STOCHASTIC GROUPS AND RELATED STRUCTURES

ULF GRENANDER
UNIVERSITY OF STOCKHOLM

1. Introduction

The last few years have witnessed an increasing interest in the probability theory of general algebraic and topological structures, especially for topological groups and linear vector spaces. This paper attempts to survey this new field and to present some of the main results together, so as to obtain as complete an exposition as possible of the present state of development. As will be painfully obvious in the following pages no unified theory exists yet, and we can answer only partially the problems that arise. Still, when the existing results are viewed together, it is hoped that the overall picture will be suggestive.

To avoid excessive length and the obscuring of general ideas by details, no proofs are given. To compensate for this much attention is given to the description of the analytical tools that are suitable for proving the sort of results that are described in the following.

The generality of the subject might give the impression that these are abstract and rather vague problems. Actually the situation is just the opposite: this is a piece of very concrete mathematics and many of the problems can be phrased in simple and direct form, although they may be far from simple to solve. The fundamental character of the questions makes this extension of classical probability theory a fascinating study for the probabilist and analyst in search of nontrivial generalizations of classical probability theory.

2. Semigroups

Consider a Hausdorff space S with a binary operation xy on $x, y \in s$, such that S is a topological semigroup. From the subsets of S we form the σ-algebra $\mathcal{B}(S)$ of Borel sets and the set $\mathcal{P}(S)$ of regular probability measure defined on $\mathcal{B}(S)$.

For two measures $P_1, P_2 \in \mathcal{P}(S)$ we define their convolution

$$P_1 * P_2(E) = \int_{xy \in E} dP_1 \times dP_2\,(x, y), \tag{2.1}$$

where $E \in \mathcal{B}(S)$ and the product measure $P_1 \times P_2$ on $S \times S$ is introduced in the usual way. With the convolution operation $\mathcal{P}(S)$ becomes a semigroup. By P^{n*} we mean the nth iterate $P * P * \cdots * P$.

We will sometimes have a useful analytical tool in the linear operator

$$Tf(x) = \int_S f(x\, x')\, dP\, (x'). \tag{2.2}$$

Defined in an appropriate functional space, this operator is a natural instrument in the study of limit theorems since the iterates T^n correspond simply to the convolutions P^{n*}.

Some of the problems to be studied are:

(a) When do *limit laws* exist, that is, when does $\Pi = \lim_{n\to\infty} P^{n*}$ exist, with the limit defined by a suitable topology? What are the limit laws?

(b) If $P = P^{2*}$, P is called an *idempotent.* Describe all idempotents.

(c) If for every $n = 1, 2, \cdots$ there is a $Q_n \in \mathcal{P}(S)$ such that $P = Q_n^{n*}$, then P is called an *infinitely divisible distribution* over S. How can they be characterized?

(d) To study these and other questions it is sometimes useful to apply various kinds of *Fourier analysis* and we then have to construct the relevant analytical instrument, playing a role similar to that of characteristic functions in R^k.

(e) For each $t > 0$ let ξ_t be a stochastic variable taking values in a semigroup S with a unit element and such that its distributions P_t satisfy

(1) $P_{t+t'} = P_t * P_{t'}$,

(2) $P_t(U) \to 0$ as $t \to 0$ for any neighborhood U of the unit element. Then ξ_t is called a *homogeneous process* and an important problem is the description of the associated semigroup P_t, its infinitesimal generator, and so on.

The above definition of our main object $\mathcal{P}(S)$ is, at least at present, sufficiently general to include the cases studied. On the other hand its generality prevents us from getting any substantial results; to be able to do this we will have to introduce more structure on S and this will be done in several steps in what follows.

2.1. *Compact semigroups.* The first specialization that we will make is that S is compact; this leads to definite results but the restrictive hypothesis of course narrows down the field of application.

We then know that the set of regular probability distributions over S is also compact, so that $\mathcal{P}(S)$ is a compact topological semigroup. Clearly the support of P^{n*} will be in $s(P)^n = s(P)s(P) \cdots s(P)$ so that we can just as well restrict ourselves to the semigroup

$$\overline{\bigcup_1^\infty s(P)^n} \tag{2.3}$$

and this will be assumed to have been done.

In passing let us mention the important relation

$$s(P_1 * P_2) = s(P_1)s(P_2), \tag{2.4}$$

which tells us how to get the support of the convolution of probability measures.

We shall say that a sequence $P_n \in \mathcal{P}(S)$ converges to $P \in \mathcal{P}(S)$ if

$$\int_S f(s)\, dP_n(s) \to \int_S f(s)\, dP(s) \tag{2.5}$$

for every $f \in C(S)$.

If a sequence P_n is convergent its limit must be an idempotent. Since $\mathcal{P}(S)$ is compact one knows that it has at least one idempotent; we shall see below how idempotents can be constructed starting from arbitrary probability distributions.

