SCALE-FREE AND POWER LAW DISTRIBUTIONS VIA FIXED POINTS AND CONVERGENCE OF (THINNING AND CONDITIONING) TRANSFORMATIONS

RICHARD ARRATIA, THOMAS M. LIGGETT, AND MALCOLM J. WILLIAMSON

Abstract. In discrete contexts such as the degree distribution for a graph, scale-free has traditionally been defined to be power-law. We propose a reasonable interpretation of scale-free, namely, invariance under the transformation of p-thinning, followed by conditioning on being positive.

For each $\beta \in (1,2)$, we show that there is a unique distribution which is a fixed point of this transformation; the distribution is power-law-$\beta$, and different from the usual Yule–Simon power law-$\beta$ that arises in preferential attachment models.

In addition to characterizing these fixed points, we prove convergence results for iterates of the transformation.

CONTENTS

1. Abstract notions of scale-free 1
2. Introduction in the context of graphs 2
2.1. “Scale-free” and “power-law” 2
2.2. Thinning by tossing a $p$-coin 3
2.3. Thinning a graph, by nodes or edges 3
2.4. Definition of power-law-$\beta$ 5
3. The (thinning and conditioning) trandformations, and their fixed points 5
4. Uniqueness 7
5. Convergence to nontrivial fixed points 9
6. Convergence to trivial fixed points 11
7. Infinite divisibility 12
References 14

1. Abstract notions of scale-free

Before turning to the statements and proofs of our results, we discuss analogous issues in the simpler continuous context. Suppose $X \geq 1$ is a continuous random variable, and ask what it means to be scale-free. A reasonable choice is to say that a scaled version $cX$ (with $c \in (0,1)$) “looks the same” as $X$. However, it cannot have

Date: June 12, 2013.
2000 Mathematics Subject Classification. Primary 60B10; Secondary 05C80.
Key words and phrases. thinning, power-law, scale-free, degree distribution, Pareto distribution.
the same distribution as $X$, so one might ask that it have the same distribution as $X$ after conditioning on $cX \geq 1$. This means that

$$P(X \geq x) = P(cX \geq x \mid cX \geq 1).$$

It is not hard to check that the only such distributions are the Pareto distributions:

$$P(X \geq x) = x^{-\alpha}, \quad x \geq 1.$$

See [12, 14] for similar observations.

In the discrete context, say $D$ taking values in \{0, 1, 2, \ldots\}, multiplying by $c \in (0, 1)$ would change the support. The natural analog of multiplying by a positive $c < 1$, while maintaining integer support, is $p$-thinning. The notion of $p$-thinning will be defined in Section 2.2. Thus we ask: what distributions on the nonnegative integers have the property that $D$ has the same distribution as ((the $p$-thinning of $D$), conditional on being strictly positive)? The preceding involved characterization of fixed points. For convergence to these fixed points, the continuous analogue of the discrete problem that we treat in this paper is the following. Take $X \geq 1$ and let $H(x) = P(X \geq x)$ be its tail probabilities. Under what conditions on $H$ does

$$\lim_{c \to 0} P(cX \geq x \mid cX \geq 1) = x^{-\alpha}, \quad x \geq 1?$$

It is not hard to check that a necessary and sufficient condition is that

$$H(x) = L(x)x^{-\alpha}$$

where $L$ is a slowly varying function.

Recall, for example from [6], that a positive function $L$ on $(0, \infty)$ is said to be slowly varying (at infinity) if

$$\lim_{t \to \infty} \frac{L(tx)}{L(t)} = 1$$

for every $x > 0$. Common examples are $L(x) = c(\log(1 + x))^\sigma$, for a constant $c \neq 0$, and any $\sigma \in (-\infty, \infty)$. Slowly varying functions are useful in discussing distributions that have power law properties without being strict power laws.

There are other results in the literature about fixed points of transformations that are the composition of two operations that change a distribution in opposite directions. Examples are [2, 5].

2. Introduction in the context of graphs

2.1. “Scale-free” and “power-law”. The notion that a degree distribution follows a power-law is concrete. It is the statement that the probability of being exactly $k$, decays as a negative power of $k$ — see [11] for the precise version, and [7] – [9] for the common extensions. The notion of a graph being scale-free is harder to pin down. Some authors, e.g., [4, 10], simply define scale-free graphs to be those where the degree distribution follows a power-law. We prefer a much broader interpretation: $G$ is scale-free if randomly chosen subgraphs $H$ are qualitatively similar to $G$.

The ambiguous phrase qualitatively similar is meant to apply to the degree distribution as well as other graph features, but in this paper we study only the degree distribution.
2.2. Thinning by tossing a $p$-coin. To $p$-thin a population means to toss a coin, independently for each individual in the population; the coin has $\mathbb{P}(\text{heads}) = p$, and only those individuals getting heads are retained. This standard but not well-known terminology comes from the study of random point processes and percolation, as in [7][9][10]. Thinning can be applied to sets and other structures, but the simplest context for thinning, corresponding to thinning a set and keeping track of its cardinality, is the following.

