ARKIV FÖR MATEMATIK Band 4 nr 25

Communicated 12 October 1960 by H. Cramér and O. Frostman

Stochastic groups

By Ulf Grenander

This paper is a continuation of "Stochastic groups" by the same author and published in the same journal. The reader is referred to the earlier parts of this paper and to the author's paper, "Stochastic groups and related structures", to appear in the fourth Berkeley Symposium on Mathematical Statistics and Probability for a full statement of the problems and for terminology and notation.

Part 6. Fourier analysis of probability distributions on locally compact groups

6.1. We now turn to the study of such stochastic groups as are assumed to be neither compact nor commutative. The probability theory of locally compact groups is at present almost entirely *terra incognita* and presents a number of challenging problems. Even for stochastic Lie groups the situation is similar, although we have some knowledge of the corresponding infinitely divisible probability distributions.[1] For a general treatment the most promising way seems to be via Fourier analysis. We then have to start from the *irreducible, unitary group representations,* construct the Fourier transform and study its properties. We hope to get a tool which will be of great help in future investigations.

For locally compact groups the unitary representations are a good deal more complicated than in the compact case, when one can appeal to the Peter–Weyl theorem. Now we are forced to use *infinite dimensional representations* with all the possible pathologies that can arise. In this section we will review some known results on unitary representations that will be needed below.

Let G be a locally compact, separable group with the generic element g. By a unitary representation $r=(\mathcal{H}, U(g))$ we understand a Hilbert space $\mathcal{H}$ and a family of unitary transformations $U(g)$, $g\in G$, in $\mathcal{H}$, satisfying the equation $U(g_1)U(g_2)=U(g_1 g_2)$. For any element $z\in\mathcal{H}$ the vector-valued function $U(g)z$ defined on $\mathcal{H}$ will be assumed to be strongly (which here is equivalent to weakly) continuous. The representation is said to be irreducible if there is no non-trivial, closed subspace of $\mathcal{H}$ left invariant by all the $U(g)$.

An important class of representations are the so-called *regular* ones. Let $\mathcal{H}$ consist of all complex valued functions $f(g)$ defined on the group and quadratically integrable with respect to left invariant Haar measure; the ordinary definition of inner product is used. Put $U(h)f(g)=f(h^{-1}g)$. It is not difficult to see that $(\mathcal{H}, U(g))$ is a unitary representation.

[1] Very recently Donald Wehn obtained some important limit theorems on Lie groups in "Limit distributions on Lie groups" (to appear).

A fundamental theorem of Gelfand and Raikov tells us that there exists a set $R=\{r\}$ of irreducible, unitary representations which is *complete* in the following sense. If $g \neq e$ is an arbitrary element of G there is a representation $r \in R$ such that $U(g) \neq I$.

A function $\varphi(g)$, $g \in G$ is called *positive definite* if for any choice of an integer $n, g_1, g_2, \ldots, g_n$ in G and of complex numbers $c_1, c_2, \ldots, c_n$ we have

$$\sum_{\nu,\mu=1}^{n} c_\nu \bar{c}_\mu \varphi(g_\mu^{-1} g_\nu) \geqslant 0.$$

We are especially interested in the normed, $\varphi(e)=1$, and continuous positive definite functions. They are related to the unitary representations in the following way. For any unitary representation $\{\mathcal{H}, U(g)\}$ and vector $z \in \mathcal{H}$, the function $(U(g)z, z)$ is continuous and positive definite. Inversely every continuous, positive definite function can be represented in this way.

It is natural to use the partial ordering $\varphi_1 \prec \varphi_2$ for two positive definite functions if the difference $\varphi_2 - \varphi_1$ is also positive definite; φ_1 is said to be subordinated to φ_2. A positive definite function $\varphi(g)$, for which the only subordinated functions are multiples, $c\varphi(g)$, is said to be *elementary*. Their importance lies in the fact that they can be used as building blocks via the *trigonometric polynomials* $c_1\varphi_1(g) + c_2\varphi_2(g) + \cdots + c_n\varphi_n(g)$. This completeness property can be expressed in either of the two following ways:

Any continuous function on G can be approximated uniformly on every compact set by trigonometric polynomials.

