12. Probability generating functions

For a non-negative discrete random variable $X$, the probability generating function contains all possible information about $X$ and is remarkably useful for easily deriving key properties about $X$.

**Definition 12.1** (Probability generating function). Let $X \geq 0$ be a discrete random variable on $\{0, 1, 2, \ldots\}$ and let $p_k := P(X = k)$, $k = 0, 1, 2, \ldots$. The probability generating function of $X$ is

\[
G_X(z) := \mathbb{E}z^X := \sum_{k=0}^{\infty} z^k p_k.
\]

For example, for $X$ taking values $1$ with probability $1/2$, $2$ with probability $1/3$, and $3$ with probability $1/6$, we have

\[
G_X(z) = \frac{1}{2}z + \frac{1}{3}z^2 + \frac{1}{6}z^3.
\]

Paraphrasing Herb Wilf, we say that the probability generating function “hangs the distribution of $X$ on a clothesline.”

Notice that $G_X(1) = 1$ and $G_X(0) = p_0$ for any random variable $X$. We summarize further properties of probability generating functions in the following theorem.

**Theorem 12.2** (Properties of generating functions). For non-negative discrete random variables $X, Y \geq 0$, let $G_X$ and $G_Y$ be their probability generating functions, respectively. Then

(i) If $Y = a + bX$ for scalars $a, b \in \mathbb{R}$, then $G_Y(z) = z^a G_X(z^b)$.

(ii) If $X$ and $Y$ are independent, then $G_{X+Y}(z) = G_X(z)G_Y(z)$.

(iii) $P(X = n) = G^{(n)}(0)/n!$, for every $n \in \{0, 1, 2, \ldots\}$, where $G^{(n)}(z) := d^n G/dz^n$.

(iv) $G_X^{(n)}(1) = \mathbb{E}X^n$, for every $n \in \{0, 1, 2, \ldots\}$. In particular, $G_X'(1) = \mathbb{E}X$.

**Proof.** These properties all follow easily from the properties of expectations. For (i), $a, b \in \mathbb{R}$ and so $z^{a+bX} = z^a z^{bX}$; whence,

\[
\mathbb{E}z^a z^{bX} = z^a \mathbb{E}(z^{bX}) = z^a G_X(z^b).
\]

For (ii), if $X$ and $Y$ are independent, then so are $z^X$ and $z^Y$; therefore,

\[
G_{X+Y}(z) := \mathbb{E}z^{X+Y} = \mathbb{E}z^X z^Y = \mathbb{E}z^X \mathbb{E}z^Y = G_X(z)G_Y(z).
\]

For (iii), we see that if $n = 0, 1, 2, \ldots$, then the terms containing $p_0, \ldots, p_{n-1}$ will become 0 after taking the $n$th derivative, because they are the coefficients of a monomial in $z$ of degree less than $n$. The $n$th derivative of $G_X(z)$ is then

\[
G_X^{(n)}(z) := \sum_{k=0}^{\infty} p_k k^n z^{n-k}.
\]

Therefore, only the first term $p_n n! z^n$ does not depend on $z$ and so

\[
G^{(n)}(0) = n! p_n.
\]

From the above expression, it is apparent that $G_X^{(n)}(z)$ evaluated at $z = 1$ gives the $n$th factorial moment of $X$, establishing (iv).

The next theorem hints at why generating functions are so important.
Theorem 12.3 (Uniqueness theorem). Let $X, Y \geq 0$ be discrete random variables. Then $X$ and $Y$ have the same distribution if and only if $G_X = G_Y$.

12.1. Generating functions of some common distributions. We now catalog generating functions of some common distributions, from which properties that we already know (and sometimes spent some effort deriving) become obvious.

To begin with the most trivial distribution, for some $k = 0, 1, 2, \ldots$, suppose $X = k$ with probability one, then $G_X(z) = z^k$.

