Ann. Inst. Statist. Math.
Vol. 49, No. 3, 555–567 (1997)

ON A CLASS OF CHARACTERIZATION PROBLEMS FOR RANDOM CONVEX COMBINATIONS

LUDWIG BARINGHAUS AND RUDOLF GRÜBEL

Institut für Mathematische Stochastik, Universität Hannover, Postfach 6009, 30060 Hannover, Germany

(Received October 20, 1995; revised November 8, 1996)

Abstract. We consider stochastic equations of the form $X =_d W_1X + W_2X'$, where (W_1, W_2), X and X' are independent, '$=_d$' denotes equality in distribution, $EW_1 + EW_2 = 1$ and $X =_d X'$. We discuss existence, uniqueness and stability of the solutions, using contraction arguments and an approach based on moments. The case of $\{0,1\}$-valued W_1 and constant W_2 leads to a characterization of exponential distributions.

Key words and phrases: Stochastic difference equations, exponential distributions, characterization problems, contractions.

1. Introduction

Let $W = (W_1, W_2)$ be a two-dimensional random vector with the property that the sum of the expected values of its components is equal to 1. The present paper deals with distributions that are invariant under linear combinations with coefficients W_1 and W_2. More precisely, we consider the class of distributions μ on the Borel subsets of the real line $\mathbb{R}$ with the property

$$X =_d W_1 \cdot X + W_2 \cdot X', \tag{1.1}$$

where W, X, X' are independent, '$=_d$' denotes equality in distribution and X, X' have distribution μ. Obviously, this class depends on the distribution of W only; furthermore, it is obvious that solutions of the stochastic difference equation (1.1) can only be unique up to a scale factor. Writing $\mathcal{L}(Z)$ for the distribution ('law') of a random quantity Z we can express the latter fact as follows: if $\mathcal{L}(X)$ solves (1.1) then so does $\mathcal{L}(cX)$ for all $c \in \mathbb{R}$. There is always a trivial solution, the unit mass δ_0 in 0.

Stochastic equations such as (1.1) have been considered by many authors, partly as being of interest in their own right, partly because they arise in the context of other problems such as from branching processes, infinite particle systems and probabilistic algorithms.

The special case of convex combinations with non-random coefficients is classical. The set of solutions to $pX + (1-p)X' =_{\mathrm{d}} X$ with some fixed p, $0 < p < 1$, has an additional invariance under shifts, i.e. $\mathcal{L}(X + c)$ is a solution if $\mathcal{L}(X)$ is, for all $c \in \mathbb{R}$. All Cauchy distributions solve this degenerate version of (1.1). For $p = 1/2$ there is another solution, not from this location-scale family, which is due to Lévy (see Feller (1971), p. 567, and Pakes (1992*a*), Section 2).

In the special case where W_1 and W_2 are independent uniform $[0, 1]$ variables and X, X' are nonnegative the only solutions (i.e. distributions of X) of (1.1) are the Erlang distributions of order 2. This was found by Huang and Chen (1989) when treating the following more general problem. Let the i.i.d. variables $W, W_1, \ldots, W_{m+n}$ be independent of the i.i.d. variables $X, X_1, \ldots, X_{m+n}$. The problem is to give the solutions $\mathcal{L}(X)$ of

$$X_1 + \cdots + X_m =_{\mathrm{d}} W_1 X_1 + \cdots + W_{m+n} X_{m+n}.$$

Huang and Chen found the nonnegative solutions in the case $\mathcal{L}(W) = \mathrm{beta}(\alpha, 1)$, $\alpha > 0$. A variant of this problem treated by Huang and Chen and also by several other authors is to identify the laws $\mathcal{L}(X)$ satisfying

$$X_1 + \cdots + X_m =_{\mathrm{d}} W(X_1 + \cdots + X_{m+n}), \tag{1.2}$$

see Steutel (1985), Kotz and Steutel (1988), Yeo and Milne (1989, 1991), and Pakes (1992*a*, 1992*b*, 1994*a*). In the simple case where $m = n = 1$, and $\mathcal{L}(W)$ is the uniform distribution on $[0, 1]$, the only nonnegative solutions are the exponential distributions. Pakes and Khattree (1992) and Pakes (1994*b*) considered the equation

$$X =_{\mathrm{d}} WX(r)$$

where X with distribution function F is nonnegative with finite moment $\mu_r = \int_0^\infty y^r F(dy)$ of order $r > 0$, $X(r)$ has the r-th-order length biased law with distribution function $\mu_r^{-1} \int_0^x y^r F(dy)$, $x \geq 0$, and W and $X(r)$ are independent. Again, in the case $r = 1$, and $\mathcal{L}(W)$ the uniform distribution on $[0, 1]$, a characterization of the exponential distribution is obtained. Other generalizations and variants of (1.1) have been considered in Pakes (1995) and van Harn and Steutel (1992, 1993).

