

# A RATE OF CONVERGENCE FOR THE VON MISES STATISTIC<sup>(1)</sup>

BY  
WALTER A. ROSENKRANTZ

**I. Introduction.** Let  $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$  denote  $n$  identically distributed, mutually independent random variables uniformly distributed over the interval  $[0, 1]$ , i.e.  $P\{0 \leq X_i \leq t\} = t, 0 \leq t \leq 1$ . We denote by  $(X_1^*(\omega), X_2^*(\omega), \dots, X_n^*(\omega))$  the set of  $n$  observations  $(X_1(\omega), \dots, X_n(\omega))$  arranged in increasing order, thus  $X_1^*(\omega) < X_2^*(\omega) < \dots < X_n^*(\omega)$ .  $X_r^*(\omega)$  is called the  $r$ th order statistic and we refer the reader to [2] or [7] for a discussion of this topic. We denote by  $F_n(t, \omega)$  the empirical distribution function corresponding to the random sample  $(X_1, \dots, X_n)$ <sup>(2)</sup>. Thus

- (1)  $nF_n(t, \omega)$  = number of observations  $X_i \leq t$ , and
- (2)  $F_n(t, \omega) = k/n$  for  $X_k^* \leq t < X_{k+1}^*$ .

The stochastic process  $\Delta_n(t, \omega)$  where

- (3)  $\Delta_n(t, \omega) = n^{1/2}(F_n(t, \omega) - t)$

has been studied by Kolmogorov [8], Doob [5], Donsker [4] and Anderson and Darling [1] in connection with the Kolmogorov-Smirnov and Cramer-Von Mises statistics. Let  $D(t, \omega) = W(t, \omega) - tW(1, \omega), 0 \leq t \leq 1$ , where  $W(t, \omega)$  is the one dimensional Brownian motion process with  $P\{W(0) = 0\} = 1$  and covariance function  $r(s, t) = \min(s, t)$ . Then, as was shown by Donsker [4], the  $\Delta_n(t)$  process tends to the process  $D(t)$  in the sense of weak convergence of stochastic processes, an account of which is to be found in references [9] and [10]. These general results of Donsker's, as is well known, include as special cases the results of Kolmogorov-Smirnov and Cramer-Von Mises. We now give two examples.

Let  $x(t)$  denote a bounded measurable function on the interval  $[0, 1]$  and define functionals  $\|x\|$  and  $\mathcal{F}(x)$  as follows<sup>(3)</sup>:

- (4)  $\|x(t)\| = \text{Max}_{0 \leq t \leq 1} |x(t)|$ ,
- (5)  $\mathcal{F}(x) = \int_0^1 (x(t))^2 dt$ .

Let us now define cumulative probability distribution functions  $K_n(\lambda), K(\lambda), V_n(\lambda), V(\lambda)$  as follows:

- (6)  $K_n(\lambda) = P\{\|\Delta_n\| \leq \lambda\}$ ,

Received by the editors July 11, 1968.

<sup>(1)</sup> This paper represents results obtained at the Courant Institute of Mathematical Sciences, New York University, New York, under the sponsorship of the National Science Foundation Grant NSF-GP-6237.

<sup>(2)</sup> Whenever it is convenient we shall suppress the  $\omega$  in the term  $X_n(\omega)$  and similar terms.

<sup>(3)</sup> The appropriate hypothesis is to assume that  $x(t) \in D[0, 1]$  and we refer the reader to [9] for a discussion of the function space  $D[0, 1]$ .

(7)  $V_n(\lambda) = P\{\mathcal{F}(\Delta_n) \leq \lambda\},$

(8)  $K(\lambda) = P\{\|D\| \leq \lambda\},$

(9)  $V(\lambda) = P\{\mathcal{F}(D) \leq \lambda\}.$

It is an immediate consequence of Donsker's theorem that  $\lim_{n \rightarrow \infty} K_n(\lambda) = K(\lambda)$  and  $\lim_{n \rightarrow \infty} V_n(\lambda) = V(\lambda)$ . It is an interesting and important problem to determine the rate of the convergence of  $K_n$  to  $K$  and  $V_n$  to  $V$ .

