THE FIRST PASSAGE PROBLEM FOR A CONTINUOUS MARKOV PROCESS[1]

By D. A. Darling and A. J. F. Siegert

Columbia University, University of Michigan, and Northwestern University

Summary. We give in this paper the solution to the first passage problem for a strongly continuous temporally homogeneous Markov process $X(t)$. If $T = T_{ab}(x)$ is a random variable giving the time of first passage of $X(t)$ from the region $a > X(t) > b$ when $a > X(0) = x > b$, we develop simple methods of getting the distribution of T (at least in terms of a Laplace transform). From the distribution of T the distribution of the maximum of $X(t)$ and the range of $X(t)$ are deduced. These results yield, in an asymptotic form, solutions to certain statistical problems in sequential analysis, nonparametric theory of "goodness of fit," optional stopping, etc. which we treat as an illustration of the theory.

1. Introduction. There are certain generalizations of the classical gambler's ruin problem which appear in various guises in numerous applications—besides statistical problems there are physical applications in the theory of noise, in genetics, etc. The exact solution of the associated random walk (or Markov chain) problem is often analytically difficult, if not impossible to obtain, and one is usually content with asymptotic solutions. The nature of the asymptotic solution is generally such that it is the solution to a Markov chain problem in which the length of the steps, and the interval between them, approach zero and which may in the limit be regarded as some sort of continuous stochastic process.

This circumstance suggests we might solve directly the associated problem with regard to the stochastic process and so obtain the asymptotic solution to the Markov chain problem without the intervention of a limiting process. Aside from the difficulty of justifying the interchange of these limiting operations, it turns out that this procedure is often quite feasible and leads to simple solutions. Using this idea Doob [7] obtained in a direct way the Kolmogorov-Smirnov limit theorems and the principle was further exploited by Anderson and Darling [1]. The general principle is, of course, quite old, and in connection with random walk problems goes back at least to Rayleigh.

A general feature of this method is that the analytical difficulties, if any, are revealed as more or less classical boundary value problems, eigenvalue problems, etc.—this intrinsic nature of the problem often being masked by the discrete approach. On the other hand, it suffers from the serious defect of giving no

Received 2/3/52, revised 3/3/53.

[1] The major part of this work was done while the authors were consultants with The RAND Corporation.

information as to the difference between the actual solution and the asymptotic one—information which is essential in the numerical applications.

In the present paper we treat the first passage (or ruin, or absorption probability) problem for a general class of Markov processes (cf. Section 2) and obtain the solution in the form of a Laplace transform (Section 3). This Laplace transform is generally given as a simple function of the solutions to an ordinary differential equation (Section 4). The methods used are similar to those used in the discrete theory by Wald [17] (fundamental identity) and Feller [9] (renewal and generating function techniques), but the analysis is considerably simplified, at least in a formal way, and not restricted to additive processes. It turns out that there is an intimate relationship between the one- and two-sided absorption probabilities, and the probability of eventual absorption in one of the boundaries.

We illustrate the theory in Section 5 by solving a problem of Wald [17] in the sequential test of the mean of a normal population against a single alternative, the derivation of a nonparametric test used by Anderson and Darling [1] and the solution to the optional stopping problem (Robbins [15]). These problems are treated by solving the associated absorption problem with the Wiener-Einstein process and the Uhlenbeck process.

In Section 6 we study the first passage moments which can be obtained by an expansion of the Laplace transforms or again through differential equations which can be explicitly solved in quadratures. There are some quite interesting relations between the moments.

In Section 7 we develop the distribution of the range which has been used by Feller [10] in a statistical study.

2. Definitions, notations, assumptions, etc. Given a stochastic process $X(t)$ with $X(0) = x$, $a > x > b$, we define the *first passage time* $T_{ab}(x)$ as the random variable

$$T = T_{ab}(x) = \sup \{t \mid a > X(\tau) > b, 0 \leqq \tau \leqq t\}.$$

We make the following assumptions about the stochastic process $X(t)$.

A) $X(t)$ has a transition probability

$$P(x \mid y, t) = \Pr\{X(t + s) < y \mid X(s) = x\}, \qquad s > 0,$$

satisfying the Chapman-Kolmogorov equation

$$P(x \mid y, t_1 + t_2) = \int_{-\infty}^{\infty} P(z \mid y, t_2)\, d_z P(x \mid z, t_1), \quad t_1 > 0, t_2 > 0;$$

that is, $X(t)$ is temporally homogeneous and stochastically definite (e.g. Markovian).

B) $X(t)$ is continuous with probability one (or is *strongly continuous*).

If $X(t)$ satisfies A) sufficient conditions on P are known that it satisfy B), cf. Doeblin [5], Fortet [11], Ito [12]. These conditions generally imply further that

P satisfies the diffusion equation of Section 4. Note that A) and B) imply the *existence* of the random variable T, and we denote by $F_{ab}(x \mid t)$ the distribution function of T, $F_{ab}(x \mid t) = Pr\{T_{ab}(x) < t\}$.

In the work to follow we shall presume P and F have derivatives p, f; these being the densities

$$p(x \mid y, t) = \frac{\partial}{\partial y} P(x \mid y, t)$$

$$f_{ab}(x \mid t) = \frac{\partial}{\partial t} F_{ab}(x \mid t),$$

the modification of the results if these conditions are not met being more or less immediate. The existence of a density for T has been established by Fortet [11] under some circumstances. In this fundamental paper of Fortet on absorption probabilities there is just one absorbing barrier, but the modification of his results for two barriers is easy.

If $a = +\infty$ or $b = -\infty$ so that we have a one-sided absorption time we write $T_c(x)$ as the corresponding random variable. That is

$$T_c(x) = \begin{cases} T_{\infty,c}(x) \text{ if } x > c \\ T_{c,-\infty}(x) \text{ if } x < c \end{cases}$$

with a corresponding distribution function $F_c(x \mid t)$ and density $f_c(x \mid t)$.

It may happen of course that absorption is not a certain event and that T is not a proper random variable, that is $Pr\{T_c(x) < \infty\} = F_c(x \mid \infty) < 1$ (or similarly for $T_{ab}(x)$) and in this case we may still meaningfully treat the conditional density of T, under the condition $T < \infty$.