Equations of the type (1.1) also arise in the theory of branching random walks; see Biggins (1977) and the references given there. Durrett and Liggett (1983) were motivated by problems in infinite particle systems; again, the references therein point to important related work. In both situations more than two summands are considered, in Biggins (1977) this number may even be random. For notational convenience we will stick to the simple form (1.1), but see Section 5 below. Both papers restrict their analysis to distributions concentrated on $[0, \infty)$, i.e. it is additionally assumed that the weights W_i and the X-variables are nonnegative. This is natural from the origin of the problems, where e.g. $\log X$ denotes the shift in position of the offspring. On a technical level this additional assumption is convenient as it allows comparison arguments and the use of Laplace transforms. Rösler (1992) studied equations of the type (1.1) with possibly infinitely many summands and allowed for an additional shift; this was motivated by the analysis of the quicksort algorithm.

In Section 2 we investigate (1.1) for general weights on using contraction arguments, an idea apparently due to Rösler (1992). The basic idea of this approach is to regard (1.1) as a fixed point equation of a suitable map on some suitable metric space of probability measures and to use a fixed point theorem. Rösler used Mallows metrics to obtain a condition for existence and uniqueness of solutions to (1.1) within the space of distributions with finite second moment. We introduce a new class of distances of probability measures on the real line and show that it can be used to obtain existence and uniqueness under weaker assumptions, and in larger spaces of distributions. We also relate this method to the general theory of probability metrics initiated by Zolotarev (see e.g. Zolotarev (1976) and Rachev (1991)).

One starting point for the present investigation was the observation that exponential distributions solve (1.1) if one of the components of W is constant and the other takes the values 0 and 1 only. This is a classical fact from queueing theory, though it has apparently never been exposed in this way. Details are given in Section 3 where we also relate this special case to a different and well understood class of characterization problems which involve geometric random sums. The general results from Section 2 can be used to obtain a characterization of exponential distributions under a mild moment condition. In the classical approach the same characterization (for the transformed problem) is obtained on solving the associated functional equation for the characteristic function of the solution, a standard technique in this area. Whereas the classical result uses slightly weaker assumptions the method of the present paper has the advantage of providing stability results at no extra cost.

In Section 4 we approach (1.1) via the associated moments; these can easily be obtained recursively. This simple method results in yet another proof of the above characterization of exponential distributions, this time under the condition that all moments exist. Further, if W takes its values in the unit square as in the case that characterizes exponential distributions, then all moments of the solutions to (1.1) exist if a moment of order r for some $r > 1$ exists. In the final Section 5 we indicate extensions of our results to more than two summands and to spaces more general than the real line.

2. Existence and uniqueness of solutions

Throughout this section we assume that, in addition to $E(W_1 + W_2) = 1$,

$$\kappa(r) := E(|W_1|^r + |W_2|^r) < 1 \quad \text{for some} \quad r \in (1,2). \tag{2.1}$$

Here, as in the following, positive powers of 0 are understood to be 0. For a given $c \in \mathbb{R}$ let $\mathcal{M}_{r,c}$ be the set of probability measures μ on the Borel subsets of $\mathbb{R}$ with the properties $\int |x|^r \mu(dx) < \infty$ and $\int x\mu(dx) = c$. For each $\mu \in \mathcal{M}_{r,c}$ let $T(\mu)$ be the distribution of $W_1X + W_2X'$ where W, X, X' are independent and $\mathcal{L}(X) = \mathcal{L}(X') = \mu$. From Minkowski's inequality and the independence assumptions it follows easily that T maps $\mathcal{M}_{r,c}$ into $\mathcal{M}_{r,c}$. Further let $\mathcal{F}_{r,c}$ be the set of Fourier transforms of probability distributions μ in $\mathcal{M}_{r,c}$. Then T can also be regarded as a map from $\mathcal{F}_{r,c}$ into $\mathcal{F}_{r,c}$, with

$$T(\phi)(t) = E(\phi(tW_1)\phi(tW_2)) \quad \text{for all} \quad \phi \in \mathcal{F}_{r,c}, \ t \in \mathbb{R}.$$

We will freely switch between measures and their Fourier transforms. The following result shows that, with a suitable notion of distance $\mathcal{F}_{r,c}$, and hence $\mathcal{M}_{r,c}$, is a complete metric space.

LEMMA 2.1. *With $r \in (1,2)$, d_r defined by*

$$d_r(\phi_1, \phi_2) := \int |\phi_1(t) - \phi_2(t)||t|^{-r-1}dt \qquad \text{for all} \quad \phi_1, \phi_2 \in \mathcal{F}_{r,c},$$

is a metric on $\mathcal{F}_{r,c}$, and $(\mathcal{F}_{r,c}, d_r)$ is complete.

