Stochastic groups

By ULF GRENANDER

With 1 figure in text

Introduction

Like other branches of mathematics probability theory of to-day is becoming more general and abstract, one tries to extend the theory as far as possible and through this process one hopes that the fundamental conceptual structure will become clearer. The road to such a development was opened some 25 years ago through Kolmogorov's axiomatic approach to probability theory considered as measure theory in a general reference space. This made is possible to create a mathematically satisfactory theory for stochastic processes choosing the reference space as an appropriate function space or as a labeling of such a space.

More recently attention has been drawn to other extensions of classical probability theory. In the work of Mourier, Fortet, and others the reference space is chosen, not as the real line, the plane or higher dimensional Euclidean spaces, but as the mathematical structures that come next in the natural order of generality, viz. the Hilbert and Banach spaces. Although this is not unrelated to earlier work in stochastic processes it marks a new direction in research because of its emphasis on the abstract character of the reference space. It is relevant to the subject of the present paper and we must discuss it at least briefly below.

The addition defined in a Banach space is commutative just as in finite dimensional vector spaces. One asks naturally what happens if this postulate is not satisfied: what can be said about probability distributions on non-commutative groups? It is interesting to observe that this question was raised and to some extent answered as long ago as 1941 by Ito and Kawada. These authors published their work in a Japanese journal during the war and their paper has not been given the attention it deserves. Recently a number of authors, apparently independent of each other, have rediscovered some of these results. We must take a brief look at these things below, since they will be of great value indicating the direction for future work in this field.

In the literature one can find a few papers dealing with others structures, e g. Riemannian manifolds, Boolean algebras, semi-groups and Lie groups. The reason why we choose to deal mainly with groups and algebras in this paper is that even with this limitation we will have such a formidable task before us with so many open and perhaps difficult problems that the subject could not possibly be treated in a definitive way in a single publication. This will be

more evident as we go along. In the future we can expect a more sophisticated and general approach, perhaps embracing some of the above structures in a unified theory. Anyway, the time seems to be ripe now for a systematic discussion of *stochastic groups* which will be attempted in this study, of which this paper contains the three first parts, and of which further parts are planned to appear in this journal.

The main stimulus for the author to study the stochastic groups has been certain problems in applied probability, some of which will be sketched in Part 1. Some have been found in the applied literature, especially in the physics journals. Others have been mentioned to the author in discussions or have appeared in his consultation work. Examination of such problems will help us with the preliminary work that has to be done, when trying to build a theory for stochastic groups. It will guide us to the essential problems and concepts of such a theory, how to choose the terminology and how to avoid trivial or inadequate generalizations.

After this examination Part 2 will contain definitions and a fairly general discussion while later parts will be denoted to a more detailed study of stochastic groups, algebras and similar structures.

Some of the topics of this paper were mentioned in an earlier paper, Grenander (1959), but only in a discursive way.

Part 1. Background

1.1. Let us start with the simplest case, the real line R_1, on which probability distributions are given, corresponding to independent stochastic variables $x_1, x_2, \ldots$. Considering R_1 as a group, with addition as the group operation, we can form the sum $x_1 + x_2$ which is a new stochastic variable, whose distribution is the convolution of the ones associated with x_1 and x_2. Similarly, the sum $S_n = x_1 + x_2 + \cdots + x_n$ has a distribution that is fully known to us, at least in principle. Much of modern probability theory tells us how S_n or its distribution behaves when n becomes large.

We know e.g. that the average S_n/n tends to the mean value m if all x_v have the same distribution with an existing expected value m. Here "tends to" can be understood as convergence in probability (the weak law of large numbers) or as almost certain convergence (the strong law).

Under certain conditions we know that the sum, appropriately normed

$$z_n = \frac{S_n - a_n}{b_n},$$

converges distributionwise to the normal distribution. More generally, if we consider a double array of independent stochastic variables

$$x_1^{(1)}$$

$$x_1^{(2)}, \ x_2^{(2)}$$

$$x_1^{(3)}, \ x_2^{(3)}, \ x_3^{(3)}$$

$$\cdots\cdots$$

and put $S_n = x_1^{(n)} + x_2^{(n)} + \cdots + x_n^{(n)}$, then z_n, with some norming, behaves asymptotically in a way that is fairly well known to us to-day. This leads us to the infinitely divisible distributions, which are, vaguely speaking, the distributions that can be represented as a sum of independent and arbitrarily small (in probability) stochastic variables.

To obtain these results many mathematical techniques have been used, the most important of which is probably the Fourier transform, the characteristic function $\varphi(z) = E \exp izx$. The importance of the characteristic function is due to its three fundamental properties

1: the correspondence between distribution and characteristic function is $1-1$.
2: the correspondence is continuous with respect to simple and natural topologies.
3: to the group operation $x_1 + x_2$ corresponds simply multiplication $\varphi_1 \cdot \varphi_2$.

1.2. Passing to n-space we have very much the same situation although of course the higher dimension makes the derivations more cumbersome, and the subject has not been covered quite as thoroughly as for $n = 1$. The characteristic function is now defined as $\varphi(z) = E \exp i(x, z)$, where (x, z) stands for the inner product of the stochastic vector x and the argument vector z. The three fundamental properties of the last section still apply, and the reason why is as obvious for arbitrary n as for $n = 1$. Indeed, the reference space here is a commutative, locally compact group with certain *characters* $\chi(x)$, and $\varphi(z) = \exp \chi(x)$. The defining property of the characters together with their completeness is, in the last analysis, what makes the characteristic function such a helpful instrument.

For related and more general questions of this type the reader is referred to Bochner's recent book, see list of references.

1.3. In 1934 Khinchin showed in his now classical research memoir "Asymptotische Gesetze der Wahrscheinlichkeitsrechnung" that the limit theorems of probability theory are in close logical relation to certain stochastic processes. A typical example is the following. Consider a sum S_n as before with

$$\left.\begin{aligned} E\,x_\nu &= 0 \\ E\,x_\nu^2 &= V_\nu \\ V_1 + V_2 + \cdots + V_n &= 1 \end{aligned}\right\}$$

If n is large and the V_ν small then S_n must behave almost as the value $W(1)$ of the Wiener process:

$$\left.\begin{aligned} &W(t) \text{ has independent increments} \\ &W(t'') - W(t') \text{ is normal with mean zero and standard deviation } |t'' - t'| \\ &W(0) = 0. \end{aligned}\right\}$$

In other words the distribution function of S_n is close to the normed, normal distribution function $\Phi(x)$. We will not go into any details here on how this

statement has to be supplemented to become complete, nor will we describe Khinchin's proof except remarking that it is built on the observation that the distribution function $\Phi\left(\dfrac{x}{\sqrt{t}}\right)$ satisfies the *heat equation*

$$\frac{\partial \Phi}{\partial t} = \frac{1}{2} \frac{\partial^2 \Phi}{\partial x^2}.$$

More generally one has to consider the stochastic process with independent and time homogeneous increments: the increments $x(t'') - x(t')$ should have distributions depending on t' and t'' only through their difference $t'' - t'$. The only possible form of these distributions are the infinitely divisible ones, which brings us back to the topic of the last section.

