Probability Distributions on Partially Ordered Sets and Positive Semigroups

Kyle Siegrist
Department of Mathematical Sciences
University of Alabama in Huntsville
siegrist@math.uah.edu

February 10, 2010

Abstract

The algebraic structure $([0,\infty), +, \leq)$, together with Lebesgue measure (and the usual topology), is the standard representation of time for a variety of probabilistic models. These include reliability theory and the aging of lifetime distributions, renewal processes, and Markov processes, to name just a few. The purpose of this monograph is to study probability models in which the standard representation of time is generalized to a partially ordered set $(S, \preceq)$ and to a natural structure $(S, \cdot, \preceq)$ known as a *positive semigroup*. The partial order $\preceq$ replaces the ordinary order $\leq$, and in the case of a positive semigroup, the operation $\cdot$ replaces the standard translation operator $+$. A reference measure λ on S replaces Lebesgue measure, and in the case of a positive semigroup, we assume that this measure is left-invariant.

Contents

Part I
Basic Theory

1 Posets and Positive Semigroups

The structures that we will study involve algebra, topology, and measure.

1.1 Algebra

The basic structure of interest is a partially ordered set $(S, \preceq)$. Topologial and measure-theoretic assumptions will be imposed in the following sections. A particularly important special case is when the partial order $\preceq$ is associated with a positive semigroup.

Definition 1. A *semigroup* $(S, \cdot)$ is a set S with an associative operation $\cdot$. A *positive semigroup* is a semigroup $(S, \cdot)$ that has an identity element e, has no non-trivial invertible elements, and satisfies the left cancellation law.

A positive semigroup has a natural partial order associated with it.

Definition 2. Suppose that $(S, \cdot)$ is a positive semigroup. We define a relation $\preceq$ on S by

$$x \preceq y \text{ if and only if } xt = y \text{ for some } t \in S$$

In the language of a multiplicative semigroup, $x \preceq y$ if and only if x divides y.

Theorem 1. The relation $\preceq$ is a partial order on S with e as the minimum element. For each $x \in S$, the mapping $t \mapsto xt$ is an order-isomorphism from S onto $\{y \in S : y \succeq x\}$.

Proof. First, $x \preceq x$ since $xe = x$. Next, suppose that $x, y \in S$ with $x \preceq y$ and $y \preceq x$. Then there exist $a, b \in S$ such that $xa = y$ and $yb = x$. Hence $xab = yb = x = xe$. By the left cancellation law, $ab = e$ and then since there are no nontrivial invertible elements, $a = b = e$. Hence $x = y$. Finally, suppose that x, y, $z \in S$ and that $x \preceq y$ and $y \preceq z$. Then there exists $a, b \in S$ such that $xa = y$ and $yb = z$. But then $xab = yb = z$ so $x \preceq z$. Finally, $ex = x$ so $e \preceq x$ for any $x \in S$. Thus, $\preceq$ is a partial order on S with e as the minimum element.

Now let $x \in S$ and consider the map $t \mapsto xt$. The map is onto $\{y \in S : y \succeq x\}$ by the definition of the order. The map is one-to-one because of the left cancellation law. Suppose that x, y, $a \in S$. If$x \preceq y$ then there exits $t \in S$ such that $xt = y$. But then $axt = ay$ so $ax \preceq ay$. Conversely suppose that $ax \preceq ay$. Then there exists $t \in S$ such that $axt = ay$. But then $xt = y$ by the left cancellation law, so $x \preceq y$. $\square$

Note 1. Generally, a triple $(S, \cdot, \preceq)$ is a *left-ordered semigroup* if $(S, \cdot)$ is a semigroup, $\preceq$ is a partial order on S, and for every $x, y, a \in S$,

$$x \preceq y \text{ if and only if } ax \preceq ay$$

Thus, if $(S, \cdot)$ is a positive semigroup and $\preceq$ the natural partial order associated with $(S, \cdot)$, then $(S, \cdot, \preceq)$ is a left-ordered semigroup.

Note 2. For a positive semigroup $(S, \cdot)$, the fact that S is order-isomorphic to $\{y \in S : y \succeq x\}$ for each $x \in S$ is a *self-similarity property*. Conversely, suppose $(S, \preceq)$ is a partially ordered set with minimum element e and with this self-similarity property. For each $x \in S$, we fix an isomorphism from S to $\{y \in S : y \succeq x\}$. and denote the value of this mapping at y by xy. Note that $\cdot$ defines a binary operation on S. Since e is the minimum element of S, xe must be the minimum element of $\{y \in S\colon x \preceq y\}$ and therefore $xe = x$. Also, we will take the isomorphism from S onto S to be the identity map and hence $ex = x$ for any $x \in S$. It follows that e is the identity. Next, since the isomorphism $y \mapsto xy$ is one-to-one, the left cancellation law must hold. There are no nontrivial invertible elements since if $xy = e$ then $x \preceq e$ and hence $x = e$ and so also $y = e$. Finally, if the isomorphisms are chosen in such a way that the operation $\cdot$ is associative then $(S, \cdot)$ is a positive semigroup. Thus, the algebraic assumptions are precisely the ones needed so that the partially ordered set $(S, \preceq)$ has the self-similarity property.

Problem 1. Make this precise. That is, if $(S, \preceq)$ has a minimum element e and satisfies the self-similarity property, then construct a positive semigroup $(S, \cdot)$ with $\preceq$ as the associated partial order.

Proposition 1. Every positive semigroup $(S, \cdot)$ is isomorphic to a positive semigroup of one-to-one mappings from S into S (with composition as the semigroup operation).

Proof. Suppose that $(S, \cdot)$ is a positive semigroup. For each $a \in S$, let $\rho_a\colon S \to S$ be defined by

$$\rho_a(x) = ax, \quad x \in S$$

Let $T = \{\rho_a\colon a \in S\}$ and consider the composition operator $\circ$ on T. First, T is closed under $\circ$ since for any a, $b \in S$,

$$\rho_a \circ \rho_b(x) = a(bx) = (ab)x = \rho_{ab}(x), \quad x \in S$$

and therefore $\rho_a \circ \rho_b = \rho_{ab}$. The mappings in T are one-to-one:

$$\rho_a(x) = \rho_a(y) \Rightarrow ax = ay \Rightarrow x = y$$

The identity map is in T since $\rho_e(x) = ex = x$ for $x \in S$. Finally, suppose that ρ_a and ρ_b are (functional) inverses. Then $\rho_a(\rho_b(x)) = x$ for any $x \in S$ or equivalently $abx = x$ for any $x \in S$. Letting $x = e$ we have $ab = e$ and therefore $a = b = e$. ☐

Proposition 2. Suppose that $(S, \cdot)$ is a semigroup satisfying the left cancellation law and the property that $xy \neq x$ for every x, $y \in S$. Then $(S, \cdot)$ can be made into a positive semigroup with the addition of an identity element.

Proof. Note that the condition $xy \neq x$ for all x, $y \in S$ implies that S does not have an identity. Thus let $T = S \cup \{e\}$ where e is a element not in S. Extend $\cdot$ to T by $xe = ex = x$ for all $x \in T$. We will show that $(T, \cdot)$ is a positive semigroup. First, the associative property $(xy)z = x(yz)$ still holds, since it holds in T and trivially holds if one of the elements is e. Next, e is the identity, by construction. There are no nontrivial inverses in T: if x, $y \in S$ then $xy \in S$ so $xy \neq e$. Hence if $xy = e$ then at least one element, and hence both, must be e. Finally, to show that the left cancellation law holds, suppose that $xy = xz$.

If x, y, $z \in S$ then $y = z$ by the left cancellation law in S. If $x = e$ then trivially $y = z$. If $x \neq e$ and $y = e$ we have $xz = x$ and hence $z = e$. Similarly, if $x \neq e$ and $z = e$ we have $xy = x$ so $y = e$. $\square$

Note 3. The algebraic assumptions of a positive semigroup do not rule out the possibility that $xy = y$ for some x, $y \in S$ with $x \neq e$. However, if this is the case, then $x^n y = y$ for all $n \in \mathbb{N}$. But then $x^n \preceq y$ for all $n \in \mathbb{N}$, so if S is countable then $(S, \cdot)$ is not locally finite (see Note 11). Ordinarily we do not consider such semigroups.

Definition 3. Suppose that $(S, \preceq)$ is a partially ordered set. We will write $x \perp y$ if x and y are *comparable*, that is, either $x \preceq y$ or $y \preceq x$. We will write $x \parallel y$ if x and y are non-comparable; that is, neither $x \preceq y$ nor $y \preceq x$.

Definition 4. A partially ordered set $(S, \preceq)$ is (algebraically) *connected* if for every x, $y \in S$, there exists a finite sequence $(x_0, x_1, \ldots, x_n)$ such that

$$x = x_0 \perp x_1 \cdots \perp x_n = y$$

The proposition below gives another type of semigroup with an associated partial order.

Proposition 3. Suppose that $(S, \cdot)$ is a commutative semigroup with the idempotent property that $x^2 = x$ for all x. Define $x \preceq y$ if an only if $xy = x$. Then $\preceq$ is a partial order on S.

Proof. First $x \preceq x$ for any $x \in S$ since $x^2 = x$ by assumption. Next suppose that $x \preceq y$ and $y \preceq x$. Then $xy = x$ and $yx = y$. But $xy = yx$ so $x = y$. Finally, suppose that $x \preceq y$ and $y \preceq z$. Then $xy = x$ and $yz = y$. Hence $xz = (xy)z = x(yz) = xy = x$ so $x \preceq z$. $\square$

We next give a number of standard definitions for subsets of a partially ordered set $(S, \preceq)$.

Definition 5. $A \subseteq S$ is *increasing*, or respectively *decreasing*, if

$$x \in A,\ x \preceq y \Rightarrow y \in A$$
$$y \in A,\ x \preceq y \Rightarrow x \in A$$

In the case of a positive semigroup $(S, \cdot)$, A is increasing if $x \in A$ implies $xu \in A$ for any $u \in S$ and A is decreasing if $xu \in A$ implies $x \in A$.

Note 4. If $A_i \subseteq S$ is increasing (respectively deceasing) for each i in a nonempty index set I, then $\bigcap_{i \in I} A_i$ is also increasing (decreasing). If $A \subseteq S$ then the increasing (decreasing) set *generated by* A is the intersection of all increasing (decreasing) subsets of S that contain A.

Definition 6. For $A \subseteq S$, define

$$\begin{aligned} I(A) &= \{x \in S \colon x \succ a \text{ for some } a \in A\} \\ D(A) &= \{x \in S \colon x \prec a \text{ for some } a \in A\} \\ I[A] &= \{x \in S \colon x \succeq a \text{ for some } a \in A\} \\ D[A] &= \{x \in S \colon x \preceq a \text{ for some } a \in A\} \end{aligned}$$

Note that $I[A]$ and $D[A]$ are the increasing and decreasing sets generated by A, respectivley. For $a \in S$, we simplify the notation to $I(a)$, $D(a)$, $I[a]$, and $D[a]$.

Definition 7. $A \subseteq S$ is *convex* if

$$a \in A,\ b \in A,\ a \preceq x \preceq b \Rightarrow x \in A$$

Note 5. If $A_i \subseteq S$ is convex for each i in a nonempty index set I, then $\bigcap_{i \in I} A_i$ is also convex. If $A \subseteq S$ then the convex set *generated by* A is the intersection of all convex subsets of S that contain A. This set is

$$C[A] = I[A] \cap D[A] = \{x \in S \colon a \preceq x \preceq b \text{ for some } a \in A,\ b \in A\}$$

If A is increasing or if A is decreasing, then A is convex.

Definition 8. If $a,\ b \in S$ and $a \preceq b$, we use interval notation in the usual way:

$$\begin{aligned} [a,\ b] &= \{x \in S \colon a \preceq x \preceq b\} \\ (a,\ b] &= \{x \in S \colon a \prec x \preceq b\} \\ [a,\ b) &= \{x \in S \colon a \preceq x \prec b\} \\ (a,\ b) &= \{x \in S \colon a \prec x \prec b\} \end{aligned}$$

Note 6. In particular, $[a,\ b] = I[a] \cap D[b]$. Each of the intervals above is convex. Generally, if A is convex if and only if $a,\ b \in A$ and $a \preceq b$ imply $[a, b] \subseteq A$. If $(S, \preceq)$ has a minimum element e then, $D[a] = [e,\ a]$.

Definition 9. Suppose that $(S, \cdot)$ is a positive semigroup. If $A, B \subseteq S$ define

$$\begin{aligned} AB &= \{x \in S \colon x = ab \text{ for some } a \in A,\ b \in B\} \\ A^{-1}B &= \{x \in S \colon ax \in B \text{ for some } a \in A\} \end{aligned}$$

If $A = \{a\}$ for some $a \in S$, we simplify the notation to aB and $a^{-1}B$, respectively. Similarly, if $B = \{b\}$ for some $b \in S$, we simplify the notation to Ab and $A^{-1}b$ respectively.

Note 7. If $(S, \cdot)$ is a positive semigroup, note that $I[A] = AS$ for any $A \subseteq S$ and in particular $I[a] = aS$ for any $a \in A$. The set $A^{-1}B$ is empty unless there exists $a \in A$ and $b \in B$ with $a \preceq b$. In particular, $x^{-1}\{y\}$ is empty unless $x \preceq y$ in which case the set contains the unique element $u \in S$ such that $xu = y$. We will denote this element by $x^{-1}y$. Generally, $x^{-1}A = \{x^{-1}a \colon a \in A \text{ and } x \preceq a\}$ and

$$AB = \bigcup_{a \in A} aB,\ A^{-1}B = \bigcup_{a \in A} a^{-1}B$$

Definition 10. Suppose that $(S, \preceq)$ is a partially ordered set and $A \subseteq S$.

1. $x \in A$ is a *minimal* element of A if no $y \in A$ satisfies $y \prec x$.
2. $x \in A$ is a *maximal* element of A if no $y \in A$ satisfies $y \succ x$.
3. $x \in A$ is the *minimum* element of A if $x \preceq y$ for all $y \in A$.
4. $x \in A$ is the *maximum* element of A if $x \succeq y$ for all $y \in A$.
5. $x \in S$ is a *lower bound* for A if $x \preceq y$ for all $y \in A$.
6. $x \in S$ is an *upper bound* for A if $y \preceq x$ for all $y \in A$.
7. $x \in S$ is the *greatest lower bound* or *infimum* of A if x is a lower bound of A and $x \succeq y$ for any lower bound y of A.
8. $x \in S$ is the *least upper bound* or *supremum* of A if x is an upper bound of A and $x \preceq y$ for any upper bound y of A.

Note 8. As the language suggests, the minimum and maximum elements of A are unique, if they exists, and are denoted $\min(A)$ and $\max(A)$, respectively. If $\min(A)$ exists, it is also a minimal element of A, and similarly, if $\max(X)$ exists, it is also a maximal element of A. Similarly, the infimum and supremum of A are unique, if they exist, and are denoted $\inf(A)$ and $\sup(A)$, respectively. In fact, $\inf(A)$ is the maximum of the set of lower bounds of A and $\sup(A)$ is the minimum of the set of upper bounds of A. If x is a lower bound of A then A is said to be *bounded below* by x, and if x is an upper bound of A then A is said to be *bounded above* by x. The set A is *bounded* (in the algebraic sense!) if A has both a lower bound x and an upper bound y; in this case $A \subseteq [x, y]$.

Definition 11. The partially ordered set $(S, \preceq)$ is a *lower semi-lattice* if $x \wedge y := \inf\{x, y\}$ exists for every x, $y \in S$. Similarly, $(S, \preceq)$ is an *upper semi lattice* if $x \vee y := \sup\{x, y\}$ exists for every x, $y \in S$. Finally, $(S, \preceq)$ is a *lattice* if it is both a lower semi-lattice and an upper semi-lattice.

Note 9. In the context of Definition 11, suppose that $A \subseteq S$ is finite and nonempty. If $(S, \preceq)$ is a lower semi-lattice then $\inf(A)$ exists. If $(S, \preceq)$ is an upper semi-lattice then $\sup(A)$ exists.

Definition 12. Suppose that $(S, \preceq)$ is a partially ordered set and that x, $y \in S$. Then y is said to *cover* x if y is a minimal element of $\{u \in S : x \prec u\}$. If $(S, \preceq)$ has a minimum element e, then an element $i \in S$ that covers e is *irreducible* . If S is countable, the *Hasse graph* or *covering graph* of $(S, \preceq)$ has vertex set S and (directed) edge set $\{(x, y) \in S^2 : y \text{ covers } x\}$. These definitions are of most value when S is countable.

Proposition 4. Suppose that $(S, \cdot)$ is a positive semigroup. An element $i \in S$ is irreducible if and only if i cannot be factored $i = xy$ except for the trivial factoring $i = ei = ie$.

Proof. If $y \in S$ is not irreducible, then there exists $x \in S$ with $e \prec x \prec y$. But then there exists $t \in S - \{e\}$ with $y = xt$ and hence y has a non-trivial factoring. Conversely, suppose that $y = xt$ for some x, $t \in S - \{e\}$, so that y has a non-trivial factoring. Then $e \prec x \prec y$ so y is not irreducible. $\square$

Proposition 5. Suppose that $(S, \cdot)$ is a positive semigroup. If x, $y \in S$ then y covers x if and only if $y = xi$ for some irreducible element i.

Proof. Suppose that y covers x. Then $x \prec y$ so there exists $i \in S - \{e\}$ with $y = xi$. If $i = ab$ for some a, $b \in S - \{e\}$ then $x \prec xa \prec xab = y$ which would be a contradiction. Thus, i has no non-trivial factorings and hence is irreducible. Conversely, suppose that $y = xi$ for some irreducible element i. Then $x \prec y$. Suppose there exists $u \in S$ with $x \prec u \prec y$. Then $u = xs$ and $y = ut$ for some s, $t \in S - \{e\}$. Thus $y = xst = xi$ so by left-cancellation, $i = st$. But this is a contradiction since i is irreducible, so y covers x. $\square$

Suppose that $(S, \preceq)$ is a partially ordered set. For $x \in S$ define

$$A_x = \{y \in S : y \text{ covers } x\}$$
$$B_x = \{w \in S : x \text{ covers } w\}$$

Thus, A_x is the set of elements immediately *after* x in the partial order while B_x is the set of element immediately *before* x in the partial order. From Proposition 5, if $(S, \cdot)$ is a positive semigroup then $A_x = xI$ where I is the set of irreducible elements. Hence A_x has the same cardinality as I for each $x \in S$.

Definition 13. Suppose that $(S, \preceq)$ is a partially ordered set and recall that for $s \in S$,

$$I(s) = \{t \in S : t \succ s\}$$

For x, $y \in S$, we say that x and y are *upper equivalent* if $I(x) = I(y)$.

Upper equivalence clearly is an equivalence relation on S; it's the equivalence relation associated with the function $s \mapsto I(s)$ from S to $\mathcal{P}(S)$.

Theorem 2. Let $\equiv$ denote the upper equivalence relation and let $\Pi = (S/\equiv)$ denote the set of equivalence classes. Define $\preceq$ on Π by $A \prec B$ if and only if there exists $y \in B$ such that $x \prec y$ for all $x \in A$. Then $\preceq$ is a partial order on Π.

Proof. We need to show that $\prec$ is irreflexive and transitive.

Suppose that $A \in \Pi$ and $A \prec A$. Then there exists $y \in A$ such that $x \prec y$ for all $x \in A$. In particular, $y \prec y$ which is a contradiction.

Next suppose that $A, B, C \in \Pi$ with $A \prec B$ and $B \prec C$. Then there exists $y \in B$ such that $x \prec y$ for all $x \in A$, and there exists $z \in C$ such that $w \prec z$ for all $w \in B$. In particular, $y \prec z$ and therefore $x \prec z$ for all $x \in A$. Hence $A \prec C$. $\square$

1.2 Topology

Suppose that $(S, \preceq)$ is a partially ordered set. Ordinarily, we need a topology on S that is compatible with the algebraic structure. Recall that, technically, $\preceq$ is the set of ordered pairs

$$\{(x, y) \in S^2 \colon x \preceq y\}$$

(sometimes also called the *graph* of $(S, \preceq)$.

Definition 14. $(S, \preceq)$ is a *topological poset* if S has a topology with the following properties:

1. $\preceq$ is closed (as a subspace of S^2 with the product topology).
2. S is locally compact, locally convex, and has a countable base.
3. $D[x]$ is compact for each $x \in S$.

In the case that $(S, \cdot)$ is a positive semigroup, we impose the additional condition that the one-to-one mapping $(x, t) \mapsto (x, xt)$ from S^2 onto R is a homeomorphism. In any case, the Borel σ-algebra will be denoted by $\mathcal{B}(S)$.

Note 10. The first assumptions means that the partially ordered set $(S, \preceq)$ is topologically closed in the sense of Nachbin [22]. In particular, this assumption implies that $I[x]$ and $D[x]$ are closed for each x, and that S is Hausdorff (so that any two points can be separated by open neighborhoods). The second assumption means that there are arbitrarily small compact neighborhoods and arbitrarily small convex neighborhoods of each point, and that S is metrizable. Local convexity also implies that the "squeeze theorem" for limits holds: if $(x_n \colon n \in \mathbb{N}_+)$, $(y_n \colon n \in \mathbb{N}_+)$, and $(t_n \colon n \in \mathbb{N}_+)$ are sequences in S with $x_n \to a$ as $n \to \infty$, $y_n \to a$ as $n \to \infty$, and $x_n \preceq t_n \preceq y_n$ eventually in n, then $t_n \to a$ as $n \to \infty$. The third assumption implies that $[a, b]$ is compact for any a, $b \in S$ with $a \preceq b$. In the case of a positive semigroup, the last assumption means that the algebraic operation$(x, t) \mapsto xt$ from S^2 onto R and its inverse, $(x, y) \mapsto x^{-1}y$ from R onto S^2 are continuous.

Note 11. Clearly $\{x\} \in \mathcal{B}(S)$ for each $x \in S$, since singletons are closed in any Hausdorff space. Hence, if $(S, \preceq)$ a topological poset and S is countable, then $\mathcal{B}(S)$ is the power set of S; all subsets of S are countable unions of singletons and hence are measurable. Conversely, suppose that $(S, \preceq)$ is a poset and that S is countable. If $D[x]$ is finite for each $x \in S$ (so that $(S, \preceq)$ is *locally finite*), then with the discrete topology, $(S, \preceq)$ is a topological positive semigroup.

Lemma 1. Suppose that S and T are LCCB spaces. If $f \colon S \to T$ is one-to-one and continuous, then $f(A) \in \mathcal{B}(T)$ for any $A \in \mathcal{B}(S)$.

Proof. Since f is one-to-one, the following properties hold for any subsets $A \subseteq S$, $B \subseteq S$, and $A_i \subseteq S$, $i \in I$ where I is any nonempty index set:

$$f\left(\bigcup_{i \in I} A_i\right) = \bigcup_{i \in I} f(A_i)$$
$$f\left(\bigcap_{i \in I} A_i\right) = \bigcap_{i \in I} f(A_i)$$
$$f(B - A) = f(B) - f(A)$$

Now let $\mathcal{F} = \{A \in \mathcal{B}(S) \colon f(A) \in \mathcal{B}(T)\}$. By the above identities, if $A_i \in \mathcal{F}$ for $i \in \mathbb{Z}_+$ then $\cup_i A_i \in \mathcal{F}$ and $\cap_i A_i \in \mathcal{F}$ and if A, $B \in \mathcal{F}$ then $B - A \in \mathcal{F}$. Moreover, since f is continuous, $\mathcal{F}$ contains the compact sets. Since S is σ-compact, $S \in \mathcal{F}$. Finally, $\mathcal{B}(S)$ is generated by the compact sets and hence $\mathcal{F} = \mathcal{B}(S)$. ☐

Note 12. If $(S, \cdot)$ is a topological positive semigroup, then the Borel sets are closed under left translations, i.e. if $A \in \mathcal{B}(S)$ and $x \in S$ then $xA \in \mathcal{B}(S)$ and $x^{-1}A \in \mathcal{B}(S)$. These statements hold since the mapping $y \mapsto xy$ from S into S is continuous and one-to-one.

Definition 15. Suppose that $(S, \preceq)$ is a toplogical poset. The *poset dimension* of S is the smallest number of chains (total orders) on S whose intersection gives $\preceq$. We denote the poset dimension by $\dim(S, \preceq)$.

Definition 16. Suppose that $(S, \cdot)$ is a positive semigroup. Then S has *semigroup dimension* $n \in \mathbb{N}_+$ if there exists $B = \{x_1, x_2, \ldots, x_n\} \subseteq S$ such that if φ is a continuous homomorphism from $(S, \cdot)$ into the group $(\mathbb{R}, +)$ with $\varphi(x_i) = 0$ for each i then $\varphi(x) = 0$ for each $x \in S$; moreover, no set with $n-1$ elements has this property. The set B will be called a *critical set.* S has semigroup dimension 0 if there is no nontrivial continuous homomorphism from S into $(\mathbb{R}, +)$. S has semigroup dimension ∞ if for every finite subset of S there exists a nontrivial continuous homomorphism from S into $(\mathbb{R}, +)$ which maps the finite subset onto 0. We denote the semigroup dimension by $\dim(S, \cdot)$. These definitions are due to Székely [31],

Problem 2. If $(S, \cdot)$ is a positive semigroup, then we have two definitions of dimension, one corresponding to the semigroup structure and one corresponding to the associated partial order. How are these definitions related? They are certainly not the same. The semigroup corresponding to the subset partial order on finite subsets of $\mathbb{N}_+$ in Chapter 15 has semigroup dimension 1 and poset dimension greater than 1.

Proposition 6. Suppose that $(S, \cdot)$ is a standard, discrete positive semigroup with I as the set of irreducible elements. If $x \in S$, then x can be factored finitely over I. That is, $x = i_1 i_2 \cdots i_n$ where $i_k \in I$ for each k.

Proof. Let $x \in S$. Since $(S, \preceq)$ is locally finite, there exists a finite path in the Hasse graph from e to x, say $(x_0, x_1, x_2, \ldots, x_n)$ where $x_0 = e$, $x_n = x$, and x_{k+1} covers x_k for each k. But then $x_{k+1} = x_k i_k$ for each k where $i_k \in I$. Hence $x = i_1 i_2 \cdots i_n$. $\square$

Note 13. Of course, the factoring of x over I is not necessarily unique, and different factorings of x over I may have different lengths.

Proposition 7. Suppose that $(S, \cdot)$ is a standard, discrete positive semigroup, with I as the set of irreducible elements. Then $\dim(S, \cdot) \le \#(I)$.

Proof. Suppose that φ is a homomorphism from $(S, \cdot)$ into $(\mathbb{R}, +)$ and that $varphi(i) = 0$ for each $i \in I$. If $x \in S$, then from Proposition 6, x can be factored over I so that $x = i_1 i_2 \cdots i_n$ where $i_k \in I$ for each k. But then

$$\varphi(x) = \varphi(i_1) + \varphi(i_2) + \cdots + \varphi(i_n) = 0$$

Hence I is a critical set and so $\dim(S, \cdot) \le \#(I)$ $\square$

Note 14. We can certainly have $\dim(S, \cdot) < \#(I)$. The semigroup corresponding to the subset partial order on finite subsets of $\mathbb{N}_+$ in Chapter 15 has infinitely many irreducible elements but semigroup dimension 1.

1.3 Measure

Our last basic ingredient is a measure that is compatible with the algebraic and topological structures. Thus, suppose that $(S, \preceq)$ is a topological poset. The term *measure* on S will refer to a positive Borel measure on $\mathcal{B}(S)$; that is, a positive measure λ on $\mathcal{B}(S)$ such that $\lambda(K) < \infty$ for compact $K \subseteq S$. Because of the topological assumptions, λ is σ-finite, so there exists a sequence $A_n \in \mathcal{B}(S)$ with $S = \bigcup_{n=1}^{\infty} A_n$ and $\lambda(A_n) < \infty$ for each n. Also, λ is regular; that is,

$$\lambda(A) = \sup\{\lambda(K)\colon K \text{ compact}, K \subseteq A\} = \inf\{\lambda(U)\colon U \text{ open}, A \subseteq U\}$$

We will also usually assume that a measure λ has support S; equivalently $\lambda(U) > 0$ for any open set U

Definition 17. A measure λ on a topological positive semigroup $(S, \cdot)$ is said to *left-invariant* (or a *left Haar measure*) if

$$\lambda(xA) = \lambda(A), \quad x \in S,\ A \in \mathcal{B}(S)$$

Recall that $t \mapsto xt$ is an order-isomorphism from S onto $xS = \{y \in S : y \succeq x\}$; thus xA is the image of A under this isomorphism. Hence, a left-invariant measure is compatible with the semigroup operation $\cdot$ and the self-similarity of the partially ordered set $(S, \preceq)$.

Most of the positive semigroups that we will study have a left-invariant measure that is unique up to multiplication by positive constants.

Definition 18. If $(S, \preceq)$ is a topological, partially ordered set and λ is a fixed reference measure, then $(S, \preceq, \lambda)$ will be referred to as a *standard* poset. If $(S, \cdot)$ is a topological positive semigroup and λ is a fixed left-invariant measure for S, then $(S, \cdot, \lambda)$ will be referred to as a *standard positive semigroup.*

Discrete posets and discrete positive semigroups are always standard. If $(S, \preceq)$ is a discrete poset, then by the topological assumptions, $D[x]$ is finite for each $x \in S$. We wil always use counting measure $\#$ as the reference measure.

Proposition 8. If $(S, \cdot)$ is a discrete positive semigroup then counting measure $\#$ is left-invariant and is unique up to multiplication by positive constants.

Proof. Recall that $\mathcal{B}(S)$ is the power set of S, since $\{x\}$ is closed (and hence a Borel set) for each $x \in S$. Next note that

$$\#(xA) = \#(A), \quad x \in S,\, A \subseteq S$$

since $u \mapsto xu$ maps A one-to-one onto xA. If μ is another left-invariant measure on S then

$$\mu(\{x\}) = \mu(x\{e\}) = \mu(\{e\})$$

Hence $\mu(A) = \mu(\{e\})\#(A)$. $\square$

Proposition 9. Suppose that $(S, \cdot)$ is a positive semigroup with left-invariant measure λ. If $\lambda\{x\} > 0$ for some $x \in S$ then $(S, \cdot)$ is discrete (and then λ is a multiple of counting measure).

Proof. Suppose that $\lambda\{x\} > 0$ for some $x \in S$. Then

$$\lambda\{x\} = \lambda(x\{e\}) = \lambda\{e\}$$

so $\lambda\{e\} > 0$ and as before,

$$\lambda\{y\} = \lambda(y\{e\}) = \lambda\{e\}, \quad y \in S$$

Hence λ is a multiple of counting measure. Since λ is regular, we must have $\lambda(K) < \infty$ for each compact K and hence compact sets must be finite. But S is locally compact and has a countable base, and thus is σ-finite. It follows that S is countable. $\square$

Much more generally, Szèkely [31] has given sufficient conditions for the existence of a left-invariant measure on a semigroup. These conditions are rather technical, but there may be more natural conditions for a positive semigroup.

Problem 3. Find sufficient conditions for the existence of a left-invariant measure on a positive semigroup, unique up to multiplication by positive constants.

Proposition 10. If $(S, \cdot)$ is a non-trivial positive semigroup with left-invariant measure λ, then $\lambda(S) = \infty$.

Proof. Suppose that $\lambda(S) < \infty$. For every $x \in S$,

$$\lambda(S - xS) = \lambda(S) - \lambda(xS) = \lambda(S) - \lambda(S) = 0$$

But $[e, x) \subseteq S - xS$ and hence $\lambda[e, x) = 0$ for every $x \in S$. Since S is non-trivial, there exists $a \in S$, $a \neq e$. But $[e, a^2)$ is a neighborhood of a and hence $\lambda[e, a^2) > 0$—a contradiction. $\square$

Proposition 11. A measure λ on the positive semigroup $(S, \cdot)$ is left-invariant if and only if

$$\int_{xS} \varphi(x^{-1}y) d\lambda(y) = \int_S \varphi(z) d\lambda(z)$$

for every $x \in S$ and every bounded measurable function $\varphi : S \to \mathbb{R}$.

Proof. Let $\varphi = \mathbf{1}_A$ where $A \in \mathcal{B}(S)$. Then for $y \in xS$, $\varphi(x^{-1}y) = 1$ if and only $x^{-1}y \in A$ if and only if $y \in xA$. Therefore

$$\int_{xS} \varphi(x^{-1}y) d\lambda(y) = \lambda(xA)$$

$$\int_S \varphi(z) d\lambda(z) = \lambda(A)$$

The general result now follows in the usual way. $\square$

Corollary 1. A measure λ on $\mathcal{B}(S)$ is left-invariant for the postive semigroup $(S, \cdot)$ if and only if

$$\int_{xA} \psi(y) d\lambda(y) = \int_A \psi(xz) d\lambda(z)$$

for every $x \in S$, $A \in \mathcal{B}(S)$, and every bounded measurable function $\psi : S \to \mathbb{R}$.

Note 15. Besides left invariance, there are other, related invariance properties which have been studied (Mukherjea and Tserpes [21]). A measure λ is *l^*-invariant* if

$$\lambda(x^{-1}A) = \lambda(A)$$

for any $x \in S$ and $A \in \mathcal{B}(S)$. A measure λ is *left contra-invariant* if

$$\lambda(x^{-1}A) \geq \lambda(A)$$

for any $x \in S$ and $A \in \mathcal{B}(S)$. It is not reasonable to expect such properties to hold for positive semigroups. In most cases, there exists an open set A and $x \in S$ such that no $y \in A$ satisfies $x \preceq y$. Thus $x^{-1}A$ is empty. Mukherjea and Tserpes [21] have a number of results which state that if a semigroup has an l^*-invariant measure then the support of the measure is a left group. Such results do not apply here.

2 Operators and Cumulative Functions

In this chapter, we assume that $(S, \preceq, \lambda)$ is a standard poset.

2.1 Lower and upper operators

Let $\mathcal{D}(S)$ denote the set of measureable functions from S into $\mathbb{R}$, that are bounded on $D[x]$ for each $x \in S$. Recall that $D[x]$ is compact for each $x \in S$.

Definition 19. Define the *lower operator* L on $\mathcal{D}(S)$ by

$$L(g)(x) = \int_{D[x]} g(t) d\lambda(t), \quad x \in S$$

For $n \in \mathbb{N}$, let L^n denote the n-fold composition of the operator L. Thus $L^0(g) = g$ and for $n \in \mathbb{N}$, $L^{n+1}(g) = L(L^n(g))$. Thus,

$$L^{n+1}(g)(x) = \int_{D[x]} L^n(g)(t) d\lambda(t)$$

The operator L is linear and is well defined on $\mathcal{D}(S)$. To see this, let $g \in \mathcal{D}(S)$. and $y \in S$. There exists C_y such that $|g(t)| \leq C$ for $t \in D[y]$. Moreover, $D[x] \subseteq D[y]$ for each $x \in D[y]$. Hence for $x \in D[y]$

$$\begin{aligned} |L(g)(x)| &= \left| \int_{D[x]} g(t) d\lambda(t) \right| \leq \int_{D[x]} |g(t)| d\lambda(t) \\ &\leq \int_{D[y]} |g(t)| d\lambda(t) \leq \int_{D[y]} C d\lambda(t) = C_y \lambda(D[y]) \end{aligned}$$

Proposition 12. If $g \in \mathcal{D}(S)$ is nonnegative, then $L(g)$ is increasing (and hence $L^n(g)$ is increasing for each $n \in \mathbb{N}_+$).

Proof. Clearly $L(g) : S \to [0, \infty)$. If $x \preceq y$,

$$L(g)(x) = \int_{D[x]} g(t) d\lambda(t) \leq \int_{D[y]} g(t) d\lambda(t) = L(g)(y)$$

□

Next, consider the usual Banach space $\mathcal{L}(S)$, consisting of measurable functions $g : S \to \mathbb{R}$, with

$$||g|| = \int_S |g(x)| d\lambda(x) < \infty$$

and recall that $\mathcal{B}(S)$ is the set of measurable functions from S into $\mathbb{R}$.

Definition 20. Define the *upper operator* $U : \mathcal{L}(S) \to \mathcal{B}(S)$ by

$$U(g)(x) = \int_{I[x]} g(y) d\lambda(y)$$

Of course, U is a linear operator. The lower operator L and the upper operator U are complementary:

$$L(g)(x) = \int_{D[x]} g(t)d\lambda(t), \quad x \in S$$
$$U(g)(x) = \int_{I[x]} g(y)d\lambda(y), \quad x \in S$$

Moreover, we have the following duality:

Theorem 3. Suppose that $f \in \mathcal{D}(S)$ and that $g \in \mathcal{L}(S)$. Then, assuming that the integrals exist,

$$\int_S L(f)(x)g(x)d\lambda(x) = \int_S f(x)U(g)(x)d\lambda(x)$$

Proof. By Fubinni's theorem,

$$\begin{aligned}
\int_S L(f)(x)g(x)d\lambda(x) &= \int_S \left(\int_{D[x]} f(t)d\lambda(t)\right) g(x)d\lambda(x) \\
&= \int_S \int_{D[x]} f(t)g(x)d\lambda(t)d\lambda(x) \\
&= \int_S \int_{I[t]} f(t)g(x)d\lambda(x)d\lambda(t) \\
&= \int_S f(t) \int_{I[t]} g(x)d\lambda(x)d\lambda(t) \\
&= \int_S f(t)U(g)(t)d\lambda(t)
\end{aligned}$$

□

Note 16. In the notation of the $\mathcal{L}^2$ inner product,

$$\langle u, v\rangle = \int_S u(x)v(x)d\lambda(x)$$

Theorem 3 becomes the adjoint condition

$$\langle L(f), g\rangle = \langle f, U(g)\rangle$$

Note 17. Both operators can be written as integral operators with a kernel function. Define $r : S \times S \to \mathbb{R}$ by

$$r(x, y) = \begin{cases} 1 & \text{if } x \preceq y \\ 0 & \text{otherwise} \end{cases}$$

so that r is the *Riemann function* in the terminology of Möbius inversion (see Section 2.4). Then

$$L(f)(x) = \int_S r(t, x) f(t) d\lambda(t), \quad x \in S$$
$$U(f)(x) = \int_S r(x, t) f(t) d\lambda(t), \quad x \in S$$

2.2 Cumulative functions

Definition 21. Define $\lambda_n = L^n(\mathbf{1})$ (where $\mathbf{1}$ is the constant function 1 on S). The function λ_n is called the *cumulative function of order* n corresponding to λ.

Thus, $\lambda_0 = \mathbf{1}$ and

$$\lambda_{n+1}(x) = \int_{D[x]} \lambda_n(t) d\lambda(t), \quad x \in S$$

By Proposition 12, λ_n is increasing for each $n \in \mathbb{N}$.

Definition 22. For $n \in \mathbb{N}_+$, let λ^n denote the n-fold product measure on S^n, corresponding to the reference measure λ on S. Let

$$D_n = \{(x_1, x_2, \ldots, x_n) \in S^n : x_1 \preceq x_2 \preceq \cdots \preceq x_n\}$$

For $x \in S$, let

$$D_n[x] = \{(x_1, x_2, \ldots, x_n) \in S^n : x_1 \preceq x_2 \preceq \cdots \preceq x_n \preceq x\}$$

Note that $D_1[x]$ is simply $D[x]$ as defined earlier.

Proposition 13. For $n \in \mathbb{N}_+$, $\lambda_n(x) = \lambda^n(D_n[x])$.

Proof. By definition,

$$\lambda_1(x) = \int_{D[x]} 1 \, d\lambda(t) = \lambda(D[x]) = \lambda(D_1[x])$$

so the result holds for $n = 1$. Suppose that the result holds for a given $n \geq 1$. Then

$$\lambda^{n+1}(D_{n+1}[x]) = \int_S \lambda^n\{\boldsymbol{u} \in D^n : (\boldsymbol{u}, t) \in D_{n+1}[x]\} \, d\lambda(t)$$

But $(\boldsymbol{u}, t) \in D_{n+1}[x]$ if and only if $t \in D[x]$ and $\boldsymbol{u} \in D_n[t]$. Hence

$$\lambda^{n+1}(D_{n+1}[x]) = \int_{D[x]} \lambda^n(D_n[t]) \, d\lambda(t) = \int_{D[x]} \lambda_n(t) \, d\lambda(t) = \lambda_{n+1}(x)$$

$\square$

Proposition 14. Suppose that $(S, \preceq)$ is a discrete standard poset (with counting measure $\#$ as the reference measure, of course). Then $x \in S$ is a minimal element if and only if $\#_1(x) = 1$, in which case $\#_n(x) = 1$ for all $n \in \mathbb{N}$.

Proof. Note that $x \in S$ is a minimal element if and only if $D[x] = \{x\}$ if and only if $\#_1(x) = 1$. Inductivley, if $\#_n(x) = 1$ then

$$\#_{n+1}(x) = \sum_{y \in D[x]} \#_n(y) = \#_n(x) = 1$$

$\square$

Definition 23. We can define a generating function of sorts. Let

$$\Lambda(x, t) = \sum_{n=0}^{\infty} \lambda_n(x) t^n$$

for $x \in S$ and for $t \in \mathbb{R}$ for which the series converges absolutely. Let $r(x)$ denote the radius of convergence, so that the series converges absolutely for $|t| < r(x)$.

Since λ_n is increasing for each $n \in \mathbb{N}$, $x \mapsto \Lambda(x, t)$ is increasing for fixed t, and hence the radius of convergence r is decreasing. The generating function Λ turns out to be important in the study of the point process associated with a constant rate distribution.

2.3 Convolution

Definition 24. Suppose that $(S, \cdot)$ is a positive semigroup and that μ and ν are measures on S. The *convolution* of μ with ν is the measure $\mu\nu$ defined by

$$\mu\nu(A) = \int_S \nu(x^{-1}A) d\mu(x), \quad A \in \mathcal{B}(S)$$

Definition 25. Suppose that λ is a measure on S and that $f, g \colon S \to \mathbb{R}$ are locally bounded (that is, bounded on compact subsets of S). The *convolution* of f with g, with respect to λ, is the function $f * g$ defined by

$$(f * g)(x) = \int_{[e,x]} f(t) g(t^{-1}x)\, d\lambda(t), \quad x \in S$$

If $f \colon S \to \mathbb{R}$ is locally bounded, then for $n \in \mathbb{N}_+$, we denote the *convolution power* of f of order n by $f^{*n} = f * f * \cdots * f$ (n times). If there is any uncertainty about the underlying measure λ, we use $*_\lambda$.

Note 18. The definition makes sense. For fixed $x \in S$, $[e, x]$ and $\{t^{-1}x : t \preceq x\}$ are compact. Hence, there exist positive constants A_x and B_x such that $|f(t)| \leq$

A_x and $|g(t^{-1}x)| \le B_x$ for $t \in [e, x]$. Therefore

$$\left|\int_{[e,x]} f(t)g(t^{-1}x)\, d\lambda(x)\right| \le \int_{[e,x]} |f(t)||g(t^{-1}x|\, d\lambda(x)$$
$$\le \int_{[e,x]} A_x B_x \, d\lambda(x) \le A_x B_x \lambda[e,x] < \infty$$

Note 19. Convolution is associative: $f * (g * h) = (f * g) * h$, and the common value at $x \in S$ is given by

$$(f * g * h)(x) = \int_{[e,x]} \int_{[e,t]} f(s)g(s^{-1}t)h(t^{-1}x)\, d\lambda(s)\, d\lambda(t), \quad x \in S$$

However, since the semigroup operation is not in general commutative, neither is the convolution operation.

Note 20. In the case of a positive semigroup $(S, \cdot)$, the lower operator L on $\mathcal{D}(S)$ can be simply expressed interms of convolution: $L(g) = g * \mathbf{1}$, and hence $L^n(g) = g * \mathbf{1}^{*n}$. In particular, $\lambda_n = \mathbf{1}^{*(n+1)}$ for $n \in \mathbb{N}$.

2.4 Möbius inversion

Suppose that $(S, \preceq)$ is a discrete standard poset. In particular $D[x]$ is finite for each $x \in S$ and hence $(S, \preceq)$ is *locally finite* [3].

Definition 26. An *arithmetic function* is a function $f\colon S \times S \to \mathbb{R}$ with the properties

$$f(x,y) \ne 0 \text{ if } x = y$$
$$f(x,y) = 0 \text{ if } x \not\preceq y$$

The set $\mathcal{A}$ of arithmetic functions on S is a group with the operation

$$(f \cdot g)(x,y) = \sum_{t \in [x,y]} f(x,t)g(t,y)$$

The identity element is the *Kronecker delta function* $\delta \in \mathcal{A}$:

$$\delta(x,y) = \begin{cases} 1, & \text{if } x = y \\ 0, & \text{if } x \ne y \end{cases}$$

The inverse of $f \in \mathcal{A}$ is the function $f^{-1} \in \mathcal{A}$ defined inductively as follows:

$$f^{-1}(x,x) = \frac{1}{f(x,x)}$$
$$f^{-1}(x,y) = -\frac{1}{f(y,y)} \sum_{t \in [x,y)} f^{-1}(x,t)f(t,y) \text{ if } x \prec y$$

Definition 27. The *Riemann function* $r \in \mathcal{A}$ is defined as follows:

$$r(x, y) = \begin{cases} 1 & \text{if } x \preceq y \\ 0 & \text{otherwise} \end{cases}$$

The *Möbius function* $m \in \mathcal{A}$ is defined inductively as follows:

$$\begin{aligned} m(x, x) &= 1 \\ m(x, y) &= -\sum_{t \in [x, y)} m(x, t) \text{ if } x \prec y \end{aligned}$$

The functions r and m are inverses of each other. The *Möbius inversion formula* states that if f and g are real-valued functions on S and

$$f(x) = \sum_{t \in D[x]} g(t), \quad x \in S$$

then

$$g(x) = \sum_{t \in D[x]} f(t) m(t, x) \quad x \in S$$

Restated in terms of the operator L, the inversion formula gives a formula for the inverse L^{-1}:

$$\begin{aligned} L(g)(x) &= \sum_{t \in D[x]} g(t), \quad x \in S \\ L^{-1}(f)(x) &= \sum_{t \in D[x]} f(t) m(t, x), \quad x \in S \end{aligned}$$

3 Basic Constructions

There are several important ways to build new posets and positive semigroups from old ones.

3.1 Isomorphism

Posets $(S, \preceq)$ and $(T, \preceq)$ are isomorphic if there exists a one-to-one function Φ from S onto T such that $x \preceq y$ in S if and only if $\Phi(x) \preceq \Phi(y)$ in T.

Positive semigroups $(S, \cdot)$ and $(T, \cdot)$ are *isomorphic* if there exists a one-to-one function Φ from S onto T such that

$$\Phi(xy) = \Phi(x)\Phi(y), \quad x, y \in S$$

It follows that the partially ordered sets $(S, \preceq_S)$ and $(T, \preceq_T)$ are isomorphic as well.

If the posets or positive semigroups are topological, we require the isomorphism Φ to be continuous and have a continuous inverse. In this case, if λ is a measure on S, then μ defined by

$$\mu(B) = \lambda\left(\Phi^{-1}(B)\right), \quad B \in \mathcal{B}(T)$$

is a measure on T, and for any bounded, measurable $f\colon T \to \mathbb{R}$,

$$\int_T f(y) d\mu(y) = \int_S f\left(\Phi(x)\right) d\lambda(x)$$

In the semigroup case, if λ is left-invariant on $(S, \cdot)$ then μ is left-invariant on $(T, \cdot)$.

Suppose now that $(S, \cdot)$ is a positive semigroup, and that $\Phi\colon S \to T$ is one-to-one and onto. We can define an operation $\cdot$ on T by

$$uv = \Phi\left(\Phi^{-1}(u)\Phi^{-1}(v)\right), \quad u, v \in T$$

Proposition 15. $(T, \cdot)$ is a positive semigroup isomorphic to $(S, \cdot)$.

Proof. Let $u, v, w \in T$. Then

$$\begin{aligned}(uv)w &= \Phi[\Phi^{-1}(uv)\Phi^{-1}(w)] = \Phi[\Phi^{-1}(u)\Phi^{-1}(v)\Phi^{-1}(w)] \\ &= \Phi[\Phi^{-1}(u)\Phi^{-1}(vw)] = u(vw)\end{aligned}$$

Thus $(T, \cdot)$ is a semigroup. Let e denote the identity in S and let $i = \Phi(e) \in T$. For $u \in T$,

$$ui = \Phi[\Phi^{-1}(u)e] = u, \; iu = \Phi[e\Phi^{-1}(u)] = u$$

and therefore $ui = iu = u$. Thus i is the identity of $(T,)$. Suppose $u, v, w \in T$ and $uv = uw$. Then

$$\Phi[\Phi^{-1}(u)\Phi^{-1}(v)] = \Phi[\Phi^{-1}(u)\Phi^{-1}(w)]$$

Therefore $\Phi^{-1}(u)\Phi^{-1}(v) = \Phi^{-1}(u)\Phi^{-1}(w)$ and hence $\Phi^{-1}(v) = \Phi^{-1}(w)$ and so $v = w$. Suppose $u, v \in T$ and $uv = i$. Then $\Phi[\Phi^{-1}(u)\Phi^{-1}(v)] = i = \Phi(e)$ and hence $\Phi^{-1}(u)\Phi^{-1}(v) = e$. Therefore $\Phi^{-1}(u) = e$, $\Phi^{-1}(v) = e$ and therefore $u = v = i$. ☐

Again, if the spaces are topological, and Φ a homeomorphism, then $(S, \cdot)$ and $(T, \cdot)$ are isomorphic as topological positive semigroups, and so our previous comments apply.

3.2 Sub-semigroups

Suppose that $(S, \cdot)$ is a positive semigroup. A non-empty subset T of S is a *sub-semigroup* if $x, y \in T$ implies $xy \in T$, so that $(T, \cdot)$ is a semigroup in its own right. If T is a sub-semigroup of S and $e \in T$ then $(T, \cdot)$ is also a positive semigroup. Note that the nonexistence of inverses, the left cancellation law, and the associative property are inherited. If T is a sub-semigroup of S and $e \notin T$, then $T \cup \{e\}$ is also a sub-semigroup, and hence the previous comments apply. Thus, unless otherwise noted, we will only consider sub-semigroups that contain the identity.

The partial orders on S that are compatible with the semigroup operation $\cdot$ can be completely characterized in terms of the sub-semigroups of $(S, \cdot)$. Specifically, if T is a sub-semigroup of S we define the relation $\preceq_T$ on S by

$$x \preceq_T y \text{ if and only if } xt = y \text{ for some } t \in T$$

Theorem 4. Suppose that $(S, \cdot)$ is a positive semigroup.

1. If T is a sub-semigroup of S then $(S, \cdot, \preceq_T)$ is a left-ordered semigroup and $\preceq_T$ is a sub-order of $\preceq_S$.
2. Conversely, if $(S, \cdot, \preceq)$ is a left-ordered semigroup and $\preceq$ is a sub-order of $\preceq_S$ then $\preceq$ is $\preceq_T$ for some sub-semigroup T of S.
3. If T and U are sub-semigroups of S then $\preceq_T$ and $\preceq_U$ are the same if and only if $T = U$.

Proof. Suppose that T is a sub-semigroup of S. If $x \in S$ then $x \preceq_T x$ since $xe = x$ and $e \in T$. Suppose that $x \preceq_T y$ and $y \preceq_T x$. Then $xs = y$ and $yt = x$ for some $s, t \in T$. Hence $xst = x$ so $st = e$ by left cancellation, and hence $s = t = e$ since there are no non-trivial inverses. Suppose that $x \preceq_T y$ and $y \preceq_T z$. Then $y = xs$ and $z = yt$ for some $s, t \in T$. Hence $z = xst$ and therefore $x \preceq_T z$ since $st \in T$. Trivially, if $x \preceq_T y$ then $x \preceq_S y$. Suppose that $x \preceq_T y$ and $a \in S$. Then $y = xt$ for some $t \in T$ so $ay = axt$ and hence $ax \preceq_T ay$. Conversely, suppose that $ax \preceq_T ay$. The $ay = axt$ for some $t \in T$. By left-cancellation, $y = xt$ so $x \preceq_T y$. Thus, $(S, \cdot, \preceq_T)$ is a left-ordered semigroup.

Conversely, suppose that $\preceq$ is a sub-order of $\preceq_S$ and that $(S, \cdot, \preceq)$ is a left-ordered semigroup. Define

$$T = \{t \in S : e \preceq t\}$$

Let $s, t \in T$. Then $e \preceq t$ and hence $s = se \preceq st$. Also $e \preceq s$, so by transitivity, $e \preceq st$ and hence $st \in T$. Of course $e \preceq e$ so $e \in T$. Thus T is a sub-semigroup of S. Suppose that $x y \in S$ and $x \preceq y$. Since $\preceq$ is a sub-order of $\preceq_S$, there exists $t \in S$ so that $y = xt$. Thus we have $x = xe \preceq xt$ and hence $e \preceq t$. Hence $t \in T$ so $x \preceq_T y$. Conversely, suppose that $x, y \in S$ and $x \preceq_T y$. There exists $t \in T$ such that $y = xt$. But $e \preceq t$ and hence $x = xe \preceq xt = y$. Thus $\preceq$ is the same as $\preceq_T$.

