Bernoulli **13**(2), 2007, 330–345
DOI: 10.3150/07-BEJ5126

Exponential and gamma distributions on positive semigroups, with applications to Dirichlet distributions

KYLE SIEGRIST

Department of Mathematical Sciences, University of Alabama in Huntsville, Huntsville AL 35899, USA. E-mail: siegrist@math.uah.edu

In this paper, we give new characterizations of exponential distributions on positive semigroups and we study the corresponding gamma distributions that govern semigroup products of independent and identically distributed exponential variables. We apply these results to the positive integers under multiplication to obtain new interpretations of results on Dirichlet distributions.

Keywords: Dirichlet distribution; exponential distribution; gamma distribution; partially ordered set; positive semigroup; zeta distribution

1. Positive semigroups

A *positive semigroup* is a semigroup $(S, \cdot)$ which has an identity e, satisfies the left cancellation law, and has no non-trivial inverses (see Siegrist 1994; Rowell 1995; Rowell and Siegrist 1998). The associated partial order $\preceq$ is defined by

$$x \preceq y \text{ if and only if } y = xu, \qquad \text{for some } u \in S.$$

If $x \preceq y$ then $u \in S$ satisfying $xu = y$ is unique, and is denoted $x^{-1}y$. The partially ordered set $(S, \preceq)$ has a *self-similarity property*: $u \mapsto xu$ is an order-isomorphism from S onto $xS = \{y \in S : x \preceq y\}$ for each $x \in S$.

Topologically, we assume that S is a locally compact Hausdorff space with a countable base. To connect the algebraic and topological structures, we make the usual assumption that $(x, y) \mapsto xy$ is a continuous mapping from $S \times S$ into S. We also assume that $[e, x] = \{y \in S; y \preceq x\}$ is compact for each x, and that if $x \prec y$ then $[e, y]$ is a neighbourhood of x and xS is a neighbourhood of y. If S is countable, we give S the discrete topology, so that the assumption that $[e, x]$ is compact means that the partially ordered set $(S, \preceq)$ is *locally finite*.

The term *measure* on S will refer to a positive Borel measure; that is, a measure λ on the Borel sets $\mathcal{B}(S)$ such that $\lambda(K) < \infty$ if $K \subseteq S$ is compact. A measure λ on S is *left-invariant* if $\lambda(xA) = \lambda(A)$ for all $x \in S$ and $A \in \mathcal{B}(S)$. Our final assumption is that $(S, \cdot)$ has a left-invariant measure that is unique up to multiplication by positive constants. This assumption is always satisfied if S is discrete; counting measure $\#$ is left-invariant measure and is essentially unique. The assumption is also satisfied if S can be embedded in a nice way as the set of positive elements of a left-ordered topological group, since such a group has a left Haar measure, essentially unique (see Halmos 1974).

Positive semigroups are the natural home for probability distributions with exponential properties (such as the memorylessness and constant failure rate properties), and for reliability concepts (such as ageing properties). This paper has two main goals. First, we study the corresponding 'gamma' distributions that govern semigroup products of independent and identically distributed (i.i.d.) exponential variables, and this in turn leads to new and interesting characterizations of exponential distributions. Second, we apply the results to the positive semigroup $(\mathbb{N}_+, \cdot)$ to reinterpret a number of recent results by Lin and Hu (2001) and by Gut (2006) for the zeta and more general Dirichlet distributions. We believe that this example, in particular, makes a strong case for the value of positive semigroups as a unifying model. See Clifford and Preston (1964, 1967) for general information on semigroups, Högnäs and Mukherjea (1995) and Ruzsa and Székely (1988) for information on probability measures on semigroups, and Azlarov and Volidin (1986) for general characterizations of the standard exponential distribution on $([0, \infty), +)$.

2. Exponential distributions

In this section, we will review some basic facts from Siegrist (1994), and obtain some new results. Our starting point is a positive semigroup $(S, \cdot)$. We will restrict our attention to random variables whose distributions are Borel probability measures with support S. Probability density functions will be relative to a fixed, left-invariant measure λ.

A random variable X has an *exponential distribution* on $(S, \cdot)$ if

$$P(X \in xA) = P(X \succeq x)P(X \in A), \qquad x \in S,\ A \in \mathcal{B}(S). \tag{1}$$

Equivalently, the conditional distribution of $x^{-1}X$ given $X \succeq x$ is the same as the distribution of X for each $x \in S$. Next, X has a *memoryless distribution* if (1) holds for all $x \in S$ and all A of the form yS where $y \in S$. Equivalently,

$$P(X \succeq xy) = P(X \succeq x)P(X \succeq y), \qquad x \in S,\ y \in S,$$

so that the conditional tail probability function of $x^{-1}X$ given $X \succeq x$ is the same as the tail probability function of X. In the language of reliability theory, X has *constant failure rate* (or simply constant rate) if X has a probability density function f that is proportional to the tail probability function:

$$f(x) = \alpha P(X \succeq x), \qquad x \in S,$$

for some positive constant α. In general, for any positive measure μ on S,

$$\int_S P(X \succeq x)\mathrm{d}\mu(x) = \mathrm{E}(\mu[e, X])$$

Thus, if X has constant rate, then the rate constant must be

$$\alpha = \frac{1}{\mathrm{E}(\lambda[e, X])} \tag{2}$$

We will need the following results from Siegrist (1994):

Theorem 1. *A probability distribution on S is an exponential distribution if and only if the distribution is memoryless and has constant rate.*

In general, however, a distribution can have one of these properties but not the other.

