Probability Generating Functions

Generating functions are widely used in mathematics and play an important role in probability theory. Probability generating functions provide a major analytical tool when working with stochastic processes on discrete state spaces, such as branching processes.

1 Definition and Basic Properties

Definition 1. Let X be a nonnegative integer valued random variable such that $P(X = n) = p_n$, $n = 0, 1, 2, \ldots$ is its probability mass function. Then the **probability generating function (pgf)** of X is defined by

$$g_X(s) = E[s^X] = \sum_{n=0}^{\infty} s^n P(X = n) = \sum_{n=0}^{\infty} p_n s^n.$$

Note that

$$g_X(1) = \sum_{n=0}^{\infty} p_n 1^n = \sum_{n=0}^{\infty} p_n = 1.$$

Since $0 \leq p_n \leq 1$, $\forall n \geq 1$, and $\sum_{n=0}^{\infty} p_n = 1$, we have that for $|s| \leq 1$

$$|p_n s^n| \leq p_n, \ \forall n \geq 1 \ \text{ and } \ \sum_{n=0}^{\infty} p_n < \infty,$$

so by the Weierstrass M-test, $g_X(s) = \sum_{n=0}^{\infty} p_n s^n$ converges uniformly for $|s| \leq 1$. Furthermore, since for each $m \geq 1$

$$f_m(s) = \sum_{n=0}^{m} p_n s^n$$

is continuous on $[-1, 1]$ and $f_m \to g_X$ uniformly on $[-1, 1]$, we also have that $g_X(s)$ is continuous on $[-1, 1]$.

The name probability generating function comes from the following property.

Proposition 1. *Let X be a nonnegative integer valued random variable with pmf $P(X = n) = p_n$, $n = 0, 1, 2, \ldots$. If X has pgf*

$$g_X(s) = \sum_{n=0}^{\infty} p_n s^n,$$

then for each $k = 0, 1, 2, \ldots$,

$$p_k = g_X^{(k)}(0) = \frac{1}{k!} \frac{d^k}{ds^k} g_X(s)\bigg|_{s=0}.$$

Proof. Note that $g_X(s)$ defines a power series with radius of convergence $R \geq 1$. Therefore, g_X is infinitely differentiable for $|s| < R$, and for each $k = 1, 2, \ldots$

$$\frac{d^k}{ds^k} g_X(s) = \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1) p_n s^{n-k}.$$

Then

$$g_X^{(0)}(0) = g_X(0) = p_0,$$

and for $k = 1, 2, \ldots,$

$$\frac{d^k}{ds^k} g_X(s)\bigg|_{s=0} = k(k-1)\cdots(k-k+1)p_k = k!p_k.$$

□

An immediate consequence of this result is that the pgf of a nonnegative integer valued random variable uniquely determines its distribution.

Corollary 1. *Let X and Y be two nonnegative integer values random variables with pgfs*

$$g_X(s) = \sum_{n=0}^{\infty} p_n s^n \quad \textit{and} \quad g_Y(s) = \sum_{n=0}^{\infty} q_n s^n.$$

Then $X \stackrel{d}{=} Y$ if and only if $g_X(s) = g_Y(s)$ for all s in some interval containing 0.

Proof. ($\Longrightarrow$) Suppose that $X \stackrel{d}{=} Y$. Then $p_n = P(X = n) = P(Y = n) = q_n$ for all $n = 0, 1, 2 \ldots,$ so

$$g_X(s) = \sum_{n=0}^{\infty} p_n s^n = \sum_{n=0}^{\infty} q_n s^n = g_Y(s)$$

for all s where the series converge. In particular, $g_X(s) = g_Y(s)$ for all $s \in [-1, 1]$.

Suppose that g_X and g_Y are pgfs such that $g_X(s) = g_Y(s)$ for all s in some interval containing 0. Then for each $k = 0, 1, 2, \ldots$

$$k!p_k = \frac{d^k}{ds^k} g_X(s)\bigg|_{s=0} = \frac{d^k}{ds^k} g_Y(s)\bigg|_{s=0} = k!q_k.$$

Thus, $p_n = P(X = n) = P(Y = n) = q_n$ for all $n = 0, 1, 2 \ldots,$ so $X \stackrel{d}{=} Y$. □

It turns out that similar to moment generating functions, we can also use a pgf to easily calculate certain moments of X. To prove this, we will need to recall a result concerning convergence of derivatives.