If μ is a bounded complex measure and

$$\int_{g\in G} \varphi(g)\, d\mu(g) = 0$$

for any elementary function φ, then $\mu=0$.

To emphasize the concrete nature of this investigation we shall illustrate the general problem by a particular group that will serve as a simple but illuminating example. Consider the group of linear transformations of the real line $x \to \alpha x + \beta$. It has one-dimensional representations of the form α^{it} where t is a real number. To construct the infinite dimensional representations consider the Hilbert space H^+ of functions $f(\lambda)$ defined on the positive real line and with the ordinary definition of inner product. Put

$$U^+(g)f(\lambda) = e^{i\lambda\beta} f(\lambda\alpha)\sqrt{\alpha}, \quad g=(\alpha,\beta),\ f(\lambda)\in H^+.$$

Similarly we introduce H^- consisting of quadratically integrable functions on the negative real line and in H we define $U^-(g)$ analogously. The operators $U^+(g)$ and $U^-(g)$ respectively are easily seen to form irreducible, unitary representations of G. Further it can be shown that, together with the one-dimensional representations, they form a *complete* set of irreducible, unitary representations.

Given a unitary representation $(\mathcal{H}, U(g))$ it can be decomposed into a direct integral of *irreducible* unitary representations

$$U(g) = \int_D \oplus\, U^d(g),$$

which means the following. To each $d \in \mathcal{D}$ corresponds a Hilbert space $\mathcal{H}^d$. On $\mathcal{D}$ there exists a measure ϱ and we can consider $\mathcal{H}$ as equivalent to the Hilbert space

$$\mathcal{H} = \int_{\mathcal{D}} \mathcal{H}^d \sqrt{d\varrho},$$

having as elements functions $x = x(d)$, $d \in \mathcal{D}$, $x(d) \in \mathcal{H}^d$ and with the inner product

$$(x, y) = \int_{\mathcal{D}} (x(d),\ y(d))\, d\varrho.$$

There exists for almost all d an irreducible unitary representation $\{\mathrm{H}^d, U^d(g)\}$ such that

$$(U(g)\, x, y) = \int_{\mathcal{D}} (U^d(g)\, x(d),\ y(d))\, d\varrho.$$

This decomposition is not unique.

For more detailed information on unitary representations, trigonometric polynomials, etc. the reader is referred to Neumark (1959) and Godement (1948).

6.2. Let G be a locally compact group with the set R of all non-equivalent, irreducible, unitary representations and with a regular, normed, $P(G) = 1$, Borel measure P. If z is an arbitrary element of $\mathcal{H}$, the vector $U(g)z$ describes a continuous curve when g runs through the group. Furthermore this curve is contained in the sphere with 0 as origin and radius $\|z\|$. The integral

$$\int_G U(g)\, z\, dP(g) = \varphi z$$

then exists in the sense of Bochner. (Note that if we use the regular representation φ coincides with the operator T studied in Part 3, and $\mathcal{H} = L_2(G)$.) It is clear that $\|\varphi z\| \leqslant \|z\|$ so that φ is a bounded linear operator in $\mathcal{H}$. We shall call φ the *Fourier transform* of P and it will sometimes be denoted by $\varphi(r)$ or $\varphi^P(r)$, $r \in R$, for the sake of clearness. We shall describe some simple and fundamental properties of the Fourier transform in the following statements (a)–(f).

(a) *$\varphi(r)$, $r \in R$, is a linear operator in $\mathcal{H}$; it reduces to I if r is the identity representation; its norm is at most one, $\|\varphi\| \leqslant 1$; if $\|\varphi z\| = \|z\|$ for $r \neq I$ and some non-trivial z then the support of P is contained in a coset of a proper subgroup of G.*

The three first statements are obvious. The fourth one can be proved as follows, using a coordinate free version of an idea applied to compact groups by Ito and Kawada (1940) and Stromberg (1960). If there is an element $z \in \mathcal{H}$ such that $\|\varphi z\| = \|z\|$, that is

$$\left\| \int_G U(g)\, z\, dP(g) \right\| = \|z\|,$$

then we must have $U(g)\, z = e^{i\alpha} z$ for all $g \in s(P)$. Introduce the sets

$$A_\alpha = \{g \mid U(g)\, z = e^{i\alpha} z\}.$$

It is clear that A_0 is a closed proper subgroup (note that $U(g)$ should be irreducible) and that $A_\alpha = \gamma A_0$ where γ is an element in A_α. This proves that $s(P) \in \gamma A_0$ as stated. A consequence of this is that, if $\varphi z = z$ occurs, we can restrict our attention to a subgroup of G; of course, we will use as our domain the smallest closed subgroup spanned by $s(P)$.