It is a consequence of a theorem of Chan-Li-Tsian, (see p. 155 of [6]), that

(10)  $\text{Max}_\lambda |K_n(\lambda) - K(\lambda)| \leq A_1 n^{-1/2}$  as  $n \rightarrow \infty$ .

The purpose of this paper is to derive the following estimate:

**THEOREM 1.**  $\text{Max}_\lambda |V_n(\lambda) - V(\lambda)| \leq A(\log n)^{3/2} \cdot n^{-1/5}$  as  $n \rightarrow \infty$ .

Our method is based on the Skorokhod representation which the author has used previously in a similar context [11]. The fact that the Skorokhod representation can be applied to the Kolmogorov statistics is due to Breiman [2] and our method is essentially an application of Breiman's idea to the Von Mises statistic  $V_n(\lambda)$ .

**II. Proof of the theorem.** We begin by considering the following sequence of random points in the plane  $(X_k^*(\omega), k/n)$ . We denote by  $\tilde{F}_n(t, \omega)$  the "random broken line" that connects  $(X_k^*, k/n)$  to  $(X_{k+1}^*, (k+1)/n)$  by a straight line. Thus  $\tilde{F}_n(t, \omega)$  is a continuous, monotone increasing function on the interval  $[0, 1]$ , and  $\tilde{F}_n(t, \omega) = F_n(t, \omega)$  when  $t = X_k^*(\omega)$ .  $F_n(t, \omega)$  itself is a step function with jumps of size  $1/n$  at the points  $t = X_k^*$ . Thus  $\tilde{F}_n(t, \omega)$  is a continuous function that approximates  $F_n(t, \omega)$ . We shall now study how good this approximation is.

**LEMMA 1.** *Let  $\epsilon_n$  denote a sequence of positive numbers decreasing to 0 at a rate of speed to be specified later. Then*

(11)  $P\{\|\Delta_n\| \geq \epsilon_n(n)^{1/2}\} \leq 2 \exp(-2n\epsilon_n^2) + A_1 n^{-1/2}.$

**Proof.** The proof is an immediate consequence of (10) and the well-known estimate

(12)  $P\{\|D\| \geq \lambda\} \leq 2 \exp(-2\lambda^2)$  (p. 396 of [5]).

We now "smooth out" the process  $\Delta_n(t, \omega)$  by introducing the following process

(13)  $\tilde{\Delta}_n(t, \omega) = n^{1/2}(\tilde{F}_n(t, \omega) - t).$

We note, without proof, the following two inequalities

(14)  $|\Delta_n(t, \omega)| \leq |\tilde{\Delta}_n(t, \omega)| + n^{-1/2}$

(15)  $|\tilde{\Delta}_n(t, \omega)| \leq |\Delta_n(t, \omega)| + n^{-1/2}.$

The proofs of (14) and (15) are elementary and are therefore omitted. The following two inequalities are immediate consequences of (14) and (15).

(16)  $\mathcal{F}(\Delta_n) - \mathcal{F}(\tilde{\Delta}_n) \leq 2n^{-1/2} \int_0^1 |\tilde{\Delta}_n(t, \omega)| dt + n^{-1},$

(17)  $\mathcal{F}(\tilde{\Delta}_n) - \mathcal{F}(\Delta_n) \leq 2n^{-1/2} \int_0^1 |\tilde{\Delta}_n(t, \omega)| dt + 3n^{-1}.$

Combining (16) and (17) we obtain

$$(18) \quad |\mathcal{F}(\Delta_n) - \mathcal{F}(\tilde{\Delta}_n)| \leq 2n^{-1/2} \int_0^1 |\tilde{\Delta}_n(t, \omega)| dt + 3n^{-1}.$$

Let us also note that

$$(19) \quad \|\Delta_n\| \geq \|\tilde{\Delta}_n\|.$$

LEMMA 2.

$$P\{|\mathcal{F}(\Delta_n) - \mathcal{F}(\tilde{\Delta}_n)| \geq 2((\log n)/n)^{1/2}\} \leq A_2 n^{-1/2}.$$

