EXTREMAL PROBLEMS FOR REGENERATIVE PHENOMENA

J.F.C. Kingman, Isaac Newton Institute

Abstract

This paper explores the possibility of a calculus of variations powerful enough to prove inequalities for the p-functions of regenerative phenomena such as that conjectured by Davidson and proved by Dai. It is shown that this is unlikely to be achieved by compactifying the space of standard p-functions, and a more promising approach is that of working in a compact subspace. The analysis leads to a class of candidate p-functions, which contains all the maxima of general functionals.

Keywords Inequalities for p-functions, calculus of variations, compact function spaces.

AMS 2000 Subject Classification – Primary 60J99
Secondary 49K77

Postal address : Isaac Newton Institute for Mathematical Sciences,
20 Clarkson Road, Cambridge CB3 0EH, UK
e-mail address: director@newton.cam.ac.uk

1. The Davidson-Dai inequality

The analytic theory of that class of random processes known as *regenerative phenomena* was largely completed thirty years ago (Kingman, 1972), but there remained a number of tantalising unsolved problems stemming largely from the work of Rollo Davidson (see Kendall and Harding, 1973), whose untimely death in 1970 deprived stochastic analysis of a remarkably original mathematician.

One example is the 1966 conjecture, proved only in (Kingman, 1996, 2003), that p^α is a p-function for any p-function p and any $\alpha > 1$. Much more important, however, is the work of Dai Yong Long and others on Davidson's oscillation problem. Davidson (1968), and independently Blackwell and Freedman (1968), proved that any standard p-function satisfies the inequality

$$p(t_1) \geqslant \tfrac{1}{2} + \left\{p(t_2) - \tfrac{3}{4}\right\}^{1/2} \tag{1.1}$$

in $0 < t_1 < t_2$, so long as $p(t_2) > \frac{3}{4}$. In particular, if $p(t_1)$ is small for some t_1, then $p(t_2)$ cannot even rise much above $\frac{3}{4}$ for any larger t_2.

Davidson knew that $\frac{3}{4}$ was not the best possible value, and conjectured that it could be replaced by e^{-1}, giving examples to show that this lower value could not be improved. Evidence for this conjecture was later provided by the work of Griffeath, Cornish, Joshi, Yu and Zou, who successively reduced the value from $\frac{3}{4}$ (the time series is given in Dai and Renshaw (2000)). Finally, Dai proved in 1994 that any standard p-function satisfies

$$p(t_2) - p(t_1) \leqslant e^{-1} \qquad (0 < t_1 < t_2)\,, \tag{1.2}$$

establishing Davidson's conjecture in a most satisfactory form.

Dai's original proof was long, complex and hard to check, and the improvement by Dai and Renshaw (2000) is still far from transparent. Moreover, it is difficult to see how to apply his methods to prove other inequalities of similar type. For instance, it seems very likely that (1.2) could be strengthened to

$$p(t_2) - p(t_1)p(t_2 - t_1) \leqslant e^{-1} \qquad (0 < t_1 < t_2)\,, \tag{1.3}$$

which would be of interest since the left hand side has a simple probabilistic meaning, but this is an open conjecture.

It would be good to have a systematic method for proving such inequalities, and an obvious route would be to set up a calculus of variations for p-functions. Thus one would regard the left hand side of (1.2) or (1.3) (for fixed t_1, t_2) as a function

$$\Phi : \mathcal{P} \to \mathbb{R} \tag{1.4}$$

defined on the set $\mathcal{P}$ of all standard p-functions. By some sort of compactness argument it might be shown that Φ attains its maximum at some (unknown) p^* in $\mathcal{P}$. A perturbation analysis of the inequality

$$\Phi(p) \leqslant \Phi(p^*) \tag{1.5}$$

for p near p^* would then give necessary conditions on p^* which might determine p^*, and (1.5) would then be the required inequality.

Unfortunately this programme falls at the first hurdle. The function

$$\Phi(p) = 1 - p(1) \tag{1.6}$$

is even simpler than those in (1.2) and (1.3), but never attains its supremum over $\mathcal{P}$. Likewise if (1.3) is true, it implies strict inequality in (1.2), so that

$$p(t_2) - p(t_1)$$

does not attain its supremum e^{-1}.

This forces us to look outside $\mathcal{P}$, at p-functions which are not standard. A good deal is known about these, and they form a space which, with a natural topology, is compact and Hausdorff. The closure of $\mathcal{P}$ in this space is a natural setting for a calculus of variations, but it seems a much deeper problem to describe this closure than to prove inequalities like (1.3) directly.

We are therefore forced to the less elegant approach of maximising Φ on compact subsets of $\mathcal{P}$, chosen so that their union is dense in $\mathcal{P}$. This approach does allow a perturbation analysis of (1.5), which limits the possible p^* to a class of *candidate p-functions* defined without reference to Φ. We are still however a long way from the solution of particular problems like the conjecture (1.3).