PROOF. Let $\phi_k \in \mathcal{F}_{r,c}$ with associated distribution μ_k, $k = 1, 2$. Putting $\psi_k(t) = 1 - \phi_k(t) + ict$ we have that $d_r(\phi_1, \phi_2) = \int |\psi_1(t) - \psi_2(t)||t|^{-r-1}dt$. From $\mathrm{Re}(\psi_k(t)) = 1 - \mathrm{Re}(\phi_k(t))$ and the identity

$$\int_{-\infty}^{+\infty} (1 - \cos(xt))|t|^{-\alpha-1}dt = K(\alpha)^{-1}|x|^{\alpha}$$

with $K(\alpha) = \pi^{-1}\Gamma(\alpha + 1)\sin(\alpha\pi/2)$ being true for $0 < \alpha < 2$ (see von Bahr and Esseen (1965)) it follows that

$$\int_{-\infty}^{+\infty} \mathrm{Re}(\psi_k)(t)|t|^{-r-1}dt = K(r)^{-1} \int |x|^r \mu_k(dx). \tag{2.2}$$

Therefore $\mathrm{Re}(\psi_k)$ is integrable with respect to the σ-finite measure ν_r which has density $|t|^{-r-1}$, $t \in \mathbb{R}$, with respect to Lebesgue measure. Using $\sin y = y - \int_0^y (1 - \cos s)ds$, $y \in \mathbb{R}$, we obtain

$$\begin{aligned} \mathrm{Im}(\psi_k(t)) = ct - \mathrm{Im}(\phi_k(t)) &= \iint_0^{xt} (1 - \cos s)ds\mu_k(dx) \\ &= \int x \int_0^t (1 - \cos(xs))ds\mu_k(dx). \end{aligned}$$

Applying Fubini's Theorem we see that

$$\begin{aligned} &\iint |x| \int_0^t (1 - \cos(xs))ds\mu_k(dx)\nu_r(dt) \\ &= 2\int_0^{\infty} \int |x| \int_0^t (1 - \cos(xs))ds\mu_k(dx)t^{-r-1}dt \\ &= 2\iint_0^{\infty} \int_s^{\infty} t^{-r-1}dt(1 - \cos(xs))ds|x|\mu_k(dx) \\ &= 2\int r^{-1} \int_0^{\infty} (1 - \cos(xs))s^{-r}ds|x|\mu_k(dx) \\ &= \int r^{-1} \int_{-\infty}^{+\infty} (1 - \cos(xs))|s|^{-r}ds|x|\mu_k(dx) \\ &= (rK(r-1))^{-1} \int |x|^r \mu_k(dx) < \infty. \end{aligned}$$

Therefore $\mathrm{Im}(\psi_k)$ is also integrable with respect to ν_r and it follows that $d_r(\phi_1, \phi_2)$ is finite, and it is then easily seen that d_r is a metric on $\mathcal{F}_{r,c}$.

To prove that the metric space $(\mathcal{F}_{r,c}, d_r)$ is complete let (ϕ_n) be a Cauchy sequence. Since the space of functions integrable with respect to ν_r is complete there is some function ψ integrable with respect to ν_r such that $\lim_{n\to\infty} \int |\psi_n - \psi| d\nu_r = 0$. It follows that $\lim_{n\to\infty} \int |\mathrm{Re}(\psi_n) - \mathrm{Re}(\psi)| d\nu_r = 0$, implying $\lim_{n\to\infty} \cdot \int \mathrm{Re}(\psi_n) d\nu_r = \int \mathrm{Re}(\psi) d\nu_r$. Thus, with μ_n denoting the distribution associated with ϕ_n, we have in view of (2.2) that

$$\sup_{n\in\mathbb{N}} \int |x|^r \mu_n(dx) < \infty. \tag{2.3}$$

This implies that the sequence of distributions μ_n is tight, hence there is a sub sequence (μ_{n_k}) such that the μ_{n_k} converge in distribution to a distribution μ as $k \to \infty$. Because of (2.3) it follows that $\int x\mu(dx) = c$. With ϕ denoting the Fourier transform of μ the continuity theorem for Fourier transforms implies

$$\lim_{k\to\infty} \psi_{n_k}(t) = 1 - \phi(t) + ict \qquad \text{for all} \quad t \in \mathbb{R}.$$

It follows that $\psi(t) = 1 - \phi(t) + ict$ for ν_r-almost all $t \in \mathbb{R}$. This proves that the metric space is complete. □

If X and Y are random variables with characteristic functions ϕ_X, ϕ_Y and distributions μ_X, μ_Y respectively we will occasionally write $d_r(X, Y)$ or $d_r(\mu_X, \mu_Y)$ instead of $d_r(\phi_X, \phi_Y)$. It is easy to see that, for any Z independent of X and Y, and any $c \in \mathbb{R}$,

$$d_r(X + Z, Y + Z) \le d_r(X, Y), \tag{2.4}$$

$$d_r(cX, cY) = |c|^r d_r(X, Y), \tag{2.5}$$

which means that d_r is an ideal probability metric in the sense of Zolotarev (1976).

The following inequality will be crucial for the proof of the main result of this section.