We need, in addition, the conditional distribution of $T_{ab}(x)$ under the condition that the absorption takes place into the barrier $\underline{a}$, which we denote by $F_{ab}^{+}(x \mid t)$:

$$F_{ab}^{+}(x \mid t) = \Pr\{T_{ab}(x) < t, T_{ab}(x) = T_a(x)\}$$

and $F_{ab}^{-}(x \mid t)$ will denote a similar expression for the lower barrier $\underline{b}$. Hence

$$F_{ab}(x \mid t) = F_{ab}^{+}(x \mid t) + F_{ab}^{-}(x \mid t)$$

and the corresponding densities are $f_{ab}^{+}(x \mid t)$ and $f_{ab}^{-}(x \mid t)$.

We denote by a circumflex over the corresponding function its Laplace transform on t; for example

$$\hat{p}(x \mid y, \lambda) = \int_0^\infty e^{-\lambda t} p(x \mid y, t)\, dt,$$

$$\hat{f}_{ab}^{+}(x \mid \lambda) = \int_0^\infty e^{-\lambda t} f_{ab}^{+}(x \mid t)\, dt,$$

etc. The continuity of the process $X(t)$ ensures the existence of these transforms.

3. The distribution of T. In this section we obtain the distribution of T in terms of the transition density p of the process. Theorem 3.1 for the one-sided barrier is due to Siegert [16] essentially.

THEOREM 3.1. *If $X(t)$ satisfies conditions* A) *and* B), *then $\hat{p}(x \mid y, \lambda)$ is a product*

$$\hat{p}(x \mid y, \lambda) = \begin{cases} u(x)u_1(y), & y > x \\ v(x)v_1(y), & y < x \end{cases}$$

and

$$\hat{f}_c(x \mid \lambda) = \begin{cases} \dfrac{u(x)}{u(c)}, & x < c \\ \dfrac{v(x)}{v(c)}, & x > c. \end{cases} \tag{3.1}$$

We note that absorption may be uncertain and $\hat{f}_c(x \mid 0) = Pr\{T_c(x) < \infty\}$ may be less than 1. A necessary and sufficient condition that absorption be certain is that $\hat{f}_c(x \mid 0) = 1$.

To prove the theorem we use a renewal principle which is very old. We have by A) and B) for $y > c > x$

$$p(x \mid y, t) = \int_0^t f_c(x \mid \tau) p(c \mid y, t - \tau)\, d\tau$$

by a direct enumeration of the paths going from x to y. On taking Laplace transforms we obtain

$$\hat{p}(x \mid y, \lambda) = \hat{f}_c(x \mid \lambda)\hat{p}(c \mid y, \lambda), \qquad y > c > x$$

and thus $\hat{p}(x \mid y, \lambda)$ is a function of x times a function of y, say $u(x)u_1(y)$ and hence for $y > c > x$ we get $\hat{f}_c(x \mid \lambda) = u(x)/u(c)$. Similarly, for $y < c < x$ we obtain $\hat{f}_c(x \mid \lambda) = v(x)/v(c)$ and hence for any c, x we obtain the conclusions to the theorem. Finally it follows by cancelling any factor which depends only on λ that $u(x)$ and $v(x)$ are uniquely determined.

THEOREM 3.2. *Let $X(t)$ satisfy* A) *and* B) *and let the functions $u(x)$ and $v(x)$ be as in Theorem* 3.1. *Then*

$$\hat{f}^{+}_{ab}(x \mid \lambda) = \frac{v(b)u(x) - u(b)v(x)}{u(a)v(b) - u(b)v(a)} \tag{3.2}$$

$$\hat{f}^{-}_{ab}(x \mid \lambda) = \frac{u(a)v(x) - v(a)u(x)}{u(a)v(b) - u(b)v(a)} \tag{3.3}$$

$$\hat{f}_{ab}(x \mid \lambda) = \frac{v(x)(u(a) - u(b)) - u(x)(v(a) - v(b))}{u(a)v(b) - u(b)v(a)}. \tag{3.4}$$

To prove the theorem we consider the two expressions

$$f_a(x \mid t) = f_{ab}^+(x \mid t) + \int_0^t f_{ab}^-(x \mid \tau) f_a(b \mid t - \tau)\, d\tau$$

$$f_b(x \mid t) = f_{ab}^-(x \mid t) + \int_0^t f_{ab}^+(x \mid \tau) f_b(a \mid t - \tau)\, d\tau$$

which are established by a direct enumeration. Considering f^+ and f^- as unknown this pair of simultaneous integral equations is solved immediately by taking Laplace transforms

$$\hat{f}_a(x \mid \lambda) = \hat{f}_{ab}^+(x \mid \lambda) + \hat{f}_{ab}^-(x \mid \lambda)\hat{f}_a(b \mid \lambda) \tag{3.5}$$

$$\hat{f}_b(x \mid \lambda) = \hat{f}_{ab}^-(x \mid \lambda) + \hat{f}_{ab}^+(x \mid \lambda)\hat{f}_b(a \mid \lambda) \tag{3.6}$$

which are 2 linear equations in 2 unknowns. On using the expressions in Theorem 3.1 for $\hat{f}_a$ and $\hat{f}_b$ we get (3.2) and (3.3) for $\hat{f}_{ab}^+$ and $\hat{f}_{ab}^-$ and the last expression (3.4) is obtained by noting $\hat{f}_{ab} = \hat{f}_{ab}^+ + \hat{f}_{ab}^-$.

A random variable closely related to T is the maximum of $X(t)$, and we define

$$M(x,t) = \sup_{0 \leq \tau \leq t} \mid X(\tau) \mid, \qquad X(0) = x. \tag{3.7}$$

Denoting the distribution of M by $G(x \mid \xi, t)$ we have clearly

$$\begin{aligned} G(x \mid \xi, t) &= Pr\{M(x, t) < \xi\} = \Pr\{T_{\xi,-\xi}(x) > t\} \\ &= 1 - F_{\xi,-\xi}(x \mid t), \qquad \xi > \mid x \mid, \end{aligned} \tag{3.8}$$

so that the distribution of M is given directly through that of T. On taking Laplace transforms of (3.8) we obtain the following corollary

Corollary 3.3. $\hat{G}(x \mid \xi, \lambda) = 1/\lambda(1 - \hat{f}_{\xi,-\xi}(x \mid \lambda))$ *for* $\hat{f}_{\xi,-\xi}(x \mid \lambda)$ *as in Theorem* 3.2.