2. The natural history of p-functions

A regenerative phenomenon (Kingman, 1972) is a random process $Z = (Z(t); t > 0)$ taking only the values 0 and 1, with the property that, for

$$n \geqslant 1\,,\ 0 < t_1 < t_2 < \ldots < t_n\,, \tag{2.1}$$

$$\mathbb{P}\{Z(t_1) = Z(t_2) = \ldots = Z(t_n) = 1\} = p(t_1)p(t_2 - t_1)\ldots p(t_n - t_{n-1})\,. \tag{2.2}$$

Here p is a function on $(0, \infty)$ called the *p-function* of Z, which determines the finite-dimensional distributions of Z by a simple inclusion-exclusion argument. In particular,

$$F(t_1, t_2, \ldots, t_n : p) = \mathbb{P}\{Z(t_1) = \ldots = Z(t_{n-1}) = 0\,, Z(t_n) = 1\} \tag{2.2}$$

is a polynomial in the values $p(t_r), p(t_r - t_s)$ $(1 \leqslant s < r \leqslant n)$. For example, the conjecture (1.3) is in this notation

$$F(t_1, t_2; p) \leqslant e^{-1} \qquad (p \in \mathcal{P})\,. \tag{2.3}$$

The Daniell-Kolmogorov theorem shows that a function $p : (0, \infty) \to \mathbb{R}$ is the p-function of some regenerative phenomenon if and only if, for every sequence (2.1),

$$F(t_1, t_2, \ldots, t_n; p) \geqslant 0 \tag{2.4}$$

and

$$\sum_{r=1}^{n} F(t_1, t_2, \ldots, t_r; p) \leqslant 1\,. \tag{2.5}$$

In fact, (2.5) can be dispensed with, since (Kingman, 2003) a function p satisfying (2.4) satisfies (2.5) if and only if it is bounded. In particular, a function

$$p : (0, \infty) \to [0, 1] \tag{2.6}$$

is a p-function if and only if it satisfies (2.4).

The most important p-functions are those that are *standard* in the sense that

$$\lim_{t \to 0} p(t) = 1\,, \tag{2.7}$$

and as above $\mathcal{P}$ denotes the class of standard p-functions. There are however many p-functions which fail (2.7) (and which we avoid calling non-standard for fear of confusion with non-standard analysis), and it is important to understand how they relate to $\mathcal{P}$. It should however be emphasised that they are significant in relation to the analytic theory of p-functions rather than to the sample function behaviour of the process Z. If we were interested in the latter, we would generalise Z in the direction of local time for processes such as Brownian motion (Kingman, 1975).

The p-functions that do not satisfy (2.7) fall naturally into three classes (Kingman, 1968, 1972).

(i) For any $p_1 \in \mathcal{P}$ and $0 < a < 1$,

$$p(t) = a p_1(t) \tag{2.8}$$

defines a p-function, and such p-functions will be described as *substandard.* Although analytically well-behaved, they correspond to processes Z that are highly irregular, possessing for instance no measurable version.

The class of p-functions which are either standard or substandard, so that they are of the form (2.8) for $0 < a \leqslant 1$, will be denoted by $\mathcal{P}_+$. An alternative description of $\mathcal{P}_+$ is as the class of strictly positive, Lebesgue measurable p-functions. We shall return to the structure of $\mathcal{P}_+$ in Section 4.

(ii) If a p-function is Lebesgue measurable but does not belong to $\mathcal{P}_+$, it must be zero almost everywhere, and will be described as *null.* (Note the convention that the zero function is null rather than substandard; it is the only continuous p-function not in $\mathcal{P}_+$.)

The simplest example of a null p-function arises if $Z(t)$ can only be 1 for integer values of t (or integer multiples of some fixed t_0). Thus $p(t) = 0$ unless t is an integer. If we write

$$f_n = F(1, 2, \ldots, n; p) \geqslant 0\,, \tag{2.9}$$

then

$$p(n) = u_n\,, \tag{2.10}$$

where (u_n) is the renewal sequence generated from (f_n) by the Feller recursion

$$u_0 = 1\,,\ u_n = \sum_{r=1}^{n} f_r u_{n-r} (n \geqslant 1)\,. \tag{2.11}$$

More generally, let C be any countable subset of $(0,\infty)$, and let $f : C \to \mathbb{R}$ satisfy

$$f(c) \geqslant 0\,, \ \sum_{c \in C} f(c) \leqslant 1\,. \tag{2.12}$$

Then the equation

$$p(t) = \sum f(c_1)f(c_2)\dots f(c_r)\,, \tag{2.13}$$

where the sum extends over all $r \geqslant 1$ and all $c_1, c_2, \dots, c_r$ (not necessarily distinct) in C, with

$$c_1 + c_2 + \dots + c_r = t\,, \tag{2.14}$$

defines a p-function which vanishes outside a countable set, and is thus a null p-function.

A p-function of the form (2.13) is said to be *of renewal type*, and not all p-functions vanishing off a countable set are of renewal type. Moreover not all null p-functions vanish outside a countable set. For example, if G is any proper additive subgroup of $\mathbb{R}$,

$$\begin{aligned} p(t) &= 1 (t > 0, t \in G) \\ &= 0 (t > 0, t \notin G) \end{aligned} \tag{2.15}$$

defines a p-function, which is null if G is measurable.

(iii) Every Lebesgue measurable p-function is either standard, substandard or null, but there are also (assuming the axiom of choice) non-measurable p-functions. For example, (2.15) defines such a function if G is not measurable. Very little is known about non-measurable p-functions; it is not for instance known if there are any that are strictly positive on $(0,\infty)$.

3. The compact space of p-functions

Denote by $\wp$ the set of all p-functions, of which $\mathcal{P}$ and $\mathcal{P}_+$ are proper subsets. Can we make $\wp$ into a topological space in which compactness arguments can be used to analyse functions

$$\Phi : \wp \to \mathbb{R} \tag{3.1}$$

such as those in (1.2) and (1.3)? A minimal requirement is that the evaluation

$$p \longmapsto p(t) \tag{3.2}$$

should be a continuous function from $\wp$ to $\mathbb{R}$ for each $t > 0$, and the weakest topology with this property is that which $\wp$ inherits as a subspace of the product space

$$[0,1]^{(0,\infty)} \tag{3.3}$$

of all functions from $(0,\infty)$ to $[0,1]$.

