Probability in the Engineering and Informational Sciences, **21**, 2007, 117–132. Printed in the U.S.A.

RANDOM FINITE SUBSETS WITH EXPONENTIAL DISTRIBUTIONS

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

Let S denote the collection of all finite subsets of $\mathbb{N}_+$. We define an operation on S that makes S into a positive semigroup with set inclusion as the associated partial order. Positive semigroups are the natural home for probability distributions with exponential properties, such as the memoryless and constant rate properties. We show that there are no exponential distributions on S, but that S can be partitioned into subsemigroups, each of which supports a one-parameter family of exponential distributions. We then find the distribution on S that is closest to exponential, in a certain sense. This work might have applications to the problem of selecting a finite sample from a countably infinite population in the most random way.

1. INTRODUCTION

The motivation for this article is the problem of selecting a finite set of positive integers in "the most random way possible." The phrase in quotes is ambiguous, of course, but we will at least obtain a partial solution by considering the collection of finite sets as a positive semigroup $(S, \cdot)$ whose associated partial order is set inclusion. Positive semigroups are the natural mathematical home for probability distributions with exponential-type properties, including the memoryless and constant failure rate properties, special properties related to uniform distributions, and maximum entropy properties. All of these relate, in some sense, to the degree of randomness of the distribution.

More specifically, our goal is to describe the semigroup $(S, \cdot)$ and study probability distributions on S that have exponential properties. We will show that there are no exponential distributions on S, but that S naturally partitions into a countable

 0269-9648/07 $16.00

collection of subsemigroups, each of which supports a one-parameter family of exponential distributions. We will then construct a two-parameter family of distributions on S that are close to exponential in a strong sense. For basic information about probability distributions on semigroups, see Högnäs and Mukherjea [2]. For properties and characterizations of the standard exponential distribution, see Azlarov and Volodin [1]. For the general theory of random sets, see Matheron [3].

2. POSITIVE SEMIGROUPS

A *semigroup* $(S, \cdot)$ consists of a set S and a binary operation $\cdot$ on S that is associative:

$$(xy)z = x(yz), \qquad x, y, z \in S.$$

A *positive semigroup* is a semigroup $(S, \cdot)$ that has an identity e, satisfies the left cancellation law, and has no nontrivial inverses; that is, for all $x, y, z \in S$, the following hold:

1. $xe = ex = x$.
2. $xy = xz$ implies $y = z$.
3. $xy = e$ implies $x = y = e$.

The relation $\preccurlyeq$ on S defined by $x \preccurlyeq y$ if and only if $y = xt$ for some $t \in S$ is a *partial order* on S; that is, $x \preccurlyeq x$ for each $x \in S$ (the reflexive property), $x \preccurlyeq y$ and $y \preccurlyeq x$ imply $x = y$ (the antisymmetric property), and $x \preccurlyeq y$ and $y \preccurlyeq z$ imply $x \preccurlyeq z$ (the transitive property). If $x \preccurlyeq y$, then $t \in S$ satisfying $xt = y$ is unique and is denoted $x^{-1}y$. For each $x \in S$, the mapping $t \to xt$ is an order isomorphism from S onto the set $xS = \{y \in S : x \preccurlyeq y\}$; that is, $x \preccurlyeq y$ if and only $xt \preccurlyeq yt$ for all $x, y, t \in S$. Indeed, the algebraic assumptions are precisely the ones needed for the partially ordered set $(S, \preccurlyeq)$ to have this self-similarity property.