Corollary 1. *Suppose that* $F : S \to (0, 1]$ *is measurable. Then* F *is the tail probability function on an exponential distribution if and only if* $F(xy) = F(x)F(y)$ *for all* $x, y \in S$, *and*

$$\int_S F(x) \mathrm{d}\lambda(x) < \infty. \tag{3}$$

Of course, when the conditions in Corollary 1 are met, then the corresponding proportional density function is $f = \alpha F$, where α is the reciprocal of the integral in (3). The following proposition gives a simple, new characterization of exponential distributions based on the probability density function.

Proposition 1. *Suppose that* f *is a probability density function on* S *and that*

$$f(x)f(y) = G(xy), \qquad x \in S, \, y \in S$$

for some measurable function $G : S \to (0, \infty)$. *Then* f *is the density of an exponential distribution.*

Proof. First we let $y = e$ to conclude that $G(x) = \alpha f(x)$ for $x \in S$, where $\alpha = f(e) > 0$. Let F denote the tail probability function. Then, using a change of variables and the fact that λ is left-invariant, we have

$$\begin{aligned} F(x) &= \int_{xS} f(y) \mathrm{d}\lambda(y) = \int_{xS} \frac{1}{\alpha} G(y) \mathrm{d}\lambda(y) = \frac{1}{\alpha} \int_S G(xu) \mathrm{d}\lambda(u) \\ &= \frac{1}{\alpha} \int_S f(x) f(u) \mathrm{d}\lambda(u) = \frac{1}{\alpha} f(x), \qquad x \in S. \end{aligned}$$

Thus, the distribution has constant rate α. Finally,

$$F(xy) = \frac{1}{\alpha} f(xy) = \frac{1}{\alpha^2} G(xy) = \frac{1}{\alpha^2} f(x) f(y) = F(x)F(y),$$

so the distribution is memoryless. □

The condition in Proposition 1 is essentially that $f(x)f(y)$ depends only on the product xy. Clearly it suffices to have this condition hold for all $x, \; y \in S$ with $xy \neq e$, since e has only the trivial factoring.

3. Gamma distributions

Suppose that X and Y are independent random variables taking values in S, with probability density functions f and g, respectively. Then XY has probability density function $f * g$, the convolution of f with g, given by

$$(f * g)(y) = \int_{[e,y]} f(x) g(x^{-1}y) \mathrm{d}\lambda(x).$$

More generally, the convolution operation makes sense for functions $f, g : S \to \mathbb{R}$, assuming that the integral exists. With our algebraic assumptions, convolution is associative, but not commutative in general.

If $(X_1, X_2, \ldots)$ is a sequence of independent random variables on S, with a common exponential distribution, then for $n \in \mathbb{N}_+$ we will say that

$$Y_n = X_1 X_2 \cdots X_n$$

has the corresponding *gamma distribution* on S of order n.

We will need a special sequence of functions. Let $\mathbf{1}$ denote the constant function 1 on S, and, for $n \in \mathbb{N}_+$, let $\mathbf{1}^n$ denote the convolution power of $\mathbf{1}$ of order n. Thus, $\mathbf{1}^1 = \mathbf{1}$, and, for $n \in \mathbb{N}_+$,

$$\mathbf{1}^{n+1}(x) = \int_{[e,x]} \mathbf{1}^n(t) \mathrm{d}\lambda(t), \qquad x \in S.$$

The following proposition gives some alternative interpretations.

Proposition 2. *For $n \in \mathbf{N}_+$, let λ^n denote the n-fold product measure on S^n corresponding to λ. For $n \in \{2, 3, \ldots\}$,*

$$\mathbf{1}^n(x) = \lambda^{n-1}\{(u_1, \ldots, u_{n-1}) \in S^{n-1} : u_1 \preceq \ldots \preceq u_{n-1} \preceq x\}, \tag{4}$$

$$= \lambda^{n-1}\{(v_1, \ldots, v_{n-1}) \in S^{n-1} : v_1 \ldots v_{n-1} \preceq x\}. \tag{5}$$

In particular, $\mathbf{1}^2(x) = \lambda[e, x]$. If $(S, \cdot)$ is discrete, then $\mathbf{1}^n(x)$ is the number of (ordered) n-factorings of x:

$$\mathbf{1}^n(x) = \#\{(v_1, v_2, \ldots, v_n) \in S^n : v_1 v_2 \ldots v_n = x\}. \tag{6}$$

Proof. By definition,

$$\mathbf{1}^2(x) = \int_{[e,x]} 1 \, \mathrm{d}\lambda(t) = \lambda[e, x],$$

so (4) and (5) hold for $n = 2$. Suppose that (4) holds for a given $n > 2$. Let

$$A_n(x) = \{(u_1, \ldots, u_n) \in S^n : u_1 \preceq \ldots \preceq u_n \preceq x\}, \qquad x \in S, n \in \mathbb{N}_+.$$

Then

$$\lambda^n(A_n(x)) = \int_S \lambda^{n-1}\{\mathbf{u} \in S^{n-1} : (\mathbf{u}, t) \in A_n(x)\}\mathrm{d}\lambda(t).$$