The operator T now maps $C(S)$ into $C(S)$. Introduce the operator and the probability measure

$$\overline{T}^n = \frac{1}{n} \sum_{\nu=1}^{n} T^\nu, \qquad \overline{P}^n = \frac{1}{n} \sum_{\nu=1}^{n} P^{\nu *} \tag{2.6}$$

There is a basic result due to Rosenblatt [20].

THEOREM 1. *The averaged sequence of probability measures $\overline{P}^n$ converges to a limit measure $\overline{P}$ such that*

$$\overline{P} * P = P * \overline{P} = \overline{P} * \overline{P} = \overline{P}. \tag{2.7}$$

This sort of result is known in the study of Markov processes on compact spaces; the proof can be based on Yosida's mean ergodic theorem and the observation that all sequences of $\overline{P}^n$ are compact.

If a limit law exists it is not difficult to see that its support must be an ideal. We have from [20] the

THEOREM 2. *If a limit law exists its support must be contained in the kernel K of the semigroup; K is the minimal ideal of S and can be written as*

$$K = \bigcap_i SiS \tag{2.8}$$

where i runs through all the idempotents of S.

For the limit distribution $\overline{P}$ it can actually be shown that its support is exactly K.

Of course we can get compactness even when it is not assumed a priori by the usual one point compactification: we add the point at infinity w, with the appropriate neighborhoods and with the multiplication rule $ws = sw = w$ for all $s \in S$. In this way the above results can be extended in an easy but not very profound manner.

Assume further that S is commutative. Then the kernel K is a commutative group and we can apply the results to be discussed in section 3.2. The support of any idempotent must be contained in K, and if it is just K the idempotent is simply the (normalized) Haar measure on K.

2.2. *Finite commutative semigroups.* The next logical step in specialization is to assume that S is a finite and commutative semigroup. Following Hewitt and Zuckerman [9] we shall study this case by Fourier analysis. We assume there exists an integer $m > 1$ such that $s^{m+1} = s$ for all $s \in S$.

A complex-valued function $\chi(s) \not\equiv 0$ is called a *semicharacter* if $\chi(s)\chi(t) = \chi(st)$ for all $s, t \in S$. The set of all semicharacters is denoted by $\hat{S}$.

Since $s^{m+1} = s$ we have $\chi^{m+1}(s) = \chi(s)$ so that $\chi(s)$ is 0 or a root of unity; $|\chi(s)| \leqq 1$.

Any probability distribution $P \in \mathcal{P}(S)$ can be completely characterized by the point masses $P(s)$ and we shall define its Fourier transform

$$\hat{P}(\chi) = \sum_{x \in S} P(x)\chi(x), \qquad \chi \in \hat{S}. \tag{2.9}$$

We have from [9]

THEOREM 3. (a) $|\hat{P}(\chi)| \leqq 1$,

(b) *if* $P = P_1 * P_2$ *then* $\hat{P}(\chi) = \hat{P}_1(\chi)\hat{P}_2(\chi)$ *for every* $\chi \in \hat{S}$,

(c) $\hat{P}(\chi)$ *determines* P *uniquely and there is an inversion formula,*

$$P(s) = \sum_{\chi \in \hat{S}} \hat{P}(\chi)\chi(s) \tag{2.10}$$

(d) *if* $P_n \to P$ *then* $\hat{P}_n(\chi) \to \hat{P}(\chi)$ *and conversely. Further if* S *is also a group and* Q *is the probability measure* $Q(s) = 1/n$, *where* n *is the order of the group, then*

$$\hat{Q}(\chi) = \begin{cases} 1 & \text{if} \quad \chi = 1, \\ 0 & \text{otherwise}. \end{cases} \tag{2.11}$$

There is also a version [9] of the theorems of Herglotz and Bochner relating the Fourier transforms of probability distributions to nonnegative definite functions on $\hat{S}$,

THEOREM 4. *A complex-valued function* $p(\chi)$ *on* $\hat{S}$ *can be written* $p(\chi) = \hat{P}(\chi)$, $P \in \mathcal{P}(S)$ *if and only if*

$$P(e) = 1, \tag{2.12}$$

$$\sum_{\chi,\psi \in \hat{S}} p(\chi\bar{\psi})C(\chi)\overline{C(\psi)} \geqq 0 \tag{2.13}$$

for any complex-valued function $C(\chi)$ *on* $\hat{S}$.

The idempotents can now be characterized completely by

THEOREM 5. *A probability measure* $P \in \mathcal{P}(S)$ *is an idempotent if and only if* $s(P)$ *is a subgroup of* S *and* P *is uniformly distributed over this subgroup:* $P(s) = P(t)$ *for* $s, t \in s(P)$.

These authors also give necessary and sufficient conditions on a distribution P over a finite commutative group (with an integer m as defined above) in order that P^{n*} should converge. Since these conditions are fairly complicated the reader is referred to [9] for the complete statement.