For a symmetrical process there is a specially simple formula for the Laplace transform of $T_{a,-a}(x)$. A process $X(t)$ is symmetrical if $p(x \mid y, t) = p(-x \mid -y, t)$ for all x, y, t. In this case $u(x) = v(-x)$ and Theorem 3.2 yields the following corollary.

Corollary 3.4. *For a symmetrical process*

$$\hat{f}_{a,-a}(x \mid \lambda) = \frac{u(x) + u(-x)}{u(a) + u(-a)}, \qquad \mid x \mid < a. \tag{3.9}$$

4. A differential equation. The function $p(x \mid y, t)$ will in most cases of interest satisfy the so-called diffusion equation

$$\frac{\partial p}{\partial t} = A(x)\frac{\partial p}{\partial x} + \tfrac{1}{2}B^2(x)\frac{\partial^2 p}{\partial x^2} \tag{4.1}$$

with initial and boundary conditions $p(\infty \mid y, t) = p(-\infty \mid y, t) = 0, p(x \mid y, 0) = \delta(x - y)$ (the Dirac function). Sufficient conditions on p, involving the infinitesimal transition moments, are known in order that p satisfy an equation of the type (4.1), and which generally further ensure the process is continuous with probability one (cf. Doeblin [5]). When A and B^2 are given a priori, conditions on them are known which ensure that (4.1) has a unique solution which is the transition density of a process continuous with probability 1. (Cf. Fortet [11]). But general necessary and sufficient conditions are not known, and it does not appear to be known under what conditions a process continuous with probability one satisfies a diffusion equation. However, for specific processes these points are generally easy to resolve.

The following theorem shows that for processes satisfying (4.1) $\underline{u}$ and $\underline{v}$ can be determined from a differential equation.

THEOREM 4.1. *If $p(x \mid y, t)$ uniquely satisfies (4.1) with the stated boundary conditions and $X(t)$ is continuous with probability one, the functions $u(x)$ and $v(x)$ can be chosen as any two linearly independent solutions of the differential equation*

$$\tfrac{1}{2}B^2(x) \frac{d^2w}{dx^2} + A(x) \frac{dw}{dx} - \lambda w = 0. \tag{4.2}$$

To prove the theorem we note that if p satisfies (4.1) its Laplace transform satisfies

$$\lambda \hat{p} = A \frac{d\hat{p}}{dx} + \tfrac{1}{2}B^2 \frac{d^2\hat{p}}{dx^2} \tag{4.3}$$

and indeed $-\hat{p}$ is the Green's solution to this equation over the infinite interval $(-\infty < x < \infty)$. As a consequence, if $u(\infty) = v(-\infty) = 0$ and $u(x)$, $v(x)$ satisfy (4.3) we obtain to a constant factor

$$\hat{p}(x \mid y, \lambda) = \begin{cases} v(x)u(y) & y \geqq x \\ v(y)u(x) & y \leqq x \end{cases}$$

so that we obtain the previous expression (3.1) for $\hat{f}_c(x \mid \lambda)$ and consequently we obtain (3.2), (3.3) and (3.4). But since (3.2), (3.3) and (3.4) are invariant under any nonsingular linear transformation of $\underline{u}$ and $\underline{v}$ we obtain Theorem 4.1.

As for (3.9) we can choose for $u(x)$ any solution to (4.2) provided $u(x)$ and $u(-x)$ are linearly independent.

The customary way to obtain the first passage probability $f_{ab}(x \mid t)$ is to solve (4.1) with the boundary conditions $p(a \mid y, t) = p(b \mid y, t) = 0$, $p(x \mid y, 0) = \delta(x - y)$ and then we should have $F_{ab}(x \mid t) = 1 - \int_b^a p(x \mid y, t)\, dy$ (cf. Fortet [11] for a proof and Lévy [14] for a general discussion). By using the Laplace transform method this will give (3.4) for $\hat{f}_{ab}(x \mid \lambda)$, but it does not appear to give $\hat{f}^+$ and $\hat{f}^-$.

Since $\hat{f}^+_{ab}(x \mid 0)$ is the probability that absorption in the barrier $\underline{a}$ occurs before absorption in $\underline{b}$, we should expect that, putting $\lambda = 0$ in (4.2), the solution to

$$\tfrac{1}{2}B^2 \frac{d^2\phi}{dx^2} + A \frac{d\phi}{dx} = 0$$

with $\phi(a) = 1$, $\phi(b) = 0$ should give this probability. Khintchine [13] has proved this result directly from the limiting case of a Markov chain without the use of a stochastic process. Barnard [3] has considered this result in connection with a sequential analysis problem.

5. A few examples.

a) *The Wiener-Einstein process.* Here $X(t)$ is the free Brownian motion; $X(t)$ is Gaussian with mean 0 and covariance $E(X(s)X(t)) = \min(s, t)$ and its transition density p satisfies the differential equation $\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2 p}{\partial x^2}$ (i.e. $A = 0$, $B^2 = 1$). Two linearly independent solutions to $\frac{1}{2}W'' - \lambda W = 0$ are $u(x) = e^{-\sqrt{2\lambda}x}$ and $v(x) = e^{\sqrt{2\lambda}x} = u(-x)$ and hence we obtain from (3.9)

$$\hat{f}_{a,-a}(x \mid \lambda) = \frac{\cosh \sqrt{2\lambda}\, x}{\cosh \sqrt{2\lambda}\, a}, \qquad |x| < a. \tag{5.1}$$

The inversion of this Laplace transform is easy, and we obtain

$$f_{a,-a}(x \mid t) = \frac{\pi}{a^2}\sum_{j=0}^{\infty}(-1)^j(j+\tfrac{1}{2})\cos\left\{(j+\tfrac{1}{2})\frac{\pi x}{a}\right\}e^{-(j+\frac{1}{2})^2\pi^2 t/2a^2}$$

and by integration on t

$$\begin{aligned} F_{a,-a}(x \mid t) &= Pr\{T_{a,-a}(x) < t\} \\ &= 1 - \frac{2}{\pi}\sum_{j=0}^{\infty}\frac{(-1)^j}{j+\frac{1}{2}}\cos\left\{(j+\tfrac{1}{2})\frac{\pi x}{a}\right\}e^{-(j+\frac{1}{2})^2\pi^2 t/2a^2}. \end{aligned}$$