Any stronger topology will have fewer compact sets, so it will be assumed without further comment that $\wp$ has been given the subspace topology of pointwise convergence. Since (2.2) is a polynomial in the values of p,

$$p \longmapsto F(t_1, t_2, \ldots, t_n; p) \tag{3.4}$$

is continuous. A function in the space (3.3) is a p-function if and only if each of the functions (3.4) is non-negative, so that $\wp$ is closed in (3.3). By Tychonov's theorem, (3.3) is a compact Hausdorff space, and hence the closed subspace $\wp$ is also compact and Hausdorff.

Unfortunately however $\wp$ is not metrisable, since the subset consisting of functions of the form (2.15) is not even first countable.

Since the function (3.2) does not attain its infimum 0 on the subset $\mathcal{P}$ of $\wp$, $\mathcal{P}$ cannot be closed in $\wp$. The closure $\mathcal{P}^-$ of $\mathcal{P}$ in $\wp$ is of course a compact subset of $\mathcal{P}$, and if it could be identified would be the natural arena for results such as those of Davidson and Dai. For instance, (1.2) holds for all $p \in \mathcal{P}^-$, with equality for some $p \in \mathcal{P}^-$.

Theorem 1 *In $\wp$, the subspaces $\mathcal{P}$ and $\mathcal{P}_+$ have the same closure $\mathcal{P}^-$, and the inclusions*

$$\mathcal{P} \subset \mathcal{P}_+ \subset \mathcal{P}^- \subset \wp \tag{3.5}$$

are all strict.

Proof For any $q, b > 0$,

$$p_0(t) = e^{-q\min(t,b)}$$

is in $\mathcal{P}$, and hence so is

$$p(t) = p_0(t)p_1(t)$$

for any $p_1 \in \mathcal{P}$. For $0 < a < 1$, let $q \to \infty$, $b \to 0$ with $qb = -\log a$, to show that

$$p(t) = ap_1(t)$$

belongs to the closure $\mathcal{P}^-$ of $\mathcal{P}$. Hence

$$\mathcal{P} \subset \mathcal{P}_+ \subseteq \mathcal{P}^- ,$$

and taking closures,

$$\mathcal{P}^- \subseteq (\mathcal{P}_+)^- \subseteq \mathcal{P}^- .$$

Thus $\mathcal{P}_+$ has the same closure $\mathcal{P}^-$ as $\mathcal{P}$, and the chain (3.5) is true with weak inclusions.

That the first inclusion is strict is clear, since the function in (2.8) satisfies

$$\lim_{t \to 0} p(t) = a ,$$

so that p belongs to $\mathcal{P}_+$ but not to $\mathcal{P}$ if $0 < a < 1$. The second is strict because $\mathcal{P}^-$ contains the zero function (as well, as we shall see below, as many other null p-functions). To prove that the third inclusion is strict, note that the Davidson-Dai inequality (1.2) extends by continuity to $\mathcal{P}^-$. If $p \in \mathcal{P}^-$ is null, we can choose t_1 arbitrarily small with $p(t_1) = 0$, so that any null p-function in $\mathcal{P}^-$ satisfies

$$p(t) \leqslant e^{-1} \quad (t > 0) . \tag{3.6}$$

There are many null p-functions that fail (3.6), such as (2.15) with $G = \mathbb{Z}$, so that $\mathcal{P}^-$ is a proper subset of $\mathcal{P}$ and the proof is complete.

The inequality (3.6) suggests that the identification of $\mathcal{P}^-$ is likely to be at least as difficult a problem as that of proving inequalities like (1.2) and (1.3). Even for p-functions of the form (2.10), it seems very hard to decide which belong to $\mathcal{P}^-$. Davidson's original construction uses the p-function

$$p(t) = \sum_{n=0}^{\infty} \pi_n \{q(t - n\tau)\} , \tag{3.7}$$

where q and τ are positive constants, and π_n denotes the Poisson probability

$$\pi_n(\mu) = \mu^n e^{-\mu}/n! \tag{3.8}$$

with $\pi_n(\mu) = 0$ for $\mu < 0$. (The provenance of this important p-function will appear in Section 4.) If in (3.7) we set $\tau = 1 - q^{-1}$ and let $q \to \infty$, we have

$$\lim_{q\to\infty} p(t) = \pi_t(t) \qquad \text{if } t = 1, 2, 3, \ldots$$
$$= 0 \qquad \textit{otherwise}\,. \tag{3.9}$$

Thus the renewal sequence

$$u_n = n^n e^{-n}/n! \tag{3.10}$$

defines by (2.10) a null p-function in $\mathcal{P}^-$. Since $p(1) = e^{-1}$, this special p-function attains the upper bound in (3.6).

This construction can be greatly generalised to produce a wide class of null members of $\mathcal{P}^-$, all of renewal type (2.13).

There is another problem about the identification of $\mathcal{P}^-$. All the known examples of elements of $\mathcal{P}^-$ are limits of sequences of elements of $\mathcal{P}$, but there may be elements of $\mathcal{P}^-$ which are not sequential limits (see for instance Kelley (1955)). In other words, the sequential closure of $\mathcal{P}$ may be a proper subset of $\mathcal{P}^-$. If this is so, there may be non-measurable p-functions in $\mathcal{P}^-$, and these cannot be ignored when maximising continuous functions on $\mathcal{P}$.