Proposition 2. *Let $\{f_n : [a, b] \mapsto \mathbb{R}, n \geq 1\}$ be differentiable on $[a, b]$ and suppose $\{f_n(x_0)\}_n$ converges for some $x_0 \in [a, b]$. If $(f_n')_n$ converges uniformly on $[a, b]$, then $f_n \to f$ uniformly on $[a, b]$ for some function f and*

$$f'(x) = \lim_{n\to\infty} f_n'(x), \ x \in [a, b].$$

Proposition 3. *Let X be a nonnegative integer valued random variable with pmf $P(X = n) = p_n$, $n = 0, 1, 2, \ldots$. If X has pgf*

$$g_X(s) = \sum_{n=0}^{\infty} p_n s^n,$$

then for each $k = 1, 2, \ldots,$

$$E[X(X-1)\cdots(X-k+1)] \textit{ exists and is finite if and only if } \frac{d^k}{ds^k} g_X(s)\bigg|_{s=1} \textit{ exists and is finite.}$$

In such a case,

$$E[X(X-1)\cdots(X-k+1)] = \frac{d^k}{ds^k} g_X(s)\bigg|_{s=1}.$$

Proof. Let R be the radius of convergence of g_X. Then $R \geq 1$. Since g_X is a power series, g_X is infinitely differentiable on $(-R, R)$ with

$$g_X^{(k)}(s) = \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1)p_n s^n, \ |s| < R, \ k = 1, 2, \ldots$$

If $R > 1$, then g_X has derivatives of all orders at $s = 1$ and

$$g_X^{(k)}(1) = \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1)p_n = E[X(X-1)\cdots(X-k+1)],$$

which exists and is finite, so the result is clear if $R > 1$

It remains to consider the case when $R = 1$. Let $k \geq 1$. If

$$E[X(X-1)\cdots(X-k+1)] = \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1)p_n < \infty,$$

then for each $1 \leq \ell < k$

$$\begin{aligned} E[X(X-1)\cdots(X-\ell+1)] &= \sum_{n=\ell}^{\infty} n(n-1)\cdots(n-\ell+1)p_n \\ &\leq \sum_{n=\ell}^{k-1} n(n-1)\cdots(n-\ell+1)p_n + \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1)p_n < \infty. \end{aligned}$$

For each $m \geq 1$, let $f_m^{(0)}(s) = \sum_{n=0}^{m} p_n s^n$, and let

$$f_m^{(j)}(s) = \begin{cases} \sum_{n=j}^{m} n(n-1)\cdots(n-j+1)p_n s^{n-j}, & j \leq m \\ 0, & j > m \end{cases}$$

for each $j = 1, \ldots, k$. Then for each $j = 1, \ldots, k$, $\sum_{n=j} n(n-1)\cdots(n-j+1)p_n < \infty$ implies that

$$f_m^{(j)}(s) \to \sum_{n=j}^{\infty} n(n-1)\cdots(n-j+1)p_n s^{n-j} := f^{(j)}(s) \ \text{ as } \ m \to \infty$$

uniformly on $[-1, 1]$ by the M-test. Since $\sum_{n=1}^{\infty} p_n = 1$, we also have that $f_m^{(0)}(s) \to g_X(s) := f^{(0)}(s)$ uniformly on $[-1, 1]$. Then for $|s| \leq 1$ and each $j = 1, \ldots, k$,

$$\frac{d}{ds} f^{(j-1)}(s) = \lim_{m \to \infty} f_m^{(j)}(s) = \sum_{n=j}^{\infty} n(n-1)\cdots(n-j+1)p_n s^{n-j} = f^{(j)}(s).$$

by Proposition 2. Since $g_X(s) = f^{(0)}(s)$, we have just shown that $g_X^{(j)} = \sum_{n=j}^{\infty} n(n-1)\cdots(n-j+1)p_n s^{n-j}$ for each $j = 1, 2, \ldots, k$. In particular,

$$g_X^{(k)}(1) = \sum_{n=j}^{\infty} n(n-1)\cdots(n-j+1)p_n = E[X(X-1)\cdots(X-k+1)] < \infty.$$