This completely solves the case of Brownian motion for general barriers, since

$$F_{a,b}(x \mid t) = F_{(a-b)/2,-(a-b)/2}\left(x - \frac{a+b}{2} \,\middle|\, t\right). \tag{5.2}$$

This result is well known (Bachelier [2], Lévy [14]) and alternatively can be obtained by the method of images with the heat equation $\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2 p}{\partial x^2}$.

b) *The Uhlenbeck process.* Here $X(t)$ is stationary, Markovian, and Gaussian, with mean 0 and covariance $E(X(s)X(t)) = e^{-|s-t|}$ and the transition density satisfies (4.1) with $B^2 = 2$, $A = -x$. Solutions to

$$\frac{d^2w}{dx^2} - x\frac{dw}{dx} - \lambda w = 0$$

are $u(x) = e^{x^2/4}D_{-\lambda}(x)$ and $v(x) = e^{x^2/4}D_{-\lambda}(-x)$ where $D_\nu(z)$ is the Weber function, (cf. Whittaker and Watson [18]). Hence (3.9) gives

$$\hat{f}_{a,-a}(x \mid \lambda) = \exp\left\{\frac{x^2}{4} - \frac{a^2}{4}\right\}\frac{D_{-\lambda}(x) + D_{-\lambda}(-x)}{D_{-\lambda}(a) + D_{-\lambda}(-a)}, \tag{5.3}$$

but it appears very difficult to invert this transform. For the particular case $x = 0$ this result (5.3) was obtained from a limiting case of an Ehrenfest urn scheme describing molecular equilibrium by Bellman and Harris [4].

c) *A problem of Wald in sequential analysis.* Let $X_1, X_2, \cdots$, be independent random variables, normally distributed with an unknown mean θ and a known variance K^2. That is, the density of X_i is

$$\phi(x, \theta) = \frac{1}{\sqrt{2\pi}\,K} e^{-(x-\theta)^2/2K^2}.$$

According to the sequential likelihood ratio test of Wald, in order to test the hypothesis H_2 that $\theta = \theta_2$ against the hypothesis $H_1 : \theta = \theta_1$ we consider random variables

$$Z_i = \log \frac{\phi(X_i, \theta_1)}{\phi(X_i, \theta_2)} = \frac{\theta_1 - \theta_2}{K^2}\left(X_i - \frac{\theta_1 + \theta_2}{2}\right)$$

and let $S_j = Z_1 + Z_2 + \cdots + Z_j$. Then for $a > 0 > b$ we study the random variable N defined as the smallest integer for which either $S_N > a$ or $S_N < b$ and determine for this N the probabilities of these outcomes.

Now

$$E(Z_i) = \frac{\theta_1 - \theta_2}{K^2}\left(\theta - \frac{\theta_1 + \theta_2}{2}\right) = \mu, \tag{5.4}$$

$$\operatorname{Var}(Z_i) = \left(\frac{\theta_1 - \theta_2}{K}\right)^2 = \sigma^2; \tag{5.5}$$

so that this suggests we study a Gaussian process $S(t)$ with independent increments and with $E(S(t)) = \mu t$ and $\operatorname{Var}(S(t)) = \sigma^2 t$ (a linear transformation of the Wiener process). Then the joint distribution of $S_1, S_2, \cdots, S_j$ is the same as the joint distribution of $S(1), S(2), \cdots, S(j)$, and in place of finding the distribution of N we approximate to it by finding the distribution of the absorption time $T_{a,b}(0)$ in connection with the process $S(t)$. It should be remarked that the nature of this approximation is quite different from Wald's approximation of "neglecting the excess" since the process $S(t)$ may leave and re-enter one of the barriers between two consecutive integer time instants.

The differential equation satisfied by the transition density p of the process $S(t)$ is

$$\frac{\partial p}{\partial t} = \mu \frac{\partial p}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 p}{\partial x^2};$$

that is, $A = \mu$, $B^2 = \sigma^2$, and (4.2) becomes

$$\frac{\sigma^2}{2} \frac{d^2 w}{dx^2} + \mu \frac{dw}{dx} - \lambda w = 0. \tag{5.6}$$

It is simple to solve this equation with constant coefficients and since the two roots of $\frac{\sigma^2}{2}\xi^2 + \mu\xi - \gamma = 0$ are

$$\xi_1 = \frac{-\mu + \sqrt{\mu^2 + 2\sigma^2\lambda}}{\sigma^2}, \qquad \xi_2 = \frac{-\mu - \sqrt{\mu^2 + 2\sigma^2\lambda}}{\sigma^2},$$

two linearly independent solutions to (5.6) are $u(x) = e^{\xi_1 x}$ and $v(x) = e^{\xi_2 x}$ and hence by Theorem 3.2 we immediately obtain $\hat{f}^+$, $\hat{f}^-$, and $\hat{f}$ and the problem is formally solved. The expressions are to be considered for $x = 0$, and (3.2) gives for $x = 0$, with this $u(x)$, $v(x)$,

$$\hat{f}^+(0 \mid \lambda) = \frac{e^{\xi_1 b} - e^{\xi_2 b}}{e^{\xi_2 a + \xi_1 b} - e^{\xi_1 a + \xi_2 b}}$$

and at $\lambda = 0$ this gives the probability of being absorbed into the barrier a before b, and we abbreviate $L^+(\theta) = \hat{f}^+_{ab}(0 \mid 0)$ for this probability. For $\lambda = 0$ we have $\xi_1 = 2\mu/\sigma^2$, $\xi_2 = 0$ so that

$$L^+(\theta) = \frac{e^{-(2\mu b/\sigma^2)} - 1}{e^{-(2\mu/\sigma^2)b} - e^{-(2\mu/\sigma^2)a}}. \tag{5.7}$$

According to the test of Wald we should choose the barriers a and b so that $L^+(\theta_1) = 1 - \beta$, $L^+(\theta_2) = \alpha$ where α, β, are given positive numbers with $\alpha + \beta < 1$. For $\theta = \theta_1$ we have $2\mu = \sigma^2$ and for $\theta = \theta_2$ we have $2\mu = -\sigma^2$ from (5.4) and (5.5). Hence from (5.7) we get as two equations for $\underline{a}$ and $\underline{b}$

$$1 - \beta = \frac{e^{-b} - 1}{e^{-b} - e^{-a}}, \qquad \alpha = \frac{e^b - 1}{e^b - e^a},$$

which are easily solved to give

$$a = \log \frac{1-\beta}{\alpha}, \qquad b = \log \frac{\beta}{1-\alpha}.$$

These are the formulas of Wald.