The “$p$-thinning” of a nonnegative integer valued random variable $D$ is the number $S_D$ of heads, in $D$ tosses of a $p$-coin. The notation $S_D$ comes from random walks. Take a sequence of independent Bernoulli random variables $X_1, X_2, \ldots$, with $p = \mathbb{P}(X_i = 1), 1 - p = \mathbb{P}(X_i = 0)$, and for $n = 0, 1, 2, \ldots$, let $S_n = X_1 + \cdots + X_n$, so that $\mathbb{P}(S_n = k) = \binom{n}{k} p^k(1-p)^{n-k}$. Finally, require that $X_1, X_2, \ldots$ be independent of $D$. With this setup, $S_D$ is the $p$-thinning of $D$, and we use the notation $S_{D,p}$ when we want to emphasize the role of $p$:

\begin{equation}
S_D \equiv S_{D,p} = X_1 + \cdots + X_D = \text{the number of heads in $D$ tosses.}
\end{equation}

In terms of the probability generating function, with $G_D(s) := \mathbb{E} s^D$, we have

\begin{equation}
G_{S_D}(s) = G_D(1 - p + ps) = G_D(1 - p(1 - s)).
\end{equation}

This makes it tempting to use “doubly complemented” generating functions, with $N_D(s) := 1 - G_D(1 - s)$, as the result of $p$-thinning is captured by the relation $N_{S_D}(s) = N_D(ps)$. See [17] for a use of the doubly complemented generating function.

2.3. Thinning a graph, by nodes or edges. Start with a simple graph $G = (V, E)$; neither multiple edges nor loops are allowed. This $G$ may be thought of as deterministic or random graph; if the graph is random, then we require all our $p$-coins to be independent of $G$.

To thin “by nodes” we take a random subset $S \subseteq V$, chosen by tossing a $p$-coin for each node. The thinned graph $H$ is the induced graph on $S$, i.e., $S$ is the set of nodes of $H$, and the edges of $H$ are all those edges of $G$ which have both endpoints in $S$.

To thin “by edges” we toss a $p$-coin for each edge. The thinned graph $H$ has the original $V$ as its set of nodes, and the edges of $H$ are those $e \in E$ whose coin comes up heads.

It is easiest to see the difference between these two ways of thinning by considering the example $G = K_n$, the complete graph with $n$ nodes and $m = \binom{n}{2}$ edges. If we thin by nodes we get a complete graph $K_k$ for some random $k$, with $0 \leq k \leq n$. If we thin by edges, we get a graph on $n$ nodes — and all $2^m$ such graphs are possible; in fact the random graph $H$ is exactly distributed according to the beloved Erdős-Rényi model $G(n, p)$.

**Theorem 1.** Let $D$ be the degree of a randomly chosen node of $G = (V, E)$, a graph with $n$ nodes and $m$ edges. For either method of thinning, by nodes or by edges, the degree $D^*$ in $H$ of a randomly chosen node of $H$ is exactly the $p$-thinning $S_D$ of $D$. [The choice of node is to be uniform among the available nodes, and otherwise independent of $G$ and $H$; in the case of node-thinning, in order to have $D^*$ well-defined, we modify the procedure by conditioning on the event, of probability $1 - (1-p)^n$, that $H$ has at least one node.]
Proof: Write \( D_x(G) \) for the degree of node \( x \) in \( G \), and \( D_x(H) \) for its degree in \( H \). Note that \( D_x(G) \) may be deterministic or random, according to whether \( G \) is deterministic or random. Write \( N_x \) for the set of neighbors of \( x \) in \( G \), so that \( D_x(G) = |N_x| \). Write \( X \) for the random node selected to represent \( H \), so that \( D^* = D_X(H) \).

Consider first the case of thinning by edge. Since \( G \) and \( H \) have the same set of nodes, \( X \) also serves as a randomly chosen node for \( G \), so that \( D = D_X(G) \) and \( X \) are independent of the \( p \)-coins used for thinning. Now \( H \) is formed by tossing a \( p \)-coin for each edge of \( G \); in particular, for every \( x \), we have tossed a \( p \)-coin for each of the \( D_x(G) \) edges from \( x \) to a node in \( N_x \), and \( D_x(H) \) is the number of heads among these. Using the independence of \( G, X \), and the coins,

\[
P(D^* = j) = \sum_{x \in V} P(X = x)P(D_x(H) = j)
\]

\[
= \sum_{x \in V, \ell \geq 0} P(X = x)P(D_x(G) = \ell)P(D_x(H) = j|D_x(G) = \ell)
\]

\[
= \sum_{x \in V, \ell \geq 0} \frac{1}{n} P(D_x(G) = \ell) {\ell \choose j} p^j (1 - p)^{\ell - j}
\]

\[
= \sum_{\ell \geq 0} P(D_X(G) = \ell) {\ell \choose j} p^j (1 - p)^{\ell - j}.
\]

(4)

Consider next the case of thinning by node. Write \( S \) for the random subset of \( V \) formed by \( p \)-thinning, so that \( D_x(H) = |S \cap N_x| \). Our goal is to show that for each \( x \in V \), \( P(X = x) = 1/n \); having this, the entire argument in the four lines of the display containing (4) applies, and we are done with the case of thinning by node!

