HEURISTIC APPROACH TO THE KOLMOGOROV-SMIRNOV THEOREMS[1]

By J. L. Doob

University of Illinois

1. Introduction and summary. Asymptotic theorems on the difference between the (empirical) distribution function calculated from a sample and the true distribution function governing the sampling process are well known. Simple proofs of an elementary nature have been obtained for the basic theorems of Komogorov[2] and Smirnov[3] by Feller,[4] but even these proofs conceal to some extent, in their emphasis on elementary methodology, the naturalness of the results (qualitatively at least), and their mutual relations. Feller suggested that the author publish his own approach (which had also been used by Kac), which does not have these disadvantages, although rather deep analysis would be necessary for its rigorous justification. The approach is therefore presented (at one critical point) as heuristic reasoning which leads to results in investigations of this kind, even though the easiest proofs may use entirely different methods.

No calculations are required to obtain the qualitative results, that is the existence of limiting distributions for large samples of various measures of the discrepancy between empirical and true distribution functions. The numerical evaluation of these limiting distributions requires certain results concerning the Brownian movement stochastic process and its relation to other Gaussian processes which will be derived in the Appendix.

2. The problem. Let $x_1, x_2, \cdots$ be mutually independent random variables with a common distribution function $F(\lambda)$,

$$F(\lambda) = Pr\{x_j \leq \lambda\}.$$

In statistical language $x_1, \cdots, x_n$ form a sample of n drawn from the distribution with distribution function $F(\lambda)$. Let $\nu_n(\lambda)$ be the number of these x_j's which are $\leq \lambda$. According to the strong law of large numbers, for each λ

$$\lim_{n\to\infty} \frac{\nu_n(\lambda)}{n} = F(\lambda) \tag{2.1}$$

with probability 1. For fixed n $\nu_n(\lambda)/_n$ is itself a distribution function (which depends on the sample values $x_1, \cdots, x_n$) the *empirical distribution function*, and an elaboration of the argument which led to (2.1) shows that (2.1) is true

[1] Research connected with a probability project at Cornell University under an ONR contract.

[2] *Inst. Ital. Atti., Giorn.*, Vol. 4 (1933), pp. 83–91.

[3] *Rec. Math.* (*Matematiceskii Sbornik*), N.S. 6, Vol. 48 (1939), pp. 3–26, *Bull. Math. Univ. Moscou*, Vol. 2 (1939), fasc. 2.

[4] *Annals of Math. Stat.*, Vol. 19 (1948), pp. 177–189.

uniformly in λ, with probability 1; that is if

$$D_n = \underset{-\infty<\lambda<\infty}{\text{L.U.B.}} \left| \frac{\nu_n(\lambda)}{n} - F(\lambda) \right|, \tag{2.2}$$

then D_n is a random variable and

$$\lim_{n\to\infty} D_n = 0$$

with probability 1.[5] This result would be of limited practical statistical importance except that the distribution of D_n does not depend on the distribution function $F(\lambda)$ if $F(\lambda)$ is continuous. In fact in that case the random variables $F(x_1)$, $F(x_2)$, $\cdots$ are mutually independent and each is uniformly distributed in the interval (0, 1); if $\hat{\nu}_n(\lambda)$ is the number of $F(x_j)$'s $\leq \lambda$, for $j \leq n$,

$$\underset{0\leq\mu\leq 1}{\text{L.U.B.}} \left| \frac{\hat{\nu}_n(\mu)}{n} - \mu \right| = \underset{-\infty<\lambda<\infty}{\text{L.U.B.}} \left| \frac{\nu_n(\lambda)}{n} - F(\lambda) \right|.$$

Thus it is no restriction, replacing x_j by $F(x_j)$ if necessary, in finding the distribution of D_n to assume that $F(\lambda) = \lambda$ for $0 \leq \lambda \leq 1$, and

$$D_n = \underset{0\leq\lambda\leq 1}{\text{L.U.B.}} \left| \frac{\nu_n(\lambda)}{n} - \lambda \right|. \tag{2.2'}$$

