Probability generating functions are useful, as we saw last time, for treating sums of independent random variables. And they also work very well for treating random sums.

\[ Z = \sum_{j=1}^{N} X_j \] where the \( X_j \) are iid, and \( N \) is also random, but independent of the \( X_j \).

We can compute the statistics of \( Z \) from the statistics of \( X_j \) and \( N \) by using probability generating function.

\[
P_Z(s) = \left[ \frac{\prod_{j=1}^{N} s^{X_j}}{\prod_{j=1}^{N} \mathbb{E}[s^{X_j}]} \right] P(N = n)
\]

(shortened: \[ \mathbb{E}\left[ \prod_{j=1}^{N} s^{X_j} | N = n \right] P(N = n) \]

\[
= \sum_{n=1}^{\infty} \mathbb{E}\left[ \prod_{j=1}^{N} s^{X_j} | N = n \right] P(N = n)
\]

\[
= \sum_{n=1}^{\infty} \mathbb{E}\left[ \prod_{j=1}^{N} s^{X_j} | N = n \right] P(N = n)
\]

\[
= \sum_{n=1}^{\infty} \mathbb{E}\left[ \prod_{j=1}^{N} s^{X_j} | N = n \right] P(N = n)
\]

\[
= \sum_{n=1}^{\infty} \mathbb{E}\left[ \prod_{j=1}^{N} s^{X_j} | N = n \right] P(N = n)
\]
More details and examples on random sums can be found in Karlin and Taylor Sec. 1.1 and Resnick Ch. 0.

So now we will introduce branching processes and see how probability generating functions help in their analysis.

A branching process is a special case of a countable-state discrete-time Markov chain; we'll focus primarily on the simplest branching process model called the Galton-Watson model.

We are tracking a number of agents which, at each epoch, will give rise to a random number of offspring at the next epoch. If an agent survives to the next epoch, we simply account for it as being one of its own offspring. The number of offspring each agent has is independent of the number of offspring of any other agent. Each agent is indistinguishable from each other.
Probability transition matrix can in principle be written down (as for any discrete-time MC) but it's not convenient. However, probability generating functions are convenient to deploy here because of the special structure of the model involving combinations of many independent random variables.

Rather than formulate the probability transition matrix, we will describe the stochastic update rule, and apply probability generating functions to it.

\[
X_{n+1} = \sum_{k=1}^{N_n} \left( \begin{array}{c} N_n \\ k \end{array} \right) \mathbb{P}(\text{offspring})^{k} \cdot \mathbb{P}(\text{offspring})^{N_n-k}
\]

**Offspring probability distribution**: \( \mathbb{P}(\text{offspring}) = \mathbb{P}(X_k = j) \) \( j \geq 0, 1, 2 \ldots \)

**Initial distribution**: \( \mathbb{P}(X_0 = j) = \mathbb{P}(X_0 = j) \) \( j \geq 0, 1, 2 \ldots \)

The Galton-Watson process can be generalized in several directions:

- multiple types of agents
- age structure for the agents
- continuous time

See for an extended discussion of more general branching process models as well as a wealth of applications:

- **Branching Processes in Biology**, Kimmel and Axelrod

Applications:

- Genealogy, genetics across generations
- Biomolecular reproduction processes such as polymerase chain reaction as well as natural processes happening in the cell
- Population of cells in cancerous tumors
- Population growth
- Disease spread in early stages
- Photomultiplier tube cascades
- Nuclear fission
- Earthquake triggering
- Queueing models
Finite-Time Statistics of Branching Processes

We’ll approach this with iteration of the probability generation function rather than of the probability transition matrix.

Notice that the stochastic update rule is in the form of a random sum.

\[
P_{X_{n+1}}(s) = P \left( \sum_{n} P_{X_{n}}(s) \right)
\]

By iteration:

\[
P_{X_{n}}(s) = P \left( \sum_{n} P_{X_{n}}^{x_{n}}(s) \right)
\]

These statistics are easily evaluated computationally by a recursive algorithm.