**Proof.** It follows from (18) that

$$P\left\{|\mathcal{F}(\Delta_n) - \mathcal{F}(\tilde{\Delta}_n)| \geq 2\left(\frac{\log n}{n}\right)^{1/2}\right\} \leq P\left\{\int_0^1 |\tilde{\Delta}_n(t, \omega)| dt \geq \varepsilon_n \cdot n^{1/2}\right\}$$

where  $\varepsilon_n = [(\log n)/n]^{1/2} - 3/2n$ . Now  $\tilde{\Delta}_n(t, \omega)$  is a continuous function of  $t$ , hence

$$P\left\{\int_0^1 |\tilde{\Delta}_n(t)| dt \geq \varepsilon_n n^{1/2}\right\} \leq P\{\|\tilde{\Delta}_n\| \geq \varepsilon_n \cdot n^{1/2}\} \\ \leq P\{\|\Delta_n\| \geq \varepsilon_n \cdot n^{1/2}\} \leq 2 \exp(-2n\varepsilon_n^2) + A_1 n^{-1/2},$$

where the last inequality is a consequence of Lemma 1. With our choice of  $\varepsilon_n$  it is easy to see that  $2 \exp(-2n\varepsilon_n^2) = O(n^{-1/2})$ , and this completes the proof of Lemma 2.

In other words, in so far as the functional  $\mathcal{F}$  is concerned we make a small error in considering the  $\tilde{\Delta}_n(t)$  process instead of the  $\Delta_n(t)$  process.

Let us now consider the process  $\bar{F}_n(t, \omega)$  obtained from the process  $\tilde{F}_n(t, \omega)$  "by reflecting the latter process across the line  $y = x$ ". More precisely,  $\bar{F}_n(t, \omega)$  is the "random broken line" connecting the points  $(k/n, X_k^*(\omega))$ ,  $k = 0, 1, 2, \dots, n$ . Note that

$$(20) \quad \bar{F}_n(t, \omega) = X_k^*(\omega) \quad \text{when } t = k/n, \quad \text{whereas } \tilde{F}_n(t, \omega) = k/n \\ \text{when } t = X_k^*(\omega).$$

To put it another way, the  $\tilde{F}_n(t)$  process has jumps of fixed size  $1/n$  at the random times  $X_k^*$  whereas the  $\bar{F}_n(t)$  process has random jumps at the fixed times  $k/n$ .

We now define a process  $D_n(t, \omega)$  as follows

$$(21) \quad D_n(t, \omega) = n^{1/2}(\bar{F}_n(t, \omega) - t).$$

The process  $D_n(t, \omega)$  is of course not the same as the process  $\tilde{\Delta}_n(t, \omega)$ , nevertheless we have the following remarkable fact.

LEMMA 3.  $\mathcal{F}(\tilde{\Delta}_n) = \mathcal{F}(D_n)$ .

**Proof.** It is easily seen that Lemma 3 is equivalent to the following result<sup>(4)</sup>: Suppose  $f(t)$  is a monotone strictly increasing function on  $[0, 1]$  with  $f(0) = 0$ ,  $f(1) = 1$ . Then, as is well known,  $f(t)$  has an inverse function  $g(t)$  with the same

<sup>(4)</sup> For the proof of Lemma 3 I am indebted to Professor Ray Mayer. The geometrical interpretation of this result is due to Professor S. R. S. Varadhan.

properties. Now  $g(t)$  is obtained geometrically by reflecting the curve  $(t, f(t))$  across the line  $y=x$  and so it is enough to show that:

$$(22) \quad \mathcal{F}((f(t) - t)) = \mathcal{F}((g(t) - t)).$$

Now a moment's reflection will show that both sides of (22) equal, except for a constant factor independent of  $f$ , the volume of the solid of revolution obtained by rotating the curve  $(t, f(t)), 0 \leq t \leq 1$ , about the diagonal line  $y=x$ .

The reader can easily verify for himself that this result extends to monotone increasing functions of the form  $\tilde{F}_n(t, \omega)$  and their "reflections  $\bar{F}_n(t, \omega)$  across the line  $y=x$ " to conclude that

$$(23) \quad \mathcal{F}(n^{1/2}(\tilde{F}_n(t, \omega) - t)) = \mathcal{F}(n^{1/2}(\bar{F}_n(t, \omega) - t))$$

and this completes the proof of Lemma 3.