The results will hold for D_n defined by (2.2) for any continuous $F(\lambda)$. We shall also consider D_n^+ and D_n^-, defined by

$$\begin{aligned} D_n^+ &= \underset{0\leq\lambda\leq 1}{\text{L.U.B.}} \left[\frac{\nu_n(\lambda)}{n} - \lambda \right], \\ D_n^- &= -\underset{0\leq\lambda\leq 1}{\text{G.L.B.}} \left[\frac{\nu_n(\lambda)}{n} - \lambda \right], \end{aligned} \tag{2.3}$$

and again the results will hold (with the obvious definitions of D_n^+ and D_n^- in the general case) for every continous $F(\lambda)$.

The problem is to find the limiting distributions of (properly normalized) D_n, D_n^+, D_n^- when $n \to \infty$.

3. Derivation of the Kolmogorov and Smirnov theorems.

Define

$$x_n(t) = n^{\frac{1}{2}} \left(\frac{\nu_n(t)}{n} - t \right), \qquad 0 \leq t \leq 1.$$

Since $\nu_n(0) = 0$ with probability 1 and $\nu_n(t) - \nu_n(s)$ is the number of successes in n independent trials, with probability $t - s$ of success in each trial, $\nu_n(t) - \nu_n(s)$ has expectation $n(t - s)$ and variance $n(t - s)[1 - (t - s)]$. Hence

$$\begin{aligned} E\{x_n(t)\} &= 0, & 0 \leq t \leq 1; \\ E\{[x_n(t) - x_n(s)]^2\} &= (t - s)[1 - (t - s)], & 0 \leq s \leq t \leq 1. \end{aligned} \tag{3.1}$$

[5] Cf. M. Fréchet, *Généralités sur les probabilités. Variables aléatoires*, Paris, 1937, pp. 260–261.

Now let $\{x(t)\}$ be a one parameter family of random variables, $0 \leq t \leq 1$ with the following properties:

(a) for each j if $0 \leq t_1 < \cdots < t_j \leq 1$ the j-variate distribution of the random variables $x(t_1), \cdots, x(t_j)$ is Gaussian;

(b) (3.1) holds, that is

$$E\{x(t)\} = 0, \qquad 0 \leq t \leq 1; \tag{3.1'}$$

$$E\{[x(t) - x(s)]^2\} = (t - s)[1 - (t - s)], \qquad 0 \leq s \leq t \leq 1.$$

(c) $Pr\{x(0) = 0\} = 1$.

According to the central limit theorem, the j variate distribution of $x_n(t_1), \cdots, x_n(t_j)$ is asymptotically that of $x(t_1), \cdots, x(t_j)$; in fact the normalizing factor $n^{\frac{1}{2}}$ in the definition of $x_n(t)$ and the choice of means and variances in (3.1′) were made precisely to bring this about. As far as first and second moments are concerned the $x_n(t)$ and $x(t)$ processes are identical; when $n \to \infty$ the distributions, or at least the j variate ones mentioned, become identical also.

We shall assume, until a contradiction frustrates our devotion to heuristic reasoning, that *in calculating asymptotic $x_n(t)$ process distributions when $n \to \infty$ we may simply replace the $x_n(t)$ processes by the $x(t)$ process*. It is clear that this cannot be done in all possible situations, but let the reader who has never used this sort of reasoning exhibit the first counter example.

The $x(t)$ process has continuous sample functions (cf. Appendix). Define

$$D = \operatorname*{Max}_{0 \leq t \leq 1} |x(t)|,$$

$$D^+ = \operatorname*{Max}_{0 \leq t \leq 1} x(t),$$

$$D^- = -\operatorname*{Min}_{0 \leq t \leq 1} x(t).$$

Then in accordance with our substitution principle $n^{\frac{1}{2}}D_n$, $n^{\frac{1}{2}}D_n^+$, $n^{\frac{1}{2}}D_n^-$ have as n becomes infinite the distributions of D, D^+, D^- respectively. (The latter two are the same because the $-x(t)$ process is stochastically identical with the $x(t)$ process.) Thus these simple qualitative considerations have led to the existence of the limiting distributions derived and evaluated by Kolmogorov, who proved:

THEOREM[6] (Kolmogorov).

$$\lim_{n \to \infty} Pr\{n^{\frac{1}{2}}D_n \geq \lambda\} = 2 \sum_1^\infty (-1)^{m+1} e^{-2m^2\lambda^2}; \tag{3.2}$$

$$\lim_{n \to \infty} Pr\{n^{\frac{1}{2}}D_n^+ \geq \lambda\} = \lim_{n \to \infty} Pr\{nD_n^- \geq \lambda\} = e^{-2\lambda^2}. \tag{3.3}$$

To complete our treatment we shall prove in the Appendix that

$$Pr\{D \geq \lambda\} = 2 \sum_1^\infty (-)^{m+1} e^{-2m^2\lambda^2}; \tag{3.2'}$$

[6] In Feller's paper (*loc. cit.*, p. 178, equation (1.4)) the factor 2 in the exponent was omitted by the printer. The same misprint occurs in Smirnov's table of the values of the series in our (3.2), *Annals of Math. Stat.*, Vol. 19 (1948), pp. 279–281.

$$(3.3') \qquad Pr\{D^+ \geq \lambda\} = Pr\{D^- \geq \lambda\} = e^{-2\lambda^2},$$

so that in fact the above considerations have led not only to the existence but to the evaluation of the asymptotic distributions. (Actually we shall prove somewhat more general results about the $x(t)$ process.)

So much for the Kolmogorov theorems. Smirnov obtained results (also independent of the given continuous distribution function $F(\lambda)$) of a somewhat different nature. Let x_1^*, x_2^*, $\cdots$ be mutually independent random variables with the same individual distributions as the x_j's, that is each distributed uniformly in the interval (0, 1); define $\nu_n^*(\lambda)$ as the number of the first n x_j's which are $\leq \lambda$. Smirnov considered the difference between empirical distribution functions,

$$D_{mn} = \underset{0 \leq \lambda \leq 1}{\text{L.U.B.}} \left| \frac{\nu_m(\lambda)}{m} - \frac{\nu_n^*(\lambda)}{n} \right|,$$

as well as D_{mn}^+ and D_{mn}^- defined in the obvious way. To avoid stressing the obvious we consider only the D_{mn}.

THEOREM (Smirnov). *If m, $n \to \infty$ in such a way that $\frac{m}{n} \to r$, and if $N = mn/(m + n)$,*

$$(3.4) \qquad \lim_{n \to \infty} Pr\{N^{\frac{1}{2}} D_{mn} \geq \lambda\} = 2 \sum_1^\infty (-1)^{m+1} e^{-2m^2\lambda^2}.$$

To derive this result define an $x^*(t)$ process stochastically identical with the $x(t)$ process but independent of it. Then if $x_n^*(t)$ is defined by

$$x_n^*(t) = n^{\frac{1}{2}} \left(\frac{\nu_n^*(t)}{n} - t \right),$$

we identify, in accordance with our heuristic principle the process with variables

$$\{x(t) - r^{\frac{1}{2}} x^*(t)\}$$

with the one with variables

$$\left\{ x_m(t) - \left(\frac{m}{n}\right)^{1/2} x_n^*(t) \right\}.$$

Doing this leads to the fact that the distribution of

$$(N)^{1/2} D_{mn} = \left(\frac{n}{m + n}\right)^{1/2} \underset{0 \leq t \leq 1}{\text{L.U.B.}} \left| x_m(t) - \left(\frac{m}{n}\right)^{1/2} x_n^*(t) \right|$$

converges to that of

$$\left(\frac{1}{1 + r}\right)^{1/2} \underset{0 \leq t \leq 1}{\text{Max}} \left| x(t) - (r)^{1/2} x^*(t) \right|.$$

Now the $x(t)$ process and the process with variables

$$\left\{\frac{x(t) - (r)^{1/2}x^*(t)}{(1 + r)^{1/2}}\right\}$$

are stochastically identical. Hence we are led to the conclusion that the distribution of $(N)^{1/2}D_{mn}$ converges to that of D, and this is Smirnov's theorem, stated above. (The method we use does not seem applicable to Smirnov's deeper theorems on the number of intersections between empirical and true distribution curves or between pairs of empirical distribution curves.)