LEMMA 2.2. *$d_r(T(\phi_1), T(\phi_2)) \le E(|W_1|^r + |W_2|^r) d_r(\phi_1, \phi_2)$ for all $\phi_1, \phi_2 \in \mathcal{F}_{r,c}$.*

PROOF. Let ϕ_1 and ϕ_2 be elements of $\mathcal{F}_{r,c}$. Then the assertion follows from

$$\begin{aligned}
&\int_{-\infty}^{+\infty} |E(\phi_1(tW_1)\phi_1(tW_2)) - E(\phi_2(tW_1)\phi_2(tW_2))| |t|^{-r-1} dt \\
&\quad \le E\left(\int_{\infty}^{+\infty} |\phi_1(tW_1) - \phi_2(tW_1)| |t|^{-r-1} dt\right) \\
&\qquad + E\left(\int_{-\infty}^{+\infty} |\phi_1(tW_2) - \phi_2(tW_2)| |t|^{-r-1} dt\right) \\
&\quad = E(|W_1|^r + |W_2|^r) \int_{-\infty}^{+\infty} |\phi_1(t) - \phi_2(t)| |t|^{-r-1} dt \\
&\quad = E(|W_1|^r + |W_2|^r) d_r(\phi_1, \phi_2).
\end{aligned}$$

□

The above proof can alternatively be formulated in terms of ideal probability metrics, using (2.4) and (2.5).

THEOREM 2.1. *Let the distribution of W be such that* (2.1) *holds, and let $c \in \mathbb{R}$ be given. Then there exists a unique solution $\mu = \mathcal{L}(X)$ of* (1.1) *that satisfies $\int |x|^r \mu(dx) < \infty$ and $\int x\mu(dx) = c$. Further, for every $\mu^{(0)} \in \mathcal{M}_{r,c}$ the sequence $(\mu^{(n)})_{n\in\mathbb{N}}$, $\mu^{(n)} := T^n(\mu^{(0)})$, converges to μ at an exponential rate with respect to d_r.*

PROOF. By Lemma 2.2 the assumption $\kappa(r) < 1$ implies that T is a strict contraction on $(\mathcal{M}_{r,c}, d_r)$, which is a complete metric space by Lemma 2.1. Hence T has a unique fixed point in $\mathcal{M}_{r,c}$, this fixed point solves (1.1), and the iterates of any element of $\mathcal{M}_{r,c}$ converge to μ at an exponential rate in this space. Also, any solution of (1.1) in $\mathcal{M}_{r,c}$ is a fixed point of T. □

COROLLARY 2.1. *If $c = 0$ then $\mu = \delta_0$ is the only solution of* (1.1) *in $\mathcal{M}_{r,c}$. If $P(W_1 \geq 0) = P(W_2 \geq 0) = 1$ holds in addition to the conditions of Theorem* 2.1, *then the support of the solution is concentrated on a half line:*

$$c > 0 \Rightarrow \mu([0,\infty)) = 1, \qquad c < 0 \Rightarrow \mu((-\infty, 0]) = 1, \qquad c = 0 \Rightarrow \mu = \delta_0.$$

PROOF. Let $\mu^{(n)} := T^n(\delta_c)$ where δ_c denotes the unit mass in c. Clearly, $\mu^{(n)} = \delta_0$ if $c = 0$. The additional assumption on the weights implies that $T(\mu)$ is concentrated on $[0,\infty)$ if μ is, and similarly for $(-\infty, 0]$. By Theorem 2.1, the sequence $\mu^{(n)}$ converges to the solution with respect to d_r, and the statement now follows on noting that convergence with respect to the metric d_r implies weak convergence. □

A noteworthy aspect of the above contraction approach to characterization problems is the fact that stability results for the characterization can be obtained along with the characterization without additional effort: if μ_0 solves (1.1) in $\mathcal{F}_{r,c}$ and if the characterization equation is approximately true for some $\mu_1 \in \mathcal{F}_{r,c}$ in the sense that $d_r(\mu_1, T(\mu_1)) \leq \delta$, then μ_1 is close to μ_0:

$$d_r(\mu_1, \mu_0) \leq \frac{\delta}{1 - E(|W_1|^r + |W_2|^r)}.$$

For general comments on the stability of characterizations and the use of probability metrics in this context we refer the reader to the seminal paper by Zolotarev (1976); see also Rachev (1991).

Pakes (1992*b*) proves the analogue of the uniqueness part of Theorem 2.1 for equation (1.2) by bounding the pointwise distance of characteristic functions; see also Athreya (1969). Rösler (1992) obtained the analogue of Theorem 2.1 for $r = 2$ on using a Mallows metric. Note that $\kappa(2) < 1$ implies $\kappa(r) < 1$ for some $r \in (1,2)$; for nonnegative weights $E(W_1 + W_2) = 1$ and the convexity of $r \to EW_i^r$, $r > 0$, imply that $\kappa(r_2) < 1$ for all $r_2 \in (1, r_1)$ if $\kappa(r_1) < 1$. Hence the above theorem gives a sufficient condition for the existence of a solution to (1.1) under assumptions which are weaker than in the aforementioned paper.

Durrett and Liggett (1983) obtain the existence of a fixed point under assumptions weaker than ours, but they restrict their analysis to measures concentrated on a half line. Note that the contractive property is stronger than the existence of a fixed point; in particular, the unique solution with mean c can be obtained by iteration starting with an arbitrary element of $\mathcal{M}_{r,c}$. They also prove uniqueness results which imply that, under the conditions of Theorem 2.1, the solution to (1.1) is unique in the space of all distributions that are concentrated on $[0,\infty)$. These results, however, require the assumption that the solution is concentrated on a half line: if $W_1 \equiv p$, $W_2 \equiv 1-p$ then $\kappa(r) < 1$ for all $r > 1$, and the unique solution in $\mathcal{M}_{r,c}$ is the unit mass in c. In the larger space of all probability measures on the real line, the Cauchy distributions appear as additional solutions, so uniqueness is lost. This phenomenon is not tied to the W-components being concentrated in a single point; indeed, as the following theorem shows, which is similar in spirit to Theorem 3.1 in Pakes (1992*b*), from any particular solution of (1.1) a whole family of solutions (with non-existing mean) can be obtained by multiplication with Cauchy variables.

