

PROGRESS AND PROBLEMS IN THE THEORY OF REGENERATIVE PHENOMENA

J.F.C. Kingman

Abstract

Regenerative phenomena were introduced some forty years ago to address problems in the theory of continuous time Markov processes. The early work in the theory left a number of difficult unsolved problems, in the classification of p -functions, oscillation and inequalities, the multiplicative theory, and the theory of unbounded semi- p -functions. Recent years have shown progress on all of these fronts, and this paper surveys these results, while drawing attention to significant problems that remain open.

2000 Mathematics Subject Classification 60J25 (primary), 26A51 (secondary)

1 The context

The theory of regenerative phenomena, as expounded for instance in [27], is a branch of the theory of continuous time Markov processes. It depends on isolating one particular point of the state space, and studying the set of time instants at which the process is at that point. The most fruitful situation is that in which that random set is of non-zero Lebesgue measure.

More precisely, let $X(t)(t \geq 0)$ be a Markov process on some arbitrary state space, and let x_0 be a point of that space. Suppose for definiteness that

$$X(0) = x_0, \tag{1.1}$$

and define a process $Z(t)$ taking values 0 and 1 by

$$\begin{aligned} Z(t) &= 1 \text{ if } X(t) = x_0 \\ &= 0 \text{ if } X(t) \neq x_0. \end{aligned} \tag{1.2}$$

In general Z is not a Markov process, but it satisfies the equation

$$\begin{aligned} \mathbb{P}\{Z(t_1) = Z(t_2) = \dots = Z(t_n) = 1\} \\ = p(t_1)p(t_2 - t_1) \dots p(t_n - t_{n-1}) \end{aligned} \tag{1.3}$$

whenever $0 < t_1 < t_2 < \dots < t_n$, where

$$p(t_1) = \mathbb{P}\{Z(t_1) = 1\}. \tag{1.4}$$

A random process Z taking values in $\{0, 1\}$ and satisfying (1.3) is called a *regenerative phenomenon* with *p-function* (1.4).

It is a consequence of a theorem in [25] (see also Section 3.4 of [27]) that, if Z is a measurable process, the function p is either almost everywhere zero, or is strictly positive and satisfies

$$\lim_{t \downarrow 0} p(t) = 1. \quad (1.5)$$

Most of the theory is concerned with p -functions satisfying (1.5), but the other case is also of great importance, including as it does processes derived from models of Brownian motion (see Section 3).

The idea of studying the process (1.2) stems from the success of Feller's theory [12, 13] of recurrent events in the study of Markov processes in discrete time. In particular, with Erdős and Pollard [11] he proved a theorem implying the Kolmogorov limit theorem [2] for such processes on a countable state space.

It was natural in the 1960s to ask whether a similar technique could be used with similar success in continuous time. The most obvious approach, that of renewal theory [14], deals only with the simplest cases. The need was for a tool effective at the level of generality adopted in Chung's magisterial account [2] of Markov processes with a countable state space, and this was provided in [23] and [24]. For Chung the fundamental objects are the transition functions

$$p_{ij}(t) = \mathbb{P}\{X(t) = j | X(0) = i\}, \quad (1.6)$$

where i and j are elements of the countable sets of states. Taking $i = j = x_0$ identifies the p -function p with the diagonal transition function p_i .

Thus we have two natural questions. What functions p can be the p -functions of regenerative phenomena? Among such functions which can arise as the diagonal case $i = j$ of 1.6? The first was answered in 1963 [23], but the second was unresolved until 1971 [26]. The theory and its applications were summarised in 1972 [27], but that account left many open questions. Some of these have been resolved only recently, while others are still undecided.

This paper does not claim to cover every related development since 1972, but it does attempt to describe the most significant advances, and in particular those that involve new techniques that might be used to attack the remaining open questions.

2 The theory of p -functions

Equation (1.3) determines all the joint distributions of the process Z in terms of the p -function p , by a simple inclusion-exclusion argument. In particular,

$$\begin{aligned} F(t_1, t_2, \dots, t_n; p) \\ = \mathbb{P}\{Z(t_1) = \dots = Z(t_{n-1}) = 0, Z(t_n) = 1\} \end{aligned} \quad (2.1)$$

is a polynomial in the values of $p(t_\alpha)$ and $p(t_\beta - t_\alpha)$ ($\alpha, \beta = 1, 2, \dots, n; \alpha < \beta$). (See equation (2.2.11) of [27] for explicit formulae.) Every p -function satisfies the inequalities

$$F(t_1, t_2, \dots, t_n; p) \geq 0 \quad (2.2)$$

and

$$\sum_{r=1}^n F(t_1, t_2, \dots, t_r; p) \leq 1. \quad (2.3)$$

Standard arguments show conversely that any function $p : (0, \infty) \rightarrow \mathbb{R}$ which satisfies (2.2) and (2.3) for all $n \geq 1, 0 < t_1 < t_2 < \dots < t_n$ is the p -function of some regenerative phenomenon.

In fact, more is true. A function p satisfying (2.2) but not necessarily (2.3) is called a *semi- p -function*. There are many semi- p -functions that are not p -functions, but they are all unbounded. Any bounded function satisfying (2.2) necessarily satisfies (2.3) as well. This has long been known [28] under the additional condition (1.5), but the general case is a recent result [36].

Thus the class φ of all p -functions is exactly the class of functions $p : (0, \infty) \rightarrow [0, 1]$ which satisfy all the inequalities (2.2). Within this can be distinguished important subclasses [36]. A p -function is *standard* (the standard but unhappy terminology) if it satisfies (1.5), and the class of standard p -functions is a proper subclass \mathcal{P} of φ . Standard p -functions are strictly positive, continuous with finite right and left derivatives on $(0, \infty)$, and (defining $p(0) = 1$) have finite or infinite right derivatives at 0.

