PROBABILITY

A HISTORY OF PROBABILITY IN THE UNITED STATES OF AMERICA BEFORE THE TWENTIETH CENTURY

WILLIAM DOWELL BATEN

This article contains a history of the material concerning probability which was written by people who lived in the United States of America. It begins with an article presented in 1791 and extends through 1900. During this period many papers on probability were presented to the various scientific organizations. The subject matter includes life insurance theory, mortality tables, derivations of the normal curve of error, the method of least squares, probable errors of observations, games of chance, the importance of probability problems, probability laws of functions of chance variables, derivations of various probability laws, inverse probability, properties of polynomials, mean values, gambler's ruin, geometric probability, test books, teaching concerning probability, applications, probability courses, etc.

This paper also lists the names of the people who presented this material and summarizes their contributions to the development of probability subject matter.

MICHIGAN STATE COLLEGE,
EAST LANSING, MICH., U. S. A.

ON ASYMPTOTIC EXPANSIONS OF PROBABILITY FUNCTIONS

HARALD BERGSTRÖM

Consider the pr. f. of a sum $X^{(1)} + X^{(2)} + \cdots + X^{(n)}$ of n independent, equally distributed random variables in the euclidean space R_k of the finite dimension k. If each variable has the pr. f. $P(E)$, the sum has the pr. f. $P^{*n}(E)$, where $P^{*n}(E)$ denotes the n-fold convolution of P with itself. In particular for 1-dimensional d.f.'s the known Edgeworth expansion has been used as an approximation of the d.f. for the sum. In the following we give a more general asymptotic expansion of P^{*n} from which the Edgeworth expansion may be determined as a special case.

If $\varphi\ (E)$ is a set function and E a set, we denote the total variation of φ in R_k by $V[\varphi]$ and put $M_E[\varphi] = \mathrm{Max}_x \mid \varphi(E - x) \mid$, where $E - x$ denotes the set obtained by E through the translation $-x$. $g(n)$ (as well as $g_1(n, s)$) may denote a decreasing function of n with $g(n/\rho) = O[g(n)]$ for every positive constant ρ.

Let now Q be a pr. f. which shall be used for the approximation of P^{*n}. Then

$$(A_1) \qquad P^{*n} = \sum_{\nu=0}^{s} \Delta_n^{(\nu)} + r_n^{(s+1)}$$

with $\Delta_n^{(\nu)} = C_{n,\nu}Q^{*n-\nu}*(P-Q)^{*\nu}$ and some remainder term $r_n^{(s+1)}$. If the condition

$$(\mathrm{C}_1) \qquad V[\Delta_n^{(1)}] < g(n)$$

is satisfied, $\Delta_n^{(\nu)} = O[g^\nu(n)]$ for any fixed integer ν. However $\Delta_n^{(\nu)}$ may be of a smaller order of magnitude than $g^\nu(n)$. Let now (C_1) hold and let $S = \sum_{n=1}^{\infty} g^{s+1}(n)/n$ converge when s is some fixed integer $\geqq 0$, and suppose that the inequalities $M_E[\Delta_n^{(s+1)}] < \mu(E)g_1(n, s)$, $M_E(P) < \mu(E)$, $M_E(Q) < \mu(E)$ and the condition

$$(\mathrm{C}_2) \qquad \lim_{p\to\infty} V[P^{*p}*(P-Q)] = 0$$

are satisfied. Then

$$(1) \qquad M_E[r_n^{(s+1)}] < C\mu(E)g_1(n, s),$$

where C depends on s and the functions $g(n)$ and $g_1(n, s)$ but not on n. When (C_1) and (C_2) are satisfied and S converges, then (A_1) has moreover the strictly asymptotical properties

$$(2) \qquad V[\Delta_n^{(\nu)}] = O[g^\nu(n)], \qquad V[r_n^{(s+1)}] = O[g^{s+1}(n)]$$

and (C_1) and (C_2) are necessary for the existence of these properties when S converges. The condition (C_2) is used only to that extent that $V[P^{*n}*(P-Q)]$ is smaller then some value depending on s and the functions $g(n)$ and $g_1(n, s)$ but not on n.