Positive semigroups are often found embedded in groups. Specifically, suppose that $(G, \cdot, \preccurlyeq)$ is a *left-ordered group*; that is, $(G, \cdot)$ is a group and $\preccurlyeq$ is a partial order on G satisfying $x \preccurlyeq y \Rightarrow zx \preccurlyeq zy$ for $x, y, z \in G$. Let e denote the identity element of G and let $S = \{x \in G : x \succcurlyeq e\}$ denote the set of *positive elements* of G. Then $(S, \cdot)$ is a positive semigroup and $\preccurlyeq$ restricted to S is the associated partial order. Conversely, suppose that $(G, \cdot)$ is a group and that S is a positive subsemigroup of G. For $x, y \in G$, define $x \preccurlyeq y$ if and only if $xt = y$ for some $t \in S$. Then $(G, \cdot, \preccurlyeq)$ is a left-ordered group with S as the set of positive elements. On the other hand, as this article hopefully illustrates, there are interesting positive semigroups that cannot be embedded in groups.

In general, topological and measure-theoretic assumptions are imposed on S as well, but in this article we will only be concerned with discrete semigroups (where S is countably infinite). In this case, the one additional assumption needed is that $[e, x] = \{t \in S : t \preccurlyeq x\}$ is finite for each $x \in S$, so that the partially ordered set $(S, \cdot)$ is *locally finite*. Note that counting measure # is *left-invariant* for $(S, \cdot)$; that is, $\#(xA) = \#(A)$ for $x \in S$ and $A \subseteq S$. Moreover, # is the unique left-invariant measure up to multiplication by positive constants.

https://doi.org/10.1017/S0269964807070088 Published online by Cambridge University Press

If T is a nonempty subset of S and is closed under the operation $\cdot$, then $(T, \cdot)$ is also a positive semigroup, since the other algebraic assumptions are simply inherited. However, the partial order associated with T is not, in general, the restriction of $\preccurlyeq$ to T, but a subpartial order of this restriction; that is, if $x, y \in T$, then $x \preccurlyeq_T y$ implies $x \preccurlyeq y$, but the converse is not true unless $x^{-1}y \in T$. If T is closed under $\cdot$ but does not contain e, then $T \cup \{e\}$ is also a closed under $\cdot$ and, hence, is a positive subsemigroup of S.

3. EXPONENTIAL DISTRIBUTIONS

Suppose that $(S, \cdot)$ is a (discrete) positive semigroup and that X is a random variable taking values in S (so that $P(X = x) > 0$ for each $x \in S$). The *tail probability function* of X is the mapping $x \to P(X \succcurlyeq x)$. In general, this function does not uniquely determine the distribution of X.

Random variable X has an *exponential distribution* if

$$P(X \in xA) = P(X \succcurlyeq x)P(X \in A), \qquad x \in S, A \subseteq S. \tag{1}$$

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

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

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

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

for some positive constant α. In general,

$$\sum_{x \in S} P(X \succcurlyeq x) = E(\#[e, X]),$$

so that if X has constant rate, then the rate constant must be

$$\alpha = \frac{1}{E(\#[e, X])}.$$

There are a number of nice properties and characterizations of exponential distributions on positive semigroups, perhaps a bit surprising given the minimal algebraic assumptions. We mention a few of these; for more details, see Siegrist [5–7]. First, X has an exponential distribution on S if and only if X is memoryless and has constant rate. In general, however, a distribution can have one of these properties (memoryless or constant rate), but not the other. Second, $F : S \to (0,1]$ is the tail

https://doi.org/10.1017/S0269964807070088 Published online by Cambridge University Press

probability function of an exponential distribution on S if and only if $F(xy) = F(x)F(y)$ for all $x, y \in S$ and $\sum_{x \in S} F(x) < \infty$ (the reciprocal of this sum is then the rate constant). This characterization prescribes a method for finding all exponential distributions on a given positive semigroup. Finally, suppose that X and Y are independent and identically distributed (i.i.d.) on S. Then the common distribution is exponential if and only if the conditional distribution of X given $XY = z$ is uniform on $[e, z]$ for each $z \in S$. Furthermore, if $X_1, X_2, \ldots$ are i.i.d. exponential and $Y_n = X_1 \cdots X_n$ is the corresponding "gamma" variable of order $n \in \mathbb{N}_+$, then the conditional distribution of $(Y_1, \ldots, Y_n)$ given $Y_{n+1} = z$ is uniform on the set

$$\{(y_1, y_2, \ldots, y_n) \in S^n : y_1 \preccurlyeq y_2 \preccurlyeq \cdots \preccurlyeq y_n\}.$$