1.4. The next logical step is clearly to consider probability distributions in infinite dimensional vector spaces. Recently Mourier has shown how this can be done in a Banach, or more specially in a Hilbert space. As can be expected the measure theoretic set-up becomes more complicated than in R_n, but a good deal of the finite dimensional theory seems to hold with appropriate modifications in this more general setting.

To get a mean value operation defined on the Banach space X it is necessary to introduce integration and in Mourier's approach this is done by using the Pettis integral assuming that every *linear functional* $x^*(x)$ is measurable (if this holds we speak of an L-space) and integrable, one studies the integral equation

$$x^*(m) = E\, x^*(x).$$

If this equation has a solution m valid for all x^*, then it is unique and m is called the mean value; this mean value operation has properties that should be demanded of it, such as additivity and monotonicity.—To actually introduce a concrete probability measure on the Banach space in question one will probably usually do this indirectly by prescribing the probability distributions of all the (numerically-valued) linear functionals $x^*(x)$. Again this makes it necessary to assume X to be an L-space. The fact that the probability distribution over X is uniquely determined in this way is a direct generalization of a theorem of Cramér and Wold to the effect that the knowledge of the probability masses in every half-space of R_n determines the whole distribution.

This definition of mean value makes it possible to prove *laws of large numbers* in different versions, e.g. the following. If the dual X^* of X is separable, if x_1, x_2, ... have the same distribution over X and are independent, if $E\|X_i\| < \infty$ and $E X_i = m$ exists then

$$\frac{1}{n} \sum_{1}^{n} x_i \to m$$

weakly with probability one. This theorem is due to Mourier as is this *central limit theorem*: If $x_1, x_2, \ldots$ are identically and independently distributed over a separable Hilbert space X with

$$E\,x_i = m \; \Big\}$$
$$E\,\|\,x_i\,\|^2 = \sigma^2 \;\Big\}$$

then
$$\frac{\sum\limits_{1}^{n} x_i - n\,m}{\sigma\,\sqrt{n}}$$

is asymptotically normal over X distributionwise. Here a *distribution over X is said to be normal if this is true for every linear functional $x^*(x)$.*

This theory is only a few years old but it is already clear that it is an essential extension of the classical results. Among the many possible applications one notes the case when the stochastic elements x are the realizations of a stochastic process. One could also mention stochastic Schwartz *distributions* as a related topic.

1.5. Considered as a group the Banach space is commutative. To get an idea of what can happen for *non-commutative* groups let us consider k objects numbered $1, 2, \ldots, k$. We can permute them in $k!$ different ways and we will denote this set of permutations by P and by p an arbitrary element of P. Suppose the k objects are mixed in a random way, i.e. we have probabilities associated with each p. Instead of the original configuration we now have a new one that can be denoted $p_1 (1, 2, \ldots, k)$. Repeating the mixing n times, each time independent of the others, we get $p_n\,p_{n-1} \ldots p_2\,p_1 (1, 2, \ldots, k)$ and we may ask how this *stochastic permutation* behaves for large values of n.

Assuming for simplicity that all the p's have positive probabilities it is easy to show (e.g. considering the mixing as a Markov chain) that the distribution after many mixings converges to the uniform distribution attributing the probability $1/k!$ to every p.

In this example P forms a group, and it matters in which order the successive permutations are performed, P is non-commutative. It is remarkable that one can state a general result, very similar to the above example, for an important class of groups, commutative or not, the compact ones. This is due to Ito and Kawada, who proved this by using a beautiful extension of Fourier analysis. Since the group operations may not commute, the numerically valued representations, the characters, are not sufficient, but we must use instead the *unitary representations* $U(g)$ of the group G, $g \in G$, where $U(g)$ is a unitary matrix and

$$U(g)\,U(h) = U(g\,h); \; g, h \in G.$$

The set of irreducible non-equivalent representations can be enumerated, $U_0 = I_1$, $U_1, U_2, \ldots$ and these will now replace the characters. Instead of the characteristic function we now have the matrix valued quantity $E\,U(g)$. After n successive operations we get the quantity $[E\,U(g)]^n$; here we have just used the above relation together with the independence of the operations. But the behaviour of a high matrix power A^n depends upon the largest eigenvalues of A. Under the present circumstances it can be shown for a case, which is not quite general but typical, that the matrix $E\,U(g)$ has eigen-values of modulus less than one

except for the identity representation $U_0 = I$. Hence

$$\lim_{n \to \infty} E\,[U\,(g)]^n = \begin{cases} I & \text{if } U = I \\ 0 & \text{otherwise,} \end{cases}$$

and one can now appeal to the Peter-Weyl theorem, guaranteeing the completeness of the set of representations, to show that the limit distribution over G must satisfy

$$E\,U\,(\gamma) = \begin{cases} I & \text{if } U = I \\ 0 & \text{otherwise.} \end{cases}$$

With the conditions that we have in mind here this leads to the invariant (Haar) measure over G normed so that $P(G) = 1$. This completes the sketch of the proof.

Recently it has been shown by Rosenblatt that one can deal successfully also with the case when G is only a semi-group.

It would be natural to ask how this limit theorem is connected with stochastic processes taking values from a compact group. This does no seem to have been done and we will return to this question in a forthcoming part of this study.

1.6. Let us consider a system consisting of a large number of links in series. The links could be of a mechanical or electrical nature or they may represent industrial operations performed in a certain order. Say for simplicity that the state of a link, the nth one, can be represented by a vector s_n in R_n and that the operations are linearly related $s_{n+1} = M_n s_n$. Then M_n is an $n \times n$ matrix and

$$s_n = M_n \ldots M_2 M_1 s_1.$$

If the *transfer matrix* M_ν takes different values according to some probability distribution we have a problem resembling the one of last section. We will have to see what group G is spanned by the possible values of M; if G is compact the result of Ito and Kawada applies immediately. E.g. if the M_ν mean rotations in n-space the limit distribution would typically be Haar measure over the orthogonal group or over one of its subgroups. If G turns out not to be compact (nor commutative making Fourier analysis possible) we have a new difficulty in front of us. One might think of using the theory of group representations for non-compact groups but, since this theory is both complicated and incomplete, this does not look too hopeful. We will see in later parts, however, that for many important concrete groups we can make use of the more detailed structure of the group implied by the problem.