When P is a nonsingular d.f. with finite absolute mean values of the order λ for some λ in the interval $2 < \lambda \leqq 3$, and Q is that (nonsingular) normal d.f. which has the same mean value vector and the same moment matrix as P, then (C_1) is satisfied with

$$g(n) = n^{-(\lambda-2)/2}\log^{k/2} n$$

and $\Delta_n^{(\nu)} = o(n^{-\nu(\lambda-2)/2})$ or $O[n^{-\nu(\lambda-2)/2}]$ according as $\lambda < 3$ or $\lambda = 3$. We have

$$(3) \qquad r_n^{(s+1)} = O[n^{-1/2} + n^{-(s+1)(\lambda-2)/2}]$$

always, and

$$(4) \qquad r_n^{(s+1)} = O[n^{-(s+1)(\lambda-2)/2}]$$

when (C_2) holds. Here (C_2) may be replaced by the less restrictive condition

$$(\mathrm{C}_3) \qquad \lim_{n\to\infty} V\{P^{*n}(x)*Q[(an)^\alpha x]*[P(x) - Q(x)]\} = 0$$

with $\alpha = (s+1)(\lambda-2)/2 - 1/2$ and a constant $a > 1$. In the 1-dimensional case, (C_3) is necessarily satisfied if

$$r_n^{(s+1)} = o(n^{-(s+1)(\lambda-2)/2}).$$

From (A_1) different expansions may be derived. Let P be a nonsingular pr. f. with finite absolute mean values of order λ for some λ in the interval $2 < \lambda \leqq 3$, and let Q be that normal pr. f. which has the same mean value vector and the same moment matrix as P. Further suppose that $M_E(P)/m(E)$ is bounded, when E belongs to any family, regular in the sense of Lebesgue, and $m(E) > 0$ is the measure of E. Then denoting the derivative in the sense of Lebesgue by $D[\varphi]$ for a set function φ, we have

$$D[\Delta_n^{(\nu)}] = O[n^{-(k+\nu(\lambda-2))/2}] \tag{5}$$

everywhere and, if the condition (C_2) is satisfied,

$$D[r_n^{(s+1)}] = O[n^{-(k+(s+1)(\lambda-2))/2}] \tag{6}$$

in all points, where DP^{*n} exists. If P is absolutely continuous, the condition (C_2) is necessary for the validity almost everywhere of that last relation.

CHALMERS UNIVERSITY OF TECHNOLOGY,
GOTHENBURG, SWEDEN.

RANDOM FUNCTIONS ON DIVISIA ENSEMBLE

MARIA CASTELLANI

In this paper we investigate a class of real measurable functions of a statistical Divisia ensemble. Let us consider a lattice space with a set D of elements which renew themselves at random during a certain time interval (t_0, t), because there is an inflow and outflow of elements. Let us assume that each element is associated with a real, measurable, bounded function $\delta(\theta - x)$ for $(\theta - x) \subseteq (t - t_0)$.

Let $p(\theta - x)$ be the probability function of an element which is not being eliminated from D, $d\mu_\theta$ the probability density of an element which is eliminated, $d\nu_x$ the elements which are flowing into D during the time interval $(x, x + dx)$, and λ_0 the elements in D at the time t_0. In several problems it might be more convenient to assume $p(\theta - x)\cdot\delta(\theta - x) = f(\theta - x)$. According to any given law of random flow, the expected value $E(\delta)$ is

$$E(\delta) = \lambda_0 f(t - t_0) + \lambda_0 \int_{t_0}^{t} f(\theta - t_0)\, d\mu_\theta + \int_{t_0}^{t} f(t - x)\, d\nu_x + \int_{t_0}^{t}\int_{x}^{t} f(\theta - x)\, d\mu_\theta\, d\nu_x .$$

We do not investigate this general integral equation but, instead, limit our attention to obtaining some fundamental equations of actuarial mathematics. When the δ is a linear function and we assume a Karup functional relation $p(\theta - x)\, d\mu_\theta = -\,dp(\theta - x)$, we then obtain a generalized Galbrun integral equation for sickness insurance. A generalized Cantelli equation for mathematical reserves is easily derived by considering the actual rate of elements flowing out.

The Benini-Des Essars equation for the average time spent in D by the flowing elements is derived under very restrictive conditions of continuity and linearity.

UNIVERSITY OF KANSAS CITY,
KANSAS CITY, MO., U. S. A.

