

## Stochastic groups

By ULF GRENANDER

### Part 4. A particular stochastic group

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 for a full statement of the problems and for terminology and notation.

**4.1.** After the general discussion of stochastic groups and algebras it may be wise to turn to specific groups in the hope of finding information of value for further work along the lines discussed previously. Such a group should be non-pathological and sufficiently simple to be amenable to analysis; on the other hand it should possess such properties that are thought typical of stochastic groups in a more general context. We should not allow the group to be Abelian, which would make Fourier analysis possible, or compact, where an analytical tool is available in the irreducible matrix valued representations.

A reasonable choice for our purpose would be a low-dimensional Lie group: we will study the well known group  $G$  consisting of the matrices

$$x = \begin{pmatrix} 1 & x_1 \\ 0 & x_2 \end{pmatrix}, \quad x_2 > 0.$$

The composition rule is for

$$y = \begin{pmatrix} 1 & y_1 \\ 0 & y_2 \end{pmatrix}, \quad z = \begin{pmatrix} 1 & z_1 \\ 0 & z_2 \end{pmatrix},$$

given by

$$z = xy = \begin{pmatrix} 1 & y_1 + x_1 y_2 \\ 0 & x_2 y_2 \end{pmatrix},$$

and the multiplication is not commutative. The unit element  $e$  of  $G$  corresponds to  $x_1 = 0$ ,  $x_2 = 1$ .  $G$  is identical with the open upper halfplane of  $R^2$ , so that it is not compact.

The right invariant Haar measure  $\mu$  of  $G$  is absolutely continuous with respect to Lebesgue measure  $m(x)$  in the halfplane, and the density is with an arbitrarily chosen norming constant

$$\frac{d\mu(x)}{dm(x)} = \frac{1}{x_2^2}.$$

In the upper halfplane  $x_2 > 0$  a probability distribution is defined; we will assume that  $x_1$  and  $x_2$  are independent with certain regularity assumptions that will be specified later on. The marginal distribution of  $\xi_1$  will be denoted by  $P_1$  with the characteristic function  $\varphi(z)$ . The distribution over  $G$  need not be symmetric. Putting

$$\xi^{(n)} = \begin{pmatrix} 1 & \xi_1^{(n)} \\ 0 & \xi_2^{(n)} \end{pmatrix} = x^{(n)} x^{(n-1)} \dots x^{(1)}$$

we are going to study the asymptotic distributions of  $\xi_1^{(n)}$  and  $\xi_2^{(n)}$  for large values of  $n$ .

**4.2.** Given a probability measure  $P$  on  $G$ , what will be the behaviour of the convolutions  $P^{n*}$  for large values of  $n$ ? From what has been said in Part 3 it is clear that we cannot expect convergence towards a limit distribution, but perhaps this can be attained by a suitable norming of the distributions.

However that may be, one thing is obvious, namely that the marginal distribution of  $\xi_2^{(n)}$  after  $n$  convolutions is asymptotically *lognormal*. Indeed,

$$\log \xi_2^{(n)} = \sum_1^n \log x_2^{(v)},$$

which is asymptotically normal with mean value  $n E \log x_2$  and variance  $n \text{Var}(\log x_2)$ , assuming only that these moments exist. Hence we can leave  $\xi_2$  aside and concentrate our attention on  $\xi_1$  and the asymptotic behavior of its marginal distribution. More completely, we may wish to find its asymptotic conditional distribution for given  $\xi_2$ .

It will be easier to understand the problem if we start by studying two special distributions over  $G$ . In the first one we let  $\xi_2$  take the two values 1 and  $1/2$ , say, with the same probability,  $1/2$ .

Now we can write

$$\xi_1^{(n)} = x_1^{(n)} + x_2^{(n)} x_1^{(n-1)} + x_2^{(n)} x_2^{(n-1)} x_1^{(n-2)} + \dots + x_2^{(n)} x_2^{(n-1)} \dots x_2^{(2)} x_1^{(1)},$$

which in the present case reduces to

$$\xi_1^{(n)} = x_1^{(n)} + 2^{-\varepsilon_n} x_1^{(n-1)} + 2^{-\varepsilon_n - \varepsilon_{n-1}} x_1^{(n-2)} + \dots + 2^{-\varepsilon_n - \varepsilon_{n-1} - \dots - \varepsilon_1} x_1^{(1)},$$

where the  $\varepsilon_v$  are stochastic variables taking the values 1 and 0 with the probabilities  $1/2$ . To study the distribution of  $\xi_1^{(n)}$  for large values of  $n$  consider the infinite series

$$y = \sum_{\nu=0}^{\infty} y_{\nu} 2^{-s_{\nu}},$$

where the  $y_{\nu}$  are independent stochastic variables with the distribution of  $\xi_1$ . The  $s_{\nu}$  are  $s_{\nu} = \delta_1 + \delta_2 + \dots + \delta_{\nu}$ ,  $s_0 = 0$ , and the  $\delta_{\nu}$  have the same distribution as  $\varepsilon_{\nu}$ . Assuming that  $E|\xi_1| < \infty$  the above series converges almost certainly, since

$$\sum_{\nu=0}^{\infty} E|y_{\nu} 2^{-s_{\nu}}| = E|\xi_1| \sum_{\nu=0}^{\infty} \left(\frac{1}{2}\right)^{\nu} < \infty.$$

From this reasoning it is also clear that the distribution of  $\xi_1^{(n)}$  converges to that of  $y$  as  $n$  increases indefinitely. The characteristic function  $\varphi_y(z)$  of the latter can be found conveniently by introducing the stochastic variables  $k_1, k_2, \dots$  so that

$$\begin{cases} \delta_1 = \delta_2 = \dots = \delta_{k_1-1} = 0, & \delta_{k_1} = 1 \\ \delta_{k_1+1} = \delta_{k_2+2} = \dots = \delta_{k_1+k_2-1} = 0, & \delta_{k_1+k_2} = 1 \\ \text{etc.} \end{cases}$$

It is well known that the  $k$ 's are independent and have a geometric distribution

$$P(k = \nu) = \frac{1}{2^\nu}, \quad \nu = 1, 2, \dots$$

But we can write

$$y = (y_0 + y_1 + \dots + y_{k_1-1}) + (y_{k_1} + \dots + y_{k_1+k_2-1}) \cdot \frac{1}{2} + (y_{k_1+k_2} + \dots + y_{k_1+k_2+k_3-1}) \cdot \frac{1}{4} + \dots$$

so that

$$\begin{aligned} \varphi_y(z) &= E \exp i y z = E \varphi^{k_1}(z) \varphi^{k_2}\left(\frac{z}{2}\right) \varphi^{k_3}\left(\frac{z}{4}\right) \dots \\ &= \sum_{k_1=1}^{\infty} \left[\frac{1}{2} \varphi(z)\right]^{k_1} \cdot \sum_{k_2=1}^{\infty} \left[\frac{1}{2} \varphi\left(\frac{z}{2}\right)\right]^{k_2} \cdot \sum_{k_3=1}^{\infty} \left[\frac{1}{2} \varphi\left(\frac{z}{4}\right)\right]^{k_3} \dots = \\ &= \frac{\frac{1}{2} \varphi(z)}{1 - \frac{1}{2} \varphi(z)} \cdot \frac{\frac{1}{2} \varphi(z/2)}{1 - \frac{1}{2} \varphi(z/2)} \cdot \frac{\frac{1}{2} \varphi(z/4)}{1 - \frac{1}{2} \varphi(z/4)} \dots \end{aligned}$$

so that

$$\varphi_y(z) = \prod_{\nu=0}^{\infty} \frac{\varphi(2^{-\nu} z)}{2 - \varphi(2^{-\nu} z)}$$

It need scarcely be said that the convergence of  $\xi_1^{(n)}$  is only distributionwise; it does not converge to any stochastic variables as can be seen from the fact that the leading terms in its expansion are independent of what happened before  $n, n-1$ , and so on.