Davidson (1968) proved that $\mathcal{P}$ is metrisable, but it does not follow that its closure is metrisable. This is an open question, and one which would be worth settling because, if $\mathcal{P}^-$ could be shown to be the sequential closure of $\mathcal{P}$, if could only contain measurable (and therefore standard, substandard or null) p-functions.

In addressing this question, it is useful to note that, as a topological space, $\mathcal{P}_+$ is somewhat better behaved than $\mathcal{P}$. This fact, and its consequences, are the topic of the next section.

4. The Volterra equation for standard and substandard p-functions

The original representation theorem for regenerative phenomena set up a one-to-one correspondence between the set $\mathcal{P}$ of standard p-functions and the set of measures μ on $(0, \infty]$ satisfying

$$\int \min(t, 1)\mu(dt) < \infty\,, \tag{4.1}$$

by the Laplace transform identity

$$\hat{p}(\theta) = \int_0^\infty p(t)e^{-\theta t}dt = \left\{\theta + \int \left(1 - e^{-\theta t}\right) \mu(dt)\right\}^{-1} \tag{4.2}$$

in $\theta > 0$. For instance, if μ concentrates all its mass q at a single point τ,

$$\hat{p}(\theta) = \left\{\theta + q - qe^{-\theta\tau}\right\}^{-1}, \tag{4.3}$$

which is easily inverted to give Davidson's formula (3.7).

Later it was realised that, for many purposes, the correspondence between p and its 'canonical measure' μ could more usefully be expressed by the Volterra equation

$$1 - p(t) = \int_0^t p(t-s)m(s)ds\,, \tag{4.4}$$

which is equivalent to (4.2) if

$$m(t) = \mu(t, \infty]\,. \tag{4.5}$$

The function m is non-negative, non-increasing and right-continuous on $(0, \infty)$, and integrable on (0,1), and any such function defines by (4.5) a measure μ which is the canonical measure of the standard p-function satisfying (4.2).

Kendall (1968) showed that the bijection (4.2) is a homeomorphism, if $\mathcal{P}$ is given the product topology described above, and the set of canonical measures is given the weakest topology making

$$\mu \longmapsto \int \min(t, 1)f(t)\mu(dt) \tag{4.6}$$

continuous for every bounded continuous f on $(0, \infty]$. Thus $\mathcal{P}$ is metrisable, and indeed Davidson (1968) showed that the product topology is equivalent to the compact-open topology on $\mathcal{P}$ (here and elsewhere using the terminology of Kelley (1955)) of functions on $[0, \infty)$. The compact subsets of $\mathcal{P}$ are exactly the closed sets on which (2.7) holds uniformly, and $\mathcal{P}$ is not a countable union of such sets, and thus fails to be σ-compact.

This whole approach extends to the larger space $\mathcal{P}_+$, which turns out also to be metrisable. However, the condition for a closed subset of $\mathcal{P}_+$ to be compact is much simpler than is the case for $\mathcal{P}$, and $\mathcal{P}_+$ is σ-compact. These facts are spelt out in Theorem 2.

They are not however enough to resolve the sequential closure problem. In a compact Hausdorff space that is not metrisable, it is perfectly possible to have a metrisable σ-compact subspace whose sequential closure is a proper subset of its closure (consider for example the Stone-Čech compactification of the integers). Thus it remains an open question whether every number of $\mathcal{P}^-$ is the limit of a sequence of standard p-functions, and in particular whether all members of $\mathcal{P}^-$ are measurable.

Theorem 2 If the function $\ell : [0,\infty) \to [0,\infty)$ *is continuous, non-decreasing and concave, then the equation*

$$1 - p(t) = p(t)\ell(0) + \int_0^t p(t-s)d\ell(s) \tag{4.7}$$

has a unique solution p*, which belongs to* $\mathcal{P}_+$*. Conversely, every* $p \in \mathcal{P}_+$ *is the unique solution of (4.7) for a unique continuous, non-decreasing, concave* ℓ*. The topology induced on the set* $\mathcal{L}$ *of all such functions by the bijection* $\lambda : \mathcal{P}_+ \to \mathcal{L}$ *defined by (4.7) is that of pointwise convergence on* $(0,\infty)$*. The space* $\mathcal{P}_+$ *is metrisable and* σ*-compact, and a closed subset* $F \subset \mathcal{P}_+$ *is compact if and only if*

$$\sup_{p\in F} \ell(1) < \infty\,. \tag{4.8}$$

Note that (4.7) can be written in the suggestive form

$$1 - p(t) = \int_{0_-}^t p(t-s)d\ell(s)\,, \tag{4.9}$$

if we adopt the convention that $\ell(0-) = 0$ and include the jump at 0 in the Stieltjes integral. Note also that the requirement that ℓ be continuous is needed only to ensure right-continuity at 0.

Proof For $\ell \in \mathcal{L}$, define

$$a = \{1 + \ell(0)\}^{-1}\ ,\ \ell_1(t) = a\{\ell(t) - \ell(0)\}\ , \tag{4.10}$$

so that $0 < a \leqslant 1$, $\ell_1 \in \mathcal{L}$ and $\ell_1(0) = 0$. Then (4.7) is equivalent to

$$1 - a^{-1}p(t) = \int_0^t a^{-1}p(t-s)d\ell_1(s) = \int_0^t a^{-1}p(t-s)m_1(s)ds\,, \tag{4.11}$$

where m_1 is the right derivative of ℓ_1. Since m_1 is non-negative, non-increasing and right-continuous on $(0, \infty)$, and integrable on $(0, 1)$, it corresponds in (4.4) to a function $p_1 \in \mathcal{P}$. Comparing (4.4) and (4.10) shows that the latter has the unique solution

$$p(t) = ap_1(t)\,, \tag{4.12}$$

and this is a p-function in $\mathcal{P}_+$. Thus $\lambda : \mathcal{P}_+ \to \mathcal{L}$ is a bijection.