AN ERGODIC THEOREM FOR STATIONARY MARKOV CHAINS WITH A COUNTABLE NUMBER OF STATES

KAI LAI CHUNG

Let X_n, $n = 0, 1, \cdots$, be the random variables of a Markov chain with a countable number of states, numbered $1, 2, \cdots$, and with stationary transition probabilities $P_{jk}^{(1)}$. As usual $P_{jk}^{(n)}$ denotes the n-step transition probability. Write $T_{jk}^{(n)} = \sum_{\nu=0}^{n} P_{jk}^{(\nu)}$. The ergodic theorem for Markov chains states that if j and k belong to the same ergodic class, then $\lim_{n\to\infty} n^{-1}T_{jk}^{(n)} = p_k$. It is known that if $p_j > 0$, then $p_k > 0$ so that the above implies $\lim_{n\to\infty} T_{jj}^{(n)}/T_{kk}^{(n)} = p_j/p_k$. If $p_j = 0$, then $p_k = 0$ and no conclusion can be drawn of the existence of the last-written limit. This question was raised by Kolmogorov and answered by Doeblin (Bull. Soc. Math. France vol. 66 (1938) pp. 210–220). More precisely he proved that if j and k are such that there exist positive integers m and m' for which $P_{jk}^{(m)} > 0$ and $P_{kj}^{(m')} > 0$, then $\lim_{n\to\infty} T_{jk}^{(n)}/T_{kk}^{(n)} = 1$, and $\lim_{n\to\infty} T_{jj}^{(n)}/T_{kk}^{(n)}$ exists and is positive and finite.

A new proof of Doeblin's theorem is given by introducing the quantities in the infinite series written below and using generating functions. This method also determines the last-written limit to be equal to λ_j/λ_k where $\lambda_j = 1 + \sum_{n=1}^{\infty} P(X_\nu \neq k \text{ for } 1 \leqq \nu \leqq n, X_n = j \mid X_0 = j)$ and λ_k is obtained by interchanging j and k in λ_j. With extraneous conditions on the analytic nature of $T_{kk}^{(n)}$ the theorem becomes a consequence of a Hardy-Littlewood-Karamata-Tauberian theorem. Our proof circumvents this invocation. An extension is made to Markov chains with a continuous parameter t. If e. g. the uniformity condition $\lim_{t\downarrow 0} P_{jj}(t) = 1$ uniformly in all j is assumed, Doeblin (Skandinavisk Aktuarietidskrift vol. 22 (1939) pp. 211–222) has shown that almost all sample functions are step functions, etc. Under these circumstances it is proved that $\lim_{t\to\infty} \int_0^t P_{jj}(s)\,ds/\int_0^t P_{kk}(s)\,ds$ exists and is positive and finite, in fact a determination of its value similar to the above is given. Contrary to hope this natural extension is not easily deducible from the discrete parameter case, but our method works with Laplace transforms instead of generating functions. The question of the existence of $\lim_{n\to\infty} P_{jj}^{(n)}/P_{kk}^{(n)}$ (or of $\lim_{t\to\infty} P_{jj}(t)/P_{kk}(t)$) seems to be open. Only in the special case where X_n is the sum of n independent, identically distributed random variables with mean 0, it has been proved by Erdös and the present author that this limit is one.

CORNELL UNIVERSITY,
ITHACA, N. Y., U. S. A.

ÉLÉMENTS ALÉATOIRES DE NATURE QUELCONQUE

R. FORTET

À la suite de travaux de M. Fréchet, je me suis intéressé, et j'ai incité quelques jeunes probabilistes à s'intéresser, à l'élaboration d'une théorie directe des éléments aléatoires de nature quelconque (ne se réduisant pas nécessairement à des variables aléatoires numèriques à une ou à un nombre fini de dimensions); des applications à la théorie des fonctions aléatoires ou à divers problèmes concrets justifient cette étude; il est naturel, surtout si on a en vue l'application aux fonctions aléatoires, d'aborder d'abord le cas d'un élément X prenant ses valeurs dans un espace linéaire $\mathfrak{X}$; un premier problème est alors d'étendre au cas d'un tel élément la notion d'espérance mathématique, ce qui revient à une théorie de l'intégration; dans le cas où $\mathfrak{X}$ est un espace de Banach, Mlle Mourier a montré (C.R. Acad. Sci. Paris t. 229 (1949) p. 1300) qu'une application de l'intégrale de Pettis permet de perfectionner notablement des résultats de Fréchet; tandis qu'une extension, non encore publiée, de la méthode de Daniell faite par M. Régnier permet de définir une espérance mathématique pour une large catégorie de f.a. mesurables au sens de Doob.