We should also take a look upon the asymptotic distribution of  $\xi_1^{(n)}$  for given  $\xi_2^{(n)}$ . But this is immediately done if we remember that  $\xi_2^{(n)} = 2^{-\varepsilon_n - \varepsilon_{n-1} - \dots - \varepsilon_2} \xi_2^{(1)}$  and that with overwhelming probability the sum  $\varepsilon_2 + \varepsilon_3 + \dots + \varepsilon_n$  will be large. But  $\xi_2^{(n)} = C$  means  $\varepsilon_2 + \varepsilon_3 + \dots + \varepsilon_n = -\log_2 C$  and it is clear that such a condition influences arbitrarily little the joint distribution of any finite number  $m$  of terms  $\varepsilon_n, \varepsilon_{n-1}, \dots, \varepsilon_{n-m+1}$  if  $n$  is sufficiently large. This implies asymptotic independence between  $\xi_1^{(n)}$  and  $\xi_2^{(n)}$ .

**4.3.** The other special distribution will be the same except that we now let  $x_2$  take the two values 1 and 2 with the probabilities  $1/2$ . The variable  $\log \xi_2^{(n)}$  is still asymptotically normal although with a large positive (instead of negative) mean value.

Instead of  $\xi_1^{(n)}$  we will consider

$$\eta^{(n)} = \frac{\xi_1^{(n)}}{x_2^{(n)} x_2^{(n-1)} \dots x_2^{(2)}} = x_1^{(1)} + \frac{1}{x_2^{(2)}} x_1^{(2)} + \frac{1}{x_2^{(2)} x_2^{(3)}} x_1^{(3)} + \dots + \frac{1}{x_2^{(2)} x_2^{(3)} \dots x_2^{(n)}} x_1^{(n)}.$$

But now  $\frac{1}{x_2^{(v)}}$  is distributed just as  $x_2^{(v)}$  in the last section. Hence  $\eta^{(n)}$  converges distribution wise, and its limit distribution is the one given above. In the present case a bit more holds: the sequence of stochastic variables  $\eta^{(1)}, \eta^{(2)}, \dots$  converges almost certainly to the stochastic variable

$$\eta = \sum_{v=1}^{\infty} \frac{1}{x_2^{(2)} x_2^{(3)} \dots x_2^{(v)}} x_1^{(v)}.$$

It follows that for large  $n$

$$\xi_1^{(n)} \sim x_2^{(n)} x_2^{(n-1)} \dots x_2^{(2)} \cdot \eta = \frac{\xi_2^{(n)}}{\xi_2^{(1)}} \eta,$$

and since we know how  $\xi_2^{(n)}$  behaves the last formula answers our question. Of course we could also formulate it in terms of  $\log |\xi_1^{(n)}|$

$$\frac{\log |\xi_1^{(n)}| - n E \log x_2}{\sqrt{n \text{Var} (\log x_2)}} \sim \frac{\log |\xi_2^{(n)}| - n E \log x_2}{\sqrt{n \text{Var} (\log x_2)}} \sim N(0, 1).$$

Since we have taken the absolute value  $|\xi_1^{(n)}|$  the above tells us nothing about the probabilities with which  $\xi_1^{(n)}$  has a positive or negative sign. However this sign is the same as that of  $\eta^{(n)}$ , and this sign is asymptotically independent of  $\xi_2^{(n)}$ ; the probabilities can then be obtained at least in principle, from  $\varphi_y(s)$ .

**4.4.** Let us sum up what we have found in these special cases. The second group coordinate  $\xi_n^{(2)}$  has a log-normal distribution under both assumptions, taking large negative values in the first and large positive values in the second case. The first group coordinate  $\xi_1^{(n)}$  behaves quite differently. When  $x_2$  takes the values 1 and 2 the variable  $\xi_1^{(n)}$  is still asymptotically log-normal (except for its sign), but when  $x_2$  is 1 or  $1/2$  the distribution of  $\xi_1^{(n)}$  converges (distribution wise) without norming. One should also note that its limit distribution depends completely upon  $\varphi(z)$ , the choice of  $P_1$ , in contrast to many classical limit theorems where we have more or less "universal" limit distributions not depending upon a detailed knowledge of the original distributions.

One could argue against this that it is more natural to norm the first coordinate in the same way as the second,  $\frac{\log |\xi_1^{(n)}| - m n}{\sigma \sqrt{n}} = u_n$ . Then the normed variable  $u_n$  will converge to  $N(0, 1)$  in the second and to 0 in the first case, and these distributions do not depend upon what  $P_1$  is. Here we touch upon a basic question in the formulation of limit theorems. Say that we have a sequence of random elements  $x_1, x_2, \dots$ , in same space  $X$ . To norm the distributions of  $x$  we map  $X$  upon some other space  $Y$  via sequence of functions  $y_n = y_n^{(x)}$ , and a limit theorem will state that the distribution of  $y_n(x_n)$  converges as  $n$  tends

to infinity. Suppose now that we can prove for two different  $\{y'_n(x)\}$  and  $\{y''_n(x)\}$  that the two limit theorems hold

$$\left. \begin{aligned} L' : \lim_{n \rightarrow \infty} \text{distribution of } y'_n(x_n) = D' \\ L'' : \lim_{n \rightarrow \infty} \text{distribution } y''_n(x) = D'' \end{aligned} \right\}$$

If  $D'$  is non-atomic and  $D''$  has at least one atom, say  $A$ , then it is practically necessary to consider  $L'$  as superior to  $L''$ . Indeed, the limit theorem  $L''$  lumps some probability in  $A$ , while this probability is resolved by  $L'$ , so that  $L'$  can be said to be more detailed than  $L''$ .

Another possibility would be to consider one limit theorem  $L'$  at least as good as  $L''$  if  $D'$  is absolutely continuous with respect to  $D''$ . This would define a partial ordering among all possible  $L$ .

Considerations of this type are likely to be of consequence for the formulation of limit theorems for more general stochastic structures. In the special case under study just now we must use a formulation with limit distributions depending upon  $P_1$ . This is sad but not wholly unexpected if we recall the discussion of Part 3. There we saw that, while for commutative groups we could use the same analytical apparatus for any distribution  $P$  over the group, the non-commutative stochastic groups would require a more flexible treatment, and it would not be surprising to find that the limit distribution depends more strongly upon  $P$ .

4.5. Now when we know what happens in the two special cases it is easy to study the case of arbitrary distributions  $P_1$  for  $x_1$ ,  $P_2$  for  $x_2$ .