12.1.1. Binomial distribution. For $0 < p < 1$, let $X$ have the Bernoulli distribution with parameter $p$. Then $p_0 = 1 - p$ and $p_1 = p$ and

$$G_X(z) := \mathbb{E}z^X = (1 - p)z^0 + pz^1 = (1 - p) + pz.$$  

For a Binomial random variable $Y$ with parameter $(n, p)$, we can express $Y = X_1 + \cdots + X_n$, where $X_1, \ldots, X_n$ are i.i.d. Bernoulli($p$). Therefore, by Theorem 12.2(ii), we have

$$G_Y(z) = G_{X_1 + \cdots + X_n}(z) = \prod_{j=1}^n G_X(z) = (1 - p + pz)^n.$$  

 Apparently, taking the first derivative of $G_Y(z)$, we get

$$G_Y'(z) = n(1 - p + pz)^{n-1}p,$$

which has $G_Y'(0) = n(1 - p)^{n-1}p = \binom{n}{1}p(1 - p)^{n-1}$ and $G_Y'(1) = n(1 - p + p)^{n-1}p = np$.

12.1.2. Poisson distribution. Let $X$ have the Poisson distribution with parameter $\lambda > 0$. Then $p_k := \lambda^k e^{-\lambda}/k!$ and

$$G_X(z) = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}z^k}{k!} = e^{-\lambda + \lambda z} \sum_{k=0}^{\infty} (\lambda z)^k e^{-\lambda z}/k! = e^{-\lambda (1-z)}.$$  

This is a very nice generating function, because we can easily express the $n$th derivative of $G_X(z)$ by

$$G_X^{(n)}(z) = \lambda^n e^{-\lambda(1-z)};$$

and, therefore,

$$P(X = n) = G_X^{(n)}(0)/n! = \lambda^n e^{-\lambda(1-0)}/n! \quad \text{and} \quad \mathbb{E}X^n = G_X^{(n)}(1) = \lambda^n e^{-\lambda(1-1)} = \lambda^n.$$  

Theorem 12.4 (Superposition and Thinning properties of Poisson RVs). The Poisson distribution has the following special properties.

- Superposition property: Let $X_1, \ldots, X_n$ be independent Poisson random variables with $X_j \sim \text{Poisson}(\lambda_j)$, for $j = 1, \ldots, n$. Then $Y \sim \text{Poisson}(\lambda_1 + \cdots + \lambda_n)$, where $Y := X_1 + \cdots + X_n$.
- Thinning property: Let $X \sim \text{Poisson}(\lambda)$ and, given $X = n$, let $B \sim \text{Binomial}(n, p)$. Then the unconditional distribution of $B$ is Poisson with parameter $\lambda p$. 

Proof. The superposition property is a consequence of the exponential form of the generating function for the Poisson distribution. For \( X_1, \ldots, X_n \) with parameters \( \lambda_1, \ldots, \lambda_n \), respectively, we see that

\[
G_{X_1 + \cdots + X_n}(z) = \prod_{j=1}^{n} e^{-\lambda_j(1-z)} = e^{-\sum_{j=1}^{n} \lambda_j(1-z)}.
\]

Proof of the thinning property is left as a homework exercise.

\[\square\]

12.1.3. Geometric and Negative Binomial distributions. Let \( X \) have the Geometric distribution with success probability \( 0 < p < 1 \). Then

\[
p_k := (1 - p)^{k-1} p
\]

and

\[
G_Z(z) = \sum_{k=1}^{\infty} (1 - p)^{k-1} p z^k = \frac{pz}{1 - (1-p)z}.
\]

Furthermore, for a positive integer \( r > 0 \), let \( Y \) have the Negative Binomial distribution with parameter \((p, r)\). Then, for \( X_1, \ldots, X_r \) i.i.d. Geometric\((p)\) random variables,

\[
G_Y(z) = G_{X_1 + \cdots + X_r}(z) = \frac{pz}{1 - (1-p)z}.
\]

12.1.4. Sum of a random number of independent random variables.

**Proposition 12.5.** Let \( N \) be a positive integer random variable and, given \( N \), let \( X_1, X_2, \ldots \) be i.i.d. independent of \( N \). Then

(i) \( S_N := X_1 + \cdots + X_N \) has generating function

\[
G_{S_N}(z) = G_N(G_X(z)),
\]

(ii) \( \mathbb{E}S_N = \mathbb{E}N \mathbb{E}X_1 \), and

(iii) \( \text{Var}(S_N) = \mathbb{E} \text{Var}(X_1) + \text{Var}(N)[\mathbb{E}X_1]^2 \).