Finally, suppose that T and U are sub-semigroups of S and $a \in T - U$. Then $a \preceq_T a^2$ but $a \not\preceq_U a^2$. ∎

Note 21. In particular, $\preceq_T$ is the same as $\preceq_S$ if and only if $T = S$. At the other extreme, if $T = \{e\}$ then $\preceq_T$ is the equality relation: $x \preceq_T y$ if and only if $x = y$. Of course, $\preceq_T$ restricted to T is the partial order associated with the positive semigroup $(T, \cdot)$. Note again however, that $\preceq_T$ restricted to T is not the same as $\preceq_S$ restricted to T in general. Specifically, for $x, y \in T$, $x \preceq_T y$ implies $x \preceq_S y$ but the converse is not true unless $x^{-1}y \in T$. This leads to our next definition.

Definition 28. A sub-semigroup T is (algebraically) *complete* in S if $x, y \in T$ and $x \preceq_S y$ imply $x^{-1}y \in T$.

Note 22. If T is complete in S then $\preceq_T$ restricted to T is the same as $\preceq_S$ restricted to T, and thus we may drop the subscripts. We will seldom need to refer to topological completeness, so the term *complete* will mean algebraically complete unless otherwise specified.

Proposition 16. Suppose that T_i is a sub-semigroup of S for each i in a nonempty index set I. Then $T = \bigcap_{i \in I} T_i$ is a sub-semigroup of S. If T_i is complete in S for each i, then so is T.

Proof. Suppose $x, y \in T$. Then $x, y \in T_i$ for each $i \in I$ and hence $xy \in T_i$ for each $i \in I$. Therefore $xy \in T$. Suppose T_i is complete in S for each $i \in I$, and let $x, y \in T$ with $x \preceq y$. Then $x, y \in T_i$ for each $i \in I$ and hence $x^{-1}y \in T_i$ for each $i \in I$. Therefore $x^{-1}y \in T$. □

Definition 29. Suppose that $A \subseteq S$. The sub-semigroup of $(S, \cdot)$ generated by A is the intersection of all sub-semigroups containing A:

$$S_A = \bigcap \{T \colon T \text{ is a sub-semigroup of } S \text{ and } A \subseteq T\}$$

The complete sub-semigroup of $(S, \cdot)$ generated by A is the intersection of all complete sub-semigroups containing A:

$$\hat{S}_A = \bigcap \{T \colon T \text{ is a complete sub-semigroup of } S \text{ and } A \subseteq T\}$$

Note 23. S_A is the smallest sub-semigroup containing A, and can be written as the set of all finite products of elements in A:

$$S_A = \{x_1 x_2 \cdots x_n \colon n \in \mathbb{N},\, x_i \in A \text{ for } i = 1, 2, \ldots n\}$$

As usual, an empty product is interpreted as e. Of course, the products

$$x_1 x_2 \cdots x_n$$

do not necessarily represent distinct elements in S_A, unless S_A is a free semigroup (see Section 13). Similarly, $\hat{S}_A$ is the smallest complete sub-semigroup containing A.

Proposition 17. Suppose that $(S, \cdot)$ is a topological positive semigroup and that T is a sub-semigroup of S. Then $\mathrm{cl}(T)$, the closure of T is also a sub-semigroup of S.

Proof. Suppose that $x,\, y \in \mathrm{cl}(T)$. Then there exist sequences $x_n \in T$ and $y_n \in T$ ($n \in \mathbb{N}_+$) such that $x_n \to x$ and $y_n \to y$ as $n \to \infty$. But T is a sub-semigroup, so $x_n y_n \in T$ for each $n \in \mathbb{N}_+$. By continuity, $x_n y_n \to xy$ as $n \to \infty$. Hence $xy \in \mathrm{cl}(T)$. $\square$

Proposition 18. Suppose that $(S,\, \cdot)$ is a topological positive semigroup and that T is a closed, complete sub-semigroup of S. Then $(T,\, \cdot)$ is also a topological, positive semigroup (where T is given the relative topology).

Proof. Since S is locally compact and T is a closed subset of S, T is locally compact, as a subspace of S. Trivially, T is Hausdorff and has a countable base, since S has these properties. Suppose that $x_n \in T$ for $n \in \mathbb{N}_+$ and that $x \in T$. Then $x_n \to x$ as $n \to \infty$ in T if and only if $x_n \to x$ as $n \to \infty$ in S. Hence the mapping $(x, y) \mapsto xy$ from T^2 to T is continuous. Similarly, because of completeness, the mapping $(x, y) \mapsto x^{-1}y$ from $\{(x, y) \in T^2 \colon x \preceq y\}$ to T is continuous. For $x \in T$, $[e, x]_T = [e, x] \cap T$ is compact in S and in T. If $x,\, y \in T$ and $x \prec y$ then $[e, y]$ is a neighborhood of x in S, so $[e, y]_T = [e, y] \cap T$ is a neighborhood of x in T, and xS is a neighborhood of y in S, so $xT = xS \cap T$ is a neighborhood of y in T. $\square$

Example 1. Suppose that $(S,\, \cdot)$ is a positive semigroup. The sub-semigroup generated by e is just $\{e\}$. If $t \neq e$, the sub-semigroup generated by t is

$$S_t = \{t^n \colon n \in \mathbb{N}\}$$

where by convention, $t^0 = e$. Note that $e = t^0 \prec t^1 \prec t^2 \prec \cdots$, so in particular the elements are distinct. Note also that S_t is complete in S: $t^m \preceq t^n$ in and only in $m \leq n$, in which case $(t^m)^{-1} t^n = t^{n-m}$.

Proposition 19. Suppose that $(S,\, \cdot)$ is a (topological) positive semigroup. If $t \neq e$ then S_t does not have a convergent subsequence. Moreover, S_t is closed and the relative topology is discrete.

Proof. Suppose that $t^{n_k} \to a$ as $k \to \infty$ where n_k is strictly increasing in k. If U is a convex neighborhood of a then $t^{n_k} \in U$ for k sufficiently large. But then $t^n \in U$ for n sufficiently large by convexity, since $t^n \in [t^{n_j}, t^{n_k}]$ for some j and k. Hence $t^n \to a$ as $n \to \infty$. But then $t^n t = t^{n+1}$, so taking limits and using continuity we have $at = a$. But then $t = e$ by left-cancellation, which is a contradiction. It now follows that S_t is closed. Finally, for each n there exists a convex neighborhood U of t^n that does not contain t^{n-1} or t^{n+1} by the Hausdorff property. But then by convexity, U does not contain t^m for $m > n+1$ or $m < n-1$. Thus, $\{t^n\}$ is open in S_t. $\square$

Proposition 20. If $A \subseteq S$ is increasing, then $T = A \cup \{e\}$ is a sub-semigroup. If T is non-trivial ($T \neq \{e\}$ and $T \neq S$) then T is not complete in S.

Proof. If $x \in A$ and $y \in S$ then $x \preceq xy$ and hence, by definition, $xy \in A$. In particular, if $x,\, y \in A$ then $xy \in A$, so T is a sub-semigroup. Suppose that T is non-trivial. Then there exists $u \notin T$ and $a \in A$, $a \neq e$. But then $a \in T$ and $x = au \in T$ and $a \preceq x$, but $a^{-1}x = u \notin T$, so T is not complete. $\square$

Example 2. Suppose that $(S, \cdot)$ is a positive semigroup and $A \subset S$. Then

$$AS = \{x \in S \colon x \succeq a \text{ for some } a \in A\}$$

is the increasing set generated by A. Thus, $T_A := \{e\} \cup AS$ is a sub-semigroup of S, and is not complete in S, unless T_A is trivial. Since T_A contains A, we have $S_A \subseteq T_A$.

Suppose that $(S, \cdot)$ is a positive semigroup and that T is a sub-semigroup of S. Recall that $\mathrm{cl}(T)$, the closure of T is also a sub-semigroup, so we will frequently assume that T is closed. We will always assume at least that $T \in \mathcal{B}(S)$ and that T with the relative topology satisfies the topological assumptions. Recall, in particular, that this is true when T is closed and complete. If λ is a measure on S and $\lambda(T) > 0$, then λ_T, the restriction of λ to $\mathcal{B}(T)$ is a measure on T. If λ is left-invariant on S, then λ_T is left-invariant on T. Recall that a locally compact topological group has a left-invariant measure, unique up to multiplication by positive constants, and that many positive semigroups are embedded in such groups.

3.3 Direct product

Suppose that $(S, \preceq_1, \mu)$ and $(T, \preceq_2, \nu)$ are standard posets. The direct product $(S \times T, \preceq, \lambda)$ is also a standard poset, where $S \times T$ is given the product topology, $\preceq$ is the product order

$$(u, v) \preceq (x, y) \text{ if and only if } u \preceq_1 x \text{ and } v \preceq_2 y$$

and where $\lambda = \mu \otimes \nu$, the product measure. Also, $\mathcal{B}(S \times T) = \mathcal{B}(S) \otimes \mathcal{B}(T)$ (the σ-algebra generated by the measurable rectangles). The cumulative functions are related as follows:

Proposition 21. $\lambda_n(x, y) = \mu_n(x)\nu_n(y)$ for $n \in \mathbb{N}$, $x \in S$, and $y \in T$,

Proof. Of course $\lambda_0(x, y) = 1 = \mu_0(x)\nu_0(y)$ for $(x, y) \in S \times T$. Assume that the result holds for $n \in \mathbb{N}$.Note that $D_{S\times T}[(x, y)] = D_S[x] \times D_T[y]$. Thus

$$\begin{aligned}
\lambda_{n+1}(x, y) &= \int_{D_{S\times T}[x,y]} \lambda_n(s, t) d\lambda(s, t) \\
&= \int_{D_S[x]} \int_{D_T[y]} \mu_n(s)\nu_n(t) d\nu(t) d\mu(s) \\
&= \mu_{n+1}(x)\nu_{n+1}(y)
\end{aligned}$$

$\square$

Note however that the generating function of $(\lambda_n : n \in \mathbb{N})$ has no simple representation in terms of the generating function of $(\mu_n : n \in \mathbb{N})$ and the generating function of $(\nu_n : n \in \mathbb{N})$:

$$\Lambda[(x,y),r] = \sum_{n=0}^{\infty} \lambda_n(x,y)r^n = \sum_{n=0}^{\infty} \mu_n(x)\nu_n(y)r^n$$
$$M(x,r) = \sum_{n=0}^{\infty} \mu_n(x)r^n, \; N(y,r) = \sum_{n=0}^{\infty} \nu(y)r^n$$

Suppose now that $(S, \cdot)$ and $(T, \cdot)$ are topological positive semigroups. The *direct product* is the semigroup $(S \times T, \cdot)$ with the binary operation $\cdot$ defined by

$$(x, y)(u, v) = (xu, yv)$$

and with the product topology. The direct product is also a topological positive semigroup. If e and ϵ are the identity elements in S and T respectively, then (e, ϵ) is the identity element in $S \times T$. If $\preceq_S$ and $\preceq_T$ are the partial orders corresponding to $(S, \cdot)$ and $(T, \cdot)$ respectively then the partial order $\preceq$ corresponding to $(S \times T, \cdot)$ is the product order described above.

Proposition 22. If μ and ν are left-invariant measures for $(S, \cdot)$ and $(T, \cdot)$, respectively, then $\mu \otimes \nu$ is left invariant for $(S \times T, \cdot)$. If $(S, \cdot)$ and $(T, \cdot)$ are standard positive semigroups, then so is $(S \times T, \cdot)$.

Proof. For $x \in S$, $y \in T$, $A \in \mathcal{B}(S)$, and $B \in \mathcal{B}(T)$,

$$\begin{aligned}(\mu \otimes \nu)[(x,y)(A \times B)] &= (\mu \otimes \nu)(xA \times yB) \\ &= \mu(xA)\nu(yB) = \mu(A)\nu(B) = (\mu \otimes \nu)(A \times B)\end{aligned}$$

Therefore, for fixed $(x,y) \in S \times T$, the measures on $S \times T$

$$\begin{aligned}C &\mapsto (\mu \otimes \nu)[(x,y)C] \\ C &\mapsto (\mu \otimes \nu)(C)\end{aligned}$$

agree on the measurable rectangles $A \times B$ where $A \in \mathcal{B}(S)$ and $B \in \mathcal{B}(T)$. Hence, these measures must agree on all of $\mathcal{B}(S \times T)$, and hence $\mu \otimes \nu$ is left-invariant on $(S \times T, \cdot)$.

Suppose now that $(S, \cdot)$ and $(T, \cdot)$ are standard, so that the left-invariant measures μ and ν are unique, up to multiplication by positive constants. Let $\mathcal{C}(T)$ denote the set of compact subsets of T. Suppose now that λ is a left-invariant measure for $(S \times T, \cdot)$. For $C \in \mathcal{C}(T)$, define

$$\mu_C(A) = \lambda(A \times C), \quad A \in \mathcal{B}(S)$$

Then μ_C is a regular measure on S (although it may not have support S). Moreover, for $x \in S$ and $A \in \mathcal{B}(S)$,

$$\mu_C(xA) = \lambda(xA \times C) = \lambda((x,\epsilon)(A \times C)) = \lambda(A \times C) = \mu_C(A)$$

so μ_C is left-invariant for $(S, \cdot)$. It follows that for each $C \in \mathcal{C}(T)$, there exists $\rho(C) \in [0, \infty)$ such that $\mu_C = \rho(C)\mu$; that is,

$$\lambda(A \times C) = \mu(A)\rho(C), \quad A \in \mathcal{B}(S),\, C \in \mathcal{C}(T) \tag{1}$$

Fix $A \in \mathcal{B}(S)$ with $0 < \mu(A) < \infty$. If C, $D \in \mathcal{C}(T)$ and $C \subseteq D$ then

$$\mu(A)\rho(C) = \lambda(A \times C) \le \lambda(A \times D) = \mu(A)\rho(D)$$

so $\rho(C) \le \rho(D)$. If C, $D \in \mathcal{C}(T)$ are disjoint then

$$\begin{aligned}\mu(A)\rho(C \cup D) &= \lambda(A \times (C \cup D)) = \lambda((A \times C) \cup (A \times D)) \\ &= \lambda(A \times C) + \lambda(A \times D) = \mu(A)\rho(C) + \mu(A)\rho(D)\end{aligned}$$

so $\rho(C \cup D) = \rho(C) + \rho(D)$. If C, $D \in \mathcal{C}(T)$ then

$$\begin{aligned}\mu(A)\rho(C \cup D) &= \lambda(A \times (C \cup D)) = \lambda((A \times C) \cup (A \times D)) \\ &\le \lambda(A \times C) + \lambda(A \times D) = \mu(A)\rho(C) + \mu(A)\rho(D)\end{aligned}$$

so $\rho(C \cup D) \le \rho(C) + \rho(D)$. Thus, ρ is a *content* in the sense of [15], and hence can be extended to a regular measure on T (which we will continue to call ρ). Thus, from (1) we have

$$\lambda(A \times C) = (\mu \otimes \rho)(A \times C), \quad A \in \mathcal{B}(S),\, B \in \mathcal{C}(T)$$

By regularity, it follows that $\lambda = \mu \otimes \rho$. Again fix $A \in \mathcal{B}(S)$ with $0 < \mu(A) < \infty$. If $y \in T$ and $B \in \mathcal{B}(T)$ then

$$\mu(A)\rho(yB) = \lambda(A \times yB) = \lambda((e, y)(A \times B)) = \lambda(A \times B) = \mu(A)\rho(B)$$

so it follows that $\rho(yB) = \rho(B)$ and hence ρ is left-invariant for $(T, \cdot)$. Thus, $\rho = c\nu$ for some positive constant c and so $\lambda = c\mu \otimes \nu$. Therefore $\mu \otimes \nu$ is the unique left-invariant measure for $(S \times T, \cdot)$, up to multiplication by positive constants. $\square$

The direct product $(S \times T, \cdot)$ has several natural sub-semigroups. First, $\{(x, \epsilon)\colon x \in S\}$ is a complete sub-semigroup isomorphic to S and $\{(e, y)\colon y \in T\}$ is a complete sub-semigroup isomorphic to T. If $S = T$, then $\{(x, x)\colon x \in S\}$ is a complete sub-semigroup isomorphic to S. On the other hand, $\{(x, y)\colon x \preceq y\}$ is a sub-semigroup that is not complete in general.

Naturally, the results in this subsection can be extended to the direct product of n positive semigroups $(S_1, \cdot)$, $(S_2, \cdot), \ldots (S_n, \cdot)$ and in particular to the n-fold direct power $(S^n, \cdot)$ of a positive semigroup $(S, \cdot)$.

Suppose that $(S_i, \cdot)$ is a discrete positive semigroup for $i = 1, 2, \ldots$. We can construct an infinite product that will be quite useful. Let

$$T = \{(x_1, x_2, \ldots) : x_i \in S_i \text{for each } i \text{ and } x_i = e_i \text{ for all but finitely many } i\}$$

As before, we define the component-wise operation:

$$(x_1, x_2, \ldots) \cdot (y_1, y_2, \ldots) = (x_1 y_1, x_2 y_2, \ldots)$$

so that the associated partial order is also component-wise: $\boldsymbol{x} \preceq \boldsymbol{y}$ if and only if $x_i \preceq_i y_i$ for each $i = 1, 2, \ldots$. The set T is countable and hence with the discrete topology and counting measure is a discrete positive semigroup.

3.4 Ordered groups

Definition 30. The triple $(G, \cdot, \preceq)$ is said to be a *left-ordered group* if $(G, \cdot)$ is a group; $\preceq$ is a partial order on G and for $x, y, z \in G$,

$$x \preceq y \Rightarrow zx \preceq zy$$

If e denote the identity of G, then $S = \{x \in G \colon e \preceq x\}$ is the set of *positive elements* of G (although more accurately, it should be called the set of *nonnegative elements* of S).

Proposition 23. If $(G, \cdot, \preceq)$ is a left-ordered group and S is the set of positive elements, then $(S, \cdot)$ is a positive semigroup, and for $x,\ y \in G$, $x \preceq y$ if and only if there exists $u \in S$ such that $xu = y$. In particular, $\preceq$ restricted to S is the partial order associated with S. Conversely, if $(G, \cdot)$ is a group and S is a positive sub-semigroup of G, then $\preceq$ defined by $x \preceq y$ if and only if there exists $u \in S$ such that $xu = y$ makes G into a left-ordered group with S as the set of positive elements.

Proof. Suppose that $(G, \cdot, \preceq)$ is a left-ordered group and S is the set of positive elements, Let $x,\ y \in S$. Then $e \preceq y$ so

$$e \preceq x = xe \preceq xy$$

and therefore $xy \in S$. Of course S inherits the associativity property from G so $(S, \cdot)$ is a semigroup. Next $e \preceq e$ so $e \in S$ so S has an identity. Also S inherits the left cancellation law from G. Suppose that $x,\ x^{-1} \in S$. Then $e \preceq x$ so $x^{-1} = x^{-1}e \preceq x^{-1}x = e$. But also $e \preceq x^{-1}$ so $x^{-1} = e$ and hence $x = e$. Therefore S has no nontrivial invertible elements.

Now let $x \in G$, $u \in S$. Then $e \preceq u$ so $x = xe \preceq xu$. Conversely, suppose $x,\ y \in G$ and $x \preceq y$. Then

$$e = x^{-1}x \preceq x^{-1}y$$

so $x^{-1}y \in S$. But $x(x^{-1}y) = y$.

Now suppose that $(G, \cdot)$ is a group and S a positive sub-semigroup. Define $x \preceq y$ if and only if there exists $u \in S$ such that $xu = y$. Then $x \preceq x$ since $e \in S$ and $xe = x$. Suppose $x \preceq y$ and $y \preceq x$. There exists $u, v \in S$ such that $xu = y$ and $yv = x$. But then

$$xuv = yv = x = xe$$

so $uv = e$ and hence $u = v = e$. Therefore $x = y$. Suppose $x \preceq y$ and $y \preceq z$. Then there exist $u,\ v \in S$ such that $xu = y$ and $yv = z$. But $uv \in S$ and $xuv = yv = z$ so $x \preceq z$. Suppose that $x \preceq y$ so that there exists $u \in S$ with $xu = y$. For any $z \in G$, $zxu = zy$ so $zx \preceq zy$. Hence, $(G, \cdot, \preceq)$ is a left-ordered group. Finally, if $e \preceq x$ then there exists $u \in S$ such that $u = eu = x$, so $x \in S$ and conversely, if $u \in S$ then $eu = u$ so $e \preceq u$. Thus S is the set of positive elements of G. $\square$

Note 24. In particular this result applies to the special case of a commutative ordered group and an ordered vector space. In any of these cases, the set of positive elements forms a positive semigroup. Conversely, if a positive semigroup $(S, \cdot)$ can be embedded in a group $(G, \cdot)$, then G can be ordered so that S is the set of positive elements. As the next result shows, this is always the case if the semigroup is commutative.

Proposition 24. Suppose that $(S, \cdot)$ is a commutative positive semigroup. Then S is isomorphic to the positive elements of a commutative ordered group.

Proof. The method is similar to the construction of the positive rationals from the positive integers. We define an equivalence relation $\sim$ on S^2 as follows:

$$(x, y) \sim (z, w) \text{ if and only if } xw = yz$$

First $(x, y) \sim (x, y)$ since $xy = yx$. If $(x, y) \sim (z, w)$ then $xw = yz$ so $zy = wx$ and hence $(z, w) \sim (x, y)$. Finally if $(u, v) \sim (x, y)$ and $(x, y) \sim (z, w)$ then $uy = vx$ and $xw = yz$ so

$$uyw = vxw = vyz$$

Hence $yuw = yvz$ so canceling gives $uw = vz$. Therefore $(u, v) \sim (z, w)$.

Let $[x, y]$ denote the equivalence class generated by (x, y) under $\sim$. Define

$$G = \{[x, y] \colon x, y \in S\}$$

Define a binary operator $\cdot$ on G by

$$[x, y][z, w] = [xz, yw]$$

Suppose that $(x, y) \sim (x_1, y_1)$ and $(z, w) \sim (z_1, w_1)$. Then $xy_1 = yx_1$ and $zw_1 = wz_1$. Hence

$$(xz)(y_1w_1) = (xy_1)(zw_1) = (yx_1)(wz_1) = (yw)(x_1z_1)$$

so $(xz, yw) \sim (x_1z_1, y_1w_1)$ and therefore the operator is well defined. Next

$$\begin{aligned}([u, v][x, y])[z, w] &= [ux, vy][z, w] = [uxz, vyw] \\ &= [u, v][xz, yw] = [u, v]([x, y][z, w])\end{aligned}$$

so the associative property holds and

$$[x, y][z, w] = [xz, yw] = [zx, wy] = [z, w][x, y]$$

so the commutative property holds. Note that $(x, x) \sim (y, y)$ for any $x, y \in S$ and

$$[x, x][y, z] = [xy, xz] = [y, z]$$

so $[x, x] =: \epsilon$ is the identity. Also

$$[x, y][y, x] = [xy, yx] = [xy, xy] = \epsilon$$

so every element of G has an inverse. Thus $(G, \cdot)$ is a group.

Next, the mapping $x \mapsto [x, e]$ defines an isomorphism from S to the following sub-semigroup of G:

$$S' = \{[x, e] \colon x \in S\}$$

Note that $x \neq y \Rightarrow [x, e] \neq [y, e]$ so the mapping is one-to-one, and$[x, e][y, e] = [xy, e]$ so the mapping is a homomorphism. It follows that $(S', \cdot)$ is a positive semigroup so the partial order $\preceq$ defined by $[x, y] \preceq [z, w]$ if and only if there exists $[u, e] \in S'$ such that $[x, y][u, e] = [z, w]$ makes G into an ordered group with S' as the set of positive elements. $\square$

3.5 Simple sums

Suppose first that $(S_i, \preceq_i, \lambda_i)$ is a standard poset for each $i \in I$, where I is a countable index set. We assume that S_i, $i \in I$ are disjoint. Now let

$$S = \bigcup_{i \in I} S_i$$

and define $\preceq$ on S by $x \preceq y$ if and only if $x, y \in S_i$ for some $i \in I$ and $x \preceq_i y$. That is, $\preceq$ is the union of the relations $\preceq_i$ over $i \in I$. We give S the topology which is the union of the topologies of S_i, over $i \in I$. We define λ on $\mathcal{B}(S)$ by

$$\lambda(A) = \sum_{i \in I} \lambda_i(S_i \cap A)$$

The poset $(S, \preceq, \lambda)$ is the *simple sum* of the posets $(S_i, \preceq_i, \lambda_i)$ over $i \in I$. In the discrete case, the covering graph of the new poset is obtained by juxtaposing the covering graphs of S_i, $i \in I$.

3.6 Lexicographic sums

Suppose that $(R, \preceq_R, \mu)$ is a standard poset and that for each $x \in R$, $(S_x, \preceq_x, \nu_x)$ is a standard poset. We define the lexicographic sum of S_x over $x \in R$ as follows: First, let

$$T = \{(x, y) : x \in R, y \in S_x\} = \bigcup_{x \in R} \{x\} \times S_x$$

Define the partial order $\preceq$ on T by

$$(u, v) \preceq (x, y) \text{ if and only if } u \prec_R x \text{ or } (u = x \text{ and } v \preceq_x y)$$

In the special case that $(R, \preceq_R)$ is a discrete antichain, the lexicographic sum reduces to the simple sum of S_x over $x \in R$, studied in Section 5.5. Since we have already studied that setting, let's assume that $(R, \preceq_R)$ is not an antichain, and that S_x is compact for each x. In this case, there is a natural topology that gives the usual topological assumptions and we can define the reference measure λ on T by

$$\lambda(C) = \int_S \nu_x(C_x) d\mu(x), \quad C \in \mathcal{B}(T)$$

where $C_x = \{y \in S_x : (x, y) \in C\}$, the cross section of C at $x \in R$. In the special case that $(S_x, \preceq_x, \nu_x) = (S, \preceq_S, \nu)$, independent of $x \in R$, we have the lexicographic *product* of $(R, \preceq_S, \mu)$ with $(S, \preceq_S, \nu)$. .

In the discrete case, the covering graph of the lexicographic sum can be constructed as follows:

1. Start with the covering graph of $(R, \preceq_R)$.
2. Replace each vertex x of the graph in step 1 with the covering graph of S_x.
3. If (x, y) is a directed edge in the graph in step 1, then in the graph in step 2, draw an edge from each maximal element of S_x to each minimal element of S_y.

We give an example of a construction with positive semigroups whose associated partial order is a lexicographic sum. This construction will be useful for counterexamples.

Example 3. Suppose that $(S, \cdot)$ is a standard discrete positive semigroup with minimum element e, and let I be a countable set with $0 \notin I$. Define

$$T = \{(e, 0)\} \cup \left(\bigcup_{x \in S - \{e\}} \{x\} \times I \right)$$

We define a binary operation $\cdot$ on T as follows: First $(e, 0)$ is the identiy element of T so that $(e, 0)(x, i) = (x, i)(e, 0) = (x, i)$ for $(x, i) \in T$. Less trivially,

$$(x, i)(y, j) = (xy, j), \quad x, y \in S - \{e\},\ i, j \in I$$

Proposition 25. In the setting of Example 3, $(T, \cdot)$ is a standard discrete positive semigroup with minimum element e and associated partial order

$$(x, i) \prec (y, j) \text{ if and only if } x \prec y$$

Moreover, this partial order corresponds to the lexicographic sum of $(I_x, \preceq_x)$ over $x \in S$ where $I_e = \{e\}$ and where $I_x = I$ and $\preceq_x$ is the equality relation for $x \in S - \{e\}$.

Proof. First, note that e is the identity of T by construction, so the basic properties of a positive semigroup need only be verified for non-identity elements. The operation $\cdot$ is associative:

$$((x, i)(y, j))(z, k) = (xy, j)(z, k) = (xyz, k)$$

and

$$(x, i)((y, j)(z, k)) = (x, i)(yz, k) = (xyz, k)$$

The left-cancellation property holds: if $(x, i)(y, j) = (x, i)(z, k)$ then $(xy, j) = (xz, k)$ so $xy = xz$ and $j = k$. But the left-cancellation law holds in S so $y = z$. Thus $(y, j) = (z, k)$. Clearly there are no non-trival inverses.

Now, suppose that $(x, i) \prec (y, j)$. Then there exists (z, k) such that

$$(x, i)(z, k) = (y, j)$$

But then $(xz, k) = (y, j)$ so in particular, $xz = y$ so $x \prec y$. Conversely, suppose that $x \prec y$ and $i, j \in I$. Then there exists $z \in S$ with $xz = y$ so

$$(x, i)(z, j) = (xz, j) = (y, j)$$

so $(x, i) \prec (y, j)$. $\square$

Note that for $x \in S - \{e\}$, the points $(x, i), i \in I$ are upper equivalent in the sense of Definition 13

3.7 Uniform posets

Definition 31. A discrete standard poset $(S, \preceq)$ is *uniform* if for every x, $y \in S$ with $x \preceq y$, all paths from x to y have the same length. We let $d(x, y)$ denote the common length.

Proposition 26. If $(S, \preceq)$ is a discrete standard poset with minimum element e, then $(S, \preceq)$ is uniform if and only if for every $x \in S$, all paths from e to x have the same length.

Proof. If $(S, \preceq)$ is uniform, then trivially all paths from e to x have the same length, for every $x \in S$. Conversely, suppose that all paths from e to x have the same length for every $x \in S$. Suppose that $x \preceq y$ and there are paths from x to y with lengths m and n. There must exist a path from e to x, since the poset is locally finite; let k denote the length of this path. Then we have two paths from e to y of lengths $k + m$ and $k + n$, so $k + m = k + n$ and hence $m = n$. $\square$

If the partially ordered set $(S, \preceq)$ is associated with a positive semigroup, then $(S, \preceq)$ is uniform if and only if, for each x, all factorings of x over I (the set of irreducible elements) have the same length. A rooted tree is always uniform, since there is a unique path from e to x for each $x \in S$.

3.8 Quotient spaces

Suppose that $(S, \cdot)$ is a standard positive semigroup with left-invariant measure λ and that T is a standard sub-semigroup of S with left-invariant measure μ. Let

$$S/T = \{z \in S : [e, z] \cap T = \{e\}\} = \bigcap_{t \in T - \{e\}} (S - tS).$$

Note that $e \in S/T$, but if $y \in T$ and $y \neq e$ then $y \notin S/T$. Thus, $T \cap S/T = \{e\}$ and so S/T is a type of quotient space. We are interested in factoring elements in S over the sub-semigroup T and the quotient space S/T.

Proposition 27. Suppose that $x \in S$. Then $x = yz$ for some $y \in T$ and $z \in S/T$ if and only if y is a maximal element (with respect to $\preceq_T$) of $[e, x] \cap T$ (and $z = y^{-1}x$).

Proof. Suppose that $x = yz$ for some $y \in T$ and $z \in S/T$. Then $y \preceq x$ by definition. Suppose that $t \in T$, $t \preceq x$, and $y \preceq_T t$. There exists $a \in S$ and $b \in T$ such that $x = ta$ and $t = yb$. Hence $x = yba$. By the left cancellation rule, $z = ba$, so $b \preceq z$. But $z \in S/T$ so $b = e$ and hence $t = y$. Hence y is maximal. Conversely, suppose that y is a maximal element of $[e, x] \cap T$. Then $y \preceq x$ so $x = yz$ for some $z \in S$. Suppose that $t \in T$ and $t \preceq z$. Then $z = tb$ for some $b \in S$ so $x = ytb$. Hence $yt \preceq x$ and $yt \in T$. Since y is maximal, $yt = y$ and so $t = e$. Therefore $z \in S/T$. ☐

For the remainder of this section, we impose the following assumptions:

Assumption 1. For each $x \in S$, $[e, x] \cap T$ has a unique maximal element $\varphi_T(x)$ (with respect to $\preceq_T$). The function $\varphi_T \colon S \to T$ is measurable.

Thus $S/T = \{z \in S \colon \varphi_T(z) = e\}$, so S/T is measurable as well. For $x \in S$, let $\psi_T(x) = \varphi_T^{-1}(x)x \in S/T$ so that $x \in S$ can be factored uniquely as $x = \varphi_T(x)\psi_T(x)$. This quotient space structure does correspond to an equivalence relation: if we define $u \sim v$ if and only if $\psi_T(u) = \psi_T(v)$, then $\sim$ is an equivalence relation on S and the elements in $S/T = \text{range}(\psi_T)$ generate a complete set of equivalence classes. However, $\sim$ is not a congruence in the sense of [13]. That is, if $u \sim v$, it is not necessarily true that $xu \sim xv$ (unless, of course, $x \in T$). Finally, note that the mapping $(\varphi_T, \psi_T) \colon S \to T \times (S/T)$ is one-to-one and onto. This mapping also preserves the measure-theoretic structure:

Proposition 28. There exists a measure ν on S/T such that

$$\lambda(AB) = \mu(A)\nu(B), \quad A \in \mathcal{B}(T),\ B \in \mathcal{B}(S/T)$$

Proof. For $C \in \mathcal{C}(S/T)$ and $A \in \mathcal{B}(T)$, let $\mu_C(A) = \lambda(AC)$. Then μ_C is a regular measure on T for each $C \in \mathcal{B}(T)$ (although μ_C may not have support T). Moreover, for $y \in T$ and $A \in \mathcal{B}(T)$,

$$\mu_C(yA) = \lambda(yAC) = \lambda(AC) = \mu_C(A)$$

so μ_C is left-invariant for $(T, \cdot)$. It follows that for each $C \in \mathcal{C}(S/T)$, there exists $\nu(C) \in [0, \infty)$ such that $\mu_C = \nu(C)\mu$; that is,

$$\lambda(AC) = \mu(A)\nu(C), \quad A \in \mathcal{B}(T),\ C \in \mathcal{C}(S/T)$$

Fix $A \in \mathcal{B}(T)$ with $0 < \mu(A) < \infty$. If C, $D \in \mathcal{C}(S/T)$ with $C \subseteq D$ then

$$\mu(A)\nu(C) = \lambda(AC) \le \lambda(AD) = \mu(A)\nu(D)$$

and hence $\nu(C) \le \nu(D)$. If C, $D \in \mathcal{C}(S/T)$ are disjoint then AC and AD are also disjoint and hence

$$\begin{aligned}\mu(A)\nu(C \cup D) &= \lambda(A(C \cup D)) = \lambda(AC \cup AD)\\ &= \lambda(AC) + \lambda(AD) = \mu(A)\nu(C) + \mu(A)\nu(D)\end{aligned}$$

Therefore $\nu(C \cup D) = \nu(C) + \nu(D)$. Finally, if $C,\, D \in \mathcal{C}(S/T)$ then

$$\begin{aligned}\mu(A)\nu(C \cup D) &= \lambda(A(C \cup D)) = \lambda(AC \cup AD) \\ &\le \lambda(AC) + \lambda(AD) = \mu(A)\nu(C) + \mu(A)\nu(D)\end{aligned}$$

and so $\nu(C \cup D) \le \nu(C) + \nu(D)$. It follows that ν is a content on S/T in the sense of [15], and hence can be extended to a measure on S/T (which we will also call ν). It now follows from regularity that

$$\lambda(AB) = \mu(A)\nu(B), \quad A \in \mathcal{B}(T),\, B \in \mathcal{B}(S/T)$$

□

The following examples should help clarify the assumptions. In particular, the assumptions are satisfied in our most important special case when $T = S_t$.

Example 4. Let $(S,\,\cdot)$ be a positive semigroup and let $t \in S - \{e\}$. The sub-semigroup $S_t = \{t^n \colon n \in \mathbb{N}\}$ has quotient space $S/S_t = S - tS$. Since $[e, x]$ is compact and S is locally convex, it is straightforward to show that $\{n \in \mathbb{N} : t^n \preceq x\}$ is finite for each x, and hence has a maximum element $n_t(x)$. Thus, Assumption 1 is satisfied and $\varphi_t(x) = t^{n_t(x)}$.

Example 5. Consider the standard positve semigroup $([0,\infty)^k, +, \lambda^k)$ where λ^k is k-dimensional Lebesgue measure, of course. The associated order $\le$ is the ordinary (product) order. Let $T = \mathbb{N}^k$, so that T is a discrete, complete, positive sub-semigroup of S. The quotient space is $S/T = [0,1)^k$ and the assumptions are satisfied. The decomposition is

$$\boldsymbol{x} = \boldsymbol{n} + \boldsymbol{t}$$

where $\boldsymbol{n}$ is the vector of integer parts of $\boldsymbol{x}$ and $\boldsymbol{t}$ is the vector of remainders. The left-invariant measure on T is counting measure, of course, and the reference measure on S/T is k-dimensional Lebesgue measure. Moreover, the partially ordered set $(S, \le)$ is the lexicographhc product of $(T, \le)$ with $(S/T, \le)$.

Example 6. Again, let $(S,\,\cdot)$ be a positive semigroup and let $t \in S - \{e\}$. The sub-semigroup $T_t = \{e\} \cup tS$ also has quotient space $S - tS$, but Assumption 1 of a unique decomposition is not satisfied in general.

Example 7. Consider the direct product $(S,\,\cdot)$ of standard positive semigroups $(S_1,\,\cdot)$ and $(S_2,\,\cdot)$, with identity elements e_1 and e_2, and with left-invariant measures λ_1 and λ_2, respectively. Let $T_1 = \{(x_1, e_2) : x_1 \in S_1\}$. Then T_1 is a closed sub-semigroup of S satisfying Assumption 1. Moreover, the quotient space $S/T_1 = \{(e_1, x_2) : x_2 \in S_2\}$ is also a positive semigroup. In this example, the spaces T_1 and $T_2 = S/T_1$ are symmetric; T_1 is isomorphic to S_1 and T_2 is isomorphic to S_2. The unique decomposition is simply $(x_1, x_2) = (x_1, e_2)(e_1, x_2)$. The measures μ and ν are

$$\begin{aligned}\mu(A) &= \lambda_1(A_1), \quad A \in \mathcal{B}(T_1) \\ \nu(B) &= \lambda_2(B_2), \quad B \in \mathcal{B}(T_2)\end{aligned}$$

where $A_1 = \{x_1 \in S_1 : (x_1, e_2) \in A\}$ and $B_2 = \{x_2 \in S_2 : (e_1, x_2) \in B\}$.

Part II
Probability Distributions

In this part we study probability distributions on partially ordered sets and positive semigroups. We are particularly interested in distributions that have exponential-type properties. The *constant rate* property requires only a partial order. The *memoryless* and *full exponential* properties require the complete semigroup structure. Interestingly, the most important secondary properties flow just from the simpler constant-rate property.

4 Preliminaries

As usual, we start with a standard poset $(S, \preceq, \lambda)$. Frequently, we will also need the additional strucutre of a positive semigroup $(S, \cdot, \lambda)$. Recall that all density functions are with respect to the reference measure λ.

4.1 Distribution functions

Definition 32. Suppose that X is a random variable taking values in S. The *upper probability function* of X is the mapping F given by

$$F(x) = \mathbb{P}(X \succeq x) = \mathbb{P}(X \in I[x]), \quad x \in S$$

In the case of a positive semigroup $(S, \cdot)$, we can also write the upper probability function as

$$F(x) = \mathbb{P}(X \in xS), \quad x \in S$$

The *lower probability function* of X is the mapping G given by

$$G(x) = \mathbb{P}(X \preceq x) = \mathbb{P}(X \in D[x]), \quad x \in S$$

Note 25. If X has probability density function f, then the upper probability function is $F = U(f)$ and the lower probability function is $G = L(f)$, where as usual, U and L are the lower and upper operators.

Note 26. We will primarily be interested in the upper probability function. Since $I[x]$ contains an open set for each $x \in S$, the upper probability function F is strictly positive if the random variable X has support S. Also, of course, $F(x) \leq 1$ for all $x \in S$, and F is decreasing on S. What other properties must it have?

Problem 4. Find conditions on $F : S \to (0, 1]$ so that F is an upper probability function. That is, find conditions on F so that there exists a random variable X with support S and $F(x) = \mathbb{P}(X \succeq x)$ for all $x \in S$.

Even when F *is* an upper probability function, the corresponding distribution is not necessarily unique. That is, the distribution of X is generally not determined by its upper probability function.

Example 8. Let A be fixed set with k elements ($k \geq 2$), and let $(S, \preceq)$ denote the lexicographic sum of the anti-chains $(A_n, =)$ over $(\mathbb{N}, \leq)$, where $A_0 = \{e\}$ and $A_n = A$ for $n \in \mathbb{N}_+$. Thus, $(0, e)$ is the minimum element of S and for n, $m \in \mathbb{N}_+$ and x, $y \in A$, $(m, x) \prec (n, y)$ if and only if $n < m$. Moreover, $(S, \preceq)$ is associated with a positive semigroup, as in Section 3.6.

Now let f be a probability density function on S with upper probability function F. Define g by

$$g(n, x) = f(n, x) + \left(-\frac{1}{k-1}\right)^n c_0 \tag{2}$$

and let

$$G(n, x) = \sum_{(m,y) \succeq (n,x)} g(m, y)$$

Then

$$\begin{aligned} G(n, x) &= g(n, x) + \sum_{m=n+1}^{\infty} \sum_{y \in A} g(m, y) \\ &= f(n, x) + \left(-\frac{1}{k-1}\right)^n c_0 + \sum_{m=n+1}^{\infty} \sum_{y \in A} \left[f(m, y) + c_0 \left(-\frac{1}{k-1}\right)^m \right] \\ &= F(n, x) + c_0 \left(-\frac{1}{k-1}\right)^n + \sum_{m=n+1}^{\infty} k c_0 \left(-\frac{1}{k-1}\right)^m \\ &F(n, x) + c_0 \left(-\frac{1}{k-1}\right)^n - c_0 \left(-\frac{1}{k-1}\right)^n = F(n, x) \end{aligned}$$

In particular, $G(0, e) = \sum_{(x,n) \in S} g(n, x) = 1$. It follows that if we can choose c_0 so that $g(n, x) > 0$ for every $(n, x) \in S$, then g is a probability density function with the same upper probability function as f. For any $k \geq 3$, there exists distinct PDFs f and g with the same upper probability function F. For example, define f by

$$\begin{aligned} f(0, e) &= \frac{6}{\pi^2} \\ f(n, x) &= \frac{6}{k\pi^2(n+1)^2}, \quad n \in \mathbb{N}_+,\ x \in A \end{aligned}$$

Thus, if (N, X) is the associated random variable, then $N + 1$ has the zeta distribution with parameter 2 and given $N = n$, X is uniformly distributed on A_n. The condition that $g(n, x) > 0$ is satisfied if

$$6(k-1)^n > c_0 k \pi^2 (n+1)^2$$

In turn, this condition will hold for any $k \geq 3$ if $0 < c_0 < \frac{1}{2\pi^2}$.

Conversely, if g is a PDF on S with upper probability function F, then it's not hard to show that g must have the form given in (2).

Note 27. However, since the partial order $\preceq$ is topologically closed, that is since

$$\{(x,y) \in S^2 : x \preceq y\}$$

is closed in S^2 (with the product topology) then a distribution on S is completely determined by its values on the increasing sets in $\mathcal{B}(S)$. That is, if μ and ν are distributions on S and $\mu(A) = \nu(A)$ for every increasing $A \in \mathcal{B}(S)$, then $\mu = \nu$ [17].

Problem 5. Find conditions sufficient for the probability distribution of X to be completely determined by its upper probability function $x \mapsto \mathbb{P}(X \succeq x)$.

Problem 6. Find conditions sufficient for the probability distribution of X to be completely determined by its lower probability function $x \mapsto \mathbb{P}(X \preceq x)$.

Proposition 29. If $(S, \preceq)$ is a discrete standard poset, then the distribution of a random variable X taking values in S is completely determined by its lower probability function.

Proof. Suppose that X has lower probability function F and probability density function f. Then

$$F(x) = \sum_{t \in D[x]} f(t), \quad x \in S$$

and hence by the Möbius inversion formula,

$$f(x) = \sum_{t \in D[x]} F(t) m(t,x), \quad x \in S$$

where m is the Möbius function. $\square$

Definition 33. We can define a more general upper probability function. If A is a finite subset of S, define

$$F(A) = \mathbb{P}(X \succeq x \text{ for each } x \in A)$$

Note 28. For a finite $A \subseteq S$, the set $\{y \in S : y \succeq x \text{ for all } x \in A\}$ is clearly increasing. So the question becomes whether a probability distribution is completely determined by its values on this special class of increasing sets.

Proposition 30. If $(S, \preceq)$ is a dscrete standard poset, then the distribution of a random variable X is completely determined by the generalized upper probability function.

Proof. Let X be a random variable with probability density function f and (generalized) upper probability function F. For $x \in S$,

$$\{X \succeq x\} = \{X = x\} \cup \{X \succ x\} = \{X = x\} \cup \bigcup_{y \in I(x)} \{X \succeq y\}$$

Suppose that $I(x)$ has n elements. By the inclusion-exclusion rule,

$$F(x) = f(x) + \sum_{k=1}^{n} (-1)^{k-1} \sum_{A \subseteq I(x), \#(A)=k} F(A)$$

Suppose that $I(x) = \{x_1, x_2, \ldots\}$ is countably infinite, and let

$$I_n(x) = \{x_1, x_2, \ldots, x_n\}$$

Then

$$F(x) = f(x) + \lim_{n \to \infty} \mathbb{P}\left(\bigcup_{y \in I_n(x)} \{X \succeq y\} \right)$$

By another application of the inclusion-exclusion rule,

$$F(x) = f(x) + \lim_{n \to \infty} \sum_{k=1}^{n} (-1)^{k-1} \sum_{A \subseteq I_n(x), \#(A)=k} F(A)$$

So in any event, $f(x)$ is determined by $F(A)$ for finte A. □

Problem 7. An extension of Problem 5 is to find the smallest n such that the distribution of X is completely determined by $F(A)$ for $A \subseteq S$ with $\#(A) \le n$. We will refer to this as the *distributional dimension* of the poset $(S, \preceq)$. In particular, distributional dimension 1 means that a probability distribution on $(S, \preceq)$ is uniquely determined by the upper probability function F.

Proposition 31. If $(S, \preceq)$ is a discrete upper semilattice, then $(S, \preceq)$ has distributional dimension 1.

Proof. If $A \subseteq S$ is finite, then $\sup(A)$ exists, so $F(A) = F(\sup(A))$. □

Proposition 32. If $(S, \preceq)$ is a rooted tree then $(S, \preceq)$ has distributional dimension 1.

Proof. For $x \in S$ recall that A_x denotes the set of elements that cover x. (These are the children of x since S is a rooted tree.) Let f be a probability density function on S with upper probability function F. If x is not maximal (not a leaf) then

$$f(x) = F(x) - \sum_{y \in A(x)} F(y)$$

If x is maximal, $f(x) = F(x)$. □

Example 9. Consider the poset $(S, \preceq)$ in Example 8. Suppose that X is a random variable taking values in S with probability density function f and upper probability function F. Then for $(n, x) \in S$,

$$f(n, x) = F(n, x) - F(\{n+1\} \times A)$$

Thus, the distributional dimension of $(S, \preceq)$ is no more than k. Is it exactly k? No. For example, suppose that $k = 4$ and $A = \{a, b, c, d\}$. For $(n, x) \in S$,

$$\begin{aligned} f(n, x) = F(n, x) &- F\{(n+1, a), (n+1, b)\} - F\{(n+1, b), (n+1, c)\} \\ &- F\{(n+1, c), (n+1, d)\} + F(n+1, b) + F(n+1, c) \end{aligned}$$

Thus, the distributional dimension is no greater than 2. But Example 8 shows that it is not 1 and hence must be exactly 2. Similar arguments show that the distributional dimension is 2 for any $k \geq 3$.

On the other hand, for $k = 2$, the distributional dimension is 1. Let $A = \{a, b\}$. For $(n, x) \in S$,

$$f(n, x) = F(n, x) + \sum_{i=1}^{\infty} (-1)^n [F(n+i, a) + F(n+i, b)]$$

Problem 8. For each $n \in \mathbb{N}_+$, construct a standard, discrete poset whose distributional dimension is n. If this cannot be done, what distributional dimensions are possible? In particular, is the distributional dimension always either 1 or 2?

Proposition 33. Suppose that $(S, \preceq)$ is a standard discrete poset, and that X is a random variable with upper probabiltiy function F. If $x_1 \prec x_2 \prec x_3 \cdots$ is an infinite chain in S then $F(x_n) \to 0$ as $n \to \infty$.

Proof. The events $\{X \succeq x_n\}$ are decreasing in n, and $\bigcap_{n=1}^{\infty} \{X \succeq x_n\} = \emptyset$ since S is locally finite. Hence $F(x_n) \to 0$ by the continuity theorem. $\square$

A consequence of the next proposition is that the upper probability function is measurable in the case of a positive semigroup.

Proposition 34. Suppose that $(S, \cdot)$ is a standard positive semigroup and that X is a random variable taking values in S. If $A \in \mathcal{B}(S)$ then $x \mapsto \mathbb{P}(X \in xA)$ is measurable.

Proof. Let $g \colon S^2 \to S^2$ be defined by $g(x, y) = (x, xy)$. Then g is continuous and one-to-one and hence $g(C) \in \mathcal{B}(S^2)$ for any $C \in \mathcal{B}(S^2)$. In particular, if $A \in \mathcal{B}(S)$

$$g(S \times A) = \{(x, xy) \colon x \in S,\, y \in A\} \in \mathcal{B}(S^2)$$

Therefore by Fubinni's Theorem,

$$x \mapsto \mathbb{E}\left(\mathbf{1}_{g(S \times A)}(x, X)\right)$$

is measurable. But for fixed $x \in S$, $(x, z) \in g(S \times A)$ if and only if $z = xy$ for some $y \in A$ if and only if $z \in xA$. Therefore $x \mapsto \mathbb{P}(X \in xA)$ is measurable. $\square$

Problem 9. Under what conditions is the mapping $x \mapsto \mathbb{P}(X \in xA)$ continuous on S for each $A \in \mathcal{B}(S)$? It is clearly not sufficient that X have a continuous distribution, so that $\mathbb{P}(X = x) = 0$ for each $x \in S$. For example, consider the positive semigroup $([0, \infty)^2, +)$, the direct power of the standard semigroup

$([0,\infty),+)$. A random variable (X,Y) could have a continuous distribution on $[0,\infty)^2$ and yet place positive probabiity on a vertical line $x = x_0$. The UPF F would be discontinous at (x_0, y) for each $y \in [0,\infty)$. There are clearly lots of other examples.

Definition 34. Suppose that X has upper probability function F and probability density function f. The *rate function* of X with respect to λ is the function $r\colon S \to (0,\infty)$ defined by

$$r(x) = \frac{f(x)}{F(x)}, \quad x \in S$$

.

Roughly speaking, $r(x)\, d\lambda(x)$ is the probability that X is in a small neighborhood of x (with measure $d\lambda(x)$), given that $X \succeq x$. If the semigroup is $([0,\,\infty),\,+)$ and λ is Lebesgue measure, then the rate function is the ordinary *failure rate function*; the value at x gives the probability density of failure at x given survival up to x. Of course, in the generality of our setting of posets and semigroups, the reliability interpretation does not have much meaning; we are using "rate" simply as a convenient term.

Definition 35. Random variable X has *constant rate* if r is constant on S, *increasing rate* if r is increasing on S and *decreasing rate* if r is decreasing on S.

We are particularly interested in distributions with constant rate; these will be studied in the next chapter.

Proposition 35. Suppose that $(S,\preceq)$ is a standard discrete poset, and that X is a random variable with support S and rate function r. Then $0 < r(x) \le 1$ for $x \in S$ and $r(x) = 1$ if and only if x is maximal.

Proof. As usual, let f denote the probability density function of X and F the upper probability function. For $x \in S$,

$$F(x) = f(x) + \mathbb{P}(X \succ x) = r(x)F(x) + \mathbb{P}(X \succ x)$$

Hence

$$\mathbb{P}(X \succ x) = [1 - r(x)]F(x) \tag{3}$$

Since X has support S, we must have $r(x) > 0$ for all x, and by (3) we must also have that $r(x) \le 1$ for all $x \in S$. If x is maximal, then $\mathbb{P}(X \succ x) = \mathbb{P}(\emptyset) = 0$ so $r(x) = 1$. Conversely, if x is not maximal then $\mathbb{P}(X \succ x) > 0$ (since X has support S) and hence $r(x) < 1$. $\square$

Suppose that the standard posets $(S,\preceq,\lambda)$ and $(T,\preceq,\mu)$ are isomorphic and let $\Phi\colon S \to T$ be an isomorphism. Thus, λ and μ are related by

$$\mu(B) = \lambda\left(\Phi^{-1}(B)\right), \quad B \in \mathcal{B}(T)$$

Proposition 36. Suppose that X is a random variable taking values in S and let $Y = \Phi(X)$. If X has upper probability function F then Y has upper probability function $G = F \circ \Phi^{-1}$. If X has probability density function f then Y has probability density function $g = f \circ \Phi^{-1}$.

Proof. First note that

$$\begin{aligned} G(y) &= \mathbb{P}(Y \succeq_T y) = \mathbb{P}(\Phi(X) \succeq_T y) \\ &= \mathbb{P}\left(X \succeq_S \Phi^{-1}(y)\right) = F\left(\Phi^{-1}(y)\right), \quad y \in T \end{aligned}$$

Next, for $B \in \mathcal{B}(T)$,

$$\begin{aligned} \mathbb{P}(Y \in B) &= \mathbb{P}(\Phi(X) \in B) = \mathbb{P}\left(X \in \Phi^{-1}(B)\right) \\ &= \int_{\Phi^{-1}(B)} f(x)\, d\lambda(x) = \int_B f\left(\Phi^{-1}(y)\right)\, d\mu(y) \end{aligned}$$

so $f \circ \Phi^{-1}$ is a density function for Y. ∎

Proposition 37. Suppose that the standard poset $(S, \preceq, \lambda)$ is a lower semilattice, and that X and Y are independent random variables taking values in S, with upper probability functions F and G respectively. The random variable $X \wedge Y$ has upper probability function FG.