The basic characterisation of \mathcal{P} was first published in [23], but in essence it is due to D.G. Kendall, and in a primitive version to M.S. Bartlett ([1], section 3.3). A continuous function $p : [0, \infty) \rightarrow [0, 1]$ belongs to \mathcal{P} if and only if there is a measure μ on $(0, \infty]$ such that, for all $\theta > 0$,

$$\int_0^\infty p(t)e^{-\theta t} dt = \left\{ \theta + \int (1 - e^{-\theta x})\mu(dx) \right\}^{-1} \quad (2.4)$$

This identity sets up a bijection between \mathcal{P} and the class \mathcal{M} of measures μ on $(0, \infty]$ with

$$\int (1 - e^{-x})\mu(dx) < \infty. \quad (2.5)$$

Every $p \in \mathcal{P}$ arises by (1.4) and (1.2) from some Markov process, but not necessarily from one with a countable state space. Thus the class of diagonal transition functions, (1.6) with $i = j$, is a proper subclass \mathcal{PM} of \mathcal{P} . The corresponding measures μ are exactly those which (apart from a possible atom at infinity) are either identically zero or admit a strictly positive lower semicontinuous density h on $(0, \infty)$ which is not too small at infinity; more precisely

$$h(x) > 0(x > 0), \quad h(x) \geq e^{-\beta x}(x \geq 1) \quad (2.6)$$

for some β . This result [26] completely characterises the function (1.6) with $i = j$. Similar results ([27], Section 6.6, [31]) deal with the case $i \neq j$ by

characterising the matrix-valued function

$$\begin{pmatrix} p_{ii}(t) & p_{ij}(t) \\ p_{ji}(t) & p_{jj}(t) \end{pmatrix}. \quad (2.7)$$

It might therefore seem that \mathcal{P} is familiar territory, but we shall see in Sections 4, 6 and 7 that this is illusory. Outside \mathcal{P} there are many puzzles. First there are p -functions for which the limit (1.5) exists but is less than 1. For any $a \in [0, 1]$ and $p \in \mathcal{P}$,

$$\bar{p}(t) = ap(t) \quad (2.8)$$

defines a p -function \bar{p} . If $0 < a < 1$, \bar{p} is called *substandard*. The class \mathcal{P}_+ of p -functions that are either standard or substandard is exactly the class of strictly positive Lebesgue measurable p -functions. It is conjectured [36] that every strictly positive p -function is measurable, no counterexample being known.

Although (2.8) defines a smooth function on $(0, \infty)$, the corresponding process Z is anything but regular. If $0 < a < 1$, Z has no measurable version [25].

There are non-measurable p -functions with zero values, and there are measurable p -functions which are not of the form (2.8). Little is known about either class, except that a measurable p -function not in \mathcal{P}_+ is necessarily zero almost everywhere [36]. A greater understanding of these functions would be of value for the problems described in Section 4.

3 Sample function behaviour

The form of (2.4) strongly suggests a connection with the theory of subordinators. If the measure μ on $(0, \infty]$ satisfies (2.5), there is a subordinator, a process $Y(s)$ ($s \geq 0$) with stationary, independent, non-negative increments, such that

$$\mathbb{E}\{e^{-\theta Y(s)}\} = \exp\left\{-s\theta - s \int (1 - e^{-\theta x})\mu(dx)\right\} \quad (3.1)$$

The term $s\theta$ represents a deterministic drift, and the measure μ is, in a sense, the ‘distribution’ of the heights of the jumps of Y . It was pointed out by Kendall [21] that the closure of the range of Y is a random set whose indicator function is a regenerative phenomenon with p -function p determined by (2.4).

This means that the sample function behaviour of any standard regenerative phenomenon can be deduced from the corresponding properties of subordinators. The process Z always has a version in which

$$\{t > 0; Z(t) = 0\} \quad (3.2)$$

is an open subset of $(0, \infty)$, and is therefore expressible in a unique way as a union of disjoint open intervals, the *excursions* of Z (corresponding in (1.2) with the excursions of X from x_0). If we describe such an interval $I = (a, b)$ by the two parameters

$$\sigma_I = b - a, \quad \tau_I = \int_0^a Z(t)dt \quad (3.3)$$

then the random points (σ_I, τ_I) form a Poisson process on the quadrant $(0, \infty)^2$ whose mean measure is the product of μ and Lebesgue measure.

We shall return to this picture in Section 7, but it is worth relating to the theory when p is almost everywhere zero. As already noted, this arises for example in models of Brownian motion. If X is the usual Wiener process and $x_0 = 0$, the process Z defined by (1.2) has trivial finite-dimensional distributions given by (1.3) with $p(t) = 0$ for all $t > 0$. It is nevertheless important in describing the excursions of X from 0. The set (3.2) is still open, but the integral in (3.3) is zero, and a more subtle way of spreading out the points σ_I is needed.

This is of course supplied by Lévy's theory of *local time* (see for instance [41] or more crudely [30]). A unified approach to both situations can be made in terms of the forward and backward recurrence times

$$\mathcal{F}(t) = \inf\{u \geq 0; Z(t+u) = 1\}, \quad (3.4)$$

$$\mathcal{B}(t) = \inf\{u \geq 0; Z(t-u) = 1\}. \quad (3.5)$$

This lies outside the scope of the present paper, and the reader is referred to [15], [29], [32].