From (5.2) and (5.3) we see that $2\mu/\sigma^2 = (2\theta - (\theta_1 + \theta_2))/(\theta_1 - \theta_2)$ which denote by $h(\theta)$. Then setting $A = (1 - \beta)/\alpha$, $B = \beta/(1 - \alpha)$ we obtain from (5.7)

$$L^+(\theta) = \frac{B^{-h(\theta)} - 1}{B^{-h(\theta)} - A^{-h(\theta)}},$$

the probability of absorption in the barrier a, which is the power of the test (i.e., the probability of rejecting $H_2 : \theta = \theta_2$ when θ is the true mean) and $1 - L^+(\theta) = L^-(\theta)$ is the expression given by Wald for the operating characteristic of test.

For the distribution of T (approximate number of observations necessary to terminate the test) we use the expression (3.4) with $x = 0$ to give

$$E(e^{-\lambda T_{ab}(0)}) = \hat{f}_{ab}(0 \mid \lambda) = \frac{(e^{\xi_1 b} - e^{\xi_2 b}) - (e^{\xi_1 a} - e^{\xi_2 a})}{e^{\xi_2 a + \xi_1 b} - e^{\xi_2 b + \xi_1 a}}$$

which can be inverted to give a rather complicated expression:

$$F_{ab}(0 \mid t) = \Pr\{T_{ab}(0) < t\}$$

$$= 1 - \frac{\sigma^2 \pi^2}{(a-b)^2} \sum_{n=1}^{\infty} \frac{n(-1)^n}{\frac{\mu^2}{2\sigma^2} + \frac{\sigma^2 n^2 \pi^2}{2(a-b)^2}} \left\{ e^{\mu b/\sigma^2} \cdot \sin \frac{n\pi a}{a-b} - e^{\mu a/\sigma^2} \cdot \sin \frac{n\pi b}{a-b} \right\}$$

$$\cdot \exp\left(-t\left\{\frac{\mu^2}{2\sigma^2} + \frac{\sigma^2 n^2 \pi^2}{2(a-b)^2}\right\}\right).$$

But the moments are easy to obtain by expanding about $\lambda = 0$, since we have the moment generating function of T (note that T is a proper random variable, that is, absorption is a certain event since $\hat{f}_{ab}(0 \mid 0) = 1$). An alternative way is to use the result of the next section which gives the moments as the solutions to differential equations. If we let $m(x) = E(T_{ab}(x))$ then from (6.6) it follows that m satisfies the differential equation $\frac{1}{2}\sigma^2 m''(x) + \mu m'(x) = -1$ with $m(a) = m(b) = 0$.

Assuming first that $\mu \neq 0$ we obtain by solving this equation

$$m(0) = E(T) = \frac{1}{\mu}(aL^+(\theta) + bL^-(\theta))$$

while for $\mu = 0$

$$E(T) = -\frac{ab}{\sigma^2} = \log\left(\frac{1-\alpha}{\alpha}\right) \log\left(\frac{1-\beta}{\beta}\right) \frac{K^2}{(\theta_1 - \theta_2)^2}.$$

Here $L^-(\theta) = 1 - L^+(\theta)$ is the probability of absorption in the barrier b, as before, and a, b, μ, σ^2 have their former significance.

It is rather remarkable that despite the differing nature of the approximations of Wald and the approximations by presuming a continuous process as here, they should give the same formulas.

d) *A nonparametric test in "goodness of fit."* In a test related to the Kolmogorov-Smirnov tests the following important absorption probability problem arose. If $X(t)$ is the Uhlenbeck process (cf. example *b*) above) calculate the probability

$$b(\xi \mid t) = \Pr\{|X(\tau)| < \xi, 0 \leqq \tau \leqq t\}$$

where $X(0)$ has its stationary distribution. Thus we have the problem of finding the distribution of the random variable $M(x, t)$ defined by (3.7) whose distribution function is $G(x \mid \xi, t)$ as in (3.8).

For $|x| < \xi$ we have $G(x \mid \xi, t) = 1 - F_{\xi,-\xi}(x \mid t)$ and since for $|x| \geqq \xi$ we have $G = 1$ we define $F = 0$ for $|x| \geqq \xi$. The stationary distribution of $X(t)$ is $N(0, 1)$, that is, has a density $\phi(x) = (2\pi)^{-\frac{1}{2}} e^{-x^2/2}$ and hence

$$b(\xi \mid t) = \int_{-\infty}^{\infty} \phi(x) G(x \mid \xi, t)\, dx = \int_{-\infty}^{\infty} \phi(x) F_{\xi,-\xi}(x \mid t)\, dx.$$

On taking Laplace transforms we get

$$\hat{b}(\xi \mid \lambda) = \frac{1}{\lambda} - \int_{-\xi}^{\xi} \phi(x) \hat{F}_{\xi,-\xi}(x \mid \lambda)\, dx = \frac{1}{\lambda}\left\{1 - \int_{-\xi}^{\xi} \phi(x) \hat{f}_{\xi,-\xi}(x \mid \lambda)\, dx\right\}$$

and substituting from (5.3) we have

$$\hat{b}(\xi \mid \lambda) = \frac{1}{\lambda}\left\{1 - \sqrt{\frac{2}{\pi}}\, \frac{e^{-\xi^2/4}}{D_{-\lambda}(\xi) + D_{-\lambda}(-\xi)} \cdot \int_0^{\xi} e^{-x^2/4}(D_{-\lambda}(x) + D_{-\lambda}(-x))\, dx\right\}.$$