In principle any information desired about \(X_n\) can now be obtained from its probability generating function. If it is too difficult to explicitly invert, you can still compute moments by differentiation:

\[
\left[ \mathbb{E} X_{n} \right]_{s=1} = s \frac{d}{ds} P_{X_{n}}(s) \bigg|_{s=1}
\]
Higher moments can be calculated similarly with more effort.

\[
\mathbb{E} X_n = \left( s \frac{d}{ds} \right)^m \mathbb{P} \left( X_n \right) \bigg|_{s=1}
\]
Long-Time Properties of Branching Processes

Let’s start with a topological analysis of branching processes.

One can see, because all states have possible transitions to the absorbing state at 0 (extinction), that the Markov chain must be transient. But there are two qualitatively different transient modes for the nonzero states. The Markov chain could either become absorbed at 0, or run off to infinity (i.e., remain in the transient class without ever getting absorbed).

- extinction vs proliferation

Key question: What is the probability of extinction, given $X_0$ (possibly random) and offspring distribution?

- The probability for the population to grow unboundedly is 1 minus this probability, because that’s the only other possible long-time behavior for the transient class.

$$a(k) = P(X_n = 0 \text{ for some } n \geq 0 | X_0 = k): \text{ the probability that a branching process (with some prescribed offspring distribution function) will become extinct if it starts with } k \text{ agents.}$$

First of all, we observe by independence of each agent that $a(k) = (a(1))^k$.

We will write $a = a(1)$.

We will use a first-step analysis to solve for $a$. This is like our absorption probability derivations for MCs involving probability transition matrices, but now we are formulating the argument in terms of probability generating functions because these are more convenient for branching processes.
for branching processes.

\[ q = \prod_{n=1}^{\infty} p \left( \bigcup_{n=1}^{\infty} \{ X_n = 0 \} \mid X_0 = 1 \right) \]

\[ \text{Law of total probability partitioned on first step} \]

\[ = \sum_{j=0}^{\infty} \prod_{n=1}^{\infty} p \left( \bigcup_{n=1}^{\infty} \{ X_n = 0 \} \mid X_1 = j, X_0 = 1 \right) p_j \]

\[ = \prod_{n=1}^{\infty} p \left( \bigcup_{n=1}^{\infty} \{ X_n = 0 \} \mid X_0 = 1 \right) p_0 + \sum_{j=1}^{\infty} \prod_{n=1}^{\infty} p \left( \bigcup_{n=1}^{\infty} \{ X_n = 0 \} \mid X_1 = j \right) p_j \]

\[ = p_0 + \sum_{j=1}^{\infty} a(j) p_j \]

\[ = p_0 + \sum_{j=1}^{\infty} a^j p_0 \]

\[ q = \sum_{j=0}^{\infty} p_j a^j \]
Does this have a unique solution? Not quite....

First the trivial cases:

\[ p_0 > 0, \quad p_1 = 1, \quad p_j = 0 \quad \text{for} \quad j \geq 2; \]

\[ p_j^{-1} P_j(a) = a \quad \text{(correct value:} \quad a = 0 \]

\[ 0 < p_0 < 1, \quad p_1 = 1 - p_0, \quad p_j > 0 \quad \text{for} \quad j \geq 2; \]

\[ p_j^{-1} P_j(a) = p_0 + (1 - p_0) a \]

\[ a > p_j^{-1} P_j(a). \]

\[ a > p_0 \frac{1 - (1 - p_0) a}{1 + p_0} \]

\[ a > \frac{1 + p_0}{1 + p_0} > 1 \quad a > 1 \]

The more general case:

\[ p_j > 0 \quad \text{for some} \quad j \geq 2. \]

\[ P_j^{-1}(s) \equiv \sum_{j=0}^{\infty} p_j s^j \]

\[ = \sum_{j=0}^{\infty} p_0 j^{(j-1)} s^{j-2} > 0 \]

Therefore the probability generating function is convex.
Therefore the probability generating function is convex.

We also know that

\[ P_X(1) = 1. \]

\[ P_X(0) = \rho_0. \]

There are essentially two possible cases:

Case \( \mu \leq 1 \): \( a = 1 \).

Case \( \mu > 1 \): \( a < 1 \) (nontrivial solution)

where \( \mu = EY = P'_X(1) \).