Proof. Note that for x, $y \in S$, $x \wedge y \succeq z$ if and only if $x \succeq z$ and $y \succeq z$. Therefore

$$\mathbb{P}(X \wedge Y \succeq z) = \mathbb{P}(X \succeq z, Y \succeq z) = \mathbb{P}(X \succeq z)\mathbb{P}(Y \succeq z) = F(z)G(z)$$

∎

In particular, if S is a lower semilattice, the collection of upper probability functions forms a semigroup under pointwise multiplication.

Problem 10. Suppose that $(S, \preceq)$ is a standard poset (with no special structure). If F and G are upper probability functions on S is it true that FG is an upper probability function? If not, under what additional conditions will this be true?

Problem 11. Suppose that $(S, \preceq)$ is a standard poset, and that F is an upper probability function on S. For what values of $r > 0$ is F^r an upper probability function?

4.2 Residual distributions

In this section, we suppose that $(S, \cdot, \lambda)$ is a standard positive semigroup. The conditional distribution of $x^{-1}X$ given $X \succeq x$ will play a fundamental role in our study of exponential distributions in the next chapter. We will refer to this distribution as the *residual distribution at x*.

Proposition 38. Suppose that X has upper probability function F. For $x \in S$, the residual distribution at x is given by

$$\mathbb{P}(x^{-1}X \in A \mid X \succeq x) = \frac{\mathbb{P}(X \in xA)}{F(x)}, \quad A \in \mathcal{B}(S)$$

The upper probability function F_x of the residual distribution at x is given by

$$F_x(y) = \frac{F(xy)}{F(x)}, \quad y \in S$$

If X has density function f, then the residual distribution at x has density f_x given by

$$f_x(y) = \frac{f(xy)}{F(x)}, \quad y \in S$$

Proof. Let $A \in \mathcal{B}(S)$. Then

$$\mathbb{P}(x^{-1}X \in A \mid X \succeq x) = \frac{\mathbb{P}(x^{-1}X \in A, X \succeq x)}{\mathbb{P}(X \succeq x)} = \frac{\mathbb{P}(X \in xA)}{\mathbb{P}(X \in xS)} = \frac{\mathbb{P}(X \in xA)}{F(x)}$$

If we let $A = yS$ in this equation we get

$$\mathbb{P}(x^{-1}X \succeq y \mid X \succeq x) = \frac{\mathbb{P}(X \in xyS)}{F(x)} = \frac{\mathbb{P}(X \succeq xy)}{F(x)} = \frac{F(xy)}{F(x)}$$

Finally, suppose that X has density f. Returning to our first equation above, and using the integral version of the left-invariance property, we have

$$\mathbb{P}(x^{-1}X \in A \mid X \succeq x) = \frac{1}{F(x)} \int_{xA} f(z)\, d\lambda(z) = \int_A \frac{f(xy)}{F(x)}\, d\lambda(y)$$

∎

Note 29. If the semigroup is $([0, \infty), +)$ and the random variable X is interpreted as the lifetime of a device, then the upper probability function is the *survival function* or *reliability function* and the residual distribution at x is the *remaining lifetime distribution at* x, i.e., the conditional distribution of $X - x$ given $X \geq x$.

4.3 Expected value

In this section, we assume that $X$is a random variable taking values in a standard poset $(S, \preceq, \lambda)$. The following proposition gives a simple but important result on the expected values of the cumulative functions. Suppose that $g \in \mathcal{D}(S)$, and recall the cumulative functions $L^n(g)$ associated with g in Definition 19

Theorem 5. For $n \in \mathbb{N}$,

$$\int_S L^n(g)(x)\mathbb{P}(X \succeq x)d\lambda(x) = \mathbb{E}[L^{n+1}(g)(X)]$$

Proof. Using Fubinni's theorem,

$$\begin{aligned}
\int_S L^n(g)(x)\mathbb{P}(X \succeq x)d\lambda(x) &= \int_S L^n(g)(x)\mathbb{E}[\mathbf{1}(X \succeq x)]d\lambda(x) \\
&= \mathbb{E}\left(\int_{D[X]} L^n(g)(x)d\lambda(x)\right) \\
&= \mathbb{E}[L^{n+1}(g)(X)]
\end{aligned}$$

□

Note 30. In the case that X has a PDF f, Theorem 5 follows from Theorem 3. The proof as the following form, where F is the UPF and f PDF:

$$\begin{aligned}
\int_S L^n(g)(x)F(x)d\lambda(x) &= \int_S L^n(g)(x)U(f)(x)d\lambda(x) \\
&= \int_S L^{n+1}(g)(x)f(x)d\lambda(x) = \mathbb{E}[L^{n+1}(g)(X)]
\end{aligned}$$

Corollary 2. For $n \in \mathbb{N}$,

$$\int_S \lambda_n(x)\mathbb{P}(X \succeq x)d\lambda(x) = \mathbb{E}[\lambda_{n+1}(X)]$$

Corollary 3. In particular, when $n = 0$, Corollary 2 gives

$$\int_S \mathbb{P}(X \succeq x)d\lambda(x) = \mathbb{E}(\lambda(D[X])) \tag{4}$$

Corollary 4. When S is discrete and $n = 0$, Corollary 2 gives

$$\sum_{x \in S} \mathbb{P}(X \succeq x) = \mathbb{E}(\#(D[X]))$$

Corollary 5. In particular, when the poset is $(\mathbb{N}, \leq)$, then we get the standard result

$$\sum_{n=0}^{\infty} P(X \geq n) = \mathbb{E}(X) + 1$$

Corollary 6. When the poset is $([0, \infty), \leq)$, λ is Lebesgue measure, and $n = 0$ then Corollary 2 reduces to the standard result

$$\int_0^{\infty} \mathbb{P}(X \geq x)\, dx = \mathbb{E}(X)$$

4.4 Moment generating function

Suppose again that $(S, \preceq, \lambda)$ is a standard poset. Recall the generating function

$$\Lambda(x,t) = \sum_{n=0}^{\infty} \lambda_n(x) t^n, \quad x \in S, \ |t| < r(x)$$

where λ_n is the cumulative function of order n and where $r(x)$ is the radius of convergence for a given $x \in S$. If X is a random variable with values in S, then we can define a moment generating function of sorts:

$$\Lambda_X(t) = \mathbb{E}[\Lambda(X,t)] = \sum_{n=0}^{\infty} \mathbb{E}[\lambda_n(X)] t^n$$

The power series will converge for $|t| < r$ where r is the radius of convergence of the series. Assuming that $r > 0$, we can compute the cumulative moments in the usual way:

$$\mathbb{E}[\lambda_n(X)] = \frac{\Lambda_X^{(n)}(0)}{n!}, \quad n \in \mathbb{N}$$

Theorem 6. Suppose that X is a random variable with values in S and upper probability function F. Then

$$\Lambda_X(t) = 1 + t \int_S \Lambda(x,t) F(x) d\lambda(x)$$

Proof.

$$\Lambda_X(t) = \mathbb{E}[\lambda_0(X)] t^0 + \sum_{n=0}^{\infty} \mathbb{E}[\lambda_{n+1}(X)] t^{n+1} = 1 + \sum_{n=0}^{\infty} \mathbb{E}[\lambda_{n+1}(X)] t^{n+1}$$

But from Corollary 2, we can rewrite this as

$$\Lambda_X(t) = 1 + \sum_{n=0}^{\infty} \left(\int_S \lambda_n(x) F(x) d\lambda(x) \right) t^{n+1}$$

From Fubinii's theorem,

$$\Lambda_X(t) = 1 + t \int_S \left(\sum_{n=0}^{\infty} \lambda_n(x) t^n \right) F(x) d\lambda(x) = 1 + t \int_S \Lambda(x,t) F(x) d\lambda(x)$$

∎

4.5 Joint distributions

Suppose that $(S, \preceq, \mu)$ and $(T, \preceq, \nu)$ are standard posets, Recall that the direct product $(S \times T, \preceq, \mu \otimes \nu)$ is also a standard poset, where $S \times T$ is given the product topology and where

$$(u,v) \preceq (x,y) \text{ in } S \times T \text{ if and only if } u \preceq x \text{ in } S \text{ and } v \preceq y \text{ in } T$$

In this subsection, suppose that X and Y are random variables taking values in S and T, respectively, so that (X, Y) takes values in $S \times T$. The first proposition is a standard result.

Proposition 39. If (X, Y) has density h then X has density f and Y has density g, where

$$\begin{aligned} f(x) &= \int_T h(x, y) d\nu(y), \quad x \in S \\ g(y) &= \int_S h(x, y) d\mu(x), \quad y \in T \end{aligned}$$

If X and Y are independent, X has density function f, and Y has density function g, then (X, Y) and density function h where

$$h(x, y) = f(x) g(y), \quad (x, y) \in S \times T$$

When the posets have minimum elements, there is a simple relationship between the upper probability functions of X, Y, and (X, Y).

Proposition 40. Suppose that S has minimum element e and T has minimum element ϵ. Let F and G be the upper probability function of X and Y, respectively, and let H denote the upper probability function of (X, Y). Then

$$F(x) = H(x, \epsilon), \; G(y) = H(e, y); \quad x \in S, \, y \in T$$

If X and Y have independent upper events then

$$H(x, y) = F(x) G(y), \quad (x, y) \in S \times T$$

Proof. For $x \in S$,

$$F(x) = \mathbb{P}(X \succeq x) = \mathbb{P}(X \succeq x, Y \succeq \epsilon) = \mathbb{P}[(X, Y) \succeq (x, \epsilon)] = H(x, \epsilon)$$

Similarly for $y \in T$,

$$G(y) = \mathbb{P}(Y \succeq y) = \mathbb{P}(X \succeq e, Y \succeq y) = \mathbb{P}[(X, Y) \succeq (e, y)] = H(e, y)$$

If X and Y have independent upper events, then for $(x, y) \in S \times T$,

$$\begin{aligned} H(x, y) &= \mathbb{P}[(X, Y) \succeq (x, y)] = \mathbb{P}(X \succeq x, Y \succeq y) \\ &= \mathbb{P}(X \succeq x) \mathbb{P}(Y \succeq y) = F(x) G(y) \end{aligned}$$

□

Note 31. The results of Propositions 39 and 40 extend in a straightforward way to the direct product of n posets $((S_i, \preceq_i) \colon i \in \{1, 2, \ldots n\})$. In particular, the results extend to the n-fold direct power of a poset $(S, \preceq)$.

4.6 Products of independent variables

In this section, we suppose that $(S, \cdot, \lambda)$ is a standard positive semigroup.

Proposition 41. Suppose that X and Y are independent random variables taking values in S. Then (X, XY) takes values in $D_2 = \{(x, z) \in S^2 : x \preceq z\}$. If X and Y have densities f and g respectively, then the density of (X, XY) (with respect to λ^2) is the function h given by

$$h(x, z) = f(x)g(x^{-1}z) \quad (x, y) \in D_2$$

Proof. Note that technically, D_2 *is* the partial order $\preceq$. Let A, $B \in \mathcal{B}(S)$. Then

$$\mathbb{P}(X \in A,\, XY \in B) = \mathbb{E}\left(\mathbb{P}(X \in A,\, XY \in B|X)\right) = \mathbb{E}\left(\mathbb{P}(Y \in X^{-1}B|X); A\right).$$

Therefore

$$\mathbb{P}(X \in A,\, XY \in B) = \int_A \left(\int_{x^{-1}B} g(y)d\lambda(y)\right) f(x)d\lambda(x)$$

But by an integral version of the left-invariance property of λ,

$$\begin{aligned}\int_{x^{-1}B} g(y)d\lambda(y) &= \int_S g(y)\mathbf{1}_{x^{-1}B}(y)d\lambda(y)\\ &= \int_{xS} g(x^{-1}z)\mathbf{1}_{x^{-1}B}(x^{-1}z)d\lambda(z)\\ &= \int_B g(x^{-1}z)\mathbf{1}_{D_2}(x, z)d\lambda(z)\end{aligned}$$

Therefore

$$\mathbb{P}(X \in A,\, XY \in B) = \int_A \int_B f(x)g(x^{-1}z)\mathbf{1}_{D_2}(x, z)d\lambda(x)d\lambda(z)$$

∎

Corollary 7. Suppose that $(X_1, X_2, \ldots)$ is a sequence of independent random variables taking values in S and that X_i has density f_i for each i. Let $Y_n = X_1 \cdots X_n$ for $n \in \mathbb{N}_+$. Then the density of $(Y_1, Y_2, \ldots Y_n)$ (with respect to λ^n) is the function h_n given by

$$h_n(y_1, y_2, \ldots y_n) = f_1(y_1)f_2(y_1^{-1}y_2) \cdots f_n(y_{n-1}^{-1}y_n), \quad (y_1, y_2, \ldots, y_n) \in D_n$$

where we recall that$D_n = \{(y_1, y_2, \ldots, y_n) \in S^n : y_1 \preceq y_2 \preceq \cdots \preceq y_n\}$.

Corollary 8. Suppose that $(X_1, X_2, \ldots)$ is a sequence of independent, identically distributed random variables taking values in S, with common probabiity density function f. Let $Y_n = X_1 \cdots X_n$ for $n \in \mathbb{N}_+$. Then $\boldsymbol{Y} = (Y_1, Y_2, \ldots)$ is a homogenous Markov chain with transtion probability density function g given by

$$g(y, z) = f(y^{-1}z), \quad y \in S,\, z \in I[y]$$

In this case, the joint density of $(Y_1, Y_2, \ldots Y_n)$ (with respect to λ^n) is the function h_n given by

$$h_n(y_1, y_2, \ldots y_n) = f(y_1) f(y_1^{-1} y_2) \cdots f(y_{n-1}^{-1} y_n), \quad (y_1, y_2, \ldots, y_n) \in D_n$$

Proposition 42. Suppose again that X and Y are independent random variables taking values in S, with densities f and g, respectively. Then XY has density $f * g$ (all densities are with respect to λ).

Proof. For $A \in \mathcal{B}(S)$,

$$\begin{aligned}
\mathbb{P}(XY \in A) &= \int_S \left(\int_{x^{-1}A} g(y) d\lambda(y) \right) f(x) d\lambda(x) \\
&= \int_S \left(\int_S g(y) \mathbf{1}_{x^{-1}A}(y) d\lambda(y) \right) f(x) d\lambda(x)
\end{aligned}$$

But by one of the integral versions of left-invariance,

$$\begin{aligned}
\int_S g(y) \mathbf{1}_{x^{-1}A}(y) d\lambda(y) &= \int_{xS} g(x^{-1}z) \mathbf{1}_{x^{-1}A}(x^{-1}z) d\lambda(z) \\
&= \int_{xS} g(x^{-1}z) \mathbf{1}_A(z) d\lambda(z)
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathbb{P}(XY \in A) &= \int_S \int_{xS} g(x^{-1}z) \mathbf{1}_A(z) f(x) d\lambda(z) d\lambda(x) \\
&= \int_S \int_{[e,z]} g(x^{-1}z) \mathbf{1}_A(z) f(x) d\lambda(x) d\lambda(z)
\end{aligned}$$

and hence

$$\mathbb{P}(XY \in A) = \int_A \int_{[e,z]} g(x^{-1}z) f(x) d\lambda(x) d\lambda(z)$$

It follows that the density of XY is the convolution $f * g$. $\square$

4.7 Ladder Variables

Suppose again that $(S, \preceq, \lambda)$ is a standard poset. Let $\boldsymbol{X} = (X_1, X_2, \ldots)$ be a sequence of independent, identically distributed random variables, taking values in S, with common upper probability function F and density function f. We define the sequence of *ladder times* $\boldsymbol{N} = (N_1, N_2, \ldots)$ and the sequence of *ladder variables* $\boldsymbol{Y} = (Y_1, Y_2, \ldots)$ associated with $\boldsymbol{X}$ as follows: First

$$\begin{aligned}
N_1 &= 1 \\
Y_1 &= X_1
\end{aligned}$$

and then recursively,

$$\begin{aligned}
N_{n+1} &= \min\{n > N_n : X_n \succeq Y_n\} \\
Y_{n+1} &= X_{N_{n+1}}
\end{aligned}$$

Proposition 43. The sequence $\boldsymbol{Y}$ is a homogeneous Markov chain with transition probability density g given by

$$g(y,z) = \frac{f(z)}{F(y)}, \quad y \in S,\ z \in I[y]$$

Proof. Let $(y_1,\ldots,y_{n-1},y,z) \in D_{n+1}$. The conditional distribution of Y_{n+1} given $\{Y_1 = y_1,\ldots,Y_{n-1} = y_{n-1}, Y_n = y\}$ corresponds to observing independent copies of X until a variable occurs with a value greater that y (in the partial order). The distribution of this last variable is the same as the conditional distribution of X given $X \succeq y$. This conditional distribution has density $z \mapsto f(z)/F(y)$ on $I[y]$. $\square$

Corollary 9. The random vector $(Y_1, Y_2, \ldots, Y_n)$ has density function h_n given by

$$h_n(y_1,y_2,\ldots,y_n) = f(y_1)\frac{f(y_2)}{F(y_1)}\cdots\frac{f(y_n)}{F(y_{n-1})}, \quad (y_1,y_2,\ldots,y_n) \in D_n$$

Proof. Note that $Y_1 = X_1$ has probability density function f. Hence by Proposition 43, $(Y_1,Y_2,\ldots,Y_n)$ has probability density function

$$\begin{aligned} h_n(y_1,y_2,\ldots,y_n) &= f(y_1)g(y_2|y_1)\cdots g(y_n|y_{n-1}) \\ &= f(y_1)\frac{f(y_2)}{F(y_1)}\cdots\frac{f(y_n)}{F(y_{n-1})}, \quad (y_1,y_2,\ldots,y_n) \in D_n \end{aligned}$$

$\square$

Note 32. The density function of $(Y_1, Y_2, \ldots, Y_n)$ can also be written in terms of the rate function r as

$$h_n(y_1,y_2,\ldots,y_n) = r(y_1)r(y_2)\cdots r(y_{n-1})f(y_n), \quad (y_1,y_2,\ldots,y_n) \in D_n$$

Corollary 10. The density function g_n of Y_n is given by

$$g_n(y) = f(y)\int_{D_{n-1}[y]} r(y_1)r(y_2)\cdots r(y_{n-1})d\lambda^{n-1}(y_1,y_2,\ldots y_{n-1}), \quad y \in S$$

Corollary 11. The conditional distribution of $(Y_1,\ldots,Y_{n-1})$ given $Y_n = y$ has density function

$$h_{n-1}(y_1,\ldots,y_{n-1}|y) = \frac{1}{C_{n-1}[y]}r(y_1)\cdots r(y_{n-1}), \quad (y_1,\ldots,y_{n-1}) \in D[y]$$

where $C_{n-1}[y] = \int_{D_{n-1}[y]} r(y_1)\cdots r(y_{n-1})d\lambda^{n-1}(y_1,\ldots y_{n-1})$ is the normalizing constant.

4.8 The point process

Suppose that $(S, \preceq, \lambda)$ is a standard poset, and suppose that $\boldsymbol{Y} = (Y_1, Y_2, \ldots)$ is an increasing sequence of random variables with values in S. One special case is when $\boldsymbol{Y}$ is the sequence of ladder variables associated with an IID sequence $\boldsymbol{X} = (X_1, X_2, \ldots)$, as in Section 4.7. When $(S, \preceq, \lambda)$ is the poset associated with a standard positive semigroup $(S, \cdot, \lambda)$, then another special case occurs when $\boldsymbol{Y}$ is the partial product sequence corresponding to an IID sequence $\boldsymbol{X} = (X_1, X_2, \ldots)$. Of course, in this latter special case, the ladder sequence and the partial product sequence will be different in general.

In any event, we are interested in the corresponding point process. Thus, for $x \in S$, let

$$N_x = \#\{n \in \mathbb{N}_+ : Y_n \preceq x\}$$

That is, N_x is the number of random points in $D[x]$. We have the usual inverse relation between the processes $(Y_n : n \in \mathbb{N}_+)$ and $(N_x : x \in S)$:

Proposition 44. $Y_n \preceq x$ if and only if $N_x \geq n$ for $n \in \mathbb{N}_+$ and $x \in S$.

Proof. Suppose that $Y_n \preceq x$. Then $Y_k \preceq x$ for $k \leq n$ so $N_x \geq n$. Conversely, suppose that $Y_n \not\preceq x$. Then $Y_k \not\preceq x$ for $k \geq n$ so $N_x < n$ ☐

For $n \in \mathbb{N}_+$, let G_n denote the *lower* probability function of Y_n, so that

$$G_n(x) = \mathbb{P}(Y_n \preceq x), \quad x \in S$$

Then by Proposition 44, for $x \in S$,

$$\mathbb{P}(N_x \geq n) = \mathbb{P}(Y_n \preceq x) = G_n(x), \quad n \in \mathbb{N}_+$$

Of course, $\mathbb{P}(N_x \geq 0) = 1$. Let's adopt the convention that $G_0(x) = 1$ for all $x \in S$. Then for fixed $x \in S$, $n \mapsto G_n(x)$ is the *upper* probability function of N_x. Thus we have the following standard results as well, for $x \in S$:

$$\mathbb{P}(N_x = n) = G_n(x) - G_{n+1}(x), \quad n \in \mathbb{N}$$
$$\mathbb{E}(N_x) = \sum_{n=1}^{\infty} G_n(x)$$

A bit more generally, we can define

$$N(A) = \#\{n \in \mathbb{N}_+ : Y_n \in A\} = \sum_{n=1}^{\infty} \mathbf{1}(Y_n \in A), \quad A \in \mathcal{B}(S)$$

Thus, $A \mapsto N(A)$ is a random, discrete measure on S that places mass 1 at Y_n for each $n \in \mathbb{N}_+$. Of course, $N_x = N(D[x])$ for $x \in S$. Also,

$$E[N(A)] = \sum_{n=1}^{\infty} \mathbb{P}(Y_n \in A)$$

4.9 Entropy

In this section, we assume that $(S, \preceq, \lambda)$ is a standard poset. Suppose that X is a random variable taking values in S and that the distribution of X has density function f. The *entropy* of X is

$$H(X) = -\mathbb{E}[\ln(f(X))] = -\int_S f(x)\ln[f(x)]d\lambda(x)$$

The following proposition gives a fundamental inequality for entropy.

Proposition 45. Suppose that X and Y are random variables on S with density functions f and g respectively. Then

$$H(Y) = -\int_S g(x)\ln[g(x)]d\lambda(x) \le -\int_S g(x)\ln[f(x)]d\lambda(x)$$

with equality if and only if $\lambda\{x \in S : f(x) \neq g(x)\} = 0$ (so that X and Y have the same distribution).

Proof. Note first that $\ln(t) \le t - 1$ for $t > 0$, so $-\ln(t) \ge 1 - t$ for $t > 0$, with equality only at $t = 1$. Hence,

$$-\ln\left(\frac{f(x)}{g(x)}\right) = -\ln[f(x)] + \ln[g(x)] \ge 1 - \frac{f(x)}{g(x)}, \quad x \in S$$

Multiplying by $g(x)$ gives

$$-g(x)\ln[f(x)] + g(x)\ln[g(x)] \ge g(x) - f(x), \quad x \in S$$

Therefore

$$\begin{aligned} -\int_S g(x)\ln[f(x)]d\lambda(x) + \int_S g(x)\ln[g(x)] &\ge \int_S g(x)d\lambda(x) - \int_S f(x)d\lambda(x) \\ &= 1 - 1 = 0 \end{aligned}$$

Equality holds if and only $f(x)/g(x) = 1$ except on a set of λ measure 0. $\square$

5 Distributions with Constant Rate

5.1 Definitions and basic properties

As usual, we start with a standard poset $(S, \preceq, \lambda)$. Let X be a random variable taking values in S. We can extend the definition of constant rate to the case where the support of X is a proper subset of S.

Definition 36. Suppose that X has upper probability function F. Then X has constant rate $\alpha > 0$ if $f = \alpha F$ is a probability density function of X.

Theorem 7. Suppose that $g \in \mathcal{D}(S)$. If X has constant rate α, then

$$\mathbb{E}[L^n(g)(X)] = \frac{1}{\alpha^n}\mathbb{E}[g(X)], \quad n \in \mathbb{N}$$

Proof. Let F denote the upper probability function of X, so that $f = \alpha F$ is a density function of X with respect to λ. By Theorem 5,

$$\begin{aligned}\mathbb{E}[L^{n+1}(g)(X)] &= \int_S L^n(g)(x)F(x)d\lambda(x) \\ &= \frac{1}{\alpha}\int_S L^n(g)(x)f(x)d\lambda(x) = \frac{1}{\alpha}\mathbb{E}[L^n(g)(X)]\end{aligned}$$

and of course, $\mathbb{E}[L^0(g)(X)] = \mathbb{E}[g(X)]$ since $L^0(g) = g$. ∎

Corollary 12. If X has constant rate α, then

$$\mathbb{E}[\lambda_n(X)] = \frac{1}{\alpha^n}, \quad n \in \mathbb{N}$$

In particular, $\mathbb{E}[\lambda_1(X)] = \mathbb{E}[\lambda(D[X])] = 1/\alpha$, so the rate constant is the reciprocal of $\mathbb{E}[\lambda_1(X)]$, which must be finite.

Corollary 13. If X has constant rate α then the generating function of X is

$$\Lambda_X(t) = \frac{\alpha}{\alpha - t}, \quad |t| < \alpha$$

Proof. Recall that $\Lambda_X(t) = \mathbb{E}[\Lambda(X,t)] = \sum_{n=0}^{\infty}\mathbb{E}[\lambda_n(X)]t^n$. Hence if X has constant rate,

$$\Lambda_X(t) = \sum_{n=0}^{\infty}\left(\frac{t}{\alpha}\right)^n = \frac{1}{1 - t/\alpha} = \frac{\alpha}{\alpha - t}, \quad |t| < \alpha$$

∎

The converse of Theorem 7 also holds, and thus gives a characterization of constant rate distributions.

Theorem 8. Suppose that random variable X satisfies

$$\mathbb{E}[L(g)(X)] = \frac{1}{\alpha}\mathbb{E}[g(X)] \tag{5}$$

for every $g \in \mathcal{D}(S)$. Then X has constant rate α.

Proof. Suppose that (5) holds for every $g \in \mathcal{D}(S)$. Let F dentoe the upper probability function of X. For $g \in \mathcal{D}(S)$,

$$\begin{aligned}\mathbb{E}[L(g)(X)] &= \mathbb{E}\left[\int_{D[X]} g(t)d\lambda(t)\right] = \mathbb{E}\left[\int_S g(t)\mathbf{1}(t \preceq X)d\lambda(t)\right] \\ &= \int_S g(t)\mathbb{P}(X \succeq t) = \int_S g(t)F(t)d\lambda(t)\end{aligned}$$

Thus (5) gives

$$\int_S g(t)\alpha F(t)d\lambda(t) = \mathbb{E}[g(X)]$$

It follows that αF is a density of X and hence X has constant density. $\square$

Theorem 9. The poset $(S, \preceq, \lambda)$ supports a distribution with constant rate α if and only if there exists a measureable function $G : S \to (0, \infty)$ that satisfies

$$\int_S G(x)d\lambda(x) < \infty \tag{6}$$

$$\int_{I[x]} G(y)d\lambda(y) = \frac{1}{\alpha}G(x), \quad x \in S \tag{7}$$

Proof. If F is the upper probability function of a distribution with constant rate α, then trivially (6) and (7) hold with $G = F$, since $f = \alpha F$ is a probability density function of the distribution. Conversely, suppose $G : S \to (0, \infty)$ satisfies (6) and (7). Let

$$C = \int_S G(x)d\lambda(x)$$

and let $F = G/(\alpha C)$. Then from (6),

$$\int_S \alpha F(x)d\lambda(x) = \frac{1}{C}\int_S G(x)d\lambda(x) = 1$$

so $f = \alpha F$ is a density function with respect to λ. Next, from (7),

$$\int_{I[x]} f(x)d\lambda(x) = \int_{I[x]} \alpha F(y) = \frac{1}{C}\int_{I[x]} G(y)d\lambda(y) = \frac{1}{\alpha C}G(x) = F(x)$$

so F is the upper probability function corresponding to f. $\square$

Note 33. If the poset $(S, \preceq)$ has a minimum element e (in particular, if the poset is associated with a a standard positive semigroup $(S, \cdot, \lambda)$, then condition (6) is unnecessary. Letting $x = e$ in (7) gives

$$\int_S G(x)d\lambda(x) = \int_{I[e]} G(y)d\lambda(y) = \frac{1}{\alpha}G(e) < \infty$$

From Theorem 9, it's clear that finding constant rate distributions is an eigenvalue-eigenvector problem for the upper operator U. From Theorem 9, $(S, \preceq, \lambda)$ supports a distribution with constant rate $\alpha > 0$ if and only if there exists a positive $g \in \mathcal{L}(S)$ satisfying $U(g) = \frac{1}{\alpha}g$. That is, if and only if there is a strictly positive eigenfunction of the operator U corresponding to the eigenvalue $\frac{1}{\alpha}$. The upper probability function F of a distribution with constant rate α is a positive eigenfunction of the operator U with the additional property that $||F|| = 1/\alpha$. It's interesting that we can characterize constant rate distributions in terms of both of the lower and upper operators.

5.2 Discrete Posets

Suppose that $(S, \preceq)$ is a standard discrete poset, so that the reference measure is counting measure $\#$. Suppose that X has constant rate α on S. Recall from Proposition 35 that if $x \in S$ then

$$(1-\alpha)\mathbb{P}(X \succeq x) = \mathbb{P}(X \succ x), \quad x \in S \tag{8}$$

If $\alpha = 1$, then $\mathbb{P}(X \succ x) = 0$ for every $x \in S$, so X must be concentrated on the set of minimal elements of S (which form an anti-chain). Conversely, any distribution concentrated on the minimial elements has constant rate 1. In particular, if S is an anti-chain, then every distribution on S has constant rate 1 since

$$P(X = x) = \mathbb{P}(X \succeq x), \quad x \in S$$

On the other hand, if $\mathbb{P}(X \succ x) > 0$ for some $x \in S$, then we must have $0 < \alpha < 1$.

Suppose now that x is a maximal element of S. Then $\mathbb{P}(X \succ x) = 0$, so from (8), either $F(x) = 0$ or $\alpha = 1$. Thus, either $F(x) = 0$ for every maximal element x of S, or $\alpha = 1$ and X is concentrated on the minimal elements of S.

Suppose that X has constant rate α on S and is not supported by the minimal elements. Thus the maximal elements of S are not in the support set and hence can be removed. Repeating the argument, we can remove the maximal elements of the new poset. Continuing in this way, we see that if x is in the support set of X then there must be an infinite chain in S containing x.

Suppose that X has constant rate α on S and upper probability function F. If x, y are upper equivalent (Definition 13), then $\mathbb{P}(X \succ x) = \mathbb{P}(X \succ y)$ so from (8), $F(x) = F(y)$. Thus, the upper probability function (and hence also the density function) are constant on the equivalence classes.

Note 34. If F is an upper probability function of a distribution with constant rate α on $(S, \preceq)$, then F is well defined on the partially ordered collection of equivalence classes $(\Pi, \preceq)$ (see Theorem 2). However in general, the mapping $[x] \mapsto F(x)$ on $\Pi = S/\equiv$ is not an upper probabiltiy function, let alone one with constant rate. In general, it's not clear how to go from constant rate distributions on $(S, \preceq)$ to constant rate distributions from $(\Pi, \preceq)$, or the other way around.

5.3 Entropy

Theorem 10. Suppose that X has constant rate and upper probability function F. Then X maximizes entropy among all random variables Y on S for which $\mathbb{E}[\ln(F(Y))]$ is constant.

Proof. Suppose that X has constant rate α, so that $f = \alpha F$ is a density of X. Suppose that Y is a random variable taking values in S, with density function

g. From Proposition 45 we have

$$\begin{aligned} H(Y) &= -\int_S g(x)\ln[g(x)]d\lambda(x) \le -\int_S g(x)\ln[f(x)]d\lambda(x) \\ &= -\int_S g(x)\{\ln(\alpha) + \ln(F(x))\}d\lambda(x) \\ &= -\ln(\alpha) - \int_S g(x)\ln[F(x)]d\lambda(x) = -\ln(\alpha) - \mathbb{E}[\ln(F(Y))] \end{aligned}$$

Of course, the entropy of X achieves the upper bound. $\square$

Note that since the upper probability function F typically has an "exponential" form of some sort, $\mathbb{E}[\ln(F(Y))]$ often reduces to a natural moment condition.

5.4 Sub-posets

Suppose that $(S, \preceq, \lambda)$ is a standard poset and that X has constant rate α on S with upper probability function F. Suppose further that $T \subseteq S$ and $\lambda(T) > 0$. In this section, we are interested in the conditional distribution of X given $X \in T$ on the poset $(T, \preceq, \lambda)$. The upper probability function of the conditional distribution of X given X in T is

$$F_T(x) = \mathbb{P}(X \succeq x | X \in T) = \frac{\mathbb{P}(X \succeq x, X \in T)}{\mathbb{P}(X \in T)}, \quad x \in T \tag{9}$$

The density of the conditional distribution is

$$f_T(x) = \frac{\alpha F(x)}{\mathbb{P}(X \in T)} = \frac{\alpha \mathbb{P}(X \succeq x)}{\mathbb{P}(X \in T)}, \quad x \in T \tag{10}$$

So the conditional distribution of X given $X \in T$ has constant rate β if and only if

$$\alpha \mathbb{P}(X \succeq x) = \beta \mathbb{P}(X \succeq x, X \in T), \quad x \in T \tag{11}$$

In particular, we see that if we take T to be the support of X then $\mathbb{P}(X \in T) = 1$ so the conditional distribution of X given $X \in T$ also has constant rate α. Thus, we are particularly interested in the case when X has support S; equivalently, we are interested in knowing when a poset S supports distributions with constant rate.

Theorem 11. Suppose that S has a minimum element e and that $e \in T$. Then the conditional distribution of X given $X \in T$ has constant rate on T if and only if $\{X \in T\}$ and $\{X \succeq x\}$ are independent for each $x \in T$. The rate constant of the conditional distribution is $\alpha/\mathbb{P}(X \in T)$.

Proof. Suppose that $\{X \in T\}$ and $\{X \succeq x\}$ are independent for each $x \in T$. Then from (9), $F_T(x) = F(x)$ for $x \in T$ and from (10),

$$f_T(x) = \alpha \frac{F(x)}{\mathbb{P}(X \in T)}, \quad x \in T$$

Thus, the conditional distribution has constant rate $\alpha/\mathbb{P}(X \in T)$. Conversely, suppose that the conditional distribution of X given $X \in T$ has constant rate β. Then taking $x = e$ in (11) we have $\alpha = \beta\mathbb{P}(X \in T)$ and then substituting back we have

$$\mathbb{P}(X \succeq x)\mathbb{P}(X \in T) = \mathbb{P}(X \succeq x, X \in T), \quad x \in T$$

so $\{X \in T\}$ and $\{X \succeq x\}$ are independent for each $x \in T$. $\square$

Theorem 12. If $t \in S$ with $F(t) > 0$, then the conditional distribution of X given $X \succeq t$ has constant rate α on $I[t] = \{x \in S : x \succeq t\}$.

Proof. From (9), the upper probability function of the conditional distribution of X given $X \succeq t$ is

$$F_t(x) = \frac{F(x)}{F(t)}, \quad x \succeq t$$

From (10), the conditional density of X given $X \succeq t$ is

$$f_t(x) = \frac{\alpha F(x)}{F(t)}, \quad x \succeq t$$

Hence $f_t(x) = \alpha F_t(x)$ for $x \succeq t$, so the conditional distribution of X given $X \succeq t$ also has constant rate α. $\square$

Proposition 46. Suppose $(S, \cdot, \lambda)$ is a standard positive semigroup and that X is a random variable taking values in S and that X has constant rate α. Then the conditional distribution of $x^{-1}X$ given $X \succeq x$ also has constant rate α for every $x \in S$.

Proof. Let F denote the upper probability function of X. Recall from Proposition 38 that the upper probability function F_x of $x^{-1}X$ given $X \succeq x$ is

$$F_x(y) = \frac{F(xy)}{F(x)}, \quad y \in S$$

If f is the density function of X, then the density function f_x of the conditional distribution is

$$f_x(y) = \frac{f(xy)}{F(x)}, \quad y \in S$$

Since X has constant rate α, we have $f = \alpha F$ and therefore

$$f_x(y) = \frac{\alpha F(xy)}{F(x)} = \alpha \frac{F(xy)}{F(x)} = \alpha F_x(y), \quad y \in S$$

$\square$

5.5 Mixtures

Many important distributions are actually parametric families of distributions. New distributions can be created by randomizing one or more of the parameters; the new distributions are *mixtures* of the given distributions. This process can be applied to a family of distributions with constant rate on a standard poset $(S, \preceq, \lambda)$; the constant rate property is preserved.

Specifically, suppose that X_i is a random variable taking values in S for each $i \in I$, where $(I, \mathcal{I})$ is a measurable space. Suppose that U is a random variable taking values in I, independent of $(X_i \colon i \in I)$. Thus, the distribution of X_U is mixture of the distributions of $(X_i \colon i \in I)$, and U is the mixing variable.

Theorem 13. If X_i has constant rate α for each $i \in I$ then X_U has constant rate α.

Proof. For $i \in I$, let F_i denote the upper probability function of X_i, so that X_i has density function αF_i. Let G denote the upper probability function of X_U. Then

$$G(x) = \mathbb{P}(X_U \succeq x) = \mathbb{E}\left[\mathbb{P}(X_U \succeq x | U)\right] = \mathbb{E}[F_U(x)], \quad x \in S$$

Next, for $A \in \mathcal{B}(S)$,

$$\begin{aligned}\mathbb{P}(X_U \in A) &= \mathbb{E}\left[\mathbb{P}(X_U \in A | U)\right] \\ &= \mathbb{E}\left[\int_A \alpha F_U(x) d\lambda(x)\right] = \int_A \alpha \mathbb{E}\left[F_U(x)\right] d\lambda(x)\end{aligned}$$

Hence X_U has density function $g(x) = \alpha \mathbb{E}\left[F_U(x)\right]$ and thus X_U has constant rate α. $\square$

We consider two applications of mixing—to simple sums and to simple joins. Suppose first that $(S_i, \preceq_i, \lambda_i)$ is a standard poset for each $i \in I$, where I is a countable index set. Let $(S, \preceq, \lambda)$ be the simple sum of $(S_i, \preceq_i, \lambda_i)$ over $i \in I$ as defined in Section 3.5.

Suppose now that X_i is a random variable taking values in S_i for each $i \in I$ and that X_i has constant rate α and upper probability function F_i, for each $i \in I$. Without loss of generality, we can assume that X_i, $i \in I$ are defined on a common probability space. Then X_i also has constant rate α considered as a random variable taking values in S (although, of course, the support of X_i is a subset of S_i). If we extend the upper probability function F_i and density f_i of X_i to all of S, then for $x \in S_i$, $f_i(x) = \alpha F_i(x)$ while for $x \notin S_i$, $F_i(x) = 0$ and $f_i(x) = 0$. Now let U take values in I, independent of $(X_i : i \in i)$. Define $Y = X_U$, so that Y takes values in S. Then Y has constant rate α on S.

This construction leads to an interpretation of mixtures generally, in the setting of Theorem 13. Let S_i be a copy of S for each $i \in I$, where S_i are disjoint. Modify X_i so that it takes values in S_i, and we are in the setting of this section. Thus, we can think of random variable U as selecting the copy S_i.

Puri and Rubin [25] have shown that in $([0, \infty)^k, +)$, the only distributions with constant rate with respect to Lebesgue measure are mixtures of exponential distributions. This result does not hold for general posets.

Suppose that $(T, \preceq)$ is a standard discrete poset with minimal element u. Suppose there exist subsets R and S of T with the property that $R \cup S = T$, $R \cap S = \{u\}$ and for $x \in R - \{u\}$ and $y \in S - \{u\}$, $x \parallel y$. Conversely, given disjoint standard discrete posets $(R, \preceq)$ and $(S, \preceq)$ with minimal elements a and b, respectively, we could create a new poset $(T, \preceq)$ by joining a and b to create a new vertex u.

Suppose now that F is the upper probability function of a distribution with constant rate α on $(R, \preceq)$ and that G is the upper probability function with rate α on $(S, \preceq)$. We can extend F to T by $F(x) = 0$ for $x \in T - R$, and similarly, we can extend G to T by $G(x) = 0$ for $x \in T - S$.

Proposition 47. *F and G are the upper probability functions of distributions with constant rate α on T (but of course, not support T in general).*

Proof. We prove the result for F; the proof for G is identical. First

$$\sum_{x \in T} \alpha F(x) = \sum_{x \in R} \alpha F(x) = 1$$

Next, for $x \in R$,

$$\sum_{y \in I[x]} \alpha F(y) = \sum_{y \in I[x] \cap R} \alpha F(y) = F(x)$$

while for $x \in S - \{u\}$,

$$\sum_{y \in I[x]} \alpha F(y) = \sum_{y \in I[x] \cap S} \alpha F(y) = 0 = F(x)$$

∎

It now follows from our general Theorem 13 that for $p \in (0, 1)$, the mixture distribution with upper probability function H given by

$$H(x) = pF(x) + (1 - p)G(x), \quad x \in T$$

has constant rate α on T. In particular, if $(R, \preceq)$ and $(S, \preceq)$ support constant rate distributions, then so does $(T, \preceq)$ (and with an additional free parameter, namely p).

5.6 Special constructions

In this section we describe two special constructions for a standard discrete poset $(S, \preceq)$.

For the first construction, we fix $v \in S$ and then add a new element u so that v covers u. Thus, we have constructed a new poset $(T, \preceq)$ where $T = S \cup \{u\}$. The new partial order is defined as follows: for $x, y \in S$, $x \preceq y$ in T if and only if $x \preceq y$ in S. For $x \in S$, $x \succ u$ in T if and only if $x \succeq v$ in S. Finally, no $x \in S$ satisfies $x \prec u$ in T. Thus, u is minimal in T.

Suppose now that F is the upper probability funciton of a distribution with constant rate α on $(S, \preceq)$. Define G on T by

$$\begin{aligned} G(x) &= \frac{1-\alpha}{1-\alpha+\alpha F(v)} F(x), \quad x \in S \\ G(u) &= \frac{F(v)}{1-\alpha+\alpha F(v)} \end{aligned}$$

Theorem 14. The function G is the upper probability function of a distribution with constant rate α on $(T, \preceq)$.

Proof. Let $C = (1-\alpha)/(1-\alpha+\alpha F(v))$. For $x \in S$,

$$\sum_{y \succeq x} \alpha G(y) = C \sum_{y \succeq x} \alpha F(y) = CF(x) = G(x)$$

Also,

$$\begin{aligned} \sum_{y \succeq u} \alpha G(y) &= \alpha G(u) + C \sum_{y \succeq v} \alpha F(y) = \alpha G(u) + CF(v) \\ &= \frac{\alpha F(v)}{1-\alpha+\alpha F(v)} + \frac{(1-\alpha)F(v)}{1-\alpha+\alpha F(v)} \\ &= \frac{F(v)}{1-\alpha+\alpha F(v)} = G(u) \end{aligned}$$

Finally,

$$\begin{aligned} \sum_{x \in T} \alpha G(x) &= \alpha G(u) + C \sum_{x \in S} \alpha F(x) = \alpha G(u) + C \\ &= \frac{\alpha F(v)}{1-\alpha+\alpha F(v)} + \frac{1-\alpha}{1-\alpha+\alpha F(v)} = 1 \end{aligned}$$

∎

Our next construction requires a definition. A point $u \in S$ is a *chain point* if for every $x \in S$, either $x \preceq u$ or $u \preceq x$. For a fixed chain point $u \in S$ we can split u into two new vertices v and w, to create a new set T. We define the partial order on T as follows:

1. For $x \in S$, if x covers u in S then x covers v and x covers w in T.
2. For $x \in S$, if u covers x in S then v covers x and w covers x in T.
3. For $x, y \in S - \{u\}$, if y covers x in S then y covers x in T.

Theorem 15. Suppose that $(S, \preceq)$ is a standard discrete poset with a chain point u. Suppose that F is the upper probability function of a distribution with constant rate α on S. Let $(T, \preceq)$ be the poset obtained by splitting S at the chain point u (into new points v and w). Define the function G on T by

$$G(x) = \begin{cases} \frac{1}{1+\alpha} F(u) & \text{if } x \in \{v, w\} \\ \frac{1}{1+\alpha} F(x) & \text{if } x \succ u \text{ in } S \\ F(x) & \text{if } x \prec u \text{ in } S \end{cases}$$

Then G is the upper probability function of a distribution with constant rate α on T.

Proof. In the definition of G note that $x \succ u$ in S if and only if $x \succ v$ in T and $x \succ w$ in T. Similarly $x \prec u$ in S if and only if $x \prec v$ in T and $x \prec w$ in T. Also, since u is a chain point in S, $x \in T$ if and only if $x \in \{v, w\}$ or $x \prec u$ in S or $x \succ u$ in S and these statements are mutually exclusive.

First we show that αG is a probability density function.

$$\begin{aligned} \sum_{x \in T} \alpha G(x) &= \alpha G(v) + \alpha G(w) + \sum_{x \succ u} \alpha G(x) + \sum_{x \prec u} \alpha G(x) \\ &= \frac{\alpha}{1+\alpha} F(v) + \frac{\alpha}{1+\alpha} F(w) + \sum_{x \succ u} \frac{\alpha}{1+\alpha} F(x) + \sum_{x \prec u} \alpha F(x) \end{aligned}$$

But since F is the upper probability function of a constant rate distribution on S, and since u is a chain point,

$$\sum_{x \prec u} \alpha F(x) = 1 - \sum_{x \succeq u} \alpha F(x) = 1 - F(u)$$

Also,

$$\begin{aligned} \sum_{x \succ u} \frac{\alpha}{1+\alpha} F(x) &= \frac{1}{1+\alpha} \left(\sum_{x \succeq u} \alpha F(x) - \alpha F(u) \right) \\ &= \frac{1}{1+\alpha} [F(u) - \alpha F(u)] = \frac{1-\alpha}{1+\alpha} F(u) \end{aligned} \tag{12}$$

Substituting gives

$$\sum_{x \in T} \alpha G(x) = \frac{2\alpha}{1+\alpha} F(u) + \frac{1-\alpha}{1+\alpha} F(u) + 1 - F(u) = 1$$

so αG is a probability density function on T.

Finally, we show the constant rate property. First

$$\begin{aligned}
\sum_{y \succeq_T v} \alpha G(y) &= \alpha G(v) + \sum_{y \succ u} \alpha G(y) \\
&= \frac{\alpha}{1+\alpha} F(u) + \sum_{y \succ u} \frac{\alpha}{1+\alpha} F(y) \\
&= \frac{1}{1+\alpha} \sum_{y \succeq u} \alpha F(y) = \frac{1}{1+\alpha} F(u) = G(v)
\end{aligned}$$

Similarly,

$$\sum_{y \succeq_T w} \alpha G(y) = G(w)$$

Next, suppose $x \succ v$ in T (so that $x \succ u$ in S). Then

$$\begin{aligned}
\sum_{y \succeq_T x} \alpha G(y) &= \sum_{y \succeq x} \alpha G(y) = \sum_{y \succeq x} \frac{\alpha}{1+\alpha} F(y) \\
&= \frac{1}{1+\alpha} \sum_{y \succeq x} \alpha F(y) = \frac{1}{1+\alpha} F(x) = G(x)
\end{aligned}$$

Finally, suppose that $x \prec u$ (so that $x \prec_T v$ and $x \prec_T w$. Then

$$\begin{aligned}
\sum_{y \succeq_T x} \alpha G(y) &= \sum_{x \preceq y \prec u} \alpha G(y) + \alpha G(v) + \alpha G(w) + \sum_{y \succ u} \alpha G(y) \\
&= \sum_{x \preceq y \prec u} \alpha F(y) + \frac{\alpha}{1+\alpha} F(u) + \frac{\alpha}{1+\alpha} F(u) + \sum_{y \succ u} \frac{\alpha}{1+\alpha} F(y) \\
&= \sum_{y \succeq x} \alpha F(y) - \sum_{y \succeq u} \alpha F(y) + 2\frac{\alpha}{1+\alpha} F(u) + \frac{1-\alpha}{1+\alpha} F(u) \\
&= F(x) - F(u) + F(u) = F(x)
\end{aligned}$$

Note that we used (12) in the next to the last line. □

5.7 Joint distributions

In this section, suppose that $(S, \preceq_S, \mu)$ and $(T, \preceq_T, \nu)$ are standard posets, and that $(S \times T, \preceq, \lambda)$ is the direct product. Suppose that X and Y are random variables taking values in S and T, respectively, so that (X, Y) takes values in $S \times T$. Let F denote the upper probability function of X, G the upper probability function of Y, and H the upper probability function of (X, Y).

Theorem 16. If (X, Y) has constant rate γ and X and Y have independent upper events, then X and Y are completely independent and there exist $\alpha > 0$ and $\beta > 0$ with $\alpha\beta = \gamma$ such that X has constant rate α and Y has constant rate β. Conversely, if X and Y are independent and have constant rates α and β, respectively, then (X, Y) has constant rate $\gamma = \alpha\beta$.

Proof. Suppose that (X, Y) has constant rate γ with respect to λ and that X and Y have independent upper events. Then $H(x, y) = F(x)G(y)$ for $x, \in S$, $y \in T$, and hence (X, Y) has density h with respect to λ given by $h(x, y) = \gamma H(x, y) = \gamma F(x)G(y)$ for $x \in S$, $y \in T$. It follows from the standard factorization theorem that X and Y are independent. Moreover, there exists α and β such that X has density f with respect to μ given by $f(x) = \alpha F(x)$; Y has density g with respect to ν given by $g(x) = \beta G(x)$; and $\alpha\beta = \gamma$.

Conversely, suppose that X and Y are independent and have constant rates α and β with respect to μ and ν, respectively. Then X has density $f = \alpha F$ with respect to μ and Y has density $g = \beta G$ with respect to ν, and by independence, (X, Y) has density h with respect to λ given by

$$h(x, y) = f(x)g(y) = \alpha F(x)\beta G(y) = \alpha\beta F(x)G(y) = \alpha\beta H(x, y), \ (x, y) \in S \times T$$

Hence (X, Y) has constant rate $\alpha\beta$ with respect to λ. □

In the second part of Theorem 16, if we just know that X and Y have constant rates and have independent upper events, then we can conclude that H is the upper probability function of a distribution with constant rate on $S \times T$. and that this distribution is the joint distribution of independent random variables. However, we cannot conclude that this distribution is the distribution of our given variables (X, Y), unless we know that upper probability functions on $S \times T$ completely determine distributions (that is, unless $(S \times T, \preceq)$ has distributional dimension 1.

In general, there will be lots of constant rate distributions on $(S \times T, \preceq)$ that do not fit Theorem 16.

Example 10. Suppose that $(S \times T)$ is a discrete poset and that S and T have minimum elements e and ϵ, respectivley. Suppose that (X, Y) is a random variable taking values in $S \times T$ with constant rate α. As usual, let F, G, and H denote the upper probability functions of X, Y, and (X, Y), respectivley, and let f, g, and h denote the corresponding density functions. The upper probability function of X is $F(x) = H(x, \epsilon)$ while the density function is

$$f(x) = \sum_{y \in T} h(x, y) = \sum_{y \in T} \alpha H(x, y)$$

It certainly seems possible that (X, Y) could have constant rate without X and Y having constant rate.

In particular, we can typically construct mixtures of constant rate distributions (Section 5.5) that do not satisfy the setting of Theorem 16.

Example 11. In the simplest example, suppose that (X_1, Y_1) and (X_2, Y_2) are random variables taking values in the discrete poset $(S \times T, \preceq)$. We assume that (X_1, Y_1) and (X_2, Y_2) are independent and both have constant rate γ. Let U be an indicator variable independent of (X_1, Y_1) and (X_2, Y_2) and let

$$(X, Y) = U(X_1, Y_1) + (1 - U)(X_2, Y_2) = (UX_1 + (1 - U)X_2, UY_1 + (1 - U)Y_2)$$

Then (X, Y) has the mixture distribution and hence it also has constant rate γ, although clearly X and Y will not be independent in general and may not have constant rate. Speciallizing further, suppose X_1 and Y_1 are independent with constant rates α_1, β_1 respectively, with $\alpha_1\beta_1 = \gamma$ and that similarly X_2 and Y_2 are independent with constant rates α_2, β_2 respectively, with $\alpha_2\beta_2 = \gamma$. Then X and Y are still not independent and would still not have constant rate, unless $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$.

The results of this section extend in an obvious way to the direct product of a finite number of posets $(S_1, \preceq_1, \mu_1), (S_2, \preceq_2, \mu_2), \ldots, (S_n, \preceq_n, \mu_n)$. On the other hand, given a countably infinite collection of posets $((S_n, \preceq_n, \mu_n) : n \in \mathbb{N}_+)$, where S_i has minimum element e_i for each i, define

$$T = \{(x_1, x_2, \ldots) : x_i \in S_i \text{ for all } i \text{ and } x_i = e_i \text{ eventually in } i\}$$

Using component-wise constructions, the result of this section extend to the poset $(T, \preceq, \lambda)$.