3. Groups

When S has also group structure we will denote it by G and for the moment we assume only that G is a *locally compact* topological group. This will not be enough to enable us to draw the sort of conclusions we want; in fact here seems to be, at present, the most serious gap in the theory that we examine.

However, we are not entirely ignorant even in this general situation. There is, for example, the well-known result on the existence of a left (and similarly right) *invariant measure* μ,

(a) μ is a regular Borel measure,

(b) $\mu(O) > 0$ for any open set O,

(c) $\mu(xE) = \mu(E)$ for every Borel set E and every $x \in G$. This measure, which is uniquely determined except for a multiplicative constant, is called the *Haar measure*. The right invariant measure will be denoted by ν. A good reference is [8].

The Haar measure can be normalized to a probability measure, $\mu(G) = 1$, only if it is finite. This happens if and only if G is compact.

It is clear that the set $\mathcal{P}(G)$ of regular and normalized Borel measures on G forms a semigroup but it should be noted that it is not a group unless G consists of the single element e. Actually an element $P \in \mathcal{P}(G)$ has an inverse if and only if P is concentrated on a single element.

A measure μ is said to be *symmetric* if $\mu(E) = \mu(E^{-1})$ for every Borel set E.

It is called *absolutely continuous* if it is absolutely continuous with respect to Haar measure.

By a *unitary representation* $U_H = \{\mathcal{H}, U(g)\}$ of the group G one understands a Hilbert space $\mathcal{H}$ and a strongly continuous function $U(g)$ taking as values unitary transformations of $\mathcal{H}$ onto itself and such that $U(g_1)U(g_2) = U(g_1g_2)$ for all $g_1, g_2 \in G$. Then, for any $x \in \mathcal{H}$, the function $\varphi(g) = [x, U(g)x]$ is *nonnegative definite*, that is

$$\sum_{\nu,\mu=1}^{n} \varphi(g_\nu^{-1}g_\mu)\bar{z}_\nu z_\mu \geqq 0 \tag{3.1}$$

for any n and arbitrary complex numbers $z_1, z_2, \cdots, z_n$. Conversely every continuous nonnegative definite function $\varphi(g)$ can be represented in the form $[x, U(g)x]$ with a suitable choice of the representation and of the element x.

The natural definition of the Fourier transform of $P \in \mathcal{P}(G)$ would then be

$$\hat{P}(U_H) = \int_G U(g)\, dP\,(G) \tag{3.2}$$

and $\hat{P}(I) = I$ (where I is the identity operation in $\mathcal{H}$) and

$$\widehat{P_1 * P_2}(U_H) = \hat{P}_1(U_H)\hat{P}_2(U_H). \tag{3.3}$$

So far this tool has been used extensively in probability theory only for commutative or compact groups.

The operator T can now be defined on the Hilbert space $L_2(G, \mu)$ of complex-valued functions on G quadratically integrable with respect to μ-measure. Assuming P to be symmetric the spectrum of T is situated in the real interval $(-1, 1)$. Conditions can be given for $|\lambda| = 1$ to belong to the spectrum; see [7] and, for the case of a denumerable group, [14].

3.1. *Compact groups.* With this assumption it is possible to get more complete and detailed results. The idempotents of $\mathcal{P}(G)$ can now be characterized completely [28]: they are the normalized Haar measures on compact subgroups of G (or rather the extension of such measures to G). Note that on a compact group the right and left invariant measures are essentially equal.

To obtain limit theorems we study the Fourier transforms $\hat{P}$. But now we need only consider the set of irreducible and nonequivalent finite dimensional representations enumerated $U_0, U_1, U_2, \cdots$, where U_0 is the identity representation $U_0(g) \equiv 1$, see [18], [27]. Using the famous theorem of Peter and Weyl one can show that $\hat{\mu}(U)$ is equal to 1 for $U = U_0$ and to 0 otherwise. The mapping $P \to \hat{P}$ is a continuous homomorphism of $P(G)$ into the set of all complex-valued $n \times n$ matrices, where n is the dimension of the representation U; the uniqueness and continuity theorem for ordinary characteristic functions in R^k is extended to the present case, enabling us to get the desired results, due essentially to Itô and Kawada, who obtained this pioneering result as early as 1940; the modified form given below is due to K. Stromberg [23].

A special case, the circle group, had been investigated earlier by Lévy [16]; see also [13].

THEOREM 6. *Let K be the smallest closed subgroup containing $s(P)$, where $P \in \mathcal{P}(G)$. The limit of P^{n*} exists if and only if $s(P)$ is not contained in any coset of any proper closed normal subgroup of K. If the limit exists it is the normalized Haar measure on K.*

The following version, see [24], is also useful.

THEOREM 7. *The limit of P^{n*} exists if and only if K is the smallest closed subgroup containing $\bigcup_{n=1}^{\infty} \{s(P)^n[s(P)^{-1}]^n\}$.*

COROLLARY. *If $e \in s(P)$ then the limit of P^{n*} exists.*

If there is convergence it is monotone.