APPENDIX

4. The Brownian movement process. Consider any Gaussian stochastic process, with random variables $\{x(t)\}$ where t varies in some interval. That is, we assume that for each t in the interval $x(t)$ is a random variable and that for any $j \geq 1$ if $t_1 < \cdots < t_j$ are in the interval the j variate distribution of $x(t_1), \cdots, x(t_j)$ is Gaussian. In the following we shall always assume that $E\{x(t)\} \equiv 0$. Then the process is determined stochastically by the covariance function

$$r(s, t) = E\{x(s)x(t)\}.$$

In particular, if the range of parameter is the interval $[0, \infty)$ and if

$$r(s, t) = \sigma^2 \text{ Min } (s, t), \qquad 0 \leq s, t < \infty,$$

the process is called the Brownian movement process, or sometimes the Wiener process; σ is a positive constant. When considering this process we shall write $\zeta(t)$ instead of $x(t)$. For the $\zeta(t)$ process

$$Pr\{\zeta(0) = 0\} = 1,$$

$$E\{[\zeta(t) - \zeta(s)]^2\} = \sigma^2 \mid t - s \mid,$$

and if $0 \leq s_1 < t_1 < s_2 < t_2$ the increments $x(t_1) - x(s_1)$ and $x(t_2) - x(s_2)$ are mutually independent. We shall use the following properties of this process, of which the first two are well known.

(a) The sample functions are everywhere continuous with probability 1. In the following we can therefore write as if all the sample curves were continuous.

(b) For fixed s

$$(4.1) \qquad Pr\{\underset{0<t\leq T}{\text{Max}} [\zeta(s + t) - \zeta(s)] \geq \lambda\} = 2Pr\{\zeta(s + T) - \zeta(s) \geq \lambda\}.^7$$

(Note that the use of a general initial value s, rather than 0, has not added to the generality and we drop this affectation below.)

(c) If $a \geq 0$, $b > 0$, $\alpha \geq 0$, $\beta > 0$, then

$$(4.2) \qquad Pr\{\underset{0\leq t<\infty}{\text{L.U.B.}} [\zeta(t) - (at + b)] \geq 0\} = e^{-2ab/\sigma^2},$$

[7] Due to Bachelier; cf. the proof by P. Lévy, *Comp. Math.*, Vol. 7 (1939), p. 293. One way to prove (*a*) is to prove (4.1) first, with L.U.B. instead of Max, and then use it to calculate the probabilities relevant to (*a*).

$$(4.3)\quad Pr\{\underset{0\leq t<\infty}{\text{L.U.B.}}\,[\zeta(t) - (at + b)] \geq 0 \quad \textit{or} \quad \text{G.L.B.}\,[\zeta(t) + \alpha t + \beta] \leq 0\},$$

$$= \sum_{m=1}^{\infty} \{e^{-2[m^2ab+(m-1)^2\alpha\beta+m(m-1)(a\beta+\alpha b)]}$$

$$+e^{-2[(m-1)^2ab + m^2\alpha\beta + m(m-1)(a\beta + \alpha b)]}$$

$$-e^{-2[m^2(ab + \alpha\beta) + m(m-1)a\beta + m(m+1)\alpha b]}$$

$$-e^{-2[m^2(ab + \alpha\beta) + m(m+1)a\beta + m(m-1)\alpha b]}\};$$

in particular ($\alpha = a$, $\beta = b$)

$$(4.3')\qquad Pr\left\{\underset{0\leq t<\infty}{\text{L.U.B.}}\,\frac{|\zeta(t)|}{at+b} \geq 1\right\} = 2\sum_{1}^{\infty}(-1)^{m+1}e^{-2m^2ab}.$$

The probability in (4.2) is the probability that a $\zeta(t)$ sample curve will ever reach the line with slope a and ordinate intercept b; the probability in (4.3) is the probability that a sample curve will ever reach either of the indicated halflines, one above and one below the t axis. Since the right hand sides are continuous functions of a, b, α, β we could write >0 instead of ≥ 0 and <0 instead of ≤ 0 on the left, so that these probabilities are also the probabilities that a sample curve will ever rise above the indicated line or leave the indicated angle.