Equation (4.12) can be written in Laplace transform terms. If $\hat{p}(\theta)$ is defined by (4.2), and

$$\hat{\ell}(\theta) = \int_0^\infty \ell(t) e^{-\theta t} dt\,, \tag{4.18}$$

this integral necessarily converging for $\theta > 0$, (4.7) leads at once to the identity

$$\theta \hat{p}(\theta) \left\{ 1 + \theta \hat{\ell}(\theta) \right\} = 1 \quad (\theta > 0)\,. \tag{4.14}$$

The space $\mathcal{L}$ is a familiar one, and it is easy to check that, with the topology of pointwise convergence, it is metrisable and that a closed set is compact if and only if it is bounded. Since $\ell \in \mathcal{L}$ satisfies

$$\ell(t) \leqslant \ell(1)(t < 1)\,,\ \ell(t) \leqslant \ell(1)t(t > 1)\,, \tag{4.15}$$

it is sufficient for $\ell(1)$ to be bounded.

The metrisability of $\mathcal{P}_+$ is more delicate. Davidson's proof for $\mathcal{P}$ depends on establishing a Lipshitz condition which uses (2.7) and is not available in $\mathcal{P}_+$. Instead we need a deeper estimate which is proved in Section 3.3 of (Kingman, 1972). The inequality (5) shows easily that, if $0 < \alpha < \beta < \infty$ and θ is chosen so that $\hat{m}_1(\theta) < 1$, then the p-function (4.12) satisfies

$$|p(t_1) - p(t_2)| \leqslant C\left(\alpha, \beta, \theta, \hat{m}_1(\theta)\right) |t_1 - t_2|\,, \tag{4.16}$$

for $\alpha \leqslant t_1 < t_2 \leqslant \beta$, where the constant C depends on p only through $\hat{m}_1(\theta)$. This shows, as in Davidson (1968), that the topology of $\mathcal{P}_+$ is equivalent to that of convergence on the rationals of $(0, \infty)$, which is metrisable.

Because both $\mathcal{P}_+$ and $\mathcal{L}$ are metrisable, the continuity of the bijection λ and its inverse will be proved if we can show that a sequence (p_n) in $\mathcal{P}_+$ converges pointwise to p in $\mathcal{P}_+$ if and only if the corresponding functions $\ell_n = \lambda(p_n)$ converge to $\ell = \lambda(p)$. By the continuity theorem for Laplace transforms:

$$p_n \to p \text{ if and only if } \hat{p}_n(\theta) \to \hat{p}(\theta) \text{ for all } \theta > 0\,,$$
$$\ell_n \to \ell \text{ if and only if } \hat{\ell}_n(\theta) \to \hat{\ell}(\theta) \text{ for all } \theta > 0\,,$$

and the equivalence of these assertions is guaranteed by (4.14).

This completes the proof, but the result is slightly unsatisfactory because (4.8) is expressed in terms of the function ℓ rather than directly in terms of the p-function. It is therefore worth noting that (4.14) shows that (4.8) is equivalent to

$$\inf_{p\in F} \hat{p}(\theta) > 0\,, \tag{4.17}$$

for some and then for all $\theta > 0$.

5. The extrapolation problem

The results of the last two sections offer us a variety of choices of compact subsets of $\wp$ in which to work. The largest is $\mathcal{P}^-$, but since this is unknown in extent it is impossible to use effectively. More useful would be a subset of $\mathcal{P}_+$ whose compactness could be assured by (4.8) or (4.17). If we prefer to stay in $\mathcal{P}$ we need to use the Kendall compactness criterion, that a closed subset F of $\mathcal{P}$ is compact if and only if

$$\lim_{t\to 0} \inf_{p\in F} p(t) = 1\,. \tag{5.1}$$

In section 6 we shall use such a compact subset, namely

$$\mathcal{P}_Q = \{p \in \mathcal{P}\,;\ -p'(0) \leqslant Q\} \tag{5.2}$$

for a constant Q. In this section, however, we make an even more drastic choice, achieving (5.1) by specifying the values of p in some neighbourhood of $t = 0$.

Thus fix $\tau > 0$ and $p_0 \in \mathcal{P}$ and consider the set

$$\mathcal{P}(\tau, p_0) = \{p \in \mathcal{P}\,;\ p(t) = p_0(t)\,,\ t \leqslant \tau\}\ . \tag{5.3}$$

What values can $p(t)$ take for $t > \tau$? This is the extrapolation problem.

If m and m_0 correspond to p and p_0 in the Volterra equation (4.4), and m^0 is the restriction of m_0 to $(0,\tau)$, the condition that $p = p_0$ on $(0,\tau]$ is equivalent to

$$m(t) = m^0(t) \quad (0 < t < \tau)\,. \tag{5.4}$$

Thus the functions of $\mathcal{P}(\tau, p_0)$ correspond in (4.4) to the extensions of m^0 to non-negative, non-increasing, right-continuous functions on $(0,\infty)$. These extensions form a convex and therefore connected set, so that the set of values of $p(t)$ for fixed $t > \tau$, compact because (5.3) is compact, is also connected, and is thus a compact interval:

$$\{p(t)\,;\; p \in \mathcal{P}(\tau, p_0)\} = [a(t,\tau,p_0)\,, b(t,\tau,p_0)]\;. \tag{5.5}$$

The extrapolation problem is that of determining $a(t,\tau,p_0)$ and $b(t,\tau,p_0)$.