5.8 Lexicographic sums

As in Section 3.6, suppose that $(R, \preceq_R, \mu)$ is a standard poset and that for each $x \in R$, $(S_x, \preceq_x, \nu_x)$ is standard poset. We let $(T, \preceq, \lambda)$ denote the lexicographic sum of $(S_x, \preceq_x)$ over $x \in R$.

Suppose now that (X, Y) is a random variable taking values in T. Then X takes values in R, and given $X = x \in R$, Y takes values in S_x. Unconditionally, Y takes values in $\cup_{x \in R} S_x$. We will let F denote the upper probability function of X and f the density function of X (relative to μ). For $x \in R$, we will let G_x and g_x denote the upper probability function and density (with respect to ν_x) of the conditional distribution of Y given $X = x$. Finally, we will let H and h denote the upper probability function and density (with respect to λ) of (X, Y). Then

$$h(x, y) = f(x) g_x(y), \quad (x, y) \in T$$

Also

$$\begin{aligned} H(x, y) &= \mathbb{P}[(X, Y) \succeq (x, y)] \\ &= \mathbb{P}[X \succ_S x \text{ or } (X = x \text{ and } Y \succeq_x y)] \\ &= \mathbb{P}(X \succ_S x) + \mathbb{P}(X = x, Y \succeq_x y) \\ &= F(x) - \mu\{x\} f(x) + \mu\{x\} f(x) G_x(y) \\ &= F(x) - \mu\{x\} f(x)[1 - G_x(y)], \quad (x, y) \in T \end{aligned}$$

There are a couple of obvious special cases. If X has a continuous distribution, so that $\mu(\{x\}) = 0$, then

$$H(x, y) = F(x), \quad x \in R, \, y \in S_x$$

If X has a discrete distribution, so that μ is counting measure, then

$$H(x, y) = F(x) - f(x) + f(x) G_x(y) = F(x) - f(x)[1 - G_x(y)], \quad (x, y) \in U$$

Suppose now that X has constant rate α on R, so that $f = \alpha F$. Then (X, Y) has density h given by

$$h(x, y) = \alpha F(x) g_x(y), (x, y) \in T$$

and upper probability function H given by

$$\begin{aligned} H(x, y) &= F(x) 1 - \alpha\mu\{x\} F(x) + \alpha\mu\{x\} F(x) G_x(y) \\ &= F(x)[1 - \alpha\mu\{x\} + \alpha\mu\{x\} G_x(y)], \quad (x, y) \in T \end{aligned}$$

If X has a continuous distribution (and constant rate α), then (X, Y) will have constant rate β if

$$\alpha F(x) g_x(y) = \beta F(x), \quad (x, y) \in T$$

or equivalently we need $g_x(y)$ constant in $(x, y) \in T$. This can only happen if $\nu_x(S_x)$ is constant in x and the conditional distribution of Y given $X = x$ is uniform on S_x. In the special case that $S_x = S$ for each x, the condition for (X, Y) to have constant rate is that X have constant rate α, X and Y be independent, and Y uniformly distributed on S.

If X has a discrete distribution (and constant rate α), then (X, Y) will have constant rate β if

$$\alpha F(x) g_x(y) = \beta F(x)[1 - \alpha + \alpha G_x(y)], \quad (x, y) \in T$$

or equivalently,

$$g_x(y) = \frac{\beta}{\alpha}[1 - \alpha + \alpha G_x(y)], \quad (x, y) \in T \tag{13}$$

That is, we need $g_x(y)/[1 - \alpha + \alpha G_x(y)]$ constant in $(x, y) \in T$. In the discrete case, when S_x is finite for each x, so we might be able to satisfy (13) by working backwards from the maximal elements of S_x.

5.9 Pairs of independent variables

First we consider independent random variables X and Y taking values in a standard poset $(S, \preceq, \lambda)$. Let F and G denote the upper probability functions of X and Y, respectively.

Proposition 48. Suppose that X has constant rate α and let

$$\frac{1}{\gamma} = \int_S F(x) G(x) d\lambda(x)$$

Then $\mathbb{P}(X \preceq Y) = \alpha/\gamma$ and the conditional density of X given $X \preceq Y$ is $\gamma F G$.

Proof. By assumption, $f = \alpha F$ is a density of X with respect to λ and therefore

$$\mathbb{P}(X \preceq Y) = \mathbb{E}[\mathbb{P}(X \preceq Y \mid X)] = \mathbb{E}[G(X)] = \int_S \alpha F(x) G(x) d\lambda(x) = \frac{\alpha}{\gamma}$$

Next, if $A \in \mathcal{B}(S)$, then

$$\begin{aligned}\mathbb{P}(X \in A, X \preceq Y) &= \mathbb{E}[\mathbb{P}(X \in A, X \preceq Y \mid X)] \\ &= \mathbb{E}[G(X), X \in A] = \int_A G(x)\alpha F(x) d\lambda(x)\end{aligned}$$

and therefore

$$\mathbb{P}(X \in A \mid X \preceq Y) = \int_A \gamma F(x) G(x) d\lambda(x)$$

so the conditional density of X given $X \preceq Y$ is γFG. $\square$

Problem 12. Find conditions on X and Y so that the conditional distribution of X given $X \preceq Y$ has constant rate. That is, find conditions on X and Y so that the conditional distribution of X given $X \preceq Y$ has upper probability function FG.

If Y has constant rate β, then $\mathbb{P}(Y \preceq X) = \beta/\gamma$. If X has constant rate α and Y has constant rate β and $\mathbb{P}(X = Y) = 0$ (which would be the case if X or Y have a continuous distribution) then

$$\mathbb{P}(X \parallel Y) = 1 - \frac{\alpha + \beta}{\gamma}$$

On the other hand, suppose that S is discrete (with counting measure as the reference measure, of course). Suppose that X has constant rate α and that Y has constant rate β. As above let

$$\gamma = \sum_{x \in S} F(x) G(x)$$

Then as above, $\mathbb{P}(X \preceq Y) = \alpha/\gamma$ and $\mathbb{P}(Y \preceq X) = \beta/\gamma$. Moreover,

$$\mathbb{P}(X = Y) = \mathbb{E}[\mathbb{P}(X = Y \mid X)] = \mathbb{E}[\beta G(X)] = \alpha\beta \sum_{x \in S} F(x) G(x) = \frac{\alpha\beta}{\gamma}$$

Therefore

$$\begin{aligned}\mathbb{P}(X \prec Y) &= \frac{\alpha(1-\beta)}{\gamma}, \quad \mathbb{P}(Y \prec X) = \frac{\beta(1-\alpha)}{\gamma} \\ \mathbb{P}(X \perp Y) &= \frac{\alpha + \beta - \alpha\beta}{\gamma}, \quad \mathbb{P}(X \parallel Y) = \frac{\gamma - \alpha + \beta - \alpha\beta}{\gamma}\end{aligned}$$

Problem 13. Recall that if $(S, \preceq)$ is a lower semi-lattice, then the upper probability function of $X \wedge y$ is FG. Find conditions on X and Y so that $X \wedge Y$ has constant rate.

5.10 Gamma distributions

As usual, assume that $(S, \preceq, \lambda)$ is a standard poset. Recall the cumulative functions λ_n, $n \in \mathbb{N}$ discussed in Section 2.2 and the discussion of ladder variables in Section 4.7.

Suppose that F is the upper probability function of a distribution on S with constant rate α. Let $\boldsymbol{X} = (X_1, X_2, \ldots)$ be a sequence of independent, identically distributed random variables with this distribution, and let $\boldsymbol{Y} = (Y_1, Y_2, \ldots)$ be the corresponding sequence of ladder variables. The following results follow easily from the general results in Section 4.7.

Proposition 49. $\boldsymbol{Y}$ is a homogeneous Markov chain on S with transition probability density g given by

$$g(z|y) = \alpha \frac{F(z)}{F(y)}, \quad y \in S, \ z \in I[y]$$

Proof. From Proposition 43, $\boldsymbol{Y}$ is a homogeneous Markov chaing with transition probability g given by

$$g(y, z) = \frac{f(z)}{F(y)}, \quad y \in S, \ z \in I[y]$$

so if X has constant rate α,

$$g(y, z) = \alpha \frac{F(z)}{F(y)}, \quad y \in S, \ z \in I[y]$$

□

Proposition 50. $(Y_1, Y_2, \ldots, Y_n)$ has probability density function h_n given by

$$h_n(y_1, y_2, \ldots, y_n) = \alpha^n F(y_n), \quad (y_1, y_2, \ldots, y_n) \in D_n$$

Proof. From Corollary 9, $(Y_1, Y_2, \ldots, Y_n)$ has probability density function h_n given by

$$h_n(y_1, y_2, \ldots, y_n) = r(y_1) f(y_2) \cdots r(y_{n-1}) f(y_n), \quad (y_1, y_2, \ldots, y_n) \in D_n$$

where r is the rate function. Hence, if X has constant rate α,

$$h_n(y_1, y_2, \ldots, y_n) = \alpha^n F(y_n), \quad (y_1, y_2, \ldots, y_n) \in D_n$$

□

Proposition 51. Y_n has probability density function g_n given by

$$g_n(y) = \alpha^n \lambda_{n-1}(y) F(y), \quad y \in S$$

Proof. From Corollary 10, Y_n has probability density function g_n given by

$$g_n(y) = f(y) \int_{D_{n-1}[y]} r(y_1)r(y_2)\cdots r(y_{n-1})d\lambda^{n-1}(y_1, y_2, \dots y_{n-1}), \quad y \in S$$

Hence, if X has constant rate α,

$$g_n(y) = \alpha^n \lambda_{n-1}(y)F(y), \quad y \in S$$

$\square$

Proposition 52. The conditional distribution of $(Y_1, Y_2, \dots, Y_{n-1})$ given $Y_n = y$ is uniform on $D_{n-1}[y]$.

Proof. By Corollary 11, $(Y_1, Y_2, \dots, Y_{n-1})$ given $Y_n = y$ has probability density function given by

$$h_{n-1}(y_1, \dots, y_{n-1}|y) = \frac{1}{C_{n-1}[y]} r(y_1)\cdots r(y_{n-1}), \quad (y_1, \dots, y_{n-1}) \in D_{n-1}[y]$$

where $C_{n-1}[y]$ is the normalizing constant. Hence, if X has constant rate, $h_{n-1}(\cdot|y)$ is constant on $D_{n-1}[y]$ (and then of course the constant is $1/\lambda_{n-1}(y)$). $\square$

We will refer to the distribution with density g_n as the *gamma distribution* of order n associated with the given constant rate distribution. When $\boldsymbol{X}$ is an IID sequence with constant rate α, the sequence of ladder variables $\boldsymbol{Y}$ is analogous to the arrival times in the ordinary Poisson process. By Proposition 52, this process defines the most random way to put a sequence of ordered points in S. The conversse of Proposition 52 is not quite true.

Example 12. Consider the poset $(S, \preceq)$ that consists of two parallel chains. That is,

$$S = \{a_0, a_1, \dots\} \cup \{b_0, b_1, \dots\}$$

where $a_0 \prec a_1 \prec \cdots$, $b_0 \prec b_1 \prec \cdots$, and $a_i \parallel b_j$ for all i, j. Let $f : S \to (0,1)$ be defined by

$$f(a_n) = p\alpha(1-\alpha)^n, \ f(b_n) = (1-p)\beta(1-\beta)^n; \quad n \in \mathbb{N}$$

where $\alpha, \beta, p \in (0,1)$ and $\alpha \neq \beta$. It's easy to see that f is a PDF with corresponding UPF given by

$$F(a_n) = p(1-\alpha)^n, \ F(b_n) = (1-p)(1-\beta)^n; \quad n \in \mathbb{N}$$

Hence the rate function r is

$$r(a_n) = \alpha, \ r(b_n) = \beta; \quad n \in \mathbb{N}$$

if $\boldsymbol{X}$ is an IID sequence with PDF f and $\boldsymbol{Y}$ the corresponding sequence of ladder variables then from Proposition 52, the conditional distribution of $(Y_1, \dots, Y_{n-1})$ given $Y_n = y$ is uniform for each $y \in S$, yet X does not (quite) have constant rate.

Proposition 53. Suppose that $(S, \preceq \lambda)$ is a standard, connected poset, and that $\boldsymbol{Y}$ is the sequence of ladder variables associated with a sequence of IID variables $\boldsymbol{X}$. If the conditional distribution of Y_1 given $Y_2 = y$ is uniform on $D[y]$, then X has a distribution with constant rate.

Proof. The conditional distribution of Y_1 given $Y_2 = y$ is

$$h_1(y_1|y) = \frac{1}{C_1[y]} r(y_1), \quad y_1 \in D[y]$$

But this is constant in $y_1 \in D[y]$ and hence r is constant on $D[y]$ for each $y \in S$. Thus, it follows that $r(x) = r(y)$ whenever $x \perp y$. Since S is connected, for any $x, y \in S$, there exists a finite squence $(x_0, x_1, \ldots, x_n)$ such that $x = x_0 \perp x_1 \cdots \perp x_n = y$. It then follows that $r(x) = r(y)$. ☐

Problem 14. Suppose that $\boldsymbol{Y} = (Y_1, Y_2, \ldots)$ is an increasing sequence of random variables in S with the property that the conditional distribution of $(Y_1, Y_2, \ldots, Y_{n-1})$ given $Y_n = y$ is uniform on $D_{n-1}[y]$ for every $n \in \mathbb{N}_+$ and $y \in S$. What, if anything, can we say about the finite dimensional distributions of $\boldsymbol{Y}$? In particular, under what additional conditions can we conclude that $\boldsymbol{Y}$ is the sequence of ladder variables associated with an IID sequence $\boldsymbol{X}$ with constant rate?

5.11 The point process

Let F be the upper probability function of a distribution with constant rate α on a standard poset $(S, \preceq, \lambda)$. Let $\boldsymbol{X} = (X_1, X_2, \ldots)$ be a sequence of independent random variables, each with the constant rate distribution, and let $\boldsymbol{Y} = (Y_1, Y_2, \ldots)$ be the corresponding sequence of ladder variables as in Section 5.10. Recall the basic notation and results from Section 4.8 on point processes. Thus, for $x \in S$, let

$$N_x = \#\{n \in \mathbb{N}_+ : Y_n \preceq x\}$$

That is, N_x is the number of random points in $D[x]$. Then $Y_n \preceq x$ if and only if $N_x \geq n$ for $n \in \mathbb{N}_+$ and $x \in S$. For $n \in \mathbb{N}_+$, let G_n denote the *lower* probability function of Y_n, so that

$$G_n(x) = \mathbb{P}(Y_n \preceq x), \quad x \in S$$

For $x \in S$,

$$\mathbb{P}(N_x \geq n) = \mathbb{P}(Y_n \preceq x) = G_n(x), \quad n \in \mathbb{N}_+$$

so that for fixed $x \in S$, $n \mapsto G_n(x)$ is the *upper* probability function of N_x. Recall also that for $x \in S$:

$$\mathbb{P}(N_x = n) = G_n(x) - G_{n+1}(x), \quad n \in \mathbb{N}$$

$$\mathbb{E}(N_x) = \sum_{n=1}^{\infty} G_n(x)$$

Now, using the probability density function of Y_n, we have

$$\begin{aligned} G_n(x) &= \int_{D[x]} \alpha^n \lambda_{n-1}(t) F(t) d\lambda(t) = \int_{D[x]} \alpha^{n-1} \lambda_{n-1}(t) f(t) d\lambda(t) \\ &= \mathbb{E}[\alpha^{n-1} \lambda_{n-1}(X), X \preceq x], \quad n \in \mathbb{N}_+, \ x \in S \end{aligned}$$

where X is a variable with the constant rate distribution (that is, X has probability density function $f = \alpha F$). Thus for fixed $x \in S$, the upper probability function of N_x is given by

$$\begin{aligned} &\mathbb{P}(N_x \geq 0) = 1 \\ &\mathbb{P}(N_x \geq n) = \alpha^{n-1} \mathbb{E}[\lambda_{n-1}(X), X \preceq x], \quad n \in \mathbb{N}_+ \end{aligned}$$

while the probability density function is given by

$$\begin{aligned} &\mathbb{P}(N_x = 0) = 1 - \mathbb{P}(X \preceq x) \\ &\mathbb{P}(N_x = n) = \alpha^{n-1} \mathbb{E}[\lambda_{n-1}(X) - \alpha \lambda_n(X), X \preceq x], \quad n \in \mathbb{N}_+ \end{aligned}$$

Also,

$$\begin{aligned} \mathbb{E}(N_x) = &\sum_{n=1}^{\infty} \mathbb{E}[\alpha^{n-1} \lambda_{n-1}(X), X \preceq x] \\ &\mathbb{E}\left[\sum_{n=1}^{\infty} \alpha^{n-1} \lambda_{n-1}(X), X \preceq x\right] = \mathbb{E}[\Lambda(X, \alpha), X \preceq x] \end{aligned}$$

where, we recall, the generating function associated with λ is

$$\Lambda(x, t) = \sum_{n=0}^{\infty} t^n \lambda_n(x)$$

The function $m : S \to [0, \infty)$ defined by

$$m(x) = \mathbb{E}(N_x) = \sum_{n=1}^{\infty} G_n(x) = \mathbb{E}[\Lambda(X, \alpha), X \preceq x]$$

is the *renewal function*.

Our next topic is *thinning* the point process. As above, suppose that $\boldsymbol{Y} = (Y_1, Y_2, \ldots)$ is the sequence of gamma variables corresponding to a distribution with constant rate α and upper probability function F. Let N have the geometric distribution on $\mathbb{N}_+$ with rate parameter $r \in (0, 1)$ so that

$$\mathbb{P}(N = n) = r(1 - r)^{n-1}, \quad n \in \mathbb{N}_+$$

Moreover, we assume that N and $\boldsymbol{Y}$ are independent. The basic idea is that we *accept* a point with probability r and *reject* the point with probability $1 - r$, so that Y_N is the first point accepted. We are interested in the distribution of Y_N.

Theorem 17. The probability denstiy function g of Y_N is is given by

$$g(x) = r\alpha\Lambda[x, \alpha(1-r)]F(x), \quad x \in S$$

where again, Λ is the generating function associated with $(\lambda_n : n \in \mathbb{N})$.

Proof. For $x \in S$,

$$\begin{aligned} g(x) = \mathbb{E}[f_N(x)] &= \sum_{n=1}^{\infty} r(1-r)^{n-1} f_n(x) \\ &= \sum_{n=1}^{\infty} r(1-r)^{n-1}\alpha^n \lambda_{n-1}(x)F(x) \\ &= \alpha r F(x) \sum_{n=1}^{\infty} [\alpha(1-r)]^{n-1}\lambda_{n-1}(x) \\ &= \alpha r F(x)\Lambda[x, \alpha(1-r)] \end{aligned}$$

□

In general, Y_N does not have constant rate.

6 Memoryless and Exponential Distributions

In this chapter, we assume that $(S, \cdot, \lambda)$ is a standard positive semigroup. Because of the algebraic, topological, and measure-theoretic assumptions, this is the natural home for exponential properties, aging properties, and related concepts.

6.1 Basic definitions

Suppose that X is a random variable taking values in S. Most characterizations of the exponential distribution (and its generalizations) in Euclidian spaces are based on the equivalence of the residual distributions with the original distribution, in some sense. In the setting of a positive semigroup, such properties take the form

$$\mathbb{P}(x^{-1}X \in A \mid X \succeq x) = \mathbb{P}(X \in A),$$

or equivalently,

$$\mathbb{P}(X \in xA) = \mathbb{P}(X \succeq x)\mathbb{P}(X \in A) \tag{14}$$

for certain $x \in S$ and $A \in \mathcal{B}(S)$.

Definition 37. We will say that X has an *exponential distribution* if (14) holds for all $x \in S$ and $A \in \mathcal{B}(S)$. We will say that X has a *memoryless distribution* if (14) holds for all $x \in S$ and all A of the form yS, $y \in S$.

Note 35. Thus, if X has a memoryless distribution then

$$\mathbb{P}(X \succeq xy) = \mathbb{P}(X \succeq x)\mathbb{P}(X \succeq y), \quad x, y \in S$$

Equivalently, if F is the upper probability function of X then

$$F(xy) = F(x)F(y), \quad x, y \in S$$

A random variable with an exponential distribution has the property that the conditional distribution of $x^{-1}X$ given $X \succeq x$ is the same as the distribution of X. If X has only a memoryless distribution, then the conditional upper probability function of $x^{-1}X$ given $X \succeq x$ is the same as the upper probability function of X. An exponential distribution is necessarily memoryless, but as we will see, a memoryless distribution may not be exponential. On the other hand, if the distributional dimension of $(S, \cdot)$ is 1, that is if the upper probability function of a distribution uniquely determines the distribution, then a memoryless distribution is necessarily also an exponential distribution.

Note 36. Recall from Halmos [15] that a positive measure μ on a locally compact topological group S is said to be *relatively invariant* if

$$\mu(xA) = F(x)\mu(A), \quad x \in S, \ A \in \mathcal{B}(S)$$

for some function $F \colon S \to (0, \infty)$. Suppose that we replace group with positive semigroup in this definition, and let μ be a probability measure. Then letting $A = S$ gives $F(x) = \mu(xS)$, so that F is the upper probability function of the distribution. Thus an exponential distribution is simply a *relatively invariant probability measure.*

6.2 Invariant pairs

Definition 38. Suppose that X is a random variable taking values in S. If $x \in S$ and $A \in \mathcal{B}(S)$, we will call (x, A) an *invariant pair* for X if the exponential property holds for (x, A):

$$\mathbb{P}(X \in xA) = \mathbb{P}(X \succeq x)\mathbb{P}(X \in A)$$

Definition 39. If X is a random variable taking values in S, let

$$S_X = \{x \in S \colon \mathbb{P}(X \in xA) = \mathbb{P}(X \succeq x)\mathbb{P}(X \in A) \text{ for all } A \in \mathcal{B}(S)\}$$

That is, S_X consists of all $x \in S$ such that (x, A) is an invariant pair for X for all $A \in \mathcal{B}(S)$.

Proposition 54. If X is a random variable taking values in S, then S_X is a complete sub-semigroup of S.

Proof. Let $x, y \in S_X$ and let $A \in \mathcal{B}(S)$. Then

$$\begin{aligned}\mathbb{P}[X \in (xy)A] &= \mathbb{P}[X \in x(yA)] = \mathbb{P}(X \succeq x)\mathbb{P}(X \in yA) \\ &= \mathbb{P}(X \succeq x)\mathbb{P}(X \succeq y)\mathbb{P}(X \in A)\end{aligned}$$

In particular, letting $A = S$ we have

$$\mathbb{P}(X \succeq xy) = \mathbb{P}(X \succeq x)\mathbb{P}(X \succeq y)$$

Substituting back we have

$$\mathbb{P}[X \in (xy)A] = \mathbb{P}(X \succeq xy)\mathbb{P}(X \in A)$$

and hence $xy \in S_X$. Now suppose that $x, y \in S_X$ and $x \preceq y$. Let $u = x^{-1}y$ so that $xu = y$. Let $A \in \mathcal{B}(S)$. Since $x \in S_X$ we have

$$\mathbb{P}(X \in xuA) = \mathbb{P}(X \succeq x)\mathbb{P}(X \in uA)$$

On the other hand, since $y = xu \in S_X$ we have

$$\mathbb{P}(X \in xuA) = \mathbb{P}(X \succeq xu)\mathbb{P}(X \in A)$$

Again, since $x \in S_X$ we have

$$\mathbb{P}(X \succeq xu) = \mathbb{P}(X \in xuS) = \mathbb{P}(X \succeq x)\mathbb{P}(X \in uS) = \mathbb{P}(X \succeq x)\mathbb{P}(X \succeq u)$$

Substituting we have

$$\mathbb{P}(X \succeq x)\mathbb{P}(X \in uA) = \mathbb{P}(X \succeq x)\mathbb{P}(X \succeq u)\mathbb{P}(X \in A)$$

Since $\mathbb{P}(X \succeq x) > 0$, we have

$$\mathbb{P}(X \in uA) = \mathbb{P}(X \succeq u)\mathbb{P}(X \in A)$$

and therefore $u \in S_X$. □

Note 37. If $\mathbb{P}(X \in S_X) > 0$, then the conditional distribution of X given $X \in S_X$ is exponential on the positive semigroup $(S_X, \cdot)$. In particular, this holds if $(S, \cdot)$ is discrete. However, it may well happen that $S_X = \{e\}$. For any random variable X on a discrete positive semigroup, the conditional distribution of X given $X = e$ is exponential.

Definition 40. If X is a random variable taking values in S, let

$$\mathcal{B}_X(S) = \{A \in \mathcal{B}(S) \colon \mathbb{P}(X \in xA) = \mathbb{P}(X \succeq x)\mathbb{P}(X \in A) \text{ for all } x \in S\}$$

That is, $\mathcal{B}_X(S)$ consists of all $A \in \mathcal{B}(S)$ such that (x, A) is an invariant pair for X for all $x \in S$.

Proposition 55. $\mathcal{B}_X(S)$ is closed under countable disjoint unions, proper differences, countable increasing unions, and countable decreasing intersections (and hence is a monotone class).

Proof. Let $(A_1, A_2, \ldots)$ be a sequence of disjoint sets in $\mathcal{B}_X(S)$ and let $x \in S$. Then $(xA_1, xA_2, \ldots)$ is a disjoint sequence and

$$\begin{aligned}\mathbb{P}\left(X \in x\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(X \in \bigcup_{i=1}^{\infty} xA_i\right) &= \sum_{i=1}^{\infty} \mathbb{P}\left(X \in xA_i\right) \\ &= \sum_{i=1}^{\infty} \mathbb{P}(X \succeq x)\mathbb{P}\left(X \in A_i\right) \\ &= \mathbb{P}(X \succeq x)\sum_{i=1}^{\infty} \mathbb{P}\left(X \in A_i\right) \\ &= \mathbb{P}(X \succeq x)\mathbb{P}\left(X \in \bigcup_{i=1}^{\infty} A_i\right)\end{aligned}$$

Hence $\bigcup_{i=1}^{\infty} A_i \in \mathcal{B}_X(S)$. Let A, $B \in \mathcal{B}_X(S)$ with $A \subseteq B$ and let $x \in S$. Then $xA \subseteq xB$ so

$$\begin{aligned}\mathbb{P}[X \in x(B - A)] = \mathbb{P}[X \in (xB - xA)] &= \mathbb{P}(X \in xB) - \mathbb{P}(X \in xA) \\ &= \mathbb{P}(X \succeq x)\mathbb{P}(X \in A) - \mathbb{P}(X \succeq x)\mathbb{P}(X \in B) \\ &= \mathbb{P}(X \succeq x)[\mathbb{P}(X \in B) - \mathbb{P}(X \in A)] \\ &= \mathbb{P}(X \succeq x)\mathbb{P}[X \in (B - A)]\end{aligned}$$

and hence $B - A \in \mathcal{B}_X(S)$. Of course $S \in \mathcal{B}_X(S)$ and hence the other results follow. ☐

Proposition 56. Suppose that X is a random variable taking values in S. Then X has an exponential distribution on $(S, \cdot)$ if and only if $\mathcal{B}_X(S)$ contains the open sets if and only if $\mathcal{B}_X(S)$ contains the compact sets.

Proof. Of course if X has an exponential distribution, then by definition, $\mathcal{B}_X(S)$ contains all Borel sets. On the other hand, if $\mathcal{B}_X(S)$ contains the open sets, then it also contains the closed sets and hence the algebra of open and closed sets. By the monotone class theorem, $\mathcal{B}_X(S) = \mathcal{B}(S)$.

Similarly, if X has an exponential distribution, then $\mathcal{B}_X(S)$ contains the compact sets. Conversely, if $\mathcal{B}_X(S)$ contains the compact sets, then it contains the algebra of compact sets and their complements. By the monotone class theorem, $\mathcal{B}_X(S) = \mathcal{B}(S)$. ☐

6.3 Basic properties and characterizations

Proposition 57. Suppose that X takes values in S. Then X has an exponential distribution on $(S, \cdot)$ if and only if

$$\mathbb{E}[\varphi(x^{-1}X) \mid X \succeq x] = \mathbb{E}[\varphi(X)]$$

or equivalently

$$\mathbb{E}[\varphi(x^{-1}X), X \succeq x] = \mathbb{P}(X \succeq x)\mathbb{E}[\varphi(X)]$$

for every bounded, measurable $\varphi \colon S \to \mathbb{R}$.

Proof. Let $\varphi = \mathbf{1}_A$ where $A \in \mathcal{B}(S)$. Then

$$\mathbb{E}[\varphi(x^{-1}X \mid X \succeq x] = \mathbb{P}(x^{-1}X \in A \mid X \succeq x) = \mathbb{P}(X \in A) = \mathbb{E}[\varphi(X)]$$

That is, equation for the expected values is precisely the exponential property. The general result now follows by the usual methods. $\square$

Corollary 14. Suppose that X has an exponential distribution on $(S, \cdot)$ and that $\psi \colon S \to \mathbb{R}$ is bounded and measurable. Then

$$\mathbb{E}[\psi(X),\, X \in xA] = \mathbb{P}(X \succeq x)\mathbb{E}[\psi(xX),\, X \in A]$$

Proof. Let $\varphi(z) = \mathbf{1}_A(z)\psi(xz)$ and apply the previous theorem. $\square$

Lemma 2. Suppose that X has an exponential distribution on S with upper probability function F. Then F is continuous.

Proof. Suppose that x, $x_n \in S$ for $n \in \mathbb{N}_+$ and $x_n \to x$ as $n \to \infty$. Let y, $z \in S$ satisfy $x \prec y \prec z$. Then $[e, y]$ is a neighborhood of x so there exists $N \in \mathbb{N}_+$ such that $x_n \in [e, y]$ for $n \geq N$. Also yS is a neighborhood of z so there exists a compact neighborhood K of z and an open neighborhood U of z with $K \subseteq U \subseteq yS$. Since S is locally compact and Hausdorff, there exists a continuous function $\varphi \colon S \to [0, 1]$ such that $\varphi = 1$ on K and $\varphi = 0$ on U^c. It follows that for $n \geq N$,

$$\mathbb{E}[\varphi(X), X \succeq x_n] = \mathbb{E}[\varphi(X), X \succeq x] = \mathbb{E}[\varphi(X), X \in U]$$

Next note that for $n \geq N$, $\mathbb{E}[\varphi(x_n X)] > 0$ and $\mathbb{E}[\varphi(xX)] > 0$ since $\varphi(x_n t) = 1$ for $t \in x_n^{-1}K$, and $\varphi(xt) = 1$ for $t \in x^{-1}K$ and these sets have nonempty interior. By continuity and bounded convergence,

$$\mathbb{E}[\varphi(x_n X)] \to \mathbb{E}[\varphi(xX)] \text{ as } n \to \infty$$

From the exponential property (see Corollary 1),

$$\mathbb{E}[\varphi(X); X \succeq x] = F(x)\mathbb{E}[\varphi(xX)]$$

But as noted above, $\mathbb{E}[\varphi(X), X \succeq x] = \mathbb{E}[\varphi(X), X \in U]$ and therefore we have

$$F(x) = \frac{\mathbb{E}[\varphi(X), X \in U]}{\mathbb{E}[(\varphi(xX)]}$$

Similarly,

$$F(x_n) = \frac{\mathbb{E}[\varphi(X), X \in U]}{\mathbb{E}[\varphi(x_n X)]}$$

Taking limits we have $F(x_n) \to F(x)$ as $n \to \infty$ and hence F is continuous. Also, since xS has nonempty interior and X has support S, $F(x) > 0$ for each x. $\square$

Corollary 15. Suppose that X has an exponential distribution on S. Then $x \mapsto P(X \in xA)$ is continuous on S for each $A \in \mathcal{B}(S)$.

Proof. This follows immediately from the previous lemma, since for fixed $A \in \mathcal{B}(S)$,

$$P(X \in xA) = F(x)P(X \in A), \quad x \in S$$

$\square$

Theorem 18. Suppose that X is a random variable taking values in S. Then X has an exponential distribution if and only if X is memoryless and has constant rate with respect to a left-invariant measure on S.

Proof. Let F denote the upper probability function of X. Suppose that X has an exponential distribution. Then trivially, X is memoryless. Let μ be the positive measure on S defined by

$$\mu(A) = \mathbb{E}\left(\frac{1}{F(X)}, X \in A\right), \quad A \in \mathcal{B}(S)$$

That is, μ has density function $1/F$ with respect to the distribution of X and hence, since F is positive, X has probability density function F with respect to μ. That is, X has constant rate 1 with respect to μ:

$$\mathbb{P}(X \in A) = \int_A F(x) d\mu(x), \quad A \in \mathcal{B}(S)$$

Let $x \in S$ and $A \in \mathcal{B}(S)$. By the expected value version of the exponential property and by the memoryless property,

$$\begin{aligned}\mu(xA) &= \mathbb{E}\left(\frac{1}{F(X)}, X \in xA\right) = F(x)\mathbb{E}\left(\frac{1}{F(xX)}, X \in A\right)\\ &= F(x)\mathbb{E}\left(\frac{1}{F(x)F(X)}, X \in A\right) = \mathbb{E}\left(\frac{1}{F(X)}, X \in A\right) = \mu(A)\end{aligned}$$

Thus, μ is left-invariant.

Conversely, suppose that X is memoryless and has constant rate α with respect to a left-invariant measure λ on S. Thus $f = \alpha F$ is a density function of X with respect to λ. Let $x \in S$ and $A \in \mathcal{B}(S)$. Then, using the memoryless property and the integral version of left-invariance, we have

$$\begin{aligned}\mathbb{P}(X \in xA) &= \int_{xA} \alpha F(y) d\lambda(y) = \int_A \alpha F(xz) d\lambda(z)\\ &= F(x) \int_A \alpha F(z) d\lambda(z) = F(x)\mathbb{P}(X \in A)\end{aligned}$$

Hence X has an exponential distribution. $\square$

Note 38. This proof is essentially the same as the characterization of relative invariant measures in locally compact groups given in Halmos. Note also that since S has a unique left-invariant measure, up to multiplication by positive constants, the exponential distribution with a given upper probability function is unique.

Theorem 19. Suppose that $F\colon S \to (0,1]$ is measurable. Then F is the upper probability function of an exponential distribution if and only if

$$F(xy) = F(x)F(y); \quad x,\, y \in S \tag{15}$$

$$\int_S F(x)d\lambda(x) < \infty \tag{16}$$

in which case the rate constant is the reciprocal of the integral in (16).

Proof. Suppose that F is the upper probability function of an exponential distribution that has rate α. Then the distribution is memoryless so (15) holds, and

$$\int_S F(x)d\lambda(x) = \int_S \frac{1}{\alpha} f(x)d\lambda(x) = \frac{1}{\alpha}$$

Conversely, suppose that $F\colon S \to (0,1]$ is measurable and that (15) and (16) hold. Let $f = \alpha F$ where α is the reciprocal of the integral in (16). Then f is a probability density function by (16). Suppose that X is a random variable with this distribution. For $x \in S$ and $A \in \mathcal{B}(S)$, using (15) and the integral version of left-invariance, we have

$$\begin{aligned}\mathbb{P}(X \in xA) &= \int_{xA} f(y)d\lambda(y) = \int_{xA} \alpha F(y)d\lambda(y) = \int_A \alpha F(xz)d\lambda(z)\\ &= F(x)\int_A \alpha F(z)d\lambda(z) = F(x)\int_A f(z)d\lambda(z) = F(x)\mathbb{P}(X \in A)\end{aligned}$$

Letting $A = S$ we see that F is the upper probability function of X, and hence X has constant rate α with respect to λ. And then, of course, X is memoryless by (15). ☐

Note 39. Since S has a unique left-invariant measure, up to multiplication by positive constants, this theorem gives a method for finding all exponential distributions on S. It also follows that the memoryless property *almost* implies the constant rate property. More specifically, if F is an upper probability function satisfying (15) and (16), then $f = \alpha F$ is the probability density function of an exponential distribution (where again, α is the reciprocal of the integral in (16)), but (if $(S, \cdot, \lambda)$ does not have distributional dimension 1), then there may be other PDFs with upper probability F that do not have constant rate. This can happen, as we will see, in Section 11.2. The only other possiblity is given in the follow problem.

Problem 15. Is it possible to have an upper probability function that satisfies the memoryless property in (15), but for which the integral in (16) is infinite?

Corollary 16. Suppose that X has an exponential distribution on S with upper probability function F and constant rate α. Let $r > 0$. If

$$\frac{1}{\alpha_r} := \int_S F^r(x) d\lambda(x) < \infty$$

(in particular if $r \geq 1$) then F^r is the upper probability function of an exponential distribution with rate α_r.

Proof. Clearly $F^r(xy) = F^r(x)F^r(y)$ for $x, y \in S$. Thus the result follows immediately from the previous theorem. $\square$

In particular, it follows that if a positive semigroup has an exponential distribution, then in fact it supports at least a one-parameter family of exponential distributions.

Proposition 58. Suppose that f is a probability density function on S, and that

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

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 $G(x) = \alpha f(x)$ where $\alpha = f(e) > 0$. Let F denote the upper probability function of f. Then

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

Thus, the distribution has constant rate α with respect to λ. 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. $\square$

Suppose that the positive semigroups $(S, \cdot)$ and $(T, \cdot)$ are isomorphic, and let $\Phi \colon S \to T$ be an isomorphism. Recall that if λ is a left-invariant measure on S, then the measure μ defined by

$$\mu(B) = \lambda\left(\Phi^{-1}(B)\right), \quad B \in \mathcal{B}(T)$$

is left-invariant on T.

Theorem 20. Suppose that X is a random variable taking values in S and let $Y = \Phi(X)$. If X is exponential or memoryless or has constant rate with respect to λ, then Y is exponential or memoryless or has constant rate with respect to $\mu = \lambda\Phi^{-1}$, respectively.

Proof. Suppose that X has an exponential distribution. For $y \in T$ and $B \in \mathcal{B}(T)$,

$$\begin{aligned}\mathbb{P}(Y \in yB) &= \mathbb{P}(\Phi(X) \in yB) = \mathbb{P}\left(X \in \Phi^{-1}(yB)\right) \\ &= \mathbb{P}\left(X \in \Phi^{-1}(y)\Phi^{-1}(B)\right) = \mathbb{P}\left(X \succeq \Phi^{-1}(y)\right)\mathbb{P}\left(X \in \Phi^{-1}(B)\right) \\ &= \mathbb{P}(Y \succeq y)\mathbb{P}(Y \in B)\end{aligned}$$

Hence Y has an exponential distribution. Suppose that X has a memoryless distribution. For $y, z \in T$,

$$\begin{aligned}G(yz) &= F\left(\Phi^{-1}(yz)\right) = F\left(\Phi^{-1}(y)\Phi^{-1}(z)\right) \\ &= F\left(\Phi^{-1}(y)\right)F\left(\Phi^{-1}(z)\right) = G(y)G(z)\end{aligned}$$

Hence Y has a memoryless distribution. Suppose that X has constant rate $\alpha > 0$ with respect to λ. Then X has density function $f = \alpha F$ with respect to λ and hence Y has density function $g = (\alpha F) \circ \Phi^{-1} = \alpha(F \circ \Phi^{-1}) = \alpha G$ with respect to μ. Thus Y has constant rate α with respect to μ. □

Note 40. Suppose that $(S, \cdot)$ is a positive semigroup and that T is a topological space which is homeomorphic to S. Let $\Phi\colon S \to T$ be a homeomorphism. Define the binary operation $\cdot$ on T by

$$uv = \Phi[\Phi^{-1}(u)\Phi^{-1}(v)]$$

Recall that $(T, \cdot)$ is a positive semigroup isomorphic to $(S, \cdot)$. In particular, the previous proposition holds.

6.4 Conditional distributions

Theorem 21. Suppose that X and Y are independent random variables taking values in S. Suppose also that X has an exponential distribution with upper probability function F and Y has a memoryless distribution with upper probability function G. Then the conditional distribution of X given $X \preceq Y$ is exponential, with upper probability function FG.

Proof. First, since both distributions are memoryless, we have

$$\begin{aligned}(FG)(xy) &= F(xy)G(xy) = F(x)F(y)G(x)G(y) \\ &= F(x)G(x)F(y)G(y) = (FG)(x)(FG)(y), \quad x, y \in S\end{aligned}$$

Since X has an exponential distribution, it has constant rate α for some $\alpha > 0$. Since $G\colon S \to (0, 1]$ we have

$$\frac{1}{\gamma} := \int_S F(x)G(x)d\lambda(x) \le \int_S F(x)d\lambda(x) = \frac{1}{\alpha} < \infty$$

Thus, from our Theorem 19, it follows that FG is the upper probability function of an exponential distribution that has constant rate γ. It remains to show that this distribution is the conditional distribution of X given $X \preceq Y$. Towards this end, note that

$$\begin{aligned}\mathbb{P}(X \preceq Y) &= \mathbb{E}[\mathbb{P}(X \preceq Y \mid X)] = \mathbb{E}[G(X)] \\ &= \int_S G(x)\alpha F(x) d\lambda(x) = \alpha \int_S G(x)F(x)d\lambda(x) = \frac{\alpha}{\gamma}\end{aligned}$$

Next, if $A \in \mathcal{B}(S)$, then

$$\begin{aligned}\mathbb{P}(X \in A, X \preceq Y) &= \mathbb{E}[\mathbb{P}(X \in A, X \preceq Y \mid X)] \\ &= \mathbb{E}[G(X), X \in A] = \int_A G(x)\alpha F(x)d\lambda(x)\end{aligned}$$

and therefore

$$\mathbb{P}(X \in A \mid X \preceq Y) = \int_A \gamma F(x)G(x)d\lambda(x)$$

so the conditional density of X given $X \preceq Y$ is γFG. In particular, the conditional upper probability function is FG:

$$\begin{aligned}\mathbb{P}(X \succeq x \mid X \preceq Y) &= \gamma \int_{xS} F(y)G(y)d\lambda(y) = \gamma \int_S F(xz)G(xz)d\lambda(z) \\ &= \gamma F(x)G(x) \int_S F(z)G(z)d\lambda(z) = F(x)G(x), \quad x \in S\end{aligned}$$

∎

Note 41. Suppose that X and Y are independent variables with values in S, and that each has an exponential distribution. Then the conditional distribution of X given $X \preceq Y$ and the conditional distribution of Y given $Y \preceq X$ are the same (both exponential with upper probability function FG).

Proposition 59. Suppose that X and Y are independent random variables taking values in S and that the distribution of Y is exponential with upper probability function G. Then given $X \preceq Y$, the variables X and $X^{-1}Y$ are conditionally independent and $X^{-1}Y$ has the same distribution as Y. The conditional distribution of X given $X \preceq Y$ is

$$\mathbb{P}(X \in A \mid X \preceq Y) = \frac{\mathbb{E}[G(X), X \in A]}{\mathbb{E}[G(X)]}, \quad A \in \mathcal{B}(S)$$

Proof. As in the previous theorem, let

$$\frac{1}{\gamma} = \mathbb{P}(X \preceq Y) = \mathbb{E}[\mathbb{P}(X \preceq Y \mid X)] = \mathbb{E}[G(X)]$$

Then for A, $B \in \mathcal{B}(S)$,

$$\begin{aligned}
\mathbb{P}(X \in A, X^{-1}Y \in B \mid X \preceq Y) &= \frac{\mathbb{P}(X \in A, X^{-1}Y \in B, X \preceq Y)}{\mathbb{P}(X \preceq Y)} \\
&= \gamma \mathbb{P}(X \in A, Y \in XB) \\
&= \gamma \mathbb{E}[\mathbb{P}(X \in A, Y \in XB \mid X)] \\
&= \gamma \mathbb{E}[\mathbb{P}(Y \in XB \mid X), X \in A] \\
&= \gamma \mathbb{E}[\mathbb{P}(Y \succeq X \mid X)\mathbb{P}(Y \in B \mid X), X \in A] \\
&= \gamma \mathbb{E}[G(X)\mathbb{P}(Y \in B), X \in A] \\
&= \gamma \mathbb{P}(Y \in B)\mathbb{E}[G(X), X \in A]
\end{aligned}$$

It follows from the standard factorization theorem that X and $X^{-1}Y$ are conditionally independent given $X \preceq Y$. Letting $A = S$,

$$\mathbb{P}(X^{-1}Y \in B \mid X \preceq Y) = \mathbb{P}(Y \in B)$$

Letting $B = S$,

$$\mathbb{P}(X \in A \mid X \preceq Y) = \frac{\mathbb{E}[G(X), X \in A]}{\mathbb{E}[G(X)]}$$

□

The following proposition gives a partial converse.

Proposition 60. Suppose that X and Y are independent random variables taking values in S, and that X and $X^{-1}Y$ are conditionally independent given $X \preceq Y$. Suppose also that $x \mapsto \mathbb{P}(Y \in xA)$ is continuous for each $A \in \mathcal{B}(S)$. Then the distribution of Y is exponential.

Proof. Let A, $B \in \mathcal{B}(S)$. Form the independence assumptions, we have

$$\mathbb{P}(X \in A, Y \in XB)\mathbb{P}(X \preceq Y) = \mathbb{P}(X \in A, X \preceq Y)\mathbb{P}(Y \in XB)$$

Let μ denote the distribution of Y, so that $\mu(A) = \mathbb{P}(Y \in A)$ for $A \in \mathcal{B}(S)$. Then the last equation can be written

$$\mathbb{E}[\mathbf{1}_A(X)\mu(XB)]\mathbb{E}[\mu(XS)] = \mathbb{E}[\mathbf{1}_A(X)\mu(XS)]\mathbb{E}[\mu(XB)]$$

or equivalently,

$$\mathbb{E}\left(\mathbf{1}_A(X)\mu(XB)\mathbb{E}[\mu(XS)]\right) = \mathbb{E}\left(\mathbf{1}_A(X)\mu(XS)\mathbb{E}[\mu(XB)]\right)$$

This holds for all $A \in \mathcal{B}(S)$. Therefore for each $B \in \mathcal{B}(S)$,

$$\mathbb{P}(\mu(XB)\mathbb{E}[\mu(XS)] = \mu(XS)\mathbb{E}[\mu(XB)]) = 1$$

Since X has support S and $x \mapsto \mu(xS)$ and $x \mapsto \mu(xB)$ are continuous, it follows that

$$\mu(xB)\mathbb{E}[\mu(XS)] = \mu(xS)\mathbb{E}[\mu(XB)], \quad x \in S,\ B \in \mathcal{B}(S)$$

Letting $x = e$ we see that

$$\mu(B) = \frac{\mathbb{E}[\mu(XB)]}{\mathbb{E}[\mu(XS)]}, \quad B \in \mathcal{B}(S)$$

Substituting back, we have

$$\mu(xB) = \mu(xS)\mu(B), B \in \mathcal{B}(S), \quad x \in S,\, B \in \mathcal{B}(S)$$

so μ is an exponential distribution on S. $\square$

The collection of homomorphisms from $(S, \cdot)$ into $((0, 1], \cdot)$ forms a commutative semigroup under pointwise multiplication. The identity element is the mapping $\mathbf{1}$ given by

$$\mathbf{1}(x) = 1, \quad x \in S$$

Of course, many of these homomorphisms (such as $\mathbf{1}$) are not upper probability functions for probability measures on S. Clearly this semigroup of homomorphisms has no nontrivial invertible elements and satisfies the left cancellation law. Hence the collection of homomorphisms is a positive semigroup.

Note 42. Suppose that X and Y are independent random variables with upper probability functions F and G, respectively. Suppose also that X is exponential and Y is memoryless. We showed earlier that the conditional distribution of X given $X \preceq Y$ is exponential, and has upper probability function FG. But from Proposition 37, if $(S, \preceq)$ is a lower semi-lattice, $X \wedge Y$ also has upper probability function FG. If $(S, \cdot)$ has distributional dimension 1, then it follows that the distribution of $X \wedge Y$ is the same as the conditional distribution of X given $X \preceq Y$, and in particular, that this distribution is exponential.

6.5 Joint distributions

Suppose that $(S, \cdot)$ and $(T, \cdot)$ are standard positive semigroups with identity elements e and ϵ, and with left-invariant measures μ and ν, respectively. Recall that the direct product $(S \times T, \cdot)$ is also a standard positive semigroup, where $S \times T$ is given the product topology and where

$$(x, y)(u, v) = (xu, yv)$$

The measure $\lambda = \mu \otimes \nu$ is left-invariant.

Suppose also that (X, Y) is a random variable with values in $S \times T$ and upper probability function H. Let F and G denote the upper probability functions of X and Y, respectively.

Lemma 3. Suppose that X and Y are random variables taking values in S and T respectively. If X and Y are independent, and (x, A) is an invariant pair for X and (y, B) is an invariant pair for Y, then $((x, y), A \times B)$ is an invariant pair for (X, Y). If $((x, \epsilon), A \times T)$ is an invariant pair for (X, Y) then (x, A) is an invariant pair for X. If $((e, y), S \times B)$ is an invariant pair for (X, Y) then (y, B) is an invariant pair for Y.

Proof. Suppose that (x, A) and (y, B) are invariant pairs for X and Y respectively. Then by independence,

$$\begin{aligned}\mathbb{P}((X,Y) \in (x,y)(A \times B)) &= \mathbb{P}((X,Y) \in xA \times yB) \\ &= \mathbb{P}(X \in xA, Y \in yB) = \mathbb{P}(X \in xA)\mathbb{P}(Y \in yB) \\ &= \mathbb{P}(X \succeq x)\mathbb{P}(X \in A)\mathbb{P}(Y \succeq y)\mathbb{P}(Y \in B) \\ &= \mathbb{P}(X \succeq x, Y \succeq y)\mathbb{P}(X \in A, Y \in B) \\ &= \mathbb{P}((X,Y) \succeq (x,y))\mathbb{P}((X,Y) \in A \times B)\end{aligned}$$

Hence $((x, y), A\times B)$ is an invariant pair for (X, Y). Suppose that $((x, \epsilon), A\times T)$ is an invariant pair for (X, Y). Then

$$\begin{aligned}\mathbb{P}(X \in xA) &= \mathbb{P}((X,Y) \in (xA) \times T) = \mathbb{P}((X,Y) \in (x,\epsilon)(A \times T)) \\ &= \mathbb{P}((X,Y) \succeq (x,\epsilon))\mathbb{P}((X,Y) \in A \times T) = \mathbb{P}(X \succeq x)\mathbb{P}(X \in A)\end{aligned}$$

Thus, (x, A) is an invariant pair for X. Note that independence of X and Y is not required. By a symmetric argument, if $((e, y), S \times B)$ is an invariant pair for (X, Y) then (y, B) is an invariant pair for Y. □

Theorem 22. (X, Y) is memoryless on $(S\times T, \cdot)$ if and only if X is memoryless on $(S, \cdot)$, Y is memoryless on $(T, \cdot)$ and

$$H(x, y) = F(x)G(y), \quad x \in S,\ y \in T \tag{17}$$

Equation (17) is equivalent to the statement that $\{X \succeq_S x\}$ and $\{Y \succeq_T y\}$ are independent for all $x \in S$, $y \in T$.

Proof. Suppose that (X, Y) is memoryless on $(S \times T, \cdot)$. For each x, $u \in S$,

$$F(xu) = H(xu, \epsilon) = H((x, \epsilon)(u, \epsilon)) = H(x, \epsilon)H(u, \epsilon) = F(x)F(u)$$

so X is memoryless. By a symmetric argument, Y is memoryless. Moreover,

$$H(x, y) = H((x, \epsilon)(e, y)) = H(x, \epsilon)H(e, y) = F(x)G(y)$$

Conversely, suppose that X and Y are memoryless and (17) holds. Then for (x, y), $(u, v) \in S \times T$,

$$\begin{aligned}H((x,y)(u,v)) &= H(xu, yv) = F(xu)G(yv) = F(x)F(u)G(y)G(v) \\ &= [F(x)G(y)][F(u)G(v)] = H(x,y)H(u,v)\end{aligned}$$

so (X, Y) is memoryless. □

Note 43. If $(S\times T, \cdot)$ has distributional dimension 1, then (X, Y) is memoryless on $(S\times T, \cdot)$ if and only if X and Y are memoryless on $(S, \cdot)$ and $(T, \cdot)$, respectively, and X and Y are independent. However, not every positive semigroup has the given property.

Theorem 23. (X, Y) is exponential on $(S \times T, \cdot)$ if and only if X is exponential on $(S, \cdot)$, Y is exponential on $(T, \cdot)$, and X, Y are independent.