THEOREM 2.2. *Suppose that* $P(W_1 \geq 0) = P(W_2 \geq 0) = 1$ *holds in addition to the conditions of Theorem* 2.1. *If* $\mu^{(0)} = \mathcal{L}(X)$ *is concentrated on a half line and solves* (1.1), *then* $\mu^{(1)} := \mathcal{L}(XZ)$ *also solves* (1.1), *where* Z *is independent of* X *and Cauchy distributed with median* 0.

PROOF. We may assume that Z has scale parameter 1 and that $\mu^{(0)}$ is concentrated on $[0,\infty)$. Then, with Z' an independent copy of Z that is also independent of X, X' and W,

$$\begin{aligned} &E(\exp(it(W_1XZ + W_2X'Z'))) \\ &\quad = E(E[\exp(it(W_1XZ + W_2X'Z')) \mid W_1, W_2, X, X']) \\ &\quad = E\exp(-|t(W_1X + W_2X')|) \\ &\quad = E\exp(-|tX|) \\ &\quad = E\exp(itXZ). \end{aligned}$$

This is the Fourier transform version of (1.1) for $\mu^{(1)}$. □

Hence, for nonnegative weights, if we let $\mathcal{M}_{1+}$ and $\mathcal{M}_{1-}$ denote the set of all distributions μ with $\int |x|^r \mu(dx) < \infty$ for some $r > 1$ and for all $r < 1$ respectively, then we have shown that (2.1) implies uniqueness up to a scale factor in $\mathcal{M}_{1+}$, and that this uniqueness is lost if we slightly enlarge the space to $\mathcal{M}_{1-}$.

3. A special case involving exponential distributions

In a stationary $M/M/1$ queue with arrival rate λ and departure rate $\tau > \lambda$ the virtual waiting time V_t of a customer arriving at time t and his service time S_t are independent, S_t is exponentially distributed with parameter τ and the distribution of V_t is a mixture of δ_0 and an exponential distribution with parameter $\tau - \lambda$. The total time $V_t + S_t$ in the system is again exponentially distributed with parameter $\tau - \lambda$ (see e.g. Bhat (1972), Section 11.3). Let $p := \lambda/\tau < 1$ be the traffic intensity of the queue and let W_1 be such that $P(W_1 = 1) = 1 - P(W_1 = 0) = p$. Then, with X, X' independent and exponentially distributed with parameter $\tau - \lambda$ and independent of W_1, we have that $W_1 X$ is equal in distribution to V_t and that $(1-p)X'$ is equal in distribution to S_t, i.e. (1.1) is satisfied with $W_2 \equiv 1 - p$. The results of Section 2 show that exponential distributions are characterized by this property within the space of all probability measures with finite moment of order r for some $r > 1$.

In the above case of $\{0,1\}$-valued W_1 and constant W_2 equation (1.1) is closely related to the problem of characterizing the solutions $\mu = \mathcal{L}(X)$ of the equation

$$X =_{\mathrm{d}} (1-p) \sum_{k=1}^{N} X_k, \tag{3.1}$$

where $N, X_1, X_2, \ldots$ are independent, $\mathcal{L}(X_k) = \mu$ for all $k \in \mathbb{N}$, and $P(N = n) = (1-p)p^{n-1}$ for all $n \in \mathbb{N}$ (geometric compounding). Indeed, any solution of (1.1) with W of this special form also solves (3.1) and vice versa. This is easily seen once both problems are rewritten in terms of the characteristic function ϕ of the solution: in both cases the same functional equation arises. The step from (1.1) to (3.1) for $\{0,1\}$-valued W_1 is also obvious if one iterates (1.1), replacing the first X-variable on the right hand side by an independent copy of the right hand side of (1.1): if $W_{1,k}$, $k \in \mathbb{N}$, are independent random variables with the same distribution as W_1 then the cumulative products $\prod_{k=1}^{n-1} W_{1,k}$, $n \in \mathbb{N}$, are equal to 1 for $n \leq N$ and vanish from $N+1$ onwards, where N has a geometric distribution as in connection with (3.1).

Stochastic equations of the form (3.1) have been considered by many authors. Arnold (1973) showed that exponential distributions are the only solutions within the space of probability distributions with finite first moment or the space of distributions concentrated on the positive half line. Arnold's proof proceeds by solving the functional equation associated with the characteristic function version of (3.1). With the method of the present paper the same characterization can be obtained, albeit under the condition that some moment of order $r > 1$ of the solution is finite. The benefit of the stronger assumption is a corresponding stability result as explained in the previous section.