It is part of the folklore of the subject that $a(t,\tau,p_0)$ is, as a function of t, itself a p-function, and is the unique minimal element of $\mathcal{P}(\tau,p_0)$. In fact, we have the following theorem.

Theorem 3 *For fixed τ, p_0, $\mathcal{P}(\tau,p_0)$ has a unique minimal element $p_{\min}$, characterised by the property that its canonical measure is concentrated on $(0,\tau)\cup\{\infty\}$, and*

$$a(t,\tau,p_0) = p_{\min}(t) \tag{5.6}$$

for $t > \tau$. If $p \in \mathcal{P}(\tau,p_0)$ has canonical measure μ, then

$$p(t) = p_{\min}(t) + \sum \int_{[n\tau,t]} p_{\min}^{(n+1)}(t-s)\mu_n(ds)\,, \tag{5.7}$$

where $p_{\min}^{(n)}$ is the n-fold convolution of $p_{\min}$ with itself, μ_n is the n-fold Stieltjes convolution with itself of the restriction of μ to $[\tau,\infty)$, and the sum extends over n with $1 \leqslant n \leqslant t/\tau$.

Proof The smallest extension of m^0 is that with $m(t) = m_0(\tau-)$ for all $t \geqslant \tau$, which corresponds to sweeping all the mass in $[\tau,\infty)$ to ∞. Define $p_{\min}$ as the corresponding p-function, with Laplace transform

$$\hat{p}_{\min}(\theta) = \left\{\theta + \int_{(0,\tau)} \left(1 - e^{-\theta t}\right)\mu_0(dt) + m_0(\tau-)\right\}^{-1}\,. \tag{5.8}$$

For any other $p \in \mathcal{P}(\tau, p_0)$ with canonical measure μ, $\mu = \mu_0$ on $(0, \tau)$, and $\mu[\tau, \infty] = \mu_0[\tau, \infty] = m_0(\tau-)$, so that

$$\hat{p}_{\min}(\theta)^{-1} - \hat{p}(\theta)^{-1} = \int_{[\tau,\infty)} e^{-\theta t} \mu(dt) \, . \tag{5.9}$$

Thus, if θ is large enough for the series to converge,

$$\hat{p}(\theta) = \hat{p}_{\min}(\theta) + \sum_{n=1}^{\infty} \hat{p}_{\min}(\theta)^{n+1} \left\{ \int_{[\tau,\infty)} e^{-\theta t} \mu(dt) \right\}^{n} ,$$

which inverts to give (5.7). Thus $p(t) \geqslant p_{\min}(t)$ for $t > \tau$, and the proof is complete.

The situation is quite different for the upper bound $b(t, \tau, p_0)$. Take for example the case $\tau < t \leqslant 2\tau$. Then only the first two terms in (5.7) are non-zero, and

$$p(t) = p_{\min}(t) + \int_{[\tau,t]} p^{(2)}_{\min}(t-s)\mu(ds) \, . \tag{5.10}$$

This is maximised when μ concentrates all its measure $m_0(\tau-)$ in $[\tau, \infty]$ at the maxima of $p^{(2)}_{\min}(t-s)$, and

$$b(t, \tau, p_0) = p_{\min}(t) + m_0(\tau-) \max \left\{ p^{(2)}(u) \, ; \; 0 \leqslant u \leqslant t - \tau \right\} . \tag{5.11}$$

The maximising p depends on t, and $b(\cdot, \tau, p_0)$ is not a p-function.

When $t > 2\tau$ the determination of $b(t, \tau, p_0)$ is a difficult non-linear optimisation problem, though it should be noted that ingenious manipulation of this function is a feature of Dai's technique (see for instance Proposition 2 of Dai and Renshaw (2000)).

6. Perturbation theory

We now turn to the second aspect of the programme, that of perturbing about an unknown maximum to get conditions that might determine the maximum. For definiteness we work in the space $\mathcal{P}_Q$ defined at (5.2), but the analysis is very similar in other compact spaces such as, for instance the space of $p \in \mathcal{P}_+$ with $\hat{p}(1) \geqslant \delta > 0$.

Recall that every standard p-function has a (finite or infinite) derivative at 0 which is related to its canonical measure by

$$q = -p'(0) = \mu(0,\infty]\,. \tag{6.1}$$

If q is finite,

$$p(t) \geqslant e^{-qt} \tag{6.2}$$

for all $t > 0$. It follows that, for any constant Q, $q \leqslant Q$ if and only if

$$p(t) \geqslant e^{-Qt} \quad \text{for all } b > 0\,. \tag{6.3}$$

The Kendall criterion then shows that the space $\mathcal{P}_Q$ of p-functions with $q \leqslant Q$ is compact. (The space of those with $q = Q$ is not compact unless $Q = 0$.)

The union of the $\mathcal{P}_Q$ for all $Q > 0$ is dense in $\mathcal{P}$. Thus if we can prove an inequality like (1.3) in $\mathcal{P}_Q$ for every Q, it extends at once to $\mathcal{P}$. This is the excuse for the inelegant restriction to $\mathcal{P}_Q$.