Fix a node \( x \in V \) and a nonempty set \( T \subset V \), and consider the event \( A_{x,T} \) that \( X = x \) and \( S = T \). Write \( k = |T| \). Since we are \( p \)-thinning the set \( V \) and then conditioning on not getting the empty set,

\[
P(S = T) = p^k (1 - p)^{n-k} (1 - (1-p)^n).
\]

Since we are choosing \( X \) uniformly from \( S \), we have \( P(A_{x,T}) = \frac{k}{n} P(S = T) \) if \( x \in T \), and \( P(A_{x,T}) = 0 \) if \( x \notin T \). There are \( \binom{n-1}{k-1} \) ways to pick a set \( T \) of size \( k \) containing \( x \). Finally, we have

\[
P(X = x) = \sum_{T \ni x} P(A_{x,T})
\]

\[
= \sum_{k=1}^{n} \binom{n-1}{k-1} \frac{1}{k} p^k (1 - p)^{n-k} \frac{1}{(1 - (1-p)^n)} = \frac{1}{n},
\]

using \( \binom{n}{k}/n = \binom{n-1}{k-1}/k \) to get the final equality. \( \square \)

**Remark 1.** The difference between the argument for edge-thinning and the argument for node-thinning is subtle. For edge-thinning, we have a particular coupling between \( D = D_X(G) \), \( D^* = D_X(H) \), based on a single random choice \( X \) of node, with \( X \) independent of the coins, but there is no such construction when thinning by node. In both cases, \( P(X = x) = 1/n \) for each node \( x \), which is the key in the step (4). For node-thinning, \( X \) cannot be independent of the coins, because tails for the coin at node \( x \) implies \( X \neq x \).
2.4. Definition of power-law-$\beta$. We say that a positive integer valued random variable $D$ satisfies a power-law-$\beta$ distribution if it satisfies the following asymptotic condition on the point probabilities:

$P(D = k) \sim c_1 k^{-\beta},$ where $\beta > 1.$

The notation $a_k \sim b_k$ denotes the asymptotic relation $\lim_{k \to \infty} a_k/b_k = 1.$

One way to broaden the definition would be to say that a positive integer valued random variable $D$ satisfies a power-law-$\beta$ distribution if it satisfies the following asymptotic condition on the tail probabilities: as $k \to \infty$,

$P(D \geq k) \sim c_2 k^{-\alpha},$ where $\alpha := \beta - 1 > 0.$

Another way to broaden the definition is to allow a slowly varying function in place of the constant in (5), and say that a positive integer valued random variable $D$ satisfies a power-law-$\beta$ distribution if

$P(D = k) = L(k) k^{-\beta},$ where $\beta > 1,$

and $L$ is slowly varying, as in (1).

The broadest natural definition of power-law-$\beta$ combines the upper-tail feature of (6) with the slowly varying feature of (7); one might say that positive integer valued random variable $D$ satisfies a power-law-$\beta$ distribution if

$P(D \geq n) = L(n) n^{1-\beta},$ with $L$ slowly varying.

3. The (thinning and conditioning) transformations, and their fixed points

If $D$ is a nonnegative integer valued random variable and $0 < p < 1$, the $p$-thinning $S_D$ of $D$ is the random variable given by

$S_D = X_1 + \cdots + X_D,$

where $X_1, X_2, \ldots$ are i.i.d. random variables (also independent of $D$) with $P(X_i = 1) = p, P(X_i = 0) = 1 - p$. The distribution of $S_D$ is then given by

$P(S_D = n) = \sum_{l=n}^{\infty} P(D = l) \binom{l}{n} p^n (1-p)^{l-n} = \left( \frac{p}{1-p} \right)^n \frac{1}{n!} \sum_{l=n}^{\infty} (l)_n (1-p)^l P(D = l).$

This uses the notation $(z)_k = z(z-1)\cdots(z-k+1)$ for the falling product.

Fix an integer $m = 1, 2, \ldots$. For $p \in (0,1)$, the transformations $T \equiv T_p \equiv T_{p,m}$ for which we consider fixed points and convergence of iterates are given by

$P(TD = l) = P(S_D = l \mid S_D \geq m).$

In Section 4 we will prove that the fixed points of the transformation are precisely those described by (11)–(15) below, and in Section 4 and 5 we will prove results where these fixed points arise as limits of iterates of the transformation.

Remark 2. We are referring here to distributions that are fixed points for all $p,$ not just for some $p$. It would be interesting to know whether these are the only fixed points for a given $p.$
For $m = 1, 2, \ldots$, the distribution with $P(D = m) = 1$ is a trivial fixed point of $T_{p,m}$.

For $m = 1$, all nontrivial fixed points have the form: for some $\alpha \in (0,1)$,
\begin{equation}
G_D(s) := E s^D = 1 - (1 - s)^\alpha =: \sum_{k \geq 0} c_k(\alpha) s^k.
\end{equation}

The right hand side of (11) defines $c_k(\alpha)$ to be the coefficient of $s^k$ in $1 - (1 - s)^\alpha$, so for any $\alpha \in \mathbb{R}$,
\begin{equation}
c_0(\alpha) = 0, c_1(\alpha) = \alpha, c_2(\alpha) = \frac{\alpha(1 - \alpha)}{2}, c_3(\alpha) = \frac{\alpha(1 - \alpha)(2 - \alpha)}{6};
\end{equation}
for $k \geq 1,$
\begin{equation}
c_k(\alpha) = \frac{(-1)^{k-1}(\alpha)_k}{k!};
\end{equation}
and then specifically for $m = 1$, with the restriction $\alpha \in (0,1)$,
\begin{equation}
P(D = k) = c_k(\alpha), \quad k = 1, 2, \ldots .
\end{equation}

In general, for $m = 1, 2, \ldots$ and $\alpha \in (0,m)$ there is a nontrivial fixed point for $T_{p,m}$, which is power-law-$\beta$ for $\beta = 1 + \alpha$, with
\begin{equation}
G_D(s) := E s^D = \frac{1 - (1 - s)^\alpha - \sum_{1 \leq k < m} c_k(\alpha) s^k}{1 - \sum_{1 \leq k < m} c_k(\alpha)},
\end{equation}
and this gives all nontrivial fixed points of $T_{p,m}$.