We now define a probability distribution function  $\tilde{V}_n(\lambda)$  as follows

$$\tilde{V}_n(\lambda) = P\{\mathcal{F}(\tilde{\Delta}_n) \leq \lambda\} = P\{\mathcal{F}(D_n) \leq \lambda\}.$$

It is an immediate consequence of Lemma 2 that

$$(24) \quad \tilde{V}_n(\lambda - \epsilon_n) - A_2 n^{-1/2} \leq V_n(\lambda) \leq \tilde{V}_n(\lambda + \epsilon_n) + A_2 n^{-1/2}, \quad \text{where } \epsilon_n = 2\left(\frac{\log n}{n}\right)^{1/2}.$$

We also note that (24) remains valid for sequences  $\epsilon'_n$  that decrease to zero no faster than does the sequence  $\epsilon_n$ .

We now study the  $D_n(t)$  process in more detail; in particular, following Breiman (Chapter 13 of [2]), we shall show that the  $D_n(t)$  process converges to the tied down Brownian motion process  $\dot{D}(t)$ , and what is most important we shall be able to estimate the rate of the convergence by techniques similar to those used in a previous paper of the author [11].

Let  $Y_1, Y_2, \dots, Y_{n+1}$  be  $n+1$  mutually independent random variables with the common distribution  $G(y) = 1 - \exp(-y)$  i.e. the random variables  $Y_i$  are exponentially distributed and we note that  $E(Y_i) = V(Y_i) = 1$ .

Let  $Z_k = \sum_{i=1}^k Y_i$ . It can be shown (see [2] or [7]) that the joint distribution of the random variables

$$(25) \quad Z_1/Z_{n+1}, Z_2/Z_{n+1}, \dots, Z_n/Z_{n+1}$$

is the same as the joint distribution of the random variables  $X_1^*, X_2^*, \dots, X_n^*$ . Let us now apply this result to the  $D_n(t)$  process defined at (21). Clearly  $D_n(t) = n^{1/2}(X_k^* - k/n)$  when  $t = k/n$ . Using (25) we can represent the  $D_n(t)$  process for  $t = k/n$  as follows:

$$(26) \quad D_n(t) = n^{1/2}\left(\frac{Z_k}{Z_{n+1}} - \frac{k}{n}\right) = \frac{n}{Z_{n+1}}\left(\frac{Z_k - k}{\sqrt{n}} - \frac{k}{n} \frac{Z_{n+1} - n}{\sqrt{n}}\right).$$

REMARK. The representation (26) is due to Breiman [2].

If we now define  $X_{ni} = (Y_i - 1)/\sqrt{n}$  it is clear that the random variables  $X_{ni}$  satisfy conditions (4) and (5) of [11], with  $a=2$ .

If we now define  $S_n(t)$  to be the following random broken line:

$$(27) \quad S_n(t) = \sum_{i=1}^k X_{ni} = \frac{Z_k - k}{\sqrt{n}}, \quad t = \frac{k}{n}$$

and define  $S_n(t)$  for  $k/n \leq t \leq (k+1)/n$  by linear interpolation, then  $D_n(t)$  can be written in terms of  $S_n(t)$  as follows:

$$(28) \quad D_n(t) = \frac{n}{Z_{n+1}}(S_n(t) - tS_n(1)) - t \frac{n}{Z_{n+1}} \cdot \frac{Y_{n+1}}{\sqrt{n}}.$$

Now by the law of large numbers  $\lim_{n \rightarrow \infty} (Z_{n+1}/n) = 1$ , with probability one and  $\lim_{n \rightarrow \infty} (t(n/Z_{n+1}))(Y_{n+1}/\sqrt{n}) = 0$ , with probability one. Thus the  $D_n(t)$  process is close to the  $S_n(t) - tS_n(1)$  process as  $n \rightarrow \infty$  and the  $S_n(t) - tS_n(1)$  process in turn is close to the tied down Brownian motion process  $D(t)$ . We now proceed to estimate how close is the  $D_n(t)$  process to the  $D(t)$  process.