Proof. Suppose that X is exponential on $(S, \cdot)$, Y is exponential on $(Y, \cdot)$, and X and Y are independent. From Lemma 3, for every $x \in S$, $y \in T$, $A \in \mathcal{B}(S)$, and $\mathcal{B}(T)$, $((x, y), A \times B)$ is an invariant pair for (X, Y). It follows that for fixed $(x, y) \in S \times T$, the probability measures on $S \times T$

$$\begin{aligned} C &\mapsto \mathbb{P}[(X, Y) \in (x, y)C] \\ C &\mapsto \mathbb{P}[(X, Y) \succeq (x, y)]\mathbb{P}((X, Y) \in C] \end{aligned}$$

agree on the measurable rectangles $A \times B$, where $A \in \mathcal{B}(S)$, $B \in \mathcal{B}(T)$. It follows that these measures must agree on all of $\mathcal{B}(S \times T)$. Conversely, suppose that (X, Y) is exponential on $(S \times T, \cdot)$. For every $x \in S$ and $A \in \mathcal{B}(S)$, $((x, \epsilon), A \times T)$ is an invariant pair for (X, Y). From Lemma 3, (x, A) is an invariant pair for X, and therefore X is exponential on $(S, \cdot)$. By a symmetric argument, Y is exponential on $(T, \cdot)$. But by Theorem 22, the upper probability function H of (X, Y) factors into the upper probability functions F of X and G of Y. Since (X, Y) is exponential, it has constant rate with respect to $\lambda = \mu \otimes \nu$. Thus (X, Y) has a probability density function h of the form

$$h(x, y) = \gamma H(x, y) = \gamma F(x)G(y), \quad (x, y) \in S \times T$$

where $\gamma > 0$ is the rate constant. From the standard factorization theorem, it follows that X and Y are independent. $\square$

Note 44. The results of this section extend in an obvious way to the direct product of a finite number of positive semigroups $((S_1, \cdot), (S_2, \cdot), \ldots, (S_n, \cdot))$. If $(S_i, \cdot)$ is a discrete semigroup for $i = 1, 2, \ldots$, then the results of this section also extend in an obvious way to the semigroup

$$T = \{(x_1, x_2, \ldots) : x_i \in S_i \text{ and } x_i = e_i \text{ for all but finitely many } i\}$$

Note 45. A generalized "exponential distribution" on a product space can be defined by requiring certain invariant pairs. For example, in the positive semigroup $([0, \infty)^2, +)$, suppose that we require the following to be invariant pairs for a random variable (X, Y):

$$((x, x), C), ((x, 0), A \times [0, \infty)), ((0, x), [0, \infty) \times A))$$

for any $x \in [0, \infty)$, $C \in \mathcal{B}([0, \infty)^2)$ and $A \in \mathcal{B}([0, \infty))$. Then (X, Y) has a "bivariate exponential distribution" in the sense of Marshall and Olkin [19]. Note that this is equivalent to requiring X and Y to have exponential distributions individually, and for (X, Y) to have the memoryless property at pairs of points of the form (x, x) and (z, w):

$$\mathbb{P}(X \geq x + z, Y \geq x + w) = \mathbb{P}(X \geq x, Y \geq x)\mathbb{P}(X \geq z, Y \geq w)$$

6.6 Gamma distributions

As usual, we assume that $(S, \cdot, \lambda)$ is a standard positive semigroup. Suppose that $\boldsymbol{X} = (X_1, X_2, \ldots)$ is an IID sequence in S, with a common eponential distribution with PDF f and UPF F. Let $\boldsymbol{Y} = (Y_1, Y_2, \ldots)$ denote the corresponding partial product sequence, so that $Y_n = X_1 X_2 \cdots X_n$.

Theorem 24. The distribution of $(Y_1, Y_2, \ldots, Y_n)$ has probability density function h_n given by

$$h_n(y_1, y_2, \ldots, y_n) = \alpha^n F(y_n), \quad (y_1, y_2, \ldots, y_n) \in D_n$$

Proof. By Corollary 8,

$$h_n(y_1, y_2, \ldots, y_n) = f(y_1) f(y_1^{-1} y_2) \cdots f(y_{n-1}^{-1} y_n), \quad (y_1, y_2, \ldots, y_n) \in D_n$$

But by the constant rate and memoryless properties,

$$h_n(y_1, y_2, \ldots, y_n) = \alpha^n F(y_1) F(y_1^{-1} y_2) \cdots F(y_{n-1}^{-1} y_n) = \alpha^n F(y_n)$$

□

It follows that the sequence of ladder variables associated with $\boldsymbol{X}$ and the partial product sequence associated with $\boldsymbol{X}$ are equivalent (that is, the two sequences have the same finite dimensional distributions).

Corollary 17. The partial product sequence $\boldsymbol{Y}$ is a homogeneous Markov chain with transition probability density g given by

$$g(y, z) = f(y^{-1} z) = \frac{f(z)}{F(y)}, \quad y \in S,\ z \in I[y]$$

Corollary 18. Y_n has the gamma distribution of order n. That is, Y_n has probability density function

$$f^{*n} = \alpha^n \lambda_{n-1} F$$

Corollary 19. The conditional distribution of $(Y_1, Y_2, \ldots, Y_n)$ given $Y_{n+1} = y$ is uniform on $D_n[y]$.

The condition in Corollary 19 characterizes the exponential distribution.

Theorem 25. Suppose that X and Y are independent, identically distributed random variables taking values in S. If the conditional distribution of X given $XY = z$ is uniform on $[e, z]$ for every $z \in S$ then the common distribution is exponential.

Proof. If $A \subseteq S$ is measurable, then

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

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

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

for each $x \succ e$. Equivalently,

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

It now follows from the characterization in Proposition 58 that f is the density of an exponential distribution. □

The equivalence of the sequence of ladder variables and the sequence of partial products also characterizes the exponential distribution.

Proposition 61. Suppose that $\boldsymbol{X} = (X_1, X_2, \ldots)$ is an IID sequence taking values in S. Let $\boldsymbol{Y}$ denote the sequence of ladder variables and $\boldsymbol{Z}$ the sequence of partial products. If $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are equivalent then the distribution of X is exponential.

Proof. As usual, let f denote the probability density function of X and F the upper probability function. Since $Y_1 = Z_1 = X_1$, the equivalence of $\boldsymbol{Y}$ and $\boldsymbol{Z}$ means that the two Markov chains have the same transition probability density. Thus,

$$\frac{f(z)}{F(y)} = f(y^{-1}z), \quad y \in S,\, z \in I[y]$$

Equivalently,

$$f(xu) = F(x)f(u), \quad x,\, u \in S$$

Letting $u = e$ we have $f(x) = f(e)F(x)$, so the distribution has constant rate $\alpha = f(e)$. But then we also have $\alpha F(xu) = F(x)\alpha F(u)$, so $F(xu) = F(x)F(u)$, and hence the distribution is memoryless as well. □

Proposition 62. Suppose that F is the upper probability function of a memoryless distribution on S. Then $(\lambda_n F) * F = \lambda_{n+1} F$.

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

$$\begin{aligned}
[(\lambda_n F) * F](x) &= \int_{[e,x]} \lambda_n(t) F(t) F(t^{-1}x) d\lambda(x) \\
&= F(x) \int_{[e,x]} \lambda_n(t) d\lambda(t) = F(x)\lambda_{n+1}(x)
\end{aligned}$$

□

6.7 Compound Poisson distributions

As usual, suppose that $(S, \cdot, \lambda)$ is a standard positive semigroup. A random variable X taking values in S has a *compound Poisson distribution* if

$$X = U_1 U_2 \cdots U_N$$

where $(U_1, U_2, \ldots)$ are independent, identically distributed random variables on S and where N is independent of $(U_1, U_2, \ldots)$ and has a Poisson distribution.

Proposition 63. Suppose that $(X_1, X_2, \ldots)$ is a sequence of independent, identically distributed variables whose common distribution is compound Poisson. Then for $n \in \mathbb{N}$, the partial product $Y_n = X_1 X_2 \cdots X_n$ is compound Poisson.

Proof. For each $i \in \{1, 2, \ldots, n\}$ we have

$$X_i = \prod_{j=1}^{N_i} U_{ij}$$

where $(U_{i1}, U_{i2}, \ldots)$ are independent and identically distributed, and where N_i is independent of $(U_{i1}, U_{i2}, \ldots)$ and has a Poisson distribution. Since $(X_1, X_2, \ldots)$ are independent and identically distributed, we can take U_{ij} to be IID over all $i, j \in \mathbb{N}_+$ and we can take $(N_1, N_2, \ldots)$ IID with common parameter λ, and independent of $(U_{ij} : i, j \in \mathbb{N}_+)$. Hence we have

$$Y_n = \prod_{i=1}^{n} \prod_{j=1}^{N_i} U_{ij}$$

But this representation is clearly equivalent to $Y_n = \prod_{k=1}^{M} V_k$ where $(V_1, V_2, \ldots)$ are IID (with the same distribution as the U_{ij}), and where $M = \sum_{i=1}^{n} N_i$. But M has the Poisson distribution with parameter $n\lambda$. Hence Y_n has a compound Poisson distribution. $\square$

We will see that exponential distributions are compound Poisson in many special cases. Whenever this is true, the corresponding gamma distributions are also compound Poisson.

Problem 16. Find conditions under which an exponential distribution on a positive semigroup is compound Poisson.

6.8 Quotient spaces

As usual, we start with a standard positive semigroup $(S, \cdot, \lambda)$. Suppose that T is a standard sub-semigroup of S and that X is a random variable taking values in S with $\mathbb{P}(X \in T) > 0$.

Proposition 64. If X has an exponential distribution on $(S, \cdot)$ then the conditional distribution of X given $X \in T$ is exponential on $(T, \cdot)$, and the upper probability function of X given $X \in T$ is the restriction to T of the upper probability function of X:

$$\mathbb{P}(X \succeq_T x \mid X \in T) = \mathbb{P}(X \succeq x), \quad x \in T$$

Proof. For $A \in \mathcal{B}(T)$ and $x \in T$,

$$\begin{aligned}\mathbb{P}(X \in xA \mid X \in T) &= \frac{\mathbb{P}(X \in xA)}{\mathbb{P}(X \in T)} \\ &= \frac{\mathbb{P}(X \succeq x)\mathbb{P}(X \in A)}{\mathbb{P}(X \in T)} \\ &= \mathbb{P}(X \succeq x)\mathbb{P}(X \in A \mid X \in T)\end{aligned}$$

In particular, letting $A = T$, we have $\mathbb{P}(X \succeq_T x \mid X \in T) = \mathbb{P}(X \succeq x)$ and therefore

$$\mathbb{P}(X \in xA \mid X \in T) = \mathbb{P}(X \succeq_T x \mid X \in T)\mathbb{P}(X \in A \mid X \in T)$$

∎

We will generalize and extend this basic result. Consider the setting of Section 3.8 and suppose that X is a random variable taking values in S. We impose Assumption 1, so that X can be decomposed uniquely as $X = Y_T Z_T$ where $Y = \varphi_T(X)$ takes values in T, and $Z_T = \psi_T(X)$ takes values in S/T. Our goal is the study of the random variables Y_T and Z_T. When $T = S_t$ for $t \in S - \{e\}$, we simplify the notation to Y_t and Z_t. Note that $Y_t = t^{N_t}$ where N_t takes values in $\mathbb{N}$. The following theorem is our first main result.

Theorem 26. Suppose that X has an exponential distribution on S. Then

1. Y_T has an exponential distribution on T.
2. The upper probability function of Y_T is the restriction to T of the upper probability function of X.
3. Y_T and Z_T are independent.

Proof. Let $y \in T$, $A \in \mathcal{B}(T)$, and $B \in \mathcal{B}(S/T)$. Then by the uniqueness of the factorization and since X has an exponential distribution,

$$\begin{aligned}\mathbb{P}(Y_T \in yA, Z_T \in B) &= \mathbb{P}(X \in yAB) = \mathbb{P}(X \succeq y)\mathbb{P}(X \in AB) \\ &= \mathbb{P}(X \succeq y)\mathbb{P}(Y_T \in A, Z_T \in B). \qquad (18)\end{aligned}$$

Substituting $A = T$ and $B = S/T$ in (18) gives

$$\mathbb{P}(Y_T \succeq_T y) = \mathbb{P}(X \succeq y)$$

so the upper probability function of Y is the restriction to T of the upper probability function of X. Returning to (18) with general A and $B = S/T$ we have

$$\mathbb{P}(Y_T \in yA) = \mathbb{P}(Y_T \succeq_T y)\mathbb{P}(Y_T \in A)$$

so Y_T has an exponential distribution on T. Finally, returning to (18) with $A = T$ and general B we have

$$\mathbb{P}(Y_T \succeq_T y, Z_T \in B) = \mathbb{P}(Y_T \succeq_T y)\mathbb{P}(Z_T \in B)$$

so Z_T and tail events of Y_T are independent. To get full independence, recall that X has constant rate α with respect to λ, so X has density f with respect to λ given by $f(x) = \alpha\mathbb{P}(X \succeq x)$ for $x \in S$. Similarly, Y_T has constant rate with respect to μ on T. Thus the function g on $S \times S/T$ given by $g(y, z) = f(yz)$ is a density function of (Y_T, Z_T) with respect to $\mu \otimes \nu$. By the memoryless property,

$$g(y, z) = f(yz) = \alpha\mathbb{P}(X \succeq yz) = \alpha\mathbb{P}(X \succeq y)\mathbb{P}(X \succeq z), \quad y \in S,\, z \in S/T$$

and so by the standard factorization theorem, Y_T and Z_T are independent. $\square$

Example 13. Consider the direct product of positive semigroups $(S_1, \cdot)$ and $(S_2, \cdot)$ with the sub-semigroup and quotient space discussed previously. In this case, Theorem 26 gives another proof of the characterization of exponential distributions: (X_1, X_2) is exponential on $S_1 \times S_2$ if and only if X_1 is exponential on S_1, X_2 is exponential on S_2, and X_1, X_2 are independent.

Proposition 65. Suppose that X is a random variable taking values in S.

1. If $\mathbb{P}(X \in T) > 0$ then the conditional distribution of X given $X \in T$ is the same as the distribution of Y_T if and only if Y_T and $\{Z_T = e\}$ are independent.

2. If $\mathbb{P}(X \in S/T) > 0$ then the conditional distribution of X given $X \in S/T$ is the same as the distribution of Z_T if and only if Z_T and $\{Y_T = e\}$ are independent.

Proof. Suppose that $A \in \mathcal{B}(T)$. Then $\{X \in A\} = \{Y_T \in A, Z_T = e\}$ and in particular, $\{X \in T\} = \{Z_T = e\}$. Thus,

$$\mathbb{P}(X \in A | X \in T) = \mathbb{P}(Y_T \in A | Z_T = e).$$

The proof of the second result is analogous. $\square$

Corollary 20. Suppose that X has an exponential distribution on S.

1. If $\mathbb{P}(X \in T) > 0$ then the conditional distribution of X given $X \in T$ is the same as the distribution of Y_T, and this distribution is exponential on T.

2. If $\mathbb{P}(X \in S/T) > 0$ then the conditional distribution of X given $X \in S/T$ is the same as the distribution of Z_T.

Corollary 21. Suppose that X has an exponential distribution on S and $t \in S - \{e\}$. Then N_t has a geometric distribution on $\mathbb{N}$ with parameter $p_t = 1 - P(X \succeq t)$:

$$\mathbb{P}(N_t = n) = p_t(1 - p_t)^n, \quad n \in \mathbb{N}.$$

Proof. By Theorem 26, Y_t has an exponential distribution on S_t and therefore N_t has a geometric distribution on $\mathbb{N}$, since $(S_t, \cdot)$ and $(\mathbb{N}, +)$ are isomorphic. Next, $N_t \geq 1$ if and only if $Y_t \succeq t$ if and only if $X \succeq t$. Thus, the rate (or success) parameter of the geometric distribution is $p_t = 1 - \mathbb{P}(N_t \geq 1) = 1 - \mathbb{P}(X \succeq t)$. $\square$

The following theorem is our second main result, and gives a partial converse to Theorem 26.

Theorem 27. Suppose X is a random variable taking values in S and that for each $t \in S - \{e\}$, Y_t and Z_t are independent, and Y_t has an exponential distribution on S_t. Then X has an exponentially distribution on S.

Proof. Let p_x denote the parameter of the geometric distribution of N_x, so that

$$\mathbb{P}(N_x = n) = p_x(1 - p_x)^n, \quad n \in \mathbb{N}.$$

Let $x \in S$ and let $A \in \mathcal{B}(S)$. Because of the basic assumptions, we have $A = \cup_{n=0}^{\infty} x^n B_n$ where $B_n \subseteq S/S_x$ for each n; the collection $\{x^n B_n : n \in \mathbb{N}\}$ is disjoint. Similarly, $xA = \cup_{n=0}^{\infty} x^{n+1} B_n$ and the collection $\{x^{n+1} B_n : n \in \mathbb{N}\}$ is disjoint. From the hypotheses,

$$\begin{aligned}
\mathbb{P}(X \in xA) &= \sum_{n=0}^{\infty} \mathbb{P}(X \in x^{n+1} B_n) = \sum_{n=0}^{\infty} \mathbb{P}(N_x = n + 1, Z_x \in B_n) \\
&= \sum_{n=0}^{\infty} p_x(1 - p_x)^{n+1} \mathbb{P}(Z_x \in B_n).
\end{aligned}$$

But also $1 - p_x = \mathbb{P}(N_x \geq 1) = \mathbb{P}(Y_x \succeq x) = \mathbb{P}(X \succeq x)$ so

$$\begin{aligned}
\mathbb{P}(X \succeq x)\mathbb{P}(X \in A) &= (1 - p_x) \sum_{n=0}^{\infty} \mathbb{P}(X \in x^n B_n) \\
&= (1 - p_x) \sum_{n=0}^{\infty} \mathbb{P}(N_x = n, Z_x \in B_n) \\
&= (1 - p_x) \sum_{n=0}^{\infty} p_x(1 - p_x)^n \mathbb{P}(Z_x \in B_n).
\end{aligned}$$

If follows that $\mathbb{P}(X \in xA) = \mathbb{P}(X \succeq x)\mathbb{P}(X \in A)$ and hence X has an exponential distribution on S. $\square$

Part III

Examples and Applications

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

The pair $([0, \infty), +)$ with the ordinary topology is a positive semigroup; 0 is the identity element. The corresponding partial order is the ordinary order $\leq$. Lebesgue measure λ is the only invariant measure, up to multiplication by positive constants. The cumulative function of order $n \in \mathbb{N}$ associated with Lebesgue measure λ is

$$\lambda_n(x) = \frac{x^n}{n!}, \quad x \in [0, \infty)$$

Hence the generating function associated with $(\lambda_n : n \in \mathbb{N}_+)$ is given by

$$\Lambda(x, t) = \sum_{n=0}^{\infty} \frac{x^n}{n!} t^n = e^{tx}$$

This semigroup has dimension 1 in every sense.

7.1 Exponential distributions

A distribution is memoryless if and only if it has constant rate with respect to Lebesgue measure if and only if it is exponential. The exponential distribution with constant rate $\alpha > 0$ has upper probability function $F(x) = e^{-\alpha x}$ and probability density function $f(x) = \alpha e^{-\alpha x}$.

Let X have the exponential distribution on $([0, \infty), +)$ with rate parameter $\alpha > 0$ and upper probability function F, as given above. If Y is a random variable taking values in $[0, \infty)$ then from our general discussion of entropy in Section 4.9,

$$H(Y) \leq -\ln(\alpha) - \mathbb{E}[\ln(F(Y))] = -\ln(\alpha) - \alpha \mathbb{E}(Y)$$

It follows that X maximizes entropy over all random variables with $\mathbb{E}(Y) = \mathbb{E}(X) = 1/\alpha$; the maximum entropy is

$$H(X) = 1 - \ln(\alpha)$$

7.2 Gamma distributions and the Point Process

Suppose that $\boldsymbol{X} = (X_1, X_2, \ldots)$ is a sequence of independent variables, each having the exponential distribution with rate α. Let $\boldsymbol{Y} = (Y_1, Y_2, \ldots)$ denote the partial sum process, or equivalently, the sequence of ladder variables. The distribution of $(Y_1, Y_2, \ldots, Y_n)$ has density

$$h_n(y_1, y_2, \ldots, y_n) = \alpha^n e^{-y_n}, \quad 0 \leq y_1 \leq y_2 \leq \cdots \leq y_n$$

The transition density of $\boldsymbol{Y}$ is

$$g(y,z) = \alpha e^{-(y-z)}, \quad 0 \le y \le z$$

Finally, Y_n has the gamma distribution with rate constant α and order n, with probability density function

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

Of course, this is the standard gamma distribution.

The counting process $\{N_x : x \in [0,\infty)\}$, of course is the standard Poisson process. To check our results, we will compute the renewal function via the formula derived in Section 5.11:

$$\begin{aligned} m(x) &= \mathbb{E}(N_x) = \mathbb{E}[\Lambda(X,\alpha), X \le x] = \mathbb{E}(e^{\alpha X}, X \le x) \\ &= \int_0^x e^{\alpha t} \alpha e^{-\alpha t} dt = \alpha x, \quad x \in [0,\infty) \end{aligned}$$

We also check the thinning result from Section 5.11: The probability density function of Y_N, the first accepted point is

$$g(x) = r\alpha\Lambda[x,(1-r)\alpha]F(x) = r\alpha e^{(1-r)\alpha} e^{-\alpha x} = r\alpha e^{-r\alpha x}, \quad x \in [0,\infty)$$

so Y_N has the exponential distribution with parameter $r\alpha$.

7.3 Sub-semigroups and quotient spaces

For $t \in (0,\infty)$, the sub-semigroup generated by t is $\{nt : n \in \mathbb{N}\}$ and the corresponding quotient space is $[0,t)$. The assumptions in Section 3.8 are satisfied; thus, $x \in [0,\infty)$ can be written uniquely as

$$x = tn_t(x) + z_t(x).$$

where $n_t(x) = \lfloor x/t \rfloor \in \mathbb{N}$ and $z_t(x) = x - tn_t(x) = x \bmod t \in [0,t)$.

Suppose that X has the exponential distribution on $[0,\infty)$ with rate parameter α. From Theorem 26,

1. N_t and Z_t are independent.
2. N_t has the geometric distribution on $\mathbb{N}$ with rate parameter $p_t = 1 - e^{-\alpha t}$.
3. The distribution of Z_t is the same as the conditional distribution of X given $X < t$ and has density function $s \mapsto \alpha e^{-\alpha s}/(1 - e^{-\alpha t})$ on $[0,t)$.

In this standard setting, we can do better than the general converse stated in Section 3.8. Suppose that X is a random variable taking values in $[0,\infty)$. Galambos & Kotz [8] (see also [1]) show that if N_t has a geometric distribution for all $t > 0$ then X has an exponential distribution. We now explore a converse based on independence properties of N_t and Z_t. Suppose that X has a continuous distribution with density function f and distribution function F. If Z_t and $\{N_t = 0\}$ are independent for each $t > 0$ then we can assume (by an appropriate choice of the density function) that

$$f(s) = F(t) \sum_{n=0}^{\infty} f(nt + s). \tag{19}$$

for $t \in (0, \infty)$ and $s \in [0, t)$. However, it is easy to see that if X has an exponential distribution, then (19) holds for all $t > 0$ and $s \geq 0$. Thus, our converse is best stated as follows:

Theorem 28. Suppose that (19) holds for $s = 0$ and for $s = t$, for all $t > 0$. Then X has an exponential distribution.

Proof. The hypotheses are that

$$f(0) = F(t) \sum_{n=0}^{\infty} f(nt), \quad t \in (0, \infty) \tag{20}$$

$$f(t) = F(t) \sum_{n=0}^{\infty} f((n+1)t), \quad t \in (0, \infty) \tag{21}$$

Thus from (21) we have

$$f(t) = F(t) \sum_{n=1}^{\infty} f(nt) = F(t) \left(\sum_{n=0}^{\infty} f(nt) - f(0) \right).$$

Applying (20) gives $f(t) = f(0) - F(t)f(0)$ for $t \in (0, \infty)$. It follows that f is differentiable on $(0, \infty)$ and $f'(t) = -f(0)f(t)$ for $s \in (0, \infty)$. Therefore $f(t) = \alpha e^{-\alpha t}$ for $t \in (0, \infty)$, where $\alpha = f(0)$, and hence X has an exponential distribution on $[0, \infty)$. □

Note 46. The quotient space here can also be viewed as a lexicographic product. That is, $([0, \infty), \leq)$ is isomorphic to the lexicographic product of $(t\mathbb{N}, \leq)$ with $([0, t), \leq)$.

7.4 Compound Poisson distributions

Suppose that X has the exponential distribution on $([0, \infty), +)$ with rate parameter α as above. Then it is well known that X has a compound Poisson distribution.

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

The pair $(\mathbb{N}, +)$ (with the discrete topology, of course) is a positive semigroup; 0 is the identity element. The corresponding partial order is the ordinary order $\leq$. Since the space is discrete, counting measure # is the only invariant measure, up to multiplication by positive constants.

Proposition 66. The cumulative function of order $n \in \mathbb{N}$ for counting measure is given by

$$\#_n(x) = \binom{n+x}{x}, \quad x \in \boldsymbol{N}_+$$

Proof. The expression is correct when $n = 0$. Assume that the expression is correct for a given value of n. Then

$$\#_{n+1}(x) = \sum_{z=0}^{x} \#_n(z) = \sum_{z=0}^{x} \binom{n+z}{x} = \binom{n+1+x}{x}$$

by a well-known combinatorial identity. □

The generating function associated with $(\#_n : n \in \mathbb{N})$ is

$$\Lambda(x,t) = \sum_{n=0}^{\infty} \binom{n+x}{x} t^n = \frac{1}{(1-t)^{x+1}}, \quad |t| < 1,\ x \in \mathbb{N}$$

This semigroup has dimension 1 in every sense.

8.1 Exponential distributions

A distribution is memoryless if and only if it has constant rate with respect to # if and only if it is exponential. The exponential distribution on $(\mathbb{N}, +)$ with constant rate $p \in (0, 1)$ has upper probability function $F(x) = (1-p)^x$ and probability mass function $f(x) = p(1-p)^x$. Of course, this distribution is the *geometric distribution* with success parameter p. This distribution governs the number of failures before the first success in a sequence of Bernoulli trials with success parameter p.

Suppose that X has the exponential distribution on $(\mathbb{N}, +)$ with rate parameter $p \in (0,1)$ and upper probability function F as given above. If Y is a random variable taking values in $\mathbb{N}$ then from our general discursion of entropy in Section 4.9

$$H(Y) \le -\ln(p) - \mathbb{E}[\ln(F(Y))] = -\ln(p) - \mathbb{E}(Y)\ln(1-p)$$

It follows that X maximizes entropy over all random variables Y with $\mathbb{E}(Y) = \mathbb{E}(X) = (1-p)/p$. The maximum entropy is

$$H(X) = -\ln(p) - (1-p)\ln(1-p)/p$$

8.2 Gamma distributions and the point process

Suppose that $\boldsymbol{X} = (X_1, X_2, \ldots)$ is a sequence of independent variables, each with the exponential distribution on $(\mathbb{N}, +)$ that has constant rate p (that is, an IID sequence of geometric variables). Let $\boldsymbol{Y} = (Y_1, Y_2, \ldots)$ denote the corresponding sequence of ladder variables, or equivalently, the sequence of partial sums. Then $(Y_1, Y_2, \ldots, Y_n)$ has PDF

$$h_n(y_1, y_2, \ldots, y_n) = p^n (1-p)^{y_n}, \quad (y_1, y_2, \ldots, y_n) \in D_n$$

The transition probability density of $\boldsymbol{Y}$ is

$$g(y, z) = p(1-p)^{z-y}, \quad y, z \in \mathbb{N},\ y \le z$$

Finally, Y_n has density function

$$f_n(x) = p^n \#_{n-1}(x) F(x) = p^n \binom{n+x-1}{x} (1-p)^x, \quad x \in \mathbb{N}$$

which of course we recognize as the negative binomial distribution with parameters n and p. This distribution governs the number of failures before the nth success in a sequence of Bernoulli trials with success parameter p.

Consider next the counting process $(N_x, x \in \mathbb{N})$ associated with $\boldsymbol{Y}$. The upper probability function of N_x is

$$G_n(x) = \sum_{k=0}^{x} p^n \binom{n+k-1}{k} (1-p)^k, \quad n \in \mathbb{N}_+$$

This does not seem to reduce to a simple closed-form expression, so let's compare our point process to one that's more familiar. The variable $Z_n = Y_n + n$ is the trial number of the nth success, in the usual Bernoulli trials formulation; Z_n has the alternative negative binomial distribution. For the corresponding counting process,

$$M_x = \#\{k \in \mathbb{N}_+ : Z_k \le x\}$$

is the number of successes in the first x trials, and of course has the binomial distribution with parameters x and p. For our point process we have

$$N_x = \#\{k \in \mathbb{N}_+ : Y_k \le x\} = \#\{k \in \mathbb{N}_+ : Z_k - k \le x\}$$

The renewal function of our point process, on the other hand, is easy to compute:

$$m(x) = \mathbb{E}(N_x) = \sum_{t=0}^{x} \frac{1}{(1-p)^{t+1}} p(1-p)^t = \frac{p}{1-p}(x+1), \quad x \in \mathbb{N}$$

For the thinned point process, the probability density function of the first accepted point Y_N is

$$\begin{aligned} g(x) &= rp\Lambda[x, (1-r)p]F(x) = rp \frac{1}{[1-(1-r)p]^{x+1}} (1-p)^x \\ &= \frac{rp}{1-p+rp} \left(\frac{1-p}{1-p+rp} \right)^x, \quad x \in \mathbb{N} \end{aligned}$$

so Y_N has the exponential distribution with constant rate $rp/(1-p+rp)$.

8.3 Compound Poisson distributions

Suppose that X has the exponential distribution on $(\mathbb{N}, +)$ with rate parameter p, as above. It is well known that X has a compound Poisson distribution. Specifically, X can be decomposed as

$$X = U_1 + U_2 + \cdots + U_N$$

where $(U_1, U_2, \ldots)$ are independent and identically distributed on $\mathbb{N}_+$ with a *logarithmic distribution*:

$$\mathbb{P}(U = n) = -\frac{(1-p)^n}{n \ln(p)}, \quad n \in \mathbb{N}_+$$

and where N is independent of $(U_1, U_2, \ldots)$ and has a Poisson distribution with parameter $-\ln(p)$. From the general theory, if Y has the corresponding gamma distribution on $(\mathbb{N}, +)$ with parameter n, then Y also has a compound Poisson distribution. Specifically,

$$Y = U_1 + U_2 + \cdots + U_N$$

where the structure is the same, except that N has the Poisson distribution with parameter $-n \ln(p)$. We will also need the following related result:

Proposition 67. Suppose that X_n has the Poisson distribution with parameter $\lambda_n = (1-p)^n/n$ for $n \in \mathbb{N}_+$ where $p \in (0, 1)$. Then $Y = \sum_{n=1}^\infty n X_n$ has the geometric distribution with rate parameter p.

Proof. The probability generating function of X_n is

$$\mathbb{E}\left(t^{X_n}\right) = \exp(\lambda_n(t-1)) = \exp\left(\frac{(1-p)^n}{n}(t-1)\right)$$

Hence the probability generating function of Y is

$$\begin{aligned}
\mathbb{E}\left(t^Y\right) &= \prod_{n=1}^\infty \mathbb{E}\left(t^{nX_n}\right) = \prod_{n=1}^\infty \exp\left(\frac{(1-p)^n}{n}(t^n - 1)\right) \\
&= \exp\left(\sum_{n=1}^\infty \frac{(1-p)^n}{n}(t^n - 1)\right) = \exp[-\ln(1-(1-p)t) + \ln(p)] \\
&= \frac{p}{1-(1-p)t}, \quad |t| < \frac{1}{1-p}
\end{aligned}$$

But this is the probability generating function of the geometric distribution with rate parameter p. ∎

8.4 Sub-semigroups and quotient spaces

For $t \in \mathbb{N}_+$, the sub-semigroup generated by t is $\{nt : n \in \mathbb{N}\}$ and the corresponding quotient space is $\{0, \ldots t-1\}$. The assumptions in Section 3.8 hold. Thus, $x \in \mathbb{N}$ can be written uniquely as

$$x = tn_t(x) + z_t(x)$$

where $n_t(x) = \lfloor x/t \rfloor \in \mathbb{N}$ and $z_t(x) = x - y_t(x) = x \bmod t \in \{0, 1, \ldots t-1\}$.

Suppose that X has the geometric distribution on $\mathbb{N}$ with rate parameter p. From Theorem 26,

1. N_t and Z_t are independent.
2. N_t has the geometric distribution on $\mathbb{N}$ with rate parameter $1-(1-p)^t$, and this is also the conditional distribution of X/t given $X \in T_t$.
3. Z_t has probability density function $k \mapsto p(1-p)^k/(1-(1-p)^t)$ on $\{0, \ldots, t-1\}$, and this is the conditional distribution of X given $X \in \{0, 1, \ldots, t-1\}$.

As before, we are interested in a converse that is stronger than the general converse in Section 3.8. Thus, suppose that X is a random variable taking values in $\mathbb{N}$, with density function f and (lower) distribution function F. Then the independence of Z_t and $\{N_t = 0\}$ is equivalent to

$$f(k) = F(t-1) \sum_{n=0}^{\infty} f(nt+k) \tag{22}$$

for all $t \in \mathbb{N}_+$ and all $k \in \{0, \ldots, t-1\}$. However, as in the continuous case, it is easy to see that if X has a geometric distribution, then (22) holds for all $k \in \mathbb{N}$, not just $k \in \{0, 1, \ldots, t-1\}$.

Theorem 29. Suppose that (22) holds for $k = 0$ and for $k = t$, for all $t \in \mathbb{N}_+$. Then X has a geometric distribution.

Proof. The hypotheses are

$$f(0) = F(t-1) \sum_{n=0}^{\infty} f(nt), \quad t \in \mathbb{N}_+ \tag{23}$$

$$f(t) = F(t-1) \sum_{n=0}^{\infty} f((n+1)t), \quad t \in \mathbb{N}_+ \tag{24}$$

From (23) and (24), it follows that

$$f(t) = f(0)[1 - F(t-1)], \quad t \in \mathbb{N}_+$$

and therefore $F(t) = 1 - (1-p)^t$ for $t \in \mathbb{N}$ where $p = f(0)$. Thus, X has a geometric distribution. ∎

However, there are non-geometrically distributed variables for which N_t and Z_t are independent. The following result is easy to verify.

Theorem 30. . Suppose that the support of X is $\{0, 1\}$ or $\{0, 2\}$ or $\{c\}$ for some $c \in \mathbb{N}$. Then N_t and Z_t are independent for all $t \in \mathbb{N}_+$.

The following theorem gives a partial converse:

Theorem 31. Suppose that X takes values in a proper subset of $\mathbb{N}$ and that Z_t and $\{N_t = 0\}$ are independent for all t. Then the support of X is one of the sets in Theorem 30.

Proof. As usual, let f denote the density function and F the distribution function of X.

First, we will use induction on $a \in \mathbb{N}$ to show that if X takes values in $\{0, 1, \ldots, a\}$, then the support of X is one of the sets in Theorem 30. If $a = 0$ or $a = 1$, the result is trivially true. Suppose that the statement is true for a given a, and suppose that X takes values in $\{0, 1, \ldots, a+1\}$. With $t = a + 1$, (22) becomes

$$f(k) = \sum_{j=0}^{a} f(j) \sum_{n=0}^{\infty} f(n(a+1)+k), \quad k \in \{0, \ldots, a\} \tag{25}$$

But $\sum_{j=0}^{a} f(j) = 1 - f(a+1)$. Hence (25) gives

$$f(0) = [1 - f(a+1)][f(0) + f(a+1)], \quad (k = 0) \tag{26}$$

$$f(k) = [1 - f(a+1)]f(k), \quad k \in \{1, \ldots, a\} \tag{27}$$

Suppose that $f(k) > 0$ for some $k \in \{1, \ldots, a\}$. Then from (27), $f(a+1) = 0$. Hence X takes values in $\{0, \ldots, a\}$, so by the induction hypothesis, the support set of X is $\{0, 1\}$, $\{0, 2\}$, or $\{c\}$ for some c. Suppose that $f(k) = 0$ for all $k \in \{1, \ldots, a\}$. Thus, X takes values in $\{0, a+1\}$. But then (22) with $t = a$ and $k = 0$ gives $f(0) = f(0)f(0)$. If $f(0) = 0$ then the support of X is $\{a+1\}$. If $f(0) = 1$ then the support of X is $\{0\}$.

To complete the proof, suppose that X takes values in a proper subset of $\mathbb{N}$, so that $f(k) = 0$ for some $k \in \mathbb{N}$. Then $F(t-1) > 0$ for t sufficiently large and hence by (22), $f(t+k) = 0$ for t sufficiently large. Thus X takes values in $\{0, 1, \ldots, a\}$ for some a, and hence the support of X is one of the sets in Theorem 30. $\square$

Problem 17. If X has support $\mathbb{N}$ and if Z_t and $\{N_t = 0\}$ are independent for each $t \in \mathbb{N}_+$, does X have a geometric distribution?

9 Positive semigroups isomorphic to $[0, \infty), +)$

9.1 The positive semigroup

Now let I be an interval of the form $[a, b)$ where $-\infty < a < b \leq \infty$ or of the form $(a, b]$ where $-\infty \leq a < b < \infty$; I has the relative topology. Let Φ be a homeomorphism from I onto $[0, \infty)$. If we define the operator $*$ on I by

$$x * y = \Phi^{-1}[\Phi(x) + \Phi(y)], \quad x, y \in I$$

then $(I, *)$ is a positive semigroup isomorphic to $([0, \infty), +)$ and Φ is an isomorphism. The partial order $\preceq$ induced by $*$ on I is the ordinary order $\leq$ if $I = [a, b)$ (Φ must be strictly increasing) and is the reverse of the ordinary order if $I = (a, b]$ (Φ must be strictly decreasing). If $x, y \in I$ and $x \preceq y$ then

$$x^{*-1} * y = \Phi^{-1}[\Phi(y) - \Phi(x)]$$

The rule for exponentiation under $*$ is $x^{*c} = \Phi^{-1}[c\Phi(x)]$ for $x \in I$ and $c \geq 0$.

A left-invariant measure μ for $(I, *)$ (unique up to multiplication by positive constants is given by

$$\mu(A) = \lambda[\Phi(A)], \quad A \in \mathcal{B}(I)$$

where λ is Lebesgue measure on $[0, \infty)$. If Φ is a smooth function with derivative ϕ then

$$d\mu(x) = |\phi(x)| d\lambda(x)$$

9.2 Exponential distributions

Because the semigroups are isomorphic, a random variable X taking values in I is exponential for the positive semigroup $(I, *)$ if and only if $\Phi(X)$ is exponential for the standard, positive semigroup $([0, \infty), +)$ (that is, $\Phi(X)$ has an exponential distribution in the ordinary sense). In particular, a distribution will be exponential relative to $(I, *)$ if and only if it is memoryless, if and only if it has constant rate with respect to μ. It therefore follows that the exponential distribution relative to $(I, *)$ which has constant rate $\alpha > 0$ with respect to μ is the distribution with upper probability function

$$F(x) = \exp[-\alpha\Phi(x)], \quad x \in I$$

The probability density function of X relative to the left-invariant measure μ, is of course,

$$f(x) = \alpha \exp[-\alpha\Phi(x)], \quad x \in I$$

by the constant rate property. The probability density function of X relative to Lebesgue measure λ is

$$g(x) = \alpha \exp[-\alpha\Phi(x)]|\phi(x)|, \quad x \in I$$

In particular, note that $\alpha|\phi|$ is the rate function of X in the ordinary sense.

Suppose that X is a random variable taking values in I with a continuous distribution (with support I of course). Then X is exponential with respect to some semigroup $(I, *)$ isomorphic to $([0, \infty), +)$. Specifically, if $I = [a, b)$, let

$$\Phi(x) = -\ln[\mathbb{P}(X \geq x)], \quad x \in [a, b)$$

and if $I = (a,\, b]$, let

$$\Phi(x) = -\ln[\mathbb{P}(X \le x)], \quad x \in (a, b]$$

This is essentially a restatement of the fact that $Y = \Phi(X)$ has a standard exponential distribution. However the semigroup formulation provides some additional insights.

Let X have the exponential distribution on $(I, *)$ with rate parameter α and upper probability function F given above. If Y is a random variable taking values in I, then from our general discussion of entropy in Section 4.9,

$$H(Y) \le -\ln(\alpha) - \mathbb{E}[\ln(F(Y))] = -\ln(\alpha) + \alpha\mathbb{E}[\Phi(Y)]$$

it follows that X maximizes entropy over all random variables with $\mathbb{E}[\Phi(Y)] = \mathbb{E}[\Phi(X)] = 1/\alpha$; the maximum entropy is $1 - \ln(\alpha)$.

9.3 Gamma distributions

Similarly, a random variable Y_n taking values in I has the gamma distribution on $(I, *)$ with rate α and order n if and only if $\Phi(Y_n)$ has the gamma distribution on $[0, \infty), +)$ with rate α and order n (that is, a gamma distribution in the usual sense). It follows from the usual change of variables formula that the density of Y_n with respect to Lebesgue measure λ is

$$g_n(x) = \alpha^n \frac{\Phi^{n-1}(x)}{(n-1)!} \exp[-\alpha\Phi(x)]\phi(x), \quad x \in I$$

Hence, the density function of Y_n with respect to the left-invariant measure μ is

$$f_n(x) = \alpha^n \frac{\Phi^{n-1}(x)}{(n-1)!} \exp[-\alpha\Phi(x)], \quad x \in I$$

Therefore, the cumulative function of order $n \in \mathbb{N}$ for μ is

$$\mu_n(x) = \frac{\Phi^n(x)}{n!}, \quad x \in I$$

Of course, we could also derive this last result directly.

Our characterization of exponential distributions based on IID variables goes like this: if X and Y are independent and identically distributed on I, then the common distribution is exponential if and only if the conditional distribution of X given $X * Y = z$ is uniform on $[a, z]$ if $I = [a, b)$ or uniform on $[z, b]$ if $I = (a, b]$.

In the remainder of this chapter we explore a number of specific examples.

9.4 The positive semigroup $((0,\, 1],\, \cdot)$

Let $I = (0,\, 1]$ and let $\Phi(x) = -\ln(x)$. Then Φ is a homeomorphism from I onto $[0, \infty)$ and the associated operation $*$ is ordinary multiplication. The associated partial order is the reverse of the ordinary order. The invariant measure μ is given by $d\mu(x) = (1/x)d\lambda(x)$. The exponential distribution with constant rate $\alpha > 0$ has density

$$f(x) = \alpha x^\alpha, \quad x \in (0, 1]$$

with respect to μ and density

$$g(x) = \alpha x^{\alpha-1}, \quad x \in (0, 1]$$

with respect to Lebesgue measure λ Note that this is the beta distribution with parameters α and 1 and the special case $\alpha = 1$ gives the uniform distribution on (0, 1].

Suppose that X has the exponential distribution on $((0, 1], \cdot)$ with rate parameter α as given above. From our general discussion, X maximizes entropy over all random variables Y taking values in $(0, 1]$ with $\mathbb{E}[-\ln(Y)] = 1/\alpha$.

The gamma distribution on $((0, 1], \cdot)$ with rate parameter α and order n has density function

$$f_n(x) = \alpha^n (-1)^{n-1} \frac{\ln^{n-1}(x)}{(n-1)!} x^\alpha, \quad x \in (0, 1]$$

with respect to μ, and density

$$g_n(x) = \alpha^n (-1)^{n-1} \frac{\ln^{n-1}(x)}{(n-1)!} x^{\alpha-1}, \quad x \in (0, 1]$$

with respect to Lebesgue measure λ. The cumulative function of order $n \in \mathbb{N}$ for μ is given by

$$\mu_n(x) = (-1)^n \frac{\ln^n(x)}{n!}, \quad x \in (0, 1]$$

9.5 The positive semigroup $([1, \infty), \cdot)$

Let $I = [1, \infty)$ and let $\Phi(x) = \ln(x)$. Then Φ is a homeomorphism from I onto $[0, \infty)$ and the associated operation $*$ is ordinary multiplication. The associated partial order is the ordinary order $\leq$. The invariant measure μ is given by $d\mu(x) = (1/x)d\lambda(x)$, where λ is Lebesgue measure. The exponential distribution with constant rate $\alpha > 0$ has upper probability function

$$F(x) = x^{-\alpha}, \quad x \in [1, \infty)$$

The distribution has probability density function

$$f(x) = \alpha x^{-\alpha}, \quad x \in [1, \infty)$$

with respect to μ and density

$$g(x) = \alpha x^{-(\alpha+1)}, \quad x \in [1, \infty)$$

with respect to Lebesgue measure λ. This is the Pareto distribution with shape parameter α.

Suppose that X has the exponential distribution on $([1,\infty),\cdot)$ with rate parameter α as given above. From our general discussion, X maximizes entropy over all random variables Y taking values in $[1,\infty)$ with $\mathbb{E}[\ln(Y)] = 1/\alpha$.

The gamma distribution on $([1,\infty),\cdot)$ with rate parameter α and order n has density function

$$f_n(x) = \alpha^n \frac{\ln^{n-1}(x)}{(n-1)!} x^{\alpha}, \quad x \in [1,\infty)$$

with respect to μ, and density

$$g_n(x) = \alpha^n \frac{\ln^{n-1}(x)}{(n-1)!} x^{\alpha-1}, \quad x \in [1,\infty)$$

with respect to Lebesgue measure λ. The cumulative function of order $n \in \mathbb{N}$ for μ is

$$\mu_n(x) = \frac{\ln^n(x)}{n!}, \quad x \in [1,\infty)$$

9.6 An application to Brownian functionals

Let $I = [0,\,1/2)$ and let $\Phi(x) = x/(1-2x)$. Then Φ is a homeomorphism from I onto $[0,\infty)$ and the corresponding semigroup operation $*$ is given by

$$x * y = \frac{x + y - 4xy}{1 - 4xy}, \quad x,\, y \in [0, 1/2)$$

The invariant measure μ is given by

$$d\mu(x) = \frac{1}{(1-2x)^2} d\lambda(x)$$

where λ is Lebesgue measure, and the associated partial order is the ordinary order $\leq$. The positive semigroup $([0,1/2),*)$ occurs in the study of generalized Brownian functionals [16]

The exponential distribution on $([0,1/2),*)$ with rate α has upper probability function

$$F(x) = \exp\left(-\alpha \frac{x}{1-2x}\right), \quad x \in [0,1/2)$$

Suppose that X has the exponential distribution on $([0,1/2),*)$ with rate parameter α as given above. From our general discussion, X maximizes entropy over all random variables Y taking values in $[0,1/2)$ with $\mathbb{E}[Y/(1-2Y)] = 1/\alpha$.

The gamma distribution on $([0,1/2),*)$ with rate α and order n has density function

$$f_n(x) = \frac{\alpha^n}{(n-1)!}\left(\frac{x}{1-2x}\right)^{n-1} \exp\left(-\alpha \frac{x}{1-2x}\right), \quad x \in [0,1/2)$$

with respect to μ and has density

$$g_n(x) = \frac{\alpha^n}{(n-1)!} \frac{x^{n-1}}{(1-2x)^{n+1}} \exp\left(-\alpha \frac{x}{1-2x}\right), \quad x \in [0, 1/2)$$

with respect to Lebesgue measure λ. The cumulative function of order $n \in \mathbb{N}$ associated with μ is

$$\mu_n(x) = \frac{1}{n!} \left(\frac{x}{1-2x}\right)^n, \quad x \in [0, 1/2)$$

9.7 Applications to reliability

The positive semigroup formulation provides a way to measure the relative aging of one lifetime distribution relative to another. In this section, we fix an interval $I = [a, b)$ with $-\infty < a < b \leq \infty$, and we consider only random variables with continuous distributions on I. If X is such a random variable then the upper probability function of X has the form

$$F(x) = P(X \geq x) = e^{-R(x)}, \quad x \in I$$

where R is a homeomorphism from I onto $[0, \infty)$. If X is interpreted as a random lifetime, then F is called the *reliability function* and R is called the *cumulative failure rate function*. (For a basic introduction to reliability, see [12].) As noted in the last section, if we define

$$x * y = R^{-1}[R(x) + R(y)], \quad x, y \in I$$

then $(I, *)$ is a standard, positive semigroup isomorphic to $([0, \infty), +)$, and X has an exponential distribution on $(I, *)$.

Suppose now that X and Y are random variables on I with cumulative failure rate functions R and S, and semigroup operations $\bullet$ and $*$, respectively. A natural way to study the relative aging of X relative to Y is to study the aging of X on $(I, *)$. Note that the cumulative failure rate function and the reliability function of X are still R and e^{-R}, respectively, when considered on $(I, *)$, because the associated partial order is just the ordinary order $\leq$. Thus, these functions are *invariants*.

First we say that X is *exponential* relative to Y if X has an exponential distribution on $(I, *)$.

Theorem 32. X is exponential relative to Y if and only if $\bullet = *$. The exponential relation defines an equivalence relation on the collection of continuous distributions on I.

Proof. Note that X is exponential relative to Y if and only if $S = cR$ for some positive constant c. □

Next we consider the increasing failure rate property, the strongest of the basic aging properties. If R and S have positive derivatives on I, then the density function of X relative to the invariant measure on $(I, *)$ is

$$f(x) = e^{-R(x)} \frac{R'(x)}{S'(x)}, \quad x \in I$$

Thus, the *failure rate function* of X on $(I, *)$ is obtained by dividing the density function by the reliability function, and hence is R'/S'.

Lemma 4. Show that R'/S' is increasing on I if and only if R is convex on $(I, *)$:

$$R(x * h) - R(x) \le R(y * h) - R(y), \text{ for every } x, y, h \in I \text{ with } x \le y$$

Thus, we will use the convexity condition for our definitions, since it is more general by not requiring that R and S be differentiable. Specifically, we say that X has *increasing failure rate* (IFR) relative to Y if R is convex on $(I, *)$, and X has *decreasing failure rate* (DFR) relative to Y if R is concave on $(I, *)$.

Theorem 33. X has increasing failure rate relative to Y if and only if Y has decreasing failure rate relative to X if and only if

$$(x * h) \bullet y \le (y * h) \bullet x \text{ for all } x, y, h \in I \text{ with } x \le y$$

The IFR relation defines a partial order on the collection of continuous distributions on I, modulo the exponential equivalence in Theorem 32.

Proof. Note that X has increasing failure rate relative to Y if and only if the distribution on $[0, \infty)$ with cumulative rate function $R \circ S^{-1}$ is IFR in the ordinary sense. $\square$

Next, the failure rate average over $[a, x)$ is the cumulative failure rate over $[a, x)$ divided by the length of this interval, as measured by the invariant measure on $(I, *)$. This length is simply $S(x)$, and hence the average failure rate function of X on $(I, *)$ is R/S. We say that X has *increasing failure rate average* (IFRA) relative to Y if R/S is increasing on I and *decreasing failure rate average* (DFRA) relative to Y if R/S is decreasing on I.

Theorem 34. X has increasing failure rate average relative to Y if and only if Y has decreasing failure rate average relative to X if and only if

$$x^{\bullet\alpha} \le x^{*\alpha} \text{ for all } x \in I \text{ and } \alpha \ge 1$$

The IFRA relation defines a partial order on the collection of continuous distributions on I, modulo the exponential equivalence in Theorem 32.

Proof. Note that X has increasing failure rate relative to Y if and only if the distribution on $[0, \infty)$ with cumulative rate function $R \circ S^{-1}$ is IFRA in the ordinary sense. $\square$

Next, X is *new better than used* (NBU) relative to Y if the conditional reliability function of $x^{*(-1)} * X$ given $X \ge x$ is dominated by the reliability function of X, equivalently

$$\mathbb{P}(X \ge x * y) \le \mathbb{P}(X \ge x)\mathbb{P}(X \ge y), \quad x, y \in [a, b) \tag{28}$$

Similarly, X is *new worse than used* (NWU) relative to Y if the inequality in (28) is reversed.

Theorem 35. X is new better than used relative to Y is and only if Y is new worst than used relative to X if and only if

$$x \bullet y \le x * y \text{ for } x, y \in I$$

The NBU relation defines a partial order on the collection of continuous distributions on I, modulo the exponential equivalence in Theorem 32.

Proof. X is new better than used relative to Y if and only if the distribution on $[0, \infty)$ with cumulative rate function $R \circ S^{-1}$ is NBU in the ordinary sense. □

Theorem 36. The relative aging properties for distributions on $[a, b)$ are related as follows:

$$\text{IFR} \Rightarrow \text{IFRA} \Rightarrow \text{NBU}$$

Equivalently, the NBU partial order extends the IFRA partial order, which in turn extends the IFR partial order.

Proof. It's well known that the standard aging properties on $[0, \infty)$ are related as stated. □

9.8 More examples

We consider several two-parameter families of distributions. In each case, α is the exponential parameter while β is an "aging parameter" that determines the relative aging.

Example 14. Let $I = [0, \infty)$ and for fixed $\beta > 0$, let R be given by $R(x) = x^\beta$. The corresponding semigroup operator $*$ is given by

$$x * y = (x^\beta + y^\beta)^{1/\beta}, \quad x, y \in [0, \infty)$$

The exponential distribution on $(I, *)$ with rate $\alpha > 0$ has reliability function

$$F(x) = \exp(-\alpha x^\beta), \quad x \in [0, \infty)$$

Of course, this is a Weibull distribution with shape parameter β. In the usual formulation, the rate parameter α is written as $\alpha = R(c) = c^\beta$ where $c > 0$ is the scale parameter.

In the ordinary sense, the Weibull distribution has decreasing failure rate if $0 < \beta < 1$, is exponential if $\beta = 1$, and has increasing failure rate if $\beta > 1$. A stronger statement, from Theorem 33 is that the Weibull distribution with shape parameter β_1 is IFR relative to the Weibull distribution with shape parameter β_2 if and only if $\beta_1 \le \beta_2$.

Example 15. Let $I = [0, \infty)$ and for fixed $\beta > 0$, let R be given by $R(x) = e^{\beta x} - 1$. The corresponding semigroup operator $*$ is given by

$$x * y = \frac{1}{\beta} \ln(e^{\beta x} + e^{\beta y} - 1), \quad x, y \in [0, \infty)$$

The exponential distribution on $(I, *)$ with rate $\alpha > 0$ has reliability function

$$F(x) = \exp[-\alpha(e^{\beta x} - 1)], \quad x \in [0, \infty)$$

This is the modified extreme value distribution with parameters α and β.