In the above example the G was a group with matrices as elements. A more general—and more difficult—assumption would be to let G consist of transformations defined on a function space. This occurs naturally in the study of *non-linear stochastic difference equations*. Let x_t, $t = 0, 1, \ldots$, be (numerically valued) stochastic variables, joined to each other through the relations

$$x_{t+1} = f_t(x_t),$$

where for given t and x the value of $f_t(x)$ is a stochastic variable. For fixed $x_1, x_2, \ldots$ we also assume that $f_1(x_1), f_2(x_2), \ldots$ are stochastically independent. —Note that this simplifies to the well-known autoregressive model if $f_t(x) = = a x + \xi_t$, where ξ_t are independent and identically distributed stochastic variables.

Often the introduction of non-linearity into a stochastic problem gives rise to non-commutative difficulties. It is gratifying though that the process is Markovian, so that techniques exist that may be of use in this context.

1.7. We will arrive at a different but similar problem starting from *differential equations with stochastic coefficients*, as is clear from the analogy with the stochastic difference equations. Consider e.g. the equation

$$x^{(p)}(t) + a_1(t) x^{(p-1)}(t) + \cdots + a_{p-1}^{(t)} x'(t) + a_p(t) x(t) = 0$$

in the interval I, $0 \leqslant t \leqslant T$. The coefficients $a_\nu(t)$ are stochastic and, for the moment, let us assume that they are "independent at different time points"; these things will be made rigorous in a later part. Anyway, if we divide our interval into n equal parts $I = i_1 + i_2 + \cdots + i_n$, $i_\nu = \left(\dfrac{\nu-1}{n} T, \dfrac{\nu}{n} T \right)$, it is clear that specifying the values of the vector $y(t) = (x(t), x'(t), \ldots x^{(p-1)}(t))$ at the left endpoint of i_ν will determine $y(t)$ at the right endpoint if the coefficients are treated as constants throughout the short interval i_ν. These two values of $y(t)$ are related linearly through a matrix $M_\nu^{(n)}$ and the $M_\nu^{(n)}$ should be considered as independent for different ν's. Hence

$$y(T) = \lim_{n \to \infty} M_n^{(n)} \ldots M_2^{(n)} M_1^{(n)} y(0),$$

which should be compared to the corresponding finite relation of last section. Just as in the classical cases we hope, so far without any real justification, that the limit theorems on stochastic groups will be connected with stochastic processes taking values in these same groups.

One task of ours will be to give a mathematical meaning to differential equations of the type mentioned above and to study their solutions. This will perhaps be of special interest for certain partial equations with physical interpretation, e.g. the equation governing *stochastic waves* in one dimension with random propagation coefficient $a(x)$

$$a(x) \frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 u}{\partial t^2}.$$

The wawes will be superpositions of components whose frequencies will depend upon the eigen values λ defined by the equation

$$a(x) \frac{d^2 v}{d x} + \lambda v = 0$$

with appropriate boundary conditions. Therefore we have to study such *stochastic spectra*, and of course we should do this for difference equations also.

In the latter case there are already some preliminary but suggestive results due to Dyson and Wigner.

In this connection we should mention problems of the following type. In some n-space we are given a differential equation E, whose coefficients may be random or certain functions. The equation should be valid in some region D, which is likely to be multiply connected in some applications to electrical conductivity and simply connected in others. However that may be, the characteristic thing about D is that it is not fixed but given in probability terms only. If the solution of E should satisfy certain conditions on the boundary of D we will again be led to study certain stochastic eigen value problems. At present it is not clear if these later important problems can be phrased conveniently within the framework of stochastic groups, or if they require some other thechnique for their solution.

1.8. Let us sum up this discussion. We have to introduce probability measures on groups and see what general relations we can get. These relations will of course be more informative the more detailed structure we are given for the group. At present the most challenging case seems to be groups which are neither commutative nor compact. For some of the cases that we have mentioned (or will arrive at later on) it is striking that they possess two fundamental operations, say addition and multiplication, and form *algebras*. Actually in many important situations they are *Banach algebras*, and this will be useful to us. This is still a very general concept and we will have to specialize more in various directions.

On the group, of whatever type it happens to be, we will study stochastic processes of independent increments. Only the time homogeneous ones will be dealt with. In the classical cases this would give a good picture of what happens in general; in the present one the lack of commutativity could possibly make this assumption of homogeneity in time too narrow, but this is not known at present and should be investigated.

With the help of the knowledge we can get for such stochastic processes we hope to be able to state and prove limit theorems for "sums" or "products" of independent stochastic group elements. Again this will require some specialization of the groups. We will try to obtain laws of large numbers, analogues of the central limit theorem and results on infinitely divisible distributions on the groups. As far as infinitely divisible laws on Lie groups are concerned the interested reader should study the important work of Hunt.

Often the groups are presented as transformation groups already in the original form of the practical problem: g transforms some space S (with elements s) into itself, $g S \subseteq S$. Instead of working with the probability distribution of g over G it may be more convenient to work with the distribution of $g s_0$ over S for an arbitrary but fixed element s_0 of S.

Part 2. General discussion

2.1. The object of our investigation will usually have three structures superimposed: an *algebraic*, a *topological*, and a *probabilistic* structure. These three cannot be chosen independently of each other. The topology must be chosen

to make the fundamental operations of the group, ring, field, or whatever it is, continuous functions of the arguments. But in the same way we must demand that the same functions be measurable so that they can be subjected to probability statements. This leads us naturally to the definition

By a stochastic group we mean a topological group G forming a probability space $(G, \mathcal{G}, P)$, *where the* σ*-algebra* $\mathcal{G}$ *of subsets of G has the property that* $ST \in \mathcal{G}$, $S^{-1} \in \mathcal{G}$, *for any* $S, T \in \mathcal{G}$. *Open sets should belong to* $\mathcal{G}$.

At present we will write the group operation as multiplication, and ST means the group operation product and not the set theoretic product, which will be denoted by $S \cap T$.

Considering stochastic groups $G_1, G_2, ..., G_n$ they are said to be *independent* if the probability measure over the Cartesian product $G_1 \times G_2 \times ... \times G$ is simply the product of the measures over the individual components.

The topology defined on G induces a topology in the set of probability distributions on G, and this gives us the notion of *convergence in probability on a stochastic group*. Given a sequence of probability distributions $P_n(g_n, g)$ defined on $G \times G$ we shall say that g_n converges to g in probability if for every neighbourhood N of the unit element e of the group we have

$$\lim_{n \to \infty} P_n \{g_n g^{-1} \in N\} = 1.$$

Given two probability distributions P_1 and P_2 on G we define their *convolution* as

$$P_1 * P_2(S) = \int_{g_1 g_2 \in S} dP_1(g_1) dP_2(g_2) =$$

$$= \int_{g \in G} dP_1(g) P_2(g^{-1} S).$$

The convolution operation will in general not be commutative. Here, as in the following, the choice between right and left multiplication will give rise to a duality in definitions and relations.