Without the assumption of finiteness of the first moment or the assumption of one-sided support uniqueness does not hold: applying Theorem 2.2 we see that the distributions with characteristic function $\phi(t) = 1/(1 + \sigma|t|)$, $\sigma > 0$, also solve (1.1) if W is as in the present section; again, this has already been noted by Arnold (1973). These distributions are special Pólya type distributions, see e.g. Lukacs (1970); they are also known as Linnik laws, see Devroye (1990) and Pakes (1995).

4. Moments

Theorem 2.1 gives a general condition for (1.1) to have a unique solution, up to a scale factor, but does not provide the solution itself. In the special case considered in the previous section the solution could be obtained from solving a functional equation. In this section we deal with the moments of solutions of (1.1) which occasionally provides a more constructive angle on the problem. Throughout this section we assume that

$$P(0 \le W_1 \le 1) = P(0 \le W_2 \le 1) = 1. \tag{4.1}$$

If (1.1) holds and all moments of X are finite, then

$$\begin{aligned}\mu_n := EX^n &= E(W_1X + W_2X')^n \\ &= \sum_{k=0}^{n} \binom{n}{k} EW_1^k W_2^{n-k} \mu_k \mu_{n-k}.\end{aligned}$$

If conditions (2.1) and (4.1) are satisfied then $EW_1^n + EW_2^n < 1$ for all $n > 1$ and it follows that

$$\mu_n(1 - EW_1^n - EW_2^n) = \sum_{k=1}^{n-1} \binom{n}{k} EW_1^k W_2^{n-k} \mu_k \mu_{n-k} \tag{4.2}$$

for all $n \ge 2$. Hence, if all moments are finite, then these can easily be obtained recursively from $\mu_0 = 1$ and any given μ_1.

This argument leads to yet another proof of the characterization of exponential distributions discussed in the previous section, now under the assumption that all moments of the solution are finite: if $P(W_1 = 1) = 1 - P(W_1 = 0) = p$ and $P(W_2 = 1-p) = 1$ then $EW_1^n = p$, $EW_2^n = (1-p)^n$ and $EW_1^k W_2^{n-k} = p(1-p)^{n-k}$ so that (4.2) becomes

$$\mu_n(1 - p - (1-p)^n) = \sum_{k=1}^{n-1} \binom{n}{k} p(1-p)^{n-k} \mu_k \mu_{n-k}.$$

It is straightforward to check that this is solved by $\mu_n = n!\mu_1^n$, and these are the moments of the exponential distribution with mean μ_1.

This approach leads us to look for conditions on W that imply the existence of moments of solutions to (1.1). This question has also been considered by Durrett and Liggett (1983) who conjecture that "the effect of the spread of W should be only to determine the size of the tails of the stationary measure" (p. 281). The following theorem shows that any solution which has a finite moment of order r for some $r > 1$ automatically has finite moments of all orders.

THEOREM 4.1. *If (4.1) holds and W_1 and W_2 are not both concentrated on $\{0, 1\}$ then any solution X to (1.1) has the following property:*

$$\exists r > 1 : E|X|^r < \infty \longrightarrow \exists \alpha > 0, C < \infty \ \forall x \ge 0 : P(|X| \ge x) \le C \exp(-\alpha x).$$

PROOF. The assumptions on W imply that $\kappa(r) := E(W_1^r + W_2^r) < 1$ for any $r > 1$, hence we may assume on using Corollary 2.1 that X is concentrated on the nonnegative half line; after suitable rescaling we may also assume that $EX = 1$ (the statement is trivially true in the degenerate case $X \equiv 0$). Finally, we may assume that $r < 2$.

Let T be as defined at the beginning of Section 2 and let $\mathcal{M}(r, t_0)$ be the set of all probability measures μ on $[0, \infty)$ with the property

$$\int e^{tx} \mu(dx) \leq 1 + t + t^r \quad \text{for} \quad 0 \leq t \leq t_0.$$

We first prove the following:

$$\exists t_0 > 0 \; \forall \mu : \mu \in \mathcal{M}(r, t_0) \Rightarrow T(\mu) \in \mathcal{M}(r, t_0). \tag{4.3}$$

Let W, X, X' be independent with $\mathcal{L}(X) = \mathcal{L}(X') = \mu$, and let $M(t) := E\exp(tX)$. Then, if $\mu \in \mathcal{M}(r, t_0)$, and with $\kappa = \kappa(r)$,

$$\begin{aligned}
\int e^{tx} T(\mu)(dx) &= EM(tW_1)M(tW_2) \\
&\leq E(1 + tW_1 + t^r W_1^r)(1 + tW_2 + t^r W_2^r) \\
&= 1 + tE(W_1 + W_2) + t^r E(W_1^r + W_2^r) \\
&\quad + t^2 EW_1 W_2 + t^{r+1} E(W_1 W_2^r + W_1^r W_2) + t^{2r} EW_1^r W_2^r \\
&\leq 1 + t + \kappa t^r + t^2 + 2t^{r+1} + t^{2r}.
\end{aligned}$$

This last term is less than or equal to $1 + t + t^r$ if $t^2 + t^{2r} + 2t^{r+1} \leq (1 - \kappa)t^r$. As $\kappa < 1$ this can be made to hold for all $t \in [0, t_0]$ by choosing $t_0 > 0$ small enough; note that t_0 does not depend on μ.