But $(\mathbf{u}, t) \in A_n(x)$ if and only if $t \in [e, x]$ and $\mathbf{u} \in A_{n-1}(t)$. Hence

$$\lambda^n(A_n(x)) = \int_{[e,x]} \lambda^{n-1}(A_{n-1}(t))\mathrm{d}\lambda(t) = \int_{[e,x]} \mathbf{1}^n(t)\mathrm{d}\lambda(t) = \mathbf{1}^{n+1}(x).$$

Similarly, suppose that (5) holds for a given $n > 2$. Let

$$B_n(x) = \{(v_1, \ldots, v_n) \in S^n : v_1 \ldots v_n \preceq x\}, \qquad x \in S,\ n \in \mathbb{N}_+.$$

Then

$$\lambda^n(B_n(x)) = \int_S \lambda^{n-1}\{\mathbf{v} \in S^{n-1} : (t, \mathbf{v}) \in B_n(x)\}\mathrm{d}\lambda(t).$$

But $(t, \mathbf{v}) \in B_n(x)$ if and only if $tv_1 \cdots v_{n-1} \preceq x$ if and only if $t \in [e, x]$ and $\mathbf{v} \in B_{n-1}(t^{-1}x)$. Hence

$$\lambda^n(B_n(x)) = \int_{[e,x]} \lambda^{n-1}\big(B_{n-1}(t^{-1}x)\big)\mathrm{d}\lambda(t) = \int_{[e,x]} \mathbf{1}^n(t^{-1}x)\mathrm{d}\lambda(t) = \mathbf{1}^{n+1}(x).$$

Finally, in the discrete case, (6) holds since there is one-to-one correspondence between $B_{n-1}(x)$ and the set

$$\{(v_1, \ldots, v_n) \in S^n : v_1 \ldots v_n = x\}.$$

□

Lemma 1. *Suppose that F is the tail probability function of a memoryless distribution on S. Then $(\mathbf{1}^n F) * F = \mathbf{1}^{n+1} F$.*

Proof. Let $x \in S$. From the memorylessness property,

$$[(\mathbf{1}^n F) * F](x) = \int_{[e,x]} \mathbf{1}^n(t)F(t)F(t^{-1}x)\mathrm{d}\lambda(t)$$

$$= F(x)\int_{[e,x]} \mathbf{1}^n(t)\mathrm{d}\lambda(t) = F(x)\mathbf{1}^{n+1}(x).$$

□

The gamma densities have a simple and elegant formulation in terms of the underlying exponential tail probability function F, the rate parameter α, and the convolution powers of $\mathbf{1}$. A preliminary version of the following theorem appeared in Rowell (1995).

Theorem 2. *Consider an exponential distribution on S with tail probability function F and rate parameter $\alpha > 0$. Then the probability density function f_n of the corresponding gamma distribution of order n is given by*

$$f_n = \alpha^n \mathbf{1}^n F. \tag{7}$$

Proof. The proof is by induction on n. First, f_1 is the probability density function of the given exponential distribution, and hence by the constant rate property, $f_1 = \alpha F$. Thus the result if true for $n = 1$. Assume that the result is true for a given n. Then

$$f_{n+1} = f_n * f = (\alpha^n \mathbf{1}^n F) * (\alpha F) = \alpha^{n+1}[(\mathbf{1}^n F) * F] = \alpha^{n+1}\mathbf{1}^{n+1} F$$

by Lemma 1. Therefore the result holds for $n + 1$. □

Theorem 3. *Suppose that F is the tail probability function of a probability distribution on S with constant rate α. Then f_n given in* (7) *is a probability density function for each $n \in \mathbb{N}_+$.*

Proof. The proof is by induction on n. First, $f_1 = \alpha F$ is a probability density function, by definition of the constant rate property. Suppose that f_n is a probability density function for some n. Then

$$\begin{aligned}
\int_S f_{n+1}(x)\mathrm{d}\lambda(x) &= \int_S \alpha^{n+1}\mathbf{1}^{n+1}(x)F(x)\mathrm{d}\lambda(x) \\
&= \int_S \alpha^{n+1}\int_{[e,x]} \mathbf{1}^n(y)\mathrm{d}\lambda(y)F(x)\mathrm{d}\lambda(x) \\
&= \int_S \int_{yS} \alpha^{n+1}\mathbf{1}^n(y)F(x)\mathrm{d}\lambda(x)\mathrm{d}\lambda(y) \\
&= \int_S \alpha^n \mathbf{1}^n(y)\left(\int_{yS} \alpha F(x)\mathrm{d}\lambda(x)\right)\mathrm{d}\lambda(y) \\
&= \int_S \alpha^n \mathbf{1}^n(y)F(y)\mathrm{d}\lambda(y) \\
&= \int_S f_n(y)\mathrm{d}\lambda(y) = 1.
\end{aligned}$$

□

Of course, if the distribution only has constant rate, we cannot conclude that the density f_n is the n-fold convolution power of f; equivalently, we cannot conclude that f_n is the density of the product of n i.i.d. variables.