Now, suppose that $g_X^{(k)}(1)$ exists and is finite. Then $g_X^{(j)}(1)$ exists for each $j = 1, 2, \dots, k-1$ as well. We will now show that for each $j = 1, 2, \dots, k$,

$$g_X^{(j)}(1) = \sum_{n=j} n(n-1)\cdots(n-j+1)p_n.$$

Since $g_X^{(1)}(1)$ exists and is finite, we have that

$$\lim_{m\to\infty} \frac{g_X(1) - g_X(s_m)}{1 - s_m}$$

for every sequence $(s_m)_m$ in the domain of g_X such that $s_m \neq 1$ for all $m \geq 1$ and $s_m \to 1$. Let $(s_m)_m$ be an increasing sequence in $(0, 1)$ such that $s_m \to 1$. Then

$$\frac{g_X(1) - g_X(s_m)}{1 - s_m} = \frac{1 - \sum_{n=0}^{\infty} p_n s_m^n}{1 - s_m} = \sum_{n=1}^{\infty} p_n \frac{(1 - s_m^n)}{1 - s_m} = \sum_{n=1}^{\infty} p_n \frac{(1 - s_m)\sum_{\ell=0}^{n-1} s_m^{\ell}}{1 - s_m} = \sum_{n=1}^{\infty} p_n \sum_{\ell=0}^{n-1} s_m^{\ell}.$$

By assumption, $\lim_{m\to\infty} \sum_{n=1}^{\infty} p_n \sum_{\ell=0}^{n-1} s_m^{\ell}$ exists and is finite and is equal to $g_X^{(1)}(s)$. We will show that

$$\lim_{m\to\infty} \sum_{n=1}^{\infty} p_n \sum_{\ell=0}^{n-1} s_m^{\ell} = \sum_{n=1}^{\infty} np_n.$$

Note that $\sum_{n=1}^{\infty} np_n$ exists (possibly infinite). Let $A < \sum_{n=1}^{\infty} np_n$. Choose an M_0 such that $\sum_{n=1}^{M_0} np_n > A$. Since $\sum_{n=1}^{\infty} p_n \sum_{\ell=0}^{n-1} s_m^{\ell} \leq \sum_{n=1}^{\infty} np_n$ for all $m \geq 1$, we have

$$\begin{aligned}
\sum_{n=1}^{\infty} np_n &\geq \lim_{m\to\infty} \sum_{n=1}^{\infty} p_n \sum_{\ell=0}^{n-1} s_m^{\ell} \\
&= \lim_{m\to\infty} \lim_{M\to\infty} \sum_{n=1}^{M} p_n \sum_{\ell=0}^{n-1} s_m^{\ell} \\
&\geq \lim_{m\to\infty} \sum_{n=1}^{M_0} p_n \sum_{\ell=0}^{n-1} s_m^{\ell} \quad \text{(since the partial sum sequence is increasing)} \\
&= \sum_{n=1}^{M_0} p_n \sum_{\ell=0}^{n-1} (1)^{\ell} \\
&= \sum_{n=1}^{M_0} np_n \\
&> A
\end{aligned}$$

Since $A < \sum_{n=1}^{\infty} np_n$ was arbitrary, $g_X^{(1)}(1) = \lim_{m\to\infty} \sum_{n=1}^{\infty} p_n \sum_{\ell=0}^{n-1} s_m^{\ell} = \sum_{n=1}^{\infty} np_n$, which is finite by assumption. Repeating the argument for each $j = 2, \dots, k$, we obtain that

$$g_X^{(j)}(1) = \sum_{n=j}^{\infty} n(n-1)\cdots(n-j+1)p_n,$$

which exists and is finite. Thus,

$$E[X(X-1)\cdots(X-k+1)] = \sum_{n=j}^{\infty} n(n-1)\cdots(n-j+1)p_n < \infty.$$

□

Corollary 2. *Let X be a nonnegative integer valued random variable with pmf $P(X = n) = p_n$, $n = 0, 1, 2, \ldots$ and pgf*

$$g_X(s) = \sum_{n=0}^{\infty} p_n s^n.$$

Then $Var(X) = g_X^{(2)}(1) + g_X^{(1)}(1) - \left[g_X^{(1)}(1)\right]^2$.