(b) *$\{\varphi^P, r \in R\}$ determines P uniquely.*

Proof. Suppose the two measures P_1 and P_2 are not identical but have the same Fourier transform. Then

$$\int_G U(g)\, d\, Q(g) = 0$$

for all $r \in R$; we have put $Q = P_1 - P_2$. Let $p(g)$ be an *elementary positive definite* function on G. It can then be represented as $p(g) = (U(g)z, z)$, where $U(g)$ is an irreducible, unitary representation of G. But then

$$\int_G p(g)\, d\, Q(g) = 0,$$

which implies $Q = 0$ (see the previous section).

(c) *Let P^- denote the probability distribution of g^{-1}, $P^-(E) = P(E^{-1})$. Then the adjoint $(\varphi^P)^*$ is equal to the Fourier transform φ^{P^-} of P^-. In particular φ is self-adjoint if and only if P is a symmetric measure. It is normal if and only if $P^- \ast P = P \ast P^-$.*

Proof. We have the obvious relations

$$\varphi^{P^-} = \int_G U(g)\, d\, P^-(g) = \int_G U(g^{-1})\, d\, P(g) = \int_G U^*(g)\, d\, P(g) = (\varphi^P)^*.$$

The last two statements follow from the uniqueness property (b). Unfortunately this makes application of spectral theory difficult except in special cases.

(d) *If $P = P_1 \ast P_2$ then $\varphi^P = \varphi^{P_1} \varphi^{P_2}$.*

This is proved just as on the real line.

(e) *If a sequence of probability measures P_n converges weakly to P then φ^{P_n} converges strongly to φ^P.*

This statement is known (see R. Godement [1]). In order that the Fourier transform should be really useful for the study of limit theorems we would need some sort of converse of (e). A solution to this problem will be given *for groups of type S*: the constant function 1 can be uniformly approximated on every compact subset G by positive definite functions vanishing outside compact sets.

To find positive definite functions approximating to 1 in the way described, we could try functions of the form $c \ast \tilde{c}(g)$ or more particularly functions of the form

$$p(g) = \frac{1}{\mu(C)} I_C \ast \tilde{I}_C(g),$$

where $I_c(g)$ is the indicator function of a compact set C and $\tilde{c}(g) = \bar{c}(g^{-1})$.

(f) *Given probability measures $P_1, P_2, \ldots$ and P on a locally compact group G of type S. If the Fourier transforms $\varphi^{P_1}, \varphi^{P_2}, \ldots$ converge strongly to φ^P, then P_n converges weakly to P.*

Proof. We can always choose a subsequence P_{n_ν} converging weakly to a measure Q, $Q(G) \leqslant 1$. To show that $Q(G) = 1$ we consider a positive definite function $p(g)$ vanishing outside of a compact set C. The unitary representation $U(g)$ corresponding to $p(g)$ can be decomposed into irreducible representations $U(g) = \int_D \oplus U^f(g)$. But for almost every $f \in \mathcal{D}$ we have

$$\int_G U^f(g)\, d\, P_{n_\nu}(g) \to \int_G U^f(g)\, d\, P(g)$$

with strong convergence. Hence we get

$$\int_G p(g)\, d\, Q(g) = \lim_{\nu\to\infty} \int_G p(g)\, d\, P_{n_\nu}(g) = \int_G p(g)\, d\, P(g).$$

But now we can approximate the function 1 by functions like $p(g)$ so that we must have $Q(G) = P(G) = 1$.

To complete the proof we observe that $Q = P$ and since this will hold for any convergent subsequence the result follows.