Proposition 3. *Suppose that X has constant rate α and that $g : S \to \mathbb{R}$ is a measurable function with* $\mathrm{E}(|g(X)|) < \infty$. *For $n \in \mathbb{N}_+$, let $g_n = g * \mathbf{1}^n$. Then*

$$\mathrm{E}(g_n(X)) = \frac{1}{\alpha^n}\,\mathrm{E}(g(X)).$$

Proof. The following computations hold by the standard change of variables formula and Fubini's theorem:

$$\begin{aligned}
\mathrm{E}(g_{n+1}(X)) &= \int_S g_{n+1}(y)f(y)\mathrm{d}\lambda(y) = \int_S \int_{[e,y]} g_n(x)\mathrm{d}\lambda(x)f(y)\mathrm{d}\lambda(y) \\
&= \int_S \int_{xS} g_n(x)f(y)\mathrm{d}\lambda(y)\mathrm{d}\lambda(x) = \int_S g_n(x)F(x)\mathrm{d}\lambda(x) \\
&= \frac{1}{\alpha}\int_S g_n(x)f(x)\mathrm{d}\lambda(x) = \frac{1}{\alpha}\mathrm{E}(g_n(X)).
\end{aligned}$$

□

Corollary 2. *Suppose that X has constant rate α. Then*

$$\mathrm{E}(\mathbf{1}^n(X)) = \frac{1}{\alpha^{n-1}}, \qquad n \in \mathbb{N}_+.$$

Note that this result generalizes (2), which corresponds to $n = 2$. The following theorem gives a basic characterization of exponential distributions.

Theorem 4. *Suppose that X and Y are i.i.d. random variables taking values in S. Then the common distribution is exponential if and only if the conditional distribution of X given $XY = z$ is uniform on $[e, z]$ for every $z \succ e$.*

Proof. Suppose first that X and Y have a common exponential distribution with tail probability function F and constant rate $\alpha > 0$. Thus $f = \alpha F$ is a density function for X and for Y. Hence, (X, XY) has probability density function h given by

$$h(x, z) = f(x)f(x^{-1}z) = \alpha^2 F(x)F(x^{-1}z) = \alpha^2 F(z), \qquad x \preceq z,$$

using the constant rate and memorylessness properties. On the other hand, by Theorem 2, the probability density function of XY is

$$f_2(z) = \alpha^2 \mathbf{1}^2(z)F(z) = \alpha^2 \lambda[e, z]F(z), \qquad z \in S.$$

Therefore, the conditional density function of X given $XY = z \succ e$ is

$$x \mapsto \frac{h(x, z)}{f_2(z)} = \frac{1}{\lambda[e, z]}, \qquad x \preceq z,$$

which is the probability density function of the uniform distribution on $[e, z]$.

Conversely, suppose that X and Y are i.i.d. variables on S and that the conditional distribution of X given $XY = z$ is uniform on $[e, z]$ for each $z \succ e$. If $A \subseteq S$ is measurable, then

$$P(X \in A) = \mathrm{E}(P(X \in A|XY)) = \mathrm{E}\left(\frac{\lambda(A \cap [e, XY]}{\lambda[e, XY]}\right),$$

and therefore X is absolutely continuous with respect to λ. Let f denote a density of X, so that $f_2 = f * f$, a density function of XY. Then, with an appropriate choice of f,

$$\frac{f(x)f(x^{-1}z)}{f_2(z)} = \frac{1}{\lambda[e, z]}, \qquad x \preceq z,$$

for each $x \succ e$. Equivalently,

$$f(x)f(y) = \frac{f_2(xy)}{\lambda[e, xy]}, \qquad x \in S,\ y \in S,\ xy \succ e.$$

It now follows from the characterization in Proposition 1 that f is the density of an exponential distribution. □

The restriction $z \succ e$ in Theorem 4 is necessary to avoid division by 0 in the continuous case. The first part of the theorem has a simple extension to any number of i.i.d. exponential variables, and this shows that the gamma distributions, in a sense, govern the most random way to place points in S. Although the proofs are simple, it is still a little surprising that so many of the basic characterizations and properties of the exponential distribution are valid with the minimal algebraic assumptions of a positive semigroup.

Corollary 3. *Suppose that* $(X_1, X_2, \ldots)$ *is a sequence of independent random variables on S, each with a common exponential distribution. For* $n \in \mathbb{N}_+$*, let* $Y_n = X_1 \cdots X_n$*. For* $n \geqslant 2$*, the conditional distribution of* $(Y_1, Y_2, \ldots, Y_{n-1})$ *given* $Y_n = z$ *is uniform on*

$$A_{n-1}(z) = \{(u_1, \ldots, u_{n-1}) \in S^{n-1} : u_1 \preceq \cdots \preceq u_{n-1} \preceq z\}$$

That is, the conditional density is the constant $1/\mathbf{1}^n(z)$ *on* $A_{n-1}(z)$*.*

4. Brief examples

In this section we briefly consider several examples. In some cases, the exponential distribution were studied in Siegrist (1994), so our main goal is to identify the gamma distributions and the convolution powers of **1**.

