RANDOM CROSSINGS OF CUMULATIVE DISTRIBUTION FUNCTIONS

MEYER DWASS

1. Introduction. Let $X_1, \cdots, X_n$ be n independent and identically distributed random variables, each with continuous c.d.f. (cumulative distribution function), $F(x)$. Let $F_n(x)$ be the empirical c.d.f. of the n random variables and let $N_1(n)$ be the number of times F_n equals F. There is no loss of generality in supposing that the X_i's are distributed uniformly over the interval (0, 1), and to be specific, $N_1(n)$ is defined by

$$N_1(n) = \text{number of indices, } i, \text{ for which } F_n(i/n) = i/n, \quad i = 1, \cdots, n.$$

Similary, let $X_1, \cdots, X_n, \cdots, Y_1, \cdots, Y_n$ be $2n$ independent random variables, each with the same continuous c.d.f., $F(x)$, and let F_n, G_n denote the empirical c.d.f.'s of the X_i's and Y_i's respectively. Let $N_2(n)$ be the number of times F_n equals G_n. That is.

$$N_2(n) = \text{number of indices } i \text{ for which } F_n(X_i) = G_n(X_i),$$
$$\text{plus}$$
$$\text{number of indices } i \text{ for which } F_n(Y_i) = G_n(Y_i), \quad i = 1, \cdots, n.$$

The purpose of this paper is to show that

$$\lim_{n\to\infty} P\left(\frac{N_1(n)}{\sqrt{2n}} < t\right) = \lim_{n\to\infty} P\left(\frac{N_2(n)}{\sqrt{4n}} < t\right) = 1 - e^{-t^2} .$$

The methods for obtaining these results are practically the same for N_1 and N_2, so the first case is treated with somewhat greater detail. In both cases, the random variables are related to other random variables on appropriate stochastic processes with independent increments, to obtain generating functions for the moments of N_i. The Karamata Tauberian theorem is then applied to describe the asymptotic behavior of these moments.

2. Some preliminaries on the Poisson process. Let $Y(t)$ be the Poisson process with stationary independent increments, $t \geqq 0$, $Y(0) = 0$, $E[Y(1)] = 1$. Consider γt, the straight line coming out of the origin with slope $\gamma > 1$. The random function $Y(t)$ can equal γt at times $1/\gamma$, $2/\gamma$, etc. The event that $Y(t) = \gamma t$ is a recurrent event in the sense of Feller [4]. Because γ is greater than 1, this recurrent event is an *uncertain* one. It was shown by Baxter and Donsker [1] that

Received February 23, 1960. Research done under contract with the U. S. Office of Naval Research. Contract Nonr-1228(10), project NR 047-021.

$$P[Y(t) < \gamma t, \text{ all positive } t] = 1 - 1/\gamma .$$

A completely elementary proof of this fact was given by Dwass [3]. In other words, the probability that the uncertain recurrent event under discussion *never* takes place is $1 - 1/\gamma$. To introduce some specific notation, let $N =$ number of times that $Y(t)$ equals γt. That is,

$$N = \text{number of indices } i \text{ for which } Y(i/\gamma) = \gamma(i/\gamma) = i, \quad i = 1, 2, \cdots.$$

The random variable, N is geometrically distributed, specificially,

$$P(N = k) = (1/\gamma)^k(1 - 1/\gamma) ,$$

and for the rth factorial moment we have,

$$EN^{(r)} = EN(N-1)(N-2)\cdots(N-r+1) = r!/(r-1)^r . \tag{2.1}$$

3. A generating function for $E[N_1^{(r)}(n)]$. The link between the random variables N and $N_1(n)$ lies in the following lemma.

LEMMA 3.1. *The conditional distribution of N given that $Y(t) = \gamma t$ for the last time at time $t = n/\gamma$ is exactly the same as the distribution of $N_1(n)$.*

Proof of Lemma 3.1. This follows directly from the well-known fact that the places where the jumps of $Y(t)$ occur in the interval $(0, a)$ are distributed as n randomly chosen points in $(0, a)$ under the condition that $Y(a) = n$.

Making use of this lemma, we can compute the rth factorical moment of N in the following iterative way. Let A_k denote the event that the last crossing of γt by $Y(t)$ takes place at time k/γ. Then

$$E(N^{(r)}) = \sum_{k=0}^{\infty} E(N^{(r)} \mid A_k)P(A_k) .$$

Since

$$P(A_k) = (k/\gamma)^k e^{-k/\gamma}(1 - 1/\gamma)/k!$$

and

$$E(N^{(r)} \mid A_k) = E[N_1^{(r)}(k)], \qquad (k, 0, 1, 2, \cdots) ,$$

we have, making use of (2.1), the following theorem.