In the ordinary sense, these distributions are IFR for all parameter values. On the other hand, a distribution with parameter β_1 is IFR with respect to a distribution with parameter β_2 if and only if $\beta_1 \le \beta_2$.

Example 16. Let $I = [0, \infty)$ and for fixed $\beta > 0$, let R be given by $R(x) = \ln(x + \beta) - \ln(\beta)$. The corresponding semigroup operator $*$ is given by

$$x * y = x + y + \frac{xy}{\beta}, \quad x, y \in [0, \infty)$$

The exponential distribution on $(I, *)$ with rate $\alpha > 0$ has reliability function

$$F(x) = \left(\frac{\beta}{x + \beta}\right)^{\alpha}, \quad x \in [0, \infty)$$

This is a two-parameter family of Pareto distribution.

In the ordinary sense, these distributions are DFR for all parameter values. On the other hand, a distribution with parameter β_1 is DFR with respect to a distribution with parameter β_2 if and only if $\beta_1 \le \beta_2$.

Example 17. Let $I = [0, 1)$ and for fixed $\beta > 0$, let R be given by $R(x) = -\ln(1 - x^{-\beta})$. The corresponding semigroup operator $*$ is given by

$$x * y = (x^\beta + y^\beta - x^\beta y^\beta)^{1/\beta}, \quad x, y \in [0, 1)$$

The exponential distribution on $(I, *)$ with rate $\alpha > 0$ has reliability function

$$F(x) = \left(1 - x^\beta\right)^{\alpha}, \quad x \in [0, 1)$$

Note that if $\alpha = 1$ or $\beta = 1$, the distribution is beta; if $\alpha = \beta = 1$, the distribution is uniform.

In the ordinary sense, these distributions are IFR if $\beta \ge 1$, but are neither NBU nor NWU if $0 < \beta < 1$. On the other hand, a distribution with parameter β_1 is DFR with respect to a distribution with parameter β_2 if and only if $\beta_1 \le \beta_2$.

9.9 Strong aging properties

Consider an "aging property", and the corresponding improvement property, for continuous distributions in the standard semigroup $([0,\infty),+)$. We are interested in characterizing those aging properties that can be extended to relative aging properties for continuous distributions on an arbitrary interval $[a,b)$. Moreover, we want the relative aging property to define a partial order on the distributions, modulo the exponential equivalence in Theorem 32, just as the IFR, IFRA, and NBU properties do. Such a characterization would seem to describe "strong" aging properties.

Definition 41. A strong aging property satisfies the following conditions:

1. A distribution both ages and improves if and only if the distribution is exponential.
2. The distribution with cumulative rate R ages if and only if the distribution with cumulative rate R^{-1} improves.
3. If the distributions with cumulative rates R and S age, then the distribution with cumulative rate $R \circ S$ ages.
4. If a distribution is IFR then the distribution ages.

The last condition is to ensure that the property does capture *some* idea of aging, and to incorporate the fact that the IFR condition is presumably the strongest aging property.

Suppose now that X and Y are random variables with continuous distributions on $[a,b)$ having cumulative rate functions R and S, respectively. We say that X ages (improves) relative to Y if the distribution on $[0,\infty)$ with cumulative rate function $R \circ S^{-1}$ ages (improves).

Theorem 37. Consider a strong aging property. The corresponding aging relation defines a partial order on the equivalence class of continuous distributions on $[a,b)$, modulo the exponential equivalence.

Proof. The proof is straightforward. □

Theorem 38. The IFR, IFRA, and NBU properties are strong aging properties.

Proof. This follows from our previous theorems on IFR, IFRA, and NBU.. □

From our point of view, the conditions in Definition 41 are natural for a strong aging property. Conditions 1 and 4 are often used in the literature to justify aging properties, but Conditions 2 and 3 seem to have been overlooked, even though they are essential for the partial order result in Theorem 37.

Not all of the common aging properties are strong. A random variable X with a continuous distribution on $[0,\infty)$ is *new better than used in expectation* (NBUE) if

$$E(X - t \mid X \geq t) \leq E(X), \quad t \in [0,\infty)$$

and of course is *new worse than used in expectation* (NWUE) if the inequality is reversed.

Proposition 68. NBUE is not a strong aging property.

Proof. Define $R : [0, \infty) \to [0, \infty)$ by

$$R(t) = \begin{cases} at, & 0 \le t < 1 \\ a + b(t-1), & 1 \le t < 2 \\ a + b(t-1) + c(t-2), & t \ge 2 \end{cases}$$

Positive constants a, b, and c can be chosen such that the distribution with cumulative rate function R is NBUE, but the distribution with cumulative rate function R^{-1} is not NWUE. ☐

9.10 Minimums of exponential distributions

If X and Y are independent variables which are exponential in $([0, \infty), *)$ and have failure rates α and β with respect to λ, then $X \wedge Y$ is exponential in $([0, \infty), *)$ and has failure rate $\alpha+\beta$. If $\alpha = \Phi(a)$, $\beta = \Phi(b)$ then $\alpha+\beta = \Phi(a*b)$.

Proof. The variables X and Y have upper probability functions F and G given by

$$F(x) = \exp[-\alpha\Phi(x)] \quad G(x) = \exp[-\beta\Phi(x)]$$

Hence $X \wedge Y$ has upper probability function FG which can be written

$$F(x)G(x) = \exp[-(\alpha + \beta)\Phi(x)]$$

and hence $X \wedge Y$ is exponential in $([0, \infty), *)$ with failure rate $\alpha + \beta$. The last statement follows from the definition of $*$. ☐

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

10.1 Definitions

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

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

and the identity element is 1. Of course, we use counting measure as the invariant measure.

Let I denote the set of prime numbers; these are the irreducible elements of $(\mathbb{N}_+, \cdot)$. Each $x \in \mathbb{N}_+$ has the canonical prime factorization

$$x = \prod_{i \in I} i^{n_i}$$

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

$$M = \{\boldsymbol{n} = (n_i : i \in I) : n_i \in \mathbb{N} \text{ for } i \in I, \ n_i = 0 \text{ eventually in } i\} \tag{29}$$

and where $+$ is componentwise addition. An isomorphism is

$$(n_i : i \in I) \mapsto \prod_{i \in I} i^{n_i}$$

Proposition 69. $\dim(\mathbb{N}_+, \cdot)$ is ∞.

Proof. By Proposition 7, $\dim(\mathbb{N}_+, \cdot) \leq \#(I) = \infty$. Suppose that $C \subset \mathbb{N}_+$ is finite. Let D denote the set of all primes in the prime factorings of the elements of C. Of course, D is also finite. Define $\varphi(i) = 0$ for $i \in D$ and $\varphi(i) = 1$ for $i \in I - D$. Extend φ to $\mathbb{N}_+$ by

$$\varphi(i_1 i_2 \cdots i_n) = \varphi(i_1) + \varphi(i_2) + \cdots + \varphi(i_n), \quad i_i, i_2, \ldots i_n \in I$$

This definition is consistent, since the factoring of x over I is unique, except for the ordering of the factors. Thus φ is a homomorphism from $(\mathbb{N}_+, \cdot)$ into $(\mathbb{R}, +)$, $\varphi(x) = 0$ for $x \in C$, but φ is not identically 0. $\square$

On the other hand, since $(\mathbb{N}_+, \preceq)$ is an upper semilattice, the distributional dimension is 1. That is, a distribtion on $\mathbb{N}_+$ is uniquely determined by its upper probability function.

For $J \subset I$, let $\mathbb{N}_J$ denote the sub-semigroup generated by J. The elements of $\mathbb{N}_J$ are positive integers whose prime factors are in J. Then $(\mathbb{N}, \cdot)$ is isomorphic to the direct product of $(\mathbb{N}_J, \cdot)$ and $(\mathbb{N}_{I-J}, \cdot)$.

Recall that the cumulative function of order k at $x \in \mathbb{N}_+$ gives the number of $k+1$ factorings of x. This is an important function in number theory, so we deviate from our standard notation and denote it by τ_k instead of $\#_k$. In particular, $\tau_1(x) = \#\{u \in \mathbb{N}_+ : u \preceq x\}$ is the number of divisors of x. In terms of the canonical factorization,

$$\tau_k\left(\prod_{i \in I} i^{n_i}\right) = \prod_{i \in I} \binom{k + n_i}{k}, \quad (n_i : i \in I) \in M$$

It's well known [23] that τ_1 is *multiplicative*, that is $\tau_1(xy) = \tau_1(x)\tau_1(y)$ if x and y are relatively prime . It follows that τ_k is multiplicative for each $k \in \mathbb{N}$.

Problem 18. Find a closed form expression for the generating function

$$\Lambda(x, t) = \sum_{n=0}^{\infty} \tau_n(x) t^n$$

If X is a random variable taking values in $\mathbb{N}_+$ then from Theorem 5 we have the interesting result

$$\sum_{x=1}^{\infty} \tau_k(x) P(X \succeq x) = E[\tau_{k+1}(X)]$$

10.2 Exponential distributions

We will first characterize the exponential distributions in terms of the prime factorization.

Theorem 39. A distribution is exponential on $(\mathbb{N}_+, \cdot)$ if and only if it has a upper probability function of the form

$$F\left(\prod_{i \in I} i^{n_i}\right) = \prod_{i \in I} (1 - p_i)^{n_i}, \quad (n_i : i \in I) \in M$$

where $p_i \in (0, 1)$ for each $i \in I$ and $\prod_{i \in I} p_i > 0$. This product is the rate constant.

Proof. The memoryless property for a upper probability function F is

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

In the language of number theory, F is *completely multiplicative.* It follows that

$$F\left(\prod_{i \in I} i^{n_i}\right) = \prod_{i \in I} [F(i)]^{n_i}, \quad \boldsymbol{n} \in M$$

Let $F(i) = 1 - p_i$ where $p_i \in (0, 1)$ for each $i \in I$. Then

$$F\left(\prod_{i \in I} i^{n_i}\right) = \prod_{i \in I} (1 - p_i)^{n_i}, \quad \boldsymbol{n} \in M$$

Next note that

$$\begin{aligned} \sum_{x \in \mathbb{N}_+} F(x) &= \sum_{\boldsymbol{n} \in M} F\left(\prod_{i \in I} i^{n_i}\right) = \sum_{\boldsymbol{n} \in M} \prod_{i \in I} (1 - p_i)^{n_i} \\ &= \prod_{i \in I} \sum_{n=0}^{\infty} (1 - p_i)^n = \prod_{i \in I} \frac{1}{p_i} \end{aligned}$$

Thus, the result follows from Theorem 26. The probability density function of this distribution is

$$f\left(\prod_{i \in I} i^{n_i}\right) = \prod_{i \in I} p_i (1 - p_i)^{n_i}, \quad (n_i : i \in I) \in M$$

∎

The exponential distribution given in Theorem 39 corresponds to independent, geometric distributions on the prime exponents. That is, if X is a random variable with this distribution then

$$X = \prod_{i \in I} i^{N_i} \tag{30}$$

where $(N_i : i \in I)$ are independent random variables and N_i has the geometric distribution with rate parameter p_i for each $i \in I$. This characterization of the exponential distributions could also be obtained from Section 6.5 and the identification of $(\mathbb{N}_+, \cdot)$ with the semigroup $(M, +)$ in (29).

The strong exponential property has the following interpretation: The conditional distribution of X/x given that x divides X is the same as the distribution of X. Thus, knowledge of one divisor of X does not help in finding other divisors of X. This property may have some practical applications.

Note 47. That X is a well defined random variable also follows from the Borel-Cantelli lemma. With probability 1, $N_i = 0$ for all but finitely many $i \in I$.

We will now give a different characterization of the exponential distributions. First recall that a *Dirichlet series* is a series of the form

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

where $a : \mathbb{N}_+ \to [0, \infty)$ is an *arithmetic function*. If the Dirichlet series converges for some $s > 0$, then 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:

$$A(s) = \prod_{i \in I} \frac{1}{1 - a(i) i^{-s}}, \quad s > s_0$$

There is a one-to-one correspondence between the coefficient function a and the series function A. Given a, we compute A, of course, as the infinite series in the definition. Conversely, given A defined on (s_0, ∞), we can recover the coefficient function a (see [11]).

Given the coefficient function a (or equivalently the series 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. The probability density function of the distribution with parameter $s > s_0$ is proportional to $a(x)/x^s$; of course the proportionality constant must then be $A(s)$. Thus, X has the Dirichlet distribution corresponding to a with parameter s if

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

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*:

$$\zeta(s) = \sum_{x=1}^{\infty} \frac{1}{x^s} = \prod_{i \in I} \frac{1}{1 - i^{-s}}, \quad s > 1$$

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

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

Theorem 40. A distribution on $(\mathbb{N}_+, \cdot)$ is exponential if and only if it is a member of a Dirichlet family of distributions with a strictly positive, completely multiplicative coefficient function.

Proof. Suppose that a is strictly positive and completely multiplicative and that X has the Dirichlet distribution corresponding to a with parameter s. Then

$$\begin{aligned}
\mathbb{P}(X \succeq x) &= \sum_{y \succeq x} \frac{a(y)y^{-s}}{A(s)} = \sum_{z=1}^{\infty} \frac{a(xz)(xz)^{-s}}{A(s)} \\
&= \sum_{z=1}^{\infty} \frac{a(x)a(z)x^{-s}z^{-s}}{A(s)} = \frac{a(x)x^{-s}}{A(s)} \sum_{z=1}^{\infty} a(z)z^{-s} \\
&= \frac{a(x)x^{-s}}{A(s)} A(s) = a(x)x^{-s}, \quad x \in \mathbb{N}_+
\end{aligned}$$

Hence X has constant rate. Also, X is memoryless, since a is completely multiplicative:

$$\mathbb{P}(X \succeq xy) = a(xy)(xy)^{-s} = a(x)a(y)x^{-s}y^{-s} = \mathbb{P}(X \succeq x)\mathbb{P}(X \succeq y)$$

Therefore X has an exponential distribution.

In a sense, the converse is trivially true. Suppose that X has an exponential distribution with upper 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

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

and so X has the Dirichlet distribution corresponding to a with parameter t. Note that since a is completely multiplicative, all members of this Dirichlet family are exponential, from the first part of the theorem. ☐

Thus, a Dirichlet distribution with completely multiplicative coefficient function (in particular, the zeta distribution) has the representation given in (30). The geometric parameters for the random prime exponents are given by

$$1 - p_i = \mathbb{P}(X \succeq i) = \frac{a(i)}{i^s}, \quad i \in I$$

This result was considered surprising by Lin and Hu [18], but is quite natural in the context of positive semigroups.

We give yet another representation in terms of independent Poisson variables. This was also obtained in [18], but we give an alternate derivation based on the exponential distribution.

Proposition 70. Suppose that X has the exponential distribution with parameter vector $(p_i \colon i \in I)$. For $x \in \{2, 3, \ldots\}$, let $\lambda_x = (1 - p_i)^n / n$ if $x = i^n$ for some $i \in I$ and $n \in \mathbb{N}_+$, and let $\lambda_x = 0$ otherwise. Then X can be written in the form

$$X = \prod_{x=2}^{\infty} x^{V_x}$$

where V_x has the Poisson distribution with parameter λ_x, and $(V_2, V_3 \ldots)$ are independent.

Proof. We start with the representation (30). By Proposition 67, we can write

$$N_i = \sum_{n=1}^{\infty} n V_{in}$$

where $\{V_{in} : i \in I,\, n \in \mathbb{N}_+\}$ are independent and V_{in} has the Poisson distribution with parameter $(1 - p_i)^n / n$. Substituting we have

$$X = \prod_{i \in I} \prod_{n=1}^{\infty} i^{n V_{in}} = \prod_{i \in I} \prod_{n=1}^{\infty} (i^n)^{V_{in}}$$

Now, for $x \in \{2, 3, \ldots\}$, let $V_x = V_{in}$ if $x = i^n$ for some $i \in I$ and $n \in \mathbb{N}_+$, and let $V_x = 0$ otherwise. Then $(V_2, V_3, , \ldots)$ are independent, V_x has the Poisson distribution with parameter λ_x given in the proposition, and

$$X = \prod_{x=2}^{\infty} x^{V_x}$$

□

The following corollary restates Proposition 70 in the language of Dirichlet distributions. Recall that the *Mangoldt function* $\Lambda : \mathbb{N}_+ \to (0, \infty)$ is defined as follows:

$$\Lambda(x) = \begin{cases} \ln(i) & \text{if } x = i^n \text{ for some } i \in I \text{ and } n \in \mathbb{N}_+ \\ 0 & \text{otherwise} \end{cases}$$

Corollary 22. Suppose that X has the Dirichlet distribution with completely multiplicative coefficient function a and parameter s. Then X can be decomposed as

$$X = \prod_{x=2}^{\infty} x^{V_x}$$

where $(V_2, V_3, \ldots)$ are independent and V_x has the Poisson distribution with parameter

$$\lambda_x = \frac{a(x)\Lambda(x)}{\ln(x) x^s}$$

10.3 Constant rate distributions

New constant rate distributions for $(N_+, \preceq)$ can be constructed by mixing exponential distributions with the same rate. In general, these distributions will not be memoryless and hence not exponential.

To illustrate the simplest possible case, suppose that X_1 has the exponential distribution with parameter vector $(p_i : i \in I)$, X_2 has the exponential distribution with parameter vector $(q_i : i \in I)$, and that

$$\prod_{i \in I} p_i = \prod_{i \in I} q_i \tag{31}$$

so that the rate constants of X_1 and X_2 agree. Of course, one way that (31) could happen would be for $(q_i : i \in I)$ to be a permutation of $(p_i : i \in I)$. We will let α denote the common product in (31). Now let X be the random variable whose distribuiton is the mixture of the distributions of X_1 and X_2, with mixing parameter $r \in (0, 1)$. By Theorem 13, X also has constant rate α and upper probability function

$$F(x) = r \prod_{i \in I} (1 - p_i)^{n_i} + (1 - r) \prod_{i \in I} (1 - q_i)^{n_i}$$

Of course, F will not in general be completely multiplicative (memoryless).

Problem 19. Are all constant rate distributions on $(\mathbb{N}_+, \preceq)$ mixtures of exponential distributions on $(\mathbb{N}_+, \cdot)$?

10.4 Moments

In this section, we assume that X has an exponential distribution on $(\mathbb{N}_+, \cdot)$. Thus, X can be characterized in terms of its geometric parameters $(p_i : i \in I)$, as in Theorem 39, or in terms of a Dirichlet series A with completely multiplicative coefficient function a, interval of convergence (s_0, ∞), and a parameter $s > s_0$, as in Theorem 40.

From Proposition 8, the expected number of divisors of X is

$$\mathbb{E}[\tau_1(X)] = \prod_{i \in I} \frac{1}{p_i} = A(s)$$

In the special case that X has a zeta distribution with parameter s, the expected number of divisors of X is

$$\zeta(s) = \mathbb{E}[\tau_1(X)]$$

From Theorem 10, X maximizes entropy over all random variables

$$Y = \prod_{i \in I} i^{U_i} \in \mathbb{N}_+$$

with $\mathbb{E}(U_i) = (1 - p_i)/p_i$ for each $i \in I$, The maximum entropy is

$$H(X) = -\sum_{i \in I} \left(\ln(p_i) + \ln(1 - p_i) \frac{1 - p_i}{p_i} \right)$$

Equivalently, X maximizes entropy over all random variables $Y \in \mathbb{N}_+$ with $\mathbb{E}[\ln(Y)] = \mathbb{E}[\ln(X)]$ and $\mathbb{E}[\ln(a(Y))] = \mathbb{E}[\ln(a(X))]$. But from [11],

$$\mathbb{E}[\ln(X)] = \frac{1}{A(s)} \sum_{x=1}^{\infty} \ln(x) a(x) x^{-s} = \frac{A'(s)}{A(s)}$$

Note 48. If $p_i < 1/i^n$ for all $i \in I$ then

$$\mathbb{E}(X^n) = \prod_{i \in I} \frac{1 - p_i}{1 - i^n p_i}$$

Equivalently, if $s > s_0 + n$ then

$$\mathbb{E}(X^n) = \frac{A(s - n)}{A(s)}$$

For the special case of the zeta distribution with parameter s we have the standard result

$$\mathbb{E}(X^n) = \frac{\zeta(s - n)}{\zeta(s)} \text{ if } s > n + 1$$

Next we obtain a result from [18]. Our proof is better because it takes advantage of the general theory of positive semigroups.

Proposition 71. Suppose that $b : \mathbb{N}_+ \to [0, \infty)$ is a nonnegative arithmetic function, not identically zero. Let B be the corresponding Dirichlet function, which we assume converges for $t > t_0$, and recall that $L(b)(x) = \sum_{y \preceq x} b(y)$ for $x \in \mathbb{N}_+$. Suppose that X has the zeta distribution with parameter $s > \max\{t_0, 1\}$. Then

$$\mathbb{E}[L(b)(X)] = B(s)$$

Proof. It follows immediately from Proposition 8 that

$$\mathbb{E}[L(b)(X)] = \zeta(s) \mathbb{E}[b(X)]$$

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

$$\zeta(s) \mathbb{E}[b(X)] = \zeta(s) \sum_{x=1}^{\infty} \frac{b(x)}{x^s \zeta(s)} = \sum_{x=1}^{\infty} \frac{b(x)}{x^s} = B(s)$$

□

10.5 Gamma distributions

Now suppose that X_n, $n \in \mathbb{N}_+$ are independent variables, each with the exponential distribution with parameters $(p_i : i \in I)$ as in Theorem 39. Let $Y_n = X_1 \cdots X_n$ be the corresponding gamma variable of order n. It follows immediately that

$$Y_n = \prod_{i \in I} i^{U_{ni}}$$

where U_{ni} has the negative binomial distribution with parameters n and p_i, and where U_{ni}, $i \in I$ are independent. Hence

$$\mathbb{P}\left(Y_n = \prod_{i \in I} i^{k_i}\right) = \prod_{i \in I}^{\infty} \binom{n + k_i - 1}{k_i} (1 - p_i)^{k_i} p_i^n, \quad (k_i : i \in I) \in M$$

Of course, the density function of the gamma variable Y_n is a special case of Proposition 51

$$\mathbb{P}(Y_n = x) = \left(\prod_{i \in I} p_i^n\right) \tau_{n-1}(x) F(x), \quad x \in \mathbb{N}_+$$

where F is the upper probability function of the exponential distribution.

We can reformulate these results in the notation of the Dirichlet distributions. Specifically, suppose that X has the Dirichlet distribution corresponding to the completely multiplicative coefficient function a, Dirichlet function A, and parameter s. Let $\boldsymbol{X} = (X_1, X_2, \ldots)$ be an IID sequence with the same distribution as X. Let $\boldsymbol{Y} = (Y_1, Y_2, \ldots)$ denote the corresponding sequence of ladder variables, or equivalently, the corresponding sequence of partial products. Then $(Y_1, Y_2, \ldots, Y_n)$ has PDF h_n given by

$$h_n(y_1, y_2, \ldots, y_n) = \frac{a(y_n)}{A^n(s) y_n^s}, \quad (y_1, y_2, \ldots, y_n) \in D_n$$

The Markov chain $\boldsymbol{Y}$ has transition density function g given by

$$g(y, yz) = f(z) = \frac{a(z)}{A(s) z^s}, \quad y,\, z \in \mathbb{N}_+$$

Finally, Y_n has PDF g_n given by

$$g_n(y) = \frac{\tau_{n-1}(y) a(y)}{A^n(s) y^s}, y \in \mathbb{N}_+$$

Thus, Y_n also has a Dirichlet distribution, but corresponding to a multiplicative coefficient function instead of a *completely* multiplicative coefficient function. It follows that

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

In the special case of the zeta distribution with parameter s, we have

$$g_n(y) = \frac{\tau_{n-1}(y)}{\zeta^n(s) y^s}, \quad y \in \mathbb{N}_+$$

so it follows that

$$\sum_{x=1}^{\infty} \frac{\tau_{n-1}(x)}{x^s} = \zeta^n(s), \quad n \in \mathbb{N}_+, \, s > 1$$

10.6 Compound Poisson distributions

Suppose that X has the exponential distribution on $(\mathbb{N}_+, \cdot)$ with parameter vector $(p_i : i \in I)$. According to Gut [11], X has a compound Poisson distribution. In our notation, X can be decomposed as

$$X = V_1 V_2 \cdots V_N$$

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

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

The random index N is independent of $(V_1, V_2, \ldots)$ and has the Poisson distribution with parameter $\ln(\mathbb{E}(\tau(X)))$.

Problem 20. Interpret this result in the context of positive semigroups.

11 Lexicographic Sums

For $n \in \mathbb{N}$, let A_n be set with k_n points, where $k_n \in \mathbb{N}_+$ for $n \in \mathbb{N}_+$. Now let $(S, \preceq)$ be the lexicographic sum of the anti-chains $\{(A_n, =) : n \in \mathbb{N}\}$ over $(\mathbb{N}, \leq)$. Thus, for $(n, a) \in S$, $(m, b) \in S$. we have $(n, a) \prec (m, b)$ if and only if $n < m$. The elements of $\{0\} \times A_0$ are minimal, and in particular if $n_0 = 1$, the element $(0, e)$ is the minimum element (where $A_0 = \{e\}$). Note also that the equivalence classes under the upper equivalence relation are $\{n\} \times A_n$ for $n \in \mathbb{N}$.

11.1 Constant rate distributions

Now let f be a probability density function on S and F the corresponding upper probability function. Then

$$F(n, a) = f(n, a) + \sum_{m=n+1}^{\infty} \sum_{b \in A_m} f(m, b)$$

Equivalently,

$$F(n,a) - f(n,a) = \sum_{m=n+1}^{\infty} \sum_{b \in A_m} f(m,b)$$

Suppose now that the distribution has constant rate α. Then $f = \alpha F$ and $F(n,a)$ is constant over $a \in A_n$; let p_n denote this constant value. Then we have

$$p_n - \alpha p_n = \sum_{m=n+1}^{\infty} \sum_{b \in A_m} \alpha p_m = \sum_{m=n+1}^{\infty} \alpha k_m p_m$$

Solving we have

$$p_n = \frac{\alpha}{1-\alpha} \sum_{m=n+1}^{\infty} k_m p_m$$

It then follows that $p_n = \frac{\alpha}{1-\alpha} k_{n+1} p_{n+1} + p_{n+1}$ or equivalently

$$p_{n+1} = \frac{1-\alpha}{\alpha k_{n+1} + 1 - \alpha} p_n, \quad n \in \mathbb{N} \tag{32}$$

Solving (32) gives

$$p_n = \frac{(1-\alpha)^n}{(1-\alpha+\alpha k_1)(1-\alpha+\alpha k_2)\cdots(1-\alpha+\alpha k_n)} p_0, \quad n \in \mathbb{N}$$

Now fix $a \in A_0$. Then

$$F(0,a) + \sum_{b \in A_0 - \{a\}} f(0,b) = 1$$

But $F(0,a) = p_0$ and $f(0,b) = \alpha p_0$ for a, $b \in A_0$. Therefore $p_0 + (k_0 - 1)\alpha p_0 = 1$. Hence

$$p_0 = \frac{1}{1 - \alpha + k_0 \alpha}$$

Thus, the upper probability function and probability density function of the distribution with constant rate α are

$$F(n,a) = \frac{(1-\alpha)^n}{(1-\alpha+\alpha k_0)(1-\alpha+\alpha k_1)\cdots(1-\alpha+\alpha k_n)}, \quad (n,a) \in S$$

$$f(n,a) = \frac{\alpha(1-\alpha)^n}{(1-\alpha+\alpha k_0)(1-\alpha+\alpha k_1)\cdots(1-\alpha+\alpha k_n)}, \quad (n,a) \in S$$

Now suppose that (X,Y) is a random variable taking values in S with this constant rate distribution. The probability density function of X is

$$g(n) = \sum_{a \in A_n} f(n,a)$$

$$g(n) = \frac{\alpha(1-\alpha)^n k_n}{(1-\alpha+\alpha k_0)(1-\alpha+\alpha k_1)\cdots(1-\alpha+\alpha k_n)}, \quad n \in \mathbb{N} \tag{33}$$

On the other hand, the conditional density of Y given $X = n$ has density function

$$h(a|n) = \frac{f(n,a)}{g(n)} = \frac{1}{k_n}, \quad a \in A_n,\ n \in \mathbb{N}$$

so given $X = n$, Y is uniformly distributed on A_n

We can verify directly that g in (33) is a probability density function on $\mathbb{N}$. Writing αk_n in the numerator of $g(n)$ as $1 - \alpha + \alpha k_n - (1-\alpha)$ and expanding we have

$$g(n) = \frac{(1-\alpha)^n}{(1-\alpha+\alpha k_0)\cdots(1-\alpha+\alpha k_{n-1})} - \frac{(1-\alpha)^{n+1}}{(1-\alpha+\alpha k_0)\cdots(1-\alpha+\alpha k_n)}$$

so $\sum_{n=0}^\infty g(n)$ is a collapsing sum that reduces to 1. It then also follows that the upper probability function of X on $\mathbb{N}$ is

$$G(n) = \frac{(1-\alpha)^n}{(1-\alpha+\alpha k_0)\cdots(1-\alpha+\alpha k_{n-1})}, \quad n \in \mathbb{N}$$

Hence the rate function is

$$\frac{g(n)}{G(n)} = \frac{\alpha k_n}{1-\alpha+\alpha k_n}, \quad n \in \mathbb{N}$$

If k_n is increasing in n, decreasing in n, or constant in n, then X has increasing rate, decreasing rate, or constant rate, respectively. In the latter case, of course, X has a geometric distribution. In the decreasing case, k_n must eventually be constant in n

Equation (33) defines an interesting class of distributions on $\mathbb{N}$. Here are some special cases:

Example 18. If $k_n = k$ for all $n \in \mathbb{N}$, then (as noted above) X has the geometric distribution with rate parameter $\alpha k/(1-\alpha+\alpha k)$. In particular, if $k = 1$, X has the geometric distribution with parameter α.

Example 19. If $k_n = n+1$ for $n \in \mathbb{N}$ we get

$$g(n) = \frac{\alpha(1-\alpha)^n(n+1)}{(1)(1+\alpha)(1+2\alpha)\cdots(1+n\alpha)} = \frac{\alpha(1-\alpha)^n(n+1)}{[1,\alpha]_{n+1}}, \quad n \in \mathbb{N}$$
$$G(n) = \frac{(1-\alpha)^n}{(1)(1+\alpha)\cdots[1+(n-1)\alpha]} = \frac{(1-\alpha)^n}{[1,\alpha]_n}, \quad n \in \mathbb{N}$$

where we are using the generalized permutation noation:

$$[a,s]_j = a(a+s)\cdots(a+(j-1)s)$$

for a, $s \in \mathbb{R}$ and $j \in \mathbb{N}$. It's easy to see that $g(n+1) \ge g(n)$ if and only if $n \le \sqrt{(1-\alpha)/\alpha}$ and hence the distribution is unimodal with mode at $\lfloor \sqrt{(1-\alpha)/\alpha} \rfloor$.

Example 20. If $\alpha = 1/2$ we get

$$g(n) = \frac{k_n}{(1+k_0)(1+k_1)\cdots(1+k_n)}, \quad n \in \mathbb{N}$$
$$G(n) = \frac{1}{(1+k_0)\cdots(1+k_{n-1})}, \quad n \in \mathbb{N}$$

Example 21. If both $k_n = n+1$ and $\alpha = 1/2$ we get

$$g(n) = \frac{n+1}{(n+2)!} = \frac{1}{(n+1)!} - \frac{1}{(n+2)!}, \quad n \in \mathbb{N}$$
$$G(n) = \frac{1}{(n+1)!}, \quad n \in \mathbb{N}$$

An easy computation shows that the probability generating function is

$$\mathbb{E}(t^X) = \frac{te^t - e^t + 1}{t^2}$$

11.2 Positive semigroups

In the general lexicographic construction considered in this section, only one case corresponds to a positive semigroup. Let $k_0 = 1$ and $k_n = k$ (constant) for all $n \in \mathbb{N}_+$. Thus, we can take $A_n = A$ for all $n \in \mathbb{N}_+$ where A is a fixed set with k elements. In this case, $(S, \preceq)$ has the self-similarity property that characterizes a positive semigroup. The correct semigroup operation is $(n,a)(m,b) = (n+m,b)$ for n, $m \in \mathbb{N}_+$ and $a, b \in A$. Of course, $(0,e)$ is the identity where $A_0 = \{e\}$. Note that we are in the setting of Example 3.

The memoryless property for a upper probability function F is

$$F(n,a)F(m,b) = F(n+m,b), \quad (n,a) \in S,\ (m,b) \in S - \{(0,e)\}$$

In particular, for each $a \in A$,

$$F(n,a)F(m,a) = F(n+m,a), \quad n,\ m \in \mathbb{N}_+$$

It follows that for each $a \in A$, there exists $q_a \in (0,1)$ such that $F(x,a) = q_a^x, x \in \mathbb{N}_+$. But then another application of the memoryless condition gives $q_a^n q_b^m = q_b^{n+m}$ for any $n, m \in \mathbb{N}_+$ and $a, b \in A$ and therefore $q_a = q_b$ for all a, b. We will denote the common value by q so that $F(n,a) = q^n$ for any $n \in \mathbb{N}_+$ and any $a \in A$.

Let f be a probability mass function associated with this upper probability function. For $a, b \in A$,

$$F(n,a) = f(n,a) + \sum_{m=n+1}^{\infty} \sum_{u \in A} f(m,u)$$
$$F(n,b) = f(n,b) + \sum_{m=n+1}^{\infty} \sum_{u \in A} f(m,u)$$

Since $F(n, a) = F(n, b)$, if follows that $f(n, a) = f(n, b)$ for $n \in \mathbb{N}_+$ and $a, b \in A$. Thus, let $\varphi(n)$ denote the common value of $f(n, a)$ for any $a \in A$ and let $\varphi(0) = f(0, e)$. It follows that

$$q^n = \varphi(n) + k \sum_{m=n+1}^{\infty} \varphi(m)$$

for any $n \in \mathbb{N}$. Subtracting gives

$$(1 - q)q^n = \varphi(n) + (k - 1)\varphi(n + 1)$$

Using this result recursively gives an explicit formula for the probability mass function of a memoryless distribution:

$$f(n, a) = \frac{(1 - q)q^n}{1 - q + kq} + \frac{(-1)^n}{(k - 1)^n} \left[\frac{-(1 - q)}{1 - q + kq} + f(0,\, e) \right], \quad (n, a) \in S$$

If we pick $f(0,\, 0) = (1 - q)/(1 - q + kq)$ then

$$f(n, a) = \frac{(1 - q)q^n}{1 - q + kq}, \quad n \in \mathbb{N}_+, a \in A$$

This is the probability mass function of the distribution with constant failure $\alpha = (1 - q)/(q - q + kq)$.

The choice of $f(0, e)$ above is not the only possible one. Indeed, any choice of $f(0, e)$ such that $f(n, a) \geq 0$ for all $(n, a) \in S$ will lead to a probability mass function whose corresponding upper probability function is $F(n, a) = q^n$. In particular, the upper probability function does not uniquely specify the probability mass function, even for a countable, locally finite semigroup like this one.

For example, suppose that $k = 4$ and that $q = 1/2$. The memoryless probability mass function

$$f(n, a) = \frac{1}{5} \left[\left(\frac{1}{2} \right)^n + \left(\frac{-1}{3} \right)^n \right], \quad (n, a) \in S$$

has upper probability function

$$F(n,\, a) = \left(\frac{1}{2} \right)^n \text{ for } (n, a) \in S$$

On the other hand, the exponential probability mass function

$$g(n, a) = \frac{1}{5} \left(\frac{1}{2} \right)^n, \quad (n, a) \in S$$

has the same upper probability function F.

Suppose that F is the upper probability function of a distribution which has constant rate $\alpha \in (0, 1)$. Then

$$F(n, a) = \frac{(1-\alpha)^n}{(1)(1-\alpha+\alpha k)^n} = \left(\frac{1-\alpha}{1-\alpha+\alpha k}\right)^n, \quad (n, a) \in S$$

Hence the distribution is memoryless as well, and hence exponential. Thus, every distribution with constant rate is memoryless (and hence exponential), but conversely, there are memoryless distributions that do not have constant rate.

This example shows that a semigroup satisfying the basic assumptions but which cannot be embedded in a group may still have exponential distributions.

Of course, if $k = 1$ then $(S, \cdot)$ is isomorphic to $(\mathbb{N}, +)$ and the distribution defined by g in (33) is the geometric distribution with rate α.

12 Trees

In this section we consider a standard discrete poset $(S, \preceq)$ whose covering graph is an ordered tree. Our interest will mostly be in the special case where S has a minimum element e, so that the covering graph is a rooted tree with root e. In this case, $B(x)$ has a single element, which we will denote by x^- for each $x \neq e$. On the other hand, $A(x)$ can be empty (if x is a leaf), nonempty and finite, or countably infinite. There is a unique path from e to x for each $x \in S$, and more generally, a unique path from x to y whenever $x \preceq y$. We let $d(x)$ denote the distance from the root e to x, and more generally, we let $d(x, y)$ denote the distance from x to y when $x \preceq y$. Using results from Section 8, The cumulative function of order $n \in \mathbb{N}$ for a rooted tree is

$$\#_n(x) = \binom{n + d(x)}{d(x)} = \binom{n + d(x)}{n}, \quad x \in S$$

The corresponding generating function is

$$\Lambda(x, t) = \frac{1}{(1-t)^{d(x)+1}}, \quad x \in S,\ |t| < 1$$

Recall also some other standard notation from the Section 3.7 on uniform posets: For $x \in S$ and $n \in \mathbb{N}$,

$$A_n(x) = \{y \in S : x \preceq y,\ d(x, y) = n\}$$

Thus, $A_0(x) = \{x\}$, and $\{A_n(x) : n \in \mathbb{N}\}$ partitions $I[x] = \{y \in S : x \preceq y\}$. When $x = e$, we write A_n instead of $A_n(e)$.

12.1 Upper Probability Functions

Suppose again that $(S, \preceq)$ is a rooted tree with root e. Let X be a random variable with values in S having probability density function f and upper probability function F. Then of course

$$F(x) = \mathbb{P}(X \succeq x) = \mathbb{P}(X = x) + \mathbb{P}(X \succ x)$$

But $\{X \succ x\} = \bigcup_{y \in A(x)} \{X \succeq y\}$, and the events $\{X \succeq y\}$ are disjoint over $y \in A(x)$. Hence

$$F(x) = \mathbb{P}(X = x) + \sum_{y \in A(x)} \mathbb{P}(Y \succeq y) = f(x) + \sum_{y \in A(x)} F(y)$$

It follows that

$$f(x) = F(x) - \sum_{y \in A(x)} F(y), \quad x \in S$$

In particular, F uniquely determines f, and so the distributional dimension of S is 1. Moreover, we can charaterize upper probability functions mathematically.

Proposition 72. Suppose that $F : S \to [0, 1]$. Then F is the upper probability function of a probability distribution on S if and only if

1. $F(e) = 1$
2. $F(x) \geq \sum_{y \in A(x)} F(y)$ for every $x \in S$.
3. $\sum_{x \in A_n} F(x) \to 0$ as $n \to \infty$.

Proof. Suppose first that F is the upper probability function of a random variable X taking values in S. Then trivially $F(e) = 1$, and as above,

$$F(x) - \sum_{y \in A(x)} F(y) = \mathbb{P}(X = x) \geq 0$$

Next, $d(X) \geq n$ if and only if $X \succeq x$ for some $x \in A_n$. Moreover the events $\{X \succeq x\}$ are disjoint over $x \in A_n$. Thus

$$\mathbb{P}[d(X) \geq n] = \sum_{x \in A_n} F(x)$$

But by local finiteness, the random variable $d(X)$ (taking values in $\mathbb{N}$) has a proper (non-defective) distribution, so $\mathbb{P}[d(X) \geq n] \to 0$ as $n \to \infty$).

Conversely, suppose that $F : S \to [0, 1]$ satisfies conditions (1)–(3) above. Define f on S by

$$f(x) = F(x) - \sum_{y \in A(x)} F(y), \quad x \in S$$

Then $f(x) \geq 0$ for $x \in S$ by (2). Suppose $x \in S$ and let $m = d(x)$. Then

$$\begin{aligned}
\sum_{k=0}^{n-1} \sum_{y \in A_k(x)} f(y) &= \sum_{k=0}^{n-1} \sum_{y \in A_k(x)} \left[F(y) - \sum_{z \in A(y)} F(z) \right] \\
&= \sum_{k=0}^{n-1} \left[\sum_{y \in A_k(x)} F(y) - \sum_{y \in A_k(x)} \sum_{z \in A(y)} F(z) \right] \\
&= \sum_{k=0}^{n-1} \left[\sum_{y \in A_k(x)} F(y) - \sum_{y \in A_{k+1}(x)} F(y) \right] \\
&= F(x) - \sum_{y \in A_n(x)} F(y)
\end{aligned}$$

since $A_0(x) = \{x\}$ and since the sum collapses. But

$$0 \leq \sum_{y \in A_n(x)} F(y) \leq \sum_{y \in A_{m+n}} F(y) \to 0 \text{ as } n \to \infty$$

Thus letting $n \to \infty$ we have

$$\sum_{y \in I[x]} f(y) = F(x), \quad x \in S$$

Letting $x = e$ we see that $\sum_{y \in S} f(y) = 1$ so f is a probability density function on S. Then we also see that F is the upper probability function of f. $\square$

Note 49. To characterize upper probability functions F with support S, we require that $F : S \to (0, 1]$ and we replace the weak inequality in (2) with strong inequality.

Note 50. Recall that $(S, \preceq)$ is a lower semi-lattice. Hence if X and Y are independent random variables with values in S, with upper probability functions F and G, repsectivley, then $X \wedge Y$ has upper probability function FG.

Lemma 5. Suppose that $F : S \to [0, 1]$ satisfies the decreasing property (2) in Proposition 72. Then for $x \in S$ and $k \in \mathbb{N}$,

$$\sum_{y \in A_k(x)} F(y) \leq F(x) \tag{34}$$

In particular, $\sum_{x \in A_n} F(x) \leq F(e)$ for $n \in \mathbb{N}$.

Proof. The proof is by induction on k. Trivially, (34) holds for $k = 0$, and by assumption for $k = 1$. Assume that (34) holds for a given $k \in \mathbb{N}$. Then

$$\sum_{y \in A_{k+1}(x)} F(y) = \sum_{t \in A_k(x)} \sum_{y \in A(t)} F(y) \leq \sum_{t \in A_k(x)} F(t) \leq F(x)$$

where the last step holds by the induction hypothesis. $\square$

Corollary 23. Suppose that $F : S \to [0,1]$ satisfies conditions (1) and (2) of Proposition 72 and let $p \in (0,1)$. Then $G : S \to [0,1]$ defined by

$$G(x) = p^{d(x)} F(x), \quad x \in S$$

is an upper probability function on S. In particular, if F is an upper probability function on S with probablitiy density function f, then G is an upper probability function with probability density function g given by

$$g(x) = p^{d(x)} f(x) + p^{d(x)}(1-p) \sum_{y \in A(x)} F(y), \quad x \in S$$

Proof. First, $G(e) = p^0 F(e) = 1$. Next, for $x \in S$,

$$\begin{aligned}\sum_{y \in A(x)} G(y) &= \sum_{y \in A(x)} p^{d(y)} F(y) = p^{d(x)+1} \sum_{y \in A(x)} F(x) \\ &\le p^{d(x)+1} F(x) \le p^{d(x)} F(x) = G(x)\end{aligned}$$

Finally,

$$\sum_{x \in A_n} G(x) = \sum_{x \in A_n} p^{d(x)} F(x) = p^n \sum_{x \in A_n} F(x) \le p^n F(e) \to 0 \text{ as } n \to \infty$$

so it follows from Proposition 72 that G is an upper probability function.

Suppose now that F is an upper probability function with probability density function f, and let g denote the probability density function corresponding to G. Then

$$\begin{aligned}g(x) &= G(x) - \sum_{y \in A(x)} G(y) = p^{d(x)} F(x) - \sum_{y \in A(x)} p^{d(y)} F(y) \\ &= p^{d(x)} F(x) - p^{d(x)+1} \sum_{y \in A(x)} F(y) \\ &= p^{d(x)} F(x) - p^{d(x)} \sum_{y \in A(x)} F(y) + p^{d(x)} \sum_{y \in A(x)} F(y) - p^{d(x)+1} \sum_{y \in A(x)} F(y) \\ &= p^{d(x)} f(x) + p^{d(x)}(1-p) \sum_{y \in A(x)} F(y)\end{aligned}$$

∎

Note that $x \mapsto p^{d(x)}$ is not itself an upper probability function, unless the tree is a path, since properties 2 and 3 will fail in general. Thus, we cannot view G simply as the product of two UPFs in general. However, We can give a probabilisitc interpretation of the construction in Corrolary 23. Suppose that X is a random variable taking values in S with UPF F and PDF f. Moreover, suppose that each edge in the tree $(S, \preceq)$, independently of the other edges, is either *working* with probability p or *failed* with probability $1-p$. Define U by

$$U = \max\{u \preceq X : \text{ the path from } e \text{ to } u \text{ is working}\}$$

Corollary 24. Random variable U has UPF G and PDF g given in Corollary 23.

Proof. First, $\mathbb{P}(U = u|X = u) = p^{d(u)}$, since the path from e to u must be working. For $u \prec x$, $\mathbb{P}(U = u|X = x) = p^{d(u)}(1-p)$ since the path from e to u must be working, but the first edge on the path from u to x must have failed. Let g denoting the PDF of U, and f and F denote the PDF and UPF of X, Conditioning on X we get

$$\begin{aligned} g(u) &= p^{d(u} f(u) + \sum_{x \succ u} f(x) p^{d(u)}(1-p) \\ &= p^{d(u)} f(u) + p^{d(u)}(1-p) \sum_{y \in A(u)} F(y), \quad u \in S \end{aligned}$$

∎

12.2 Rate Functions

Next we are interested in characterizing rate functions of distributions that have support S. If r is such a function, then as noted earlier, $0 < r(x) \le 1$ and $r(x) = 1$ if and only if x is a leaf. Moreover, if F is the upper probability function, then

1. $F(e) = 1$
2. $\sum_{y \in A(x)} F(y) = [1 - r(x)] F(x)$

Conversely, these conditions give a prescription for constructing an upper probability function corresponding to a given rate function. Specifically, suppose that $r : S \to (0, 1]$ and that $r(x) = 1$ for every leaf $x \in S$. First, we define $F(e) = 1$. Then if $F(x)$ has been defined for some $x \in S$ and x is not a leaf, then we define $F(y)$ for $y \in A(x)$ arbitrarily, subject only to the requirement that $F(y) > 0$ and that condition (2) holds. Note that F satisfies the first two conditions in Proposition 72. Hence if F satisfies the third condition, then F is the upper probability function of a distribution with support S and with the given rate function r. It seems complicated, and probably not worthwhile, to completely characterize functions r for which the third condition in Proposition 72 is satisfied. However, the following proposition gives a simple sufficient condition.

Proposition 73. Suppose that $r : S \to [0, 1)$ and that $r(x) = 1$ for each leaf $x \in S$. If there exists $\alpha > 0$ such that $r(x) \ge \alpha$ for all $x \in S$, then r is the rate function of a distribution with support S.

Proof. Let $F : S \to (0, 1]$ be any function constructed according to the prescription above. Then as noted above, F satisfies the first two conditions in Proposition 72, so we just need to verify the third condition. We show by induction that

$$\sum_{x \in A_n} F(x) \le (1 - \alpha)^n, \quad n \in \mathbb{N} \tag{35}$$

Inequality (35) holds trivially if $n = 0$. Suppose that (35) holds for a given $n \in \mathbb{N}$. Then

$$\begin{aligned}\sum_{x \in A_{n+1}} F(x) &= \sum_{t \in A_n} \sum_{x \in A(t)} F(x) = \sum_{t \in A_n} [1 - r(t)] F(t) \\ &\le \sum_{t \in A_n} (1 - \alpha) F(t) = (1 - \alpha)^{n+1}\end{aligned}$$

□

Note 51. Condition (35) means that the distribution of $d(X)$ is stochastically smaller than the geometric distribution with rate constant α.

Note 52. If $(S, \preceq)$ is not a path, then the rate function does not uniquely determine the distribution. Indeed, if x has two or more children, then there are infinitely many ways to perform step 2 in the prescription above.

Example 22. Suppose that F is an upper probability function on S, with corresponding rate function r. Let G denote the upper probability function constructed in Corollary 23, and let ρ denote the corresponding rate function. Then

$$\rho(x) = r(x) + \frac{1 - p}{F(x)} \sum_{y \in A(x)} F(y), \quad x \in S$$

12.3 Constant rate distributions

If $(S, \preceq)$ has leaves, then the only constant rate distribution has rate constant 1, and then the distribution must be point mass at the root e.

Corollary 25. Suppose that $(S, \preceq)$ has no leaves. Then $F : S \to (0, 1]$ is the upper probability function of a distribution with constant rate α if and only if $F(e) = 1$ and

$$\sum_{y \in A(x)} F(y) = (1 - \alpha) F(x), \quad x \in S \tag{36}$$

Proof. This follows immediately from Proposition 73 □

Corollary 26. Suppose that X has constant rate α on $(S, \preceq)$. Then $d(X)$ has the geometric distribution on $\mathbb{N}$ with rate α.

Proof. For $n \in \mathbb{N}$,

$$\mathbb{P}[d(X) \ge n] = \sum_{x \in A_n} \mathbb{P}(X \succeq x) = \sum_{x \in A_n} F(x) = (1 - \alpha)^n$$

So $\mathbb{P}[d(X) = n] = \alpha(1 - \alpha)^n$. □

As a special case of our previous comment, we can construct the upper probability functions of constant rate distributions on $(S, \preceq)$ recursively.

1. Start with $F(e) = 1$.
2. If $F(x)$ is defined for a given $x \in S$, then define $F(y)$ for $y \in A(x)$ arbitrarily, subject only to the conditions $F(y) > 0$ and that (36) holds.

Equivalently, we can realize the distributions with constant rate as the invariant distributions of certain upward run chains on $(S, \preceq)$. Thus, let $\boldsymbol{X} = (X_0, X_1, \ldots)$ be an upward run chain on $(S, \preceq)$ with transition probability function P. Suppose that $\boldsymbol{X}$ has the property that $P(x, e) = \alpha \in (0, 1)$ for each $x \in S$. Let $F(x) = \mathbb{P}_e(T_x \le T_e)$, where as usual, T_z is the first return time to z for $z \in S$. Note that $F(x) = P^n(e, x)$ if $x \in A_n$. Thus, for $x \in A_n$,

$$\begin{aligned}\sum_{y \in A(x)} F(y) &= \sum_{y \in A(x)} P^{n+1}(e, y) = \sum_{y \in A(x)} P^n(e, x) P(x, y) \\ &= P^n(e, x) \sum_{y \in A(x)} P(x, y) = (1 - \alpha) P^n(e, x) = (1 - \alpha) F(x)\end{aligned}$$

Hence (36) holds, so F is the upper probability function of a distribution with constant rate α. Moreover, $f = \alpha F$ is the invariant distribution of the chain $\boldsymbol{X}$.

12.4 Gamma distributions and the point process

Again, let $(S, \preceq)$ be a rooted tree with root e. Let F be the upper probability function of a distribution with constant rate α, so that F satisfies the conditions in Corollary 26. The gamma distribution of order $n \in \mathbb{N}_+$ has probability density function

$$f_n(x) = \alpha^n \binom{n + d(x) - 1}{d(x)} F(x), \quad x \in S$$

Note that if $x \in S$ and $y \in A(x)$ then $d(y) = d(x) + 1$. It then follows from (36) that

$$\begin{aligned}\sum_{y \in A(x)} f_n(y) &= \sum_{y \in A(x)} \alpha^n \binom{n + d(y) - 1}{d(y)} F(y) \\ &= \alpha^n \binom{n + d(x)}{d(x) + 1} \sum_{y \in A(x)} F(y) = \alpha^n (1 - \alpha) \binom{d(x) + n}{d(x) + 1} F(x)\end{aligned}$$

Consider now the thinned point process associated with $(Y_1, Y_2, \ldots)$, where a point is accepted with probability r and rejected with probability $1 - r$, independently from point to point. The probability density function of the first accepted point is

$$\begin{aligned}g(x) &= r\alpha \Lambda(x, (1 - r)\alpha) F(x) = r\alpha \frac{1}{[1 - (1 - r)\alpha]^{d(x)+1}} F(x) \\ &= \frac{r\alpha}{1 - \alpha + r\alpha} \frac{F(x)}{(1 - \alpha + r\alpha)^{d(x)}}, \quad x \in S\end{aligned}$$

Consider the function $G : S \to (0, 1]$ given by

$$G(x) = \frac{F(x)}{(1 - \alpha + r\alpha)^{d(x)}}, \quad x \in S$$

Note that $G(e) = 1$ and for $x \in S$

$$\begin{aligned}
\sum_{y \in A(x)} G(y) &= \sum_{y \in A(x)} \frac{F(y)}{(1 - \alpha + r\alpha)^{d(y)}} \\
&= \frac{1}{(1 - \alpha + r\alpha)^{d(x)+1}} \sum_{y \in A(x)} F(y) \\
&= \frac{1 - \alpha}{(1 - \alpha + r\alpha)^{d(x)+1}} F(x) \\
&= \frac{1 - \alpha}{1 - \alpha + r\alpha} G(x)
\end{aligned}$$

Hence, the distribution of the first accepted point has constant rate

$$\frac{r\alpha}{1 - \alpha + r\alpha}$$

Note 53. The UPF F is related to the UPF G by the construction in Corollay 23 (so the notation is reversed). That is, suppose Y denotes the first accepted point in the thinned process. Then the basic random variable X that we started with can be constructed as

$$X = \max\{x \preceq Y : \text{ there is a working path from } e \text{ to } x\}$$

where each edge is working, independently, with probability $1 - \alpha + r\alpha$.