The convolution corresponds of course to *multiplication of independent groups*. For several independent groups $G_1, G_2, ...$ we define their *partial product* $\gamma_n = g_1 g_2 ... g_n$ with its probability measure $\pi_n = P_1 * P_2 * \cdots * P_n$. The study of this measure for large values of n is one of the major problems in the theory of stochastic groups.

Sometimes we can introduce a numerical quantity measuring the spread of a distribution over a group. More generally we will define a partial ordering, saying that P_1 is *more concentrated* than P_2 if there is a non degenerate distribution P_3 such that $P_2 = P_1 * P_3$.

Consider a family of probability measures $P_t(g)$ over g, indexed by a positive argument t and such that

$$P_{t+s} = P_t * P_s \quad \text{for every } s, t > 0.$$

This will be called a *homogeneous stochastic process* with values in G. Usually

we will assume that the process has some continuity property, e.g. that it is *continuous in probability* so that

$$\lim_{t \downarrow 0} P_t (g \in N) = 1$$

for every neighbourhood N of the identity e.

Another important concept is that of an infinitely divisible distribution. We shall say that a distribution P over the stochastic group G is *infinitely divisible* if for every positive integer n there is a distribution Q such that $P = Q^{n}{}^{*}$.

We shall say that a probability distribution P over G is symmetric if $P(E^{-1}) = P(E)$ for every $E \in \mathcal{G}$.

2.2. Given a stochastic group G with a distribution P, what can we say about the behaviour of $\pi_n = P^{n}{}^{*}$ when n increases indefinitely? We have already mentioned in 1.5 that Ito and Kawada have shown essentially, that if G is compact, then π_n converges (weakly) to the invariant normed measure on some subgroup of G. Actually there is also one other possibility which is however of little concern to us at present.

If we deal with a stochastic group for which $\pi_n \underset{\text{weakly}}{\to} \pi$ then it follows that the limit measure π must satisfy $\pi \ast \pi = \pi$; it must be *idempotent*. However if this is the case one can show that G is compact.

At first glance it may appear surprising that the seemingly innocent assumption on compactness should (essentially) guarantee convergence of π_n. A moments reflection makes this plausible. Indeed, if G is compact, then the set of probability distributions over G is also compact, and any sequence of probability distributions on G, say $\pi_1, \pi_2, \pi_3, \ldots$, must at least contain a convergent subsequence; the limit of this will be the invariant measure.

Leaving the compact groups it is clear that we cannot expect limit theorems of the simple type that we have just discussed. To get results analogue to the classical limit theorems, say the law of large numbers, we have to norm or transform the partial products $g_1 g_2 \ldots g_n$ in a suitable manner. The probability space may be given a linear structure, and this can be done in many ways. In the next section we will describe *one* way that we have chosen mainly because it seems to correspond naturally to the practical problems of Part 1 that have motivated the present study.

Before we do this, let us remark that it is possible to obtain *asymptotic* expressions for π_n in a general context. We will return to this important problem in a later part of this publication.

2.3. We need an algebraic-topological concept that is general enough to include most of the applied situations, but at the same time it should have a sufficiently detailed structure to make possible the statement of limit theorems of the sort we have in mind. Such a concept seems to be the *Banach algebra*.

We shall say that X is a stochastic Banach algebra if it is a Banach algebra and forms a probability space $(X, \mathcal{X}, P)$, where the σ-ring $\mathcal{X}$ includes the neighbourhoods of the space and is invariant with respect to the algebraic operations of the algebra.

In this paper we shall always assume 1) that X is separable and 2) that it has a unit element e.

Within the framework it is easy to formulate and prove a natural extension of the law of large numbers. If x_1, x_1, ... are independent and identically distributed stochastic elements in a Banach algebra, we should norm the product $\pi_n = x_1 x_2 \ldots x_n$ or its factors so that π_n converges in probability to a constant element of X. One way of doing this is the following.

Let y_ν be independent and identically distributed stochastic elements of X such that $E\|y\|$ exists and is finite. Let $x_\nu = e + \dfrac{1}{n} y_\nu$ so that

$$\pi_n = \left(e + \frac{1}{n}y_1\right)\left(e + \frac{1}{n}y_2\right)\ldots\left(e + \frac{1}{n}y_n\right) = e + S_1^{(n)} + S_2^{(n)} + \cdots + S_n^{(n)},$$

where

$$\begin{cases} S_1^{(n)} = \dfrac{1}{n}\sum_\nu y_\nu \\[2mm] S_2^{(n)} = \dfrac{1}{n^2}\sum_{\nu<\mu} y_\nu y_\mu \\[2mm] \text{etc.} \end{cases}$$

Applying a theorem of Mourier to the sum $S_1^{(n)}$ we know that it converges almost certainly strongly to an element $m \in X$. This limit element m is the *expected value* of y, $m = \int\limits_X y\, dP(y)$, where this integral is interpreted in the sense of Pettis. Under the given conditions the expected value exists.

Similarly

$$S_2^{(n)} = \frac{1}{n}\sum_{\mu=1}^{n}\frac{\mu-1}{n}S_1^{(\mu)}y_\mu = \frac{1}{n}\sum_{\mu=1}^{n}\frac{\mu-1}{n}m y_\mu + \frac{1}{n}\sum_{\mu=1}^{n}\frac{\mu-1}{n}\varepsilon_\mu y_\mu$$

where $S_1^{(\mu)} = m + \varepsilon_\mu$. Hence, with probability one,

$$S_2^{(n)} = \frac{m^2}{2} + \varepsilon,$$

with

$$\|\varepsilon\| \to 0 \text{ as } n \to \infty.$$

This means that $S_2^{(n)}$ converges almost certainly strongly to $m^2/2$. In this way we can prove that $S^{(n)}$ converges almost certainly to $m^\nu/\nu!$.