For general positive semigroups, counting measure # is replaced by a left-invariant measure λ; sums become integrals with respect to λ; and density functions and uniform distributions are with respect to λ as well. At least in terms of the algebraic structure, exponential distributions specify the most random way to select elements of S. Of course, the motivating example for this theory is the positive semigroup $([0,\infty),+)$. The associated partial order is the ordinary order and Lebesgue measure is the invariant measure. The exponential distributions for this positive semigroup are the ordinary exponential distributions. Rowell and Siegrist [4] explore positive semigroups isomorphic to the standard one, in the context of reliability theory. We briefly mention a few discrete examples.

Example 1: Let $\mathbb{N}$ denote the set of nonnegative integers. Then $(\mathbb{N},+)$ is a positive semigroup, and the associated partial order is the ordinary order $\leq$. The exponential distribution with rate parameter $p \in (0,1)$ has tail probability function and density function given by

$$P(X \geq x) = (1-p)^x, \qquad P(X = x) = p(1-p)^x, \qquad x \in \mathbb{N}.$$

Of course, this is the standard geometric distribution with success parameter p.

Example 2: Let $\mathbb{N}_+$ denote the set of positive integers. Then $(\mathbb{N}_+,\cdot)$ is a positive semigroup where $\cdot$ is ordinary multiplication. The associated partial order is the division order: $x \preccurlyeq y$ if and only if x divides y. The exponential distributions turn out to be precisely the Dirichlet distributions with completely multiplicative coefficient functions; the zeta distribution is an important special case (see Siegrist [7]).

Example 3: Let A be a finite alphabet and let $S = A^*$ denote the space of all finite words with letters from A. The free semigroup $(S,\cdot)$ is a positive semigroup where $\cdot$ is the concatenation operation. The identity is the "empty word" e, and $x \preccurlyeq y$ if and only if x is a prefix of y. An exponentially distributed variable has the form

$$X = X_1 \ldots X_L,$$

where the random letters $X_1, X_2, \ldots$ are i.i.d. on the alphabet A and where the length L is independent of $(X_1, X_2, \ldots)$ and has a geometric distribution on $\mathbb{N}$.

https://doi.org/10.1017/S0269964807070088 Published online by Cambridge University Press

4. THE POSITIVE SEMIGROUP OF FINITE SUBSETS

Let S denote the set of all finite subsets of $\mathbb{N}_+$. It is easy to see that the partially ordered set $(S, \subseteq)$ satisfies the self-similarity property described in Section 2. Our goal in this section is to define and study the corresponding positive semigroup.

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 or with domain $\mathbb{N}_+$ if x is infinite. 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. Thus, $\#(x) \le \max(x)$ for every x. If x and y are nonempty subsets of $\mathbb{N}_+$ with $\max(y) \le \#(x)$, we let $x \circ y$ denote the subset whose function is the ordinary 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.

We now define a 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. The operation might seem contrived, but it is not. Up to a relabeling of the positive integers, there is only one way to associate a positive semigroup with the subset partial order.

THEOREM 1: *$(S,\cdot)$ 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

$$x\emptyset = x \cup (x^c \circ \emptyset) = x \cup \emptyset = x,$$

$$\emptyset x = \emptyset \cup (\mathbb{N}_+ \circ x) = \emptyset \cup x = x.$$

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. However, then $y = z$. There are no nontrivial 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 $xu = y$. ■

https://doi.org/10.1017/S0269964807070088 Published online by Cambridge University Press

Note that the irreducible elements of $(S, \cdot)$ are the singletons $\{i\}$, where $i \in \mathbb{N}_+$. Note also that

$$\{i\}\{i\} = \{i, i+1\},$$
$$\{i+1\}\{i\} = \{i, i+1\},$$
$$\{i\}\{i+1\} = \{i, i+2\}.$$