If  $E x_2 < 1$  we can reason just as in 4.2. and we see that for large values of  $n$  the first group coordinate  $\xi_1^{(n)}$  is asymptotically distributed as

$$y = \sum_{\nu=0}^{\infty} y_{\nu} \cdot z_1 \dots z_{\nu},$$

where the  $y_{\nu}$ ,  $z_{\mu}$  are independent and the distribution of  $y_{\nu}$  given by  $P_1$  and the distribution of  $z_{\nu}$  by  $P_2$ . The series converges almost certainly since

$$\sum_{\nu=0}^{\infty} E(|y_{\nu}| z_1 \dots z_{\nu}) = E|y_1| \sum_{\nu=0}^{\infty} (E z_1)^{\nu} < \infty.$$

On the other hand if  $E \frac{1}{x_2} < 1$  we have as before the relation

$$\xi_1^{(n)} = x_2^{(2)} x_2^{(3)} \dots x_2^{(n)} \left[ x_1^{(1)} + \frac{1}{x_2^{(n)}} x_1^{(2)} + \dots + \frac{1}{x_2^{(n)} x_2^{(n-1)} \dots x_2^{(2)}} x_1^{(n)} \right]$$

and we find as before that  $\log |\xi_1^{(n)}|$  has asymptotically a log-normal distribution.

It remains only to investigate the case  $E(x_2) \geq 1$ ,  $E\left(\frac{1}{x_2}\right) \geq 1$ . Introduce the geometric mean value of  $x_n$

$$\gamma = \exp E(\log x_2),$$

and assume to start with that  $\gamma < 1$ . According to the strong law of large numbers we have almost certainly

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \log x_2^{(i)} = \log \gamma < 0.$$

Hence the products  $z_1 z_2 \dots z_n$  defined above will behave asymptotically as  $\gamma^n$ , decreasing exponentially. The series expansion for  $y$  will then have coefficients for  $y^v$  that decrease so fast that the series will converge almost certainly. Hence we can conclude that for  $\gamma < 1$  we get the same limit result for  $E x_2 < 1$ ; see above.

If instead the geometric mean  $\gamma > 1$  we know that the products  $z_1 z_2 \dots z_n$  behave asymptotically as  $\gamma^n$  almost certainly. A slight modification of the previous reasoning shows that we get the same limiting behaviour as for  $E 1/x_2 < 1$ .

The case  $\gamma = 1$  will be left open; it is conjectured that the asymptotic distribution of  $\log |\xi_1^{(n)}|$  will be asymptotically the same as that of  $\max_{1 \leq v \leq n} s_v$ , where the  $s_v$  are the partial sums corresponding to the independent stochastic variables  $\log \xi_2^{(1)}, \log \xi_2^{(2)}, \dots, \log \xi_2^{(n)}$ .

**4.6.** It is interesting to study what happens if the variables  $x_1^{(v)}, x_2^{(v)} - 1$  are made small together with  $1/n$ . In the limiting case, when these variables are made infinitesimal, we can even find the explicit form of the equilibrium distribution when this exists. Since

$$\xi_1^{(v+1)} = x_1^{(v)} + x_2^{(v)} \xi_1^{(v)}, \quad \Delta \xi_1^{(v)} = x_1^{(v)} + (x_2^{(v)} - 1) \xi_1^{(v)}$$

we have

$$\left. \begin{aligned} E [\Delta \xi_1^{(v)} | \xi_1^{(v)} = \xi] &= A_1 + A_2 \xi \\ \text{Var} [\Delta \xi_1^{(v)} | \xi_1^{(v)} = \xi] &= B_1 + B_2 \xi^2 \end{aligned} \right\}$$

where

$$\left. \begin{aligned} E x_1 &= A_1 \\ E X_2 &= 1 + A_2 \\ \text{Var } x_1 &= B_1 \\ \text{Var } x_2 &= C_2 \end{aligned} \right\}$$

and  $x_1$  and  $x_2$  are assumed independent as before. Letting  $n$  tend to infinity while  $A_i = a_i/n$ ,  $B_i = b_i/n$ ,  $i = 1, 2$ , the limiting frequency function  $p(t, \xi)$  will satisfy

$$\frac{\partial p}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial \xi^2} (b_1 + b_2 \xi^2) p - \frac{\partial}{\partial \xi} (a_1 + a_2 \xi) p.$$

In analogy with what we saw in the last section we can expect the equilibrium distribution  $p(\xi)$  to exist when the geometric mean of  $x_2$  is less than, or possibly equal to one. As here  $x_2 = 1 - \varepsilon$ , with  $\varepsilon$  small, we get

geometric mean of  $x_2 = \exp E \log (1 - \varepsilon) \sim \exp \left( -E \varepsilon - \frac{1}{2} E \varepsilon^2 \right) \sim \exp \left( a_2 - \frac{1}{2} \right) \frac{1}{n}$ .

Therefore equilibrium can be expected for  $a_2 < \frac{1}{2}$  and possibly even for  $a_2 = \frac{1}{2}$ .

Assuming for convenience that  $b_1 = b_2 = 1$  we should have, if an equilibrium density  $p(\xi)$  exists,

$$\frac{1}{2} \frac{\partial^2}{\partial \xi^2} (1 + \xi^2) p(\xi) - \frac{\partial}{\partial \xi} (a_1 + a_2 \xi) p = 0$$

and hence

$$\frac{1}{2} \frac{\partial}{\partial \xi} (1 + \xi^2) p - (a_1 + a_2 \xi) p = \frac{1}{2} (1 + \xi^2) p' - (a_1 + (a_2 - 1) \xi) p = C_1.$$

The general solution of this equation is then

$$p(\xi) = (1 + \xi^2)^{a_2+1} \exp 2a_1 \operatorname{arctg} \xi \left[ 2C_1 \int_{\xi_0}^{\xi} (1 + u^2)^{-a_2} \exp (-2a_1 \operatorname{arctg} u) du + C_2 \right].$$

If  $a_2 \geq \frac{1}{2}$  the above expression is not integrable over  $(-\infty, \infty)$  and hence not a frequency function. If  $a_2 < \frac{1}{2}$  the integral inside the brackets behaves like  $\xi^{-2a_2+1}$  so that to get a possible frequency function, we must put  $C_1 = 0$ . Then we obtain the equilibrium distribution

$$p(\xi) = C_2 (1 + \xi^2)^{a_2-1} \exp 2a_1 \operatorname{arctg} \xi.$$

Since  $p'(\xi) = 0$  occurs only for  $\xi = \frac{a_1}{1 - a_2}$  this distribution is unimodal. It is symmetric around  $\xi = 0$  if  $a_1 = 0$ .

**4.7.** For this particular group there is another natural norming that lies close at hand when studying limit distributions. This is via the  $n$ -th roots which exist and are unique in  $G$ . Indeed, let  $x$  be an arbitrary element of  $G$  and find a  $y \in G$  such that  $y^n = x$ ,  $n$  positive integer. In coordinates this relation is, at least for  $y_2 \neq 1$ ,

$$\begin{cases} x_1 = y_1 + y_2 y_1 + y_2^2 y_1 + \dots + y_2^{n-1} y_1 = y_1 \frac{1 - y_2^n}{1 - y_2} \\ x_2 = y_2^n \end{cases}$$

so that there is a unique solution

$$\left. \begin{aligned} y_1 &= \frac{1 - x_2^{1/n}}{1 - x_2} x_1 \\ y_2 &= x_2^{1/n} \end{aligned} \right\}$$

If  $y_2 = 1$  we get instead  $y_1 = \frac{1}{n} x_1$ .