Un second problème est d'envisager l'addition d'éléments aléatoires X_1, $X_2, \cdots, X_n, \cdots$ à valeurs dans $\mathfrak{X}$; pour cela la notion de caractéristique semble devoir être utile; si $\mathfrak{X}$ est un espace de Banach, on peut définir la caractéristique $\varphi(x^*)$ d'un élément aléatoire X, comme fonction de la fonctionnelle linéaire x^* variable dans le dual $\mathfrak{X}^*$ de $\mathfrak{X}$, par $\varphi(x^*) = E[x^*(X)]$; des propriétés de $\varphi(x^*)$ s'établissent aisément, dont la principale est que $\varphi(x^*)$ est "définie positive"; mais une fonction définie positive $\varphi(x^*)$, même continue avec la topologie faible dans $\mathfrak{X}^*$, n'est pas forcément une caractéristique; si $\mathfrak{X}$ est séparable et réflexif, on peut donner une condition nécessaire et suffisante pour que $\varphi(x^*)$, supposée définie positive, soit la caractéristique d'un élément aléatoire X "proprement dit" (c'est-à-dire tel que Pr $[\,\|X\| < +\infty] = 1$); on peut définir un élément laplacien X par la condition que, en supposant $E(X) = \theta$ (θ étant l'élément 0 de $\mathfrak{X}$), le logarithme de la caractéristique $\varphi(x^*)$ est de la forme: $-(1/2)E[x^*(Z)]$, où Z est un élément aléatoire proprement dit quelconque à valeurs dans $\mathfrak{X}$.

UNIVERSITY OF CAEN,
CAEN, FRANCE.

A NOTE ON THE GENERAL CHEBYCHEFF INEQUALITY

SALEM H. KHAMIS

Let $\Phi(x)$, $a \leqq x \leqq b$, be a cumulative distribution function, differentiable, never-decreasing and possessing moments up to order $2n$. Let $\Phi(x)$ be defined as 0 for $x \leqq a$ and 1 for $x \geqq b$, and assume that it has at least $n + 1$ points of increase. If $\Psi(x)$, $a' \leqq x \leqq b'$, be another cumulative distribution function with similar properties and identical moments up to order $2n$, then, by the Chebycheff inequality, $|\Phi(x) - \Psi(x)| \leqq [Q_n(x)Q'_{n+1}(x) - Q'_n(x)Q_{n+1}(x)]^{-1} = \Omega_{2n}(x)$, say, where, $Q_r(x)$, $r = 0, 1, 2, \cdots$, or n, is the denominator of the rth convergent of the continued fraction associated with $\int_{-\infty}^{+\infty} d\Phi(t)/(x - t)$ or $\int_{-\infty}^{+\infty} d\Psi(t)/(x - t)$ and is a polynomial of exact degree r possessing orthogonal properties with respect to $d\Phi(x)/dx$ and $d\Psi(x)/dx$ as weight functions.

We have shown that the right-hand side of the above inequality may be improved upon, obtaining for it the value $A\Omega_{2n}(x)$, where A is given by

$$0 \leqq A = 1 + \min\left[\max\left\{-\frac{\Psi'(x)}{\Phi'(x)}\right\}, \quad \max\left\{-\frac{\Phi'(x)}{\Psi'(x)}\right\}\right] \leqq 1,$$

provided the maxima are finite.

We have shown also that when the property "never-decreasing" does not hold, the difficulty, under certain general conditions, could be overcome, thus making it possible to apply the Chebycheff inequality to approximations for distribution functions by series expansions involving negative components. Further, the general relationship between the polynomials $Q_r(x)$ and the corresponding orthogonal polynomials, as usually defined in treatise on orthogonal polynomials, is given.

The methods used in this paper are illustrated, particularly, in the case of the Pearson type III distribution function.

Statistical Office, United Nations,
Lake Success, N. Y., U. S. A.

UNLIKELY EVENTS IN GENERAL STATIONARY-TRANSITION MARKOFF CHAINS

Bernard Osgood Koopman

A system is considered whose possible states form a perfectly general abstract set X, the observable subsets of which form a σ-algebra Σ. It evolves stochastically through a sequence of n stages, at the kth of which it is in the state $x_k \in X$ $(k = 1, \cdots, n)$. The transition probabilities $p_{n,k}(x, E) = \text{prob}(x_k \in E \mid x_{k-1} = x)$ are one-step Markoffian $(x \in X, E \in \Sigma)$. For a certain $S \in \Sigma$, the event $[x_k \in S]$ is called "success on the kth trial". The object is to find $P(s) = \lim_{n\to\infty} P_n(s)$, where $P_n(s) =$ probability of s successes in first n trials.