Thus, the binary operation is not commutative and the right-cancellation law does not hold, so $(S, \cdot)$ 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

$$\{i_n\}\{i_{n-1}\} \ldots \{i_1\} = \{i_1, i_2, \ldots, i_n\}.$$

PROPOSITION 1: *For $x, y \in S$,*

$$\#(xy) = \#(x) + \#(y), \tag{2}$$

$$\max(xy) = \begin{cases} \max(x) & \text{if } \max(y) \le \max(x) - \#(x) \\ \max(y) + \#(x) & \text{if } \max(y) > \max(x) - \#(x). \end{cases} \tag{3}$$

PROOF: $\#(xy) = \#[x \cup (x^c \circ y)] = \#(x) + \#(x^c \circ y)$ since x and $x^c \circ y$ are disjoint. However, clearly, $\#(x^c \circ y) = \#(y)$.

Equation (3) 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 $\max(xy) = \max(x^c \circ y) = x^c(\max(y)) = n + (\max(y) - (n - i)) = \max(y) + i$. ■

The semigroup $(S, \cdot)$ has an interesting structure, but to see it we need some additional notation. For $k \in \mathbb{N}$, let

$$S_k = \{x \in S : \max(x) - \#(x) = k\}$$

and let $T_k = \{\emptyset\} \cup S_k$. For $(n, k) \in \{(0,0)\} \cup (\mathbb{N}_+ \times \mathbb{N})$, let

$$S_{n,k} = \{x \in S : \#(x) = n, \max(x) = n + k\} = \{x \in S_k : \#(x) = n\}.$$

Of course, $S_{0,0} = \{\emptyset\}$. If $n \in \mathbb{N}_+$ and $k \in \mathbb{N}$, then

$$\#(S_{n,k}) = \binom{n+k-1}{n-1} \tag{4}$$

https://doi.org/10.1017/S0269964807070088 Published online by Cambridge University Press

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 (4) is valid for $n = k = 0$ also.

THEOREM 2: *S_k is a subsemigroup of S, and hence T_k is a positive subsemigroup of S, for each $k \in \mathbb{N}$. The associated partial order on T_k is the subset partial order.*

PROOF: We first show that $xy \in S_k$ for $x, y \in S_k$. The result is trivial if $x = \emptyset$ or $y = \emptyset$ (which can only happen when $k = 0$). Thus, 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 Proposition 1, $\max(xy) = \max(y) + \#(x)$. Hence,

$$\max(xy) - \#(xy) = (\max(y) + \#(x)) - (\#(x) + \#(y)) = \max(y) - \#(y) = k.$$

Therefore, $xy \in S_k$. To prove the last statement in the theorem, it suffices to show that if $x, y \in S_k$ and $x \subset y$, then $x^{-1}y \in S_k$. Thus, suppose that $x \in S_{m,k}$, $y \in S_{n,k}$, where $m < n$, and $xu = y$ for some $u \in S$ (so that $x \subset y$). Then $\max(y) = n + k > m + k = \max(x)$, so by another application of Proposition 1, $\max(y) = \max(u) + m$ and, hence, $\max(u) = n - m + k$. However, $\#(u) = n - m$ and, hence, $u \in S_k$. ■

Note that $S_0 = \{\{1,2,\ldots,m\} : m \in \mathbb{N}\}$. If $y \in S$ and $\#(y) = n$, 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 $(\mathbb{N},+)$, the positive semigroup in Example 1, and $x \mapsto \#(x)$ is an isomorphism. 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 2: *The set of minimal element of S_k is*

$$M_k = \{x \in S_k : x(i) \le k \text{ for all } i < \#(x)\}.$$

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 $x \in S_k \backslash M_k$, 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 $x(i) - i \le k$ and, hence, $x(i) - k - 1 \le i - 1$, 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 $uv = x$. By Theorem 2, $v \in S_k$ and, hence, x is not a minimal element of S_k.

https://doi.org/10.1017/S0269964807070088 Published online by Cambridge University Press

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,

$$\#(M_k) = \sum_{n=1}^{k+1} \binom{k}{n-1} = 2^k. \qquad \blacksquare$$

For example, the minimal elements of S_2 are $\{3\}$, $\{1,4\}$, $\{2,4\}$, and $\{1,2,5\}$.