THEOREM.

$$\sum_{k=0}^{\infty} \frac{e^{-k}k^k}{k!} EN_1^{(r)}(k) \left(\frac{e^{1-1/\gamma}}{\gamma}\right)^k = \frac{r!\,\gamma}{(\gamma-1)^{r+1}} . \tag{3.1}$$

REMARKS.

(a) In (3.1), $e^{-k}k^k/k!$ should be understood to be 1 when $k = 0$.

(b) $u = e^{1-1/\gamma}$ is a strictly decreasing function of $1/\gamma$, for $\gamma \geqq 1$, and maps $(1, \infty)$ onto $(0, 1)$. Let $1/\gamma = P(u)$ denote the inverse function. Then (2.2) can be rewritten,

$$\sum_{k=0}^{\infty} \frac{e^{-k}k^k}{k!} EN_1^{(r)}(k)u^k = \frac{r!}{P(u)[P^{-1}(u) - 1]^{r+1}} = h(u) , \qquad 0 \leqq u < 1 . \tag{3.2}$$

Since

$$\lim_{x \to 1} \frac{1 - xe^{1-x}}{(x-1)^2} = 1/2 ,$$

or equivalently,

$$\lim_{u \to 1} \frac{1 - u}{(P^{-1}(u) - 1)^2} = 1/2 ,$$

it follows that

$$\lim_{u \to 1} (1 - u)^{(r+1)/2} h(u) = \frac{r!}{2^{(r+1)/2}} . \tag{3.3}$$

If the coefficients of u^k in $h(u)$ form an increasing sequence, then Karamata's Tauberian theorem is applicable and we could conclude that the sum of the first k coefficients of powers of u in $(1 - u)h(u)$ is asymptotically equal to

$$k^{(r-1)/2} \frac{r!}{2^{(r+1)/2}\Gamma\left(\frac{r+1}{2}\right)} ,$$

or equivalently,

$$\lim_{k \to \infty} \frac{\frac{e^{-k}k^k}{k!} EN_1^{(r)}(k)}{k^{(r-1)/2}} = \frac{r!}{2^{(r+1)/2}\Gamma\left(\frac{r+1}{2}\right)}$$

$$\lim_{k \to \infty} \frac{EN_1^{(r)}(k)}{k^{r/2}} = \frac{r!\sqrt{\pi}}{2^{r/2}\Gamma\left(\frac{r+1}{2}\right)} = 2^{r/2}\Gamma\left(\frac{r}{2} + 1\right)$$

by the "duplication formula" for the gamma function (p. 240 [6]).

Since the asymptotic behavior of the rth factorial moment is the same as that of the rth ordinary moment, we would have finally,

$$\lim_{k \to \infty} E\left(\frac{N_1(k)}{\sqrt{2k}}\right)^r = \Gamma\left(\frac{r}{2} + 1\right) = \int_0^{\infty} x^r f(x) dx$$

where $f(x)$ is the probability density function

$$f(x) = \begin{cases} 2xe^{-x^2}, & x \geqq 0 \\ 0, & x < 0 \end{cases} ,$$

However, it is not at all clear that the usual conditions for Karamata's theorem to hold are applicable, and a slightly more delicate argument is required.

4. The limiting distribution of $N_1(n)$. Following the discussion in the last section, the main effort which remains is to justify the applicability of Karamata's theorem.

LEMMA 4.1. *Let $a_i(u)$, $(i = 1, \cdots, r)$ be power series having positive, non-decreasing coefficients. Then $a(u) = \prod_i a_i(u)$ has the same property.*

Proof of Lemma 4.1. $a_i(u)$ has positive, non-decreasing coefficients means that the coefficients of $(1 - u)a_i(u)$ are non-negative.

$$(1 - u) \prod_i a_i(u) = \prod_i [(1 - u)a_i(u)](1 - u)^{-(r-1)}$$

is a product of power series all with non-negative coefficients, which completes the proof.

LEMMA 4.2.

(a) $$\sum_{k=0}^{\infty} \frac{k^k e^{-k}}{k!} u^k = f(u) = 1/(\gamma - 1)$$

(b) $$c(1 - u)^{-1/2} - f(u) = g(u)$$

is a power series with positive, non-decreasing coefficients if c is sufficiently large.