Now we just have to complement this reasoning with a simple uniformity argument. We have of course

$$\|S_\nu^{(n)}\| \leqslant \frac{1}{n^\nu}\sum_{k_1<k_2<\cdots<k_\nu}\|y_{k_1}\|\cdot\|y_{k_2}\|\ldots\|y_{k_\nu}\|$$

so that

$$E\|S_\nu^{(n)}\| \leqslant \frac{1}{\nu!}(E\|y\|)^\nu.$$

Combining this with the above we have shown that *the normed partial products* π_n *converge strongly in probability to the constant element*

$$e^m = e + m + \frac{m}{2} + \cdots \in X.$$

2.4. Let $y(t)$, $0 < t < \infty$, be a homogeneous process with values in X and continuous in probability. The word homogeneous should here refer to *addition*, which is a commutative operation on X. In practical applications $y(t)$ will usually be something wellknown to us, a Wiener- or Poisson-process, a process of independent increments on R_n or on something else, of this nature. From $y(t)$ we would like to form a process $x(t)$ homogeneous with respect to *multiplication*. We could do something like this: let us divide the interval $(0, t)$ into n disjoint intervals $(0, t_1)$, (t_1, t_2), ... (t_{n-1}, t) with the corresponding increments $\Delta_\nu y = y(t_\nu) - y(t_{\nu-1})$. Form the product

$$\pi_n(t) = (e + \Delta_1 y)(e + \Delta_2 y) \dots (e + \Delta_n y).$$

As we make n large the products $\pi_n(t)$ could perhaps be proved to converge in probability to some $\pi(t)$ and this would be our homogeneous (multiplicative) process.

We will do this in the following way resembling the expansion of π_n of last section. Let us define

$$x(t) = e + \int_0^t dy(s) + \int_{0 < s_1 < s_2 < t} dy(s_1)\, dy(s_2) + \int_{0 < s_1 < s_2 < s_3 < t} dy(s_1)\, dy(s_2)\, dy(s_3) + \cdots.$$

and we will call $x(t)$ *the multiplicative homogeneous process generated by* $y(t)$. We must of course give a precise meaning to the terms of this expansion and show that it converges.

The integrals will be defined as the limit of Riemann-Stieltjes sums in the L_1-topology over G. Let us consider the double integral only; the higher ones are dealt with in the same manner. Divide the interval $(0, t)$ into $(0, t_1)$, (t_1, t_2), ..., (t_{n-1}, t) and form the Riemann-Stieltjes sum

$$S = \sum_{\nu < \mu} [y(t_\nu) - y(t_{\nu-1})]\,[y(t_\mu) - y(t_{\mu-1})].$$

As the division is made finer these sums will converge to a limit which does not depend upon what sequence of divisions we have chosen. To see this let us take another division $(0, t_1')$, (t_1', t_2'), ..., (t_{m-1}', t) and denote the corresponding sum by S'. The combined division $(0, t_1'')$, (t_1'', t_2''), ..., (t_{n+m-1}'', t) gives rise to to a sum S''.

As an illustration consider the figure below. To each rectangle of the figure corresponds one term in the respective sum. In the difference $S'' - S$ only the shaded rectangles of the last figure occur, which follows from $(x + y)(z + u) = = xz + xu + yz + yu$. It is now clear what happens in general: the difference $S'' - S$ consists only of rectangles close to the diagonal and its norm is dominated by
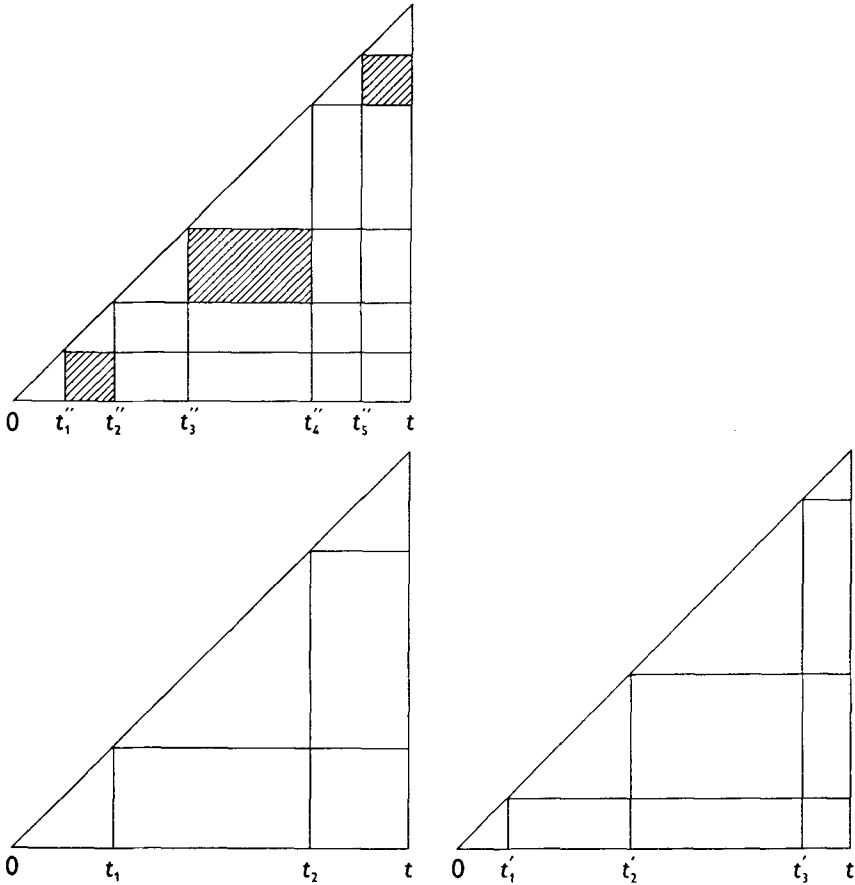
Fig. 1.

$$\| S'' - S \| \leqslant \sum \| [y\,(t_\nu'') - y\,(t_{\nu-1}'')]\,[y\,(t_{\nu+1}'' - y\,(t_\nu'')] \| \leqslant$$
$$\leqslant \sum \| (t_\nu'') - y\,(t_{\nu-1}'') \| \cdot \| y\,(t_{\nu+1}'') - y\,(t_\nu'') \|. \tag{1}$$

Using the independence of the increments of the y-process the L_1-norm is bounded by

$$E \| S'' - S \| \leqslant \sum E \| y\,(t_\nu'') - y\,(t_{\nu-1}'') \| \cdot E \| y\,(t_{\nu+1}'') - y\,(t_\nu'') \|.$$

We will assume that $E \| y\,(s+h) - y\,(s) \|$ exists, is continuous and that the sums

$$\sum E \| y\,(t_\nu) - y\,(t_{\nu-1}) \| \leqslant M \tag{2}$$

are uniformly bounded for arbitrary divisions of the fixed interval $(0, t)$. Then, if the division is sufficiently fine, S differs arbitrarily little from S'' and hence also from S' which completes the proof.