4.1. The positive semigroup $([0, \infty), +)$

Of course, $([0, \infty), +)$ is a positive semigroup and is the motivation for this theory. The associated partial order is the ordinary order $\leqslant$, Lebesgue measure λ is invariant, and the exponential distributions are the ordinary exponential distributions. The convolution power of **1** of order $n \in \mathbb{N}_+$ is given by

$$\mathbf{1}^n(x) = \frac{x^{n-1}}{(n-1)!}, \qquad x \in [0, \infty).$$

The gamma distribution with rate $\alpha \in (0, \infty)$ and order $n \in \mathbb{N}_+$ is the ordinary gamma distribution, with density function

$$f_n(x) = \alpha^n \frac{x^{n-1}}{(n-1)!} e^{-\alpha x}, \qquad x \in [0, \infty).$$

4.2. The positive semigroup $(\mathbb{N}, +)$

The pair $(\mathbb{N}, +)$ is a positive semigroup. The associated partial order is the ordinary order $\leqslant$, and of course counting measure $\#$ is invariant. The exponential distributions for this semigroup are the ordinary geometric distributions. The convolution power of $\mathbf{1}$ of order $n \in \mathbf{N}_+$ is given by

$$\mathbf{1}^n(x) = \binom{n+x-1}{x}, \qquad x \in \mathbb{N},$$

which, in light of Proposition 2, we recognize as the number of ordered partitions of length n for the integer x. The gamma distribution with rate $1 - p \in (0, 1)$ and order $n \in \mathbb{N}_+$ is the ordinary negative binomial distribution, with density function

$$f_n(x) = (1-p)^n \binom{n+x-1}{x} p^x, \qquad x \in \mathbb{N}.$$

4.3. The free semigroup on a finite set

Let A be a finite set, thought of as an alphabet, and let $S = A^*$ be the set of all finite strings of letters from A, thought of as words. Then $(S, \cdot)$ is a positive semigroup, where $\cdot$ is the concatenation operation (the 'empty word' is the identity). For the associated partial order, $x \preceq y$ if and only if x is prefix of y. Let $N(x)$ denote the number of letters in x. If Y_n is a gamma variable of order n, then $N(Y_n)$ has a gamma distribution of order n for the positive semigroup $(\mathbb{N}, +)$ (that is, a negative binomial distribution), and given $N(Y_n) = k$, the k letters of Y_n are i.i.d. on the alphabet A. The convolution power of $\mathbf{1}$ of order n is given by

$$\mathbf{1}^n(x) = \binom{N(x)+n-1}{N(x)}, \qquad x \in S.$$

4.4. Finite subsets of $\mathbf{N}_+$

Let S denote the collection of all finite subsets of $\mathbb{N}_+$. Clearly the partially ordered set $(S, \subseteq)$ has the self-similar property noted in Section 1. In fact, the operation $\cdot$ defined by

$$xy = x \cup \{i\text{th smallest element of } x^c : i \in y\}$$

makes S into a discrete positive semigroup with $\varnothing$ as the identity and set inclusion as the associated partial order. Aside from the inherent mathematical interest of choosing a finite subset of $\mathbb{N}_+$ in the most random way, this example is interesting because only the minimal algebraic assumptions are satisfied. In particular, the operation is non-commutative and the right cancellation law does not hold, so $(S, \cdot)$ cannot be embedded in a group.

This positive semigroup is studied in detail in Siegrist (2007). Here we simply note that the convolution power of $\mathbf{1}$ of order n for the semigroup $(S, \cdot)$ is given by $\mathbf{1}^n(x) = n^{\#(x)}$. This result follows from a simple induction argument, and of course when $n = 2$, we obtain the usual formula for the cardinality of a power set.

4.5. Other self-similar structures

A positive semigroup S is self-similar in a special way: for each $x \in S$, $xS = \{y \in S : x \preceq y\}$ looks like the entire space S, both in an algebraic sense (the operator $\cdot$ and the partial order $\preceq$) and in a measure-theoretic sense (the left-invariant measure λ). Conversely, every partially ordered set $(S, \preceq)$ with the property that $\{y \in S : x \preceq y\}$ is order-isomorphic to S for each $x \in S$ essentially corresponds to a positive semigroup.

Of course, there are a variety of other self-similar structures in probability that do not fit our model, even though there may be semigroups lurking in the background. In fact, the term *self-similar* usually refers to random structures with distributions that are invariant under certain types of scaling. See Pitman (2006) for a variety of examples involving random partitions, fragmentation trees, and other combinatorial structures.

To give a brief illustration, consider the set S of partitions of $\mathbb{N}_+$. An element of S can be thought of as a function $x : \mathbb{N}_+ \times \mathbb{N}_+ \to \{0, 1\}$: i and j are in the same partition block if and only if $x(i, j) = 1$ (x must have the obvious reflexive, symmetric and transitive properties). The product (or minimum) operation makes S into a commutative semigroup with the idempotent property $x^2 = x$ for $x \in S$. (Thus, the left-cancellation law fails rather completely.) The relation $x \preceq y$ if and only if $xy = x$ is the natural partial order associated with this type of semigroup, and in fact corresponds in this example to partition refinement. The space S is the home of the Ewens partition process, which is exchangeable and self-similar (although the semigroup formulation may not be helpful).

5. The positive semigroup $(\mathbb{N}_+, \cdot)$

The pair $(\mathbb{N}_+, \cdot)$ is a positive semigroup, where $\cdot$ is ordinary multiplication. The corresponding partial order is the division partial order

$$x \preceq y \Leftrightarrow x \text{ divides } y.$$

The exponential distributions for this semigroup were noted briefly in Siegrist (1994). The purpose of this section is to study the semigroup in more depth and, in particular, to identify the gamma distributions and note connections with the recent work of Lin and Hu (2001) and Gut (2006).