Introduce the deviation d_n, between P^{n*} and a measure μ such that $\mu * P = \mu$, by means of the definition

$$d_n = d\,(P^{n*}, \mu) = \sup_E |P^{n*}(E) - \mu(E)|. \tag{3.4}$$

Then we have

THEOREM 8. $d_{n+1} \leqq d_n$.

3.1.1. *Compact, commutative groups.* A probability distribution P is called a *Poisson distribution* over G if

$$P(E) = P(x_0^k \in E) \quad \text{for all} \quad E \in \mathcal{B}(G), \tag{3.5}$$

where $x_0 \in G$ and k is an ordinary Poisson variable.

More generally if P can be written as

$$P = \sum_{\nu=0}^{\infty} \frac{Q^{\nu *}}{\nu!} e^{-Q(G)}, \tag{3.6}$$

where Q is a regular finite measure defined on $\mathcal{B}(G)$, then P is said to be a *compound Poisson distribution.*

The following result has been obtained by Urbanik [25].

THEOREM 9. *P is a compound Poisson distribution if and only if there is a sequence $P_n \in P(G)$ with*

$$P = P_n^{n*} \tag{3.7}$$

$$\lim_{n\to\infty} P_n(e) = 1. \tag{3.8}$$

Urbanik also gives a similar condition for P to be a Poisson distribution.

3.1.2. *Finite groups.* The idempotents are immediately obtained from Wendel's theorem (see 3.1); $P \in P(G)$ is an idempotent if and only if there is a subgroup H, say of order h, and $P(g) = h^{-1}$ if $g \in H$ and $= 0$ otherwise (see also [3]).

The class of *infinitely divisible distributions on a finite group* have been described exhaustively by Böge [3] through the following representation

THEOREM 10. *Let H be an arbitrary subgroup and I_H the corresponding idempotent measure. Form the set K_H of H-invariant real-valued measures m over G and such that*

$$m(g) \geqq 0, \qquad \text{if} \quad g \in H \tag{3.9}$$

$$\sum_{g \in G} m(g) = 0. \tag{3.10}$$

Then

$$P = I_H + \sum_{\nu=1}^{\infty} \frac{1}{\nu!} m^{\nu*}, \qquad m \in K_H \tag{3.11}$$

is the general representation of an infinitely divisible distribution over G.

The representation is unique only in the very special case that all elements of G are of order 2.

Böge also points out that the set of infinitely divisible distributions is closed under convolutions if and only if G is commutative.

3.2. *Commutative and locally compact groups.* The Fourier transform $\hat{P}(\gamma)$ of a distribution on G is now defined as

$$\hat{P}(\gamma) = \int_G (g, \gamma)\, dP\,(g), \tag{3.12}$$

where γ is an arbitrary character, that is, a complex-valued, continuous function on G taking values (g, γ) such that $|(g, \gamma)| = 1$ and $(g_1 + g_2, \gamma) = (g_1, \gamma)(g_2, \gamma)$, where we now write the group operation as addition. The γ form a set Γ, the dual group of G, if addition of characters is defined by $(g, \gamma_1 + \gamma_2) = (g, \gamma_1)(g, \gamma_2)$. For further information, see [27]. There is an extension of Bochner's theorem on characteristic functions,

THEOREM 11. *The class of all functions $\hat{P}(\gamma)$, when P runs through $P(G)$, is identical with the class of continuous, nonnegative definite functions normed to be equal to unity for $\gamma = 0$.*

If P is an idempotent, then $\hat{P}(\gamma) = 1$ or 0; in other words one has to study those sets of Γ whose indicator functions are Fourier transforms of probability distributions.

The idempotents $P \in P(G)$ are described by the following

THEOREM 12. *Any idempotent measure P on a commutative and locally compact group is concentrated on a compact subgroup. P is then the normalized Haar measure on this subgroup.*

This result also tells us something about what sort of limit we can expect for this type of groups.

3.3. *Lie groups.* For a group that is neither commutative nor compact (but is locally compact) very little seems to be known about such things as limit laws. The reason for this is obviously that the sort of Fourier transform defined at the beginning of section 3, if it is at all the appropriate analytical tool, is not so easily mastered as for compact or commutative groups.

However, for Lie groups Hunt [10] has given a characterization, complete in a certain sense, of the associated homogeneous processes; this generalizes the well-known representation due to Khinchin, Lévy, and others of the processes with independent increments on the real line.