Thus, the special case $m = 1$ of (13) was given by (11), under the restriction $0 < \alpha < 1$; the special case $m = 2$ of (11) is
\begin{equation}
G_D(s) := E s^D = \frac{1 - (1 - s)^\alpha - \alpha s}{1 - \alpha},
\end{equation}
for $0 < \alpha < 2$; and the special case $m = 3$ of (11) is
\begin{equation}
G_D(s) := E s^D = \frac{1 - (1 - s)^\alpha - \alpha s - \frac{\alpha(1 - \alpha)}{2} s^2}{1 - \alpha - \frac{\alpha(1 - \alpha)}{2}},
\end{equation}
for $0 < \alpha < 3$.

A unified description of the fixed points (for all $p$) of $T_{p,m}$, including both the trivial fixed point, obtained by taking $\alpha = m$, is: $1 + \alpha = \beta \in (1, m + 1]$, $P(D \in \{m, m + 1, m + 2, \ldots \}) = 1$, $P(D = m) > 0$, and
\begin{equation}
\frac{P(D = k + 1)}{P(D = k)} = \frac{k - \alpha}{k + 1}, \quad k \geq m.
\end{equation}
or equivalently, shifting the dummy variable $k$ by 1,
\begin{equation}
\frac{P(D = k)}{P(D = k - 1)} = \frac{k - \beta}{k}, \quad k > m.
\end{equation}

The Yule–Simon distribution for power-law-$\beta$ has point probabilities given by $P(D = k) = (\beta - 1) \Gamma(k) \Gamma(\beta)/\Gamma(k + \beta)$, and hence ratios
\begin{equation}
\frac{P(D = k)}{P(D = k - 1)} = \frac{k - 1}{k - 1 + \beta}.
\end{equation}
In comparison with (15), both formulas have denominator minus numerator = $\beta$, for every $k$, but for non-integer $\beta$, (15) has the integer in the denominator, while the Yule–Simon ratio (16) has the integer in the numerator.
Remark 3 (Iterates of the combined transformation). It may or may not be intuitively obvious that the four steps: $p$-thinning to get $S_D$, then conditioning on $S_D \geq m$ to get $T_p D$, then $q$-thinning — to get say $Y$, and conditioning again, on $Y \geq m$, to get $T_q T_p D$. The intuition is reasonable, since the event $Y \geq m$ is a subset of the event $S_D \geq m$. For the case $m = 1$, an easy way to see that $T_p$ followed by $T_q$ equals $T_{pq} D$ is to use the “doubly complemented” generating functions from the end of Section 2.2 for which the distribution of $T_p D \equiv T_{pq} D$ is determined by

\begin{equation}
N_{T_p D}(s) = N_D(ps) / N_D(p).
\end{equation}

This allows a proof via the calculation

\[ \frac{N_{T_p D}(qs)}{N_{T_p D}(q)} = \frac{N_D(ps) / N_D(p)}{N_D(pq) / N_D(p)} = N_D(pqs) / N_D(pq) = N_{T_{pq} D}(s). \]

That $T_q \circ T_p = T_{pq}$ for all $m$ is true, and can be proved using a well-known coupling $p$-coins, $q$-coins, and $pq$-coins; we omit the details of this coupling proof.

From $T_q \circ T_l = T_{st}$ it follows that the $k$-fold iterate $(T_q)^k$ of $T_q$ is $T_p$ with $p = q^k$. Theorem 4 allows $p \to 0$ with only the restriction $p > 0$, and the special case where $p$ goes to zero along a geometric sequence $q^k$ yields convergence for iterates of the transformation $T_q$, for one fixed $q$.

4. Uniqueness

The goal is to show that, for $m = 1, 2, \ldots$, any distribution $D$ on the nonnegative integers which is unchanged by $p$-thinning followed by conditioning on being at least $m$, for all $p \in (0, 1)$, is either the constant $D = m$ or else, as specified by (14), the law with $1 < \beta < m + 1$ and ratios $\mathbb{P}(D = k) / \mathbb{P}(D = k - 1) = (k - \beta) / k$ for $k \geq m + 1$.

Lemma 2. Suppose $A$ and $B$ are two nonnegative integer valued random variables with probability generating functions $G_A$, $G_B$. Let $m$ be a positive integer. Assume $\mathbb{P}(A \geq m) > 0$ and $\mathbb{P}(B \geq m) > 0$. Consider the statements

(a): $\mathbb{P}(A = k) = \mathbb{P}(B = k)$ for all $k \geq m$.

(b): $(A | A \geq m)$ and $(B | B \geq m)$ have the same distribution.

(c): $G_A^{(m)}(s) = G_B^{(m)}(s)$ for all $s \in [0, 1]$.

(d): $G_A^{(m)}(s) = c G_B^{(m)}(s)$ for all $s \in [0, 1)$, for some constant $c > 0$.

Then (a) if and only if (c), and (b) if and only if (d).

Proof. Let $a_k := \mathbb{P}(A = k)$ and $b_k := \mathbb{P}(B = k)$ so that $G_A(s) = \sum_{k \geq 0} a_k s^k$ and likewise for $G_B$. These are power series with radius of convergence $\geq 1$, hence differentiable term-by-term, with $G_A^{(m)}(s) = \sum_{k \geq 0} k(m) a_k s^{k-m}$ for $|s| < 1$, and likewise for $G_B$. This immediately shows that (a) implies (c); to see that (c) implies (a), given $k \geq m$, differentiate $k - m$ times and evaluate at $s = 0$.