To see that (4.3) implies the statement of the theorem we use the argument from the end of the proof of Theorem 2.1: if μ is the (unique) solution of (1.1) in $\mathcal{M}_{r,1}$ then μ is the weak limit of $\mu^{(n)}$ as $n \to \infty$ where $\mu^{(n)}$ arises by applying T n times to the unit mass in 1. All these are in $\mathcal{M}(r, t_0)$ for a suitable fixed $t_0 > 0$, which means, by Markov's inequality, that

$$\mu^{(n)}([x, \infty)) \leq C \exp(-t_0 x) \quad \text{for all} \quad x \geq 0, \; n \in \mathbb{N}$$

with some suitable constant $C > 0$ not depending on n. Weak convergence implies that the same inequality holds for μ, hence all moments of μ are finite. □

The arguments of the proof can be used to obtain a suitable α if κ is given explicitly.

Again, the case of constant weights with the Cauchy distributions as solutions, shows that the assumption that $E|X|^r$ is finite for some $r > 1$ cannot be dropped. Rösler ((1992), Section 4) obtained a result similar to Theorem 4.1 but under the (implicit) assumption that $EX^2 < \infty$.

The case of $\{0,1\}$-valued weights, which is excluded in the above theorem, is interesting as it provides us with two extreme examples: if, additionally, $P(W_1 + W_2 = 1) = 1$, then all probability distributions solve (1.1), if $P(W_1 = W_2 = 1) > 0$ (e.g. weights determined by coin-tossing), then there are no non-trivial solutions. This latter fact can easily be seen on writing down the functional equation for the Fourier transform of a potential solution (this contradicts a statement made by Durrett and Liggett ((1983), p. 280), but does not affect their theorems).

5. Extensions

Biggins (1977), Durrett and Liggett (1983) and Rösler (1992) considered analogues of (1.1) with more than two summands on the right hand side. Our results carry over to this situation, the crucial step being an extension of the inequality given in Lemma 2.2. Let $W = (W_0, W_1, \dots, W_N)$ be a random element of the set $\bigcup_{n=1}^{\infty} \mathbb{R}^n$ of finite sequences of real numbers. We assume that $EW_0 = 0$ and $E(\sum_{n=1}^{N} W_n) = 1$. Let $r \in (1,2)$ and $c \in \mathbb{R}$ be given and consider the transformation

$$T : \mathcal{M}_{r,c} \to \mathcal{M}_{r,c}, \qquad T(\mu) := \mathcal{L}\left(W_0 + \sum_{k=1}^{N} W_k X_k \right),$$

where $(X_k)_{k\in\mathbb{N}}$ is a sequence of independent random variables with distribution μ, independent of W. Using the elementary fact

$$\left| \prod_{k=1}^{n} a_k - \prod_{k=1}^{n} b_k \right| \le \sum_{k=1}^{n} |a_k - b_k| \qquad \text{for all} \quad n \in \mathbb{N},\ a_k, b_k \in \{z \in \mathbb{C} : |z| \le 1\}$$

we obtain for any $\mu, \nu \in \mathcal{M}_{r,c}$

$$\begin{aligned} d_r(T(\mu), T(\nu)) &= \int \left| E\left(e^{itW_0} \sum_{n=1}^{\infty} 1_{\{N=n\}} \left(\prod_{k=1}^{n} \phi_\mu(tW_k) - \prod_{k=1}^{n} \phi_\nu(tW_k) \right) \right) \right| |t|^{-r-1} dt \\ &\le E\left(\sum_{n=1}^{\infty} 1_{\{N=n\}} \int \sum_{k=1}^{n} |\phi_\mu(tW_k) - \phi_\nu(tW_k)| |t|^{-r-1} dt \right) \\ &= E\left(\sum_{n=1}^{\infty} 1_{\{N=n\}} \sum_{k=1}^{n} |W_k|^r \int |\phi_\mu(t) - \phi_\nu(t)| |t|^{-r-1} dt \right) \\ &= E\left(\sum_{n=1}^{N} |W_n|^r \right) d_r(\mu, \nu), \end{aligned}$$

hence the analogue of Lemma 2.2 holds.

A second straightforward generalization concerns the range space of the X-variables in (1.1). Let $\mathbb{H}$ be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$. A random quantity X with values in $\mathbb{H}$ has mean $c \in \mathbb{H}$ if $E\langle t, X\rangle = \langle t, c\rangle$ for all $t \in \mathbb{H}$, its characteristic function is

$$\phi : \mathbb{H} \to \mathbb{C}, \qquad \phi(t) := E \exp(i\langle t, X\rangle).$$

Let $\mathcal{M}_{r,c}$ be the space of probability measures $\mu = \mathcal{L}(X)$ on $\mathbb{H}$ with $E\|X\|^r < \infty$ and $EX = c$; $\mathcal{F}_{r,c}$ denotes the set of the associated Fourier transforms. Let

$$d_r(\phi_1, \phi_2) := \sup_{x\in\mathbb{H}, \|x\|=1} \int |\phi_1(tx) - \phi_2(tx)||t|^{-r-1} dt.$$

The bounds obtained in the proof of Lemma 2.1 show that this is finite for all $\phi_1, \phi_2 \in \mathcal{F}_{r,c}$. It is easy to see that d_r is a metric on $\mathcal{F}_{r,c}$ (and therefore on $\mathcal{M}_{r,c}$). Let T denote the Hilbert space extension of the operator T defined at the beginning of Section 2; it is straightforward to show that the analogue of the inequality in Lemma 2.2 holds. Hence, if W satisfies (2.1), then the solution of (1.1) must be unique in $\mathcal{M}_{r,c}$.