It is shown that the class $\mathfrak{B}$ of all complex-valued functions $f = f(x, E)$, defined and bounded over $(x, E) \in (X, \Sigma)$, σ-additive in E (x fixed) and measurable (Σ) in x (E fixed), is a Banach algebra, provided linear combinations with complex coefficients are defined as usual, the $\mathfrak{B}$-product $fg = (fg)(x, E)$ as the abstract integral of $g(y, E)$ with respect to the measure $f(x, F_y)((y, F_y)$

being variables of integration), and norm $N(f)$ as $\sup_{x \in X}$ [absolute variation of $f(x, E)$ on X]. The unit element $u \in \mathfrak{B}$, where $u(x, E) = \chi_E(x)$ = characteristic function of set E. Accents and $\cap$ denote complements and intersections in X; and the ordinary (non $\mathfrak{B}$-) product is written $f(x, E)g(x, E)$.

The following hypotheses involve the stationarity of transitions, the improbability of success, and the Markoffian regularity of $\lim_{m \to \infty} h^m = g$ in (4):

(1) Success $\to$ success: $\chi_S(x)p_{n,k}(x, S \cap E) = a(x, E)$, $N(a) < 1$.
(2) Success $\to$ failure: $\chi_S(x)p_{n,k}(x, S' \cap E) = b(x, E)$.
(3) Failure $\to$ success: $\chi_{S'}(x)p_{n,k}(x, S \cap E) = c(x, E)/n$.
(4) Failure $\to$ failure: Writing $\chi_{S'}(x)p_{n,k}(x, S' \cap E)$ as $g_{n,k}(x, E)p_{n,k}(x, S')$,

where $g_{n,k}(x, E) = \text{prob}(x_k \in E \mid x_{k-1} = x, x_k \in S')$, we assume: (i) $g_{n,k}(x, E) = g(x, E)$; (ii) $g(x, E) \geqq \chi_{S'}(x)\nu(E)$, where $\nu(E)$ is σ-additive over Σ, $\nu(E) \geqq 0$, $\nu(S') > 0$, $\nu(S' \cap E) = \nu(E)$. By classical reasoning $g^m \to h$, where $h(x, E) = \chi_{S'}(x)\mu(E)$, and $\mu(E)$ is like $\nu(E)$, but $\mu(S') = 1$.

Then it is shown that $P(s)$ is generated by the function $\varphi(t) = \sum_{s=0}^{\infty} P(s)t^s$ obtained by first constructing the Banach-algebraic expression

$$f_t = p_0[(u - at)^{-1}b + u]\, g \exp\,[(t - 1)c(u - at)^{-1}h],$$

where $h(x, E) = \chi_S(x)\mu(E)$, and $p_0(x, E) = p_0(E)$ is the initial probability that $x_1 \in E$; and then setting $\varphi(t) = f_t(x, X)$ (which is independent of x).

An equally explicit but more complicated result is obtained in the non-stationary case.

Columbia University,
New York, N. Y., U. S. A.

SOME NON-NEGATIVE TRIGONOMETRIC POLYNOMIALS CONNECTED WITH A PROBLEM IN PROBABILITY

Eugene Lukacs and Otto Szász

Let $0 < b_1 < b_2 < \cdots < b_n$ be n integers and let $g(\theta)$ be the Vandermonde determinant formed from $b_1^2, b_2^2, \cdots, b_n^2$ with the first row replaced by $\sin^2 b_i\theta/2$ ($i = 1, \cdots, n$). The function $g(\theta)$ is then a cosine polynomial. In connection with a problem in probability the question arose as to when $g(\theta)$ is a non-negative trigonometric polynomial. The following results were obtained:

(A) If the b_i are the first n consecutive integers, then $g(\theta)$ is non-negative.

(B) If the b_i are the first n consecutive odd integers, then $g(\theta)$ is non-negative.

(C) If the numbers $b_1, \cdots, b_n$ are obtained from the first $(n + 1)$ consecutive integers by omitting the integer k ($1 \leqq k \leqq n$), then the trigonometric polynomial $g(\theta)$ is non-negative if and only if $2k^2 \geqq n + 1$.