It follows from (28) that

$$(29) \quad D_n(t) - (S_n(t) - tS_n(1)) = \left(1 - \frac{Z_{n+1}}{n}\right) D_n(t) - t \frac{Y_{n+1}}{\sqrt{n}}.$$

Hence,

$$(30) \quad \|D_n(t) - (S_n(t) - tS_n(1))\| \leq |1 - Z_{n+1}/n| \cdot \|D_n(t)\| + |Y_{n+1}/\sqrt{n}|.$$

We note that  $P\{|Y_{n+1}/\sqrt{n}| \geq (\log n)/\sqrt{n}\} = 1/n$ ; this is because  $P\{Y_{n+1} > y\} = \exp(-y)$ .

It follows from (30) that

$$(31) \quad P\{\|D_n(t) - (S_n(t) - tS_n(1))\| \geq \varepsilon_n\} \leq P\{|1 - Z_{n+1}/n| \cdot \|D_n(t)\| \geq \gamma_n\} + 1/n,$$

where  $\gamma_n = \varepsilon_n - (\log n)n^{-1/2}$ . We shall choose  $\varepsilon_n = \gamma(\log n)n^{-1/2}$  with  $\gamma > 1$ . Thus  $\gamma_n = (\gamma - 1)(\log n)n^{-1/2}$ . We now write  $\gamma_n = \alpha_n \cdot \beta_n$  with  $\alpha_n = (\gamma - 1)[(\log n)/n]^{1/2}$ , and  $\beta_n = (\log n)^{1/2}$ .

LEMMA 4. (i)  $P\{|Z_{n+1}/n - 1| \geq \alpha_n\} \leq 5(n/2\pi)^{1/2} \exp(-n\alpha_n^2/8) = r_1(n)$ ,

(ii)  $P\{\|D_n\| > \beta_n\} \leq 2 \exp(-2\beta_n^2) + A_1 n^{-1/2} = r_2(n)$ .

**Proof.** We prove (i) first.

$$P\{|Z_{n+1}/n - 1| > \alpha_n\} = P\{Z_{n+1} > n(1 + \alpha_n)\} + P\{Z_{n+1} < n(1 - \alpha_n)\}.$$

Now the probability density function of  $Z_{n+1}$  is  $(x^n/n!) \exp(-x)$ ,  $x > 0$ . Thus

$$P\{Z_{n+1} > n(1 + \alpha_n)\} = \frac{1}{n!} \int_{n(1 + \alpha_n)}^{\infty} x^n \exp(-x) dx.$$

The change of variable  $x = n\tau$  transforms the above integral into:

$$\frac{n^{n+1}e^{-n}}{n!} \int_{1 + \alpha_n}^{\infty} \exp[-n(\tau - 1 - \log \tau)] dt.$$

From Stirling's formula we conclude that  $n^{n+1}e^{-n}/n! \leq (n/2\pi)^{1/2}$ , hence

$$P\{Z_{n+1} > n(1 + \alpha_n)\} \leq \left(\frac{n}{2\pi}\right)^{1/2} \int_{1+\alpha_n}^{\infty} \exp[-n(\tau - 1 - \log \tau)] dt,$$

and a similar estimate holds for the other term.

Now we are assuming  $\alpha_n \downarrow 0$ , so by a well-known estimate of Courant-Hilbert (p. 523 of [3]) we conclude that:

$$\int_0^{1-\alpha_n} \exp[-n(\tau - 1 - \log \tau)] dt + \int_{1+\alpha_n}^{\infty} \exp[-n(\tau - 1 - \log \tau)] dt \leq 5 \exp(-n\alpha_n^2/8).$$

It follows that

$$P\{|Z_{n+1}/n - 1| > \alpha_n\} \leq 5(n/2\pi)^{1/2} \exp(-n\alpha_n^2/8),$$

and this completes the proof of (i).

We now prove (ii) by noting that

$$P\{\|D_n\| > \beta_n\} = P\{\|\tilde{D}_n\| > \beta_n\} \leq 2 \exp(-2\beta_n^2) + A_1 n^{-1/2},$$

the last inequality is a consequence of (19), (10) and (12).