6.3. Now let us return for a moment to the group of linear transformations of the real line. Consider for simplicity the probability distribution P over G with all its mass on $\alpha = \frac{1}{2}$ and a distribution D for the values of β. Then the Fourier transform associated with the $U^+(g)$ representation takes the form

$$\varphi f(\lambda) = D(\lambda) f\left(\frac{\lambda}{2}\right) \sqrt{\tfrac{1}{2}} = \mathcal{D}\mathcal{J},$$

where $D(\lambda)$ is the characteristic function belonging to D, and where the operators $\mathcal{D}$ and $\mathcal{J}$ are defined by

$$\left.\begin{aligned} \mathcal{D} f(\lambda) &= D(\lambda) f(\lambda) \\ \mathcal{J} f(\lambda) &= f\left(\frac{\lambda}{2}\right)\sqrt{\tfrac{1}{2}} \end{aligned}\right\},$$

To show that this group is of type S let us consider the indicator function $c(g)$ of the compact set

$$S = \left\{ g \,\middle|\, \frac{1}{B} < \alpha < B, \quad -A < \beta < A \right\}.$$

The convolution of $c(g)$ with $\tilde{c}(g)$ with respect to right invariant measure ν, $d\nu(\alpha, \beta) = d\alpha d\beta/\alpha$ is clearly a positive definite function. Introduce the normed positive definite function $p(g) = \frac{1}{\nu(S)}\, c \ast \tilde{c}(g)$, $p(e) = 1$. For any fixed g it can be shown that $p(g) \to 1$ when $A \to \infty$ with $B = 0(A)$. This proves the assertion.

Multiplying independent stochastic group elements $g_1, g_2, \ldots$ having the same distribution P, we get the product $\gamma_n = g_1 g_2 \ldots g_n$. The Fourier transform of the distribution of γ_n is then

$$\varphi^n f = \underbrace{\mathfrak{D}\mathfrak{J}\mathfrak{D}\mathfrak{J}\ldots\mathfrak{D}\mathfrak{J}}_{n \text{ times}} f.$$

If h is the group element with $\alpha = 2$, $\beta = 0$ the Fourier transform of $\gamma_n h^n$ is

$$\Phi_n f(\lambda) = \mathfrak{D}\mathfrak{J}\mathfrak{D}\mathfrak{J}\ldots\mathfrak{D}\mathfrak{J}\mathfrak{J}^{-n} f(\lambda)$$
$$= \varphi(\lambda)\,\varphi\left(\frac{\lambda}{2}\right)\cdots\varphi\left(\frac{\lambda}{2^n}\right) f(\lambda).$$

If the mean value of D exists, the infinite product

$$\Phi(\lambda) = \prod_{n=0}^{\infty} \varphi\left(\frac{\lambda}{2^n}\right)$$

converges and $\Phi_n f(\lambda) \to \Phi(\lambda) f(\lambda) = \Phi f(\lambda)$. It is not difficult to show that Φ is the Fourier transform of a probability distribution Q. Hence the stochastic group element $\gamma_n h^n$ converges distribution-wise.

While this example is very simple and can be treated by a direct method (see Part 4) it may give some hints of what can be expected in more complicated situations.

6.4. In a quite general stochastic group we can obviously not have an analogue of the law of large numbers. On the real line (or in R^k or in a Banach space) the law of large numbers tells us that $(1/n)x_1 + (1/n)x_2 + \cdots + (1/n)x_n$ converges in some probabilistic sense. For a general group we do not necessarily have operations corresponding to multiplications by the factor $1/n$. In order to get any further in this direction we must therefore *assume that n-th roots are uniquely defined on G*: for any group element g there is one and only one element γ such that $\gamma^n = g$; we then write $\gamma = g^{1/n}$. Such groups are sometimes called divisible R-groups.

Then we can speak of the powers g^r where r is any rational number. Assuming that $g^r \to e$ if $r \to 0$, we can extend the defnition to arbitrary exponents. We will therefore start from the following

Assumption. To any real t and element $g \in G$ there is an element $g^t \in G$ with the following properties:

(i) $g^0 = e$, $g^1 = g$,
(ii) g^t is a continuous function of g and t,
(iii) $g^{t+s} = g^t g^s$.