**Proof.** For (i),

\[
G_{S_N}(z) = \mathbb{E}z^{X_1 + \cdots + X_N}
= \sum_{n=0}^{\infty} \mathbb{E}z^{X_1 + \cdots + X_n} | N = n) \mathbb{P}(N = n)
= \sum_{n=0}^{\infty} \mathbb{E}z^{X_1 + \cdots + X_n} \mathbb{P}(N = n)
= \sum_{n=0}^{\infty} [G_X(z)]^n \mathbb{P}(N = n)
= G_N(G_X(z)).
\]

For (ii), we use the properties of generating functions to obtain \( \mathbb{E}S_N = G'_{S_N}(1) \). Thus,

\[
G'_{S_N}(z) = G'_{X_1}(z)G'_{N}(G_X(z)),
\]
by the chain rule, and
\[ G'_{S_N}(1) = G'_{X_1}(1)G'_{X_N}(G_{X_1}(1)) = \mathbb{E}X_1 G'_{X_N}(1) = \mathbb{E}X_1 \mathbb{E}X_N. \]

Item (iii) is a good exercise that you should definitely do. □

12.2. **Infinitely divisible distributions.** The class of *infinitely divisible distributions* comprises an important family of probability distributions. We say a probability distribution \( F \) is *infinitely divisible* if, for every \( n \in \mathbb{N} \), there exists a sequence \( X_1, \ldots, X_n \) of i.i.d. random variables so that \( S_n := X_1 + \cdots + X_n \) has distribution \( F \). Many of the distributions we have encountered so far are infinitely divisible. The easiest way to see this is through the probability generating function.

For example, the constant distribution is infinitely divisible, for if \( X = c \) with probability one, we can write
\[ X = \frac{c}{n} + \cdots + \frac{c}{n} \text{ (n times),} \]
for every \( n \geq 1 \).

Less trivially, by the superposition property, we know that the Poisson distribution with parameter \( \lambda \) can be obtained as the sum of \( n \) i.i.d. Poisson random variables with parameter \( \lambda/n \). Therefore, the Poisson distribution is infinitely divisible. On the other hand, the Binomial distribution is not infinitely divisible. To see this, we prove that the Bernoulli distribution is not infinitely divisible.

**Proposition 12.6.** The Bernoulli distribution is not infinitely divisible for any \( 0 < p < 1 \).

**Proof.** First of all, if \( p = 0 \), then \( X \) is deterministic and the above argument says that \( X \) is trivially infinitely divisible.

For \( 0 < p < 1 \), let \( X \) be a Bernoulli random variable with parameter \( p \). We have already observed that the generating function for \( X \) is \( G(z) = 1 - p + pz \). Now, if \( X \) is infinitely divisible, then there must be a random variable \( Y \) with pgf
\[ G_Y(z) := \sqrt{G(z)} = (1 - p + pz)^{1/2} \]
so that \( X = Y_1 + Y_2 \), where \( Y_1, Y_2 \) are i.i.d. with the same distribution as \( Y \). However, we see that
\[ G_Y''(z) = \frac{1}{2} \left( -\frac{1}{2} \right) p^2 (1 - p + pz)^{-3/2}; \]
and so
\[ G_Y''(0) = -\frac{1}{4} p^2 (1 - p)^{-3/2} < 0, \]
which is not a probability. □

When we let the parameter \( r \) in the Negative Binomial distribution take any non-negative real value, the Negative-Binomial(\( p, r \)) distribution is infinitely divisible. To be specific, let \( X \) have the Negative Binomial distribution with parameter \( (p, r) \). Then \( X \) is equal in distribution to \( X_1 + \cdots + X_n \), where \( X_1, \ldots, X_n \) are i.i.d. with Negative-Binomial(\( p, r/n \)) distribution. We have not yet verified that the Negative Binomial distribution permits non-integer values of \( r \), but you have been left with this task as a homework exercise.