Now let us use the norming

$$\eta^{(n)} = \begin{pmatrix} 1 & \eta_1^{(n)} \\ 0 & \eta_2^{(n)} \end{pmatrix} = (\xi^{(n)})^{1/n}.$$

As  $n \rightarrow \infty$  we have

$$\log \eta_2^{(n)} = \frac{1}{n} \sum_{\nu=1}^n \log x_2^{(\nu)} \rightarrow E \log x_2,$$

assuming only that the mean value on the right side exists. To deal with  $\eta_1^{(n)}$  we separate cases and first assume that the geometric mean value  $\gamma = \text{Geom. mean } x_2 < 1$ . Then, for large values of  $n$  the first group coordinate behaves as

$$\eta_1^{(n)} \sim (1 - \gamma) y \text{ distribution wise}$$

where the stochastic variable  $y$  is the one introduced in 4.5. In the opposite case,  $\gamma > 1$ , we get instead

$$\eta_1^{(n)} \sim (\gamma - 1) \left[ \frac{x_1^{(1)}}{x_2^{(1)}} + \frac{x_1^{(2)}}{x_2^{(1)} x_2^{(2)}} + \dots + \frac{x_1^{(n)}}{x_2^{(1)} x_2^{(2)} \dots x_2^{(n)}} \right]$$

distribution wise.

Hence in both cases the distribution of  $\eta^{(n)} = (\xi^{(n)})^{1/n}$  converges as  $n$  tends to infinity; the first group coordinate to a non-degenerate distribution, the second one to a constant.

### Part. 5. Some stochastic algebras

**5.1.** We have seen in Part 2 that additive stochastic processes in a stochastic Banach algebra induce multiplicative stochastic processes in a natural way, and also that the additive limit theorems that may be available give us corresponding multiplicative limit theorems. The derivations were carried out in the  $L_1$ -topology and hold under certain conditions that are satisfied for several interesting cases, e.g. the Poisson-like situations.

On the other hand these conditions were not formulated so as to suit other important stochastic algebras, e.g. those that correspond to Wiener processes on the real line. To study this we will have to modify the earlier derivations, using now instead the  $L_2$ -topology. Since a good deal of what was said in Part 2 still goes through, we can be quite brief here and refer the reader back to Part 2 for more details. Let us point out that the space dealt with should be an algebra and a Banach-space, but we need not have  $\|xy\| \leq \|x\| \|y\|$  so that it is not necessarily a Banach algebra.

As before we start with an additive, time-homogeneous, stochastic process  $y(t)$ ,  $t \geq 0$ , and try to define a multiplicative process on the same stochastic algebra through the series of integrals in section 2.4. Consider the region  $S_n: 0 < t_1 < t_2 < \dots < t_r < t$  of  $R^r$ . In  $S$ , we will consider rectangles with sides parallel to the coordinate axes, or sums of such rectangles. To any such sum  $\sigma$  we associate a stochastic element  $y(\sigma)$  of the stochastic algebra as we did before;  $y(\sigma)$  is an additive stochastic set function and  $y(R)$ , where  $R$  is a rectangle in  $S$ , with the sides  $I_1, I_2, \dots, I_r$ , is given by  $y(R) = \Delta y(I_1) \Delta y(I_2) \dots \Delta y(I_r)$ , where



the order of multiplication is important. Note that  $y(\sigma)$  is not a stochastic set function with independent increments.

We will need a bound for  $E \|y(\sigma)\|^2$  and it is sufficient for the present purpose to assume that

$$E \|y(\sigma)\|^2 \leq C^r \cdot \text{volume of } \sigma, \tag{1}$$

where  $C$  is some constant and the volume is Lebesgue measure in  $r$ -space. If this is so, it follows almost immediately that the stochastic integrals in question are uniquely defined. Indeed, the difference between different Riemann sums can be expressed through the values of  $y(\sigma)$ , where the elementary rectangles of  $\sigma$  together cover a small fraction only of the volume of  $S_r$ , and this shows uniqueness. The norm of such an integral is also bounded (again limit of Riemann sums) by

$$E \|J_r(t)\|^2 = E \left\| \int_{S_r} dy(t_1) dy(t_2) \dots dy(t_r) \right\|^2 \leq C^r \cdot \text{volume of } S_r = C^r \frac{t^r}{r!}.$$

Then

$$\sum_{r=0}^{\infty} \{E \|J_r(t)\|^2\}^{\frac{1}{2}} \leq \sum_{r=0}^{\infty} \frac{(Ct)^{r/2}}{\sqrt{r!}} < +\infty,$$

so that the series

$$x(t) = \sum_{r=0}^{\infty} J_r(t), \quad J_0(t) = e,$$

converges in the mean.—In a similar way we can also modify the derivation of the multiplicative limit theorems in Part 2.

We now return to condition (1). Although this condition will have to be verified from case to case, we can sometimes simplify it a little. Indeed if the norm is introduced via an inner product  $(x; y)$ , we have for two disjoint rectangles  $R' = \Delta'_1 \times \Delta'_2 \times \dots \times \Delta'_r$  and  $R'' = \Delta''_1 \times \Delta''_2 \times \dots \times \Delta''_r \subset S_r$

$$E(y(R'); y(R'')) = E(y(\Delta'_1) y(\Delta'_2) \dots y(\Delta'_r); y(\Delta''_1) y(\Delta''_2) \dots y(\Delta''_r)).$$

At least for some  $i$  we must have  $\Delta'_i \cap \Delta''_i = 0$ . If we assume that the expected value of the  $y$ -process exists and is zero the above inner product will vanish. In such a case we need only verify condition (1) for  $\sigma =$  an arbitrary rectangle.

Irrespective of the particular form that we have given to condition (1) it is clear that the essential content of any such condition should be some sort of absolute continuity of the set function  $E \|y(\sigma)\|^2$  with respect to Lebesgue volume. It is of some importance to note that this should hold in the *interior* of  $S_r$ . On and around the boundary, say  $t_1 = t_2 = \dots = t_r$ , this need not hold, as is evident already from the scalar case when  $r=2$  and  $y(t)$  is the Wiener process on  $R^1$ . Indeed the stochastic variable

$$y_n = \sum_{i=1}^n [y(t_{i+1}^{(n)}) - y(t_i^{(n)})]^2$$

does not tend to zero as the division is made finer; instead it converges, as is well known, to the variance of  $y(t)$ . Here lies the basic reason why conditions of type (1) are needed.

In this connection it should be pointed out that our construction of the multiplicative  $x$ -process has proceeded via the finite products

$$\prod_{\nu=1}^n (e + \Delta_{\nu} y)$$

evaluated in the appropriate order. We could also have started with

$$\prod_{\nu=1}^n \exp (\Delta_{\nu} y)$$

but the corresponding limiting processes would not be the same in general under the conditions used in this section. While this is only a choice of definition one should be aware of the different results of the two definitions.