For $i \in \mathbb{N}_+$, let σ_i denote the ith prime number (in the usual order); these are the irreducible elements of the semigroup. Each $x \in \mathbb{N}_+$ has the canonical prime factorization

$$x = \prod_{i=1}^{\infty} \sigma_i^{n_i},$$

where $n_i \in \mathbb{N}$ for each $i \in \mathbb{N}_+$ and $n_i = 0$ for all but finitely many i. Thus, $(\mathbb{N}_+, \cdot)$ is isomorphic to the positive semigroup $(M, +)$, where

$$M = \{(n_1, n_2, \ldots) : n_i \in \mathbb{N} \text{ for each } i \text{ and } n_i = 0 \text{ for all but finitely many } i\}$$

and where $+$ is pointwise addition.

Note that the convolution power $\mathbf{1}^2(x)$ is the number of divisors of $x \in \mathbb{N}_+$ (or equivalently, the number of ordered 2-factorings of x). This is an important function in number theory, and is usually denoted $\tau(x)$. In particular, τ is *multiplicative*:

$$\tau(xy) = \tau(x)\tau(y) \qquad \text{if } x, y \in \mathbb{N}_+ \text{ are relatively prime;}$$

hence $\mathbf{1}^k$ is multiplicative for each $k \in \mathbb{N}_+$. By Proposition 2, $\mathbf{1}^k(x)$ is the number of ordered k-factorings of x. In terms of the canonical factorization,

$$\mathbf{1}^k\left(\prod_{i=1}^{\infty} \sigma_i^{n_i}\right) = \prod_{i=1}^{\infty} \binom{k + n_i - 1}{n_i}, \qquad (n_1, n_2, \ldots) \in M.$$

5.1. Exponential distributions

Proposition 4. $F : \mathbb{N}_+ \to (0, \infty)$ *is the tail probability function of an exponential distribution on* $(\mathbb{N}_+, \cdot)$ *if and only if*

$$F\left(\prod_{i=1}^{\infty} \sigma_i^{n_i}\right) = \prod_{i=1}^{\infty} p_i^{n_i}, \qquad (n_1, n_2, \ldots) \in M, \tag{8}$$

where $p_i \in (0, 1)$ *for each* i *and* $\prod_{i=1}^{\infty} (1 - p_i) > 0$. *This infinite product is the rate constant of the distribution.*

Proof. The memorylessness property for a tail probability function F is

$$F(xy) = F(x)F(y), \qquad \text{for } x, y \in \mathbb{N}_+.$$

That is, F is *completely multiplicative*. It follows that F has the form given in (8), where $p_i = F(\sigma_i)$ for each $i \in \mathbb{N}_+$. Next note that

$$\sum_{x \in \mathbb{N}_+} F(x) = \sum_{(n_1, n_2, \ldots) \in M} F\left(\prod_{i=1}^{\infty} \sigma_i^{n_i}\right) = \sum_{(n_1, n_2, \ldots) \in M} \prod_{i=1}^{\infty} p_i^{n_i}$$

$$= \prod_{i=1}^{\infty} \sum_{n=0}^{\infty} p_i^n = \prod_{i=1}^{\infty} \frac{1}{1 - p_i}.$$

Thus, the result follows from Corollary 1. □

The density function of the exponential distribution in Proposition 4 is

$$f\left(\prod_{i=1}^{\infty} \sigma_i^{n_i}\right) = \prod_{i=1}^{\infty} p_i^{n_i}(1 - p_i), \qquad (n_1, n_2, \ldots) \in M.$$

If X is a random variable with the distribution then

$$X = \prod_{i=1}^{\infty} \sigma_i^{N_i}, \tag{9}$$

where $(N_i : i \in \mathbb{N}_+)$ are independent random variables and N_i has the geometric distribution with rate parameter $1 - p_i$. This characterization of the exponential distributions could also be obtained from the identification of $(\mathbb{N}_+, \cdot)$ with the positive semigroup $(M, +)$ given above, and the results in Section 4.2. The memorylessness property has the following interpretation: the conditional probability that y divides X/x given that x divides X is the same as the probability that y divides X. Thus, knowledge of one divisor of X does not help in finding other divisors of X. This property might have some practical applications.

Recall now that a (non-negative) *Dirichlet series* is a series of the form

$$A(s) = \sum_{x=1}^{\infty} \frac{a(x)}{x^s},$$

where $a : \mathbb{N}_+ \to [0, \infty)$; the series converges absolutely for s in an interval of the form (s_0, ∞). If the coefficient function a is completely multiplicative, then the function A also has a product expansion:

$$A(s) = \prod_{i=1}^{\infty} \frac{1}{1 - a_i \sigma_i^{-s}}, \qquad s > s_0,$$

where $a_i = a(\sigma_i)$ for $i \in \mathbb{N}_+$. There is a one-to-one correspondence between the coefficient function a and the series function A.