This result was given, without proof, in [1].

e) *The optional stopping problem.* In [15] Robbins outlined the optional stopping problem. Let, as in example c) above, the problem be that of testing the mean of a normal universe with known variance, say σ^2, but instead of testing the hypotheses H_1 and H_2 of example c) we have the hypothesis H_1 : $\theta = 0$ to test against $H_2 : \theta \neq 0$ (Robbins considers $H_2 : \theta > 0$). As sketched by Robbins the basic problem is to calculate the probability, when $\theta = 0$, $S_n = X_1 + X_2 +, \cdots, + X_n$,

$$g(n_1, n_2, \alpha) = \Pr\{|S_n| < \alpha\sigma\sqrt{n}, n_1 \leqq n \leqq n_2\}$$

for given α, n_1 and n_2. For the case of S_n instead of $|S_n|$ Robbins gave an inequality, and here we give an approximate and an asymptotic result.

The random variables $\{S_n/\sigma\sqrt{n}\}$, $n = n_1, n_1 + 1, \cdots, n_2$ have mean 0, variance 1, are normally distributed and form a Markov chain with covariance

$$E\left\{\frac{S_j}{\sigma\sqrt{j}} \cdot \frac{S_n}{\sigma\sqrt{n}}\right\} = \frac{\min(j, n)}{\sqrt{jn}} = e^{-|\frac{1}{2}\log j - \frac{1}{2}\log n|}.$$

Hence their joint distribution is the same as the joint distribution of $X(\frac{1}{2}\log n_1)$, $X(\frac{1}{2}\log(n_1 + 1))$, $\cdots$, $X(\frac{1}{2}\log n_2)$ where $X(t)$ is the Uhlenbeck process; (cf. examples b) and d) above). Hence we have, using approximations like those in example c),

$$g(n_1, n_2, \alpha) \cong \Pr\{|X(t)| < \alpha, \tfrac{1}{2}\log n_1 \leqq t \leqq \tfrac{1}{2}\log n_2\} = b\left(\alpha \mid \tfrac{1}{2}\log\frac{n_2}{n_1}\right),$$

where $b(\xi \mid t)$ is the function of example d) and of which we have the Laplace transform.

It is also possible to give an exact asymptotic result which is applicable even if the variables are not normally distributed, but merely have mean 0, variance σ^2, and obey the central limit theorem (e.g., if they are identically distributed). Let $n_1 \to \infty$, $n_2 \to \infty$, $n_1/n_2 \to t$, $0 < t < 1$, and consider a sequence $\{t_n\}$, $n = n_1, n_1 + 1, \cdots, n_2$ defined for fixed n_2 by $t_n = n/n_2$; this sequence depends upon n_2, $\{t_n\}_{n_2}$, and for $n_2 \to \infty$ becomes everywhere dense in the interval $t \leqq \tau \leqq 1$. That is, given any $\tau(t \leqq \tau \leqq 1)$ we can choose an element τ_k from $\{t_n\}_k$ such that $\lim_{k\to\infty}\tau_k = \tau$.

Then since

$$g(n_1, n_2, \alpha) = \Pr\left\{\frac{|S_{n_2 t_n}|}{\sigma\sqrt{n_2}} < \alpha\sqrt{t_n}, n_1 \leqq n \leqq n_2\right\}$$

it will follow from a theorem of Donsker [6] that the limiting distribution g can be expressed as the distribution of the corresponding Wiener functional. Hence for $n_1 \to \infty$, $n_2 \to \infty$, $n_1/n_2 \to t$, $0 < t < 1$,

$$g(n_1, n_2, \alpha) \to \Pr\{|W(\tau)| < \alpha\sqrt{\tau}, t \leqq \tau \leqq 1\},$$

where $W(t)$ is the Wiener-Einstein process (cf. example a) above).

Now if $X(t)$ is the Uhlenbeck process (cf. example b)) we can write $W(t) = \sqrt{t}X(\frac{1}{2}\log t)$ (Doob [8]) and thus

$$\begin{aligned}\lim g &= \Pr\{|X(\tfrac{1}{2}\log\tau)| < \alpha, t \leqq \tau \leqq 1\}\\ &= \Pr\left\{|X(\tau)| \leqq \alpha, 0 \leqq \tau \leqq \tfrac{1}{2}\log\frac{1}{t}\right\} = b\left(\alpha \,\middle|\, \tfrac{1}{2}\log\frac{1}{t}\right),\end{aligned}$$

and since $1/t = \lim n_2/n_1$ we obtain $g \sim b(\alpha \mid \frac{1}{2}\log n_2/n_1)$, the approximate expression deduced above. It seems somewhat striking that these two expressions should agree, being deduced from essentially distinct principles.

6. On the moments of *T*. In the preceding work the distributions were generally expressed as Laplace transforms which are often difficult to invert but which give immediate information about the moments of T.

In the present section we suppose that $\Pr\{T < \infty\} = 1$, that is, that T is a proper random variable, as otherwise the moments will not exist. If the corresponding Laplace transform is 1 for $\lambda = 0$ the variable is proper. Let us put

$$t_{ab}^{(n)}(x) = E(T_{ab}^n(x)), \qquad t_c^{(n)}(x) = E(T_c^n(x))$$

which we suppose to *exist* for $n \leqq n_0$. We have by a series expansion

$$(6.1)\qquad \begin{aligned}\hat{f}_{ab}(x \mid \lambda) &= \sum_{n=0}^{n_0} \frac{t_{ab}^{(n)}(x)}{n!}(-\lambda)^n + o(\lambda^{n_0}), &\lambda \to 0\\ \hat{f}_c(x \mid \lambda) &= \sum_{n=0}^{n_0} \frac{t_c^{(n)}(x)}{n!}(-\lambda)^n + o(\lambda^{n_0}), &\lambda \to 0\end{aligned}$$

from which the moments are determined.