(D) If the numbers $b_1, \cdots, b_n$ are obtained from the first $(n + 1)$ consecutive odd integers by omitting the kth odd integer $2k - 1$ $(1 \leqq k \leqq n)$, then the trigonometric polynomial $g(\theta)$ is non-negative if

1. $k \geqq k_1$ where $k_1 = \dfrac{1 + (1 + 2n)^{1/2}}{2}$ or if
2. $k < k_1$ but if $\int_{-x_0}^{1} (1 - t^2)^{n-1}(t^2 - x_0^2)\, dt \geqq 0$ where

$$x_0 = \left(\frac{n + 2k - 2k^2}{2n(n + 1) - 2k(k - 1)}\right)^{1/2}.$$

If neither of these conditions is satisfied, $g(\theta)$ assumes also negative values.

(E) If the numbers $b_1, \cdots, b_n$, are chosen from the first $(n + 2)$ consecutive integers by omitting two integers k and p $(1 \leqq k \leqq p - 1 \leqq n)$, the non-negativeness depends on the discussion of a quadratic polynomial whose coefficients are functions of n, k, and p.

National Bureau of Standards,
Washington, D. C., U. S. A.
National Bureau of Standards,
Los Angeles, Calif., U. S. A.
University of Cincinnati,
Cincinnati, Ohio, U. S. A.

INFORMATION RETRIEVAL VIEWED AS TEMPORAL SIGNALLING

Calvin N. Mooers

The problem of directing a user to stored information, some of which may be unknown to him, is the problem of "information retrieval". Signalling theories can be applied, though this is a form of temporal signalling, which distinguishes it from the point-to-point signalling currently under study by others.

In information retrieval, the addressee or receiver rather than the sender is the active party. Other differences are that communication is temporal from one epoch to a later epoch in time, though possibly at the same point in space; communication is in all cases unidirectional; the sender cannot know the particular message that will be of later use to the receiver and must send all possible messages; the message is digitally representable; a "channel" is the physical document left in storage which contains the message; and there is no channel noise because all messages are presumed to be completely accessible to the receiver. The technical goal is finding in minimum time those messages of interest to the receiver, where the receiver has available a selective device with a finite digital scanning rate.

Classification and indexing schemes are ruled out because of gross topological difficulties, and the Batten system because of no reasonable large-scale mechanization. To avoid scanning all messages in entirety, each message is characterized by N independently operating digital descriptive terms (representing ideas) from a vocabulary V, and a selection is prescribed by a set of S terms. Conventional assignment of a digital configuration to a message requires $N \log_2 V$ binary digit places. It is a nonsingular transformation from ideas to configuration. Prescription of a selection requires $S \log_2 V$ digit places. Although only $\log_2 M$ digit places are required to differentiate (or enumerate) M messages, $S \log_2 V$ may be many times greater, indicating a digital redundancy and waste in coding. This redundancy can be removed by recoding the descriptive ideas into the digital configuration representing the message by a singular transformation (Zatocoding). The message can then be characterized with $(N/S) \log_2 M$ digit places, with a possible increase in scanning rate per message. Because Zatocoding represents each idea by a pattern ranging over the digit places for a message, and superposes these in the same coordinate frame by Boolean addition, the selection process includes the desired messages and gives a statistical exclusion of the undesired messages.

Zator Company,
Boston, Mass., U. S. A.

