \to d$ with probability 1, so that $P(N < \infty) = 1$. Hence if we agree to reject H_0 as soon as we observe that $N < \infty$, while if $N = \infty$ we do not reject H_0, then the test certainly has power 1 when H_0 is false. It remains to consider the Type I error probability.

It can be shown that when H_0 is true, no matter what F_x and F_y may be in other respects, the inequality

$$P\left(D^+(F_x^n, F_y^n) \geqq \frac{r}{n}\right) \leqq \frac{(n!)^2}{(n-r)!\,(n+r)!} \leqq e^{-(r^2/n+1)} \qquad (r = 0, 1, \cdots, n)$$

always holds, and hence the crude inequality

$$(34) \qquad P(N<\infty) \leqq \sum_{n=m}^{\infty} P\left(D^{+}(F_x{}^n, F_y{}^n) \geqq \frac{f(n)}{n}\right) \leqq \sum_{n=m}^{\infty} e^{-[f^2(n)/n+1]}.$$

holds. Choosing $f(n)$ to be $\sim [(1+\varepsilon)n\log n]^{\frac{1}{2}}$ will suffice to make the series converge and hence will guarantee an arbitrarily small Type I error probability when H_0 is true if m is chosen sufficiently large. For example, if $f(n) = [(n+1)(\log 4 + 2\log n)]^{\frac{1}{2}}$, $m = 6$ then $P(N<\infty) < .05$ whenever H_0 is true.

The law of the iterated logarithm for the sequence $D^{+}(F_x{}^n, F_y{}^n)$ shows that taking $f(n) \sim [(1+\varepsilon)n\log_2 n]^{\frac{1}{2}}$ would also suffice to ensure an arbitrarily small value of $P(N<\infty)$ under H_0 for large m. Upper bounds for $P(N<\infty)$ in such cases have recently been obtained by Richard Stanley (unpublished).

Concerning the value of EN when H_0 is false, it can be shown that it is always finite and that the inequality

$$EN \leqq g(d-m/g(d)), \qquad d = D^{+}(F_x, F_y)$$

holds, where $g(x)$ is the function inverse to $f(x)/x$.

Similar tests are available for various other non-parametric hypotheses such as

$H_1: F_x = F_y$

$H_2: F_x \leqq F$; $\quad F$ an arbitrary specified df,

$H_3: F_x \in \mathscr{F}$; $\quad \mathscr{F}$ any class of df's closed under the D metric (e.g., the set $N(\mu, \sigma^2)$ with $-\infty < \mu < \infty$, $0 \leqq \sigma^2 < \infty$).

In each case the power is 1 and the expected sample size is finite under any alternative, while an arbitrarily small upper bound for the type I error can be guaranteed. The currently available bounds are crude, however, being based on inequalities similar to (34).

7. Concluding remark. The ideas involved here seem to be a natural extension (or contraction) of Wald's sequential analysis; cf. also [1], [7], and [9]. My own work has been done in collaboration with D. A. Darling in the first instance and later with D. Siegmund, to whom I wish to express my deep appreciation.

REFERENCES

[1] Blum, J. R. and Rosenblatt, J. (1966). On some statistical problems requiring purely sequential sampling schemes. *Ann. Inst. Statist. Math.* **18** 351–354.

[2] Darling, D. A. and Robbins, H. (1967a). Iterated logarithm inequalities. *Proc. Nat. Acad. Sci.* **57** 1188–1192.

[3] Darling, D. A. and Robbins, H. (1967b). Inequalities for the sequence of sample means. *Proc. Nat. Acad. Sci.* **57** 1577–1580.

[4] Darling, D. A. and Robbins, H. (1967c). Confidence sequences for mean, variance, and median. *Proc. Nat. Acad. Sci.* **58** 66–68.

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

[6] DARLING, D. A. and ROBBINS, H. (1968b). Some nonparametric sequential tests with power 1. *Proc. Nat. Acad. Sci.* **61** 804–809.

[7] FABIAN, V. (1956). A decision function. *Czechoslovak Math. J.* **6** 31–41.

[8] FARRELL, R. H. (1962). Bounded length confidence intervals for the zero of a regression function. *Ann. Math. Statist.* **33** 237–247.

[9] FARRELL, R. H. (1964). Asymptotic behavior of expected sample size in certain one sided tests. *Ann. Math. Statist.* **35** 36–72.

[10] LEHMANN, E. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

[11] ROBBINS, H. and SIEGMUND, D. (1968). Iterated logarithm inequalities and related statistical procedures. *Mathematics of the Decision Sciences* **2**. American Mathematical Society, Providence 267–279.

[12] ROBBINS, H. and SIEGMUND, D. (1969a). Probability distributions related to the law of the iterated logarithm. *Proc. Nat. Acad. Sci.* **62** 11–13.

[13] ROBBINS, H. and SIEGMUND, D. (1969b). Boundary crossing probabilities for the Wiener process. Stanford Univ. Department of Statistics Tech. Report No. 4.

[14] ROBBINS, H. and SIEGMUND, D. (1970). Boundary crossing probabilities for the Wiener process and sample sums. *Ann. Math. Statist.* **41** 1410–1429.

[15] ROBBINS, H., SIEGMUND, D. and WENDEL, J. (1968). The limiting distribution of the last time $S_n \geqq n\varepsilon$. *Proc. Nat. Acad. Sci.* **61** 1228–1230.

[16] VILLE, J. (1939). *Étude Critique de la Notion de Collectif.* Gauthier-Villars, Paris.

[17] WALD, A. (1947). *Sequential Analysis.* Wiley, New York.