Our first task is to define in an adequate way the mean value of a probability distribution over G. To do so let us note that for a fixed g and $(\mathcal{H}, U(g))$ the operators

$$V_t = U(g^t), \quad -\infty < t < \infty,$$

form a continuous group of unitary transformations. According to a well-known theorem of Stone, there then exists a resolution of the identity, $E_g(\lambda)$; $-\infty < \lambda < \infty$, such that

$$V_t = \int_{-\infty}^{\infty} e^{i\lambda t}\, d E_g(\lambda).$$

We can then write
$$U(g^t) = \exp itH_g,$$

where H_g is the self-adjoint, possibly unbounded operator associated with the spectral representation

$$H_g = \int_{-\infty}^{\infty} \lambda\, d\, E_g(\lambda).$$

If $g_1, g_2, \ldots, g_n$ are stochastically independent elements from our group, let us form the "average" $\gamma_n = x_1^{1/n} x_2^{1/n} \ldots x_n^{1/n}$; note that this "average" in general depends upon the order of the g's. The Fourier transform of γ_n is

$$\varphi_n = E\, U(\gamma_n) = [E\, U^{1/n}(g)]^n,$$

and we have to study its behavior for large values of n.

Heuristics. Since approximately

$$U^{1/n}(g) \cong I + \frac{i}{n} H(g),$$

we have
$$E\, U^{1/n}(g) \cong I + \frac{i}{n} H,$$

where the new operator H is defined by

$$H = \int_G H(g)\, dP(g).$$

As n tends to infinity we should then expect that

$$\varphi_n \to \exp iH,$$

and if there is an element γ such that $U(\gamma) = \exp iH$ for all irreducible, unitary representations, then γ_n converges in probability, to the fixed element γ. This leads us to define mean value as follows.

Definition. *Suppose that for $z \in D$, where D is everywhere dense in $\mathcal{H}$, the operators $H(g)$ are defined and that*

$$\int_G \| H(g) z \| \, dP(g) < \infty;$$

then the operator
$$H = \int_G H(g)\, dP(g)$$

is defined in $\mathcal{D}$. If there exists an element $\bar{g} \in G$ such that

$$U(\bar{g}) = \exp iH$$

for all unitary, irreducible representations $\{\mathcal{H}, U(g)\}$, *then* $\bar{g}$ *is said to be the* mean value *of the stochastic group.*

Because of the lack of commutativity we should not expect too much similarity to the ordinary mean value operation. The following properties are easily proved though.

(i) If P has all its mass in g_0, then its mean value is g_0.

(ii) Let g be the element of a stochastic group; then

$$\overline{(g^t)} = (\bar{g})^t.$$

(iii) If the two operators $H_i = \int H(g)\, d\, P_i(g)$, $i = 1, 2$, are such that they commute, have mean values g_i and $p_1 + p_2 = 1$, $p_i \geqslant 0$, then the mean value of the distribution $p_1 P_1 + p_2 P_2$ is $g_1^{p_1} g_2^{p_2}$.

In the definition of mean value we could have used instead the defining relation $H(\bar{g}) = H$, which should be valid for any $H(g)$ associated with an irreducible, unitary representation.

To transform the above heuristic discussion into a theorem, one must impose some conditions on the stochastic group. This is done below, but the author suspects that the theorem holds in much greater generality than our very restrictive conditions might lead one to believe. To remove these restrictions seems to be an important task in future work on stochastic groups of this type.

Theorem. *Let there be given an increasing sequence of subspaces* $\mathcal{D}_1 \subset \mathcal{D}_2 \subset \mathcal{D}_3 \ldots \subset \mathcal{H}$, *together forming an everywhere dense subset of* $\mathcal{H}$. *Suppose that for* $z \in \mathcal{D}_m$ *the element* $H(g_1) H(g_2) \ldots H(g_p) z$ *is defined and of a norm of the order* $O(c_m^p)$. *Here* c_m^p *is a constant and the statement should hold for all positive and integral* m *and* p.

The operator

$$H = \int_G H(g)\, d\, P(g)$$

is then well defined in $\mathcal{D}_n$. *Suppose that there is a group element* $\bar{g}$ *(which is then the uniquely defined mean value of the stochastic group) such that* $U(\bar{g}) = \exp i H$. *Then the average*

$$\gamma_n = g_1^{1/n}\, g_2^{1/n} \ldots g_n^{1/n}$$

converges in probability to $\bar{g}$.