**5.2.** There is a simple specialization of the multiplicative processes with values in an algebra that seems to correspond directly to a class of important practical problems. This is when the algebra consists of matrices  $x = \{x_{ij}; i, j = 1, 2, \dots, k\}$ , with addition and multiplication defined in the usual way and with an appropriately chosen norm. For a description and classification of the infinitely divisible distributions corresponding to such processes the reader should consult the paper by Hunt referred to in Part 1.

Two situations will be studied. The first is when the additive matrix-valued process  $y(t)$  is of Poisson type, so that the infinitesimal probability structure is defined by

$$y(t+h) = \begin{cases} y(t) + hA & \text{with probability } 1 - \lambda h + o(h) \\ y(t) + hA + B & \text{with probability } \lambda h + o(h). \end{cases}$$

Let us fix the initial value as  $y(0) = 0$ . As norm of the matrices we can choose e.g. the ordinary norm of the matrices considered as linear operators in  $R^k$ . Since

$$\sum E \|y(t_{\nu}) - y(t_{\nu-1})\| \leq \sum [(t_{\nu} - t_{\nu-1}) (\|A\| + \lambda \|B\| + o(1))]$$

these sums are uniformly bounded and condition (2) of 2.4. is satisfied.

Elementary calculations give us the mean values and covariances of the  $x$ -process. Putting

$$M(t) = \{m_{ij}(t); i, j = 1, 2, \dots, k\} = \{E x_{ij}(t); i, j = 1, 2, \dots, k\}$$

we get

$$M(t+h) = E x(t+h) = M(t) (I + hN + o(h)),$$

where  $N = A + \lambda B$ . Hence  $M'(t) = M(t)N$  and  $M(t) = \exp tN$ . Introduce the  $k^2 \times k^2$  matrix

$$\begin{aligned} C(t) &= \{c_{i\gamma, \alpha\beta}(t); i, j, \alpha, \beta = 1, 2, \dots, k, \text{ say in lexicographic order}\} \\ &= \{\text{cov}[x_{ij}(t), x_{\alpha\beta}(t)]\}. \end{aligned}$$

Denoting the Kronecker product of two matrices  $R$  and  $S$  by  $R \times S$  we have  $C(t) = E x(t) \times x(t)$  and

$$C(t) = \{E x_{ij}(t) x_{\alpha\beta}(t) - m_{ij}(t) m_{\alpha\beta}(t)\} \\ = E x(t) \times x(t) - M(t) \times M(t).$$

Putting  $S(t) = E x(t) \times x(t)$  we get

$$S(t+h) = E x(t+h) \times x(t+h) = E [x(t) + x(t) \Delta y] \times [x(t) + x(t) \Delta y] + \\ + \text{terms of smaller order than } \Delta y$$

where we have put  $\Delta y = y(t+h) - y(t)$ . Then

$$S(t+h) = S(t) + h E x(t) \times x(t) N + h E x(t) N \times x(t) + E x(t) \Delta y \times x(t) \Delta y + \dots$$

so that 
$$\frac{dS(t)}{dt} = S(t) (I \times N) + S(t) (N \times I) + \lambda S(t) (B \times B).$$

We get the desired result

$$S(t) = \exp t \{I \times N + N \times I + \lambda B \times B\}.$$

**5.3.** Now let  $y(t)$  be the Gaussian process of independent increments and taking as values  $k \times k$ -matrices with mean values zero and covariances

$$E y_{ij}(t) y_{\alpha\beta}(t) = t b_{ij, \alpha\beta}.$$

The norm of matrices will have to be introduced through an inner product; if  $U = \{u_{ij}\}$  and  $V = \{v_{ij}\}$  are two  $k \times k$  matrices we will define

$$(U, V) = \sum u_{ij} v_{ij}, \quad \|U\|^2 = (U, U).$$

To be able to verify condition (1) of 5.1. we need in this case only study  $y(R)$  when  $R$  is a rectangle in  $S_r$ . But then

$$E \|y(R)\|^2 = \sum E y_{i_1 i_1}^{(1)} y_{i_1 i_1}^{(2)} \dots y_{i_r i_{r+1}}^{(r)} y_{i_1 i_1}^{(1)} y_{i_1 i_1}^{(2)} \dots y_{i_r i_{r+1}}^{(r)}$$

where the sum is extended over all values between 1 and  $k$  of the  $i$ 's and  $j$ 's, and where  $y_{ij}^{(v)}$  is the  $(i, j)$ -element of the matrix increment of  $y(t)$  over the  $v$ th side of length  $\Delta_v$  of the rectangle  $R$ . Hence

$$E \|y(R)\|^2 = \sum E y_{i_1 i_1}^{(1)} y_{i_1 i_1}^{(1)} E y_{i_1 i_1}^{(2)} y_{i_1 i_1}^{(2)} \dots E y_{i_r i_{r+1}}^{(r)} y_{i_r i_{r+1}}^{(r)} = \\ = O(k^{2r+1} b^r \Delta_1 \Delta_2 \dots \Delta_r) = O(\text{const}^r m(R)),$$

where  $b = \max |b_{ij, \alpha\beta}|$ , so that our condition is satisfied and the corresponding multiplicative process  $x(t)$  is well defined.

The infinitesimal covariances of the  $x(t)$  process are easily calculated as

$$\lim_{h \downarrow 0} \frac{1}{h} \text{conditional Cov} [x_{ij}(t+h) - x_{ij}(t); x_{\alpha\beta}(t+h) - x_{\alpha\beta}(t)] = \\ = \sum_{i, \lambda=1}^k x_{i\lambda}(t) x_{\alpha\lambda}(t) b_{ij, \lambda\beta} = q_{ij, \alpha\beta}(x).$$

U. GRENNANDER, *Stochastic groups*

The  $q_{ij, \alpha\beta}(x)$  are quadratic forms in the  $x$ -components and the matrix  $q$  can be written as  $[x(t) \times x(t)] B$ . The Fokker-Planck equation is then

$$\frac{\partial p}{\partial t} = \frac{1}{2} \sum \frac{\partial^2}{\partial x_{ij} \partial x_{\alpha\beta}} [q_{ij, \alpha\beta}(x) p]$$

to which appropriate initial values should be prescribed, let us say  $x(0) = I$ .

For the case of non-vanishing mean values

$$E y_{ij}(t) = t a_{ij}, \quad A = \{a_{ij}\},$$

the above modifies to

$$\frac{\partial p}{\partial t} = \frac{1}{2} \sum \frac{\partial^2}{\partial x_{ij} \partial x_{\alpha\beta}} [q_{ij, \alpha\beta}(x) p] - \sum \frac{\partial}{\partial x_{ij}} [l_{ij}(x) p]$$

with  $L = \{l_{ij}(x)\}$  and  $L = x(t) A$ , so that the  $l_{ij}^{(x)}$  are linear forms in the  $x_{ij}$ .