To sketch the main result in Hunt's paper, let G be a Lie group of dimension d and with a homogeneous process g_t with the probability measures P_t and the corresponding operator T_t (see section 1). To study the infinitesimal generator M of the semigroup T_t, defined for $t > 0$, we consider the Banach space $\overline{C}$ of functions $f(g)$ which are bounded and left uniformly continuous on G; the usual definition of norm is used. If Y is an element of the left invariant Lie algebra we define

$$Yf(g) = \lim_{t \downarrow 0} \frac{1}{t} [f(g \exp tY) - f(g)] \tag{3.13}$$

for those $f \in \overline{C}$ for which the limit exists. Introduce the Banach space $\overline{C}_2$ as the set of those f for which $Y_1Y_2 f$ makes sense for any choice of Y_1, Y_2 in the Lie algebra and with the definition of the norm

$$||f||_2 = ||f|| + \sum_i ||X_if|| + \sum_{i,j} ||X_iX_jf||, \tag{3.14}$$

where $X_1, X_2, \cdots, X_d$ is a fixed basis of the algebra. Then M is defined on $\overline{C}_2$ and can be written as

$$\begin{aligned} Mf(g) = \sum_i a_iX_if(g) + \sum_{i,j} a_{ij}X_iX_jf(g) \\ + \int \Big[f(gh) - f(g) - \sum_i X_if(g)x_i(h)\Big]\, dG\,(h). \end{aligned} \tag{3.15}$$

Here a_{ij} is a symmetric and nonnegative definite matrix, the integral is extended over the set $G - e$, the $x_i(g)$ are functions in $\overline{C}_2$ such that $x_i(e) = 0$ and $X_ix_j(e) = \delta_{ij}$, and dG is a measure over $G - e$ such that $\int \phi(g)\, dG\,(g) < \infty$, where $\phi(g)$ is a nonnegative function in $\overline{C}_2$ behaving like $\sum x_i^2(g)$ when g is near e. The converse statement is also true.

This representation can be decomposed as $M = M_1 + M_2$, where M_1 consists of the two sums and M_2 is the integral. If the sample space of the process is suitably chosen, almost all the sample functions are continuous if M_2 vanishes; conversely such a sample space can be found only if $M_2 = 0$.

If $M = M_1$ the process is called a *Brownian motion* on the Lie group in analogy with the definition for $G = R^k$. It can also be obtained as the solution of a stochastic differential equation of a form similar to M_1 (see [12]).

4. Linear spaces

The reader will have noticed that the sort of limit results studied above are not quite of the same type as those belonging to classical probability theory. We have discussed no natural extension of the law of large numbers or of the central limit theorem, we have no stable laws and, more generally, we lack a generalization of the linear theory of stochastic variables that plays such a fundamental role in the classical theory. To remedy this we must obviously introduce a linear structure in our group. This can be done in various ways, but so far the Banach and Hilbert spaces seem to be the only ones investigated thoroughly as domains of probability distributions, and we will confine our discussion to these two important linear structures.

Such probabilistic structures were first studied systematically by E. Mourier in an important paper [17]. In a more general direction, Fréchet has studied probabilities in metric spaces and his work should be consulted by the interested reader.

4.1. *Banach spaces.* Let X be a Banach space (real or complex as the case may be) with elements x and the norm $||x||$. A regular probability P is given on the Borel sets of X; this corresponds to a stochastic variable taking values on X. Linear combinations of given stochastic variables $\xi_1, \xi_2, \cdots$ as well as limits of such sequences are then also well-defined stochastic variables on X.

In this way all linear functionals $x^*(x)$ become Borel measurable. For the actual construction of probability measures it may be more convenient to start with given finite dimensional simultaneous distributions of $x_1^*(x), x_2^*(x), \cdots, x_n^*(x)$. In other words we start with the set algebra generated by cylinder sets of the form

$$\{x|[x_1^*(x), x_2^*(x), \cdots, x_n^*(x)] \in E\}, \tag{4.1}$$

where E is an arbitrary Borel set in R^n, and n an arbitrary positive integer. Applying Kolmogorov's extension theorem we extend the probability measure to the σ-algebra. If X is separable, it can be shown that this σ-algebra includes all open sets, so that the domain of the measure is wide enough for us.

Following E. Mourier [17] we define the *mean value* $E\xi = m$ as a Pettis integral: it is the unique element m, if it exists, satisfying the integral equation

$$x^*(m) = Ex^*(\xi) \tag{4.2}$$

for all linear functionals x^* in the conjugate space X^*. Note that $x^*(\xi)$ is a numerically valued stochastic variable in the ordinary sense.

With this definition the mean value operation has the usual properties:
(a) if $E\xi_1$ and $E\xi_2$ exist then so does

$$E(C_1\xi_1 + C_2\xi_2) = C_1E\xi_1 + C_2E\xi_2, \tag{4.3}$$

(b) if $E||\xi|| < \infty$ and $E\xi$ exists then $||E\xi|| \leqq E||\xi||$.

There are various sufficient conditions ensuring the existence of $E(\xi)$. The following is very useful: if X is separable and if $E||\xi|| < \infty$ then $E\xi$ exists, [17].