LEMMA 5.  $P\{|Z_{n+1}/n - 1| \cdot \|D_n\| > \alpha_n \cdot \beta_n\} \leq r_1(n) + r_2(n)$ .

**Proof.** Let  $B_1 = \{\omega : |Z_{n+1}/n - 1| \leq \alpha_n\}$  and  $B_2 = \{\omega : \|D_n\| \leq \beta_n\}$ . Then the set  $B_1 \cap B_2$  is contained in the set  $\{\omega : |Z_{n+1}/n - 1| \cdot \|D_n\| \leq \alpha_n \cdot \beta_n\}$ . Taking complement ( $B'$  denotes the set theoretic complement of  $B$ ) we conclude

$$P\{|Z_{n+1}/n - 1| \cdot \|D_n\| > \alpha_n \cdot \beta_n\} \leq P\{B_1'\} + P\{B_2'\} \leq r_1(n) + r_2(n).$$

We now apply these estimates to (31) and conclude

LEMMA 6.  $P\{\|D_n(t) - (S_n(t) - tS_n(1))\| \geq \epsilon_n\} \leq r_1(n) + r_2(n) + 1/n$  where

$$\epsilon_n = \gamma(\log n)n^{-1/2}, \quad \gamma > 1;$$

$r_1(n) = 5(2\pi)^{-1/2}n^{4-(\gamma-1)^2/8}$ , and  $r_2(n) = 2n^{-2} + A_1n^{-1/2} \leq A_3n^{-1/2}$ , where  $A_3 > A_1$ .

REMARK.  $r_1(n)$  and  $r_2(n)$  are computed by substituting  $(\gamma - 1)[(\log n)/n]^{1/2}$  for  $\alpha_n$  and  $(\log n)^{1/2}$  for  $\beta_n$  in Lemma 4.

If we now choose  $\gamma \geq 4$  then  $r_1(n) = O(n^{-1/2})$ .

The next step is to use the Skorokhod representation, as in [11], to estimate

$$P\{\|S_n(t) - tS_n(1) - D(t)\| \geq \epsilon_n\}.$$

To this end we define  $W_n(t)$  as follows:

$$\begin{aligned} (32) \quad W_n(t) &= W\left(\frac{k}{n}\right) \quad \text{for } t = \frac{k}{n}, \\ &= W\left(\frac{k}{n}\right) + n\left[W\left(\frac{k+1}{n}\right) - w\left(\frac{k}{n}\right)\right]\left(t - \frac{k}{n}\right) \quad \text{for } \frac{k}{n} \leq t \leq \frac{k+1}{n}. \end{aligned}$$

In other words,  $W_n(t)$  is a polygonal approximation to the Brownian motion path  $W(t)$ . We define

$$(33) \quad \hat{D}_n(t) = W_n(t) - tW_n(1), \quad 0 \leq t \leq 1.$$

LEMMA 7.  $P\{\|\hat{D}_n(t) - D(t)\| \geq \epsilon_n\} \leq 2 \exp(-2n\epsilon_n^2)$ .

**Proof.**  $\hat{D}_n(t) - D(t) = (W_n(t) - W(t))$ , because  $W_n(1) = W(1)$  and now apply Lemma 3 of [11].

LEMMA 8.  $P\{\|(S_n(t) - tS_n(1)) - \hat{D}_n(t)\| \geq 4(\log n)^{1/2}n^{-1/5}\} \leq A_4n^{-1/5}$ .

**Proof.**  $S_n(t) - tS_n(1) - \hat{D}_n(t) = S_n(t) - W_n(t) + t(W_n(1) - S_n(1))$ . Thus

$$\begin{aligned} \|S_n(t) - tS_n(1) - \hat{D}_n(t)\| &\leq \|S_n(t) - W_n(t)\| + |W_n(1) - S_n(1)| \\ &\leq 2\|S_n(t) - W_n(t)\|. \end{aligned}$$

Hence

$$P\{\|(S_n(t) - tS_n(1)) - \hat{D}_n(t)\| \geq \epsilon_n\} \leq P\{\|S_n(t) - W_n(t)\| \geq \epsilon_n/2\}.$$

It is a consequence of Lemma 6 of [11] that for  $\epsilon_n/2 = 2(\log n)n^{-1/5}$  we have the following estimate:

$$(34) \quad P\{\|S_n(t) - W_n(t)\| \geq \epsilon_n/2\} = O(n^{-1/5}).$$

This completes the proof of Lemma 8.