To find the expected values of the  $x$ -process a partial integration gives us

$$\frac{\partial}{\partial t} E x_{ij}(t) = \frac{\partial}{\partial t} m_{ij}(t) = E l_{ij}(x)$$

so that

$$\frac{\partial m(t)}{\partial t} = m(t) A, \quad m(t) = \exp t A.$$

For the second order moments  $c_{ij, \alpha\beta}(t)$  we get

$$\frac{\partial}{\partial t} c_{ij, \alpha\beta}(t) = E q_{ij, \alpha\beta}(x) + E l_{ij}(x) x_{\alpha\beta} + E l_{\alpha\beta}(x) x_{ij}$$

so that

$$\frac{\partial}{\partial t} c(t) = c(t) B + c(t) [A \times I + I \times A]$$

and

$$c(t) = \exp t \{B + A \times I + I \times A\}.$$

It will be convenient to drop the double subscripts and use variables denoted  $\xi_\lambda$ ,  $\lambda = 1, 2, \dots, s$ , and a corresponding equation

$$\frac{\partial p}{\partial t} = \frac{1}{2} \sum_{\lambda, \mu} \frac{\partial^2}{\partial \xi_\lambda \partial \xi_\mu} [q_{\lambda\mu}(\xi) p] - \sum_\lambda \frac{\partial}{\partial \xi_\lambda} [l_\lambda(\xi) p].$$

We are sometimes interested in the behavior of the density  $p = p(\xi, t)$  for large values of  $t$ , and especially we would like to get conditions ensuring that the distribution contracts to a singular one (concentrated on some linear subspace of  $R^s$ ), converges to a limit distribution or spreads out more and more thinly over any finite part of  $R^s$ . A criterion useful for this purpose is the quadratic integral

$$I(t) = \int_{R^s} p^2(\xi, t) d\xi = \|p\|^2.$$

Knowledge of how  $I(t)$  behaves gives us some partial information on the overall behavior of  $p$ .

But we have

$$\frac{1}{2} \frac{\partial I(t)}{\partial t} = (p', p) = \frac{1}{2} (Qp, p) - (Lp, p).$$

Here prime denotes differentiation with respect to time and  $Q$  and  $L$  are the two differential operators on the right hand side of the Fokker-Planck equation. Through partial integration

$$\begin{aligned} (Qp, p) &= - \sum_{R^s} \frac{\partial p}{\partial \xi_\lambda} \frac{\partial}{\partial \xi_\mu} (q_{\lambda\mu} p) d\xi = \\ &= - \sum \int_{R^s} q_{\lambda\mu} \frac{\partial p}{\partial \xi_\lambda} \frac{\partial p}{\partial \xi_\mu} d\xi - \sum \int_{R^s} p \frac{\partial p}{\partial \xi_\lambda} \frac{\partial q_{\lambda\mu}}{\partial \xi_\mu} d\xi = -P + \frac{1}{2} \sum_{R^s} \int p^2 \frac{\partial^2 q_{\lambda\mu}}{\partial \xi_\mu \partial \xi_\lambda} d\xi. \end{aligned}$$

But any  $q_{\lambda\mu}$  is a quadratic form in the  $\xi$ 's so that its second order derivatives are constant. The quadratic differential operator  $P$  is non-negative. We also get

$$(Lp, p) = - \sum_{R^s} \int l_\lambda p \frac{\partial p}{\partial \xi_\lambda} d\xi = \frac{1}{2} \sum_{R^s} \int \frac{\partial l_\lambda}{\partial \xi_\lambda} p^2 d\xi,$$

where the first derivatives of  $l_\lambda(\xi)$  are constant. This implies that

$$\frac{\partial I(t)}{\partial t} \leq \left[ \frac{1}{2} \sum \frac{\partial^2 q_{\lambda\mu}}{\partial \xi_\lambda \partial \xi_\mu} - \sum \frac{\partial l_\lambda}{\partial \xi_\lambda} \right] I(t) = \delta I$$

so that if  $\delta$  is negative then  $I(t)$  decreases to zero as  $t$  tends to infinity. The decrease is at least exponential. For any set  $S \subset R^s$  of finite volume we then have

$$P_t(S) = \int_S p(\xi, t) d\xi \leq \left( \text{vol. } S \int_S p^2(\xi, t) d\xi \right)^{1/2} \rightarrow 0.$$

so that the distribution tends to leave the finite part of the space.

The condition obtained is sufficient but not necessary. To obtain sharper conditions we would have to study the extremal properties of the operator  $P$ .

We know that the solution of the particular Fokker-Planck equation studied in this section appears as a limit distribution in certain applications of practical interest. It is therefore a task of considerable importance to find methods enabling us to handle the solutions numerically. Straightforward numerical solution of the equation is of course possible but not very practical because of the great number of parameters appearing in the equation. Indeed we have  $s$  parameters for the infinitesimal mean values and  $\frac{s(s+1)}{2}$  parameters for the infinitesimal covariances.

In all we have

$$N = s \left[ \frac{s}{2} + \frac{3}{2} \right] = r^2 \left[ \frac{r^2}{2} + \frac{3}{2} \right]$$

so that for  $2 \times 2$  matrices  $N = 14$  and for the physically important  $3 \times 3$  matrices  $N$  takes the formidable value of 54 parameters. This makes it almost necessary to use an analytical approach, say of representing the solution in terms of tabulated transcendental functions or by using series expansions or integral representations.

**5.4.** For a given initial distribution of  $x(t)$  for  $t=0$  the probability mass is spread out according to the Fokker-Planck equation as  $t$  increases. Sometimes the mass will be restricted for all  $t$  to a proper subset of the entire space, sometimes it will fill out the whole space, and as we shall see the latter happens in the more general situations.

To fix ideas let us consider the two-dimensional problem so that we deal with the distributions of vectors in the plane when stochastic  $2 \times 2$  matrices  $M$  operate on these vectors. First, if we assume that the mean values of  $M$  vanish, we have  $\Delta x = x(t+h) - x(t) \sim Mx(t)$  for small values of  $h$  and the conditional distribution of  $\Delta x \in R^2$  is described by the infinitesimal covariances

$$q_{ij}(x) = \text{Cov} [(Mx(t))_i, (Mx(t))_j]$$

evaluated for a given  $x(t) = x$ . If the matrix  $Q(x) = \{q_{ij}(x); i, j = 1, 2\}$  is non-singular at  $x$ , diffusion takes place in all directions from the point  $x$ . If we require that the probability mass should be restricted to a subset of  $R^2$ , the matrix  $Q(x)$  must become singular for certain values of  $x$ , say in a set  $S$ . Since  $M$  has four elements it can be represented as

$$M = \sum_{\nu=1}^r \xi_{\nu} A_{\nu}$$

where the  $A_{\nu}$  are fixed matrices, the  $\xi_{\nu}$  are uncorrelated stochastic variables with variance one and  $r \leq 4$ . Then

$$q_{ij}(x) = \sum_{\nu=1}^r (A_{\nu} x)_i (A_{\nu} x)_j$$

so that for a column vector  $z$

$$z^* Q z = \sum_{\nu=1}^r (A_{\nu} x, z)^2.$$

If this quadratic form vanishes it means that no probability mass diffuses in the direction of the vector  $z$ .