175

The reasoning also gives us a bound for the norm of the integrals

$$E \left\| \int\limits_{0 < S_1 < \cdots < S_k < t} d\,y\,(S_1)\,d\,y\,(S_2) \dots d\,y\,(S_k) \right\| \leqslant \frac{M^k}{k!},$$

which shows that the sum defining $x\,(t)$ converges in the L_1-norm. We have then proved that *if $y\,(t)$ is an additive homogeneous process continuous on $(0, t)$ in the L_1-norm and if the sums (1) are uniformly bounded then the expansion (2) converges in the same norm to a multiplicative process $x\,(t)$.* Only the last statements remains to be verified which is very simple. Indeed putting

$$z\,(t,\,t+h) = e + \int\limits_{t}^{t+h} d\,y\,(s) + \int\limits_{t < s_1 < s_2 < t+h}\!\!\!\int d\,y\,(s_1)\,d\,y\,(s_2 + \cdots$$

it follows that $x\,(t+h) = x\,(t) \cdot z\,(t,\,t+h)$ which can be seen by going back to the Riemann-Stieltjes sums. Clearly $z\,(t,\,t+h)$ is independent of $x\,(t)$ and has the same distribution as $x\,(h)$. The same holds for the increments over several intervals.

The relation between $x\,(t)$ and $y\,(t)$ can also be given the form of a *functional equation*

$$x\,(t) = e + \int\limits_0^t x\,(s)\,d\,y\,(s),$$

which can conveniently be written symbolically as

$$d\,x\,(t) = x\,(t)\,d\,y\,(t).$$

The functional equation also shows that $x\,(t)$ is continuous in the L_1-norm.

In one particular case the relation between the x- and y-process is especially simple. This is when the values of $y\,(t)$ commute so that we can write.

$$x\,(t) = e + \int\limits_0^t d\,y\,(s) + \frac{1}{2} \int\limits_0^t \int\limits_0^t d\,y\,(s_1)\,d\,y\,(s_2) +$$

$$+ \frac{1}{6} \int\limits_0^t \int\limits_0^t \int\limits_0^t d\,y\,(s_1)\,d\,y\,(s_2)\,d\,y\,(s_3) + \cdots =$$

$$= \exp \int\limits_0^t d\,y\,(s) = \exp y\,(t).$$

The reader may have noted that the construction of $x\,(t)$ from $x\,(t)$ resembles that of multiplicative integrals.

2.5. Let us now derive an *approximation theorem* useful when studying the asymptotic probability distribution of the product of a large number of stochastic factors, each of which does not deviate too much from the unit element. First we must define what should be meant by convergence distributionwise on X. Let $P_1, P_2, \ldots$ be a sequence of distributions on X with the same domain of definition. We will say that P_ν *converges distributionwise* to a distribution P if the distribution functions $f(x_k)$ converges to the distribution function of $f(x)$ in the usual sense. Here $x_1, x_2, \ldots$ and x are stochastic group elements with the distributions $P_1, P_2, \ldots$ and P respectively, and $f(x)$ is an arbitrary real valued and uniformly continuous function defined on X.

We will consider a double array of stochastic group elements

$$\left\{ \begin{array}{l} y_{11} \\ y_{21}, \ y_{22} \\ y_{31}, \ y_{32}, \ y_{33} \\ \quad . \quad . \quad . \quad . \quad . \end{array} \right.$$

where the elements in each row are independent and have the same distribution. Further we shall assume that their norm is integrable and that the sum

$$\sum_{\nu=1}^{n} E\|y_{n\nu}\| \leqslant M < \infty$$

is uniformly bounded.

Let $y(t)$ be an additive homogeneous process defined on X with the properties described in the last section and with the associated multiplicative homogeneous process $x(t)$. Assume that for every proportion c between 0 and 1 we have

$$\sum_{\nu=1}^{[cn]} y_{n\nu} \xrightarrow[\text{distributionwise}]{} y(c).$$

We are going to show *that*

$$\pi_n = (e + y_{n1})(e + y_{n2}) \ldots (e + y_{nn})$$

converges distributionwise to $x(1)$.

Indeed we have

$$\left\{ \begin{array}{l} \pi_n = e + S_1^{(n)} + S_2^{(n)} + \cdots + S_n^{(n)} \\ x(1) = e + S_1 + S_2 + \cdots, \end{array} \right.$$

where

$$\left\{ \begin{array}{l} S_1^{(n)} = \sum_{1}^{n} y_{n\nu} \\ S_2^{(n)} = \sum_{\nu<\mu} y_{n\nu} y_{n\mu} \\ \quad . \quad . \quad . \quad . \quad . \quad . \quad . \end{array} \right.$$

and

$$\begin{cases} S_1 = \int_0^1 dy(t) = y(1) \\[2mm] S_2 = \iint\limits_{0 < s_1 < s_2 < 1} dy(s_1)\, dy(s_2) \\[2mm] \cdots\cdots\cdots\cdots\cdots\cdots \end{cases}$$

Working in the L_1-topology over X we know that if the expression for $x(1)$ is replaced by a partial sum, little is changed if the truncation point m of the sum is chosen sufficiently far out. We also know that if we use a fine division $0 < t_1 < t_2 < \cdots < t_k = 1$ of the unit intervals and replace the integrals S_1, S_2, ..., S_m by the Riemann-Stieltjes sums the effect is small. Now we can do exactly the same thing with the expression for π_n: truncate the sum and replace the terms by sums which are summed over blocks (rectangles) of the indices. But each such sum over a block converges distributionwise to the analog quantity expressed in terms of increments of the y-process. Using the independence of the y_{nv}'s and increments of $y(t)$ respectively the stated results follows.

The importance of this result lies in the fact that *it enables us to pass from additive to multiplicative limit theorems*. Additive limit theorems are available both from the classical theory and from its extension to linear spaces due to Mourier and Fortet.

However, the result is limited by the assumption that the individual factors contribute little to the product. Also we may want to examine problems having group structure instead of being an algebra. For this other methods will be used. It should also be noted that there are important situations where the sums (2) do not stay bounded. It will then be convenient to use an L_2-argument instead, and this will be done in Part 5 of this study.

Part III. Remarks on limit theorems

3.1. In this part we will return to the fairly general situation of stochastic groups with no other structure superimposed. One important task here is to study the asymptotic behaviour of convolutions P^{n*} for a given probability distribution P over G. We will assume throughout that G is locally compact although a part of what follows does not depend upon this condition.

If the support S of the measure P on G does not coincide with G it is clear that we loose nothing in generality by confining us to the group $\bar{S} \subset G$. Here $\bar{S}$ is the algebraic-topological closure of S. This will always be assumed done.

Let us assume that there exists a σ-finite *right invariant measure* μ, $\mu(Ex) = \mu(E)$ for every $x \in G$, $E \in \mathcal{G}$ and that P is absolutely continuous with respect to μ

$$P(E) = \int_E p(x)\, d\mu(x).$$

Here $p(x)$ has the properties of a generalized frequency function.