Proof. Note that

$$\begin{aligned} g_X^{(2)}(1) + g_X^{(1)}(1) - \left[g_X^{(1)}(1)\right]^2 &= E[X(X-1)] + E[X] - (E[X])^2 \\ &= EX^2 - (E[X])^2 \\ &= \text{Var}(X). \end{aligned}$$

□

Example 1. Let $X \sim \text{Bernoulli}(p)$. Then

$$p_0 = P(X = 0) = 1 - p \text{ and } p_1 = P(X = 1) = p,$$

and $p_k = P(X = k) = 0$ for all $k \geq 2$. The pgf of X is then

$$g_X(s) = E[s^X] = \sum_{n=0}^{1} p_n s^n = (1 - p) + ps.$$

Note that

$$p_0 = g_X(0) = 1 - p \text{ and } p_1 = g_X^{(1)}(0) = p,$$

and

$$EX = g_X^{(1)}(1) = p.$$

Example 2. Let $X \sim \text{Poisson}(\lambda)$. Then

$$p_n = P(X = n) = \frac{e^{-\lambda}\lambda^n}{n!}, \ n = 0, 1, 2, \ldots$$

The pgf is given by

$$g_X(s) = \sum_{n=0}^{\infty} \frac{e^{-\lambda}\lambda^n}{n!} s^n = e^{-\lambda(1-s)} \sum_{n=0}^{\infty} \frac{e^{-\lambda s}(\lambda s)^n}{n!} = e^{-\lambda(1-s)}.$$

Note that

$$g_X^{(k)}(s) = \lambda^k e^{-\lambda(1-s)}, \ k = 1, 2, \ldots,$$

so

$$g_X^{(k)}(0) = \lambda^k e^{-\lambda} = k! p_k, \ k = 0, 1, 2, \ldots,$$

and

$$g_X^{(1)}(1) = \lambda = EX.$$

2 Convolutions

Let X and Y be two independent nonnegative integer valued random variables with distributions

$$p_n = P(X = n) \text{ and } q_n = P(Y = n),\ n = 0, 1, 2 \ldots.$$

Then the joint pmf of (X, Y) is

$$P(X = i, Y = j) = p_i q_k,\ i, j = 0, 1, 2, \ldots.$$

What is the distribution of $Z = X + Y$? First, note that the support Z is also the nonnegative integers. Then for each $n = 0, 1, 2, \ldots,$

$$\begin{aligned}
P(Z = n) &= P(X + Y = n) \\
&= \sum_{j=0}^{\infty} P(X + Y = n, Y = j) \quad \text{(Law of Total Probability)} \\
&= \sum_{j=0}^{\infty} P(X = n - Y, Y = j) \\
&= \sum_{j=0}^{\infty} P(X = n - j, Y = j) \\
&= \sum_{j=0}^{n} P(X = n - j, Y = j) \quad \text{(Since } P(X = n - j) = 0 \text{ for } j > n) \\
&= \sum_{j=0}^{n} P(X = n - j)P(Y = j) \quad \text{(By independence)} \\
&= \sum_{j=0}^{n} p_{n-j} q_j.
\end{aligned}$$

The distribution of Z is the convolution of the distribution of X with the distribution of Y. Let $r_n = P(Z = n)$. The pgf of Z is then

$$\begin{aligned}
g_Z(s) &= \sum_{n=0}^{\infty} r_n s^n \\
&= \sum_{n=0}^{\infty} \sum_{j=0}^{n} p_{n-j} q_j s^n \\
&= \sum_{j=0}^{\infty} q_j \sum_{n=j}^{\infty} p_{n-j} s^n \\
&= \sum_{j=0}^{\infty} q_j \sum_{n=0}^{\infty} p_n s^{n+j} \\
&= \sum_{j=0}^{\infty} q_j s^j \sum_{n=0}^{\infty} p_n s^n \\
&= g_X(s) \cdot g_Y(s)
\end{aligned}$$

Thus, for a sum of two independent random variables, the pgf of the sum is the product of the two pgfs. This, result generalizes to sums of more than two independent random variables as follows.