On the other hand  $S$  is the set where the determinant of  $Q$  vanishes

$$S = \{x \mid \det Q(x) = 0\}$$

and

$$0 = \det Q(x) = \sum_{\nu, \mu} \begin{vmatrix} (A_{\nu} x)_1 & (A_{\nu} x)_1 & (A_{\mu} x)_1 & (A_{\mu} x)_2 \\ (A_{\nu} x)_2 & (A_{\nu} x)_1 & (A_{\mu} x)_2 & (A_{\mu} x)_2 \end{vmatrix} = \sum_{\nu < \mu} \begin{vmatrix} (A_{\nu} x)_1 & (A_{\mu} x)_1 \\ (A_{\nu} x)_2 & (A_{\mu} x)_2 \end{vmatrix}^2$$

so that 
$$\begin{vmatrix} (A_\nu x)_1 & (A_\mu x)_1 \\ (A_\nu x)_2 & (A_\mu x)_2 \end{vmatrix} = 0 \text{ for all } \nu \text{ and } \mu,$$

which is a set of homogeneous equations of the second order in  $x_1$  and  $x_2$ .  $S$  can be the isolated point  $x=0$ , one or two straight lines or the entire plane.

Let us start with  $r=1$ . Then  $S=R^2$  and the quadratic form vanishes if  $(Ax, z)=0$ . This means that diffusion can only take place in a direction parallel to  $Ax$ , which leads us to consider the differential equation

$$dx = Ax \text{ or } \frac{dx_2}{dx_1} = \frac{(Ax)_2}{(Ax)_1}$$

which is discussed in all elementary text books on differential equations. Its solutions form a one-parameter family of curves  $C_\alpha$  which can consist of straight lines, ellipses, spirals and various types of hyperbolas and parabolas. If the initial distribution is concentrated to a  $C_\alpha$  the mass never leaves this curve but can only diffuse along it. A case of extreme degeneracy occurs if  $A$  is singular and if  $x$  is any vector annihilating  $A$ ,  $Ax=0$ . Then if the initial distribution is concentrated to  $x$  it stays there indefinitely.

If  $r=2$  we get for  $S$

$$\det Q = \begin{vmatrix} (A_1 x)_1 & (A_2 x)_1 \\ (A_1 x)_2 & (A_2 x)_2 \end{vmatrix} = 0$$

which implies  $A_1 x = \lambda A_2 x$ . This is possible for a non trivial vector  $x$  only if  $\det(A_1 - \lambda A_2) = 0$  so that only two values of  $\lambda$  are possible. Excluding the case  $A_1 = \lambda A_2$ , which corresponds to  $r=1$ , we see that  $(A_1 - \lambda A_2)x$  has solutions  $x = \alpha u$  where  $u \in R^2$  and  $\alpha$  is an arbitrary scalar. We only have to investigate when such straight lines can be singular in the manner intended. If only diffusion in the direction  $u$  at a point  $x = \alpha u$  should be possible we must have

$$0 = (A_1 x, n)^2 + (A_2 x, n)^2 = \alpha^2 (A_1 u, n)^2 + \alpha^2 (A_2 u, n)^2,$$

where  $n$  is a vector perpendicular to  $u$ , so that

$$\left. \begin{aligned} A_1 u &= \mu_1 u \\ A_2 u &= \mu_2 u \end{aligned} \right\}$$

This means that  $A_1$  and  $A_2$  should have one or two right eigen vectors in common, and we thus get one or two singular straight lines.

It is now easy to see that for  $r=3$  or  $4$  we get similar answers. Singular lines occur only if the  $A_\nu$  matrices have eigenvectors in common; otherwise the probability mass will fill out the entire plane.

Suppose that  $E \subset R^2$  is a region whose finite part is bounded by singular curves and that the initial distribution assigns the probability one to  $E$ . Then no mass will ever leave  $E$ ,  $P_t(E) \equiv 1$ . The boundary can have many different shapes as we have seen, but let us assume for concreteness that it consists of two semi-infinite rays  $L_1$  and  $L_2$  extending from the origin. The other possibilities are dealt with almost immediately. To prove the statement it is convenient to introduce polar coordinates,  $\rho$  and  $\theta$ . We get

U. GRENANDER, *Stochastic groups*

$$\begin{cases} \Delta \theta = (M (\cos \theta, \sin \theta); (-\sin \theta, \cos \theta)) \\ \Delta \varrho = \varrho (M (\cos \theta, \sin \theta); (\cos \theta, \sin \theta)) \end{cases}$$

and the infinitesimal covariances  $K_{ij}(\theta, \varrho)$  can be written

$$\left. \begin{aligned} K_{11}(\theta, \varrho) &= k_1(\theta) \\ K_{12}(\theta, \varrho) &= \varrho k_2(\theta) \\ K_{22}(\theta, \varrho) &= \varrho^2 k_3(\theta) \end{aligned} \right\}$$

where  $k_i(\theta)$ ,  $i = 1, 2, 3$ , are second order trigonometric polynomials. The Fokker-Planck equation becomes

$$\frac{\partial}{\partial t} p(\theta, \varrho; t) = \frac{1}{2} \frac{\partial^2}{\partial \theta^2} [k_1(\theta) p] + \frac{\partial^2}{\partial \theta \partial \varrho} [\varrho k_2(\theta) p] + \frac{1}{2} \frac{\partial^2}{\partial \varrho^2} [\varrho^2 k_3(\theta) p].$$

The singular rays  $L_1$  and  $L_2$  correspond to two angles  $\theta_1$  and  $\theta_2$  which are roots to the equation  $k_i(\theta) = 0$ ,  $i = 1, 2$ . Introducing

$$P_i(E) = \int_{\theta_1}^{\theta_2} \int_0^\infty \varrho p(\theta, \varrho; t) d\theta d\varrho$$

we get after a few reductions

$$\frac{\partial}{\partial t} P_i(E) = 0$$

as stated. Here we have made use of the fact that  $k_1(\theta)$  has second order zeroes as  $\theta_1$  and  $\theta_2$  and  $k_2(\theta)$  has first order zeroes.

The discussion of the present section can be carried out with simple modifications also when the mean value of  $M$  does not vanish. Actually this tends to make the existence of singular lines more scarce.

**5.5.** In certain applications that will be discussed in a later part of this publication one is interested especially in the distribution of angles, say in  $\theta$ . In this case the mean values play a part that we do not want to neglect and therefore we introduce

$$\left. \begin{aligned} l_1(\varrho, \theta) &= k_4(\theta) \\ l_2(\varrho, \theta) &= \varrho k_5(\theta) \end{aligned} \right\}$$

where  $k_4(\theta)$  and  $k_5(\theta)$  are trigonometric polynomials of the first order. If we write down the Fokker-Planck equation in  $\varrho$  and  $\theta$  with the above coefficients and integrate  $\varrho p$  by parts over the range of  $\varrho$  we get

$$\frac{\partial p(\theta; t)}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial \theta^2} [k_1(\theta) p] - \frac{\partial}{\partial \theta} [k_4(\theta) p].$$

Another simpler way of obtaining this equation is to note that  $\theta(t)$  is also a Markov process, since  $\Delta x = Mx$  is a homogeneous function, and then  $\theta(t)$  must satisfy the appropriate Fokker-Planck equation in  $\theta$  only.



