Remark 1.2. It is worth pointing out that two random variables, $X$ and $Y$, may well have the property $X \overset{d}{=} Y$ and yet $X(\omega) \neq Y(\omega)$ for all $\omega$. A very simple example is the following: Toss a fair coin once and set

$$X = \begin{cases} 1, & \text{if the outcome is heads,} \\ 0, & \text{if the outcome is tails,} \end{cases}$$

and

$$Y = \begin{cases} 1, & \text{if the outcome is tails,} \\ 0, & \text{if the outcome is heads.} \end{cases}$$

Clearly, $X \in \text{Be}(1/2)$ and $Y \in \text{Be}(1/2)$, in particular, $X \overset{d}{=} Y$. But $X(\omega)$ and $Y(\omega)$ differ for every $\omega$. \hfill \square

2 The Probability Generating Function

Definition 2.1. Let $X$ be a nonnegative, integer-valued random variable. The (probability) generating function of $X$ is

$$g_X(t) = E t^X = \sum_{n=0}^{\infty} t^n \cdot P(X = n).$$

Remark 2.1. The generating function is defined at least for $|t| \leq 1$, since it is a power series with coefficients in $[0, 1]$. Note also that $g_X(1) = \sum_{n=0}^{\infty} P(X = n) = 1$. \hfill \square

Theorem 2.1. Let $X$ and $Y$ be nonnegative, integer-valued random variables. If $g_X = g_Y$, then $p_X = p_Y$. \hfill \square

The theorem states that if two nonnegative, integer-valued random variables have the same generating function then they follow the same probability law. It is thus the uniqueness theorem mentioned in the previous section. The result is a special case of the uniqueness theorem for power series. We refer to the literature cited in Appendix A for a complete proof.

Theorem 2.2. Let $X_1, X_2, \ldots, X_n$ be independent, nonnegative, integer-valued random variables, and set $S_n = X_1 + X_2 + \cdots + X_n$. Then

$$g_{S_n}(t) = \prod_{k=1}^{n} g_{X_k}(t).$$

Proof. Since $X_1, X_2, \ldots, X_n$ are independent, the same is true for $t^{X_1}, t^{X_2}, \ldots, t^{X_n}$, which yields

$$g_{S_n}(t) = E t^{X_1 + X_2 + \cdots + X_n} = E \prod_{k=1}^{n} t^{X_k} = \prod_{k=1}^{n} E t^{X_k} = \prod_{k=1}^{n} g_{X_k}(t).$$ \hfill \square
This result asserts that adding independent, nonnegative, integer-valued random variables corresponds to multiplying their generating functions (recall Example 1.1(a)).

A case of particular importance is given next.

**Corollary 2.2.1.** If, in addition, $X_1, X_2, \ldots, X_n$ are equidistributed, then

$$g_{S_n}(t) = (g_X(t))^n.$$  \hfill $\square$

Termwise differentiation of the generating function (this is permitted (at least) for $|t| < 1$) yields

$$g'_X(t) = \sum_{n=1}^{\infty} nt^{n-1} P(X = n), \quad (2.1)$$

$$g''_X(t) = \sum_{n=2}^{\infty} n(n-1)t^{n-2} P(X = n), \quad (2.2)$$

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

$$g^{(k)}_X(t) = \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1)t^{n-k} P(X = n). \quad (2.3)$$

By putting $t = 0$ in (2.1)–(2.3), we obtain $g^{(n)}_X(0) = n! \cdot P(X = n)$, that is,

$$P(X = n) = \frac{g^{(n)}_X(0)}{n!}. \quad (2.4)$$

The probability generating function thus generates the probabilities; hence the name of the transform.

By letting $t \to 1$ in (2.1)–(2.3) (this requires a little more care), the following result is obtained.

**Theorem 2.3.** Let $X$ be a nonnegative, integer-valued random variable, and suppose that $E|X|^k < \infty$ for some $k = 1, 2, \ldots$. Then

$$EX(X - 1)\cdots(X - k + 1) = g^{(k)}_X(1). \quad \square$$

**Remark 2.2.** Derivatives at $t = 1$ are throughout to be interpreted as limits as $t \to 1$. For simplicity, however, we use the simpler notation $g'(1), g''(1),$ and so on. \hfill $\square$

The following example illustrates the relevance of this remark.
Example 2.1. Suppose that $X$ has the probability function

$$p(k) = \frac{C}{k^2}, \quad k = 1, 2, 3, \ldots,$$

(where, to be precise, $C^{-1} = \sum_{k=1}^{\infty} 1/k^2 = \pi^2/6$). The divergence of the harmonic series tells us that the distribution does not have a finite mean.

Now, the generating function is

$$g(t) = \frac{6}{\pi^2} \sum_{k=1}^{\infty} \frac{t^k}{k^2}, \quad \text{for } |t| \leq 1,$$

so that

$$g'(t) = \frac{6}{\pi^2} \sum_{k=1}^{\infty} \frac{t^{k-1}}{k} = -\frac{6}{\pi^2} \frac{\log(1-t)}{t} \xrightarrow{t \to 1} +\infty.$$

This shows that although the generating function itself exists for $t = 1$, the derivative only exists for all $t$ strictly smaller than 1, but not for the boundary value $t = 1$. 

For $k = 1$ and $k = 2$ we have, in particular, the following result:

Corollary 2.3.1 Let $X$ be as before. Then

(a) $E|X| < \infty \implies E X = g_X'(1),$ and

(b) $E X^2 < \infty \implies \text{Var } X = g''_X(1) + g'_X(1) - (g'_X(1))^2.$ 

Exercise 2.1. Prove Corollary 2.3.1.