It will be convenient to describe a line by its slope and ordinate intercept; the line $[u, v]$ is the line with slope u and ordinate intercept v. We shall take $\sigma = 1$ in the proof; this is no essential restriction since $\zeta(t)/\sigma$ is the random variable of a process of the same type whose σ is 1.

To prove (4.2) let $\varphi(a, b)$ be the probability on the left, the probability that a sample curve will reach the line $[a, b]$. If $b = b_1 + b_2$, $b_i > 0$, a sample curve which is to reach $[a, b]$ must first reach $[a, b_1]$ and then move up to meet a line with slope a, b_2 units above the first meeting with $[a, b_1]$. Then

$$\varphi(a, b_1 + b_2) = \varphi(a, b_1)\,\varphi(a, b_2).$$

Now $\varphi(a, b) \geq Pr\{\zeta(1) \geq a + b\} > 0$ and $\varphi(a, b)$ is monotone non-increasing in b, for fixed a. The only solution of the functional equation with these properties is

$$\varphi(a, b) = e^{-\psi(a)b}.$$

Now $\varphi(a, b)$ is the probability of reaching $[0, b]$ at some first time s and then going on to the line $[a, b]$ which from the vantage point of the first common point $(s, \zeta(s))$ is the line $[a, as]$. In other words, using (4.1)

$$e^{-\psi(a)b} = -\int_0^\infty e^{-\psi(a)as}\, d_s\, Pr\{\underset{0\leq t\leq s}{\text{Max}}\, \zeta(t) \geq b\}$$

$$= \int_0^\infty e^{-\psi(a)as}\, \frac{be^{-(b^2)/2s}}{(2\pi)^{1/2}s^{3/2}}\, ds$$

$$= \frac{2}{\pi^2}\int_0^\infty \exp\left[-s^2 - \frac{b^2 a\psi(a)}{2s^2}\right] ds$$
$$= e^{-b(2a\psi(a))^{1/2}},$$

from which it follows that $\psi(a) = 2a$, and this yields (4.2).

To prove (4.3) we consider first the following general problem: Let $[u_1, v_1]$, $[u_2, v_2]$, $\cdots$, $u_j \geq 0, v_j \geq 0$ be a sequence of lines; let $t = t_1$ be the first value of t, if any, at which a sample curve meets $[u_1, v_1]$; if t_1 is defined for a sample curve let t_2 be the first value of $t > t_1$, if any at which the curve meets $[-u_2, -v_2]$; if t_2 is defined for a sample curve, let t_3 be the first value of $t > t_2$, if any, at which the curve meets $[u_3, v_3]$, and so on. Let π_n be the probability that there is a point t_n, in other words the probability that a sample curve meets the lines $[u_1, v_1]$, $[-u_2, -v_2]$, $\cdots$ $[(-1)^{n+1}u_n, (-1)^{n+1}v_n]$ in at least n successive points. We write

$$\pi_n = \pi_n(u_1, v_1, \cdots, u_n, v_n).$$

In particular, according to (4.2)

$$\pi_1(u_1, v_1) = e^{-2u_1v_1}. \tag{4.4}$$

To evaluate π_n, let Q be the point $(t_{n-1}, \zeta(t_{n-1}))$ on the sample curve, and suppose for definiteness that n is even. Starting at Q, if there is a t_n, the curve must finally reach $[-u_n, -v_n]$, that is it must go to a line of slope $-u_n$, which is $u_{n-1}t_{n-1} + v_{n-1} + u_nt_{n-1} + v_n$ units vertically below its initial position Q when $t = t_{n-1}$. According to (4.2) the probability of doing this is