Proposition 4. *Let $X_1, X_2, \ldots, X_n$ be independent nonnegative integer valued random variables with pgfs $g_{X_i}(s)$, $i = 1, 2, \ldots, n$. Then the pgf of*

$$Z = X_1 + X_2 + \cdots + X_n$$

is given by

$$g_Z(s) = g_{X_1}(s) \cdots g_{X_n}(s) = \prod_{i=1}^{n} g_{X_i}(s).$$

Proof. For the general case, we will argue using properties of expectations of independent random variables. Note that

$$g_Z(s) = E[s^Z] = E[s^{X_1+\cdots+X_n}] = E\left[\prod_{i=1}^{n} s^{X_i}\right] = \prod_{i=1}^{n} E s^{X_i} = \prod_{i=1}^{n} g_{X_i}(s).$$

□

Corollary 3. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. nonnegative integer valued random variables with common pgf $g_X(s)$, $i = 1, 2, \ldots, n$. Then the pgf of*

$$Z = X_1 + X_2 + \cdots + X_n$$

is given by

$$g_Z(s) = [g_X(s)]^n.$$

Example 3. Let $Z \sim \text{Binomial}(n, p)$. The we can represent Z as the sum of n i.i.d. Bernoulli(p) random variables. Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} \text{Bernoulli}(p)$, then

$$Z \overset{d}{=} X_1 + \cdots + X_n,$$

and by the previous corollary, Z has pgf

$$g_Z(s) = [1 - p + ps]^n.$$

Example 4. Show that if $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$ with X and Y independent. Then $X + Y \sim \text{Poisson}(\lambda + \mu)$.

Solution. Recall that the pgfs of X and Y are given by

$$g_X(s) = e^{-\lambda(1-s)} \quad \text{and} \quad g_Y(s) = e^{-\mu(1-s)}.$$

Since X and Y are independent, the pgf of $Z = X + Y$ is

$$g_Z(s) = g_X(s) \cdot g_Y(s) = e^{-(\lambda+\mu)(1-s)}$$

which is the pgf of a Poisson($\lambda + \mu$) random variable. Since the pgf uniquely determines the distribution, we obtain the desired result. □

In applications, we can run in to random sums of random variables. This is called a compound process. Let N be a nonnegative integer valued random variable, and let $X_1, X_2, \ldots$ be a sequence of i.i.d. nonnegative integer valued random variables. A quantity of interest in many applications is a random sum of the form

$$Z_N = \sum_{j=1}^{N} X_j = X_1 + \cdots + X_N.$$

Finding the distribution of such a compound process is difficult, but the pgf is simple as the following proposition shows.

Proposition 5. *Let N be a nonnegative integer valued random variable with pgf $h(s)$, and let $X_1, X_2, \ldots$ be a sequence of i.i.d. nonnegative integer valued random variables with common pgf $g(s)$, where N is independent of $X_1, X_2, \ldots$. Then the pgf of the compound process*

$$Z_N = \sum_{j=1}^{N} X_j = X_1 + \cdots + X_N$$

is given by

$$G(s) = h(g(s)).$$

Proof. The pgf of Z_N is given by

$$\begin{aligned}
G(s) &= Es^{Z_N} \\
&= \sum_{k=0}^{\infty} P(Z_N = k)s^k \\
&= \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} P(Z_N = k, N = \ell)s^k \quad \text{(Law of total probability)} \\
&= \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} P(Z_N = k | N = \ell)P(N = \ell)s^k \quad \text{(Conditional probability)} \\
&= \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} P(X_1 + \cdots + X_\ell = k | N = \ell)P(N = \ell)s^k \\
&= \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} P(X_1 + \cdots + X_\ell = k)P(N = \ell)s^k \quad \text{(Because } N \text{ is independent of } \{X_i\}_{i=1}^{\infty}) \\
&= \sum_{\ell=0}^{\infty} P(N = \ell)\sum_{k=0}^{\infty} P(X_1 + \cdots + X_\ell = k)s^k \\
&= \sum_{\ell=0}^{\infty} P(N = \ell)[g(s)]^\ell \\
&= h(g(s)).
\end{aligned}$$