The Fourier transform of a stochastic variable ξ or of its probability distribution P is now defined as

$$\hat{P}(x^*) = E \exp ix^*(x), \qquad x^* \in X^*. \tag{4.4}$$

We have
(a) $\hat{P}(x^*) \equiv 1$ if $P\{x = 0\} = 1$,
(b) $\widehat{P_1 * P_2}(x^*) = \hat{P}_1(x^*)\hat{P}_2(x^*)$,
(c) $\hat{P}(x^*)$ is continuous in the weak topology of X^*,
(d) $\hat{P}(x^*)$ is nonnegative definite, that is

$$\sum_{\nu,\mu=1}^{n} P(x_\nu^* - x_\mu^*)z_\nu\bar{z}_\mu \geqq 0 \tag{4.5}$$

for any n and complex constants $z_1, z_2, \cdots, z_n$.

For a discussion of the uniqueness and continuity theorems, see [17]. Note that the space will in general not be locally compact.

The *law of large numbers* can now be formulated in many ways, one of which is given in [17],

THEOREM 13. *Let X be separable. If $\xi_1, \xi_2, \cdots$ is a sequence of independent and identically distributed stochastic variables with values in X and if $E||\xi_\nu|| < \infty$, then the average*

$$\eta_n = \frac{1}{n}\sum_{\nu=1}^{n} \xi_\nu \tag{4.6}$$

converges almost certainly strongly toward the element $E\xi$.

We remark that the related but more general ergodic theorem also holds under similar conditions; the limit is of course no longer necessarily a constant element of the space.

4.2. *Hilbert spaces.* Let X be a separable Hilbert space. The existence of an inner product makes possible a more detailed analysis.

A stochastic variable ξ taking values in X is said to have a *normal distribution*, if the numerically valued stochastic variables $x^*(\xi)$ for $x^* \in X^*$, are all normal. Actually this definition can be used also in a Banach space but in the present context it is especially simple since the linear functionals are easier to describe: $x^*(\xi) = (x^*, \xi)$, $X^* = X$. One can show that $E\xi$ and $E||\xi||^2$ exist and that the Fourier transform of the distribution P associated with ξ is

$$\hat{P}(x^*) = \exp\left\{iEx^*(\xi) - \frac{1}{2}\operatorname{Var}[x^*(\xi)]\right\}. \tag{4.7}$$

Conversely any such expression is the Fourier transform of a normal stochastic variable if ξ is a stochastic variable in the Hilbert space with finite $E||\xi||^2$. It is also possible to find an orthogonal system $x_n \in X$ for $n = 1, 2, \cdots$; such that

$$\xi = \sum_{n=1}^{\infty} C_n x_n, \tag{4.8}$$

where the coordinates C_n are independent normal stochastic variables.

The following central limit theorem holds [5]:

THEOREM 14. *Let $\xi_1, \xi_2, \cdots$ be independent variables taking values in X and identically distributed with $E\xi = 0$ and $E||\xi||^2 < \infty$. Consider the normed partial sums $\eta_n = n^{-1/2} \sum_1^n \xi_\nu$. If $f(x)$ is a function uniformly continuous in every sphere $||x|| < R$, then $f(\eta_n)$ has asymptotically the same distribution as $f(\eta)$, where η is a normal stochastic variable in X with the Fourier transform*

$$\hat{P}_\eta(x^*) = \exp\left\{-\frac{1}{2} E[x^*(\eta)]^2\right\}. \tag{4.9}$$

5. Banach algebras

If our stochastic variables take values in an algebra A, say, with unit element e, then it is possible to proceed further. We now have two operations, addition and multiplication, and to each there corresponds one definition of a homogeneous process, of limit theorems and so on. The probabilistic relation between the additive and multiplicative concepts should be investigated. Since addition is commutative the corresponding concept could be the simpler one, so that one may wish to start with it to construct or study the other one.

Let Z be a given space and A an algebra of operators defined on Z. Given a probability measure on A we have to deal with *stochastic operators*. Already the case when $Z = R^k$ and A is an algebra of linear operators leads to nontrivial problems, the solutions of which are only partially known at present.

A multiplicative version of the law of large numbers is the following. Let A be a separable Banach algebra and consider a sequence $\eta_1, \eta_2, \cdots$ of independent and identically distributed stochastic variables with values in A and such that $E||\eta|| < \infty$. Form the product

$$\Pi_n = \left(e + \frac{1}{n}\eta_1\right)\left(e + \frac{1}{n}\eta_2\right) \cdots \left(e + \frac{1}{n}\eta_n\right), \tag{5.1}$$

and consider the behavior of Π_n as n tends to infinity.

THEOREM 15. *The above product converges strongly in probability to the fixed element* $\exp(m) \in A$, *where $m = E\eta$, as n tends to infinity.*

This is proved using the (additive) law of large numbers (see section 4.1 and [7]).