To avoid the existence of singular lines let us assume that the coefficient  $k_1(\theta)$  is positive. We have a simple version of the Sturm-Liouville problem before us and we know that  $p(\theta; t)$  converges to some function, say  $p(\theta)$ , when  $t$  tends to infinity. We have

$$\frac{1}{2} \frac{\partial}{\partial \theta} (k_1 p) - k_4 p = \text{constant} = a$$

with the solutions

$$p(\theta) = \frac{1}{k_1(\theta)} \left[ b + 2a \int_0^\theta e^{-2 \int_0^u \frac{k_4(v)}{k_1(v)} dv} \right] e^{2 \int_0^\theta \frac{k_4(u)}{k_1(u)} du},$$

where the constants  $a$  and  $b$  should be determined so that  $p(\theta)$  is periodic and so that  $p(\theta)$  is a frequency function. This is possible since  $p(0) = b/k_1(0)$  and

$$p(2\pi) = \frac{1}{k_1(0)} [b + 2a I_1] e^{I_2}$$

where

$$\left. \begin{aligned} I_1 &= \int_0^{2\pi} e^{-2 \int_0^u \frac{k_4(v)}{k_1(v)} dv} du > 0 \\ I_2 &= 2 \int_0^{2\pi} \frac{k_4(u)}{k_1(u)} du. \end{aligned} \right\}$$

To get  $p(0) = p(2\pi)$  we should have

$$b = (b + 2a I_1) e^{I_2},$$

so that

$$a = \frac{1 - e^{I_2}}{2 I_1 e^{I_2}} b.$$

Choosing  $b$  positive this determines  $a$  so that  $p(\theta) > 0$  for all  $\theta$  since

$$b + 2a I_1 = \frac{b}{e^{I_2}} > 0.$$

We can now determine  $b$  so that  $\int_0^{2\pi} p(\theta) d\theta = 1$ .

It is interesting to study the behavior of  $\theta(t)$  for large values of  $t$ . Starting with a given value  $\theta_0$  for  $t=0$  the angle changes continuously and we will no longer identify values of  $\theta$  that are congruent mod  $2\pi$ . For a given  $t$  the value of  $\theta$  is the total number of rotations around the origin of the  $x$ -process. We still have the same form of the Fokker-Planck-equation but now we let  $\theta$  take all real values. We have

$$\frac{\partial}{\partial t} E\theta(t) = \int_{-\infty}^{+\infty} p_1(\theta, t) k_4(\theta) d\theta,$$

U. GRENANDER, *Stochastic groups*

where  $p_1$  is the frequency function of  $\theta$  on  $(-\infty, \infty)$ . But since  $k_4(\theta)$  is periodic  $2\pi$  this is the same as

$$\frac{\partial}{\partial t} E \theta(t) = \int_0^{2\pi} p(\theta, t) k_4(\theta) d\theta,$$

and as  $t$  tends to infinity we get

$$\lim_{t \rightarrow \infty} \frac{\partial}{\partial t} E \theta(t) = \int_0^{2\pi} p(\theta) k_4(\theta) d\theta = -2\pi a.$$

This means that 
$$\lim_{t \rightarrow \infty} E \frac{\theta(t)}{t} = -2\pi a.$$

We now need a bound for the variance  $V(t)$  of  $\theta(t)$ . Since

$$V(t + \Delta t) = V(t) + \text{Var}(\Delta \theta) + \text{Cov}(\theta, \Delta \theta)$$

and since the infinitesimal variances and means of the  $\theta$ -process are bounded we see that  $\left| \frac{\Delta V}{\Delta t} \right|$  is bounded. Hence

$$\lim_{t \rightarrow \infty} \text{Var} \frac{\theta(t)}{t} = \lim_{t \rightarrow \infty} \frac{V(t)}{t^2} = 0,$$

which shows that the stochastic variables  $\theta(t)/t$  tends to the limit  $-2\pi a$  in the mean as time increases indefinitely. This is related to a result of S. Lloyd concerning another angular process.

**5.6.** For a more detailed study of the particular Fokker-Planck-equation studied it is convenient to change variables and we will discuss this briefly for  $s=2$ .

Starting from the equation in polar coordinates  $(\theta, \rho)$

$$\frac{\partial p}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial \theta^2} [k_1(\theta) p] + \frac{\partial^2}{\partial \theta \partial \rho} [\rho k_2(\theta) p] + \frac{1}{2} \frac{\partial^2}{\partial \rho^2} [\rho^2 k_3(\theta) p]$$

we note the homogeneous way in which the variable  $\rho$  appears. It is obvious that the coefficients will become simpler in terms of the coordinates  $(u, \theta)$ ,  $u = \log \rho$ . Indeed, we then get

$$\frac{\partial p}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial \theta^2} [k_1(\theta) p] + \frac{\partial^2}{\partial \theta \partial u} [k_2(\theta) p] + \frac{1}{2} \frac{\partial^2}{\partial u^2} [k_3(\theta) p].$$

Let us now introduce  $v = v(u, \theta)$  so that the term with the mixed derivative vanishes. Since  $dv = v'_u du + v'_\theta d\theta$  we get

$$E dv d\theta = [v'_u k_2(\theta) + v'_\theta k_3(\theta)] dt = 0.$$

This first order partial differential equation has the solution

$$v = u - \int_0^\theta \frac{k_2(\varphi)}{k_3(\varphi)} d\varphi$$

and the new coordinate  $v$  is a one-valued function of  $u$  and  $\theta$ ; to see this it is enough to note that

$$\int_0^{2\pi} \frac{k_2(\varphi)}{k_3(\varphi)} d\varphi = 0$$

since  $k_2(\varphi + \pi) = -k_2(\varphi)$  and  $k_3(\varphi + \pi) = k_3(\varphi)$ .

Since

$$E(dv)^2 = E(du)^2 - 2 \frac{k_2(\theta)}{k_3(\theta)} E d\theta du + \left[ \frac{k_2(\theta)}{k_3(\theta)} \right]^2 E(d\theta)^2$$

$$= \frac{k_1(\theta) k_3(\theta) - k_2^2(\theta)}{k_3(\theta)} dt = \frac{D(\theta)}{k_3(\theta)} dt$$

the new equation becomes

$$\frac{\partial p}{\partial t} = \frac{1}{2} \frac{D(\theta)}{k_3(\theta)} \frac{\partial^2 p}{\partial u^2} + \frac{1}{2} \frac{\partial^2}{\partial \theta^2} [k_3(\theta) p].$$

Finally we can make the coefficient of the last term constant by introducing a new angle  $\varphi$  by

$$\varphi(\theta) = \int_0^\theta \frac{d\theta}{\sqrt{k_3(\theta)}}$$

and get

$$\frac{\partial p}{\partial t} = \frac{1}{2} \frac{D[\varphi]}{k_3[\varphi]} \frac{\partial^2 p}{\partial u^2} + \frac{1}{2} \frac{\partial^2 p}{\partial \varphi^2}$$

which is of a fairly simple form.

In a special case the solutions can be expressed as integrals of Mathieu functions but otherwise the author does not know of any way of representing the solutions in terms of standard transcendental functions. Of course when the coefficient of  $\frac{\partial^2 p}{\partial u^2}$  is constant the solutions will be composed of normal frequency functions.

Tryckt den 5 augusti 1960

Uppsala 1960. Almqvist & Wiksells Boktryckeri AB