5. EXPONENTIAL DISTRIBUTIONS ON T_k

THEOREM 3: *There are no memoryless distributions on S and, hence, no exponential distributions.*

PROOF: Suppose that X is a random variable taking values in S and that X has a memoryless distribution. By the memoryless property,

$$P(\{i\}\{i\} \subseteq X) = P(i \in X)P(i \in X),$$

$$P(\{i+1\}\{i\} \subseteq X) = P(i+1 \in X)P(i \in X).$$

However, $\{i\}\{i\} = \{i+1\}\{i\}$, as noted earlier, so we must have

$$P(i+1 \in X) = P(i \in X)$$

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,

$$\begin{aligned} P(i_1 \in X, i_2 \in X, \ldots, i_n \in X) &= P(\{i_1, i_2, \ldots, i_n\} \subseteq X) \\ &= P(\{i_n\}\{i_{n-1}\}\ldots\{i_1\} \subseteq X) \\ &= P(i_1 \in X)P(i_2 \in X)\ldots P(i_n \in X). \end{aligned}$$

It therefore follows that the events $\{\{i \in X\} : i \in \mathbb{N}_+\}$ are i.i.d. Hence, infinitely many of the events must occur with probability 1, so X is infinite—a contradiction. $\blacksquare$

Although there are no exponential distributions on S, the subsemigroup T_k has a one-parameter family of exponential distributions for each $k \in \mathbb{N}$.

THEOREM 4: *A random variable X taking values in T_k has an exponential distribution if and only if the tail probability function F and density function f have the following form, for some $\alpha \in (0,1)$:*

https://doi.org/10.1017/S0269964807070088 Published online by Cambridge University Press

$$F(x) = \alpha^{\#(x)}, \qquad x \in T_k, \tag{5}$$

$$f(x) = \frac{(1-\alpha)^{k+1}}{(1-\alpha)^{k+1} + \alpha}\, \alpha^{\#(x)}, \qquad x \in T_k. \tag{6}$$

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,

$$\begin{aligned}
\sum_{x \in T_k} F(x) &= 1 + \sum_{n=1}^{\infty} \sum_{x \in S_{n,k}} F(x) \\
&= 1 + \sum_{n=1}^{\infty} \sum_{x \in S_{n,k}} \alpha^n \\
&= 1 + \sum_{n=1}^{\infty} \binom{n+k-1}{n-1} \alpha^n \\
&= \frac{(1-\alpha)^{k+1} + \alpha}{(1-\alpha)^{k+1}}.
\end{aligned}$$

It follows that F and f as given earlier are the tail probability function and density function, respectively, of an exponential distribution.

Conversely, suppose that F is the tail probability function of a memoryless distribution on T_k. As noted earlier, T_0 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 irreducible, 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 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, $x \cup (x^c \circ x) = y \cup (y^c \circ x)$. 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 $j \in y^c \circ x$, 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 $j+1 \in x^c \circ x$, 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 $y \in S_k$ that is not minimal, but with $\#(y) = \#(x)$. It then follows that $F(x) = F(y) = \alpha^{\#(y)} = \alpha^{\#(x)}$ and the proof is complete. ■

https://doi.org/10.1017/S0269964807070088 Published online by Cambridge University Press

To illustrate the last part of the proof, suppose that $x = \{3,4,5,8,15\} \in T_{10}$. Then $j = 5$, $y = \{3,4,6,8,15\}$, and $xx = yx = \{3,4,5,6,7,8,9,12,15,20\}$.