4 Oscillations of p -functions

Many questions in the theory of continuous time Markov processes, particularly those with countable state space, can be expressed in terms of the p -function p_{ii} or p -matrices like (2.7). For this reason there has been considerable interest in the extent to which p -functions can oscillate. The best early results were due to Rollo Davidson, and are recorded in the book [22] dedicated to his memory. He drew attention in [10] to the p -function

$$p_{q,a}(t) = e^{-qt} + \sum_{n=1}^{\lfloor t/a \rfloor} \frac{1}{n!} \{q(t-na)\}^n e^{-q(t-na)} \quad (4.1)$$

which corresponds in (2.4) with the measure μ of mass $q > 0$ concentrated at the point $a > 0$. This oscillates repeatedly before converging to a limit

$$p_{q,a}(\infty) = (1 + qa)^{-1}. \quad (4.2)$$

If $qa \geq 1$,

$$p_{q,a}(a) = e^{-qa}, \quad p_{q,a}(a + q^{-1}) = (1 + e^{-qa})e^{-1}, \quad (4.3)$$

so that this p -function can be very small for some values of t and then rise above e^{-1} for larger values of t .

Davidson conjectured that e^{-1} was the largest value for which this is possible; a p -function which takes small values cannot for larger values of t rise much above e^{-1} . Successive authors proved successively stronger results, but it was

not until 1994 that Dai Yong Long established the full Davidson conjecture by proving the remarkable inequality that, for $p \in \mathcal{P}$, $t_1 < t_2$,

$$p(t_2) - p(t_1) \leq e^{-1}. \quad (4.4)$$

The original proof [6] is long, complex and hard to check, and a more accessible version is in [7], which also recounts the history of the result. It is still not easy to see what makes it work, or how to use Dai's techniques to prove other inequalities. For instance, it is plausible to conjecture that (4.4) can be strengthened to

$$F(t_1, t_2; p) = p(t_2) - p(t_1)p(t_2 - t_1) \leq e^{-1}, \quad (4.5)$$

but this is still open.

What would be very useful would be a 'calculus of variations' to provide a systematic way of proving inequalities like (4.5). An inconclusive attempt to do this was made in [36]. This depends on the fact that all p -functions, standard or not, are functions from $(0, \infty)$ to $[0, 1]$, and so can be regarded as points in the space $[0, 1]^{(0, \infty)}$ which by Tychonov's theorem is compact in the product topology. An element of this product space is a p -function if and only if it satisfies all the inequalities (2.2), each of which defines a closed subspace. Thus the set \wp of all p -functions is a closed subspace, and therefore compact Hausdorff (but not metrisable).

The inclusions

$$\mathcal{PM} \subset \mathcal{P} \subset \mathcal{P}_+ \subset \wp \quad (4.6)$$

are all strict. Taking their closures in \wp ,

$$cl\mathcal{PM} \subseteq cl\mathcal{P} \subseteq cl\mathcal{P}_+ \subseteq \wp \quad (4.7)$$

and it is easy to show that only the third of these closures is strict. Thus \mathcal{PM} , \mathcal{P} and \mathcal{P}_+ have the same closure \mathcal{P}^- , and we have the chain of strict inclusions

$$\mathcal{PM} \subset \mathcal{P} \subset \mathcal{P}_+ \subset \mathcal{P}^- \subset \wp. \quad (4.8)$$

The class \mathcal{P}^- consists of all p -functions of the form (2.8) for $0 \leq a \leq 1$, together with some but not all of the *null* (measurable and almost everywhere zero) p -functions, and possibly (though this is not known) some of the non-measurable p -functions.

Unfortunately, there is no known structure theorem for null p -functions, nor any way of deciding if a given null p -function is in \mathcal{P}^- . An inequality like (4.4) extends by continuity to \mathcal{P}^- . Thus a null member of \mathcal{P}^- necessarily has

$$p(t) \leq e^{-1} \quad (4.9)$$

for all $t > 0$. There is a close link between extremal problems in \mathcal{P} and the characterisation of \mathcal{P}^- .

A calculus of variations needs to start from a compact space on which continuous functions attain their bounds. The space \mathcal{P}^- is compact but unknown, and [36] takes the less ambitious line of using compact subsets of \mathcal{P} , like

$$\mathcal{P}_Q = \{p \in \mathcal{P}; \mu(0, \infty) \leq Q\}. \quad (4.10)$$

If for instance (4.5) could be proved in \mathcal{P}_Q for any $Q > 0$, it would be true for all $p \in \mathcal{P}^-$ because the union of the \mathcal{P}_Q is dense in \mathcal{P}^- .

More generally, let $\Phi : \mathcal{P} \rightarrow \mathbb{R}$ be a continuous functional on \mathcal{P} of the form

$$\Phi(p) = \phi\{p(\tau_1), p(\tau_2), \dots, p(\tau_k)\}, \quad (4.11)$$

where $\tau_1, \tau_2, \dots, \tau_k$ are fixed nodes and $\phi : \mathbb{R}^k \rightarrow \mathbb{R}$ is differentiable. Define

$$\Psi(p, t) = \sum_{j=1}^k \phi_j\{p(\tau_1), \dots, p(\tau_k)\} p^{(2)}(\tau_j - t), \quad (4.12)$$

where ϕ_j is the partial derivative of ϕ with respect to its j th argument and

$$p^{(2)}(t) = \begin{cases} \int_0^t p(s)p(t-s)ds & (t > 0), \\ 0 & (t \leq 0). \end{cases} \quad (4.13)$$

Then an easy perturbation argument shows that any p that maximises $\Phi(p)$ on the compact \mathcal{P}_Q has a function $\Psi(p, \cdot)$ that attains its maximum on $[0, \infty)$ and is equal to that maximum except on a set of μ -measure zero. Here μ is the measure corresponding to p in (2.4), augmented if necessary with an atom at 0 to bring its total mass up to Q .