Given the coefficient function a, we can define a one-parameter family of probability distributions on $\mathbb{N}_+$, parameterized by $s > s_0$. This family is called the *Dirichlet* family of probability distributions corresponding to a. Specifically, X has the Dirichlet distribution corresponding to a with parameter s if

$$P(X = x) = \frac{a(x)x^{-s}}{A(s)}, \qquad x \in \mathbb{N}_+.$$

The most famous special case occurs when $a(x) = 1$ for all $x \in \mathbb{N}_+$ (note that a is completely multiplicative); then the Dirichlet series gives *the Riemann zeta function*,

$$\zeta(s) = \sum_{x=1}^{\infty} \frac{1}{x^s} = \prod_{i=1}^{\infty} \frac{1}{1 - \sigma_i^{-s}}, \qquad s > 1.$$

The corresponding one-parameter family of probability distributions on $\mathbb{N}_+$ is the *zeta family of distributions*,

$$P(X = x) = \frac{x^{-s}}{\zeta(s)}, \qquad x \in \mathbb{N}_+.$$

Theorem 5. *Suppose that a is positive and completely multiplicative. Then the Dirichlet distributions corresponding to a are exponential distributions on* $(\mathbb{N}_+, \cdot)$. *Conversely, every exponential distribution* $(\mathbb{N}_+, \cdot)$ *is a member of a Dirichlet family of distributions corresponding to a positive, completely multiplicative coefficient function.*

Proof. Suppose that a is positive and completely multiplicative and that X has the Dirichlet distribution corresponding to a with parameter s. It follows immediately from Proposition 1 that X has an exponential distribution. The tail probability function is

$$P(X \succeq x) = a(x)x^{-s}, \qquad x \in \mathbb{N}_+,$$

so the rate constant is $1/A(s)$. In a sense, the converse is trivially true. Suppose that X has an exponential distribution with tail probability function F. For *fixed* $t > 0$, let $a(x) = x^t F(x)$ for $x \in \mathbb{N}_+$, and let $A(s) = \sum_{x=1}^{\infty} a(x)x^{-s}$. Then a is completely multiplicative and A is the corresponding series function. Moreover, t is in the interval of convergence. The probability density function of X is

$$P(X = x) = \frac{a(x)x^{-t}}{A(t)}, \qquad x \in \mathbb{N}_+,$$

and so X has the Dirichlet distribution corresponding to a with parameter t. □

In particular, the zeta family of probability distributions are exponential distributions on the positive semigroup $(\mathbb{N}_+, \cdot)$. Suppose that X has a Dirichlet distribution with completely multiplicative coefficient function a and parameter s as above, Then of course X has the representation in (9). The geometric parameters for the random prime exponents of X are given by

$$p_i = P(X \succeq \sigma_i) = \frac{a_i}{\sigma_i^s}.$$

This representation was obtained by Lin and Hu (2001), who referred to the result as ‘striking’. However, from our point of view, the result is very simple and natural. From (2), the expected number of divisors of X is

$$A(s) = \mathrm{E}[\tau(X)].$$

This result was also in Lin and Hu (2001), but our point again is to show that the result is a special case of a much more general theorem. In fact, from Corollary 2,

$$A^k(s) = \mathrm{E}[\mathbf{1}^{k+1}(X)], \qquad k \in \mathbb{N};$$

that is, $A^k(s)$ is the expected number of $k+1$ factorings of X. We give an alternate proof of the following result, also from Lin and Hu (2001), to reinforce the point.

Proposition 5. *Suppose that* $a : \mathbb{N}_+ \to [0, \infty)$ *is non-negative and not identically zero. Let* A *be the corresponding Dirichlet function, which we assume converges for* $s > s_0$, *and let* $a_1(x) = \sum_{y \preceq x} a(y)$ *for* $x \in \mathbb{N}_+$. *Suppose that* X *has the zeta distribution with parameter* $s > \max\{s_0, 1\}$. *Then*

$$\mathrm{E}[a_1(X)] = A(s).$$

Proof. It follows immediately from Proposition 3 that

$$\mathrm{E}[a_1(X)] = \zeta(s)\mathrm{E}[a(X)],$$

since $1/\zeta(s)$ is the rate constant of the exponential distribution of X. But

$$\zeta(s)\mathrm{E}[a(X)] = \zeta(s)\sum_{x=1}^{\infty} \frac{a(x)}{x^s\zeta(s)} = \sum_{x=1}^{\infty} \frac{a(x)}{x^s} = A(s).$$

□

Exponential distributions on a positive semigroup usually maximize entropy subject to some natural moment conditions.

Proposition 6. *Suppose that* X *has the Dirichlet distribution with completely multiplicative coefficient function* a *and parameter* s. *Then* X *maximizes entropy over all random variables* Y *on* $\mathbb{N}_+$ *with*

$$\mathrm{E}[\ln(Y)] = \mathrm{E}[\ln(X)], \qquad \mathrm{E}[\ln(a(Y))] = \mathrm{E}[\ln(a(X))].$$

Proof. We use the usual inequality for entropy: if f and g are probability density functions of random variables X and Y, respectively, taking values in $\mathbb{N}_+$, then

$$H(Y) = -\sum_{x=1}^{\infty} g(x)\ln[g(x)] \leqslant -\sum_{x=1}^{\infty} g(x)\ln[f(x)]. \tag{10}$$

If X has the given Dirichlet distribution then an upper bound for the entropy of Y is

$$\begin{aligned} H(Y) &\leqslant -\ln\left[\frac{1}{A(s)}\right] - \mathrm{E}[\ln(a(Y)Y^{-s})] \\ &= \ln[A(s)] + s\mathrm{E}[\ln(Y)] - \mathrm{E}[\ln(a(Y))]. \end{aligned}$$