$$e^{-2u_n(u_{n-1}t_{n-1}+v_{n-1}+u_nt_{n-1}+v_n)}.$$

Now we replace the line $[-u_n, -v_n]$ by a line which depends on t_{n-1} but which leaves this probability unchanged; the new line has slope $-(u_{n-1} + u_n)$ and is

$$h = \frac{u_n}{u_{n-1} + u_n}(u_{n-1}t_{n-1} + v_{n-1} + u_nt_{n-1} + v_n)$$

units below Q when $t = t_{n-1}$. Finally we reflect this new line in the line parallel to the t axis through Q. These two changes do not affect the probability we are discussing because the changes of $\zeta(t)$ after t_{n-1} are independent of the changes before and have symmetric distributions. The final line has slope $u_{n-1} + v_{n-1}$ and is h units above Q when $t = t_{n-1}$; it is the line

$$\left[u_{n-1} + u_n, \quad \frac{u_{n-1}v_{n-1} + u_nv_n + 2u_nv_{n-1}}{u_{n-1} + u_n}\right]$$

which does not depend on t_{n-1}. This line lies above $[u_{n-1}, v_{n-1}]$ in the first quadrant, so that if a sample curve reaches it the curve must also intersect $[u_{n-1}, v_{n-1}]$. We have thus proved that

$$\pi_n(u_1, v_1; \cdots; u_n, v_n) = \pi_{n-1}\left(u_1, v_1; \cdots; u_{n-2}, v_{n-2}; u_{n-1} + u_n, \frac{u_{n-1}v_{n-1} + u_nv_n + 2u_nv_{n-1}}{u_{n-1} + u_n}\right). \tag{4.5}$$

The fundamental identity (4.5) makes it possible to reduce the evaluation of π_n to π_1 in $n - 1$ steps; π_1 is evaluated in (4.4). Thus successive meetings with n lines have been reduced to a meeting with a single line. As a first example suppose

$$u_1 = \cdots = u_n = u, \qquad v_1 = \cdots = v_n = v.$$

Then we have

$$\begin{aligned}\pi_n(u, v; \cdots ; u, v) &= \pi_{n-1}(u, v; \cdots ; 2u, 2v) = \cdots \\ &= \pi_1(nu, nv),\end{aligned}$$

so that

$$\pi_n(u, v; \cdots ; u, v) = e^{-2n^2uv} \tag{4.6}$$

More generally suppose

$$\begin{aligned} u_1 = u_3 = \cdots = a, \qquad v_1 = v_3 = \cdots = b, \\ u_2 = u_4 = \cdots = \alpha, \qquad v_2 = v_4 = \cdots = \beta. \end{aligned}$$

Then we show that for suitably chosen $C_j^{(n)}$'s we have according as n is even or odd

$$\pi_n(a, b; \cdots ; \alpha, \beta) = \pi_1\left[\frac{n}{2}(a + \alpha), \frac{C_1^{(n)}ab + C_2^{(n)}\alpha\beta + C_3^{(n)}a\beta + C_4^{(n)}\alpha b}{\frac{n}{2}(a + \alpha)}\right];$$

$$\begin{aligned}&\pi_n(a, b; \cdots ; a, b) \\ &= \pi_1\left[\frac{n+1}{2}a + \frac{n-1}{2}\alpha, \frac{C_1^{(n)}ab + C_2^{(n)}\alpha\beta + C_3^{(n)}a\beta + C_4^{(n)}\alpha b}{\frac{n+1}{2}a + \frac{n+1\alpha}{2}}\right].\end{aligned} \tag{4.7}$$