This motivates the following definition. Fix Q and nodes $\tau_1, \tau_2, \dots, \tau_k$. Then $p \in \mathcal{P}_Q$ is called a *candidate* if there are numbers $\beta_1, \beta_2, \dots, \beta_k$ such that the function

$$\psi(t) = \sum_{j=1}^k \beta_j p^{(2)}(\tau_j - t) \quad (4.14)$$

on $[0, \infty)$ is equal to its maximum except on a set of μ -measure zero, p and μ being linked by (2.4). Then any p that maximises any functional (4.9) with the given nodes is a candidate.

It can be shown that (4.1) defines a candidate p -function if $q = Q$ and $\tau_j \leq 2a$ for all j , but there are no other candidates known. On the other hand, there must be other candidates. The simple functional

$$\Phi(p) = p(3) - p(1)^3 \quad (4.15)$$

satisfies $\Phi(p) \leq e^{-1}$ for all p of the form (4.1), but the p -function

$$p(t) = e^{-q \min(t, 1)}$$

has $\Phi(p) = 2/3\sqrt{3} > e^{-1}$ when

$$q = \frac{1}{2} \log 3,$$

so that whatever p -function maximises (4.13) must be a candidate not of the form (4.1).

5 The Jurkat programme

In the classical theory of continuous time Markov processes with a countable state space S , the transition functions $p_{ij}(\cdot)(i, j \in S)$ defined by (1.6) are supposed to satisfy

$$p_{ij}(t) \geq 0, \quad (5.1)$$

$$\lim_{t \downarrow 0} p_{ij}(t) = \delta_{ij}, \quad (5.2)$$

$$\sum_{j \in S} p_{ij}(t) \leq 1, \quad (5.3)$$

and the Chapman-Kolmogorov equation

$$p_{ij}(s+t) = \sum_{k \in S} p_{ik}(s)p_{kj}(t). \quad (5.4)$$

Most of the techniques used to analyse the implications of these conditions [2] make heavy use of (5.3), and it was therefore very surprising that in 1959 Jurkat [16][17] showed how virtually all the analytical properties of the $p_{ij}(\cdot)$ could be proved without this condition. It was therefore natural to ask to what extent later developments, such as the characterisation of \mathcal{PM} , could be generalised.

There is no real problem if (5.3) is weakened to the condition that there is strict inequality in (5.1) and there exists β such that

$$p_{ij}(t) = O(e^{\beta t}) \quad (5.5)$$

for some, and then for all, i and j in S . The problem comes if (5.5) holds for no finite β , a possibility shown to exist by Cornish [3]. Classical methods then break down, in particular because the p_{ij} no longer have Laplace transforms.

The theory of regenerative phenomena does however have a generalisation adequate for this situation, as was shown in [28] and [4]. The function $p = p_{ii}$ satisfies (2.2), but not necessarily (2.3). It is therefore unbounded, and may be more than exponentially large as $t \rightarrow \infty$, in which case (2.4) is meaningless.

The way to overcome this difficulty is to show that (2.4) is formally identical to the Volterra equation

$$1 - p(t) = \int_0^t p(t-s)m(s)ds \quad (5.6)$$

where the function $m(t) = \mu(t, \infty]$ is non-negative, non-increasing and right continuous on $(0, \infty)$, and is integrable on $(0,1)$.

The first result [28] is that any standard semi- p -function (a function satisfying (1.5) and (2.2)) satisfies (5.6) for a non-increasing, right continuous function on $(0, \infty)$ integrable over $(0,1)$. The function p determines uniquely, and is determined uniquely by, the function m , and the only difference from the theory of \mathcal{P} is that m can take negative values.

This result generalises to semi- p -matrices, which are the appropriate abstraction of submatrices

$$(p_{ij}(t); i, j \in I) \quad (5.7)$$

for finite subsets I of S . With these results it is possible to establish a full analogue of the characterisation of \mathcal{PM} . Thus [31] a semi- p -function can arise from some family satisfying (5.1), (5.2) and (5.4) if and only if the function m in (5.6) is either constant or satisfies

$$m(s) - m(t) = \int_s^t h(u) du \quad (5.8)$$

where h is lower semicontinuous with

$$0 < h(t) \leq \infty, \quad (5.9)$$

$$h(t) \leq e^{-\beta t} (t \geq 1) \text{ for some } \beta, \quad (5.10)$$

and

$$\int_0^t uh(u) < \infty \text{ for } 0 < t < \infty. \quad (5.11)$$

This characterises the diagonal functions p_{ii} , and the non-diagonal case is dealt with as before, by a similar characterisation of the semi- p -matrix (2.7).

6 Multiplicative theory

If Z_1 and Z_2 are independent regenerative phenomena with p -functions p_1 and p_2 , then it is immediate from (1.3) that

$$Z(t) = Z_1(t)Z_2(t) \quad (6.1)$$

defines a regenerative phenomenon with p -function

$$p(t) = p_1(t)p_2(t). \quad (6.2)$$

Thus the product of two p -functions is a p -function, and in fact \emptyset , \mathcal{P}_+ , \mathcal{P} and \mathcal{PM} are all closed under pointwise multiplication. Note that it is apparently impossible to prove this very simple fact from the structure theorem (2.4) for \mathcal{P} .

Since \mathcal{P} is a commutative semigroup with identity under the operation (6.2), one can ask about its algebraic properties. Kendall and Davidson were struck by a similarity between this semigroup and the semigroup of probability measures on under convolution, and they used this as the basis for a theory of what they called *delphic semigroups*.

Their conclusions are contained in the collection [22] (especially [8], [9] and [19]), and the theory has advanced little since Rollo Davidson's tragically early death. Most of the obvious questions are open, and probably lack useful answers. An exception however is Kendall's identification of the infinitely divisible

elements of \mathcal{P} . He showed that any continuous function $p : [0, \infty) \rightarrow (0, 1]$ with $p(0) = 1$ and such that $\log p$ is convex belongs to \mathcal{P} . Clearly any such p has the property that

$$p^\alpha : t \mapsto \{p(t)\}^\alpha \tag{6.3}$$

is also in \mathcal{P} for any $\alpha > 0$, and he showed that only these functions have this property.