Now, let $\eta(t)$ for $t \geqq 0$, be an additive homogeneous process with values in A. It seems plausible that one could construct a multiplicative homogeneous process by starting from products

$$\xi_n(t) = (e + \Delta_1 \eta)(e + \Delta_2 \eta) \cdots (e + \Delta_n \eta), \tag{5.2}$$

where

$$\Delta_\nu \eta = \eta(t_\nu) - \eta(t_{\nu-1}) \quad \text{and} \quad t_0 = 0 < t_1 < t_2 < \cdots < t_n = t, \tag{5.3}$$

is a division of the interval $(0, t)$. Assuming that $E||\eta(t)||$ exists for $t \geqq 0$ and that the sums

$$\sum E||\Delta_\nu \eta|| \leqq M < \infty \tag{5.4}$$

are uniformly bounded, then $\xi_n(t)$ converges in the L_1-norm to a multiplicative process. Symbolically we can write $d\xi(t) = \xi(t)\, d\eta(t)$ with the initial condition $\xi(0) = e$.

It is now possible to prove multiplicative limit theorems. Indeed, consider the double array of stochastic variables taking values in A

$$\begin{matrix} \eta_{11} \\ \eta_{21}, \eta_{22} \\ \eta_{31}, \eta_{32}, \eta_{33} \\ \cdots\cdots \end{matrix} \tag{5.5}$$

where the variables in each row are independent and identically distributed. Assume that, uniformly,

$$\sum_{\nu=1}^{n} E||\eta_{n\nu}|| \leqq M < \infty, \tag{5.6}$$

and that $\eta(t)$ is an additive homogeneous process such that for every fraction c between 0 and 1 we have

$$\sum_{\nu=1}^{[cn]} \eta_{n\nu} \to y(c) \tag{5.7}$$

distributionwise.

THEOREM 16. *Under the stated conditions the product*

$$\Pi_n = (e + \eta_{n1})(e + \eta_{n2}) \cdots (e + \eta_{nn}) \tag{5.8}$$

converges distributionwise to $\xi(1)$.

From the point of view of applications we get a more interesting sort of limit theorem using the L_2-topology instead of the L_1-topology; we refer the reader to [7] for a more detailed statement.

In this connection one should mention a related but different limit problem arising when A is an *algebra of operators* mapping a space Z into itself. These operators need not be linear; actually the most interesting situations seem to appear when this is not the case. Consider a sequence of independent and identically distributed stochastic operators $\cdots, \alpha_{n-1}, \alpha_n, \alpha_{n+1}, \cdots$ and form the product $\beta_n^{(m)} = \alpha_n \alpha_{n-1} \cdots \alpha_{n-m}$. Can it happen that $\beta_n^{(m)}$ converges to a nontrivial stochastic operator β_n as m tends to infinity?

If this is so consider the stochastic element (in Z) $Z_n = \beta_n z$ where $z \in Z$. It must have the same distribution as $\alpha_{n+1} z_n$ if an equilibrium distribution exists. It is clear that with a compactness assumption on Z equilibrium distributions exist, but this can also be guaranteed in noncompact situations if suitable conditions are imposed on the operators belonging to the algebra.

While the general case does not seem to have been studied it is easy to deal with certain special cases. Let us for example choose Z as the real line or a part of it, and define the mappings by a function $h(z, w)$, where w introduces the randomness via a probability distribution P over a set Ω of points w. The needed measure-theoretic assumptions are made as usual. Suppose that

$$\sup_{Z',Z''} \left| \frac{h(Z'', w) - h(Z', w)}{Z'' - Z'} \right| = \rho < 1, \tag{5.9}$$

and $Eh^2(Z_0, w) < \infty$.

THEOREM 17. *In the situation described above convergence holds and an equilibrium distribution exists.*

The above theorem generalizes easily to the case when Z is a linear vector case.

Starting with this equilibrium distribution for the elements z, we can define a stationary, real-valued stochastic process $y_n = \alpha_n \alpha_{n-1} \cdots \alpha_1 z$. Taking $h(z, w)$ as a linear function $c(w) + z\rho$, we get a moving average representation, taking it as a fractional linear function we get a continued fraction representation. Other choices of $h(z, w)$ will be of interest in the study of nonlinear random mechanisms.

A problem that has been given a good deal of attention is the *random ergodic theorem*. It can be phrased as follows, leaving out the measure-theoretic details. Consider two measure spaces Z and Ω with finite, normed measures m and P respectively, $m(Z) = P(\Omega) = 1$. Let φ_ω, with $\omega \in \Omega$, be a family of m-measure preserving transformations defined on Z. Form the iterated stochastic transformation

$$\Phi_n = \varphi_{\omega_n} \cdots \varphi_{\omega_2} \varphi_{\omega_1}, \tag{5.10}$$

where the elements $\omega_1, \omega_2, \cdots$ are independent stochastic elements on Ω with the probability distribution P. Let $f(z)$ belong to $L_1(z)$.

THEOREM 18 (Kakutani). *The average*

$$\frac{1}{n} \sum_{\nu=1}^{n} f[\Phi_n z] \tag{5.11}$$

converges almost certainly (P) *for almost all* (m) *values of* z.