We will also assume that P is a *symmetric distribution* (see 2.1) which makes simpler the use of the Hilbert space method below. It seems desirable to get rid of this restriction, which is not inherent in the original formulation of the problem.

As a tool for the examination of P^n* we will use the transformation

$$T f(x) = \int_{y \in G} f(xy) \, dP(y)$$

which is well defined at least for functions $f(x)$ taking a finite number of values

$$f(x) = \begin{cases} f_\nu & \text{if } x \in E_\nu, \ \mu(E_\nu) < \infty, \ E_\nu \text{ disjoint} \\ 0 & \text{if } x \notin \bigcup_1^n E_\nu, \end{cases}$$

$$T f(x) = \sum_1^n f_\nu P(x^{-1} E_\nu).$$

But using the L_2-metric

$$\|g\|^2 = \int_G |g(x)|^2 \, d\mu(x),$$

we have
$$\|T f(x)\|^2 = \int_G \left| \sum_1^n f_\nu P(x^{-1} E_\nu) \right|^2 d\mu(x) =$$

$$= \int_G \left| \sum_1^n f_\nu \int_G \varphi_\nu(xy) \, p(y) \, d\mu(y) \right|^2 d\mu(x),$$

where $\varphi_\nu(u)$ is the indicator function of E_ν. Hence
$$\|T f(x)\|^2 =$$

$$= \int_G \int_G \int_G \sum f_\nu \varphi_\nu(xy) \sum f_\mu \varphi_\mu(xz) \, p(y) \, p(z) \, d\mu(y) \, d\mu(z) \, d\mu(x) \leqslant$$

$$\leqslant \int_{y \in G} \int_{z \in G} \sqrt{\int_G |\sum f_\nu \varphi_\nu(xy)|^2 \, d\mu(x)} \sqrt{\int_G |\sum f_\mu \varphi_\mu(xz)|^2 \, d\mu(x)} \, p(y) \, p(z) \, d\mu(y) \, d\mu(z) =$$

$$= \|f\|^2 \int_{y \in G} \int_{z \in G} dP(y) \, dP(z) = \|f\|^2.$$

For the last equality we have used

$$\int_G |\sum f_\nu \varphi_\nu(xy)|^2 \, d\mu(x) = \int_G |\sum f_\nu \varphi_\nu(u)|^2 \, d\mu(uy^{-1}) =$$

$$= \int |\sum f_\nu \varphi_\nu(u)|^2 \, d\mu(u) = \|f\|^2.$$

But since $\|T f\| \leqslant \|f\|$ for finite valued functions $f(x)$ the transformation can be extended uniquely to functions $f(x)$ belonging to the space of quadratically functions $L_2(G)$ (with respect to the invariant measure).

The bounded linear operator T defined in $L_2(G)$ will actually have exactly the norm $\|T\| = 1$ under mild conditions. Suppose that it is possible to find an increasing sequence of sets $E_1, E_2, \ldots$ such that

$$\left. \begin{array}{c} \mu(E_\nu) < \infty \\[2mm] E_\nu \uparrow G \\[2mm] \lim\limits_{\nu \to \infty} \dfrac{\mu(E_\nu \cap E_\nu z)}{\mu(E_\nu)} = 1, \ z \in G. \end{array} \right\}$$

The last condition means that the effect of any finite translation $z \in G$ has a small effect measured relative to $\mu(E_\nu)$.

Under this condition let $\varphi_\nu(x)$ be the indicator function of E_ν. Then

$$T \varphi_\nu(x) = \int\limits_G \varphi_\nu(xy)\, p(y)\, d\mu(y)$$

and
$$\|T \varphi_\nu(x)\|^2 = \int\limits_G \int\limits_G \int\limits_G \varphi_\nu(xy)\, \varphi_\nu(xz)\, p(y)\, p(z)\, d\mu(x)\, d\mu(y)\, d\mu(z) =$$

$$= \int\limits_G \int\limits_G p(y)\, p(z)\, g_\nu(y, z)\, d\mu(y)\, d\mu(z),$$

with
$$g_\nu(y, z) = \int\limits_{x \in G} \varphi_\nu(xy)\, \varphi_\nu(xz)\, d\mu(x) = \mu(E y^{-1} \cap E z^{-1})$$

and
$$\left. \begin{array}{c} 0 \leqslant \dfrac{g_\nu(y, z)}{\mu(E_\nu)} \leqslant 1 \\[4mm] \lim\limits_{\nu \to \infty} \dfrac{g_\nu(y, z)}{\mu(E_\nu)} = 1 \end{array} \right\} ;$$

this implies bounded convergence of the integrals

$$\lim\limits_{n \to \infty} \frac{\|T \varphi_n(x)\|}{\|\varphi_n(x)\|} = 1$$

proving the assertion.

The operator T is *self adjoint* since

$$(T f, g) = \int\limits_G \int\limits_G f(xy)\, \overline{g(x)}\, dP(y)\, d\mu(x) =$$

$$= \int\limits_G \int\limits_G f(u)\, \overline{g(uv)}\, dP(v^{-1})\, d\mu(u) =$$

$$= \int\limits_G \int\limits_G f(u)\, \overline{g(uv)}\, dP(v)\, d\mu(u) = (f, T g).$$

The spectrum of T is hence situated in the interval $(-1, 1)$ and one asks

whether the endpoints ± 1 can be discrete eigen values. If such is the case we have

$$\pm f(x) = T f(x) = \int_G f(xy) \, dP(y)$$

for some $f(x) \in L_2(G)$, and according to Schwarz' inequality

$$\|f\|^2 = \iiint f(xy) \, \overline{f(xz)} \, d\mu(x) \, dP(y) \, dP(z) \leqslant$$

$$\leqslant \iint \|f\|^2 \, dP(y) \, dP(z) = \|f\|^2$$

where we have used the right invariance of μ. But equality is possible only if

$$f(xy) = c f(xz), \quad |c| = 1,$$

almost everywhere with respect to $d\mu(x)$, $dP(y)$, $dP(z)$. We now make use of the restriction of G to the closure of P and deduce that $|f(x)| = $ constant almost everywhere with respect to $d\mu(x)$. Such a function belongs to $L_2(G)$ if and only if $\mu(G)$ is finite, which implies that G is compact. Conversely if G is compact the function $f(x) \equiv$ constant belongs to $L_1(G)$ and is an eigenfunction associated with the eigen value $\lambda = 1$.

It is possible to give a sufficient and not very strong condition ensuring that the spectrum of T is bounded away from $\lambda = -1$ and situated in an interval $(c, 1)$ with $-1 < c < 1$.