This raises the question: which subsets of $(0, \infty)$ can be of the form

$$\{\alpha > 0; p^\alpha \in \mathcal{P}\} \tag{6.4}$$

for some $p \in \mathcal{P}$? Clearly this set is topologically closed, closed under addition and contains every positive integer. It was conjectured that every set of the form (6.4) must contain $[1, \infty)$, i.e. that $p \in \mathcal{P}$ implies $p^\alpha \in \mathcal{P}$ for every $\alpha > 1$.

This conjecture was eventually proved in [34], and implies that \mathcal{P}_+ and \mathcal{P}^- are also closed under the operation

$$p \mapsto p^\alpha \ (\alpha > 1). \tag{6.5}$$

It is more difficult to show that \wp is similarly closed, but this was done in [35]. It is probable that \mathcal{PM} has the same property, but the evidence though strong is not conclusive [35].

What makes these problems difficult is that (6.5) has no probabilistic interpretation when α is not an integer, so that one is thrown back on the analytic characterisation (2.4) in terms of the *canonical measure* μ of p . There is no simple form of the canonical measure of p^α even when α is an integer, as will be seen in the next section.

Before moving on, it is worth mentioning another open problem. A function $\varphi : (0, 1] \rightarrow (0, 1]$ may be said to preserve \mathcal{P} if

$$\varphi(p) : t \mapsto \varphi\{p(t)\} \tag{6.6}$$

is in \mathcal{P} whenever $p \in \mathcal{P}$. Thus we know that

$$\varphi(x) = x^\alpha \tag{6.7}$$

preserves \mathcal{P} for $\alpha \geq 1$ but not for $0 < \alpha < 1$. Are there any other functions (apart from the trivial $\varphi(x) = 1$) which preserve \mathcal{P} ?

If φ is such a function, since e^{-t} is in \mathcal{P} ,

$$p(t) = \varphi(e^{-t}) \tag{6.8}$$

is in \mathcal{P} , which shows that φ is continuous, and can be extended to a continuous function from $[0, 1]$ to $[0, 1]$, with $\varphi(0) = 1$. It has right and left derivatives in $(0, 1)$. The inequalities (2.2) translate into complicated functional inequalities for φ , but whether these imply that φ is of the form (6.7) is not known. It is not even known if a simple function like

$$\varphi(x) = \frac{1}{2}x + \frac{1}{2}x^2 \tag{6.9}$$

preserves \mathcal{P} .

7 The canonical measure of a product of p -functions

Equation (2.4) sets up a bijection between \mathcal{P} and the class \mathcal{M} of measures on $(0, \infty]$ satisfying (2.5). The semigroup operation (6.2) of pointwise multiplication in \mathcal{P} induces a semigroup operation

$$(\mu_1, \mu_2) \mapsto \mu \quad (7.1)$$

in \mathcal{M} , where μ is the canonical measure of the product p -function. The problem is to give a usable formula for this binary operation.

The corresponding problem in Feller's discrete time theory warns of the difficulties ahead. He defines a *renewal sequence* as a sequence $u = (u_n; n \geq 0)$ generated by a recurrence relation

$$u_0 = 1, \quad u_n = \sum_{r=1}^n f_r u_{n-r} \quad (n \geq 1) \quad (7.2)$$

from some sequence $f = (f_n; n \geq 1)$ with

$$f_n \geq 0, \quad \sum_{n=1}^{\infty} f_n \leq 1. \quad (7.3)$$

If $u^{(1)}$ and $u^{(2)}$ are renewal sequences, so is their product

$$u_n = u_n^{(1)} u_n^{(2)}. \quad (7.4)$$

For ease of notation write the corresponding f -sequences as (a_n) and (b_n) . Then each element of the f -sequence of u can be expressed in terms of those of $u^{(1)}$ and $u^{(2)}$. For example

$$f_3 = a_3 b_3 + 2a_1 a_2 b_1 b_2 + 2a_1 a_2 b_3 + 2a_3 b_1 b_2 + a_1^3 b_3 + a_3 b_1^3, \quad (7.5)$$

but the expressions for f_n rapidly become more complicated as n increases. It is true that they can be interpreted probabilistically, which is why all the terms are positive, but it is difficult to see how to exploit them.

If we seek a similar analysis in continuous time, this should start from the interpretation of the canonical measure in terms of the intervals that are the connected components of the open set (3.2). If, for $i = 1, 2$, Z_i is a version for which

$$G_i = \{t > 0; Z_i(t) = 0\} \quad (7.6)$$

is an open subset of $(0, \infty)$, the product phenomenon Z has

$$G = \{t > 0; Z(t) = 0\} = G_1 \cup G_2 \quad (7.7)$$

also open. For vividness we call the connected components of G_1 *red intervals* and those of G_2 *blue intervals*. Notice that, by Fubini's theorem, there is zero probability that any red interval is also a blue interval.

The red and the blue intervals taken together form a covering of G , but a covering with the very special property that it admits a unique minimal sub-covering of G . Call a red interval *essential* if it is not wholly contained in some blue interval, and a blue interval essential if not wholly contained in a red. It is easy to see that the essential red and blue intervals cover G , and that every covering of G by red and blue intervals must include every essential interval.

Now consider any bounded connected component J of G . Then J is a bounded open interval that can be expressed in a unique way as the union of the essential intervals that meet it. By compactness only finitely many of these intervals meet any closed subinterval of J , but there may be infinitely many intervals near each endpoint.