Proof of *Lemma* 4.2. Part (a) follows from (3.1) for $r = 0$. The coefficients of $(1 - u)^{-1/2}$ are of the order of $1/\sqrt{k}$ and strictly positive. The coefficients of $-f(u)$ are strictly increasing and also of the order of $1/\sqrt{k}$. Hence choosing c sufficiently large will guarantee the result.

Finally, we want to state the following form of Karamata's theorem.

LEMMA 4.3. *Let $a(u) = \sum_{k=0}^{\infty} a_k u^k$ where $\{a_k\}$ is a non-decreasing sequence, and suppose $(1 - u)^{\gamma} a(u) \to A$ as $u \to 1$, for $\gamma \geqq 0$. Then*

$$\frac{a^k}{k^{\gamma-1}} \to \frac{A}{\Gamma(\gamma)}$$

as $k \to \infty$.

Proof of Lemma 4.3. For $\gamma > 1$ the result follows from the conventional form of Karamata's theorem (for example see Theorem 4.3, p. 192, [5]) by considering that

$$(1-u)^{\gamma-1}(1-u)a(u) \to A$$

and that the partial sums of coefficients in $(1-u)a(u)$ are a_k.

For $0 \leqq \gamma < 1$ we have that

$$a_1 + \cdots + a_k \sim \frac{A}{\Gamma(\gamma+1)} k^{\gamma}$$

and we can apply Hilfassatz 3, p. 517, Doetsch, [2], to conclude that

$$a_k \sim \frac{A\gamma}{\Gamma(\gamma+1)} k^{\gamma} = \frac{A}{\Gamma(\gamma)} k^{\gamma} .$$

We can now prove the following.

THEOREM.

$$\lim_{n\to\infty} P\left(\frac{N_1(n)}{\sqrt{2n}} < t\right) = 2\int_0^t xe^{-x^2}dx = 1 - e^{-t^2}$$

Proof. The limiting distribution is determined by its moments, hence it is sufficient to show that

$$\lim_{n\to\infty} E\left(\frac{N_1(n)}{\sqrt{2n}}\right)^r = 2\int_0^{\infty} x^{r+1}e^{-x^2}dx = \Gamma(r/2+1) , \qquad r = 1, 2, \cdots .$$

Referring to (3.2) and to Lemma 4.2, we can write

$$(4.1) \qquad \begin{aligned} h(u) &= r!\,[1+f(u)][f(u)]^r \\ &= [c(1-u)^{-1/2} - g(u) + 1][c(1-u)^{-1/2} - g(u)]^r . \end{aligned}$$

Since $g(u)$ has positive and increasing coefficients then by Lemma 4.1 so does $(1-u)^{-m/2}[g(u)]^n$ for m, n positive integers, because

$$(1-u)[(1-u)^{-m/2}(g(u))^n] = (1-u)^{-m/2}(1-u)[g(u)]^n$$

has positive coefficients. Hence by Karamata's theorem, since

$$(1-u)^{(m+n)/2}(1-u)^{-m/2}[g(u)]^n \to \left(c - \frac{1}{\sqrt{2}}\right)^n ,$$

the coefficients of $(1-u)^{-m/2}[g(u)]^n$ are asymptotically equivalent to

$$\frac{(c-1/\sqrt{2})^n}{\Gamma\left(\frac{m+n}{2}\right)} k^{(m+n)/2} .$$

On expanding the right side of (4.1), an elementary computation yields the result that the coefficients of $h(u)$ are asymptotically equivalent to

$$\frac{r!}{2^{(r+1)/2}\Gamma\left(\frac{r+1}{2}\right)} k^{(r-1)/2} .$$

According to the discussion in § 4, we conclude then that

$$\lim_{k\to\infty} E\left(\frac{N_1(k)}{\sqrt{2k}}\right)^r = \Gamma\left(\frac{r}{2}+1\right) ,$$

which completes the proof of the theorem.

5. The limiting distribution of $N_2(n)$. In this section we prove the following.

THEOREM.

$$\lim_{n\to\infty} P\left(\frac{N_2(n)}{\sqrt{4n}} < t\right) = 1 - e^{-t^2}$$

The main points of the proof are essentially the same as in the preceding theorem, so we offer an outline of the method only.

Let $X_1, X_2, \cdots$ be a sequence of independent, identically distributed random variables such that

$$X_i = \begin{cases} 1 \text{ with probability } p, \\ 0 \text{ with probability } 1-p. \end{cases}$$

and let S_n denote the sum of the first n random variables.