The equivalence of (b) and (d) follows, with $c = \mathbb{P}(B \geq m) / \mathbb{P}(A \geq m)$. \hfill \Box

We apply this with $A = D$ and $B = S_D$. We are looking for a fixed point of $D \to T$, where $T = T_{D, p, m} := (S_D | S_D \geq m)$ and $S_D$ is the $p$-thinning of $D$. Since $1 = \mathbb{P}(T \geq m)$, we can have $D$ and $T$ equal in distribution only if $1 = \mathbb{P}(D \geq m)$. Thus we assume that $1 = \mathbb{P}(D \geq m)$, so that $D = (D | D \geq m)$, and now we have a
fixed point of \( D \rightarrow T \) if and only if \( (D|D \geq m) = (S_D|S_D \geq m) \). Combine Lemma 2 with 3, so that the two generating functions of interest as \( G_A(s) = G(s) \) and \( G_B(s) = G(1 - p(1 - s)) \).

Write \( f \) for the \( m \)th derivative of \( G \), so that \( G_B^{(m)}(s) = (G(1 - p + ps))^{(m)} = p^m \times f(1 - p(1 - s)) \). Assuming that \( 1 = P(D \geq m) \), we have a fixed point of \( D \rightarrow T_{D,p,m} \) if and only if

\[
f(s) = c p^m \times f(1 - p(1 - s)), \quad \text{for all } s \in [0, 1).
\]

**Lemma 3.** Let \( f \) be a continuous function from \([0, 1)\) to \((0, \infty)\), with \( f(0) = 1 \), and let \( p \mapsto c(p) \) be any function on \((0, 1)\). If

\[
∀ p \in (0, 1), ∀ s \in [0, 1), \quad f(1 - p(1 - s)) = c(p)f(s),
\]

then for some constant \( d \) we have \( f(s) = (1 - s)^{-d} \).

**Proof.** First let \( s = 1 - t \) so that (18) becomes

\[
∀ p \in (0, 1), ∀ t \in (0, 1), \quad f(1 - pt) = c(p)f(1 - t),
\]

and then consider \( g(t) := f(1 - t) \) so that the system to solve becomes

\[
∀ p \in (0, 1), ∀ t \in (0, 1), \quad g(pt) = c(p)g(t),
\]

with \( g(1) = 1 \). Plugging in \( t = 1 \) we see that \( c(p) = g(p) \), and (19) becomes \( g(pt) = g(p)g(t) \). To change from multiplication to addition, we take \( h(x) = \log(g(e^{-x})) \) for \( x \in [0, \infty) \). The system (19) implies

\[
∀ x, y \in (0, \infty), \quad h(x + y) = h(x) + h(y).
\]

This implies first that for all \( x \in (0, \infty) \), \( h(nx) = nh(x) \) for natural numbers \( n \), then that \( h(rx) = rh(x) \) for positive rational \( r \), and in particular, with \( x = 1, d := h(1) \), that \( h(r) = ar \). Finally, using the continuity of \( f \), we have \( h(x) = dx \) for \( x \in (0, \infty) \). Unraveling the substitutions yields \( f(s) = (1 - s)^{-d} \) for \( s \in [0, 1) \). □

**Theorem 4.** Let \( m \) be a positive integer, and let \( D \) be a nonnegative integer valued random variable, with \( P(D \geq m) > 0 \). The following are equivalent:

- The distribution of \( D \) is fixed by the transformation \( D \rightarrow T_{D,p,m} \) of \( p \)-thinning followed by conditioning on being at least \( m \), for all \( p \in (0, 1) \).
- Either \( D \equiv m \) is constant, or else \( D \) has power-law-\( \beta \), with \( \beta = \alpha + 1 \), \( 0 < \alpha < m \), \( P(D < m) = 0 \), and

\[
P(D = k + 1)/P(D = k) = (k - \alpha)/(k + 1) \quad \text{for } k \geq m.
\]

**Proof.** Start by assuming that \( D \) is a fixed point. We combine Lemmas 2 and 3 as in the remarks before 3, so that \( G(s) = sP(D) \), \( f \) is the \( m \)th derivative of \( G \), and the conclusion of Lemma 3 applied to \( f(s)/f(0) \) is that \( f(s) = c(1 - s)^{-d} \) with \( c > 0 \).

[We have \( c = f(0) > 0 \) because \( P(D \geq m) > 0 \) implies \( P(S_D = m) > 0 \), hence \( P(D = m|D \geq m) = P(S_D = m|S_D \geq m) > 0 \), hence \( c = mP(D = m) > 0 \).]

In case \( d = 0 \), we have \( f \) is constant and \( D \equiv m \). We cannot have \( d \) negative, since then the coefficient of \( s^1 \) in \( f \) is \( d \), while \( G \) has nonnegative coefficients. In case \( d > 0 \), writing \( [s^k]f(s) \) for the coefficient of \( s^k \) in \( f \), so that \( [s^k]G(s) = P(D = k) \), we have for \( k \geq m \)

\[
k_{(m)}P(D = k) = [s^{k-m}]f(s) = [s^{k-m}](c(1 - s)^{-d}) = c(-1)^{k-m}\frac{(-d)(k-m)}{(k-m)!}.
\]
Hence for \( k \geq m \)
\[
\mathbb{P}(D = k) = c (-1)^{k-m} \frac{(-d)(k-m)}{k!}
\]
and
\[
\frac{\mathbb{P}(D = k+1)}{\mathbb{P}(D = k)} = \frac{-(-d - (k-m))}{k+1} = \frac{k - \alpha}{k+1},
\]
with \( \alpha = m - d < m \). The requirement \( \sum \mathbb{P}(D = k) < \infty \) implies that \( \alpha > 0 \).