There are a number of possibilities. The interval J may itself be a red interval, necessarily essential, and possibly containing blue intervals, none of them essential; call this Case 1. Case 2 is similarly that in which J is a blue interval. Or J may be the union of an essential red and an essential blue interval, which we call Case 12 if the red is to the left of the blue and Case 21 if the opposite. Proceeding in this way we can define Cases

$$1, 2, 12, 21, 121, 212, 1212, 2121, 12121, \dots \tag{7.8}$$

describing finite alternations of essential red and blue intervals, each of which may contain inessential intervals of the other colour. Each essential interval overlaps its differently coloured neighbours, but no interval overlaps another of the same colour.

If the minimal covering of J is not finite, we shall still have alternations of red and blue essential intervals, but these will be unbounded at one or both ends, i.e.

$$\begin{array}{ll} 121212\dots, & \dots 1212121 \\ 212121\dots, & \dots 2121212 \end{array}$$

or

$$\dots\dots 12121212\dots\dots$$

Can such ‘wild’ coverings occur with positive probability? We do not know, but we do know that they occur in the analogous ‘local time’ situation.

Thus let X_1 and X_2 be independent Wiener processes, so that

$$X = (X_1, X_2) \tag{7.9}$$

is the usual Brownian motion in \mathbb{R}^2 . An excursion of X from $(0,0)$ encircles the origin infinitely often, and in particular crosses the coordinate axes infinitely often as it approaches $(0,0)$. The same is true at the beginning of the excursion, so that every excursion is, with probability 1, wild at both ends.

At the other extreme, consider the situation in which the canonical measures of Z_1 and Z_2 are finite, so that

$$q_i = -p'_i(0) = \mu_i(0, \infty] < \infty \tag{7.10}$$

and therefore

$$q = -p'(0) = q_1 + q_2 = \mu(0, \infty] < \infty. \quad (7.11)$$

Then the same function behaviour is very simple, each process being a step function between the values 0 and 1. For instance, $Z(t)$ has jump discontinuities at random points $0 = T_0 < T_1 < T_2 < \dots$, with

$$Z(t) = 1 \text{ on } (0, T_1] \cup [T_2, T_3] \cup \dots. \quad (7.12)$$

The random variables $T_n - T_{n-1}$ are independent, having a negative exponential distribution with mean $1/q$ if n is odd, and the distribution μ/q if n is even.

In particular, μ/q is the distribution of the length of $J = (T_1, T_2)$. Under (7.10) there are only finitely many red and blue intervals in any finite interval, and so (if $T_2 < \infty$) only the *tame* coverings (7.8) can occur. The joint distribution of T_1 and T_2 , on the event that a particular case of (7.8) occurs, can then be written down in terms of the distributions of the discontinuities of Z_1 and Z_2 . For instance, the event that Case 121 occurs is the event that there are V_1, V_2, V_3, V_4 with

$$T_1 < V_1 < V_2 < V_3 < V_4 < T_2, \quad (7.13)$$

such that (T_1, V_2) and (V_3, T_2) are red intervals and (V_1, V_4) a blue interval. The outcome is a decomposition formula for μ according to the type of the covering:

$$\begin{aligned} \mu(dt) = & p_2(t)\mu_1(dt) + p_1(t)\mu_2(dt) \\ & + \mu_{12}(dt) + \mu_{21}(dt) + \mu_{121}(dt) + \dots. \end{aligned} \quad (7.14)$$

The first two terms correspond to Case 1 and Case 2, and the others are complicated multiple integrals which the interested reader can reconstruct.

The formula (7.13) makes sense even if (7.10) is violated, but is only valid if, with probability 1, only tame coverings occur. It is conceivable that this is true for all regenerative phenomena (in contradistinction to the Brownian motion situation already mentioned). Certainly it is necessary only to assume (7.10) for one of $i = 1, 2$, but it is not known if (7.13) has wider validity. If wild coverings occur, equality must be replaced by inequality \geq .

Note also that (7.13) only gives information about μ on $(0, \infty)$, and the atom at infinity needs to be dealt with separately, using the formula

$$\int_0^\infty p(t)dt = \mu\{\infty\}^{-1}, \quad (7.15)$$

which follows from (2.4) on letting $\theta \rightarrow 0$. It is perfectly possible to have $\mu\{\infty\} > 0$ even if $\mu_1\{\infty\} = \mu_2\{\infty\} = 0$.

8 Kink analysis

Even if the formula (7.13) were shown to be valid for all p -functions, its form is so complicated that it is of little use. There is therefore good reason to seek

techniques that, even if they give only partial information about the binary operation (7.1), are simple enough to be useful. One such was introduced in [35] under the title of *kink analysis*. It depends on the fact that, if μ has an atom

$$\mu\{t\} > 0 \quad (0 < t < \infty) \quad (8.1)$$

it causes a kink in the graph of the p -function. In fact, p has right and left derivatives with

$$p'(t+) - p'(t-) = \mu\{t\}. \quad (8.2)$$

This shows at once that, if $p = p_1 p_2$, the corresponding measures have

$$\mu\{t\} = p_2(t)\mu_1(t) + p_1(t)\mu_2\{t\}. \quad (8.3)$$

If μ_1 and μ_2 are purely atomic, (8.3) determines the atoms of μ , but μ also has a non-atomic component. However, (8.3) does show that

$$\mu(dt) \geq p_2(t)\mu_1(dt) + p_1(t)\mu_2(dt), \quad (8.4)$$

and by continuity this inequality is valid for all p_1, p_2 in \mathcal{P} . The analysis of Section 7 explains the probabilistic meaning of (8.4).