For $n = 1$ this form is correct with

$$C_1^{(1)} = 1, \qquad C_2^{(1)} = C_3^{(1)} = C_4^{(1)} = 0.$$

If now n is even and if the equations are true for n,

$$\pi_{n+1}(a, b; \cdots; a, b) = \pi_2\left(a, b; \frac{n}{2}(\alpha + a), \frac{C_1^{(n)}\alpha\beta + C_2^{(n)}ab + C_3^{(n)}\alpha b + C_4^{(n)}a\beta}{\frac{n}{2}\alpha + a}\right)$$

$$= \pi_1\left(\frac{n+2}{2}a + \frac{n}{2}\alpha, \frac{ab + C_1^{(n)} + C_2^{(n)}ab + C_3^{(n)}\alpha b + C_4^{(n)}a\beta + n(\alpha + a)b}{\frac{n+2}{2}a + \frac{n}{2}\alpha}\right),$$

and comparing this with (4.7) we find that

$$C_1^{(n+1)} = C_2^{(n)} + n + 1,$$

$$\begin{aligned} C_2^{(n+1)} &= C_1^{(n)}, \\ C_3^{(n+1)} &= C_4^{(n)}, \\ C_4^{(n+1)} &= C_3^{(n)} + n, \end{aligned} \qquad (n \text{ even}).$$

If n is odd we find similarly that

$$\begin{aligned} C_1^{(n+1)} &= C_2^{(n)} + n, \\ C_2^{(n+1)} &= C_1^{(n)}, \\ C_3^{(n+1)} &= C_4^{(n)}, \\ C_4^{(n+1)} &= C_3^{(n)} + n + 1. \end{aligned}$$

The solution of these equations is

n even	*n odd*
$C_1^{(n)} = \frac{n^2}{4}$	$C_1^{(n)} = \frac{(n+1)^2}{4}$
$C_2^{(n)} = \frac{n^2}{4}$	$C_2^{(n)} = \frac{(n-1)^2}{4}$
$C_3^{(n)} = \frac{n(n-2)}{4}$	$C_3^{(n)} = \frac{n^2-1}{4}$
$C_4^{(n)} = \frac{n(n+2)}{4}$	$C_4^{(n)} = \frac{n^2-1}{4}$

Then

$$\text{(4.8)} \qquad \begin{aligned} \pi_n &= e^{-\frac{1}{2}[n^2ab + n^2\alpha\beta + n(n-2)a\beta + n(n+2)\alpha b]} && (n \text{ even}), \\ \pi_n &= e^{-\frac{1}{2}[(n+1)^2ab + (n-1)^2\alpha\beta + (n^2-1)a\beta + (n^2-1)\alpha b]} && (n \text{ odd}). \end{aligned}$$

We can now prove (4.3). In fact the left side is equal to

$$\pi_1(a, b) + \pi_1(\alpha, \beta) - \pi_2(a, b; \alpha, \beta) - \pi_2(\alpha, \beta; a, b) + \cdots,$$

which gives (4.3), on substituting (4.8). Only (4.3′), which follows from the simple (4.6), is used in the application to the Kolmogorov-Smirnov theorems.

5. Transformations of Gaussian processes to the Brownian movement process. The $\zeta(t)$ process studied in section 4 is so simple that it is important to be able to reduce others to it by elementary changes of variable. For example if the covariance function of a Gaussian process has the form

$$\text{(5.1)} \qquad r(s, t) = u(s)v(t), \qquad s < t,$$

for s, t in some interval, and if the ratio

$$\frac{u(t)}{v(t)} = a(t)$$

is continuous and monotone increasing, with inverse function $a_1(t)$. We define

$$\zeta(t) = \frac{u[a_1(t)]}{v[a_1(t)]}.$$

With this definition the ζ process is Gaussian and since if $s < t$

$$E\{\zeta(s)\zeta(t)\} = \frac{u[a_1(s)]v[a_1(t)]}{v[a_1(s)]v[a_1(t)]} = a[a_1(s)] = s = \text{Min}(s, t),$$

the ζ process is the Brownian movement process with $\sigma = 1$. This transformation from the x to the ζ process is effected by a combination of a change of variable in t and the application of a variable scaling factor. (Conversely, if such a transformation is applied to the Brownian movement process it is trivial to verify that the new covariance function will have the form (5.1). The Gaussian processes with covariance functions of this form are easily seen to be the Gaussian Markov processes.)