ON A CLASS OF TWO-DIMENSIONAL MARKOV PROCESSES

Murray Rosenblatt

Let $x(t)$ be an element of Wiener space and let $V(t, x)$ satisfy the local Hölder condition $| V(t, x) - V(t + \Delta t, x + \Delta x) | \leqq M(t, x) \{ | \Delta t |^{\alpha(t,x)} + | \Delta x |^{\alpha(t,x)} \}$, $\alpha(t, x) > 0$ everywhere except on a curve rectifiable in every finite rectangle of the (t, x) plane. Let $V(t, x)$ be bounded in every finite rectangle of the (t, x) plane. The study of the two-dimensional Markov process $(x(t), \int_0^t V(\tau, x(\tau))\, d\tau)$ is related to the study of certain differential and integral equations. Let $Q(t, x) = E\{e^{-u\int_0^t V(\tau,x(\tau))d\tau} \mid x(t) = x\} e^{-x^2/2t}/(2\pi t)^{1/2}$. If $0 \leqq V(t, x)$, $Q(t, x)$ is the solution of the differential equation $(1/2)\partial^2 Q/\partial x^2 - \partial Q/\partial t - uV(t, x)Q = 0$, subject to the conditions $Q(t, x) \to 0$, $x \to \pm\infty$, $\lim_{t\to 0} \int_{-\epsilon}^{\epsilon} Q(t, x)dx = \lim_{t\to 0} \int_{-\infty}^{\infty} Q(t, x)dx = 1$ for all $\epsilon > 0$. $Q_x(t, x)$ is continuous in (t, x) for all $(t, x) \neq (0, 0)$. Let $G(x, y, t) = \Pr\{\int_0^t V(\tau, x(\tau))d\tau \leqq y \mid x(t) = x\}$. Then $\int_y^{y+\lambda} (E(\alpha) - G(x, \alpha, t))\, d\alpha = e^{x^2/2t}(2\pi t)^{1/2} \int_0^t \int_{-\infty}^{\infty} e^{-(x-\xi)^2/2(t-\tau)-\xi^2/2\tau}/2\pi(\tau(t - \tau))^{1/2} V(\tau, \xi)\{G(\xi, y + \lambda, \tau) - G(\xi, y, \tau)\}\, d\xi\, d\tau$ where $E(y) = 1$ if $y \geqq 0$, and $=0$ if $y < 0$. This last equation is valid if one assumes that $V(t, x)$ is bounded and Borel measurable.

Cornell University,
Ithaca. N. Y., U. S. A.

DISTRIBUTION FOR THE ORDINAL NUMBER OF SIMULTANEOUS EVENTS WHICH LAST DURING A FINITE TIME

HERMANN VON SCHELLING

If we know (1) that an event of a given class may occur with the probability p per unit of time; (2) that, when it has happened, it will not occur again during the next k units of time; (3) that such an event was observed at the nth moment; we ask for the probability $w(m; n, k, p)$ that this event is the mth one since zero time.

Relative probabilities can be found immediately. Their sum $s_n(x)$ with $x = p/q^{k+1}$ satisfies the following recursion formula

$$s_{n+1}(x) - s_n(x) - xs_{n-k}(x) = 0$$

with $s_1(x) = s_2(x) = \cdots = s_{k+1}(x) = 1$. Its solution depends upon the roots of the characteristic equation

$$f(z) = z^{k+1} - qz^k - 1 + q = 0.$$

It has to be proved that (a) $f(z) = 0$ does not have multiple roots; (b) $z_1 = +1$ is the only root of $f(z) = 0$ on the unit circle; (c) the roots $z_2, z_3, \cdots, z_{k+1}$ of $f(z) = 0$ are located inside the unit circle.

Exact formulas for $w(m; n, k, p)$ and for the first and the second moment of this distribution are given as functions of the roots of $f(z) = 0$. According to (a), (b), (c) it is possible to present simple approximations which do not depend on the solution of the characteristic equation and may be evaluated easily.

The distributions $w(m; n, k, p)$ are branching from the binomial type which corresponds to the particular case $k = 0$. The final results are elementary. It is to be expected that the new distributions will become useful in many fields of science and biology since the assumption (2) is frequently adequate to observed conditions.

NAVAL MEDICAL RESEARCH LABORATORY,
NEW LONDON, CONN., U. S. A.

INFORMATION AND THE FORMAL SOLUTION OF MANY-MOVED GAMES

L. S. SHAPLEY

A game is said to be in *normalized form* if it is so formulated that each player has just one move—the selection of a *strategy*—made independently of the other players' selections. A 2-person zero-sum game with n moves in which no information passes between moves can easily be formulated in this fashion, and its solution is obtained by considering:

$$(1) \qquad \max_{\varphi} \min_{\psi} \iint H(x_1, \cdots, x_n)\, d\varphi(x_{i_1}, \cdots, x_{i_p})\, d\psi(x_{j_1}, \cdots, x_{j_{n-p}})$$

where φ is a probability distribution over the set of strategies $(x_{i_1}, \cdots, x_{i_p})$ of the maximizing player, and ψ is the same for the minimizer. We suppose that the identity of the player making the ith move, as well as the set A_i of his alternatives, does not depend on the previous course of play. We may abbreviate (1) to

$$\operatorname*{mix}_{S|I-S} H(x_1, \cdots, x_n)$$

where I is the set of integers $\{1, \cdots, n\}$ and S the set of moves assigned to the maximizer.