From equations (3.5) and (3.6) it is possible to express $\hat{f}_{ab} = \hat{f}_{ab}^{+} + \hat{f}_{ab}^{-}$ in terms of the transforms $\hat{f}_c$ and from this fact we can express the moments $t_{ab}^{(n)}$ in terms of the one-sided first passage moments $t_c^{(n)}$. We get in fact from (3.5) and (3.6)

$$(6.2) \qquad \hat{f}_{ab}(x \mid \lambda) = \frac{\hat{f}_a(x \mid \lambda)(\hat{f}_b(a \mid \lambda) - 1) + \hat{f}_b(x \mid \lambda)(\hat{f}_a(b \mid \lambda) - 1)}{\hat{f}_a(b \mid \lambda)\hat{f}_b(a \mid \lambda) - 1},$$

and it follows that $t_{ab}^{(n)}(x)$ will be given by an algebraic combination of $t_a^{(k)}(x)$ and $t_b^{(j)}(x)$ for $k \leqq n$, $j \leqq n$, provided these moments exist. But it should be remarked that $t_{ab}^{(n)}(x)$ will exist in general for finite $\underset{\sim}{a}$, $\underset{\sim}{b}$, even though $t_c^{(k)}(x)$ may not, as the simple Wiener-Einstein process, for which $t_c^{(k)}(x) = \infty$ for $k \geqq 1$, shows.

In particular for $n = 1$, where we put $t^{(1)} = t$, we get for the mean first passage time by a simple expansion of (6.2),

$$(6.3) \qquad t_{ab}(x) = \frac{t_a(x)t_b(a) + t_b(x)t_a(b) - t_a(b)t_b(a)}{t_a(b) + t_b(a)}.$$

This formula leads to interesting consequences. Let $\underset{\sim}{a}$ and $\underset{\sim}{b}$ be such that $t_a(b) = t_b(a)$. Then since $t_b(a) = t_x(a) + t_b(x)$ (6.3) becomes simply

$$(6.4) \qquad t_{ab}(x) = \frac{t_a(x) - t_x(a)}{2}.$$

The right-hand side of (6.4) is independent of $\underset{\sim}{b}$, and since $t_{ab}(x) \geqq 0$ we have the result that *when $a > x > b$ and $t_a(b) = t_b(a)$ then $t_a(x) \geqq t_x(a)$.* Thus it is possible in a stationary process that the mean length of time it takes to go from a less probable state to a more probable state for the first time is longer than that it takes to reverse the journey. It is simple to construct processes for which this result obtains, for example, one in which the stationary density is symmetric and bimodal.

It is possible also to express the probability of absorption in the barrier a before b by means of the one-sided first passage moments. Since $\hat{f}_{ab}^{+}(x|0)$ is this probability we obtain from (3.5) and (3.6)

$$\hat{f}_{ab}^{+}(x \mid \lambda) = \frac{\hat{f}_a(b \mid \lambda)\hat{f}_b(x \mid \lambda) - \hat{f}_a(x \mid \lambda)}{\hat{f}_b(a \mid \lambda)\hat{f}_a(b \mid \lambda) - 1};$$

hence letting $\lambda \to 0$ we obtain the conclusion that *if the first passage moments exist the probability of absorption in a before b is given by $P = (t_a(b) + t_b(x) - t_a(x))/(t_a(b) + t_b(a))$.*

Since the expressions $\hat{f}$, $\hat{f}^{+}$, and $\hat{f}^{-}$ satisfy the differential equation (4.2) if the corresponding transition density p satisfies (4.1) it is possible to find the moments $t^{(n)}$ directly through a differential equation, and this often affords a method that is computationally more feasible than a direct evaluation of $\hat{f}$. We have in fact the following theorem.

THEOREM 6.1. *Let $X(t)$ satisfy the hypotheses of Theorem 4.1. Then if $T = T_{ab}(x)$*

is a proper random variable whose moments of order $n \leqq n_0$ exist, $t^{(n)} = t_{ab}^{(n)}(x)$ satisfies the system

$$\tag{6.5} \begin{aligned} \tfrac{1}{2}B^2 \frac{d^2t^{(n)}}{dx^2} + A \frac{dt^{(n)}}{dx} &= -nt^{(n-1)}, & n \leqq n_0 \\ t^{(0)} &\equiv 1 & \\ t_{ab}^{(n)}(a) = t_{ab}^{(n)}(b) &= 0, & n > 0. \end{aligned}$$

To prove the theorem we merely substitute the expansion (6.1) in the differential equation (4.2) and equate the coefficient of λ^n to zero.

The system (6.5) is particularly easy to solve since the substitution $Z^{(n)} = dt^{(n)}/dx$ renders each equation linear of the first order, and the solution can be written immediately in quadratures. Starting with $n = 1$ each $t^{(n)}$ can be obtained in turn in quadratures from the previous $t^{(k)}(k < n)$. In particular for $n = 1$ we have

$$\tag{6.6} \begin{aligned} \tfrac{1}{2}B^2 \frac{d^2t}{dx^2} + A \frac{dt}{dx} = -1, \qquad t = t_{ab}^{(1)}(x) \\ t(a) = t(b) = 0, \end{aligned}$$

a result we have used already in Section 5, Example c).

7. The range of $X(t)$. In this section we develop a formula for the distribution of the random variable

$$R(x, t) = \sup_{0 \leqq \tau \leqq t} X(\tau) - \inf_{0 \leqq \tau \leqq t} X(\tau)$$

which is called the *range* of $X(t)$, or the *oscillation* of $X(t)$, and we denote its distribution by $\Phi(x \mid r, t) = \Pr\{R(x, t) < r\}$. Note that this probability exists if $X(t)$ satisfies conditions A) and B) of Section 2.

A treatment of the random variable R for the Wiener-Einstein case has been given by Feller [10] in a statistical application, and the present section solves a problem he posed on finding the distribution of R for other processes.

Again we presume the existence of a density for R, say $\phi(x \mid r, t) = \partial\Phi(x \mid r, t)/\partial r$ only to expedite the analysis. It is not difficult to show that the existence of a density for T implies that for R.