Next we consider some special distributions:

The Bernoulli distribution. Let $X \in \text{Be}(p)$. Then

$$g_X(t) = q \cdot t^0 + p \cdot t^1 = q + pt, \quad \text{for all } t,$$

$$g'_X(t) = p, \quad \text{and } g''_X(t) = 0,$$

which yields

$$E X = g'_X(1) = p$$

and

$$\text{Var } X = g''_X(1) + g'_X(1) - (g'_X(1))^2 = 0 + p - p^2 = p(1-p) = pq.$$

The binomial distribution. Let $X \in \text{Bin}(n, p)$. Then

$$g_X(t) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (pt)^k q^{n-k} = (q + pt)^n,$$

for all $t$. Furthermore,
\[ g'(t) = n(q + pt)^{n-1} \cdot p \quad \text{and} \quad g''(t) = n(n - 1)(q + pt)^{n-2} \cdot p^2, \]

which yields
\[ E X = np \quad \text{and} \quad \text{Var} \ X = n(n - 1)p^2 + np - (np)^2 = npq. \]

We further observe that
\[ g \binom{n}{p}(t) = (g \be(p)(t))^n, \]

which, according to Corollary 2.2.1, shows that if \( Y_1, Y_2, \ldots, Y_n \) are independent, \( \be(p) \)-distributed random variables, and \( X_n = Y_1 + Y_2 + \cdots + Y_n \), then
\[ g_{X_n}(t) = g \binom{n}{p}(t). \]

By Theorem 2.1 (uniqueness) it follows that \( X_n \in \binom{n}{p} \), a conclusion that, alternatively, could be proved by the convolution formula and induction.

Similarly, if \( X_1 \in \binom{n_1}{p} \) and \( X_2 \in \binom{n_2}{p} \) are independent, then, by Theorem 2.2,
\[ g_{X_1 + X_2}(t) = (q + pt)^{n_1 + n_2} = g \binom{n_1 + n_2}{p}(t), \]

which proves that \( X_1 + X_2 \in \binom{n_1 + n_2}{p} \) and hence establishes, in a simple manner, the addition theorem for the binomial distribution.

Remark 2.3. It is instructive to reprove the last results by actually using the convolution formula. We stress, however, that the simplicity of the method of generating functions is illusory, since it in fact exploits various results on generating functions and their derivatives. \( \square \)

The geometric distribution. Let \( X \in \ge(p) \). Then
\[ g_X(t) = \sum_{k=0}^{\infty} t^k pq^k = p \sum_{k=0}^{\infty} (tq)^k = \frac{p}{1 - qt}, \quad |t| < \frac{1}{q}. \]

Moreover,
\[ g_X'(t) = -\frac{p}{(1 - qt)^2} \cdot (-q) = \frac{pq}{(1 - qt)^2} \]
and
\[ g_X''(t) = -\frac{2pq}{(1 - qt)^3} \cdot (-q) = \frac{2pq^2}{(1 - qt)^3}, \]

from which it follows that \( E X = q/p \) and \( \text{Var} \ X = q/p^2 \).

Exercise 2.2. Let \( X_1, X_2, \ldots, X_n \) be independent \( \ge(p) \)-distributed random variables. Determine the distribution of \( X_1 + X_2 + \cdots + X_n \). \( \square \)
The Poisson distribution. Let $X \in \text{Po}(m)$. Then

$$g_X(t) = \sum_{k=0}^{\infty} t^k e^{-m} \frac{m^k}{k!} = e^{-m} \sum_{k=0}^{\infty} \frac{(mt)^k}{k!} = e^{m(t-1)}.$$  

**Exercise 2.3.** (a) Let $X \in \text{Po}(m)$. Show that $EX = \text{Var} \, X = m$.
(b) Let $X_1 \in \text{Po}(m_1)$ and $X_2 \in \text{Po}(m_2)$ be independent random variables. Show that $X_1 + X_2 \in \text{Po}(m_1 + m_2)$. \hfill \box

### 3 The Moment Generating Function

In spite of their usefulness, probability generating functions are of limited use in that they are only defined for nonnegative, integer-valued random variables. Important distributions, such as the normal distribution and the exponential distribution, cannot be handled with this transform. This inconvenience is overcome as follows:

**Definition 3.1.** The moment generating function of a random variable $X$ is

$$\psi_X(t) = E e^{tX},$$

provided there exists $h > 0$, such that the expectation exists and is finite for $|t| < h$.

**Remark 3.1.** As a first observation we mention the close connection between moment generating functions and Laplace transforms of real-valued functions. Indeed, for a nonnegative random variable $X$, one may define the Laplace transform

$$E e^{-sX} \quad \text{for} \quad s \geq 0,$$

which thus always exist (why?). Analogously, one may view the moment generating function as a two-sided Laplace transform.

**Remark 3.2.** Note that for nonnegative, integer-valued random variables we have $\psi(t) = g(e^t)$, for $|t| < h$, provided the moment generating function exists (for $|t| < h$). \hfill \box

The uniqueness and multiplication theorems are presented next. The proofs are analogous to those for the generating function.

**Theorem 3.1.** Let $X$ and $Y$ be random variables. If there exists $h > 0$, such that $\psi_X(t) = \psi_Y(t)$ for $|t| < h$, then $X \overset{d}{=} Y$. \hfill \box

**Theorem 3.2.** Let $X_1, X_2, \ldots, X_n$ be independent random variables whose moment generating functions exist for $|t| < h$ for some $h > 0$, and set $S_n = X_1 + X_2 + \cdots + X_n$. Then

$$\psi_{S_n}(t) = \prod_{k=1}^{n} \psi_{X_k}(t), \quad |t| < h.$$  

\hfill \box