The implication in the opposite direction is easy, again by combining Lemmas 2 and 3.

5. Convergence to nontrivial fixed points

We will prove

**Theorem 5.** Suppose the distribution of \( D \) is power-law-\( \beta \), as specified by (7) or more generally (8), allowing a slowly varying function. Then for every integer \( k \geq \beta \)

\[
\lim_{p \to 0^+} \frac{\mathbb{P}(S_D = k)}{\mathbb{P}(S_D = k - 1)} = \frac{k - \beta}{k}.
\]

Before proving Theorem 5 we state part of a Tauberian theorem that can be found on page 447 of [6]. Many other Tauberian theorems can be found in [3].

**Theorem 6.** Tauberian Theorem. Let \( q_l \geq 0 \) and suppose

\[
Q(s) = \sum_{l=0}^{\infty} q_l s^l
\]

converges for \( 0 \leq s < 1 \). If \( L \) is slowly varying, \( \rho > 0 \), and

\[
q_l \sim l^{\rho-1} L(l),
\]

then

\[
Q(s) \sim \frac{\Gamma(\rho)}{(1-s)^\rho} L \left( \frac{1}{1-s} \right)
\]
as \( s \uparrow 1 \).

**Remark.** In his more general statement, Feller assumes that \( q_l \) is monotonic, but as he points out in the middle of page 446, this monotonicity is not needed for the special case quoted above.

**Proof.** Proof of Theorem 5 First, we assume (7). Apply the Tauberian theorem to

\[
q_l = \begin{cases} 
(l)_n P(D = l) \sim l^{n-\beta} L(l) & \text{for } l \geq n, \\
0 & \text{for } l < n,
\end{cases}
\]

\( \rho = n - \beta + 1 \), and \( s = 1 - p \). Using (4), the result is that if \( n > \beta - 1 \), then

\[
P(S_D = n) \sim \frac{\Gamma(n - \beta + 1)}{n!} L^{\rho-1}(p^{-1})
\]
as \( p \downarrow 0 \). If \( k > \beta \), this can be applied to both \( n = k \) and \( n = k - 1 \) to obtain (22).

If \( k = \beta \), this can be applied to \( n = k \), but not to \( n = k - 1 \), since then the corresponding \( \rho \) is zero. When \( k = \beta \) and \( n = k - 1 \), \( q_l \sim L(l)/l \). Theorem 1(b) on page 281 of [4] implies in this case that

\[
L^*(l) = q_1 + \cdots + q_l
\]
is slowly varying, and satisfies
\[ \lim_{l \to \infty} \frac{L(l)}{L^*(l)} = 0. \]
Using Theorem 5 on page 447 of [6], we see that
\[ P(S_D = k - 1) \sim \frac{p^{k-1}}{(k-1)!} L^*(p^{-1}). \]
Combining the above results implies that the limit in (22) is zero, as required.

Now instead of (7) we assume (8). Write
\[ H = \sum_{k \geq 1} \left( 1 - p \right)^k H^*(k) + \sum_{l=1}^{\infty} (l-1)(1-p)^{l-1} H(l) \]
\[ = \sum_{k \geq 1} \left( 1 - p \right)^k H^*(k) - \sum_{l=1}^{\infty} (l-1)(1-p)^{l-1} H(l) \]
\[ \sim k \Gamma(k - \beta + 1) p^{\beta-k-1} L(p^{-1}) - \Gamma(k - \beta + 2) p^{\beta-k-1} L(p^{-1}) \]
\[ = \Gamma(k - \beta + 1)(\beta - 1) p^{\beta-k-1} L(p^{-1}), \]
provided that \( k - \beta + 1 > 0 \). This gives (15) if \( k > \beta \). If \( k = \beta \), the above computation with \( k \) replaced by \( k - 1 \) gives
\[ \sum_{l=1}^{\infty} l(l-1) \cdots (l-k+2)(1-p)^{l-1} P(D = l) \sim (k-1)! H(k-1) + (k-1) L^*(p^{-1}), \]
so (22) holds in this case as well.

Convergence of the ratios of probabilities in (22) does not immediately imply tightness of the distributions of \( (S_D \mid S_D \geq m) \) as \( p \downarrow 0 \). This tightness is needed to conclude that the iterates of the transformation converge to the appropriate fixed point. We therefore now turn our attention to that issue.

**Theorem 7.** Take \( m \geq \beta - 1 \), and suppose the distribution of \( D \) is such that (24) holds for \( k \geq \beta \). Then the distributions of \( (S_D \mid S_D \geq m) \) are tight as \( p \downarrow 0 \). It follows that these distributions have a limit as \( p \downarrow 0 \), which is the fixed point described in (15) in case \( \beta < m + 1 \), or \( \mathbb{P}(D = m) = 1 \) in case \( \beta = m + 1 \).

**Proof.** Tightness of these conditional distributions means that
\[ \lim_{k \to \infty} \limsup_{p \to 0^+} \frac{P(S_D \geq k)}{P(S_D \geq m)} = 0. \]
Thus we need to deduce the asymptotics of ratios of tail probabilities from the asymptotics of ratios of point probabilities.