The same technique is applied in [35] to powers of p -functions. If $p \in \mathcal{P}$ has canonical measure μ , and (for $\alpha > 1$) μ_α denotes the canonical measure of p^α , then the atoms of μ_α occur at exactly the same points as those of μ , and

$$\mu_\alpha\{t\} = \alpha p(t)^{\alpha-1} \mu\{t\} \quad (0 < t < \infty). \quad (8.5)$$

It follows that, if a measure ν_α on $(0, \infty)$ is defined by

$$\mu_\alpha(dt) = \alpha p(t)^{\alpha-1} \nu_\alpha(dt), \quad (8.6)$$

then ν_α increases with α :

$$\mu \leq \nu_\alpha \leq \nu_\beta \quad (1 < \alpha < \beta). \quad (8.7)$$

These results come tantalisingly close to proving that, if $p \in \mathcal{PM}$ and $\alpha > 1$, then $p^\alpha \in \mathcal{PM}$. This would settle a question raised by David Williams [35].

9 Two conjectures of D.G. Kendall

In this final section, we draw attention to two problems, both raised by Kendall, which are still open. Neither refers explicitly to regenerative phenomena, but it is quite possible that their solution may draw on the theory.

The first of these is the *Markov group conjecture* [18]. A family of functions p_{ij} satisfying the conditions (5.1)–(5.4) defines a semigroup of operators $P_t(t > 0)$ on the space l_1 by the recipe

$$(xP_t)_j = \sum_{i \in S} x_i p_{ij}(t) \quad (9.1)$$

when $(x_i; i \in S)$ is a sequence with

$$\|x\| = \sum_{i \in S} |x_i| < \infty. \quad (9.2)$$

Equation (5.4) shows that

$$P_{s+t} = P_s P_t \quad (s, t > 0) \quad (9.3)$$

and (5.2) implies that the semigroup is strongly continuous in the sense that

$$\lim_{t \downarrow 0} \|x P_t - x\| = 0 \quad (9.4)$$

for all $x \in l_1$.

Equation (9.4) does not necessarily imply the stronger condition

$$\lim_{t \downarrow 0} \|P_t - I\| = 0, \quad (9.5)$$

which is equivalent to (5.2) holding uniformly in i, j . This holds if and only if

$$P_t = \exp(Qt) \quad (9.6)$$

for some bounded operator Q on l_1 . Kendall noted that (9.6) implies that P_t is defined for all t , positive or negative, although the p_{ij} for $t < 0$ will often be negative.

The semigroup (P_t) can be extended to a group of operators satisfying (9.3) if and only if, for some and then for all t , P_t has an inverse as a bounded operator on l_1 . The Markov group conjecture asserts that this can only happen when (9.5) holds. Partial results have been obtained by Speakman [39], Williams [40], Cuthbert [5] and Mountford [37], and these link the problem with the oscillation of the p -functions p_{ii} . However, even the recent advances described in Section 4 fail to resolve the conjecture. It has been argued [33] that the problem is really one about finite positive matrices, in which case regenerative phenomena may not play a part in any eventual solution.

The second conjecture has to do with what is sometimes called the *germ problem* for Markov semigroups. However, the word ‘germ’ is used in quite another way in Markov theory, and Reuter [38] talks instead of ‘ 0^+ -equivalence’. Suppose that there are two different families $p_{ij} = a_{ij}$ and $p_{ij} = b_{ij}$ of solutions to (5.1)–(5.4), and that for each i, j there exists $\gamma_{ij} > 0$ such that

$$a_{ij}(t) = b_{ij}(t) \quad (0 < t \leq \gamma_{ij}). \quad (9.7)$$

Is it then the case that (9.7) holds for all $t > 0$?

Kendall ([20], Section 30) conjectured a positive answer, and some partial results can be found in [38] and [42]. In particular, the answer is positive if γ_{ij} depends only on i [42] (or by a duality argument only on j). No counterexample is known.

The link with regenerative phenomena is the p -matrix

$$(p_{ij}(t); i, j \in I), \quad (9.8)$$

where I is an arbitrary finite subset of S . A knowledge of $p_{ij}(t)$ for $t \leq \gamma_{ij}$ determines this matrix for $t \leq \gamma_I$, where

$$\gamma_I = \min \{\gamma_{ij}; i, j \in I\}. \quad (9.9)$$

Now there is a matrix analogue of the Volterra equation (5.6) for p -matrices, in which p and m are replaced by $I \times I$ matrices p_I and m_I , and it can be shown that $m_I(t)$ is determined by the values of p_I on $(0, t)$. Hence we know $m_I(t)$ for $t \leq \gamma_I$.

The matrix-valued functions on m_I for different I are related to one another in a complicated way. It does seem possible that a sufficiently subtle use of these relationships might show that the knowledge of m_I can be extended to an interval not depending on I , in which case the conjecture would be solved.

These arguments are very much of the flavour of the *Lévy dichotomy*, which states [2] that, under (5.1)–(5.4), each function p_{ij} is either always or never zero. Reuter's analysis suggests a further conjecture that would strengthen the Lévy dichotomy, that if (a_{ij}) and (b_{ij}) satisfy (5.1)–(5.4) and (for a particular pair i, j),

$$a_{ij}(t) \geq b_{ij}(t) \quad (9.10)$$

for all $t > 0$, then either $a_{ij}(t) = b_{ij}(t)$ for all t , or $a_{ij}(t) > b_{ij}(t)$ for all t .

References

- [1] M.S. BARTLETT Recurrence and first passage times, *Proc. Camb. Philos. Soc.* **49** (1953) 263–275.
- [2] K.L. CHUNG *Markov Chains with Stationary Transition Probabilities* (Springer, Berlin, 1967).
- [3] A.G. CORNISH Some properties of semi-groups of non-negative matrices and their resolvents, *Proc. London Math. Soc.* (3) **23** (1971) 643–667.
- [4] A.G. CORNISH Measurable semi- p -functions, *Bull London Math. Soc.* **5** (1973) 91–99.
- [5] J.R. CUTHBERT An inequality with relevance to the Markov group problem, *J. London Math. Soc.* (2) **11** (1975) 104–106.
- [6] Y. DAI The Markov oscillation problem, *Acta Math. Sinica* **10** (1994) 99–140.
- [7] Y. DAI & E. RENSHAW The Markov oscillation problem in discrete time, *J. London Math. Soc.* (2) **61** (2000) 301–314.