□

Example 5. Let N be a nonnegative integer valued random variable with pgf $h(s)$, and let $X_1, X_2, \ldots$ be a sequence of i.i.d. nonnegative integer valued random variables with common pgf $g(s)$, where N is independent of $X_1, X_2, \ldots$. Show that

$$EZ_N = E[N]E[X_1].$$

Proof. Since the pgf of Z_N is $G(s) = h(g(s))$, we have

$$EZ_N = G'(1) = h'(g(1))g'(1) = h'(1)g'(1) = E[N]E[X_1].$$

□

3 Application: Galton-Watson Branching Processes

Let $\{X_{i,j}, i, j \geq 1\}$ be a collection of i.i.d. nonnegative integer valued random variables with common distribution $\{p_n, n \geq 0\}$, where $p_n = P(X_{1,1} = n)$. A Galton-Watson branching process can be constructed as follows. Define $Z_0 = 1$ and for $n \geq 1$ recursively define

$$Z_n = \sum_{i=1}^{Z_{n-1}} X_{n,i}.$$

The branching process $\{Z_n, n \geq 0\}$ describes the following mechanism. An individual (the 0th generation) gives birth to $X_{1,1} = j$ offspring with probability p_j, $j \geq 0$, which together constitute the first generation. Then each of these j individuals give birth to some number of offspring $X_{2,1}, X_{2,2}, \ldots, X_{2,j}$, which constitute the second generation. The reproduction continues until everyone in a generation has no offspring and the population goes extinct. The random variables Z_n, then represent the size of generation n. Throughout, we will assume that the mean and variance of the offspring distribution are finite.

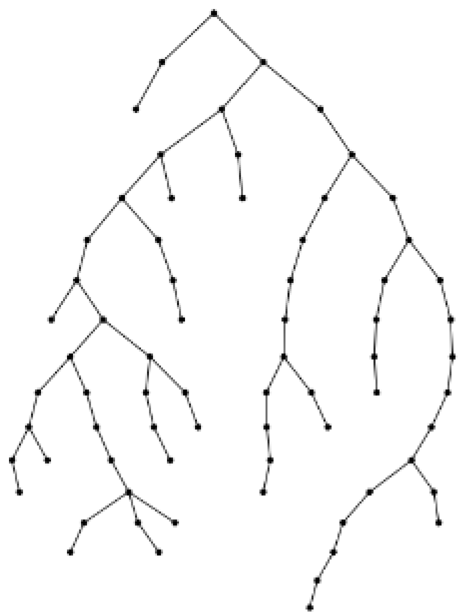


Figure 1: A tree realization of the Galton-Watson process with a Binomial(3, 0.4) offspring distribution from Nils Berglund. Martingales and Stochastic Calculus. 2013.

As an application of a real world process that can be modeled by this branching prcoess, consider an infectious disease process. There is initially $Z_0 = 1$ infected individual. The "offspring" of this individual are the people who are infected by this individual, so Z_1 represents the number of people infected by generation zero, who then each infect some number of additional individuals to form generation 2 of Z_2 infected individuals. If the discrete time points at which we measure the new infections is large enough, so that the previous generation is no longer infectious, then the Z_n individuals are the only remaining infectious individuals at time n and all new infections can be attributed to them (i.e. the new infections are their "offspring"). In this process, one primary question of interest is will the infection stop spreading and go extinct? In this branching process model, this happens as soon as

$Z_n = 0$ for some n. The probability of extinction is defined as

$$\pi_0 = P(\cup_{n=1}^{\infty}[Z_n = 0]).$$

Let $g(s) = \sum_{n=0}^{\infty} p_n s^n$ be the pgf of the offspring distribution, and let $g_n(s) = \sum_{k=0}^{\infty} P(Z_n = k)s^k$ be the pgf of Z_n for $n = 1, 2, 3, \ldots$. Since $P(Z_0 = 1) = 1$, note that the pgf of Z_0 is $g_0(s) = \sum_{k=0}^{\infty} P(Z_0 = k)s^k = s$. Since Z_n is a compound process, it follows that for each $n = 1, 2, \ldots$,

$$g_n(s) = g_{n-1}(g(s)).$$

Since $g_0(s) = s$, we have

$$g_1(s) = g_0(g(s)) = g(s) \implies g_2(s) = g_1(g(s)) = g(g(s)) \implies g_3(s) = g_2(g(s)) = g(g(g(s))).$$