Suppose that X has the exponential distribution on T_k given in Theorem 4. From the general theory in Section 3, the expected number of subsets of X in T_k is the reciprocal of the rate parameter in the density function. Thus,

$$E(\#[\emptyset, X]) = 1 + \frac{\alpha}{(1-\alpha)^{k+1}}.$$

If $k = 0$ (recall that $S_0 = T_0$), note that

$$P(X = x) = (1-\alpha)\alpha^{\#(x)}, \qquad x \in S_0. \tag{7}$$

On the other hand, suppose that $k \in \mathbb{N}_+$. Then

$$P(X \in S_k) = 1 - P(X = \emptyset) = 1 - g(0) = \frac{\alpha}{(1-\alpha)^{k+1} + \alpha}.$$

Thus, the conditional distribution of X given $X \in S_k$ has density function

$$P(X = x | X \in S_k) = \frac{P(X = x)}{P(X \in S_k)} = (1-\alpha)^{k+1}\alpha^{\#(x)-1}, \qquad x \in S_k. \tag{8}$$

The density function of X depends on $x \in T_k$ only through $\#(x)$. The following corollary gives the distribution of $\#(X)$.

Corollary 1: *Suppose that X has the exponential distribution in Theorem 4 and let $U = \#(X)$. Then*

$$P(U = n) = \frac{(1-\alpha)^{k+1}}{(1-\alpha)^{k+1} + \alpha} \binom{n+k-1}{k} \alpha^n, \qquad n \in \mathbb{N}, \tag{9}$$

$$E(U) = \frac{\alpha}{1-\alpha} \frac{\alpha(1+k\alpha)}{(1-\alpha)^{k+1} + \alpha}, \tag{10}$$

where we interpret the binomial coefficient as 1 when $n = 0$.

When $k = 0$, (9) gives $P(U = n) = (1-\alpha)\alpha^n$ for $n \in \mathbb{N}$, so U has a geometric distribution on $\mathbb{N}$. In general, U has a modified negative binomial distribution. It is easy to see from (10) that for each $k \in \mathbb{N}$, $E(U)$ is a strictly increasing function of α and maps $(0,1)$ onto $(0,\infty)$. Thus, the exponential distribution on T_k can be reparameterized by expected cardinality. Moreover, the exponential distribution maximizes entropy with respect to this parameter:

Corollary 2: *The exponential distribution in Theorem 4 maximizes entropy over all distributions on T_k with expected value given by (10).*

https://doi.org/10.1017/S0269964807070088 Published online by Cambridge University Press

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

$$-\sum_{x\in T_k} g(x)\ln[g(x)] \le -\sum_{x\in T_k} g(x)\ln[f(x)]. \tag{11}$$

If X has the exponential distribution in Theorem 4 and if $E(\#(Y)) = E(\#(X))$, then substituting into the right-hand side of (11), 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 (6) and $\mu_{k,\alpha}$ is the mean cardinality in (10). Of course, the entropy of X achieves this upper bound. ■

6. 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 (7) and (8). Thus, suppose that X takes values in S with probability mass function

$$P(X=x) = \begin{cases} \beta_0(1-\alpha_0)\alpha_0^{\#(x)}, & x \in S_0 \\ \beta_k(1-\alpha_k)^{k+1}\alpha_k^{\#(x)-1}, & x \in S_k, k \in \mathbb{N}_+, \end{cases} \tag{12}$$

where $\alpha_k, \beta_k \in (0,1)$ 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. In fact, we cannot 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 subsemigroups S_k, and these semigroups partition S.