□

5.2. Gamma distributions

Now suppose that X_n, $n \in \mathbb{N}_+$, are i.i.d. variables, each with the exponential distribution with parameter vector $(p_1, p_2, \ldots)$, as in Proposition 4. Let $Y_n = X_1 \cdots X_n$ be the corresponding gamma variable of order $n \in \mathbb{N}_+$. It follows immediately that

$$Y_n = \prod_{i=1}^{\infty} \sigma_i^{U_{ni}},$$

where U_{ni} has the negative binomial distribution with rate parameter $1 - p_i$ and order n, and where $(U_{ni};\, i \in \mathbb{N}_+)$ are independent. Hence

$$P\left(Y_n = \prod_{i=1}^{\infty} \sigma_i^{k_i}\right) = \prod_{i=1}^{\infty} \binom{n + k_i - 1}{k_i} p_i^{k_i}(1 - p_i)^n, \qquad (k_1, k_2, \ldots) \in M.$$

We can reformulate the result in terms of Dirichlet distributions. Thus, suppose that X has the Dirichlet distribution corresponding to the positive, completely multiplicative coefficient function a and with parameter s. Then from Theorems 2 and 5,

$$P(Y_n = x) = \frac{\mathbf{1}^n(x)a(x)x^{-s}}{A^n(s)}, \qquad x \in \mathbb{N}_+.$$

Thus, Y_n also has a Dirichlet distribution, but corresponding to the multiplicative coefficient function $x \mapsto \mathbf{1}^n(x)a(x)$, which in general is not completely multiplicative.

5.3. Compound Poisson distributions

Not all results on Dirichlet distributions have simple interpretations in terms of exponential distributions on positive semigroups. Suppose that X has a Dirichlet distribution with completely multiplicative coefficient function a (and thus X has an exponential distribution on $(\mathbb{N}_+, \cdot)$). Gut (2006) showed that X has a *compound Poisson* distribution. In our notation,

$$X = V_1 V_2 \cdots V_N,$$

where $(V_1, V_2, \ldots)$ are i.i.d. on the set of prime powers $\{\sigma_i^n : i, n \in \mathbb{N}_+\}$, with common probability density function

$$P(V = \sigma_i^n) = \frac{p_i^n}{n \ln(\mathrm{E}(\tau(X)))}, \qquad i, n \in \mathbb{N}_+.$$

The random index N is independent of $(V_1, V_2, \ldots)$ and has the Poisson distribution with parameter $\ln(\mathrm{E}(\tau(X)))$. In particular, it follows that X is infinitely divisible. The infinite divisibility of X was also shown by Lin and Hu (2001). In our setting, we can also conclude that gamma variables are compound Poisson and infinitely divisible. The method of proof in Gut (2006) and in Lin and Hu (2001) is via characteristic functions, and thus the method does not have a clear generalizationto the setting of positive semigroups.

This result, however, leads to an interesting question: for a general positive semigroup

$(S, \cdot)$, under what conditions are exponential distributions (and hence also gamma distributions) always compound Poisson? It is well known that the ordinary exponential distribution (corresponding to the positive semigroup in Section 4.1) is compound Poisson, as is the ordinary geometric distribution (corresponding to the positive semigroup in Section 4.2). The compound Poisson property of the geometric distribution in turn leads to the same property for exponential distributions on the free semigroup in Section 4.3. These results, along with the result in Gut (2006), suggest a possibly general answer. Another curious connection (which may be superficial) is that the study of infinitely divisible distributions on $\mathbb{R}$ leads to semigroups of operators (see Feller 1971).

References

Azlarov, T.A. and Volodin, N.A. (1986) *Characterization Problems Associated with the Exponential Distribution*. Berlin: Springer-Verlag.

Clifford, A.H. and Preston, G.B. (1964) *The Algebraic Theory of Semigroups*, Vol. I, 2nd edn. Providence, RI: American Mathematical Society.

Clifford, A.H. and Preston, G.B. (1967) *The Algebraic Theory of Semigroups*, Vol. II. Providence, RI: American Mathematical Society.

Feller, W. (1971) *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd edn. New York: Wiley.

Gut, A. (2006) Some remarks on the Riemann zeta distribution. *Rev. Roumaine Math. Pures Appl.*, **51**, 205–217.

Halmos, P.R. (1974) *Measure Theory*. Berlin: Springer-Verlag.

Högnäs, G. and Mukherjea, A. (1995) *Probability Measures on Semigroups*. New York: Plenum Press.

Lin, G.D. and Hu, C.Y. (2001) The Riemann zeta distribution. *Bernoulli*, **7**, 817–828.

Pitman, J. (2006) *Combinatorial Stochastic Processes*. Berlin: Springer-Verlag.

Rowell, G.H. (1995) Probability distributions on temporal semigroups. PhD dissertation, University of Alabama in Huntsville.

Rowell, G.H. and Siegrist, K. (1998) Relative aging of distributions. *Probab. Engrg. Inform. Sci.*, **12**, 469–478.

Ruzsa, I.Z. and Székely, G.J. (1998) *Algebraic Probability Theory*. New York: Wiley.

Siegrist, K. (1994) Exponential distributions on semigroups. *J. Theoret. Probab.*, **7**, 725–737.

Siegrist, K. (2007) Random finite subsets with exponential distributions. *Probab. Engrg. Inform. Sci.*, **21**, 117–132.

Received September 2005 and revised September 2006