The strong assumptions make the proof of this convergence quite easy for us. Let $z \in \mathcal{D}_m$; then writing

$$\varphi_n = \left(I + \frac{i}{n} H + \Delta_n\right)^n,$$

we have

$$\varphi_n z = \left(I + \frac{i}{n} H\right)^n z + \varrho_n$$

with

$$\|\varrho_n\| \leqslant \sum_{p=0}^{n-1} \binom{n}{p} \left(1 + \frac{c}{n}\right)^p \frac{c^{n-p}}{n^{2(n-p)}} \to 0$$

as $n \to \infty$. But

$$\left(I+\frac{i}{n}H\right)^n z = \exp iHz + \sum_{p=0}^{n}\left[\binom{n}{p}\frac{1}{n^p}-\frac{1}{p!}\right](iH)^p z - \sum_{n+1}^{\infty}\frac{(iH)^p}{p!}z$$

and both the two sums have norms tending to zero. Hence

$$\varphi_n z \to \exp iHz$$

for z's forming an everywhere dense set in $\mathcal{H}$. But the operators φ_n and $\exp iH$ are of bounded norm (at most one), so that convergence holds throughout $\mathcal{H}$. Now we just have to appeal to (f) of 6.2 not forgetting, of course, that the group should be of the type described in (f).

6.5. Let us return to our example. It is easy to verify that nth roots exist and have the properties required. For an element $g=(\alpha,\beta)\in G$ we have $g^t=(\alpha_t,\beta_t)$ with

$$\left.\begin{aligned}\alpha_t&=\alpha^t\\ \beta_t&=\beta\frac{1-\alpha^t}{1-\alpha}\end{aligned}\right\}$$

if $\alpha\neq 1$ and

$$\left.\begin{aligned}\alpha_t&=1\\ \beta_t&=\beta t\end{aligned}\right\}$$

if $\alpha=1$.

To determine the operators $H(g)$ (see last section), we consider the group of unitary representations, t real,

$$U^+(g^t)f(\lambda)=\exp\left(i\lambda\beta(1-\alpha^t)/(1-\alpha)\right)f(\lambda\alpha^t)\,\alpha^{t/2}.$$

To find the infinitesimal operator of this group, we study the above expression for small values of t and obtain

$$iH(g)f(\lambda)=\left[i\lambda\beta\frac{\log\alpha}{\alpha-1}+\frac{\log\alpha}{2}\right]f(\lambda)+\lambda\log\alpha f'(\lambda),$$

say for $f\in\mathcal{D}=$ the set of all functions vanishing outside of finite intervals together with their derivative.

Let P be a probability measure over G such that α and β are independent and

$$\int_G|\beta|\,dP(g)<\infty,\quad \int_G|\log\alpha|\,dP(g)<\infty.$$

Introduce

$$\left.\begin{aligned}\int_G\log\alpha\,dP(g)&=a\\ \int_G\frac{\log\alpha}{\alpha-1}\,dP(g)&=b\end{aligned}\right\}$$

so that

$$i\int_G H(g)\,dP(g)=ibM+\frac{a}{2}I+aMD,$$

where M is the operator consisting of multiplication by λ and D is the differentiation operator. Then it is easily seen that

$$i \int_G H(g)\, d P(g) = i H(\gamma),$$

if we choose $\gamma = (\xi, \eta)$ as

$$\left.\begin{aligned} \xi &= e^a \\ \eta &= b \frac{e^a - 1}{a} \end{aligned}\right\}$$

for $a \neq 0$ and as

$$\left.\begin{aligned} \xi &= 1 \\ \eta &= b \end{aligned}\right\}$$

for $a = 0$. This element γ is hence the mean value of G with respect to the given distribution.

For this simple case we can verify directly that the average converges to the mean value γ. We have, putting $\gamma_n = (\alpha^{(n)}, \beta^{(n)})$, $g_\nu = (\alpha_\nu, \beta_\nu)$.

$$\begin{cases} \alpha^{(n)} = (\alpha_1 \alpha_2 \dots \alpha_n)^{1/n} \\ \beta^{(n)} = \beta_1 \dfrac{1 - \alpha_1^{1/n}}{1 - \alpha_1} + \beta_2 \dfrac{1 - \alpha_2^{1/n}}{1 - \alpha_2} \alpha^{1/n} + \dots + \end{cases}$$

$$+ \beta_n \frac{1 - \alpha_n^{1/n}}{1 - \alpha_n} \alpha_1^{1/n} \alpha_2^{1/n} \dots \alpha_{n-1}^{1/n}.$$