We note that  $(\log n)^\beta n^{-1/5}$  decreases to zero at a much slower rate than quantities of the form  $(\log n)^\beta n^{-1/2}$ . Thus Lemmas 2, 4, 5, 6 and 7 remain valid for sequences  $\epsilon_n = O((\log n)^\beta n^{-1/5})$ .

We now define sets  $B_1, B_2, B_3$  as follows:

$$\begin{aligned} B_1 &= \{\|D_n(t) - (S_n(t) - tS_n(1))\| < \epsilon_n\}, \\ B_2 &= \{\|(S_n(t) - tS_n(1)) - \hat{D}_n(t)\| < \epsilon_n\}, \\ B_3 &= \{\|\hat{D}_n(t) - D(t)\| < \epsilon_n\}. \end{aligned}$$

Then we have  $\{\|D_n(t) - D(t)\| > 3\epsilon_n\} \subset B_1' \cup B_2' \cup B_3'$ . Combining Lemmas 6, 7, 8 yields

LEMMA 9.  $P\{\|D_n(t) - D(t)\| > 12(\log n)^{1/2}n^{-1/5}\} \leq A_5n^{-1/5}$ .

LEMMA 10.  $P\{|\mathcal{F}(D_n) - \mathcal{F}(D)| \geq \epsilon'_n\} \leq A_6n^{-1/5}$  where  $\epsilon'_n = 24(\log n)^{3/2}n^{-1/5}$ .

**Proof.** Let  $C_n = \{\|D_n(t) - D(t)\| < 12(\log n)^{1/2}n^{-1/5}\}$  and note that  $P\{C_n'\} \leq A_5n^{-1/5}$ , by Lemma 9. We observe that

$$|D_n(t, \omega)| \leq |D(t, \omega)| + \epsilon_n, \quad \omega \in C_n,$$

and

$$|D(t, \omega)| \leq |D_n(t, \omega)| + \epsilon_n, \quad \omega \in C_n, \epsilon_n = 12(\log n)^{1/2}n^{-1/5}.$$

From these two inequalities we conclude

$$(35) \quad |\mathcal{F}(D_n(t, \omega)) - \mathcal{F}(D(t, \omega))| \leq 2\epsilon_n \int_0^1 |D(t, \omega)| dt + 3\epsilon_n^2, \quad \omega \in C_n.$$

Applying the same reasoning used in Lemma 2 yields the following estimate:

$$\begin{aligned} P\{|\mathcal{F}(D_n) - \mathcal{F}(D)| \geq \epsilon'_n\} &\leq P\left\{\int_0^1 |D(t)| \geq \lambda_n\right\} + P\{C'_n\} \\ &\leq P\{\|D(t)\| \geq \lambda_n\} + A_5 n^{-1/5} \leq 2 \exp(-2\lambda_n^2) + A_5 n^{-1/5}, \end{aligned}$$

where  $\lambda_n = \epsilon'_n/2\epsilon_n - 3\epsilon_n/2$ . Clearly we want  $\lambda_n \rightarrow +\infty$ , so we choose  $\epsilon'_n = 2(\log n)\epsilon_n = 24(\log n)^{3/2}n^{-1/5}$ . With this choice of  $\epsilon'_n$  we obtain the estimate:

$$P\{|\mathcal{F}(D_n) - \mathcal{F}(D)| \geq 24(\log n)^{3/2}n^{-1/5}\} \leq A_6 n^{-1/5}$$

and this completes the proof of Lemma 10.