When working in $\mathcal{P}_Q$ for fixed Q, it is convenient to extend the canonical measure to $[0,\infty]$ by giving it an atom

$$\mu\{0\} = Q - q\,, \tag{6.4}$$

so that (4.2) can be written

$$\hat{p}(\theta) = \left\{\theta + Q - \int_{[0,\infty]} e^{-\theta t}\mu(dt)\right\}^{-1} \tag{6.5}$$

in $\theta > 0$. For precision, the extended measure on $[0,\infty]$ will be called the *Q-canonical measure* of p. Equation (6.5) expresses the homeomorphism between $\mathcal{P}_Q$ and the compact space of measures of mass Q on $[0,\infty]$.

Let p_0 and p_1 be two members of $\mathcal{P}_Q$ with Q-canonical measures μ_0 and μ_1, and write ν for the signed measure $\nu = \mu_1 - \mu_0$ with $\nu[0,\infty] = 0$. For $0 < \epsilon < 1$,

$$\mu_\epsilon = (1-\epsilon)\mu_0 + \epsilon\mu_1 = \mu_0 + \epsilon\nu \tag{6.6}$$

is the Q-canonical measure of a p-function p_ϵ in $\mathcal{P}_Q$, and (6.5) shows that

$$\hat{p}_\epsilon(\theta)^{-1} = \hat{p}_0(\theta)^{-1} - \epsilon\int e^{-\theta t}\nu(dt)\,.$$

Thus

$$\hat{p}_\epsilon(\theta) = \hat{p}_0(\theta) + \epsilon \hat{p}_0(\theta)\hat{p}_\epsilon(\theta) \int e^{-\theta t}\nu(dt)\,,$$

which inverts to give

$$p_\epsilon(t) = p_0(t) + \epsilon \int\int_{u+v\leqslant t} p_0(t-u-v)p_\epsilon(u)\nu(dv)\,. \tag{6.7}$$

This is the key perturbation formula in $\mathcal{P}_Q$, for it implies that, as $\epsilon \to 0$,

$$p_\epsilon(t) = p_0(t) + \epsilon \int_{[0,t]} p_0^{(2)}(t-v)\nu(dv) + O(\epsilon^2)\,, \tag{6.8}$$

where

$$p^{(2)}(t) = \int_0^t p(t-u)p(u)du\,. \tag{6.9}$$

Suppose that we are trying to maximise over $\mathcal{P}_Q$ the continuous functional

$$\Phi(p) = \phi\left\{p(\tau_1)\,, p(\tau_2)\,, \ldots, p(\tau_k)\right\}\,, \tag{6.10}$$

where ϕ is a differentiable function of k variables and the nodes τ_j are fixed. An example would be the left-hand side of (1.3), with $k = 3$, $\tau_1 = t_1$, $\tau_2 = t_2 - t_1$, $\tau_3 = t_2$. If Φ attains its maximum at p_0, and p_1 is any other member of $\mathcal{P}_Q$, we can substitute (6.8) into the inequality

$$\Phi(p_\epsilon) \leqslant \Phi(p_0) \tag{6.11}$$

and let $\epsilon \to 0$ to give the linearised inequality

$$\sum_{j=1}^{k} \phi_j\left\{p_0(t_1)\,, \ldots, p_0(\tau_k)\right\} \int_{[0,\tau_j]} p_0^{(2)}(\tau_j - v)\nu(dv) \leqslant 0\,, \tag{6.12}$$

where ϕ_j is the partial derivative of ϕ with respect to its jth variable.

Thus with each functional Φ of the form (6.11) we can associate a functional $\Psi : \mathcal{P} \times [0,\infty) \to \mathbb{R}$ by

$$\Psi(p,t) = \sum_{j=1}^{k} \phi_j\left\{p(\tau_1)\,, \ldots, p(\tau_k)\right\} p^{(2)}(\tau_j - t)\,, \tag{6.13}$$

with the convention that $p^{(2)}(t) = 0$ for $t < 0$. Any p-function that maximises Φ over $\mathcal{P}_Q$ must satisfy

$$\int \Psi(p, v)\nu(dv) \leqslant 0 \tag{6.14}$$

for every signed measure ν on $[0, \infty]$ with $\nu[0, \infty] = 0$ such that $\mu + \nu$ is a Q-canonical measure (μ being the Q-canonical measure of p). This last will be true if

$$\mu + \nu \geqslant 0\,. \tag{6.15}$$

In particular, if $|\nu| \leqslant \mu$, (6.15) holds both for ν and for $-\nu$, so that

$$\int \Psi(p, v)\nu(dv) = 0$$

for such ν. This shows that $\Psi(p, t)$ is a constant C outside a set of μ-measure zero. Then (6.14) shows that

$$\int \{\Psi(p, v) - C\}(\mu + \nu)(dv) \leqslant 0$$

whenever $\mu + \nu$ is a positive measure on $[0, \infty]$ with total mass Q, showing that

$$\Psi(p, t) \leqslant C$$

for all $t \epsilon [0, \infty]$. Thus we have proved the following result, in which we write $\psi(t)$ for $\Psi(p, t)$ when we wish to emphasise dependence on t. Note that $\psi(t) = 0$ for large t.

Theorem 4 A standard p-function (with canonical measure μ) that maximises Φ on $\mathcal{P}_Q$ has a function $\psi(t) = \Psi(p, t)$ given by (6.13) which attains its maximum in $t \geqslant 0$ and which is equal to that maximum almost anywhere modulo μ.