12.5 General trees

Corollary 27. Every ordered tree without maximal elements supports a constant rate distribution.

Proof. Suppose that $(S, \preceq)$ is a standard discrete poset without maximal elements, whose covering graph is a tree. Then $(S, \preceq)$ can be constructed from a rooted tree by successively adding points, as in Theorem 14 and joining other trees, as in Section 5.5. Thus, the result follows from Theorems 25, 14, and 13. ☐

13 The Free Semigroup

13.1 Definitions

Let I be a countable alphabet of *letters*, and let S denote the set of all finite length *words* using letters from I. In particular, e denotes the empty word of length 0. Let $\cdot$ denote the usual concatenation operation on elements of S: if $x = x_1 \cdots x_m \in S$ and $y = y_1 \cdots y_n \in S$, then

$$xy = x_1 \cdots x_m y_1 \cdots y_n$$

Clearly $(S, \cdot)$ is a discrete, positive semigroup, and is known as the *free semigroup* generated by I. For the associated partial order, $x \preceq y$ if and only if x is a prefix of y, and then $x^{-1}y$ is the corresponding suffix. The digraph of the partially ordered set is a regular tree rooted at e in which each element x has distinct successors xi, $i \in I$. In particular, if $\#(I) = k$, then the digraph is a regular k-tree. Specializing further, when $k = 2$ and $I = \{0, 1\}$, the elements of S are bit strings and the corresponding digraph is the binary tree. The elements in I are the irreducible elements of S.

For $x \in S$ and $i \in I$, let $N_i(x)$ denote the number of times that letter i occurs in x, and let $N(x) = \sum_{i \in I} N_i(x)$ denote the length of x. Note that $N_i(x) = 0$ for all but finitely many $i \in I$. Note also that $N(x)$ is the distance from e to x, denoted $d(x)$ earlier for uniform posets. Let

$$M = \left\{ (n_i \colon i \in I) \colon n_i \in \mathbb{N} \text{ for } i \in I \text{ and } \sum_{i \in I} n_i < \infty \right\}$$
$$M_n = \left\{ (n_i \colon i \in I) \colon n_i \in \mathbb{N} \text{ for } i \in I \text{ and } \sum_{i \in I} n_i = n \right\}, \quad n \in \mathbb{N}$$

Thus, $\{M_n : n \in \mathbb{N}\}$ partitions M. We define the multinomial coefficients by

$$\begin{aligned} C(n_i \colon i \in I) &= \# \{x \in S \colon N_i(x) = n_i \text{ for } i \in I\} \\ &= \frac{(\sum_{i \in I} n_i)!}{\prod_{i \in I} n_i!}, \quad (n_i \colon i \in I) \in M \end{aligned}$$

The free semigroup has the property that the elements of $[e, x]$ are totally ordered for each x. Moreover, it is the only discrete positive semigroup with this property.

Theorem 41. Suppose that $(S, \cdot)$ is a discrete positive semigroup with the property that $[e, x]$ is totally ordered for each $x \in S$. Then $(S, \cdot)$ is isomorphic to a free semigroup on an alphabet.

Proof. Let I denote the set of irreducible elements of $(S, \cdot)$. If $x \in S - \{e\}$ then $i_1 \preceq x$ for some $i_1 \in I$ and hence $x = i_1 y$ for some $y \in S$. If $y \succ e$ then we can repeat the argument to write $y = i_2 z$ for some $i_2 \in I$ and $z \in S$. Note that $x = i_1 i_2 z$ and hence $i_1 i_2 \preceq x$. Moreover, i_1 and $i_1 i_2$ are distinct. Since the semigroup is locally finite, $[e, x]$ is finite and hence the process must terminate. Thus, we can write x in the form $x = i_1 i_2 \cdots i_n$ for some $n \in \mathbb{N}_+$ and some $i_1, i_2, \ldots, i_n \in I$. Finally, we show that the factorization is unique. Suppose that $x = i_1 i_2 \cdots i_n = j_1 j_2 \cdots j_m$ where $i_1, \ldots, i_n \in I$ and $j_1, \ldots, j_m \in I$. Then $i_1 \preceq x$ and $j_1 \preceq x$. Since the elements of $[e, x]$ are totally ordered, we must have $i_1 = j_1$. Using left cancellation we have $i_2 \cdots i_n = j_2 \cdots j_m$. Continuing in this fashion we see that $m = n$ and $i_1 = j_2$, $i_2 = j_2, \ldots i_n = j_n$ ∎

Proposition 74. Suppose that $(S, \cdot)$ is a discrete positive semigroup with the property that every $x \in S$ has a unique finite factoring over the set of irreducible elements I. Then $(S, \cdot)$ is isomporphic to the free semigroup on I.

Proof. Form Proposition 6, every $x \in S$ has a finite factoring over I, so that $x = i_1 i_2 \cdots i_n$. If this factoring is unique, then clearly $(i_1, i_2, \ldots i_n) \mapsto i_1 i_2 \cdots i_n$ is an isomorphism from the free semigroup $(I^*, \cdot)$ onto $(S, \cdot)$. □

From Theorem 41 and our discussion in Section 8, it follows that the cumulative function of order n associated with counting measure $\#$ is given by

$$\#_n(x) = \binom{N(x)+n}{n}, \quad x \in S$$

If X is a random variable with values in S then from Theorem 5,

$$\sum_{x \in S} \binom{N(x)+n}{n} \mathbb{P}(X \succeq x) = \mathbb{E}\left[\binom{N(X)+n+1}{n+1}\right]$$

Since the digraph of the free semigroup is a tree, there is a unique distribution with a given upper probability function The probability density function f of a distribution is related to its upper probability function F as follows:

$$f(x) = F(x) - \sum_{j \in I} F(xj) \tag{37}$$

Thus, the distributional dimension of the free semigroup is 1. On the other hand, the semigroup dimension is the number of letters.

Theorem 42. $\dim(S, \cdot) = \#(I)$.

Proof. Note first that a homomorphism ϕ from $(S, \cdot)$ into $(\mathbb{R}, +)$ can uniquely be specified by defining $\phi(i)$ for all $i \in I$ and then defining

$$\phi(i_1 i_2 \cdots i_n) = \phi(i_1) + \phi(i_2) + \cdots \phi(i_n)$$

Thus, if ϕ is a such a homomorphism and $\phi(i) = 0$ for all $i \in I$, then $\phi(x) = 0$ for all $x \in S$. Now suppose that $B \subseteq S$ and $\#(B) < \#(I)$. We will show that there exists a nonzero homomorphism from $(S, \cdot)$ into $(\mathbb{R}, +)$ with $\phi(x) = 0$ for all $x \in B$. Let I_B denote the set of letters contained in the words in B. Suppose first that I_B if a proper subset of I (and note that this must be the case if I is infinte). Define a homomorphism ϕ by $\phi(i) = 0$ for $i \in I_B$ and $\phi(i) = 1$ for $i \in I - I_B$. Then $\phi(x) = 0$ for $x \in B$, but ϕ is not the zero homomorphism. Suppose next that $I_B = I$. Thus I is finite, so let $k = \#(B)$ and $n = \#(I)$, with $k < n$. Denote the words in B by

$$i_{j1} i_{j2} \cdots i_{jm_j}, \quad j = 1, 2, \ldots, k$$

The set of linear, homogeneous equations

$$\phi(i_{j1}) + \phi(i_{j2}) + \cdots \phi(i_{jm_j}) = 0, \quad j = 1, 2, \ldots, k$$

has n unkowns, namely $\phi(i)$ for $i \in I$, but only k equations. Hence there exists a non-trivial solution. The homomorphism so constructed satisfies $\phi(x) = 0$ for $x \in B$. □

13.2 Exponential distributions

Theorem 43. An exponential distribution on $(S, \cdot)$ has upper probability function F of the form

$$F(x) = (1-\alpha)^{N(x)} \prod_{i \in I} p_i^{N_i(x)}, \quad x \in S$$

where $\alpha \in (0,1)$ is the rate constant, and where $p_i \in (0,1)$ for each $i \in I$ with $\sum_{i \in I} p_i = 1$.

Proof. We apply Theorem 26. Let $F(i) = \beta_i$ for $i \in I$. Then $\beta_i > 0$ for each $i \in I$ and the memoryless condition requires that

$$F(x) = \prod_{i \in I} \beta_i^{N_i(x)}, \quad x \in S$$

The constant rate property requires that $\sum_{x \in S} F(x) < \infty$ (and the reciprocal of this sum is the rate parameter). But from the multinomial theorem

$$\sum_{x \in S} F(x) = \sum_{n=0}^{\infty} \sum \left\{ C(n_i \colon i \in I) \prod_{i \in I} \beta_i^{n_i} \colon (n_i \colon i \in I) \in M_n \right\} = \sum_{n=0}^{\infty} \beta^n$$

where $\beta = \sum_{i \in I} \beta_i$. Hence we must have $\beta \in (0,1)$ and the rate constant is $1 - \beta$. Finally, we re-define the parameters: let $\alpha = 1 - \beta$ and let $p_i = \beta_i / \beta$ for $i \in I$. $\square$

The probability density function f of the exponential distribution in Theorem 43 is

$$f(x) = \alpha(1-\alpha)^{N(x)} \prod_{i \in I} p_i^{N_i(x)}, \quad x \in S \tag{38}$$

We could also obtain (38) directly from (37). In particular, in the free semigroup, every memoryless distribution has constant rate and hence is exponential. There is a simple interpretation of the exponential distribution in terms of multinomial trials.

Corollary 28. Consider a sequence of IID variables $(X_1, X_2, \ldots)$ taking values in I with $\mathbb{P}(X_n = i) = p_i$ for $n \in \mathbb{N}_+$ and $i \in I$. where $0 < p_i < 1$ for each i and $\sum_{i \in I} p_i = 1$. Let L be independent of $(X_1, X_2, \ldots)$ and have a geometric distribution with rate parameter $\alpha \in (0,1)$. Let X be the random variable in S defined by $X = X_1 \cdots X_L$. Then X has the exponential distribution with parameters α and $(p_i \colon i \in I)$.

Proof. Let $x \in S$. Then

$$\mathbb{P}(X = x) = \sum_{n=0}^{\infty} \mathbb{P}(X = x \mid L = n) \mathbb{P}(L = n)$$

The conditional probability is 0 unless $n = N(x)$ so we have

$$\begin{aligned}
\mathbb{P}(X = x) &= \mathbb{P}[X_1 \cdots X_{N(x)} = x \mid L = N(x)]\alpha(1-\alpha)^{N(x)} \\
&= \mathbb{P}(X_1 \cdots X_{N(x)} = x)\alpha(1-\alpha)^{N(x)} \\
&= \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_{N(x)} = x_{N(x)})\alpha(1-\alpha)^{N(x)} \\
&= \alpha(1-\alpha)^{N(x)} \prod_{i \in I} p_i^{N_i(x)}
\end{aligned}$$

$\square$

In the following corollaries, suppose that X has the exponential distribution on $(S, \cdot)$. with parameters α and $(p_i : i \in I)$.

Corollary 29. The joint distribution of $(N_i(X) : i \in I)$ given $N(X) = n \in \mathbb{N}$ is multinomial with parameters n and $(p_i : i \in I)$:

Corollary 30. The distribution of $N_i(X)$ is geometric with rate parameter $\alpha/[\alpha + (1-\alpha)p_i]$ for each $i \in I$.

Corollary 31. Random variable X maximizes entropy over all random variables $Y \in S$ with

$$\mathbb{E}[N_i(Y)] = \mathbb{E}[N_i(X)] = \frac{1-\alpha}{\alpha} p_i, \quad i \in I$$

13.3 Gamma distribution

From Corollary 28, it's easy to construct the gamma variables corresponding to the exponential distribution in Theorem 43. Let $(X_1, X_2, \ldots)$ be IID variables in I with $\mathbb{P}(X_j = i) = p_i$ for $i \in I$. Let $(J_1, J_2, \ldots)$ be IID variables each with the geometric distribution on $\mathbb{N}$ with rate α, and with $(J_1, J_2, \ldots)$ independent of $(Y_1, Y_2, \ldots)$. Thus $K_n = \sum_{i=1}^n J_i$ has the negative binomial distribution with parameters α and n. Then $Y_n = X_1 \cdots X_{K_n}$ has the gamma distribution of order n. The probability density function is

$$\mathbb{P}(Y_n = y) = \binom{N(y) + n - 1}{n - 1} \alpha^n (1-\alpha)^{N(y)} \prod_{i \in I} p_i^{N_i(y)}, \quad y \in S$$

Of course, this also follows from Proposition 51.

Corollary 32. The conditional distribution of $[N_i(Y_n) : i \in I]$ given $N(Y_n) = m \in \mathbb{N}$ is multinomial with parameters m and $(p_i : i \in I)$:

Corollary 33. The distribution of $N_i(Y_n)$ is negative binomial with parameters $\alpha/[\alpha + (1-\alpha)p_i]$ and n for each $i \in I$.

13.4 Distributions with constant rate

Corollary 34. Suppose that $F\colon S \to (0, 1]$ and $0 < \alpha < 1$. The F is the upper probability function of a distribution which has constant rate α if and only if $F(e) = 1$ and

$$(1-\alpha)F(x) = \sum_{i \in I} F(xi), \quad x \in S \tag{39}$$

Proof. This follows directly from Theorem 26 ☐

Using the characterization in Corollary 34, $F(x)$ can be defined recursively for $x \in S$ to produce many constant failure rate distributions other than the exponential ones.

Example 23. Suppose that $I = \{0, 1\}$, so that S consists of bit strings. Let $a > 0$, $b > 0$, $a \neq b$, and $a + b < 1$. Define $F\colon S \to (0, 1]$ as follows:

$$F(e) = 1,\ F(0x) = a^{N_1(x)+1} b^{N_0(x)},\ F(1x) = a^{N_1(x)} b^{N_0(x)+1}, \quad x \in S$$

Then F is a upper probability function for a distribution with constant rate $1-a-b$ but this distribution is not exponential or even a mixture of exponential distributions.

In summary, every distribution on the free semigroup which is memoryless has constant rate (and hence is exponential), but there are distributions with constant rate that are not memoryless (and hence not exponential).

13.5 Compound Poisson distributions

Proposition 75. Suppose that X has the exponential distribution on $(S, \cdot)$ with parameters α and $(p_i\colon i \in I)$ as above. Then X has a compound Poisson distribution.

Proof. Recall from Corollary 28 that X can be decomposed as

$$X = X_1 X_2 \cdots X_L$$

where $(X_1, X_2, \ldots)$ are independent and identically distributed on the alphabet I, with probability density function $(p_i\colon i \in I)$, and where L is independent of $(X_1, X_2, \ldots)$ and has the geometric distribution on $\mathbb{N}$ with rate parameter α. But from Section 8, L has a compound Poisson distribution and can be written in the form

$$L = M_1 + M_2 + \cdots + M_K$$

where $(M_1, M_2, \ldots)$ are independent and identically distributed on $\mathbb{N}_+$ with the logarithmic distribution

$$\mathbb{P}(M = n) = -\frac{(1-\alpha)^n}{n \ln(\alpha)}, \quad n \in \mathbb{N}_+$$

and where K is independent of $(M_1, M_2, \ldots)$ and has the Poisson distribution with parameter $-\ln(\alpha)$. It follows that we can take $(X_1, X_2, \ldots)$, $(M_1, M_2, \ldots)$, and K mutually independent, and thus

$$X = U_1 U_2 \cdots U_K$$

where $U_i = X_{M_{i-1}+1} \cdots X_{M_i}$. Note that $(U_1, U_2, \ldots)$ are independent and identically distributed on S and thus X has a compound Poisson distribution. $\square$

14 Positive Semigroups with Two Generators

In this chapter we consider discrete, positive semigroups with two irreducible elements a and b. In a sense, we start with the free semigroup on the alphabet $\{a, b\}$, and impose "equations" that must hold.

Example 24. If we impose no conditions, we are left with the free semigroup itself.

Example 25. If we impose the condition $ab = ba$ (the commutative law), then the resulting semigroup $(S, \cdot)$ is isomporhic to the standard semigroup $(\mathbb{N}^2, +)$. To see this, note that by the commutative and associative laws, every element in S can be written uniquely in the form $a^m b^n$ for some m, $n \in \mathbb{N}$. With this canonical representation, the semigroup operation is

$$(a^i b^j) \cdot (a^m b^n) = a^{i+m} b^{j+n}$$

Thus, the mapping $(m, n) \mapsto a^m b^n$ is an isomorphism from $(\mathbb{N}^2, +)$ to $(S, \cdot)$.

Example 26. Suppose we impose the equations $ab = b^2$, $ba = a^2$. Then the resulting positive semigroup $(S, \cdot)$ is isomporphic to the lexicographic semigroup in Section 11.2, with $k = 2$. To see this, note that that every element in S can be written as x^n where $x \in \{a, b\}$ and $n \in \mathbb{N}$. With this canonical representation, the semigroup operation (on non-identity elements) is

$$a^m \cdot b^n = b^{m+n}, \; b^n \cdot a^m = a^{m+n}$$

Thus, the mapping that takes $(0, e)$ to e and takes (n, x) to x^n (where $x \in \{a, b\}$ and $n \in \mathbb{N}_+$) is an isomorphism from the lexicographic semigroup of Section 11.2 to $(S, \cdot)$.

Example 27. Suppose that we impose the condition $ba = a^2$ (but not the complmentary condition $ab = b^2$). It's easy to see that every element in S can be written uniquely in the form $a^m b^n$ for some m, $n \in \mathbb{N}$. With this canonical representation, the semigroup operation is

$$(a^i b^j) \cdot (a^m b^n) = a^{i+j+m} b^n \text{ if } m \in \mathbb{N}_+$$
$$(a^i b^j) \cdot (b^n) = a^i b^{j+n}$$

It's straightforward to verify directly that $(S, \cdot)$ is a positive semigroup. Moreover, $(S, \cdot)$ is clearly isomorphic to the semigroup $(\mathbb{N}^2, \cdot)$ where

$$\begin{aligned}(i,j)\cdot(m,n) &= (i+j+m,n) \text{ if } m \in \mathbb{N}_+\\ (i,j)\cdot(0,n) &= (i,j+n)\end{aligned}$$

The corresponding partial order is given by

$$a^i b^j \prec a^m b^n \text{ if and only if } i+j < m \text{ or } (i = m \text{ and } j < n)$$

Note that the elements that cover $a^m b^n$ are a^{m+n+1} and $a^m b^{n+1}$.

Suppose now that F is the upper probability function of an exponential distribution on $(S, \cdot)$. Then from the defining condition of this semigroup we have

$$F(b)F(a) = F(ba) = F(a^2) = [F(a)]^2$$

and therefore $F(a) = F(b)$. Let $p \in (0,1)$ denote the common value. Then $F(a^m b^n) = p^{m+n}$ and

$$\sum_{x \in S} F(x) = \sum_{(m,n)\in\mathbb{N}^2} p^{m+n} = \frac{1}{(1-p)^2}$$

Hence the corresponding probability density function f is given by

$$f(a^m b^n) = (1-p)^2 p^{m+n}, \quad (m,n) \in \mathbb{N}^2$$

thus, a random variable $U = a^X b^Y$ has an exponential distribution on $(S, \cdot)$ with rate $(1-p)^2$ if and only if X and Y are IID geometric (i.e. exponential) variables on $(\mathbb{N}, +)$ with rate $1-p$.

Example 28. If we impose the condiiton $ba^2 = a^2$, then $b^m a^n = a^n$ for every $m \in \mathbb{N}$ and $n \in \{2, 3, \ldots\}$. In particular, $b^n a \preceq a^2$ for every $n \in \mathbb{N}$, so $(S, \cdot)$ is not locally finite.

15 Finite Subsets of $\mathbb{N}_+$

Let S denote the set of all finite subsets of $\mathbb{N}_+$. Clearly, $(S, \subseteq)$ is a standard discrete poset. Interestingly, this poset is associated with a standard discrete semigroup as well.

15.1 Preliminaries

We identify a nonempty subset x of $\mathbb{N}_+$ with the function given by

$$x(i) = i\text{th smallest element of } x$$

with domain $\{1, 2, \ldots, \#(x)\}$ if x is finite and $\mathbb{N}_+$ if x is infinite. We will sometimes refer to $x(i)$ as the element of *rank* i in x. If x is nonempty and finite, max(x) denotes the maximum value of x; by convention we take $\max(\emptyset) = 0$ and $\max(x) = \infty$ if x is infinite. Note that $\#(x) \leq \max(x)$ for every x. If x and y are nonempty subsets of $\mathbb{N}_+$ with $\max(y) \leq \#(x)$, we let $x \circ y$ denote the subset whose function is the composition of x and y:

$$(x \circ y)(i) = x(y(i))$$

We also define $x \circ \emptyset = \emptyset$ for any $x \subseteq \mathbb{N}_+$. Note that $x \circ y$ is always defined when x is infinite. The results in the following two propositions are simple:

Proposition 76. Suppose that x and y are subsets of $\mathbb{N}_+$ with $\max(y) \leq \#(x)$. Then

1. $x \circ y \subseteq x$
2. $\#(x \circ y) = \#(y)$
3. if y is nonempty and finite then $\max(x \circ y) = x(\max(y))$
4. if x is infinite then $(x \circ y)^c = x^c \cup (x \circ y^c)$

Proposition 77. Suppose that x, y, and z are subsets of $\mathbb{N}_+$. Assuming that the operations are defined,

1. $x \circ (y \circ z) = (x \circ y) \circ z$.
2. $x \circ (y \cup z) = (x \circ y) \cup (x \circ z)$.
3. $x \circ (y \cap z) = (x \circ y) \cap (x \circ z)$.
4. If $x \circ y = x \circ z$ then $y = z$.

Note that the right distributive laws cannot possibly hold; $(x \cup y) \circ z$ and $(x \circ z) \cup (y \circ z)$ do not even have the same cardinality in general, and neither do $(x \cap y) \circ z$ and $(x \circ z) \cap (y \circ z)$. Similarly, the right cancellation law does not hold: if $x \circ z = y \circ z$, we cannot even conclude that $\#(x) = \#(y)$, let alone that $x = y$. Note that $\mathbb{N}_+$ is a left-identity: $\mathbb{N}_+ \circ x = x$ for any $x \subseteq \mathbb{N}_+$.

15.2 The positive semigroup

Recall that S denotes the collection of all finite subsets of $\mathbb{N}_+$ (represented as functions as in Section 15.1). We define the binary operation $\cdot$ on S by

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

Note that the operation is well-defined since x^c is infinite. Essentially, xy is constructed by adding to x those elements of x^c that are indexed by y (in a sense those elements form a copy of y that is disjoint from x).

Theorem 44. $(S, \cdot)$ is a positive semigroup with the subset partial order.

Proof. The associative rule holds, and in fact

$$x(yz) = (xy)z = x \cup (x^c \circ y) \cup (x^c \circ y^c \circ z)$$

The empty set is the identity:

$$\begin{aligned} x\emptyset &= x \cup (x^c \circ \emptyset) = x \cup \emptyset = x \\ \emptyset x &= \emptyset \cup (\mathbb{N}_+ \circ x) = \emptyset \cup x = x \end{aligned}$$

The left cancellation law holds: Suppose that $xy = xz$. Then $x \cup (x^c \circ y) = x \cup (x^c \circ z)$ by definition and hence $x^c \circ y = x^c \circ z$ since the pairs of sets in each union are disjoint. But then $y = z$. There are no non-trivial inverses: if $xy = \emptyset$ then $x \cup (x^c \circ y) = \emptyset$. Hence we must have $x = \emptyset$ and therefore also $x^c \circ y = \mathbb{N}_+ \circ y = y = \emptyset$.

Finally, the associated partial order is the subset order. Suppose first that $xu = y$. Then $x \cup (x^c \circ u) = y$ so $x \subseteq y$. Conversely, suppose that $x \subseteq y$. Let $u = \{i \in \mathbb{N}_+ : x^c(i) \in y\}$. Then $x \cup (x^c \circ u) = y$ so $xu = y$. □

Note 54. The irreducible elements of $(S, \cdot)$ are the singletons $\{i\}$ where $i \in \mathbb{N}_+$. Note also that

$$\{i\}\{i\} = \{i, i+1\} \tag{40}$$

$$\{i+1\}\{i\} = \{i, i+1\} \tag{41}$$

$$\{i\}\{i+1\} = \{i, i+2\} \tag{42}$$

Comparing (41) and (42) we see that the semigroup is not commutative, and comparing (40) and (41) we wee that the right cancellation law does not held. Thus, $(S, \cdot)$ just satisfies the minimal algebraic assumptions of a positive semigroup; in particular, S cannot be embedded in a group. Finally, if $i_1 < i_2 < \cdots < i_n$ then

$$\{i_n\}\{i_{n-1}\} \cdots \{i_1\} = \{i_1, i_2, \cdots, i_n\}$$

Proposition 78. $\dim(S, \cdot) = 1$

Proof. suppose that φ is a homomorphism from $(S, \cdot)$ into $(\mathbb{R}, +)$ with $\varphi(\{1\}) = 0$. Then using (40) and 41) it follows that $\varphi(\{i\}) = 0$ for each $i \in \mathbb{N}_+$, and then $\varphi(x) = 0$ for every $x \in S$. On the other hand, there do exist non-trivial homomorphism—for example, the cardinality function $\#$, as shown in the following proposition. □

Proposition 79. For $x, y \in S$,

$$\#(xy) = \#(x) + \#(y) \tag{43}$$

$$\max(xy) = \begin{cases} \max(x), & \text{if } \max(y) \le \max(x) - \#(x) \\ \max(y) + \#(x), & \text{if } \max(y) > \max(x) - \#(x) \end{cases} \tag{44}$$

Proof. $\#(xy) = \#[x \cup (x^c \circ y)] = \#(x) + \#(x^c \circ y)$ since x and $x^c \circ y$ are disjoint. But from Proposition 76, $\#(x^c \circ y) = \#(y)$.

Equation 44 is trivial if x or y is the identity ($\emptyset$), so we will assume that x and y are nonempty. Note that, by definition,

$$\max(xy) = \max[x \cup (x^c \circ y)] = \max\{\max(x), \max(x^c \circ y)\}$$

Let $i = \#(x)$ and $n = \max(x)$. Then $n \in x$ and the remaining $i - 1$ elements of x are in $\{1, 2, \ldots, n-1\}$. Hence, x^c contains $n - i$ elements of $\{1, 2, \ldots, n-1\}$, together with all of the elements of $\{n+1, n+2, \ldots\}$. If $\max(y) \le n - i$, then $\max(x^c \circ y) = x^c(\max(y)) \le n-1$ so $\max(xy) = n = \max(x)$. If $\max(y) > n-i$, then $\max(xy) = \max(x^c \circ y) = x^c(\max(y))$—the element of rank $\max(y)$ in x^c. Given the structure of x^c noted above, this element is $n + (\max(y) - (n-i)) = \max(y) + i$. ◻

Note 55. Since the cardinality function is a homomorphism, the poset $(S, \subseteq)$ is uniform. That is, if $x \in S$ can be factored into singletons

$$x = u_1 u_2 \cdots u_n$$

where $\#(u_i) = 1$ for each i, then $n = \#(x)$.

The following proposition explores the relationship between a probability density function and the corresponding upper probability function. In particular, a upper probability function completely determines a distribution on S.

Proposition 80. Suppose that f is a probability density function on S, and let F denote the corresponding upper probability function:

$$F(x) = \sum_{y \in xS} f(y), \quad x \in S$$

For $x \in S$ and $n \in \mathbb{N}$, let $A_n(x) = \{y \in S : y \supseteq x, \#(y) = \#(x) + n\}$. Then

$$f(x) = \sum_{n=0}^{\infty} (-1)^n \sum_{y \in A_n(x)} F(y), \quad x \in S$$

Proof. First note that

$$\sum_{n=0}^{\infty} (-1)^n \sum_{y \in A_n(x)} F(y) = \sum_{n=0}^{\infty} (-1)^n \sum_{y \in A_n(x)} \sum_{z \in yS} f(z)$$

Now consider $z \in S$ with $z \supset x$ and $\#(z) = \#(x) + k$, where $k > 0$. For $i \in \{0, 1, \ldots, k\}$, there are $\binom{k}{i}$ subsets y with $x \subseteq y \subseteq z$ and $\#(y) = \#(x) + i$. Hence the total contribution to the sum for the subset z is

$$(-1)^{\#(x)} f(z) \sum_{i=0}^{k} \binom{k}{i} (-1)^i = 0$$

Hence the only surviving term in the sum (corresponding to $n = 0$ and $y = z = x$) is $f(x)$. ◻

Proposition 81. The cumulative function of order $n \in \mathbb{N}$ for counting measure $\#$ on $(S, \cdot)$ is given by

$$\#_n(x) = (n+1)^{\#(x)}, \quad x \in S$$

Proof. The result is trivially true when $n = 0$, since $\#_0(x) = 1$ for $x \in S$. Thus, assume the result holds for a given n. Using the binomial theorem,

$$\begin{aligned}
\#_{n+1}(x) &= \sum_{t \subseteq x} \#_n(t) = \sum_{k=0}^{\#(x)} \sum_{t \subseteq x,\, \#(t)=k} (n+1)^k \\
&= \sum_{k=0}^{\#(x)} \binom{\#(x)}{k} (n+1)^k = (n+2)^{\#(x)}, \quad x \in S
\end{aligned}$$

□

When $n = 1$, we get the usual formula for the number of subsets of x: $\#_1(x) = 2^{\#(x)}$.

Proposition 82. the Möbius function is

$$m(x, y) = (-1)^{\#(y) - \#(x)}, \quad x,\, y \in S,\, x \subseteq y \tag{45}$$

Proof. The proof is by induction on $\#(y) - \#(x)$. First, $m(x, x) = 1$ by definition. Suppose that (45) holds for $x \subseteq y$ when $\#(y) = \#(x) + n$. Suppose that $x \subseteq y$ with $\#(y) \le \#(x) + n + 1$. Then

$$\begin{aligned}
m(x, y) = &= -\sum_{t \in [x, y)} m(x, t) = -\sum_{k=0}^{n} \sum \{m(x, t) : t \in [x, y), \#(t) = \#(x) + k\} \\
&= -\sum_{k=0}^{n} \sum \{(-1)^k : t \in [x, y), \#(t) = \#(x) + k\} \\
&= -\sum_{k=0}^{n} \binom{n+1}{k} (-1)^k \\
&= -\sum_{k=0}^{n+1} \binom{n+1}{k} (-1)^k + (-1)^{n+1} = 0 + (-1)^{n+1}
\end{aligned}$$

We can now verify the Möbius inversion formula in this special case, with

$\#_n(x) = (n+1)^{\#(x)}$: For $x \in S$,

$$\begin{aligned}
\sum_{t \in D[x]} \#_{n+1}(t) m(t, x) &= \sum_{t \in D[x]} (n+2)^{\#}(t)(-1)^{\#(x)-\#(t)} \\
&= \sum_{k=0}^{\#(x)} \sum \{(n+2)^{\#(t)}(-1)^{\#(x)-\#(t)} : t \subseteq x, \#(t) = k\} \\
&= \sum_{k=0}^{\#(x)} \binom{\#(x)}{k} (n+2)^k (-1)^{\#(x)-k} \\
&= (n+1)^{\#(x)} = \#_n(x)
\end{aligned}$$

□

15.3 The sub-semigroups

The semigroup $(S, \cdot)$ has an interesting structure. In particular, it can be partitioned into sub-semigroups where the difference between the maximum element and the cardinality is constant.

Definition 42. For $k \in \mathbb{N}$, let

$$\begin{aligned}
S_k &= \{x \in S : \max(x) - \#(x) = k\} \\
T_k &= \{\emptyset\} \cup S_k
\end{aligned}$$

For $(n, k) \in \{(0, 0)\} \cup (\mathbb{N}_+ \times \mathbb{N})$, let

$$\begin{aligned}
S_{n,k} &= \{x \in S : \#(x) = n, \max(x) = n + k\} = \{x \in S_k : \#(x) = n\} \\
T_{n,k} &= \{\emptyset\} \cup S_{n,k}
\end{aligned}$$

Note 56. Of course, $S_{0,0} = \{\emptyset\}$. If $n \in \mathbb{N}_+$ and $k \in \mathbb{N}$, then

$$\#(S_{n,k}) = \binom{n+k-1}{n-1} \tag{46}$$

since $x \in S_{n,k}$ must contain the element $n + k$ and $n - 1$ elements from $\{1, 2, \ldots, n+k-1\}$. If we interpret the binomial coefficient $\binom{-1}{-1}$ as 1, then (46) is valid for $n = k = 0$ also. Suppose that $x \in S_{n,k}$ and $y \in S_{m,j}$. If $m + j \le k$ (so that $j \le k - m$, then from our previous work, $\max(xy) = \max(x) = n + k$ and $\#(xy) = n + m$. Therefore

$$\max(xy) - \#(xy) = (n + k) - (n + m) = k - m \tag{47}$$

so $xy \in S_{n+m,k-m}$. On the other hand, if $m + j > k$ (so that $j > k - m$ then $\max(xy) = \max(y) + \#(x) = m + j + n$ and as before, $\#(xy) = n + m$. Therefore

$$\max(xy) - \#(xy) = (m + j + n) - (n + m) = j \tag{48}$$

so $xy \in S_{n+m,j}$.

Proposition 83. T_k is a complete sub-semigroup of S for each $k \in \mathbb{N}$.

Proof. We first need to show that $xy \in T_k$ for x, $y \in T_k$. The result is trivial if $x = \emptyset$ or $y = \emptyset$, so we will assume that x and y are nonempty. Then $\max(y) > \max(x) - \#(x)$, since the left-hand side is $k + \#(y)$ and the right-hand side is k. By (44), $\max(xy) = \max(y) + \#(x)$. Hence

$$\max(xy) - \#(xy) = (\max(y) + \#(x)) - (\#(x) + \#(y)) = \max(y) - \#(y) = k$$

Therefore $xy \in T_k$.

To show completeness, Suppose that $x, y \in T_k$ and $x \subseteq y$ so that $xu = y$ for some $u \in S$. If $x = y$ then $u = \emptyset \in T_k$ and if $x = \emptyset$ then $u = y \in T_k$. Thus, suppose that x is a proper subset of y (so that $u \neq \emptyset$). If $\max(u) \leq k$ then from (44), $\max(y) = \max(x)$ and from (43), $\#(y) = \#(x) + \#(u)$, so $\max(y) - \#(y) = k - \#(u) < k$, a contradiction. Thus, $\max(u) > k$. From (44), $\max(u) = \max(y) - \#(x)$ and from (43), $\#(u) = \#(y) - \#(x)$, so $\max(u) - \#(u) = \max(y) - \#(y) = k$. Thus, $u \in T_k$. ☐

Note 57. Note that

$$S_0 = \{\{1, 2, \ldots, m\} : m \in \mathbb{N}\}$$

If $y \in S$ with $\#(y) = n$, then

$$\{1, 2, \ldots, m\}y = \{1, 2, \ldots, m\} \cup \{m + y(1), m + y(2), \ldots, m + y(n)\}$$

In particular,

$$\{1, 2, \ldots, m\}\{1, 2, \ldots, n\} = \{1, 2, \ldots, m + n\}$$

so $(S_0, \cdot)$ is isomorphic to $(\mathbb{N}, +)$ with isomorphism $x \mapsto \#(x)$. Finally, note that $\emptyset \in S_0$ so $T_0 = S_0$.

To characterize the exponential distributions on T_k, we must first characterize the minimal elements of S_k (which are the irreducible elements of T_k).

Proposition 84. The set of minimal element of S_k is

$$M_k = \{x \in S_k : x(i) \leq k \text{ for all } i < \#(x)\}$$

There are 2^k minimal elements.

Proof. First we show that if $x \in S_k$ is not a minimal element of S_k then $x \notin M_k$. Thus, suppose that $x = uv$ where u, $v \in S_k$ are nonempty. Then $\max(u) > k$ and $\max(u) \in u \subseteq uv = x$. Moreover, $\max(u) < \max(x)$, so the rank of $\max(u)$ in x is less than $\#(x) = \#(u) + \#(v)$. Therefore $x \notin M_k$.

Next we show that if $x \notin M_k$ then x is not a minimal element of S_k. Thus, suppose that $x \in S_k$ and $x(i) > k$ for some $i < \#(x)$. Construct $u \in S$ as follows: $x(i) \in u$ and u contains $x(i) - k - 1$ elements of x that are smaller than $x(i)$. This can be done since $x(i) - i \leq k$, and hence $x(i) - k - 1 \leq i - 1$, and by definition, x contains $i-1$ elements smaller than $x(i)$. Now note that $\max(u)-\#(u) = x(i)-(x(i)-k) = k$ so $u \in S_k$. By construction, $u \subseteq x$ so there exists $v \in S$ such that $uv = x$. Recall that v is the set of ranks of the elements of $x-u$ in u^c. But u^c contains k elements less than $x(i)$ together with the elements $x(i)+1, x(i)+2, \ldots$. The largest element of $x-u$ is $\max(x) = \#(x)+k$ which has rank greater than k in u^c. Therefore $\max(v) > k = \max(u) - \#(u)$ so by Proposition 79, $\max(x) = \max(v) + \#(u)$. Therefore

$$\max(v) - \#(v) = (\max(x) - \#(u)) - (\#(x) - \#(u)) = \max(x) - \#(x) = k$$

so $v \in S_k$. Therefore x is not a minimal element of S_k.

Next, note that if $x \in S_k$ and $\#(x) \geq k+2$, then $x \notin M_k$, since one of the $k+1$ elements of x of rank less than $\#(x)$ must be at least $k+1$. For $n \leq k+1$, the number of elements $x \in M_k$ with $\#(x) = n$ is $\binom{k}{n-1}$, since x must contain $n+k$ and $n-1$ elements in $\{1, 2, \ldots, k\}$. Hence

$$\#(M_k) = \sum_{n=1}^{k+1} \binom{k}{n-1} = 2^k$$

$\square$

Example 29. The minimal elements of S_1 are $\{2\}$ and $\{1,3\}$. The minimal elements of S_2 are $\{3\}$, $\{1,4\}$, $\{2,4\}$, and $\{1,2,5\}$.

Example 30. The number of elements in a factoring of an element in T_k into irreducible elements is not necessarily unique. For example, in T_1 we have

$$\{1,3\}\{1,3\} = \{2\}\{2\}\{1,3\} = \{1,2,3,5\}$$

Thus, $(T_k, \subseteq)$ is not a uniform poset.

Since $(T_0, \cdot)$ is isomorphic to $(\mathbb{N}, +)$, the cumulative function of order $n \in \mathbb{N}$ on T_0 corresponding to counting measure $\#$ is

$$\#_{0,n}(x) = \binom{\#(x)+n}{n}, \quad x \in T_0$$

Problem 21. Find the cumulative functions for T_k, corresponding to counting measure, when $k \geq 1$.

15.4 Exponential distributions on T_k

Theorem 45. There are no memoryless distributions on S, and hence no exponential distributions.

Proof. Suppose that X is a random variable on S with a memoryless distribution. Thus, X is a random, finite subset of $\mathbb{N}_+$. By the memoryless property,

$$\mathbb{P}(\{i\}\{i\} \subseteq X) = \mathbb{P}(i \in X)\mathbb{P}(i \in X)$$

$$\mathbb{P}(\{i+1\}\{i\} \subseteq X) = \mathbb{P}(i+1 \in X)\mathbb{P}(i \in X)$$

But $\{i\}\{i\} = \{i+1\}\{i\}$ as noted above, so we must have

$$\mathbb{P}(i+1 \in X) = \mathbb{P}(i \in X)$$

for every $i \in \mathbb{N}_+$. Next, note that if $i_1 < i_2 < \cdots < i_n$ then by another application of the memoryless property,

$$\begin{aligned}
\mathbb{P}(i_1 \in X, i_2 \in X, \ldots, i_n \in X) &= \mathbb{P}(\{i_1, i_2, \ldots, i_n\} \subseteq X) \\
&= \mathbb{P}(\{i_n\}\{i_{n-1}\} \cdots \{i_1\} \subseteq X) \\
&= \mathbb{P}(i_1 \in X)\mathbb{P}(i_2 \in X) \cdots \mathbb{P}(i_n \in X)
\end{aligned}$$

It therefore follows that the events $\{\{i \in X\} : i \in \mathbb{N}_+\}$ are independent and identically distributed. By the Borel-Cantelli lemma, infinitely many of the events must occur with probability 1, so X is infinite—a contradiction. $\square$

Although there are no exponential distributions on S, each of the sub-semigroups T_k has a one-parameter family of exponential distributions.

Theorem 46. A random variable X taking values in T_k has an exponential distribution if and only if the upper probability function F and density function f have the following form, for some $\alpha \in (0,1)$:

$$F(x) = \alpha^{\#(x)}, \quad x \in T_k \tag{49}$$

$$f(x) = \frac{(1-\alpha)^{k+1}}{(1-\alpha)^{k+1} + \alpha} \alpha^{\#(x)}, \quad x \in T_k \tag{50}$$

Proof. The function $F(x) = \alpha^{\#(x)}$ takes values in $(0,1]$ and satisfies $F(xy) = F(x)F(y)$ for all $x\, y \in T_k$. Moreover

$$\begin{aligned}
\sum_{x \in T_k} F(x) &= \sum_{n=0}^{\infty} \sum_{x \in T_{k,n}} F(x) = \sum_{n=0}^{\infty} \sum_{x \in T_{n,k}} \alpha^n \\
&= 1 + \sum_{n=1}^{\infty} \binom{n+k-1}{n-1} \alpha^n = \frac{(1-\alpha)^{k+1} + \alpha}{(1-\alpha)^{k+1}}
\end{aligned}$$

It follows that F and f as given above are the upper probability function and density function, respectively, of an exponential distribution.

Conversely, suppose now that F is the upper probability function on T_k with the memoryless property. T_0, as noted earlier, is isomorphic to $(\mathbb{N}, +)$, with $\#$ an isomorphism. Thus, if $k = 0$, F must have the form $F(x) = \alpha^{\#(x)}$ where $\alpha = F(\{1\}) \in (0,1)$. For general k, we will show by induction on $\#(x)$ that $F(x) = \alpha^{\#(x)}$ where $\alpha = F(\{k+1\}) \in (0,1)$. The result is trivially true if $\#(x) = 0$, since $x = \emptyset$. The result is also trivially true if $\#(x) = 1$, since the only such $x \in T_k$ is $x = \{k+1\}$. Suppose now that $F(x) = \alpha^{\#(x)}$ for all $x \in T_k$ with $\#(x) \le n$. Let $x \in T_k$ with $\#(x) = n+1$. If x is not an irreducible element, then $x = uv$ where $u, v \in T_k$, $\#(u) \le n$, $\#(v) \le n$, and $\#(u) + \#(v) = \#(x)$. In this case,

$$F(x) = F(u)F(v) = \alpha^{\#(u)}\alpha^{\#(v)} = \alpha^{\#(x)}$$

On the other hand, if x is irreducible, let $j = \min\{i \in x : i + 1 \notin x\}$. Note that $j < \#(x)$ since $\max(x) = \#(x) + k$. Now let $y \in T_k$ be obtained by from x by replacing j with $j + 1$. Note that $\#(y) = \#(x)$ and moreover, y^c can be obtained from x^c by replacing $j + 1$ with j. We claim that $xx = yx$; that is, $x \cup x^c y = y \cup y^c x$. To see this, note first that if $i \ne j$ and $i \ne j+1$, then $i \in x$ if and only if $i \in y$, and $i \in x^c$ if and only if $i \in y^c$. On the other hand, $j \in x$ and $j \in y^c x$ since $j = y^c(x(1))$ (by definition, there are $x(1) - 1$ elements less than $x(1)$ in y^c; the next element in y^c is j). Similarly, $j + 1 \in y$ and $j + 1 \in x^c x$ since $j + 1 = x^c(x(1))$. Since $xx = yx$, it follows from the memoryless property that $F(x) = F(y)$. Continuing this process, we find that $F(x) = F(y)$ for some $y \in T_k$ that is not irreducible, but with $\#(y) = \#(x)$. It then follows that $F(x) = F(y) = \alpha^{\#(y)} = \alpha^{\#(x)}$ and the proof is complete. □

Example 31. To illustrate the last part of the proof, let $x = \{3, 4, 5, 8, 15\}$, so that $x \in T_{10}$. Then $j = 5$, $y = \{3, 4, 6, 8, 15\}$ and

$$xx = yx = \{3, 4, 5, 6, 7, 8, 9, 12, 15, 20\}$$

.

Note 58. Suppose that X has the exponential distribution on T_k given in Theorem 46. From the general theory of exponential distributions, the expected number of subsets of X in T_k is the reciprocal of the rate parameter in the density function. Thus,

$$\mathbb{E}(\#[\emptyset, X]) = 1 + \frac{\alpha}{(1-\alpha)^{k+1}}$$

If $k = 0$ (recall that $S_0 = T_0$) note that

$$\mathbb{P}(X = x) = (1 - \alpha)\alpha^{\#(x)}, \quad x \in S_0 \tag{51}$$

On the other hand, suppose that $k \in \mathbb{N}_+$. Then

$$\mathbb{P}(X \in S_k) = 1 - \mathbb{P}(X = \emptyset) = 1 - g(0) = \frac{\alpha}{(1-\alpha)^{k+1} + \alpha}$$

Thus, the conditional distribution of X given $X \in S_k$ has density function

$$\mathbb{P}(X = x | X \in S_k) = \frac{\mathbb{P}(X = x)}{\mathbb{P}(X \in S_k)} = (1 - \alpha)^{k+1}\alpha^{\#(x)-1}, \quad x \in S_k \tag{52}$$

The density function of X depends on $x \in T_k$ only through $\#(x)$, so it is natural to study the distribution of $\#(X)$. Of course, by definition $\max(X) = \#(X) + k$ on T_k, so the distribution of $\#(X)$ determines the distribution of $\max(X)$.

Corollary 35. Suppose that X has the exponential distribution on T_k given in Theorem 46, Then

$$\mathbb{P}[\#(X) = n] = \frac{(1-\alpha)^{k+1}}{(1-\alpha)^{k+1}+\alpha}\binom{n+k-1}{k}\alpha^n, \quad n \in \mathbb{N} \tag{53}$$

$$\mathbb{E}[\#(X)] = \frac{\alpha}{1-\alpha}\frac{\alpha(1+k\alpha)}{(1-\alpha)^{k+1}+\alpha} \tag{54}$$

When $k = 0$, (53) gives $\mathbb{P}[\#(X) = n] = (1-\alpha)\alpha^n$ for $n \in \mathbb{N}$, so $\#(X)$ has a geometric distribution on $\mathbb{N}$. In general, $\#(X)$ has a modified negative binomial distribution.

Corollary 36. Given $\#(X) = n$, X is uniformly distributed on $S_{n,k}$

Proof. Let $x \in S_{n,k}$. Using the probability density functions of X and $\#(X)$ we have

$$\mathbb{P}[X = x \mid \#(X) = n] = \frac{\mathbb{P}(X = x)}{\mathbb{P}[\#(X) = n]} = \frac{1}{\binom{n+k-1}{n-1}}$$

□

It is easy to see from (54) that for each $k \in \mathbb{N}$, $\mathbb{E}[\#(X)]$ is a strictly increasing function of α and maps $(0,1)$ onto $(0,\infty)$. Thus, the exponential distribution on T_k can be re-parameterized by expected cardinality. Moreover, the exponential distribution maximizes entropy with respect to this parameter:

Corollary 37. The exponential distribution in Theorem 46 maximizes entropy over all distributions on T_k with expected value given by 54.

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 T_k, then

$$-\sum_{x \in T_k} g(x)\ln[g(x)] \le -\sum_{x \in T_k} g(x)\ln[f(x)] \tag{55}$$

If X has the exponential distribution in Theorem 46, and $\mathbb{E}(\#(Y)) = \mathbb{E}(\#(X))$ then substituting into the right-hand side of equation 55 we see that the entropy of Y is bounded above by

$$-\ln(c_{k,\alpha}) - \mu_{k,\alpha}\ln(\alpha)$$

where $c_{k,\alpha}$ is the rate parameter of the exponential density in equation 50 and $\mu_{k,\alpha}$ is the mean cardinality in equation 54. Of course, the entropy of X achieves this upper bound. □

Theorem 47. Suppose that X has the exponential distribution on T_k with parameter α. If $x \in S_j$ $(x \neq \emptyset)$, then

$$\mathbb{P}(X \supseteq x) = \frac{\alpha^{\#(x)+1}}{(1-\alpha)^{k+1}+\alpha}, \quad j < k$$

$$\mathbb{P}(X \supseteq x) = \frac{(1-\alpha)^{k+1}}{(1-\alpha)^{k+1}+\alpha}\binom{j}{j-k}\alpha^{\#(x)+j-k}[1+\alpha H(j,k,\alpha)], \quad j > k$$

where $H(j,k,\alpha) = \text{hyper}([1, 1+j], [1+j-k], \alpha)$.

Proof. Let $x \in S$. Using the distribution of $\#(X)$ and the fact that X is uniformly distributed on $S_{n,k}$, given $\#(X) = n$, we have

$$\begin{aligned}
\mathbb{P}(X \supseteq x) &= \sum_{n=0}^{\infty} \mathbb{P}(\#(X) = n)\mathbb{P}(X \supseteq x) \mid \#(X) = n) \\
&= \frac{(1-\alpha)^{k+1}}{(1-\alpha)^{k+1}+\alpha} \sum_{n=0}^{\infty} \binom{n+k-1}{k} \alpha^n \frac{\#\{y \in S_{n,k} : y \supseteq x\}}{\binom{n+k-1}{k}} \\
&= \frac{(1-\alpha)^{k+1}}{(1-\alpha)^{k+1}+\alpha} \sum_{n=0}^{\infty} \alpha^n \#\{y \in S_{n,k} : y \supseteq x\}
\end{aligned}$$

Suppose first that $x \in S_{m,j}$ where $j < k$. Then $\{y \in S_{n,k} : y \supseteq x\} = \emptyset$ if $n \le m$. Suppose $n > m$ so that $m + j < n + k$ To construct $y \in S_{n,k}$ with $y \supseteq x$, we must add the element $n + k$ and then add $n - m - 1$ elements from $\{1, \dots, n+k-1\} - x$. The number of such subsets is $\binom{n+k-1-m}{n-m-1}$. Therefore

$$\begin{aligned}
\mathbb{P}(X \supseteq x) &= \frac{(1-\alpha)^{k+1}}{(1-\alpha)^{k+1}+\alpha} \sum_{n=m+1}^{\infty} \alpha^n \binom{n+k-m-1}{n-m-1} \\
&= \frac{(1-\alpha)^{k+1}}{(1-\alpha)^{k+1}+\alpha} \alpha^{m+1} \sum_{i=0}^{\infty} \binom{i+k}{i} \alpha^i \\
&= \frac{\alpha^{m+1}}{(1-\alpha)^{k+1}+\alpha}
\end{aligned}$$

Next, suppose that $x \in S_{m,j}$ with $j > k$. If $m + j > n + k$ then $\{y \in S_{n,k} : y \supseteq x\} = \emptyset$. Suppose $m + j = n + k$. To construct $y \in S_{n,k}$ with $y \supseteq x$, we must add $n - m$ elements from $\{1, \dots, n+k-1\} - x$. The number of such sets is $\binom{n-m+k}{n-m} = \binom{j}{j-k}$ Suppose $m + j < n - k$. To construct $y \in S_{n,k}$ with $y \supseteq x$, we must add $n + k$ and $n - m - 1$ elements from $\{1, \dots, n+k-1\} - x$. The number of such sets is $\binom{n-m+k-1}{n-m-1}$. Therefore

$$\begin{aligned}
\mathbb{P}(X \supseteq x) = &\frac{(1-\alpha)^{k+1}}{(1-\alpha)^{k+1}+\alpha} \times \\
&\left[\binom{j}{j-k} \alpha^{m+j-k} + \sum_{n=m+j-k+1}^{\infty} \binom{n-m-1+k}{n-m-1} \alpha^n \right]
\end{aligned}$$

Simplifying,

$$\mathbb{P}(X \supseteq x) = \frac{(1-\alpha)^{k+1}}{(1-\alpha)^{k+1}+\alpha} \binom{j}{j-k} \alpha^{m+j-k} [1 + \alpha H(j,k,\alpha)]$$

∎

Problem 22. Find a closed form expression for the gamma density function of order n on T_k. This is equivalent to the problem of finding the cumulative function of order $n-1$, corresponding to counting measure $\#$.

Problem 23. Determine whether the exponential distributions of T_k are compound Poisson (and therefore infinitely divisible relative to the semigroup operation).

Problem 24. If X has an exponential distribution on T_k, compute the hitting probability function on S. That is, compute $\mathbb{P}(X \cap x \neq \emptyset)$ for $x \in S$.