For other results on stochastic operators see [2] and papers by Hanš and Špaček referred to in [2].

REFERENCES

[1] R. BELLMAN, "Limit theorems for non-commutative operations. I," *Duke Math. J.*, Vol. 21 (1954), pp. 491–500.

[2] A. T. BHARUCHA-REID, "On random solutions of Fredholm integral equations," *Bull. Amer. Math. Soc.*, Vol. 66 (1960), pp. 104–109.

[3] W. Böge, "Über die Charakterisierung unendlich teilbarer Wahrscheinlichkeitsverteilungen," *J. Reine Angew. Math.*, Vol. 201 (1959), pp. 150–156.

[4] A. Dvoretzky and J. Wolfowitz, "Sums of random integers reduced modulo *m*," *Duke Math. J.*, Vol. 18 (1951), pp. 501–507.

[5] R. Fortet and E. Mourier, "Les fonctions aléatoires comme éléments aléatoires dans les espaces de Banach," *Studia Math.*, Vol. 15 (1955), pp. 62–79.

[6] I. Glicksberg, "Convolution semigroups of measures," *Pacific J. Math.*, Vol. 9 (1959), pp. 51–67.

[7] U. Grenander, "Stochastic groups," three papers to appear in *Ark. Math.*

[8] P. R. Halmos, *Measure Theory*, New York, Van Nostrand, 1950.

[9] E. Hewitt and H. S. Zuckerman, "Arithmetic and limit theorems for a class of random variables," *Duke Math. J.*, Vol. 22 (1955), pp. 595–615.

[10] G. A. Hunt, "Semi-groups of measures on Lie groups," *Trans. Amer. Math. Soc.*, Vol. 81 (1956), pp. 264–293.

[11] K. It and Y. Kawada, "On the probability distribution on a compact group," *Proc. Phys.-Math. Soc. Japan*, Vol. 22 (1940), pp. 977–988.

[12] K. Itô, "Stochastic differential equations in a differentiable manifold," *Nagoya Math. J.*, Vol. 1 (1950), pp. 35–47.

[13] T. Kakehashi, "Stationary periodic distributions," *J. Osaka Inst. Sci. Tech.*, Vol. 1 (1949), pp. 21–25.

[14] H. Kesten, "Symmetric random walk on groups," *Trans. Amer. Math. Soc.*, Vol. 92 (1959), pp. 336–353.

[15] B. M. Kloss, "Limiting distributions of sums of independent random variables taking values from a bicompact group," *Dokl. Akad. Nauk SSSR*, Vol. 109 (1956), pp. 453–455. (In Russian.)

[16] P. Lévy, "L'addition des variables aléatoires définies sur une circonférence," *Bull. Soc. Math. France*, Vol. 67 (1939), pp. 1–41.

[17] E. Mourier, "Éléments aléatoires dans un espace de Banach," *Ann. Inst. H. Poincaré*, Vol. 13 (1953), pp. 161–244.

[18] L. Pontrjagin, *Topological Groups*, Princeton, Princeton University Press, 1946.

[19] A. Prékopa, A. Rényi, and K. Urbanik, "On the limiting distribution of sums of independent random variables in bicompact commutative topological groups," *Acta Math. Acad. Sci. Hungar.*, Vol. 7 (1956), pp. 11–16. (In Russian, with English summary.)

[20] M. Rosenblatt, "Limits of convolution sequences of measures on a compact topological semigroup," *J. Math. Mech.*, Vol. 9 (1960), pp. 293–305.

[21] W. Rudin, "Measure algebras on Abelian groups," *Bull. Amer. Math. Soc.*, Vol. 65 (1959), pp. 227–247.

[22] ———, "Idempotent measures on Abelian groups," *Pacific J. Math.*, Vol. 9 (1959), pp. 195–209.

[23] K. Stromberg, "Probabilities on a compact group," *Trans. Amer. Math. Soc.*, Vol. 94 (1960), pp. 295–309.

[24] K. Urbanik, "On the limiting probability distribution on a compact topological group," *Fund. Math.*, Vol. 44 (1957), pp. 253–261.

[25] ———, "Poisson distributions on compact Abelian topological groups," *Colloq. Math.*, Vol. 6 (1958), pp. 13–24.

[26] N. N. Vorobev, "Addition of independent random variables on finite Abelian groups," *Mat. Sb.*, Vol. 34 (1954), pp. 89–126. (In Russian.)

[27] A. Weil, *L'Intégration dans les Groupes Topologiques et ses Applications*, Paris, Hermann, 1940.

[28] J. G. Wendel, "Haar measure and the semigroup of measures on a compact group," *Proc. Amer. Math. Soc.*, Vol. 5 (1954), pp. 923–929.

[29] J. W. Woll, Jr., "Homogeneous stochastic processes," *Pacific J. Math.*, Vol. 9 (1959), pp. 293–325.