The introduction of *information* (i.e. knowledge, at one move, of the outcome of another) quickly makes the normalized form useless in practice, since the number of strategies increases at a prohibitive rate. Yet, in the extreme case of *perfect information* (knowledge, at each move, of all preceding moves), the game is relatively easily solved one move at a time, e.g.:

$$(2) \qquad \max_{x_1 \in A_1} \min_{x_2 \in A_2} \min_{x_3 \in A_3} \cdots \max_{x_n \in A_n} H(x_1, x_2, x_3, \cdots, x_n).$$

It is of interest to ask what conditions on the informational structure are required for a game to be decomposable into a sequence of separated moves (as in (2)), or, more generally, separated subgames in normalized form. The solution in such a case would be given by:

$$(3) \qquad \operatorname*{mix}_{T_1|U_1} \operatorname*{mix}_{T_2|U_2} \cdots \operatorname*{mix}_{T_m|U_m} H(x_1, \cdots, x_n)$$

where the T_k and U_k partition S and $I - S$ respectively. Using McKinsey's notion of equivalence of patterns of information, we find that a necessary and sufficient condition that a game be solvable by (3), regardless of the nature of the function H or of the sets A_i, is that it be possible to order the moves in time so that (i) the information pattern never implies knowledge of the future, (ii) each player's knowledge of his opponent's actions increases monotonically, and (iii) the accessions of information occur only between certain pairs of consecutive moves, on which occasions both players are apprised of all their opponent's actions to date. No condition is imposed on a player's awareness of his own previous actions.

THE RAND CORPORATION,
SANTA MONICA, CALIF., U. S. A.

STOCHASTIC PROCESSES BUILT FROM FLOWS

KÔSAKU YOSIDA

By virtue of the theory of semi-groups due to E. Hille and the author, we may construct stochastic processes in a separable measure space R from flows in R. A flow (= one-parameter family of equi-measure transformations in R) $F_t x$ induces a one-parameter group T_t of transition operators: $(T_t f)(x) = f(F_t x)$, $-\infty < t < \infty$. Here a linear operator T on the Banach space $L_1(R)$ to $L_1(R)$ is called a transition operator if $f(x) \geqq 0$ implies $(Tf)(x) \geqq 0, \int (Tf)(x)\, dx = \int f(x)\, dx$. T_t admits infinitesimal generator A:

$$T_t f = \exp\,(tA)f = \text{strong}\lim_{n\to\infty} \exp\,(nt((I - n^{-1}A)^{-1} - I))f, \; -\infty < t < \infty.$$

Since $(I - n^{-1}A)^{-1}$ and $(I + n^{-1}A)^{-1}$ must be transition operators for $n > 0$, $(I - n^{-1}A^2)^{-1}$ exists as a transition operator. Hence A^2 is the infinitesimal generator of a one-parameter semi-group $S_t = \exp\,(tA^2)$, $t \geqq 0$, of transition operators. Thus the Fokker-Planck equation in a Riemannian space R: $\partial f(t, x)/\partial t = A^2 f(t, x)$, $f(0, x) = f(x) \in L_1(R)$, $t \geqq 0$, is integrable stochastically if $A = p^i(x)\partial/\partial x^i$ is the infinitesimal transformation of a one-parameter Lie group of equi-measure transformations of R.

Extensions. i) If R admits several flows $\exp\,(tA_i)$, $i = 1, 2, \cdots, m$, with mutually commutative A_i, then $\exp\,(t \sum_i A_i^2)$ defines a stochastic process in R. ii) Let the group of motions in a Riemannian space R be a semi-simple Lie group with infinitesimal transformations $X_1, X_2, \cdots, X_m$ transitive in R. Then the Casimir operator $C = g^{ij}X_iX_j$ $((g^{ij}) = (g_{ij})^{-1}$, $g_{ij} = c^{\sigma}_{i\rho}c^{\rho}_{j\sigma}$, $[X_i, X_j] = c^k_{ij}X_k)$ is commutative with every X_i. In this case, at least when R is compact, $\exp\,(tC)$ defines a temporally and spatially homogeneous, continuous stochastic process in R—a Brownian motion in the homogeneous space R.

NAGOYA UNIVERSITY,
NAGOYA, JAPAN.