15.5 Almost exponential distributions on S

There are no exponential distributions on S. However, we can define distributions that are "close" to exponential by forming mixtures of the distributions in (51) and (52). Thus, suppose that X takes values in S with probability mass function

$$\mathbb{P}(X = x) = \begin{cases} \beta_0(1-\alpha_0)\alpha_0^{\#(x)}, & x \in S_0 \\ \beta_k(1-\alpha_k)^{k+1}\alpha_k^{\#(x)-1}, & x \in S_k,\ k \in \mathbb{N}_+ \end{cases} \tag{56}$$

where α_k, $\beta_k \in (0,\, 1)$ for each $k \in \mathbb{N}$ and $\sum_{k=0}^{\infty} \beta_k = 1$. Thus, the conditional distribution of X given $X \in S_k$ is the same as the corresponding conditional distribution of an exponential variable on T_k (with parameter α_k). Note that the conditional distribution of X on T_k itself is not exponential. Nor can we construct a distribution on S by requiring that the conditional distributions on T_k be exponential for each k, essentially because these semigroups share $\emptyset$ and thus are not disjoint. The distribution of X is as close to exponential as possible, in the sense that X is essentially exponential on each of the sub-semigroups S_k, and these semigroups partition S.

The probability mass function in (56) is equivalent to

$$\mathbb{P}(X = x) = c_k \alpha_k^{\#(x)}, \quad x \in S$$

where $\alpha_k \in (0,1)$ for each $k \in \mathbb{N}$ and where $c_k \in (0,1)$, $k \in \mathbb{N}$ are chosen so that $\sum_{x \in S} c_k \alpha_k^{\#(x)} = 1$. Indeed, the equivalence is

$$\begin{aligned} c_0 &= (1-\alpha_0)\beta_0 \\ c_k &= \frac{(1-\alpha_k)^{k+1}}{\alpha_k}, \quad k \in \mathbb{N}_+ \end{aligned}$$

There is not much that we can say about the general distribution in (56). In the remainder of this section we will study a special case with particularly nice properties. For our first construction, we define a random variable X on S by first selecting a geometrically distributed population size N, and then selecting a sample from $\{1, 2, \ldots, N\}$ in an IID fashion. Of course, geometric distributions are the exponential distributions for the positive semigroup $(\mathbb{N}, +)$.

More precisely, let N be a random variable taking values in $\mathbb{N}$, and having the geometric distribution with parameter $1-r \in (0,1)$:

$$\mathbb{P}(N=n) = (1-r)r^n, \quad n \in \mathbb{N}$$

Next, given $N = n$, X is distributed on the subsets of $\{1,2,\ldots,n\}$ so that $i \in X$, independently, with probability p for each $i \in \{1,2,\ldots,n\}$. Of course, if $N = 0$, then $X = \emptyset$.

Theorem 48. For $x \in S$,

$$\mathbb{P}(X=x) = \frac{1-r}{1-r+rp}(rp)^{\#(x)}[r(1-p)]^{\max(x)-\#(x)} \tag{57}$$

$$\mathbb{P}(X \supseteq x) = p^{\#(x)} r^{\max(x)} \tag{58}$$

Proof. For $x \in S$,

$$\mathbb{P}(X=x) = \sum_{n=0}^{\infty} \mathbb{P}(N=n)\mathbb{P}(X=x|N=n)$$

If $n < \max(x)$ then x is not a subset of $\{1,2,\ldots,n\}$, so $\mathbb{P}(X = x|N = n) = 0$. If $n \geq \max(x)$ then x is a subset of $\{1,2,\ldots,n\}$ and by assumption, $\mathbb{P}(X = x|N = n) = p^{\#(x)}(1-p)^{n-\#(x)}$. Substituting gives

$$\mathbb{P}(X=x) = \sum_{n=\max(x)}^{\infty} (1-r)r^n p^{\#(x)}(1-p)^{n-\#(x)}$$

which simplifies to (57). By a similar argument,

$$\mathbb{P}(X \supseteq x) = \sum_{n=\max(x)}^{\infty} (1-r)r^n p^{\#(x)}$$

which simplifies to (58). □

Not surprisingly, the distribution of X depends on $x \in S$ only through $\#(x)$ and $\max(x)-\#(x)$. As before, let $U = \#(X)$ and now let $V = \max(X)-\#(X)$. The following corollaries will explore the relationships between the distributions of U, V, and X, and provide another way of constructing the distribution of X.

Corollary 38. For $(n,k) \in \{(0,0)\} \cup (\mathbb{N}_+ \times \mathbb{N})$,

$$\begin{aligned}\mathbb{P}(U=n, V=k) &= \mathbb{P}(X \in S_{n,k}) \\ &= \frac{1-r}{1-r+rp}\binom{n+k-1}{n-1}(rp)^n[r(1-p)]^k\end{aligned}$$

Corollary 39. For $(n,k) \in \{(0,0)\} \cup (\mathbb{N}_+ \times \mathbb{N})$, the conditional distribution of X given $U = n$, $V = k$ is uniform on $S_{n,k}$.

Corollary 40. For $n \in \mathbb{N}$, the conditional distribution of V given $U = n$ is negative binomial with parameters n and $r(1-p)$:

$$\mathbb{P}(V = k | U = n) = \binom{n+k-1}{n-1} [r(1-p)]^k (1 - r + rp)^n, \quad k \in \mathbb{N} \tag{59}$$

Corollary 41. The distribution of U is geometric with parameter $(1-r)/(1-r+rp)$. Thus,

$$\mathbb{P}(U = j) = \frac{1-r}{1-r+rp} \left(\frac{rp}{1-r+rp} \right)^j, \quad j \in \mathbb{N}$$

Of course, Corollaries 39, 40, and 41 determine the distribution of X. In fact these results give an alternate way of constructing the distribution of X in the first place: We first give U a geometric distribution with a parameter $a \in (0,1)$; given $U = n$ we give V a negative binomial distribution with parameters n and $b \in (0,1)$; and finally, given $U = n, V = k$, we give X the uniform distribution on $S_{n,k}$. Our original construction, although simple, is perhaps unsatisfactory because the population variable N is hidden (not directly observable from X). The alternate construction has no hidden variables, and moreover, the geometric distribution of U and the conditional uniform distribution for X given $U = n, V = k$ are natural. On the other hand, the conditional negative binomial distribution of V given $U = n$ is somewhat obscure. The two constructions are equivalent, since there is a one-to-one correspondence between the pairs of parameters:

$$a = \frac{rp}{1-r+pr} \qquad b = r(1-p)$$

$$r = a(1-b) + b \qquad p = \frac{a(1-b)}{a(1-b)+b}$$

Our next goal is to study the distribution of the random subset X on the sub-semigroups S_k. First note that

$$\frac{\mathbb{P}(X = x)}{\mathbb{P}(X \supseteq x)} = \frac{1-r}{1-r+rp} (1-p)^{\max(x) - \#(x)}$$

Thus, for $k \in \mathbb{N}$, X has constant rate $\frac{1-r}{1-r+rp}(1-p)^k$ on the sub-semigroup S_k. In particular, for $x \in S_0$,

$$\mathbb{P}(X = x) = \frac{1-r}{1-r+rp} (rp)^{\#(x)}$$

$$\mathbb{P}(X \supseteq x) = (rp)^{\#(x)}$$

Hence, X has the memoryless property on S_0 (in addition to the constant rate property). That is, for $x, y \in S_0$,

$$\begin{aligned}\mathbb{P}(X \supseteq xy) &= (rp)^{\#(xy)} = (rp)^{\#(x)+\#(y)} \\ &= (rp)^{\#(x)} (rp)^{\#(y)} = \mathbb{P}(X \supseteq x) \mathbb{P}(X \supseteq y)\end{aligned}$$

To find the conditional distribution of X given $X \in S_k$, we first need $\mathbb{P}(X \in S_k)$, or equivalently, the probability density function of V.

Corollary 42. V has a modified geometric distribution:

$$\mathbb{P}(X \in S_0) = \mathbb{P}(V = 0) = \frac{1-r}{(1-r+rp)(1-rp)}$$

$$\mathbb{P}(X \in S_k) = \mathbb{P}(V = k) = \frac{(1-r)rp}{(1-r+rp)(1-rp)} \left(\frac{r(1-p)}{1-rp}\right)^k, \quad k \in \mathbb{N}_+$$

Corollary 43. The conditional distributions of X on S_k are as follows:

$$\mathbb{P}(X = x | X \in S_0) = (1-rp)(rp)^{\#(x)}, \quad x \in S_0 \tag{60}$$

$$\mathbb{P}(X = x | X \in S_k) = (1-rp)^{k+1}(rp)^{\#(x)-1}, \quad x \in S_k, k \in \mathbb{N}_+ \tag{61}$$

Thus, X has an almost exponential distribution in the sense of (56), with $\alpha_k = 1 - rp$ for each $k \in \mathbb{N}$, and with the mixing probabilities given in Corollary 42.

Recall that by Theorem 45, no exponential distribution on S exists because the events $\{\{i \in X\} : i \in \mathbb{N}_+\}$ would have to be independent with a common probability. The next corollary explores these events for the random variable in Theorem 48.

Corollary 44. Suppose that X has the distribution in Theorem 48.

1. $\mathbb{P}(i \in X) = pr^i$ for $i \in \mathbb{N}_+$.
2. If $i_1, i_2, \ldots, i_n \in \mathbb{N}_+$ with $i_1 < i_2 < \cdots < i_n$ then

$$\begin{aligned}\mathbb{P}(i_n \in X | i_1 \in X, \ldots, i_{n-1} \in X) &= \mathbb{P}(i_n \in X | i_{n-1} \in X) \\ &= \mathbb{P}(i_n - i_{n-1} \in X) = pr^{i_n - i_{n-1}}\end{aligned}$$

3. For $j \in \mathbb{N}_+$, the events $\{1 \in X\}, \{2 \in X\}, \ldots, \{j-1 \in X\}$ are conditionally independent given $\{j \in X\}$ with $\mathbb{P}(i \in X | j \in X) = p$ for $i < j$.

Property 3 in Corollary 44 is clearly a result of the original construction of X. Property 2 is reminiscent of a Markov property. This property implies that the events $\{\{i \in X\} : i \in \mathbb{N}_+\}$ are positively correlated, but asymptotically uncorrelated. In fact the correlation decays exponentially since

$$\mathbb{P}(i + j \in X | i \in X) = \mathbb{P}(j \in X) = pr^j \to 0 \text{ as } j \to \infty$$

Problem 25. Characterize all random subsets of $\mathbb{N}_+$ that satisfy the "partial Markov property" above.

From Corollaries 41 and 40, we can compute the expected value of $U = \#(X)$ and $W = \max(X) = U + V$:

$$\mathbb{E}(U) = \frac{rp}{1-r} \tag{62}$$

$$\mathbb{E}(W) = \frac{rp}{(1-r)(1-r+rp)} \tag{63}$$

It is easy to see from (62) and (63) that $(\mathbb{E}(U), \mathbb{E}(W))$, as a function of (r, p) maps $(0,1)^2$ one-to-one and onto $\{(c,d) : 0 < c < d < \infty\}$. Thus, the distribution of X can be re-parameterized by expected cardinality and expected maximum. Explicitly, if $0 < c < d$, then the values of r and p that yield $\mathbb{E}(U) = c$ and $\mathbb{E}(W) = d$ are

$$r = 1 - \frac{c}{d(1+c)}, \quad p = \frac{c^2}{d + cd - c}$$

Moreover, the distribution of X maximizes entropy with respect to these parameters. The proof of the following corollary is essentially the same as the proof of Corollary 37

Corollary 45. The distribution in Theorem 48 maximizes entropy among all distributions on S with expected cardinality given by (62) and expected maximum given by (63).

Proposition 85. Let $Z = \min(X)$ if $X \neq \emptyset$, and let $Z = 0$ if $X = \emptyset$. Then Z has a modified geometric distribution on $\mathbb{N}$:

$$\begin{aligned}
\mathbb{P}(Z = 0) &= \frac{1-r}{1-r+rp} \\
\mathbb{P}(Z = k) &= rp[r(1-p)]^{k-1}, \quad k \in \mathbb{N}_+
\end{aligned}$$

Of fundamental importance in the general theory of random sets (see Matheron [20]) is the hitting probability function G:

$$G(x) = \mathbb{P}(X \cap x \neq \emptyset), \quad x \subseteq \mathbb{N}_+$$

This function completely determines the distribution of a random set. Note that G is defined for all subsets of the positive integers, not just finite subsets.

Theorem 49. Suppose that X has the almost exponential distribution with parameters p and r. Then

$$G(x) = \sum_{i=1}^{\#(x)} p(1-p)^{i-1} r^{x(i)}, \quad x \subseteq \mathbb{N}_+$$

where as usual, $x(i)$ is the i'th smallest element of x.

Proof. Suppose first that x is finite (so that $x \in S$). From the standard inclusion-exclusion formula (or from [20]),

$$G(x) = \sum_{k=1}^{\#(x)} (-1)^{k-1} \sum_{y \subseteq x, \#(y)=k} F(y)$$

Hence, substituting the formula for $F(y)$ we have

$$\begin{aligned} G(x) &= \sum_{k=1}^{\#(x)} (-1)^{k-1} \sum_{y \subseteq x, \#(y)=k} p^{\#(y)} r^{\max(y)} \\ &= \sum_{k=1}^{\#(x)} (-1)^{k-1} p^k \sum_{i=k}^{\#(x)} \sum_{y \subseteq x, \#(y)=k, \max(y)=x(i)} r^{x(i)} \\ &= \sum_{k=1}^{\#(x)} (-1)^{k-1} p^k \sum_{i=k}^{\#(x)} \binom{i-1}{k-1} r^{x(i)} \\ &= \sum_{i=1}^{\#(x)} r^{x(i)} \sum_{k=1}^{i} \binom{i-1}{k-1} (-1)^{k-1} p^k \\ &= \sum_{i=1}^{\#(x)} p(1-p)^{i-1} r^{x(i)} \end{aligned}$$

For infinite x, the formula holds by the continuity theorem. □

Example 32. An easy computation from Theorem 49 gives

$$G(\{n, n+1, \ldots\}) = \frac{pr^n}{1 - r + rp}, n \in \mathbb{N}_+$$

This is also $\mathbb{P}(\max(X) \geq n)$, and agrees with our earlier results. In particular, letting $n = 1$, we get $G(\mathbb{N}_+) = \mathbb{P}(X \neq \emptyset)$:

$$G(\mathbb{N}_+) = \frac{rp}{1 - r + rp}$$

This agrees with $1 - \mathbb{P}(X = \emptyset)$ using the probability density function. The probability that X contains an even integer is

$$G(\{2, 4, \ldots\}) = \frac{pr^2}{1 - r^2 + pr^2}$$

The probability that X contains an odd integer is

$$G(\{1, 3, \ldots\}) = \frac{rp}{1 - r^2 + pr^2}$$

Example 33. The following list gives 10 simulations of X with $r = 0.95$ and $p = 0.65$, corresponding to $\mathbb{E}(U) = 12.35$ and $\mathbb{E}(W) = 18.5$.

1. $X = \{1, 2, 3\}$, $U = 3$, $W = 3$
2. $X = \{1, 2, 3, 4, 5, 6, 9, 10, 11, 12, 16, 17, 18, 19, 20, 21, 22, 25, 26, 27, 32\}$, $U = 21$, $W = 32$
3. $X = \{2, 3, 5\}$, $U = 3$, $W = 5$

4. $X = \{1, 4, 5, 9, 10, 11, 12\}$, $U = 7$, $W = 12$
5. $X = \{2, 5, 6, 7, 8, 12, 13, 15, 17, 18, 20, 23, 24, 25, 26, 27, 29, 30, 31, 32, 35, 37, 38, 40, 41\}$, $U = 25$, $W = 41$
6. $X = \{1, 2, 6, 8, 10, 12, 15, 16, 17, 18, 20, 23, 25, 28, 29, 30, 31, 32, 34\}$, $U = 19$, $W = 34$
7. $X = \{1, 2, 3\}$, $U = 3$, $W = 3$
8. $X = \{1, 3, 4, 7, 10, 11, 12, 13, 14, 21, 30, 32, 33, 34, 35, 37, 38, 40, 41, 42, 45, 46, 47, 49, 50, 51, 52, 53\}$, $U = 28$, $W = 53$
9. $X = \{1, 4, 5, 6, 9\}$, $U = 5$, $W = 9$
10. $X = \{5, 7\}$, $U = 2$, $W = 7$

Problem 26. Determine how the random sets in this section relate to the random closed sets studied by Matheron [20] and others.

Problem 27. Compute the convolution powers of the probability density function of the almost exponential distribution with parameters r and p. That is, if $X_1, X_2, \ldots$ are IID variables with this distribution, find the density function of $Y_n = X_1 X_2 \cdots X_n$.

15.6 Constant rate distributions on S

Recall that the standard poset $(S, \subseteq)$ is a uniform; in the notation of Section 3.7, if $x \subseteq y$ then $d(x, y) = \#(y) - \#(x)$. Let

$$\begin{aligned} A_n &= \{x \in S : \#(x) = n\} \\ A_n(x) &= \{y \in S : x \subseteq y,\ \#(y) = \#(x) + n\} \end{aligned}$$

So $\{A_n : n \in \mathbb{N}\}$ is a partition of S and $\{A_n(x) : n \in \mathbb{N}\}$ is a partition of $I[x]$.

Recall that $(S, \cdot)$ does not have any exponential distributions. But does $(S, \preceq)$ have constant rate distributions? We give some examples of distributions that are *not* constant rate. These models are not mutually exclusive.

Example 34 (Independent elements)**.** Suppose that X is a random variable with values in S and the property that $i \in X$ with probability p_i, independently over $i \in \mathbb{N}$. We must have $\prod_{i \in \mathbb{N}_+} p_i = 0$ so that X is finite with probability 1. For $x \in S$,

$$\mathbb{P}(X = x) = \prod_{i \in x} p_i \prod_{i \in x^c} (1 - p_i)$$

while

$$\mathbb{P}(X \supseteq x) = \prod_{i \in x} p_i$$

Hence X cannot have constant rate.

Example 35 (Sampling models)**.** Suppose that we pick a random population size $N \in \mathbb{N}$ with probability density function g and upper probability function G. Given $N = n$, we put $i \in X$ independently with probability p for $i \in 1, \dots, n$. Then for $x \in S$,

$$\begin{aligned}\mathbb{P}(X \supseteq x) &= \sum_{n=\max(x)}^{\infty} \mathbb{P}(N = n)\mathbb{P}(X \supseteq x|N = n) \\ &= \sum_{n=\max(x)}^{\infty} g(n)p^{\#(x)} = G(\max(x))p^{\#(x)}\end{aligned}$$

whereas

$$\begin{aligned}\mathbb{P}(X = x) &= \sum_{n=\max(x)}^{\infty} \mathbb{P}(N = n)\mathbb{P}(X = x|N = n) \\ &= \sum_{n=\max(x)}^{\infty} g(n)p^{\#(x)}(1-p)^{n-\#(x)} \\ &= p^{\#(x)} \sum_{n=\max(x)}^{\infty} g(n)(1-p)^{n-\#(x)}\end{aligned}$$

So X does not have constant rate. If we generalize this model so that, given $N = n$, $i \in X$ with probability p_i independently for $i \in \{1, \dots, n\}$ then

$$\mathbb{P}(X \supseteq x) = G(\max(x)) \prod_{i \in x} p_i$$

whereas

$$\mathbb{P}(X = x) = \prod_{i \in x} p_i \sum_{n=\max(x)}^{\infty} g(n) \prod_{i \in \{1, \dots n\} - x} (1 - p_i)$$

so again, it would seem impossible for X to have constant rate.

Example 36 (Conditionally independent elements)**.** Suppose that X is a random variable taking values in S, satisfying a generalization of the last condition in Corollary 44. Specifically, for $j \in \mathbb{N}_+$ the events $\{1 \in X\}, \{2 \in X\}, \dots, \{j-1 \in X\}$ are conditionally independent with

$$\mathbb{P}(i \in X|j \in X) = p_{ij}, \quad i \in \{1, \dots j-1\}$$

Let $g(i) = \mathbb{P}(i \in X)$ for $i \in \mathbb{N}_+$. Then for $x \in S$,

$$\mathbb{P}(X \supseteq x) = g(\max(x)) \prod_{i \in a(x)} p_{i,\max(x)}$$

$$\mathbb{P}(x = x) = g(\max(x)) \prod_{i \in a(x)} p_{i,\max(x)} \prod_{j \in b(x)} (1 - p_{i,\max(x)})$$

where $a(x) = x - \{\max(x)\}$ and $b(x) = \{1, \dots \max(x) - 1\} - a(x)$. Again, it's hard to see how X could have constant rate.

Suppose that X has probability density function f and upper probabiltiy function F. Recall that the upper probability function F determines the density function f:

$$f(x) = \sum_{n=0}^{\infty} (-1)^n \sum_{y \in A_n(x)} F(y), \quad x \in S \tag{64}$$

Thus, if Xhas constant rate α then

$$f(x) = \frac{1}{\alpha} \sum_{n=0}^{\infty} (-1)^n \sum_{y \in A_n(x)} f(y), \quad x \in S \tag{65}$$

15.7 The cardinality of a constant rate variable

Suppose that X takes values in S, and that X has constant rate α on the poset $(S, \subseteq)$ (although of course, we do not know that such variables exist). Our goal in this subsection is to study the distribution of $U = \#(X)$.

Lemma 6. The distribution of U satisfies

$$\alpha \mathbb{P}(U = k) = \mathbb{E}\left[(-1)^{U+k} \binom{U}{k}\right], \quad k \in \mathbb{N} \tag{66}$$

Proof. Let f denote the probability density function of X, as above. From (65)

$$\mathbb{P}(U = k) = \sum_{x \in A_k} f(x) = \frac{1}{\alpha} \sum_{n=0}^{\infty} (-1)^n \sum_{x \in A_k} \sum_{y \in A_n(x)} f(y)$$

The last two sums are over all $x, y \in S$ with $\#(y) = n + k$, and $x \subseteq y$. Interchanging the order of summation gives

$$\begin{aligned}
\mathbb{P}(U = k) &= \frac{1}{\alpha} \sum_{n=0}^{\infty} (-1)^n \sum_{y \in A_{n+k}} \sum \{f(y) : x \in A_k,\, x \subseteq y\} \\
&= \frac{1}{\alpha} \sum_{n=0}^{\infty} (-1)^n \sum_{y \in A_{n+k}} \binom{n+k}{k} f(y) \\
&= \frac{1}{\alpha} \sum_{n=0}^{\infty} (-1)^n \binom{n+k}{k} \mathbb{P}(U = n + k)
\end{aligned}$$

Equivalently (with the substitution $j = n + k$),

$$\alpha \mathbb{P}(U = k) = \sum_{j=k}^{\infty} (-1)^{j-k} \binom{j}{k} \mathbb{P}(U = j)$$

With the usual convention on binomial coefficients, that is $\binom{a}{b} = 0$ if $b < 0$ or $b > a$, we have

$$\alpha \mathbb{P}(U = k) = \sum_{j=0}^{\infty} (-1)^{j+k} \binom{j}{k} \mathbb{P}(U = j) = \mathbb{E}\left[(-1)^{k+U} \binom{U}{k}\right]$$

$\square$

In particular, when $k = 0$, (66) gives

$$\alpha \mathbb{P}(U = 0) = \mathbb{E}[(-1)^U] = \mathbb{P}(U \text{ is even}) - \mathbb{P}(U \text{ is odd}) = 2\mathbb{P}(U \text{ is even}) - 1$$

But also, $\mathbb{P}(U = 0) = f(\emptyset) = \alpha F(\emptyset) = \alpha$. Thus, letting h denote the probability density function of U, we have the curious property that

$$h^2(0) = 2\sum_{n=0}^{\infty} h(2n) - 1$$

Now let G denote the probability generating function of U.

Lemma 7. The distribution of U satisfies (66) if and only if

$$G(t - 1) = \alpha G(t) = G(0)G(t), \quad t \in \mathbb{R} \tag{67}$$

Proof. Assume that (66) holds. Then

$$\alpha h(k) t^k = \mathbb{E}\left[(-1)^{U+k}\binom{U}{k} t^k\right], \quad t \in \mathbb{R}$$

and hence

$$\alpha \sum_{k=0}^{\infty} h(k) t^k = \mathbb{E}\left[(-1)^U \sum_{k=0}^{\infty} (-1)^k \binom{U}{k} t^k\right] \tag{68}$$

The sum on the right converges for all $t \in \mathbb{R}$ and

$$\alpha \mathbb{E}(t^U) = \mathbb{E}\left[(-1)^U \sum_{k=0}^{U} (-t)^k \binom{U}{k}\right] \tag{69}$$

$$= \mathbb{E}[(-1)^U (1 - t)^U] = \mathbb{E}[(t - 1)^U] \tag{70}$$

But also, $\alpha = h(0) = G(0)$ so (67) holds. Conversely, suppose that G satisfies (67). Then (70), (69), and (68) hold with $\alpha = G(0)$. Equating coefficients give (66). $\square$

Corollary 46. The distribution of U satisfies (67) if and only if

$$G(t - n) = \alpha^n G(t) = G^n(0)G(t), \quad t \in \mathbb{R},\ n \in \mathbb{Z} \tag{71}$$

Proof. Suppose that G satisfies (67). Let $t \in \mathbb{R}$. By a simple induction, $G(t - n) = \alpha^n G(t)$ for $n \in \mathbb{N}$. But then for $n \in \mathbb{N}$,

$$G(t) = G(t + n - n) = \alpha^n G(t + n)$$

or equivalently, $G(t + n) = \alpha^{-n} G(t)$. Conversely, (71) clearly implies (67). $\square$

Lemma 8. The distribution of U satisfies (67) if and only if

$$\mathbb{P}(U = n) = \alpha \mathbb{E}\left[\binom{U}{n}\right], \quad n \in \mathbb{N} \tag{72}$$

Proof. Suppose that (67) holds. Then $\alpha G^{(n)}(t) = G^{(n)}(t-1)$ for $n \in \mathbb{N}$ and hence $\alpha G^{(n)}(1) = G^{(n)}(0)$ for $n \in \mathbb{N}$. But $G^{(n)}(1) = \mathbb{E}[(U)_n]$ and $G^{(n)}(0) = n!\mathbb{P}(U = n)$. Hence we have

$$\mathbb{P}(U = n) = \alpha \frac{\mathbb{E}[(U)_n]}{n!} = \alpha \mathbb{E}\left[\binom{U}{n}\right]$$

Conversely suppose that (72) holds and let G denote the probability generating function of U. Then we have $\alpha G^{(n)}(1) = G^{(n)}(0)$ for $n \in \mathbb{N}$. Let $H(t) = G(0)G(t+1) = \alpha G(t+1)$. Then $H^{(n)}(0) = G^{(n)}(0)$ for $n \in \mathbb{N}$, so $G = H$. Equivalently, $G(t-1) = \alpha G(t)$ for $t \in \mathbb{R}$. □

Combining Lemmas 6 and 8 we have the curious property that

$$(-1)^n \mathbb{E}\left[(-1)^U \binom{U}{n}\right] = \alpha^2 \mathbb{E}\left[\binom{U}{n}\right]$$

if U satisfies any of the equivalent conditions in Lemmas 6, 7, and 8. Let

$$\mu(n) = \mathbb{E}\left[\binom{U}{n}\right] = \mathbb{E}\left[\frac{(U)_n}{n!}\right]$$

This is the *binomial moment* of order n. Thus, the probability density function h of U and the binomial moment function μ are related by

$$h(n) = h(0)\mu(n), \quad n \in \mathbb{N}$$

Theorem 50. The Poisson distribution with parameter $\lambda > 0$ satisfies the equivalent conditions in Lemmas 6, 7 and 8 with $\alpha = e^{-\lambda}$. Conversely, the Poisson distribution is the only distribution that satisfies these conditions.

Proof. Suppose that U has the Poisson distribution with parameter $\lambda > 0$ and $\alpha = e^{-\lambda}$, recall that

$$G(t) = e^{\lambda(t-1)}, \quad t \in \mathbb{R}$$

so clearly

$$G(t-1) = e^{\lambda(t-2)} = e^{-\lambda}e^{\lambda(t-1)} = e^{-\lambda}G(t)$$

Conversely, it's easy to show that the equivalent conditions are in turn equivalent to the famous Rao-Rubin characterization of the Poisson distribution [26]. □

15.8 Constant rate distirbutions on S, continued

Now let g_k denote the conditional probability density function of X given that $U = k$, so that

$$g_k(x) = \mathbb{P}(X = x | U = k), \quad x \in A_k$$

and suppose that U has the Poisson distribution with parameter λ. Then condition (65) for X to have constant rate $e^{-\lambda}$ is

$$e^{-2\lambda} \frac{\lambda^k}{k!} g_k(x) = \sum_{n=0}^{\infty} (-1)^n \sum_{y \in A_n(x)} e^{-\lambda} \frac{\lambda^{n+k}}{(n+k)!} g_{n+k}(y), \quad x \in A_k$$

or equivalently,

$$e^{-\lambda} \frac{g_k(x)}{k!} = \sum_{n=0}^{\infty} \frac{(-\lambda)^n}{(n+k)!} \sum_{y \in A_n(x)} g_{n+k}(y), \quad x \in A_k \tag{73}$$

Lemma 9. Condition (73) holds if

$$\sum_{y \in A_n(x)} g_{n+k}(y) = \binom{n+k}{k} g_k(x), \quad x \in A_k,\ k, n \in \mathbb{N} \tag{74}$$

Proof. Suppose that (74) holds. Then for $x \in A_k$,

$$\begin{aligned} \sum_{n=0}^{\infty} \frac{(-\lambda)^n}{(n+k)!} \sum_{y \in A_n(x)} g_{n+k}(y) &= \sum_{n=0}^{\infty} \frac{(-\lambda)^n}{(n+k)!} \frac{(n+k)!}{n!k!} g_k(x) \\ &= \frac{g_k(x)}{k!} \sum_{n=0}^{\infty} \frac{(-\lambda)^n}{n!} = e^{-\lambda} \frac{g_k(x)}{k!} \end{aligned}$$

and thus (73) holds. ∎

Condition (74) is equivalent to

$$g_k(x) = \frac{1}{\binom{n}{k}} \sum_{y \in B_n(x)} g_n(y), \quad k \in \mathbb{N}, n \in \mathbb{N}, k \le n, x \in B_k \tag{75}$$

where $B_k = \{x \in S : \#(x) = k\}$ and $B_n(x) = \{y \in S : \#(y) = n, x \subseteq y\}$. Condition (75) has the following interpretation:

Proposition 86. Suppose that Y is a random variable taking values in B_n with probability density function g_n. Let X be a randomly chosen subset of Y of size $k \le n$. Then the density function g_k of X satisfies (75).

Proof. The precise meaning of the hypothesis is that given $Y = y$, $y \in B_n$, the conditional distribution of X is uniform on the collection of subsets of y of size k. Thus, for $x \in B_k$,

$$g_k(x) = \mathbb{P}(X = x) = \sum_{y \in B_n(x)} \mathbb{P}(y = y)\mathbb{P}(X = x | Y = y) = \frac{1}{\binom{n}{k}} \sum_{y \in B_n(x)} g_n(y)$$

∎

Note 59. In particular, (75) is consistent. If g_n is a probability density function on B_n, for a given $n \in \mathbb{N}$, then we could *define* g_k by (75) for $k = 0, \ldots, n-1$; these would also be probability density functions.

Proposition 87. No sequence of probability density functions $(g_n : n \in \mathbb{N})$ satisfying (75) exists.

Proof. Suppose that such a sequence does exist. Fix $k \in \mathbb{N}$ and $x \in B_k$ and then let $n \to \infty$ in (75). Since

$$\sum_{y \in B_n(x)} g_n(y) \leq 1$$

we must have $g_k(x) = 0$. $\square$

16 A Positive Sub-semigroup of $(\mathbb{C}, \cdot)$

16.1 Definitions

Let $\mathbb{C}$ denote the set of complex numbers $z = x + iy$ with the topology of $\mathbb{R}^2$, and let $\cdot$ denote ordinary complex multiplication. Let $\mathbb{C}_0 = \mathbb{C} - \{0\}$ and let

$$S = \{z \in \mathbb{C}: \mid z \mid > 1\} \cup \{1\}$$

both with the relative topology.

Proposition 88. $(S, \cdot)$ is a positive sub-semigroup of the abelian group $(\mathbb{C}_0, \cdot)$, The induced partial order is given by

$$z \prec w \text{ if and only if } |z| < |w|$$

Proof. $z \prec w$ if and only if there exists $u \in S$, $u \neq 1$, such that $zu = w$. But this occurs if and only if $|u| = |w/z| = |w|/|z| > 1$. $\square$

Proposition 89. The measure λ defined by

$$d\lambda(z) = \frac{1}{|z|^2} dxdy = \frac{1}{x^2 + y^2} dxdy$$

is left-invariant on the group $(\mathbb{C} - \{0\}, \cdot)$. Hence λ restricted to S is left-invariant on S.

Proof. Let $f: S \to \mathbb{R}$ be measurable. It suffices to show that for $w \in \mathbb{C}_0$,

$$\int_{\mathbb{C}_0} f(wz) d\lambda(z) = \int_{\mathbb{C}_0} f(z) d\lambda(z)$$

Let $z = re^{i\theta}$ where (r, θ) denote ordinary polar coordinates in $\mathbb{C}$, $0 \leq \theta < 2\pi$. Then

$$d\lambda(z) = \frac{1}{r^2} dxdy = \frac{1}{r^2} r dr d\theta = \frac{1}{r} dr d\theta$$

Let $w = \rho e^{i\phi}$. Then

$$\int_{\mathbb{C}_0} f(wz) d\lambda(z) = \int_0^{2\pi} \int_0^\infty f(r\rho e^{i(\theta+\phi)}) \frac{1}{r} dr d\theta$$

Now let $\hat{r} = r\rho$, $\hat{\theta} = \theta + \phi$. Then $dr = (1/\rho) d\hat{r}$ and $d\theta = d\hat{\theta}$ so

$$\int_{\mathbb{C}_0} f(wz) d\lambda(z) = \int_0^{2\pi} \int_0^\infty f\left(\hat{r} e^{i\hat{\theta}}\right) \frac{1}{\hat{r}} d\hat{r} d\theta = \int_{\mathbb{C}_0} f(z) d\lambda(z)$$

□

16.2 Exponential distributions

Theorem 51. $F : S \to (0, 1]$ is the upper probability function of an exponential distribution on $(S, \cdot)$ if and only if $F(z) = |z|^{-\beta}$ for some $\beta > 0$. The corresponding density function (with respect to μ) is

$$f(z) = \frac{\beta}{2\pi} |z|^{-\beta}, \quad z \in S$$

Proof. Suppose that $F : S \to (0, 1]$ satisfies $F(zw) = F(z)F(w)$ for z, $w \in S$. If x, $y \in [1, \infty)$ then $F(xy) = F(x)F(y)$; that is, F is a homomorphism restricted to the multiplicative semigroup $[1, \infty)$. Hence there exists $\beta > 0$ such that $F(x) = x^{-\beta}$ for $x \in [1, \infty)$. If $x > 1$ then $[F(-x)]^2 = F(x^2) = x^{-2\beta}$ so $F(-x) = x^{-\beta}$ and therefore $F(x) = |x|^{-\beta}$ for $x \in (-\infty, 1) \cup [1, \infty)$. Next, if $r > 1$, $m \in \mathbb{N}$ and $n \in \mathbb{N}_+$ then

$$[F(re^{i\pi m/n})]^n = F[(re^{i\pi m/n})^n] = F(r^n e^{im\pi}) = r^{-n\beta}$$

Hence $F(re^{im\pi/n}) = r^{-\beta}$. By continuity, $F(re^{i\theta}) = r^{-\beta}$ for $r > 1$ and any θ. Next note that

$$\begin{aligned} \int_S F(z)\, d\lambda(z) &= \int_S |z|^{-\beta} \frac{1}{|z|^2}\, dx\, dy \\ &= \int_0^{2\pi} \int_1^\infty r^{-(\beta+2)} r\, dr\, d\theta = \int_0^{2\pi} \int_1^\infty r^{-(\beta+1)}\, dr\, d\theta = \frac{2\pi}{\beta} \end{aligned}$$

From the basic existence theorem, F is the upper probability function of an exponential distribution on $(S, \cdot)$ and that f given above is the corresponding density function (with respect to the left-invariant measure λ). □

Note 60. Suppose that Z has the exponential distribution on $(S, \cdot)$ with parameter $\beta > 0$, as specified in the previous theorem. Then the density function with respect to Lebesgue measure is

$$z \mapsto \frac{\beta}{2\pi} |z|^{-(\beta+2)}$$

In terms of polar coordinates, the density is

$$(r, \theta) \mapsto \frac{\beta}{2\pi} r^{-(\beta+1)}$$

It follows that $Z = Re^{i\Theta}$ where $R > 1$ and $0 < \Theta < 2\pi$; R has density $r \mapsto \beta r^{-(\beta+1)}$ (with respect to Lebesgue measure); Θ is uniformly distributed on $(0, 2\pi)$; and R and Θ are independent.

17 Positive Sub-semigroups of $GL(2)$

17.1 Definitions

Recall the *general linear group* $GL(n)$ of invertible matrices in $\mathbb{R}^{n \times n}$ under the operation of matrix multiplication and with the relative topology. A subgroup of $GL(2)$ is

$$G = \left\{ \begin{bmatrix} x & y \\ 0 & 1 \end{bmatrix} : x > 0, y \in \mathbb{R} \right\}$$

Clearly G can be identified with $(0, \infty) \times \mathbb{R}$, with the operation given by

$$(x, y)(u, v) = (xu, xv + y)$$

In this notation, identity is $e = (1, 0)$ and the inverse of (x, y) is $(1/x, -y/x)$. According to Halmos [15], the left-invariant measure of G (unique up to multiplication by positive constants) is given by

$$d\lambda(x, y) = \frac{1}{x^2} dx dy$$

Let $T = \{(x, y) : x \geq 1, y \geq 0\}$.

Lemma 10. T is a sub-semigroup of G. Moreover, T is a positive semigroup.

Proof. Clearly T is closed: If $(x, y) \in T$ and $(u, v) \in T)$ then $(x, y)(u, v) = (xu, xv + y) \in T$. The identity element $e \in T$. If $\boldsymbol{x} \in T$ but $\boldsymbol{x} \neq e$ then $\boldsymbol{x}^{-1} \notin T$ so T has no non-trivial inverses. The other properties are inherited. □

Lemma 11. The partial order associated with $(T, \cdot)$ is the ordinary product order:

$$(x, y) \preceq (z, w) \text{ if and only if } x \leq z,\ y \leq w$$

Proof. Suppose that $(x, y) \preceq (z, w)$. Then there exists $(s, t) \in T$ such that $(x, y)(s, t) = (z, w)$. That is, $xs = z$ and $xt + y = w$. But $s \geq 1$ so $x \leq z$ and $xt \geq 0$ so $y \leq w$. Conversely, suppose that $(x, y) \in T$, $(z, w) \in T$ and that $x \leq z$ and $y \leq w$. Let $s = z/x$ and $t = (w - y)/x$. Then $s \geq 1$, $t \geq 0$ and $xs = z$, $xt + y = w$. Thus $(s, t) \in T$ and $(x, y)(s, t) = (z, w)$, so $(x, y) \preceq (z, w)$. □

Thus, the poset $(T, \preceq)$ is associated with two very different positive semigroups. One is the direct product of $([1, \infty), \cdot)$ and $([0, \infty), +)$, that is,

$$(x, y)(u, v) = (xu, y + v)$$

The other is the semigroup considered in this chapter, corresponding to matrix multiplication

$$(x, y)(u, v) = (xu, xv + y)$$

17.2 Exponential distributions

Proposition 90. T has no memoryless distributions and hence no exponential distributions.

Proof. Note that $(1, k) \prec (1, k + 1)$ for $k \in \mathbb{N}$ and hence

$$(1, k)T \downarrow \emptyset \text{ as } k \to \infty$$

To see this, note that if $(x, y) \in (1, k)T$ for all $k \in \mathbb{N}$ then $y < k$ for all $k \in \mathbb{N}$. Now suppose that F is a nontrivial continuous homomorphisms from $(T, \cdot)$ into $((0, 1], \cdot)$. Then F must satisfy

$$F(xu, xv + y) = F(x, y)F(u, v), \quad (x, y) \in T, \, (u, v) \in T \tag{76}$$

Letting $x = u = 1$ we have $F(1, y + v) = F(1, y)F(1, u)$ for $y \geq 0$, $v \geq 0$ so there exists $\beta > 0$ such that $F(1, y) = e^{-\beta y}$ for $y \geq 0$. Next letting $v = y = 0$ we have $F(xu, 0) = F(x, 0)F(u, 0)$ for $x \geq 1$, $u \geq 1$ so there exists $\alpha > 0$ such that $F(x, 0) = x^{-\alpha}$ for $x \geq 1$. But then $(x, y) = (1, y)(x, 0)$ so

$$F(x, y) = F(1, y)F(x, 0) = x^{-\alpha}e^{-\beta}, \quad x \geq 1, y \geq 0$$

But then one more application of the general homomorphism condition (76) gives

$$(xu)^{-\alpha}e^{-\beta(xv+y)} = [x^{-\alpha}e^{-\beta y}][u^{-\alpha}e^{-\beta v}] = (xu)^{-\alpha}e^{-\beta(y+v)}$$

and this forces $\beta = 0$. Therefore $F(x, y) = x^{-\alpha}$ for $(x, y) \in T$. In particular, $F(1, k) = 1$ But in order for F to be the upper probability function for a measure on T we would have to have $F(1, k) \to 0$ as $k \to \infty$ Therefore, there are no memoryless, and hence no exponential distributions in S. $\square$

Any positive semigroup of $GL(2)$ that contains T as a sub-semigroup will also fail to have memoryless and hence exponential distributions. For example, let

$$S = \left\{ \begin{bmatrix} x & y \\ 0 & z \end{bmatrix} : x \geq 1, \, xz \geq 1 \right\}$$

Then S is a positive sub-semigroup (with nonempty interior) of the subgroup of upper triangular nonsingular matrices. This subgroup has left-invariant measure given by

$$d\lambda \begin{bmatrix} x & y \\ 0 & z \end{bmatrix} = \frac{1}{x^2 z} dx \, dy \, dz$$

However, S has no memoryless distributions. Next let

$$S = \left\{ \begin{bmatrix} x & y \\ w & z \end{bmatrix} : x \ge 1, \, xz - wy \ge 1 \right\}$$

Then S satisfies the basic assumptions and is a sub-semigroup (with nonempty interior) of the group $GL(2)$ itself. This group has left-invariant measure given by

$$d\lambda \begin{bmatrix} x & y \\ w & z \end{bmatrix} = \frac{1}{(xz - wy)^2} dx\, dy\, dw\, dz$$

However, S has no memoryless distributions.

Theorem 52. T has no constant rate distributions. (I think)

Proof. Let F denote the upper probabiliy function of a distribution with constant rate α. Then the density function relative to Lebesuge measure is given by $f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$. Hence the density function relative to λ is $g(x, y) = x^2 f(x, y)$. Hence F must satisfy

$$\frac{\partial^2 F}{\partial x \partial y} F(x, y) = \frac{\alpha}{x^2} F(x, y), \quad x > 1, y > 0$$

with boundary conditions $F(1, 0) = 1$, $F(x, y) \to 0$ as $x \to \infty$ or as $y \to \infty$. I don't think there are any solutions. $\square$

17.3 Discrete positive semigroups

Now let $T = \mathbb{N}_+ \times \mathbb{N}$ with the same operation as before:

$$(x, y)(u, v) = (xu, xv + y)$$

As before this corresponds to matrix multiplication if we identify $(x, y) \in T$ with the matrix $\begin{bmatrix} x & y \\ 0 & 1 \end{bmatrix}$. So, we have matrices with integer entries. For future reference, note that

$$\begin{aligned}
&(x, 0)(y, 0) = (y, 0)(x, 0) = (xy, 0), \quad x, \, y \in \mathbb{N}_+ \\
&(1, x)(1, y) = (1, y)(1, x) = (1, x + y), \quad x, \, y \in \mathbb{N} \\
&(1, 1)(x, y) = (x, y + 1), \; (x, y)(1, 1) = (x, x + y), \quad x \in \mathbb{N}_+, \; y \in \mathbb{N} \\
&(n, 0)(x, y) = (nx, ny), \; (x, y)(n, 0) = (nx, y), \quad n, \, x \in \mathbb{N}_+; \; y \in \mathbb{N}
\end{aligned}$$

Theorem 53. $(T, \cdot)$ is a discrete positive semigroup. The associated partial order is

$$(x, y) \preceq (z, w) \text{ if and only if } x | z, \; y \le w, \; x | (w - y)$$

The minimum element is $(1, 0)$ and the set of irreducible elements is

$$I = \{(1, 1)\} \cup \{(n, 0) : n \text{ is prime } \}$$

18 Trees

There are a variety of semigroup operations that can be imposed on trees and other graphs.

18.1 Merging trees

Consider the collection S of finite, rooted trees. The *merge* operation is defined on S as follows: for x, $y \in S$ let $x + y$ be the new tree obtained by merging (identifying) the roots of x and y; this merged vertex is the root of the new tree.

Proposition 91. $(S, +)$ is a commutative positive semigroup.

Proof. Clearly the operation is associative and commutative. The (degenerate) tree e which consists of a single vertex is the identity. The left-cancellation law holds: if x, y, and z are rooted trees and the merged trees $x+y$ and $x+z$ are the same, then y and z must be the same. Finally, there are no non-trivial inverses: if x merged with y produces the single vertex e then clearly $x = y = e$. □

Let $m(x)$ denote the number of edges in $x \in S$, so that $m(x)+1$ is the number of vertices. Clearly, m is additive with respect to the semigroup operation:

$$m(x + y) = m(x) + m(y)$$

To understand the associated partial order, suppose that x, $y \in S$ and $x \neq e$. Then $x \preceq y$ if and only if x is a subtree of y, with the same root, and the root of y has degree at least 2.

Thus, the irreducible elements of $(S, +)$ are the trees whose roots have degree 1. There are lots of these; in fact clearly, the set of irreducible trees can be put into one-to-one correspondence with S itself. We can associate with $x \in S$ a unique irreducible tree $\hat{x}$ by adding a new vertex to the root of x by a new edge. The new vertex is the root of $\hat{x}$.

19 Other Examples

We briefly describe related models that are not positive semigroups.

19.1 Equivalence relations and partitions

Consider a countably infinite set, which might as well be $\mathbb{N}_+$, the set of positive integers. Consider the set S of equivalence relations on $\mathbb{N}_+$, or equivalently, the collection of partitions of S. We can identify elements of S with functions $x : \mathbb{N}_+ \times \mathbb{N}_+ \to \{0, 1\}$ that have the properties

1. $x(i, i) = 1$ for each $i \in \mathbb{N}_+$ (reflexive property).
2. $x(i, j) = x(j, i)$ for all $i, j \in \mathbb{N}_+$ (symmetric property).

3. $x(i,j) = 1$ and $x(j,k) = 1$ imply $x(i,k) = 1$ for all $i, j, k \in \mathbb{N}_+$ (transitive property).

The equivalence relation associated with x is $i \sim j$ if and only if $x(i,j) = 1$. Equivalently, $x(i,\cdot)$ is the indicator function of the equivalence class generated by i.

A natural operation on S is

$$(xy)(i,j) = x(i,j)y(i,j) = \min\{x(i,j), y(i,j)\}, \quad i, j \in \mathbb{N}_+$$

If we view x as a partition (that is, a collection of non-empty, disjoint subsets whose union is $\mathbb{N}_+$, then

$$xy = \{a \cap b : a \in x, b \in y, a \cap b \neq \emptyset\}$$

In terms of equivalence relations on $\mathbb{N}_+$,

$$i \sim_{xy} j \text{ if and only if } i \sim_x j \text{ and } i \sim_y j$$

Clearly the operation makes S a commutative semigroup. There is an identity e given by $e(i,j) = 1$ for all i, j; This corresponds to a partition with just one element—the set $\mathbb{N}_+$; as an equivalence relation, all elements are equivalent. There are no non-trivial invertible elements, since $xy = e$ clearly implies $x = y = e$. However, the left-cancellation law does not hold. In particular, $x^2 = x$ for any $x \in S$.

A natural partial order on S is *refinement*: $x \preceq y$ if and only if y refines x; that is, $y(i,j) = 1$ implies $x(i,j) = 1$ or equivalently, if $b \in y$ then there exists $a \in x$ with $b \subseteq a$. The partial order is compatible with the semigroup operation in the sense of a previous proposition: $x \preceq y$ if and only if $xy = x$.

Also, S has a maximum element m defined by $m(i,j) = 1$ if and only if $i = j$; m corresponds to the equality relation $=$ and partitions $\mathbb{N}_+$ into singletons.

In summary, $(S, \cdot, \preceq)$ has some of the properties of a positive semigroup, but not all. Clearly, no sub-semigroup of S could be a positive semigroup either.

References

[1] TA Azlarov and NA Volodin. *Characterization Problems Associated with the Exponential Distribution.* Springer-Verlag, Berlin, 1986.

[2] AP Basu. "Bivariate failure rate". *Journal of the American Statistical Association*, **66** (1971), 103—104.

[3] C Berge. *Principles of Combinatorics*, Academic Press, New York, 1971.

[4] HW Block and AP Basu. "A continuous bivariate exponential extension". *Journal of the American Statistical Association*, **69** (1974), 1031—1037.

[5] AH Clifford and GB Preston. *The Algebraic Theory of Semigroups.* American Mathematical Society, Providence, Volume I, 2'nd edition, 1964; Volume II, 1867.

[6] F Downton. "Bivariate exponential distributions in reliability theory". *Journal of the Royal Statistical Society*, Series B, **32** (1970), 408—417.

[7] JE Freund. "A bivariate extension of the exponential distribution". *Journal of the American Statistical Association*, **56** (1961), 971—977.

[8] J Galambos and S Kotz, "Some characterizations of the exponential distribution via properties of the geometric distribution" in *Contributions in Statistics*, PK Sen Editor, North Holland (1983), 159–162.

[9] IB Gertxbakh. *Statistical Reliability Theory.* Marcel Dekker, New York, 1989.

[10] EJ Gumbel. "Bivariate exponential distributions". *Journal of the American Statistical Association*, **55** (1960), 698—707.

[11] A Gut, "Some remarks on the Riemann zeta distribution" *Uppsala University Department of Mathematics Report 2005:6* (2005)

[12] IB Gertsbakh. *Statistical Reliability Theory.* Marcel Dekker, New York, 1989.

[13] G Högnäs and A Mukherjea, *Probability Measures on Semigroups*, Plenum Press (1995).

[14] AG Hawkes. "A bivariate exponential distribution with applications to reliability". *Journal of the Royal Statistical Society*, Series B, **34** (1972), 129—131.

[15] PR Halmos. *Measure Theory.* Springer-Verlag, Berlin, 1974.

[16] T Hida and H Kuo. "Semigroups associated with generalized Brownian functionals". *Semigroup Forum*, **45**, 261—263.

[17] T Kamae and U Krengel. "Stochastic Partial Ordering". *The Annals of Probability*, **6** (1978), 1044—1049.

[18] GD Lin and CY Hu, "The Riemann zeta distribution", *Bernoulli* **7** (2001), 817—828

[19] AW Marshall and I Olkin. "A multivariate exponential distribution". *Journal of the American Statistical Association*, **62** (1967), 30—44.

[20] G Matheron. *Random Sets and Integral Geometry*, Wiley, 1975.

[21] A Mukherjea and NA Tserpes. *Measures on Topological Semigroups.* Lecture Notes in Mathematics 547, Springer-Verlag, Berlin, 1976.

[22] L Nachbin. *Topology and Order.* Van Nostrand, 1965.

[23] I Niven, HS Zuckerman, and HL Montgomery. *An Introduction to the Theory of Numbers*, 5th edition, Wiley 1991.

[24] M Pendergrass, private communication

[25] PS Puri and H Rubin. “On a characterization of the family of distributions with constant multivariate failure rates”. *The Annals of Probability*, **2** (1974), 738—740.

[26] CR Rao and H Rubin, “On a characterization of the Poisson distribution” *Sankhya, Ser. A*, **41** (1964), 295—298.

[27] GH Rowell. *Probability Distributions on temporal Semigroups*, PhD Dissertation, University of Alabama in Huntsville, 1995.

[28] GH Rowell and K Siegrist. “Relative aging of distributions”, *Probability in the Engineering and Informational Sciences* **12** (1998), 469—478.

[29] D Roy. “A characterization of Gumbel’s bivariate exponential and Lindley and Singpurwalla’s bivariate Lomax distributions”. *Journal of Applied Probability*, **27** (1989), 886—891.

[30] IZ Ruzsa and GJ Székely. *Algebraic Probability Theory.* Wiley, New York, 1988.

[31] GJ Székely. “Harr measure on semigroups”, *Colloquia Mathematica Societatis Jànos Bolyai* **49** (1965), 917—928.

[32] SK Sarkar. “A continuous univariate exponential distribution”. *Journal of the American Statistical Association*, **82** (1987), 667—675.

[33] K Siegrist. “Exponential distributions on semigroups”, *Journal of Theoretical Probability* **7** (1994), 725—737

[34] K Siegrist “Decomposition of exponential distributions on positive semigroups”, *Journal of Theoretical Probability.* **19** (2006), 204–220.

[35] K Siegrist “Random, finite subsets with exponential distributions”, *Probability in the Engineering and Informational Sciences*, **21** (2007), 117—131

[36] K Siegrist “Exponential and gamma distributions on positive semigroups, with applications to Dirichlet distributions”, *Bernoulli*, **13** (2007), 330—345.