There is not much that we can say about the general distribution in (12). In the remainder of this section we will study a special case with particularly nice properties. For our first construction, let N have a geometric distribution on $\mathbb{N}$ with rate parameter $1-r \in (0,1)$, as in Example 1. Next, given $N=n$, random variable 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 5: *For $x \in S$,*

$$P(X=x) = \frac{1-r}{1-r+rp}(rp)^{\#(x)}[r(1-p)]^{\max(x)-\#(x)}, \tag{13}$$

$$P(X \supseteq x) = p^{\#(x)}r^{\max(x)}. \tag{14}$$

https://doi.org/10.1017/S0269964807070088 Published online by Cambridge University Press

PROOF: For $x \in S$,

$$P(X = x) = \sum_{n=0}^{\infty} P(N = n) P(X = x \mid N = n).$$

If $n < \max(x)$, then x is not a subset of $\{1, 2, \ldots, n\}$, so $P(X = x \mid N = n) = 0$. If $n \geq \max(x)$, then x is a subset of $\{1, 2, \ldots, n\}$ and, by assumption, $P(X = x \mid N = n) = p^{\#(x)}(1 - p)^{n - \#(x)}$. Substituting gives

$$P(X = x) = \sum_{n=\max(x)}^{\infty} (1 - r) r^n p^{\#(x)} (1 - p)^{n - \#(x)},$$

which simplifies to (13). By a similar argument,

$$P(X \supseteq x) = \sum_{n=\max(x)}^{\infty} (1 - r) r^n p^{\#(x)},$$

which simplifies to (14). ■

The distribution of X depends on $x \in S$ only through $\#(x)$ and $\max(x) - \#(x)$. As before, let $U = \#(X)$ and now let $V = \max(X) - \#(X)$.

COROLLARY 3: *For* $(n, k) \in \{(0,0)\} \cup (\mathbb{N}_+ \times \mathbb{N})$,

$$P(U = n, V = k) = P(X \in S_{n,k}) = \frac{1 - r}{1 - r + rp} \binom{n + k - 1}{n - 1} (rp)^n [r(1 - p)]^k.$$

COROLLARY 4: *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 5: *For* $n \in \mathbb{N}$, *the conditional distribution of* V *given* $U = n$ *is negative binomial with parameters* n *and* $r(1 - p)$ *(when* $n = 0$, *the conditional distribution of* V *is point mass at* 0*).*

COROLLARY 6: *The distribution of* U *is geometric with parameter* $(1 - r)/(1 - r + rp)$.

Of course, Corollaries 4–6 determine the distribution of X and give an alternate way of constructing the distribution 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 $U = n$

https://doi.org/10.1017/S0269964807070088 Published online by Cambridge University Press

and $V = k$, we give X the uniform distribution on $S_{n,k}$. The two constructions are equivalent, since there is a one-to-one correspondence between the pairs of parameters (r, p) and (a, b).

Our next goal is to study the distribution of the random subset X on the subsemigroups S_k. First, note that

$$\frac{P(X = x)}{P(X \supseteq x)} = \frac{1 - r}{1 - r + rp} (1 - p)^{\max(x) - \#(x)}.$$

Thus, for $k \in \mathbb{N}$, X has constant rate

$$\frac{1 - r}{1 - r + rp} (1 - p)^k$$

on the subsemigroup S_k. In particular, for $x \in S_0$,

$$P(X = x) = \frac{1 - r}{1 - r + rp} (rp)^{\#(x)},$$

$$P(X \supseteq x) = (rp)^{\#(x)}.$$

Hence, X has the memoryless property on S_0 (in addition to the constant rate property). To find the conditional distribution of X given $X \in S_k$, we first need $P(X \in S_k)$ or, equivalently, the probability density function of V.

COROLLARY 7: *V has a modified geometric distribution:*

$$P(X \in S_0) = P(V = 0) = \frac{1 - r}{(1 - r + rp)(1 - rp)},$$

$$P(X \in S_k) = P(V = k) = \frac{(1 - r)rp}{(1 - r + rp)(1 - rp)} \left(\frac{r(1 - p)}{1 - rp} \right)^k, \qquad k \in \mathbb{N}_+.$$

COROLLARY 8: *The conditional distributions of X on S_k are as follows:*

$$P(X = x | X \in S_0) = (1 - rp)(rp)^{\#(x)}, \qquad x \in S_0, \tag{15}$$

$$P(X = x | X \in S_k) = (1 - rp)^{k+1} (rp)^{\#(x)-1}, \qquad x \in S_k, k \in \mathbb{N}_+. \tag{16}$$

Thus, X has an almost exponential distribution in the sense of (12), with $\alpha_k = 1 - rp$ for each $k \in \mathbb{N}$ and with the mixing probabilities given in Corollary 7.