7. Candidate p-functions

Theorem 4 is difficult to apply in particular cases. Suppose for instance we wish to prove the conjecture (1.3). We should then try to maximise

$$\Phi(p) = p(t_2) - p(t_1)p(t_2 - t_1)\,. \tag{7.1}$$

The corresponding function Ψ is

$$\psi(t) = p^{(2)}(t_2 - t) - p(t_2 - t_1)p^{(2)}(t_1 - t) - p(t_1)p^{(2)}(t_2 - t_1 - t)\,. \tag{7.2}$$

This vanishes for $t \geqslant t_2$ and is strictly positive if

$$\max(t_1, t_2 - t_1) \leqslant t < t_2\,,$$

so that its maximum C is strictly positive. If we knew that ψ had a single maximum, Theorem 4 would show that ψ is concentrated at this point, and is thus a Davidson p-function of the form (3.7). But $\{t\,; \psi(t) = C\}$ might contain more than one point, even a non-trivial interval.

If we could exclude these possibilities, the maximisation problem would simply be one of optimising q and τ in (3.7), and this would yield (1.3) almost at once. The key to such progress must be the function $p^{(2)}$. This has Laplace transform

$$\begin{aligned}\hat{p}(\theta)^2 &= \left\{\theta + q - \int e^{-\theta t}\mu(dt)\right\}^{-2} \\ &= \sum_{n=0}^{\infty}(n+1)(\theta+q)^{-n-2}\int e^{-\theta t}\mu_n(dt)\,,\end{aligned}$$

so that

$$p^{(2)}(t) = te^{-qt} + \sum_{n=1}^{\infty}\frac{1}{n!}\int_{(0,t)}(t-s)^{n+1}e^{-q(t-s)}\mu_n(ds)\,, \tag{7.3}$$

where μ_n is now the n-fold Stieltjes convolution with itself of the canonical measure on $(0,\infty]$.

Another useful equation for $p^{(2)}$ is a Volterra equation similar to that satisfied by p. It is easily checked by taking Laplace transforms that, for $p \in \mathcal{P}$,

$$tp(t) = p^{(2)}(t) + \int_{(0,t)} sp^{(2)}(t-s)\mu(ds). \tag{7.4}$$

Substituting this into (7.2) and simplifying gives

$$\int_{[0,t_2]} s\psi(s)\mu(ds) = t_2\Phi(p) - \psi(0)\,. \tag{7.5}$$

If p maximises Φ, $\psi = C$ almost anywhere modulo μ, so that

$$t_2\Phi(p) = \psi(0) + C\int_{[0,t_2]} s\mu(ds)\,. \tag{7.6}$$

The significance of this curious equality is obscure.

Although Theorem 4 gives no easy access to problems like the proof of (1.3), it does turn a spotlight on a particular class of p-functions. Whatever the functional Φ (with nodes $\tau_1, \tau_2, \ldots, \tau_k$), its maxima on $\mathcal{P}_Q$ must be attained by p-functions having the very special property that, for some numbers β_j, the function

$$\psi(t) = \sum_{j=1}^{k} \beta_j p^{(2)}(\tau_j - t) \tag{7.7}$$

is equal to its maximum almost everywhere modulo the Q-canonical measure. Such functions might properly be called *candidate p-functions*.

The only known examples of candidate p-functions are those of the Davidson class (3.7), but there must be others. To see this consider the functional

$$\Phi(p) = p(3) - p(1)^3\,, \tag{7.8}$$

which like (7.1) has a simple probabilistic meaning. It is easy to check that $\Phi(p) \leqslant e^{-1}$ for any p of the form (3.7). However, the p-function

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

has

$$\Phi(p) = e^{-q} - e^{-3q}$$

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

$$\Phi(p) = 2\sqrt{3}/q > e^{-1}\,.$$

However, for $Q \geqslant \frac{1}{2}\log 3$, the maximising p must be a candidate p-function that is not of the form (3.7).

References

BLACKWELL, D. and FREEDMAN, D. (1968). On the local behaviour of Markov transition probabilities. *Ann. Math. Statist.* **39**, 2123–2127.

DAI, Y. (1994). The Markov oscillation problem. *Acta Math. Sinica* **10**, 99–140.

DAI, Y. and RENSHAW, E. (2000). The Markov oscillation problem in discrete time. *J. London Math. Soc.* **(2)61**, 301–314.

DAVIDSON, R. (1968). Arithmetic and other properties of certain delphic semigroups: II. *Z. Wahrsch' theorie & verw. Geb.* **10**, 146–172. [Reprinted in Kendall and Harding (1973), pp. 150–182.]

KELLEY, J.L. (1955). *General Topology.* Van Nostrand, New York.

KENDALL, D.G. (1968). Delphic semigroups, infinitely divisible regenerative phenomena, and the arithmetic of p-functions. *Z. Wahrsch' theorie & verw. Geb.* **9**, 163–195. [Reprinted in Kendall and Harding (1973) pp. 73–114.]

KENDALL, D.G. and HARDING, E.F. (1973). *Stochastic Analysis.* Wiley, London.

KINGMAN, J.F.C. (1968). On measurable p-functions. *Z. Wahrsch' theorie & verw. Geb.* **11**, 1–8.

KINGMAN, J.F.C. (1972). *Regenerative Phenomena.* Wiley, London.

KINGMAN, J.F.C. (1975). Anticipation processes. *Perspectives in Probability and Statistics.* 201–215.

KINGMAN, J.F.C. (1996). Powers of renewal sequences. *Bull. London Math. Soc.* **28**, 527–532.

KINGMAN, J.F.C. (2003). Powers and products of regenerative phenomena (to appear).