It is immediately clear that $\alpha^{(n)} \to \xi = e^a$ in probability. Put

$$\begin{cases} b_1^{(n)} = \sum_1^n \beta_\nu \dfrac{1 - \alpha_\nu^{1/n}}{1 - \alpha_\nu} \\ b_2^{(n)} = \dfrac{1}{n} \sum_1^n \beta_\nu \dfrac{\log \alpha_\nu}{\alpha_\nu - 1}. \end{cases}$$

Of course, $b_2^{(n)} \to b$ in probability. But

$$b_1^{(n)} - b_2^{(n)} = \sum \beta_\nu \frac{\exp(1/n \log \alpha_\nu) - 1 - \frac{1}{n} \log \alpha_\nu}{\alpha_\nu - 1}$$

so that $$E\,|b_1^{(n)} - b_2^{(n)}| \leqslant E\,|\beta|\, n\, E \left| \frac{\exp(1/n \log \alpha) - 1 - \frac{1}{n} \log \alpha}{(\alpha - 1)} \right| \to 0$$

so that $b_1^{(n)} \to b$ in probability. To complete the argument, we split up the sum defining $\beta^{(n)}$ into many long blocks, such that the factors $(\alpha_1 \alpha_2 \dots \alpha_\nu)^{1/\nu}$ are nearly constant in each block, $\cong \exp(\nu / n a)$, and apply the above to each block; this proves the convergence statement.

We have seen in Part 4 that for this particular group we also know that another "average", $(g_1 g_2 \ldots g_n)^{1/n}$, converges distribution wise. The limit is not a constant element but a non-degenerate distribution. This leads us further to ask if such a convergence can be proved more generally, but the author has not succeeded in doing this so far.

6.6. Let us study a limit problem with a different norming. If the mean value of P is γ, the "reduced" distribution $Q = \frac{1}{2}P + \frac{1}{2}\varepsilon_{\gamma^{-1}}$ has mean value zero. By $\varepsilon_{\gamma^{-1}}$ we mean the degenerate distribution assigning the probability 1 to the element γ^{-1}. This follows from what was said in 6.4 and the fact that H, of course, commutes with $-H$.

Let $h_1, h_2, \ldots, h_n$ be independent, stochastic group elements drawn from the "reduced" distribution Q. Introduce the normed variable

$$\delta_n = h_1^{1/\sqrt{n}}\, h_2^{1/\sqrt{n}} \ldots h_n^{1/\sqrt{n}}.$$

We have a version of the central limit theorem.

Theorem. *Assume the same conditions as in* **6.4** *and that there is a distribution* Π *such that its Fourier transform*

$$\varphi^{\Pi} = \int_G U(g)\, d\,\Pi(g) = \exp -\tfrac{1}{2} H_2,$$

where
$$H_2 = \int_G H^2(g)\, d\,Q(g).$$

Then the distribution of δ_n *converges to* Π *as* n *tends to infinity.*

The proof is carried out almost in the same way as for the previous theorem; we now have the Fourier transform for the distribution of δ_n

$$(E_Q\, U^{1/\sqrt{n}}(g))^n = \left(I - \frac{H^2}{2n} + B_n\right)^n,$$

and we get the desired result by expansion.

As an illustration we study what happens in our example. For simplicity let us deal with the case where α and β are independent and $E\beta = 0$. Then the mean value $\gamma = e$. Further let us put

$$\begin{cases} E(\log \alpha)^2 = 1, \\ E\,\beta^2 \left(\dfrac{\log \alpha}{\alpha - 1}\right) = c < \infty. \end{cases}$$