A key identity that allows for this transition is

\[ \frac{d}{dp} P(S_D \geq k) = k p^{-1} P(S_D = k). \]

Students of the theory of percolation will recognize this as a very simple form of Russo’s formula – see page 35 of [7], for example. The proof of (24) is also simple: Use (9) to write

\[ P(S_D \geq k) = \sum_{l=0}^{\infty} P(D = l) \left[ 1 - \sum_{n=0}^{k-1} \binom{l}{n} p^n (1-p)^{l-n} \right]. \]

Differentiating gives

\[ \frac{d}{dp} P(S_D \geq k) = p^{-1} \sum_{l=0}^{\infty} P(D = l) \sum_{n=0}^{k-1} \binom{l}{n} p^n (1-p)^{l-n-1} (lp - n). \]

To prove (24) one needs to check

\[ \sum_{n=0}^{k-1} \binom{l}{n} p^n (1-p)^{l-n-1} (lp - n) = k \binom{l}{k} p^k (1-p)^{l-k}. \]

The easiest way to check this is to note that the two sides of (26) agree for \( k = 0 \), and differences of the two sides of (26) for successive values of \( k \) also agree.

By L’Hospital’s Rule, whenever (22) holds, it follows from (24) that

\[ \lim_{p \to 0^+} \frac{P(S_D \geq k)}{P(S_D \geq k - 1)} = \frac{k - \beta}{k - 1} \]

Using (27) repeatedly gives

\[ \lim_{p \to 0^+} \frac{P(S_D \geq m + k)}{P(S_D \geq m)} = \prod_{j=1}^{k} \frac{m + j - \beta}{m + j - 1}. \]

Now (28) follows from this and the fact that

\[ \sum_{j} \frac{\beta - 1}{m + j - 1} = \infty. \]

\[ \square \]

6. CONVERGENCE TO TRIVIAL FIXED POINTS

Next we consider what happens in the less interesting regime \( m < \beta - 1 \).

**Theorem 8.** Suppose \( ED^{k-1} < \infty \). Then

\[ \lim_{p \to 0} \frac{P(S_D \geq k)}{P(S_D = k - 1)} = 0, \]

provided that the denominator above is strictly positive. As a consequence, if \( ED^m < \infty \) and \( P(D \geq m) > 0 \), then

\[ \lim_{p \to 0} P(S_D = m \mid S_D \geq m) = 1. \]

**Remark** If \( \Box \) holds with \( m = \beta - 1 \), then Theorems 5 and 7 provide the above conclusion even though \( ED^m \) may be infinite.
Proof. From (9) with $n = k - 1$ and the dominated convergence theorem, we see that

\[(29) \quad P(S_D = k - 1) \sim p^{k-1}E\left(\frac{D}{k - 1}\right) \quad \text{as} \quad p \downarrow 0.\]

We need to show that

\[
\lim_{p \to 0} \frac{P(S_D \geq k)}{p^{k-1}} = 0.
\]

This will follow from (25) and the dominated convergence theorem provided that

\[(30) \quad 1 - \sum_{n=0}^{k-1} \binom{l}{n} p^n (1-p)^{l-n} \leq C(lp)^{k-1}
\]

for some $C$ depending only on $k$, and

\[(31) \quad 1 - \sum_{n=0}^{k-1} \binom{l}{n} p^n (1-p)^{l-n} = o(p^{k-1})
\]

as $p \to 0$ for each $l$. Both (30) and (31) follow from (32)

\[(32) \quad 1 - \sum_{n=0}^{k-1} \binom{l}{n} p^n (1-p)^{l-n} \leq C(lp)^k
\]

for some (different) constant $C$, again depending only on $k$; (32) is a Chernoff bound; see [8, formula (12)]. That (31) follows from (32) is immediate. To deduce (30) from (32) write

\[(33) \quad 1 - \sum_{n=0}^{k-1} \binom{l}{n} p^n (1-p)^{l-n} = 1 - \sum_{n=0}^{k-2} \binom{l}{n} p^n (1-p)^{l-n} - \binom{l}{k-1} p^{k-1} (1-p)^{l-k+1}
\]

and apply (32) to the first part of (33) with $k$ replaced by $k - 1$.

The final statement follows from (25) with $k = m + 1$. \hfill \square

7. Infinite Divisibility

We will show that the distributions in (13) are infinitely divisible; this is relatively easy, thanks to a result from renewal theory. The history of the log-convexity – infinite divisibility connection is complicated; see [17, 18] and [15, Thm. 51.3; Notes on p. 426], in addition to references cited at the end of our proof of Proposition 9.

Proposition 9. Suppose the sequence $\{u(n), n \geq 0\}$ satisfies $u(0) = 1$,

\[(34) \quad u(n) > 0, u(n - 1)u(n + 1) \geq u^2(n) \quad \text{for} \quad n \geq 1 \quad \text{and} \quad \lim_{n \to \infty} \frac{u(n)}{u(n + 1)} > 0.
\]

Let

\[(35) \quad \log \left(\sum_{n=0}^{\infty} u(n) s^n\right) = \sum_{n=1}^{\infty} \lambda(n) s^n.
\]

Then $\lambda(n) \geq 0$ for $n \geq 1$. 

Proof. Let \( \{f(n), n \geq 1\} \) be the sequence associated to \( u(\cdot) \) by the renewal equation:

\[
\begin{align*}
  u(n) &= \sum_{k=1}^{n} f(k)u(n-k),
\end{align*}
\]

and consider the two generating functions

\[
U(s) = \sum_{n=0}^{\infty} u(n)s^n \quad \text{and} \quad F(s) = \sum_{n=1}^{\infty} f(n)s^n.
\]

Multiplying (36) by \( s^n \) and summing for \( n \geq 0 \) gives

\[
U(s) = 1 + U(s)F(s), \quad \text{or equivalently} \quad U(s) = \frac{1}{1 - F(s)}.
\]

Therefore, (35) can be written as

\[
\log[U(s)] = -\log(1 - F(s)) = \sum_{n=1}^{\infty} \frac{[F(s)]^n}{n}.
\]

Kaluza (11) proved that \( f(k) \geq 0 \) for all \( k \geq 1 \). (See [13, Theorem 1] for generalizations of this statement; see also [16].) Therefore the series in (35) has nonnegative coefficients. \( \square \)

**Corollary 10.** For \( m = 1, 2, \ldots \), and \( \alpha \in (0, m) \), the probability distribution for \( D \) specified by (13) and (14) is infinitely divisible.