Continuing in this fashion, we get that for each $n \geq 1$

$$g_n(s) = \underbrace{(g \circ g \circ \cdots \circ g)}_{\text{n times}}(s) \implies g_n = g(g_{n-1}(s)).$$

The expected value of Z_n, i.e. the expected size of generation n, is given by

$$E[Z_n] = g_n'(1) = g'(g_{n-1}(1))g_{n-1}'(1) = g'(1)g_{n-1}'(1) = E[X]E[Z_{n-1}],$$

where X is an i.i.d. copy of the offspring distribution. Repeating this argument, we obtain that for each $n \geq 1$.

$$E[Z_n] = (E[X])^n .$$

Let $\theta = EX$. Note that

$$\lim_{n\to\infty} E[Z_n] = \begin{cases} 0, & \theta < 1 \\ 1, & \theta = 1 \\ \infty, & \theta > 1 \end{cases},$$

and

$$\mathrm{Var}(Z_n) = \begin{cases} \sigma^2\theta^{n-1}\left(\dfrac{1-\theta^n}{1-\theta}\right), & \theta \neq 1 \\ n\sigma^2, & \theta = 1 \end{cases}.$$

The following result shows that in this model, extinction is guaranteed when $\theta \leq 1$, but even though $E[Z_n] \to \infty$ for $\theta > 1$, there is a non-zero chance of extinction.

Proposition 6. *Suppose that $p_0 > 0$ and $p_0 + p_1 < 1$. Then*

a) π_0 is the smallest nonnegative solution to the equation

$$s = \sum_{k=0}^{\infty} p_k s^k = g(s),$$

and

b) $\pi_0 = 1$ if and only if $\theta = EX = g'(1) \leq 1$. In the case, where $\theta > 1$, $0 < \pi_0 < 1$.

Proof. First, note that

$$\pi_0 = P(\cup_{n=1}^{\infty}[Z_n = 0]) = \sum_{k=0}^{\infty} P(\cup_{n=1}^{\infty}[Z_n = 0]|X_{1,1} = k)P(X_{1,1} = k) = \sum_{k=0}^{\infty} P(\cup_{n=1}^{\infty}[Z_n = 0]|X_{1,1} = k)p_k.$$

Now, given that $X_{1,1} = k$, i.e. there are k individuals in the first generation, the population with die out if and only if each of the k families starting from these k individuals in the first generation eventually die out. Since each of these k families are independent and evolve in the same way as the original branching process, each of these k families defines and i.i.d. copy of the original branching process. Thus, the probability that each individual family dies out is also π_0 and by independence, the probability that all k families die out eventually is π_0^k, i.e.

$$P(\cup_{n=1}^{\infty}[Z_n = 0]|X_{1,1} = k) = \pi_0^k.$$

Thus,

$$\pi_0 = \sum_{k=0}^{\infty} p_k \pi_0^k,$$

which shows that π_0 is a solution to the equation $s = g(s)$ on $[0, 1]$. To prove a), it remains to show that π_0 is the smallest such solution. Let $\pi \geq 0$ be a solution to the equation $s = g(s)$. We'll first show by induction that $\pi \geq P(Z_n = 0)$ for all $n \geq 1$. Now,

$$\pi = \sum_{k=0}^{\infty} p_k \pi^k \geq p_0 \pi^0 = p_0 = P(Z_1 = 0).$$

Next, assume that $\pi \geq P(Z_n = 0)$ for some $n \geq 1$. Then

$$\begin{aligned} P(Z_{n+1} = 0) &= \sum_{k=0}^{\infty} P(Z_{n+1} = 0|X_{1,1} = k)P(X_{1,1} = k) \\ &= \sum_{k=0}^{\infty} [P(Z_n = 0)]^k p_k \quad (k \text{ independent families go extinct after } n \text{ generations}) \\ &\leq \sum_{k=0}^{\infty} \pi^k p_k \quad \text{(by induction hypothesis)} \\ &= \pi. \end{aligned}$$