From Theorem 3, 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 5.

https://doi.org/10.1017/S0269964807070088 Published online by Cambridge University Press

COROLLARY 9: *Suppose that X has the distribution in Theorem 5.*

1. $P(i \in X) = pr^i$ *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*

$$\begin{aligned} P(i_n \in X | i_1 \in X, \ldots, i_{n-1} \in X) &= P(i_n \in X | i_{n-1} \in X) \\ &= P(i_n - i_{n-1} \in X) \\ &= pr^{i_n - i_{n-1}}. \end{aligned}$$

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* $P(i \in X | j \in X) = p$ *for* $i < j$.

Property 3 in Corollary 9 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

$$P(i + j \in X | i \in X) = P(j \in X) = pr^j \to 0 \text{ as } j \to \infty.$$

From Corollaries 5 and 6, we can compute the expected value of $U = \#(X)$ and $W = \max(X) = U + V$:

$$E(U) = \frac{rp}{1 - r}, \tag{17}$$

$$E(W) = \frac{rp}{(1 - r)(1 - r + rp)}. \tag{18}$$

It is easy to see from (17) and (18) that $(E(U), E(W))$, 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 reparameterized by expected cardinality and expected maximum. 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 2

COROLLARY 10: *The distribution in Theorem 5 maximizes entropy among all distributions on S with expected cardinality given by (17) and expected maximum given by (18).*

Of fundamental importance in the general theory of random sets [3] is the hitting probability function G:

$$G(x) = P(X \cap x \neq \emptyset), \qquad x \subseteq \mathbb{N}_+$$

This function completely determines the distribution of a random set, and in the general setting (which lacks the algebraic structure that we have here), it plays the

https://doi.org/10.1017/S0269964807070088 Published online by Cambridge University Press

role of a "distribution function." Note that G is defined for all subsets of the positive integers, not just finite subsets.

THEOREM 6: *Suppose that X has the almost exponential distribution with parameters p and r given in Theorem 5. Then*

$$G(x) = \sum_{i=1}^{\#(x)} p(1-p)^{i-1} r^{x(i)}, \qquad x \subseteq \mathbb{N}_+,$$

where, as usual, $x(i)$ is the ith smallest element of x.

PROOF: Suppose first that x is finite (so that $x \in S$). From the standard inclusion–exclusion formula (or from [3]),

$$G(x) = \sum_{k=1}^{\#(x)} (-1)^{k-1} \sum_{y \subseteq x, \#(y)=k} P(X \supseteq y).$$

Hence, substituting the result in (14), 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 passing to the limit and using the continuity of probability. ■

Acknowledgment

I am grateful to Marcus Pendergrass for suggesting this model.

References

1. Azlarov, A.T. & Volodin, N.A. (1986). *Characterization problems associated with the exponential distribution*. Berlin: Springer-Verlag.
2. Högnäs, G. & Mukherjea, A. (1995). *Probability measures on semigroups*. New York: Plenum Press.
3. Matheron, G. (1975). *Random sets and integral geometry*. New York: Wiley.

4. Rowell, G.H. & Siegrist, K. (1998). Relative aging of distributions. *Probability in the Engineering and Informational Sciences* 12: 469–478.
5. Siegrist, K. (1994). Exponential distributions on semigroups. *Journal of Theoretical Probability* 7: 725–737.
6. Siegrist, K. (2006). Decomposition of exponential distributions on positive semigroups. *Journal of Theoretical Probability* 19: 204–220.
7. Siegrist, K. (2007). Exponential and gamma distributions on positive semigroups, with applications to Dirichlet distributions. *Bernoulli* (to appear).