We have

$$-H^2(g)\, f(\lambda) = \left\{ i\lambda\beta \frac{\log\alpha}{\alpha-1} + \frac{\log\alpha}{2} + \lambda \log\alpha\, D \right\} \left\{ i\lambda\beta \frac{\log\alpha}{\alpha-1} + \frac{\log\alpha}{2} + \lambda\log\alpha\, D \right\} f(\lambda)$$

$$= A(\alpha,\beta)\, f(\lambda) + B(\alpha,\beta)\, f'(\lambda) + C(\alpha,\beta)\, f''(\lambda)$$

with
$$\begin{cases} A(\alpha,\beta) = (\log\alpha)^2\left[-\dfrac{\lambda^2\beta^2}{(\alpha-1)^2}+\dfrac{i\lambda\beta}{\alpha-1}+\tfrac{1}{4}\right], \\ B(\alpha,\beta) = (\log\alpha)^2\left[\dfrac{2i\lambda^2\beta}{\alpha-1}+2\lambda\right], \\ C(\alpha,\beta) = \lambda^2(\log\alpha)^2. \end{cases}$$

Hence
$$-H_2 = -\int_G H^2(g)\,dP(g) = AI + BD + CD^2,$$

where
$$\begin{cases} A = a_2\lambda^2 + a_1\lambda + a_0 \\ B = b_2\lambda^2 + b_1\lambda \\ C = c_2\lambda^2 \end{cases}$$

with
$$\begin{cases} a_0 = \tfrac{1}{4}E(\log\alpha)^2 \\ a_1 = i\,E(\log\alpha)^2\,\dfrac{\beta}{\alpha-1} \\ a_2 = -E(\log\alpha)^2\left(\dfrac{\beta}{\alpha-1}\right)^2 \\ b_1 = 2E(\log\alpha)^2 \\ b_2 = 2i\,E(\log\alpha)^2\,\dfrac{\beta}{\alpha-1} \\ c_2 = E(\log\alpha)^2 \end{cases}$$

so that with our choice of constants

$$\begin{cases} A = -c\lambda^2 + \tfrac{1}{4} \\ B = 2\lambda \\ C = \lambda^2. \end{cases}$$

For a sufficiently well-behaved $f(\lambda)$ we put

$$f(\lambda,t) = \exp\left(-\frac{t}{2}H_2\right)f(\lambda)$$

so that we have the parabolic equation

$$\frac{\partial f}{\partial t} = \tfrac{1}{8}f - \frac{c}{2}\lambda^2 f + \tfrac{1}{2}(\lambda^2 f')'.$$

The operator $f(\lambda)\to f(\lambda,1)$ should be expressed in terms of the unitary representations to give the required limit distribution.

However, in this simple case we know, with a minor modification of what was done in Part 5, that the limit distribution of $g_1^{1/\sqrt{n}} g_2^{1/\sqrt{n}} \ldots g_\nu^{1/\sqrt{n}}$, $t = \nu/n$, is governed by the equation

$$\frac{\partial p}{\partial t} = \tfrac{1}{2} \frac{\partial^2}{\partial \alpha^2} (\alpha^2 p) + c \, \tfrac{1}{2} \frac{\partial^2}{\partial \beta^2} (\alpha^2 p) - \tfrac{1}{2} \frac{\partial}{\partial \alpha} (\alpha p) = L p;$$

it should be noted that the first order term corresponding to the infinitesimal mean value does not vanish. Introduce the Fourier transforms

$$\varphi_t = \int_G U(g) \, d p(g, t),$$

which obviously form a semigroup with

$$\lim_{t \downarrow 0} \frac{\varphi_t - I}{t} f(\lambda) = \int L^* U(g) f(\lambda) \, d \varepsilon(g),$$

where $\varepsilon(g)$ is the probability measure with all its probability in the unit element (1,0). But this reduces to the same second order differential operator as given above for $\partial f / \partial t$. This verifies that the limit distribution is the one stated.

REFERENCES

R. Godement, Les fonctions de type positif et la théorie des groupes. Trans. Amer. Math. Soc., 63, 1948.

U. Grenander, Stochastic groups. Ark. Mat., 4, no. 12 and no. 14, 1960.

U. Grenander, Stochastic groups and related structures. The Fourth Berkeley Symposium on Mathematical Statistics and Probability. To appear.

K. Ito and Y. Kawada, On the probability distribution on a compact group. Proc. Phys. Math. Soc., 22, 1940.

M. A. Neumark, Normierte Algebren. VEB Deutscher Verlag der Wissenschaften, Berlin 1959.

K. Stromberg, Probabilities on a compact group. Trans. Amer. Math. Soc., 94, no. 2, 1960.

Tryckt den 25 maj 1961

Uppsala 1961. Almqvist & Wiksells Boktryckeri AB