The event that for a positive integer n, $S_{2n} = n$, is a well-known recurrent event, representing return to the origin, in a discrete random walk on the line. Suppose $p < 1/2$. Then the probability that the recurrent event never takes place is $1 - 2p$. (See Feller, p. 288, [4].) Using N exactly as above, let $N =$ number of indices i for which $S_{2i} = i$, $i = 1, 2, \cdots$. As before, N is a geometric random variable, such that

$$P(N = k) = (1-2p)(2p)^k ,$$

and hence

$$EN^{(r)} = \frac{r!}{\left(\frac{1}{2p} - t\right)^r} .$$

Analogous to Lemma 3.1 is the following combinatorial lemma.

LEMMA 5.1. *The conditional distribution N given that $S_{2i} = i$ for the last time when $i = n$ is exactly the same as the distribution of*

$N_2(n)$. We omit the proof which is elementary.

Let A_k denote the event that $S_{2i} = i$ for the last time when $i = k$. Then

$$EN^{(r)} = \sum_{k=0}^{\infty} E(N^{(r)} \mid A_k)P(A_k)$$
$$= \sum_{k=0}^{\infty} EN_2^{(r)}(k)\binom{2k}{k}p^k(1-p)^k(1-2p) .$$

Hence

$$f(u) = \sum_{k=0}^{\infty} EN_2^{(r)}(k)\binom{2k}{k}4^k u^k = \frac{r!}{2h(u)\left(\dfrac{1}{2h(u)} - 1\right)^{r+1}} \tag{5.1}$$

where $4p(1-p) = u$, $0 \leqq p \leqq 1/2$, is an increasing function of p which maps $(0, 1/2)$ onto $(0, 1)$, and where $p = h(u)$ is the inverse function.

We next notice that

$$\lim_{u \to 1} (1-u)^{(r+1)/2} f(u) = r! ,$$

This follows from the fact that

$$\lim_{p \to 1} \frac{1 - 4p(1-p)}{\left(\dfrac{1}{2p} - 1\right)^2} = 1 .$$

The application of the Karamata theorem can now be justified exactly as before. In fact if $g(u)$ is defined in terms of $f(u)$ as in § 4, then the details go through exactly word for word. Hence we conclude that

$$\lim_{k \to \infty} E \frac{N_2^{(r)}(k)\binom{2k}{k}4^{-k}}{k^{(r-1)/2}} = \frac{r!}{\Gamma\left(\dfrac{r+1}{2}\right)} ,$$

hence

$$\lim_{k \to \infty} E \frac{N_2^{(r)}(k)}{(4k)^{r/2}} = \frac{r! \sqrt{\pi}}{2^r \Gamma\left(\dfrac{r+1}{2}\right)} = \Gamma\left(\frac{r}{2} + 1\right)$$

which completes the proof.

6. Final remarks. The asymptotic distribution of $N_1(n)$ has been studied by N. V. Smirnov in "Sur les écarts de la courbe de distribution empirique", Mat. Sbornik, 6 (48), pp. 3–26 (1939), (Russian, French summary). His methods are not based on the Karamata Tauberian theorem and seem considerably more complicated than those of this paper, though he actually dealt with a more general situation. Also, the referee has kindly pointed out that the random variable $N_2(n)$ is related to a random variable studied by W. Feller in "The number of zeros and of changes of sign in a symmetric random walk", L'Enseignement Mathématique, III, 3, (1957), 229–235.

REFERENCES

1. G. Baxter and M. D. Donsker, *On the distribution of the supremum functional for processes with stationary independent increments*, Trans. Amer. Math. Soc., **85** (1957), 73-87.
2. G. Doetsch, *Handbuch der Laplace-Transformation*, Band I, Basel: Birkhäuser Verlag, 1950.
3. Meyer Dwass, *The distribution of a generalized D_n^+ statistic*, Ann. Math. Stat. 30, December, (1958), 1024-1028.
4. William Feller, *An introduction to probability theory and its applications*, 2nd edition, New York: John Wiley and Sons, 1957.
5. D. V. Widder, *The Laplace Transform*, Princeton: Princeton University Press, 1946.
6. E. T. Whittaker, and G. N. Watson, *A Course of Modern Analysis*, 4th ed. Cambridge: Cambridge Press, 1950.

NORTHWESTERN UNIVERSITY