3.2. The *iterates* T^n of T have a simple probabilistic meaning. Consider $n = 2$

$$T^2 f(x) = \int_G \int_G f(xyz) \, dP(y) \, dP(z) = \int_G f(xu) \, dP^{2*}(u),$$

and generally $T^n f$ is the transform of f with respect to the n-th convolution. Thus we can get a convenient representation of the probability $P^{n*}(x^{-1}E)$ of any set E of finite μ-measure. If $\varphi(x)$ is the indicator function of E we have $T^n \varphi(x) = P^{n*}(x^{-1}E)$, so that e.g. $P^{n*}(E) = T^n \varphi(e)$.

Using the self-adjoint nature of T we can represent the operator in terms of a resolution $E(\lambda)$ of the identity

$$T = \int_{-1}^{1} \lambda \, dE(\lambda)$$

so that $\qquad P^{n*}(x^{-1}E) = T^n \varphi(x) = \int_{-1}^{1} \lambda^n \, dE(\lambda) \varphi(x).$

This relation also holds pointwise in x. To make this rigorous we note that $P^{n*}(x^{-1}E)$ is a continuous function of x.

We are interested in what happens to $P^{n*}(x^{-1}E)$ for large values of n. One thing we can say immediately; the probability tends to zero in the mean. Indeed

$$\|P^{n*}(x^{-1}E)\|^2 = \int_{-1}^{1} \lambda^{2n} \, d\|E(\lambda)\varphi\|^2$$

and the funation of bounded variation $\|E(\lambda)\varphi\|^2$ is continuous at $\lambda = \pm$; the opposite case was excluded by us in the last section. But then we can apply Lebesgue's theorem on bounded convergence to the above integral showing that it converges to zero.

To get a sharper statement on the asymptotic decrease of $P^{n*}(x^{-1}E)$ we must of course start from more detailed assumptions, e.g. as follows. Assume that the derivative

$$\frac{d}{d\lambda} E(\lambda)\,\varphi(x) = a(\lambda,\,x)$$

exists, is continuous in $(-1, 1)$ and behaves asymptotically as $a(1-\lambda)^{\alpha}$, $\alpha > -1$, at the point $\lambda = 1$. Then

$$P^{n*}(x^{-1}E) \sim a \int_0^1 \lambda^n (1-\lambda)^\alpha \, d\lambda = a\,\frac{\Gamma(n+1)\,\Gamma(\alpha+1)}{\Gamma(n+\alpha+2)} \sim \frac{a\,\Gamma(\alpha+1)}{n^{\alpha+1}}.$$

"Local" limit results of this type could be extended to cover the more general situation, where we let x move out as n increases, as soon as we have constructed the family $E(\lambda)$ of projection operators. In specific situations this may be difficult to achieve and we will see later that more direct methods are available in specific situations.

3.3. In the present context our main interest is concentrated on the non-commutative case, but it may prove instructive to see how the above simplifies considerably on a commutative group.

Consider for a fixed but arbitrary $y \in G$ the operator

$$S_y\,f(x) = \tfrac{1}{2}\,[f(xy) + f(xy^{-1})].$$

It has most of the properties that T has. It is certainly bounded, $\|S_y\| \leqslant 1$, and self-adjoint; the proof of this is left to the reader. For any y, $z \in G$ we have

$$4\,S_z\,S_y f(x) = f(xyz) + f(xyz^{-1}) + f(xy^{-1}z) + f(xy^{-1}z^{-1}) =$$
$$= f(xzy) + f(xzy^{-1}) + f(xz^{-1}y) + f(xz^{-1}y^{-1}) =$$
$$= 4\,S_y\,S_z f(x).$$

The set of operators $\{S_y\,;\,y \in G\}$ is then a family of bounded, self-adjoint and commuting operators. It is then known that they can all be represented as "functions" of one single operator associated with the operator family of projections $F(\lambda)$,

$$S_y = \int \psi_y(\lambda)\,dF(\lambda),$$

so that

$$T = \int_G S_y\,dP(y) = \int \psi(\lambda)\,dF(\lambda)$$

with

$$\psi(\lambda) = \int_G \psi_y(\lambda)\,dP(y)$$

Hence

$$T^n f = \int \psi^n(\lambda)\,dF(\lambda),$$

which, after a change of variable, is the same representation as before. Once we have found the spectral representation for S_y it can be used to construct the resolution of the identity for T, *and this holds whatever p is.* The existence of such a universal (on G) spectral representation makes the commutative case especially simple. At present we do not know if something like this holds for the non-cummutative groups, but it seems unlikely.

The reader may have observed that for a commutative G the differential solutions of S_y can be expressed in terms of the *group characters.*

For a general group let us study the transformation $S = S_y$ with y fixed. This leads us to the cyclic group y^n; suppose it is of infinite order. Starting from its characters we can immediately write down the spectral representation of S and the value of $P^{n*}(E)$. The simple form of the expression obtained indicates that we should be able to derive it using less brutal methods.

Indeed, taking for example E to consist of the unit element e only, we have

$$P^{n*}(E) = P(y^{\pm 1} y^{\pm 1} \ldots y^{\pm 1} \in E)$$

where the $+$ and $-$ signs are independent and have the probability $1/2$ each. Hence we get the binomial probability

$$P^{n*}(E) = \binom{n}{\frac{n}{2}} 2^{-n}$$

if n is even and the probability zero if n is odd.

The approach of the present part may be of some use for the general discussion, but for the actual detailed study of particular stochastic groups more direct algorithms may be found.

PUBLICATIONS REFERRED TO IN THE TEXT

BELLMAN, R.: Limit theorems for non-commutative operations, I. Duke Math. J. 21, 1954.
BOCHNER, S.: Harmonic analysis and the theory of probability, California Monographs in Mathematical Sciences, Berkeley 1956.
DYSON, F. J.: The dynamics of a disordered linear chain, Phys. Rev., 92, 1953.
FORTET, R.–MOURIER, E.: Convergence de la répartition empirique vers la répartition théorique, Ann. Ec. Norm. Sup. 60, 1953.
GRENANDER, U.: Some non linear problems in probability theory, The Harald Cramér volume, Stockholm, 1959.
HUNT, G. A.: Semi-groups of measures on Lie groups, Trans. Am. Math. Soc., 81, 1956.
ITO, K.–KAWADA, Y.: On the probability distribution on a compact group, Proc. Phys. Math. Soc. Japan, 22, 1940.
MOURIER, E.: Les éléments aléatoires dans un espace de Banach, Ann. Inst. Henri Poincaré, 13, 1953.
ROSENBLATT, M.: Limits of convolution sequences of measures on a compact topological semigroup, to appear.
WIGNER, E. P.: On the distribution of the roots of certain symmetric matrices, Ann. Math., 67, 1958.