- [8] R. DAVIDSON Arithmetic and other properties of certain Delphic semi-groups, *Z. Wahrscheinlichkeitsch.* **10** (1968) 120–172.
- [9] R. DAVIDSON More Delphic theory and practice, *Z. Wahrscheinlichkeitsch.* **13** (1969) 191–203.
- [10] R. DAVIDSON Smith’s phenomenon and ‘jump’ p -functions, *Stochastic Analysis* **22** 234–247.
- [11] P. ERDOS, W. FELLER & H. POLLARD A theorem on power series, *Bull. Amer. Math. Soc.* **55** (1949) 201–204.
- [12] W. FELLER Fluctuation theory of recurrent events, *Trans. Amer. Math. Soc.* **67** (1949) 98–119.
- [13] W. FELLER *An Introduction to Probability Theory and its Applications* (Wiley, New York, 1950).
- [14] W. FELLER *An Introduction to Probability Theory and its Applications, Volume II* (Wiley, New York, 1966).
- [15] J. HOFFMANN-JORGENSEN Markov sets, *Math. Scand.* **24** (1970) 145–166.
- [16] W.B. JURKAT On semi-groups of positive matrices, *Scripta Math.* **24** (1959) 123–131 and 207–218.
- [17] W.B. JURKAT On the analytic structure of semigroups of positive matrices, *Math. Z.* **73** (1960) 346–365.
- [18] D.G. KENDALL On Markov groups, *Proc. Fifth Berkeley Sympos. Math. Statist. Prob.* **2** (1965/66) 165–173.
- [19] D.G. KENDALL Delphic semigroups, infinitely divisible regenerative phenomena, and the arithmetic of p -functions, *Z. Wahrscheinlichkeitsch.* **9** (1968) 163–195.
- [20] D.G. KENDALL An introduction to stochastic analysis, *Stochastic Analysis* **22** 3–43.
- [21] D.G. KENDALL On the non-occurrence of a regenerative phenomenon in given intervals, *Stochastic Analysis* **22**, 294–308.
- [22] D.G. KENDALL & E.F. HARDING *Stochastic Analysis* (Wiley, London, 1973).
- [23] J.F.C. KINGMAN A continuous time analogue of the theory of recurrent events, *Bull. Amer. Math. Soc.* **69** (1963), 268–272.
- [24] J.F.C. KINGMAN The stochastic theory of regenerative events, *Z. Wahrscheinlichkeitsth.* **2** (1964) 180–224.

- [25] J.F.C. KINGMAN On measurable p -functions, *Z. Wahrscheinlichkeitsch.* **11** (1968) 1–8.
- [26] J.F.C. KINGMAN Markov transition probabilities V, *Z. Wahrscheinlichkeitsch.* **17** (1971) 89–103.
- [27] J.F.C. KINGMAN *Regenerative Phenomena* (Wiley, London, 1972).
- [28] J.F.C. KINGMAN Semi- p -functions, *Trans. Amer. Math. Soc.* **174** (1972) 257–273.
- [29] J.F.C. KINGMAN Homecomings of Markov processes, *Adv. Appl. Prob.* **5** (1973) 66–102.
- [30] J.F.C. KINGMAN An intrinsic description of local time, *J. London Math. Soc.* **6** (1973) 725–731.
- [31] J.F.C. KINGMAN On the Chapman-Kolmogorov equation, *Phil. Trans. Roy. Soc. A* **276** (1974) 341–369.
- [32] J.F.C. KINGMAN Anticipation processes, *Perspectives in Probability and Statistics* (Applied Probability Trust, Sheffield, 1975) 201–215.
- [33] J.F.C. KINGMAN Three unsolved problems in discrete Markov theory, *Probability, Statistics and Analysis* (Cambridge, 1983) 180–191.
- [34] J.F.C. KINGMAN Powers of renewal sequences, *Bull. London Math. Soc.* **28** (1996) 527–532.
- [35] J.F.C. KINGMAN Powers and products of regenerative phenomena, *Austral. & N. Zealand J. Statist.* **46** (2004) 79–86.
- [36] J.F.C. KINGMAN Extremal problems for regenerative phenomena, *J. Appl. Prob.* **41A** (2004) 333–346.
- [37] D. MOUNTFORD On the Markov group problem, *Bull. London Math. Soc.* **9** (1977) 57–60.
- [38] G.E.H. REUTER On Kendall’s conjecture concerning $0+$ equivalence of Markov transition functions, *J. London Math. Soc.* **35** (1987) 377–384.
- [39] J.M.O. SPEAKMAN Some problems relating to Markov groups, *Proc. Fifth Berkeley Sympos. Math. Statist. Prob.* **2** (1965/66) 175–186.
- [40] D. WILLIAMS On operator semigroups and Markov groups, *Z. Wahrscheinlichkeitsch.* **13** (1969) 280–285.
- [41] D. WILLIAMS *Diffusions, Markov Processes and Martingales* (Wiley, London, 1979).
- [42] YAO QI YU On Kendall’s “germ” conjecture for Markov chains, *Bull. London Math. Soc.* **22** (1990) 386–392.

J.F.C. Kingman
Isaac Newton Institute for Mathematical Sciences,
20 Clarkson Road,
Cambridge CB3 0EH,
UK

director@newton.cam.ac.uk