**Proof.** Let \( X = D - m \), and define \( u(n) = \mathbb{P}(X = n)/\mathbb{P}(X = 0) \) for \( n \geq 0 \). This yields

\[
\sum_{n=0}^{\infty} u(n)s^n = \left(1 - s\right)^{\alpha} - \sum_{k=0}^{m-1} (-1)^k \frac{(\alpha)_k}{k!} s^k \left(1 - s\right)^{\alpha - m},
\]

and

\[
u(n) = (-1)^n \frac{m!(\alpha-m)(\alpha-m-1)\cdots(\alpha-n-m+1)}{(n+m)!}
\]

\[
= \frac{(m-\alpha)(m-\alpha+1)\cdots(m-\alpha+n-1)}{(m+1)(m+2)\cdots(m+n)}.
\]

Then \( u(0) = 1, u(n) > 0 \) for all \( n > 0 \)

\[
u(n) = \frac{m+n+1}{m+n-\alpha}
\]

which is decreasing in \( n \), so that (34) is satisfied. The probability generating function of \( X \) is

\[
G_X(s) := \mathbb{E}s^X = \mathbb{P}(X = 0) \sum_{n=0}^{\infty} u(n)s^n
\]

\[
= \mathbb{P}(X = 0) \exp \left( \sum_{n=1}^{\infty} \lambda(n)s^n \right),
\]

and Proposition 9 shows that \( \lambda(n) \geq 0 \) for \( n = 1, 2, \ldots \). Hence \( X \) is equal in distribution to \( \sum_{n \geq 1} nZ_n \), where \( Z_1, Z_2, \ldots \) are independent, and \( Z_n \) is Poisson distributed with parameter \( \lambda(n) \). \( \square \
References

[1] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. Rev. Modern Phys., 74(1):47–97, 2002.
[2] David J. Aldous and Antar Bandyopadhyay. A survey of max-type recursive distributional equations. Ann. Appl. Probab., 15(2):1047–1110, 2005.
[3] N. H. Bingham, C. M. Goldie, and J. L. Teugels. Regular variation, volume 27 of Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, 1989.
[4] Béla Bollobás and Oliver M. Riordan. Mathematical results on scale-free random graphs. In Handbook of graphs and networks, pages 1–34. Wiley-VCH, Weinheim, 2003.
[5] Richard Durrett and Thomas M. Liggett. Fixed points of the smoothing transformation. Z. Wahrsch. Verw. Gebiete, 64(3):275–301, 1983.
[6] William Feller. An introduction to probability theory and its applications. Vol. II. Second edition. John Wiley & Sons Inc., New York, 1971.
[7] Geoffrey Grimmett. Percolation. Springer-Verlag, New York, 1989.
[8] Torben Hagerup and Christine Rüb. A guided tour of Chernoff bounds. Inform. Process. Lett., 33(6):305–308, 1990.
[9] Olav Kallenberg. Random measures. Akademie-Verlag, Berlin, third edition, 1983.
[10] Olav Kallenberg. Foundations of modern probability. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.
[11] Th. Kaluza. Über die Koeffizienten reziproker Potenzreihen. Math. Z., 28(1):161–170, 1928.
[12] N. Krishnaji. Characterization of the Pareto distributions through a model of underreported incomes. Econometrica, 38(2):251–255, 1970.
[13] Thomas M. Liggett. Total positivity and renewal theory. In Probability, statistics, and mathematics, pages 141–162. Academic Press, Boston, MA, 1989.
[14] G. P. Patil and M. V. Ratnaparkhi. On additive and multiplicative damage models and the characterizations of linear and logarithmic exponential families. Canad. J. Statist., 7(1):61–64, 1979.
[15] Ken-iti Sato. Lévy processes and infinitely divisible distributions, volume 68 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 1999. Translated from the 1990 Japanese original, Revised by the author.
[16] D. N. Shanbhag. On renewal sequences. Bull. London Math. Soc., 9(1):79–80, 1977.
[17] F. W. Steutel. Preservation of infinite divisibility under mixing and related topics., volume 33 of Mathematical Centre Tracts. Mathematisch Centrum, Amsterdam, 1970.
[18] W. D. Warde and S. K. Katti. Infinite divisibility of discrete distributions. II. Ann. Math. Statist., 42:1088–1090, 1971.

(Richard Arratia) Department of Mathematics, University of Southern California, Los Angeles CA 90089.
E-mail address: rarratia@usc.edu
URL: http://dornsife.usc.edu/cf/faculty-and-staff/faculty.cfm?pid=1003062

(Thomas M. Liggett) Department of Mathematics, University of California Los Angeles CA 90095-1555.
E-mail address: tml@math.ucla.edu
URL: http://www.math.ucla.edu/~tml/

(Malcolm J. Williamson) Center for Communications Research, 4320 Westerra Ct., San Diego, CA 92121-1969.
E-mail address: malcolm@ccrwest.org