6. The Gaussian process with r(s, t) = s(1 − t). In section 3 the Kolmogorov-Smirnov theorems were reduced to properties of a Gaussian process with parameter t, $0 \leq t \leq 1$, for which

$$Pr\{x(0) = 0\} = 1;$$

$$E\{x(t)\} = 0;$$

$$E\{[x(t) - x(s)]^2\} = (t - s)[1 - (t - s)], \quad 0 \leq s < t \leq 1.$$

Now these equations imply that

$$E\{x(t)^2\} = t(1 - t), \qquad E\{x(s)^2\} = s(1 - s),$$

and combining the set we find that

$$r(s, t) = E\{x(s)x(t)\} = s(1 - t), \qquad 0 \leq s < t \leq 1.$$

This covariance function has the form studied in section 5, and using the transformation of that section

$$\zeta(t) = (t + 1)x\left(\frac{t}{t + 1}\right), \qquad 0 \leq t < \infty,$$

defines a Brownian movement process (with $\sigma = 1$). Then if D, D^+, D^- are defined as in section 3, we have from (4.3′)

$$Pr\{D \geq \lambda\} = Pr\left\{\underset{0 \leq t < \infty}{\text{L. U. B.}} \left| \frac{\zeta(t)}{t + 1} \right| \geq \lambda\right\} = \sum_1^\infty (-1)^{m+1} e^{-2m^2\lambda^2},$$

and from (4.2)

$$Pr\{D^+ \geq \lambda\} = Pr\{D^- \geq \lambda\} = e^{-2\lambda^2}.$$

This proves (3.2′) and (3.3′). Note that we could go beyond these results, because of our detailed knowledge of the $x(t)$ process. For example we can evaluate

$$\lim_{n\to\infty} Pr\{(n)^{\frac{1}{2}}D_n^- \leq \lambda_1, \qquad (n)^{\frac{1}{2}}D_n^+ \leq \lambda_2\}.$$

If $\lambda_1 = \lambda_2 = \lambda$ the probability is the probability that $(n)^{1/2}D_n \leq \lambda$ which we have already treated. In general it is, in the limit,

$$\begin{aligned} Pr\{&\operatorname*{Min}_{0\leq t\leq 1} x(t) \geq -\lambda_1, \operatorname*{Max}_{0\leq t\leq 1} x(t) \leq \lambda_2\} \\ &= Pr\left\{\operatorname*{G.L.B.}_{0\leq t\leq\infty} \frac{\zeta(t)}{t+1} \geq -\lambda_1 \operatorname*{L.U.B.}_{0\leq t<\infty} \frac{\zeta(t)}{t+1} \leq \lambda_2\right\} \\ &= 1 - \sum_{m=1}^{\infty} \{e^{-2[m^2\lambda_2^2+(m-1)^2\lambda_1^2+2m(m-1)\lambda_1\lambda_2]} + e^{-2[(m-1)^2\lambda_2^2+m^2\lambda_1^2+2m(m-1)\lambda_1\lambda_2]} \\ &\qquad - e^{-2[m^2(\lambda_1^2+\lambda_2^2)+m(m-1)\lambda_1\lambda_2+m(m+1)\lambda_1\lambda_2]} - e^{-2[m^2(\lambda_1^2+\lambda_2^2)+m(m+1)\lambda_1\lambda_2+m(m-1)\lambda_1\lambda_2]}\} \\ &= 1 - \sum_{m=1}^{\infty} \{e^{-2[m\lambda_2+(m-1)\lambda_1]^2} + e^{-2[(m-1)\lambda_2+m\lambda_1]^2} - 2e^{-2m^2(\lambda_1+\lambda_2)^2}\} \end{aligned}$$

obtained by setting $a = b = \lambda_2$, $\alpha = \beta = \lambda_1$ in (4.3).