THEOREM 7.1. *Let $X(t)$ satisfy conditions A) and B) and let $\phi(x \mid r, t)$ be the density of $R(x, t)$. Then for $\hat{f}_{ab}(x \mid \lambda)$ as in (3.4) we have*

$$\tag{7.1} \hat{\phi}(x \mid r, \lambda) = -\frac{1}{\lambda}\frac{\partial^2}{\partial r^2}\int_{x-(r/2)}^{x+(r/2)} \hat{f}_{v+(r/2),v-(r/2)}(x \mid \lambda)\, dv.$$

We note that $\Phi(x \mid r, \lambda)$, being merely $\int_0^r \hat{\phi}(x \mid u, \lambda)\, du$, is given immediately since $\hat{\phi}$ is expressed as a derivative.

The starting point of the proof is the formula

$$\phi(x \mid r, t) = \int_{x-r}^{x} \left[-\frac{\partial^2}{\partial a\, \partial b}(1 - F_{ab}(x \mid t))\right]_{a=b+r} db$$

which is established readily by an enumeration of cases. The existence of the derivative (under the integral sign) follows from the existence of the density of $X(t)$ at a and b, for when $\delta > 0$

$$F_{ab}(x \mid t) - F_{a+\delta,b}(x \mid t) = \Pr\{a < X(t) < a + \delta, X(\tau) > b, 0 \leqq \tau \leqq t\}.$$

On taking the Laplace transform of the preceding expression (which can be done under the integration and differentiation operations) we obtain

$$\hat{\phi}(x \mid r, \lambda) = \frac{1}{\lambda} \int_{x-r}^{x} \left[\frac{\partial^2}{\partial a \, \partial b} \hat{f}_{ab}(x \mid \lambda) \right]_{a=b+r} db$$

and the conclusions to the theorem follow by noting the identity

$$\frac{\partial^2}{\partial a \, \partial b} \hat{f}_{ab}(x \mid \lambda) \Big|_{a=b+r} \equiv \frac{\partial^2}{\partial b \, \partial r} \hat{f}_{b+r}(x \mid \lambda) - \frac{\partial^2}{\partial r^2} \hat{f}_{b+r,b}(x \mid \lambda).$$

As an application we consider the Wiener-Einstein process for which we have shown ((5.1) and (5.2))

$$\hat{f}_{ab}(x \mid \lambda) = \frac{\cosh \sqrt{2\lambda} \left(x - \frac{a+b}{2} \right)}{\cosh \sqrt{2\lambda} \left(\frac{a-b}{2} \right)},$$

and here (7.1) gives on performing the integration,

$$\hat{\phi}(x \mid r, \lambda) = - \sqrt{\frac{2}{\lambda^3}} \frac{\partial^2}{\partial r^2} \tanh \sqrt{\frac{\lambda}{2}} \, r$$

independent of x since the process is spatially homogeneous. This latter transform is easy to invert, and we have

$$\phi(x \mid r, t) = \frac{2}{\pi^2} \frac{\partial^2}{\partial r^2} \left\{ r \sum_{j=0}^{\infty} \frac{1}{(j + \frac{1}{2})^2} \exp \left(\frac{-2\pi^2 t (j + \frac{1}{2})^2}{r^2} \right) \right\}$$

$$= \frac{8}{\sqrt{2\pi t}} \sum_{j=1}^{\infty} (-1)^{j-1} j^2 e^{-(j^2 r^2 / 2t)},$$

these two expressions being related by Theta function identities, and the second being given by Feller [10]. For the moments we get from (7.2) immediately $E(R^n) = c_n t^{n/2}$ where

$$c_n = - \frac{2^{n/2}}{\Gamma\left(\frac{n}{2} + 1\right)} \int_0^{\infty} \rho^n \frac{d^2}{d\rho^2} \tanh \rho \, d\rho$$

so that, for example, $E(R) = \sqrt{8t/\pi}$, $E(R^2) = 4t \log 2$, etc.

REFERENCES

[1] T. W. ANDERSON AND D. A. DARLING, "Asymptotic theory of certain 'goodness of fit criteria' based on stochastic processes," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 193–212.

[2] L. BACHELIER, "Théorie mathématique du jeu," *Ann. Scuola Norm. Super. Pisa*, Vol. 18 (1901), pp. 143-210.

[3] G. A. BARNARD, "Sequential tests in industrial statistics," *J. Roy. Stat. Soc.*, Vol. 8 (1946), pp. 27-41.

[4] R. BELLMAN AND T. HARRIS, "Recurrence times for the Ehrenfest model," *Pacific J. Math.*, Vol. 1 (1951), pp. 179-193.

[5] W. DOEBLIN, "Sur l'équation de Kolmogoroff," *C. R. Acad. Sci. Paris*, Vol. 207 (1938), pp. 705-707.

[6] M. D. DONSKER, "An invariance principle for certain probability limit theorems," *Mem. Amer. Math. Soc.*, No. 6 (1951), 12 pp.

[7] J. L. DOOB, "Heuristic approach to the Kolmogoroff-Smirnov theorems," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 393-403.

[8] J. L. DOOB, "The brownian motion and stochastic equations," *Ann. Math.*, Vol. 2 (1942), pp. 351-369.

[9] W. FELLER, *An Introduction to Probability Theory and Its Application*, John Wiley and Sons, 1950.

[10] W. FELLER, "The asymptotic distribution of the range of sums of independent random variables," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 427-432.

[11] R. FORTET, Les fonctions aléatoire du type de Markoff associées à certaines équations linéares aux dérivées partielles du type parabolique," *J. Math. Pures Appl.*, Vol. 22 (1943), pp. 177-243.

[12] K. ITO, "On stochastic differential equations," *Mem. Amer. Math. Soc.* No. 4 (1951).

[13] A. KHINCHINE, "Asymptotische Gesetze der Wahrscheinlichkeitsrechnung," *Ergebnisse der Math.* Vol. 2 (1933).

[14] P. LÉVY, *Processus Stochastique et Mouvement Brownien*, Gauthier-Villars, Paris, 1948.

[15] H. E. ROBBINS, "Some aspects of the sequential design of experiments," *Bull. Amer. Math. Soc.*, Vol. 58 (1952), pp. 527-535.

[16] A. J. F. SIEGERT, "On the first passage time probability problem," *Physical Review*, Vol. 81 (1951), pp. 617-623.

[17] A. WALD, *Sequential Analysis*, John Wiley and Sons, 1947.

[18] E. T. WHITTAKER AND G. N. WATSON, *Modern Analysis*, Cambridge University Press, 1927.