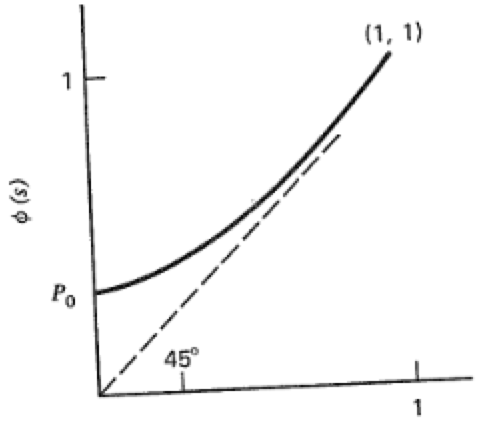
Hence, by induction, $P(Z_n = 0) \leq \pi$ for all $n \geq 1$. Note that $[Z_n = 0] \subset [Z_{n+1} = 0]$ for all $n \geq 1$, so $\cup_{n=1}^{\infty}[Z_n = 0] = \lim_{n\to\infty}[Z_n = 0]$ and

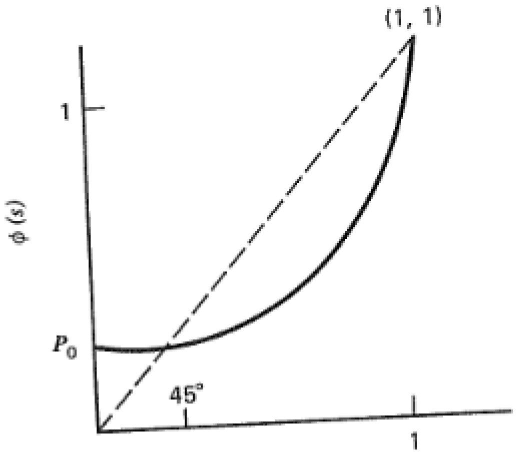
$$\pi_0 = P(\cup_{n=0}^{\infty}[Z_n = 0]) = \lim_{n\to\infty} P(Z_n = 0) \leq \pi.$$

To prove *b*), let $g(s) = \sum_{n=0}^{\infty} p_k s^k$ be the pgf of the offspring distribution. Since $p_0 + p_1 < 1$ and the variance of the offspring distribution is finite,

$$g^{(2)}(s) = \sum_{n=2}^{\infty} n(n-1)p_n s^{n-2} > 0, \ \forall s \in (0, 1).$$

Thus, $g'(s)$ is strictly increasing on $[0, 1]$. The following images, illustrate the two cases that can happen with the pgfs.





Note that $g'(0) = 0$ and $g'(1) = \theta$, that g is also strictly increasing on $[0,1]$ with $g(0) = p_0 > 0$ and $g(1) = 1$. Suppose that $\pi_0 = 1$, then $g(s) > s$ for all $s \in (0,1)$, since by a), π_0 is the smallest nonnegative solution of $g(s) = s$ and $g(0) = p_0 > 0$. (Otherwise, if $g(s_0) \leq s_0$ for some $s_0 \in (0,1)$, then either s_0 is another smaller solution or by the IVT, $\exists x \in (0, s_0)$ such that $g(x) = x$.) Thus,

$$0 < s < 1 \implies g(s) > s \implies 1 - g(s) < 1 - s \implies \frac{g(1) - g(s)}{1 - s} < 1 \implies g'(1) \lim_{s \uparrow 1} \frac{g(s) - g(1)}{s - 1} \leq 1.$$

Now, suppose that $g'(1) \leq 1$, and suppose by contradiction that $\pi_0 < 1$. Then $g(\pi_0) = \pi_0 < 1$. Thus, by the MVT, $\exists s_0 \in (\pi_0, 1)$ such that

$$g'(s_0) = \frac{g(1) - g(\pi_0)}{1 - \pi_0} = 1.$$

Since $g^{(2)}(s) > 0$ for all $s \in (0,1)$, g' is strictly increasing, so that $g'(1) > g'(s_0) = 1$, which contradicts our assumption that $g'(1) \leq 1$.

Note that in the case where $\theta > 1$, we know that $\pi_0 < 1$, but since g is increasing and $g(0) = p_0 > 0$, we have

$$0 < p_0 = g(0) \leq g(\pi_0) = \pi_0.$$

□