Now it follows from results of Anderson and Darling (p. 202 of [1]) that  $V(\lambda)$  is a nice smooth function with the property

$$(36) \quad |V(\lambda+h) - V(\lambda)| \leq L|h|$$

where  $L$  is a constant independent of  $h$ . And it follows from Lemma (10) that

$$(37) \quad V(\lambda - \epsilon'_n) - A_6 n^{-1/5} \leq \tilde{V}_n(\lambda) \leq V(\lambda + \epsilon'_n) + A_6 n^{-1/5}.$$

We now apply (36) to (37) to conclude:

$$(38) \quad V(\lambda) - L\epsilon'_n - A_6 n^{-1/5} \leq \tilde{V}_n(\lambda) \leq V(\lambda) + L\epsilon'_n + A_6 n^{-1/5}$$

where  $A_6$  and  $L$  are independent of  $\lambda$ .

It follows from (24) and (38) that

$$(39) \quad V(\lambda) - 2L\epsilon'_n - A_7 n^{-1/5} \leq V_n(\lambda) \leq V(\lambda) + 2L\epsilon'_n + A_7 n^{-1/5}$$

or equivalently

$$(40) \quad |V_n(\lambda) - V(\lambda)| \leq 2L\epsilon'_n + A_7 n^{-1/5}.$$

It is clear from our construction that the right-hand side of (40) is

$$O((\log n)^{3/2}n^{-1/5})$$

and this completes the proof of our theorem.

**III. Open problems<sup>(5)</sup>.** The rate of convergence obtained in Theorem 1 is certainly not the best possible. By suitable juggling with the Skorokhod representation it is possible to improve the exponent slightly from 1/5 to 1/4. The real problem of course is to improve the rate of convergence to  $(\log n)^\delta n^{-1/2}$ . This means sharpening

---

<sup>(5)</sup> **Added in proof.** The author has recently applied these results to the study of the asymptotic distribution of linear combinations of order statistics. In particular we can derive results similar to those recently obtained by Chernoff, Gastwirth and Johns.

the estimate of Lemma 8, because all the other estimates are  $O((\log n)^\beta n^{-1/2})$  for some  $\beta > 0$ .

Another interesting problem is to extend this result to other functionals of the  $\Delta_n(t, \omega)$  process e.g.

$$\mathcal{F}(\Delta_n(t, \omega)) = \int_0^1 (\Delta_n(t, \omega))^2 d\psi(t)$$

where  $d\psi(t)$  is a suitable measure on  $[0, 1]$  (cf. [1]).

#### REFERENCES

1. D. A. Darling and T. W. Anderson, *Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes*, Ann. Math. Statist. **23** (1952), 193–212.
2. L. Breiman, *Probability*, Addison-Wesley, Reading, Mass., 1968.
3. R. Courant and D. Hilbert, *Methods of mathematical physics*, Vol. 1, Interscience, New York, 1963.
4. M. Donsker, *Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems*, Ann. Math. Statist. **23** (1952), 277–281.
5. J. L. Doob, *Heuristic approach to the Kolmogorov-Smirnov theorems*, Ann. Math. Statist. **20** (1949), 393–403.
6. B. V. Gnedenko, V. S. Korolyuk and A. V. Skorokhod, *Asymptotic expansions in probability theory*, Proc. Fourth Berkeley Sympos. Math. Stat. and Prob. (1960), Vol. II, Univ. California Press, Berkeley, Calif., 1961, pp. 153–170.
7. S. Karlin, *A first course in stochastic processes*, Academic Press, New York, 1966.
8. A. N. Kolmogorov, *Sulla determinazione empirica delle leggi di probabilita*, Giorn. Ist. Ital. Attuari **4** (1933), 1–11.
9. K. R. Parthasarathy, *Probability measures on metric spaces*, Academic Press, New York, 1967.
10. Yu. V. Prochorov, *Convergence of random processes and limit theorems in probability*, Theor. Probability Appl. **1** (1956), 157–214.
11. W. A. Rosenkrantz, *On rates of convergence for the invariance principle*, Trans. Amer. Math. Soc. **129** (1967), 542–552.
12. A. V. Skorokhod, *Studies in the theory of random processes*, Addison-Wesley, Reading, Mass., 1965.

COURANT INSTITUTE OF MATHEMATICAL SCIENCES, NEW YORK UNIVERSITY,  
NEW YORK, NEW YORK