It is easy to obtain such a solution for any given $c \in \mathbb{H}$: let X_0 be the unique solution with mean 1 of the corresponding one-dimensional equation, let ϕ_0 be its characteristic function. Then

$$\phi : \mathbb{H} \to \mathbb{C}, \qquad \phi(t) := \phi_0(\langle t, c\rangle),$$

which is the transform of cX_0, satisfies the transform version of (1.1):

$$\begin{aligned} E\phi(tW_1)\phi(tW_2) &= E(\phi_0(\langle t, c\rangle W_1)\phi_0(\langle t, c\rangle W_2)) \\ &= \phi_0(\langle t, c\rangle) = \phi(t). \end{aligned}$$

Hence the solution is concentrated on the scalar multiples of the preassigned mean; this can also be seen by a more direct argument. An extension of this approach to weights that are random linear operators is currently under investigation.

Acknowledgements

We would like to thank Professors A. Pakes, U. Rösler and F. W. Steutel and the referees for providing us with some most relevant references and for their stimulating comments on previous versions of this paper.

REFERENCES

Arnold, B. C. (1973). Some characterizations of the exponential distribution by geometric compounding, *SIAM J. Appl. Math.*, **24**, 242–244.

Athreya, K. B. (1969). On the supercritical one dimensional age dependent branching process, *Ann. Math. Statist.*, **40**, 743–763.

Bhat, U. N. (1972). *Elements of Applied Stochastic Processes*, Wiley, New York.

Biggins, J. D. (1977). Martingale convergence in the branching random walk, *J. Appl. Probab.*, **14**, 25–37.

Devroye, L. (1990). A note on Linnik's distribution, *Statist. Probab. Lett.*, **9**, 305–306.

Durrett, R. and Liggett, Th. M. (1983). Fixed points of the smoothing transformation, *Z. Wahrsch. Verw. Gebiete*, **64**, 275–301.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Wiley, New York.

Huang, W-J. and Chen, L-S. (1989). Note on a characterization of gamma distributions, *Statist. Probab. Lett.*, **8**, 485–487.

Kotz, S. and Steutel, F. W. (1988). Note on a characterization of exponential distributions, *Statist. Probab. Lett.*, **6**, 201–203.

Lukacs, E. (1970). *Characteristic Functions*, Hafner, New York.

Pakes, A. G. (1992*a*). A characterization of gamma mixtures of stable laws motivated by limit theorems, *Statist. Neerlandica*, **46**, 209–218.

Pakes, A. G. (1992*b*). On characterizations through mixed sums, *Austral. J. Statist.*, **34**, 323–339.

Pakes, A. G. (1994*a*). Necessary conditions for characterization of laws via mixed sums, *Ann. Inst. Statist. Math.*, **46**, 797–802.

Pakes, A. G. (1994*b*). Characterization by invariance of the total lifetime under random scaling, Research Report 41, University of Western Australia.

Pakes, A. G. (1995). Characterization of discrete laws via mixed sums and Markov branching processes, *Stochastic Process Appl.*, **55**, 285–300.

Pakes, A. G. and Khattree, R. (1992). Length biasing, characterization of laws and the moment problem, *Austral. J. Statist.*, **34**, 307–322.

Rachev, S. T. (1991). *Probability Metrics and the Stability of Stochastic Models*, Wiley, New York.

Rösler, U. (1992). A fixed point theorem for distributions, *Stochastic Process Appl.*, **42**, 195–214.

Steutel, F. W. (1985). Problem 159, *Statist. Neerlandica*, **39**, 57–58.

van Harn, K. and Steutel, F. W. (1992). A generalization of the life-time pairs of renewal theory, Report WS-395, Vrije Universiteit Amsterdam.

van Harn, K. and Steutel, F. W. (1993). Stability equations for processes with stationary independent increments using branching processes and Poisson mixtures, *Stochastic Process Appl.*, **45**, 209–230.

von Bahr, B. and Esseen, C.-G. (1965). Inequalities for the rth absolute moment of a sum of random variables, $1 \le r \le 2$, *Ann. Math. Statist.*, **36**, 299–303.

Yeo, G. F. and Milne, R. K. (1989). On characterizations of exponential distributions, *Statist. Probab. Lett.*, **7**, 303–305.

Yeo, G. F. and Milne, R. K. (1991). On characterizations of beta and gamma distributions, *Statist. Probab. Lett.*, **11**, 239–342.

Zolotarev, V. M. (1976). Metric distances in spaces of random variables and their distributions, *Math. USSR-Sb.*, **30**, 373–401.