Chapter 3
Generating Functions

It is a fundamental principle of mathematics to map a class of objects that are of interest into a class of objects where computations are easier. This map can be one to one, as with linear maps and matrices, or it may map only some properties uniquely, as with matrices and determinants.

In probability theory, in the second category fall quantities such as the median, mean and variance of random variables. In the first category, we have characteristic functions, Laplace transforms and probability generating functions. These are useful mostly because addition of independent random variables leads to multiplication of the transforms. Before we introduce characteristic functions (and Laplace transforms) later in the book, we want to illustrate the basic idea with probability generating functions that are designed for $\mathbb{N}_0$-valued random variables.

In the first section, we give the basic definitions and derive simple properties. The next two sections are devoted to two applications: The Poisson approximation theorem and a simple investigation of Galton–Watson branching processes.

3.1 Definition and Examples

Definition 3.1 (Probability generating function) Let X be an $\mathbb{N}_0$-valued random variable. The (probability) *generating function* (p.g.f.) of $\mathbf{P}_X$ (or, loosely speaking, of X) is the map $\psi_{\mathbf{P}_X} = \psi_X$ defined by (with the understanding that $0^0 = 1$)

$$\psi_X : [0, 1] \to [0, 1], \quad z \mapsto \sum_{n=0}^{\infty} \mathbf{P}[X = n] z^n. \tag{3.1}$$

Theorem 3.2

(i) ψ_X *is continuous on* $[0, 1]$ *and infinitely often continuously differentiable on* $(0, 1)$. *For* $n \in \mathbb{N}$, *the* n*th derivative* $\psi_X^{(n)}$ *fulfills*

A. Klenke, *Probability Theory*, Universitext, DOI 10.1007/978-1-4471-5361-0_3,

$$\lim_{z\uparrow 1}\psi_X^{(n)}(z)=\sum_{k=n}^{\infty}\mathbf{P}[X=k]\cdot k(k-1)\ldots(k-n+1), \tag{3.2}$$

where both sides can equal ∞.

(ii) *The distribution* $\mathbf{P}_X$ *of* X *is uniquely determined by* ψ_X.

(iii) *For any* $r\in(0,1)$, ψ_X *is uniquely determined by countably many values* $\psi_X(x_i)$, $x_i\in[0,r]$, $i\in\mathbb{N}$. *If the series in* (3.1) *converges for some* $z>1$, *then the statement is also true for any* $r\in(0,z)$ *and we have*

$$\lim_{x\uparrow 1}\psi_X^{(n)}(x)=\psi_X^{(n)}(1)<\infty \quad \textit{for } n\in\mathbb{N}.$$

In this case, ψ_X *is uniquely determined by the derivatives* $\psi_X^{(n)}(1)$, $n\in\mathbb{N}$.

Proof The statements follow from the elementary theory of power series. For the first part of (iii), see, e.g. [148, Theorem 8.5]. □

Theorem 3.3 (Multiplicativity of generating functions) *If* $X_1,\ldots,X_n$ *are independent and* $\mathbb{N}_0$*-valued random variables, then*

$$\psi_{X_1+\ldots+X_n}=\prod_{i=1}^{n}\psi_{X_i}.$$

Proof Let $z\in[0,1)$ and write $\psi_{X_1}(z)\,\psi_{X_2}(z)$ as a Cauchy product

$$\begin{aligned}\psi_{X_1}(z)\,\psi_{X_2}(z)&=\left(\sum_{n=0}^{\infty}\mathbf{P}[X_1=n]z^n\right)\left(\sum_{n=0}^{\infty}\mathbf{P}[X_2=n]z^n\right)\\&=\sum_{n=0}^{\infty}z^n\left(\sum_{m=0}^{n}\mathbf{P}[X_1=m]\mathbf{P}[X_2=n-m]\right)\\&=\sum_{n=0}^{\infty}z^n\sum_{m=0}^{n}\mathbf{P}[X_1=m,\ X_2=n-m]\\&=\sum_{n=0}^{\infty}\mathbf{P}[X_1+X_2=n]z^n=\psi_{X_1+X_2}(z).\end{aligned}$$

Inductively, the claim follows for all $n\geq 2$. □

Example 3.4

(i) Let X be $b_{n,p}$-distributed for some $n\in\mathbb{N}$ and let $p\in[0,1]$. Then

$$\psi_X(z)=\sum_{m=0}^{n}\binom{n}{m}p^m(1-p)^{n-m}z^m=\big(pz+(1-p)\big)^n. \tag{3.3}$$

(ii) If X, Y are independent, $X \sim b_{m,p}$ and $Y \sim b_{n,p}$, then, by Theorem 3.3,

$$\psi_{X+Y}(z) = \big(pz + (1-p)\big)^m \big(pz + (1-p)\big)^n = \big(pz + (1-p)\big)^{m+n}.$$

Hence, by Theorem 3.2(ii), $X + Y$ is $b_{m+n,p}$-distributed and thus (by Theorem 2.31)

$$b_{m,p} * b_{n,p} = b_{m+n,p}.$$

(iii) Let X and Y be independent Poisson random variables with parameters $\lambda \geq 0$ and $\mu \geq 0$, respectively. That is, $\mathbf{P}[X = n] = e^{-\lambda}\lambda^n/n!$ for $n \in \mathbb{N}_0$. Then

$$\psi_{\mathrm{Poi}_\lambda}(z) = \sum_{n=0}^{\infty} e^{-\lambda} \frac{(\lambda z)^n}{n!} = e^{\lambda(z-1)}. \tag{3.4}$$

Hence $X + Y$ has probability generating function

$$\psi_{\mathrm{Poi}_\lambda}(z) \cdot \psi_{\mathrm{Poi}_\mu}(z) = e^{\lambda(z-1)} e^{\mu(z-1)} = \psi_{\mathrm{Poi}_{\lambda+\mu}}(z).$$

Thus $X + Y \sim \mathrm{Poi}_{\lambda+\mu}$. We conclude that

$$\mathrm{Poi}_\lambda * \mathrm{Poi}_\mu = \mathrm{Poi}_{\lambda+\mu}. \tag{3.5}$$

(iv) Let $X_1, \ldots, X_n \sim \gamma_p$ be independent geometrically distributed random variables with parameter $p \in (0, 1)$. Define $Y = X_1 + \ldots + X_n$. Then, for any $z \in [0, 1]$,

$$\psi_{X_1}(z) = \sum_{k=0}^{\infty} p(1-p)^k z^k = \frac{p}{1-(1-p)z}. \tag{3.6}$$

By the generalized binomial theorem (see Lemma 3.5 with $\alpha = -n$), Theorem 3.3 and (3.6), we have

$$\begin{aligned} \psi_Y(z) = \psi_{X_1}(z)^n &= \frac{p^n}{(1-(1-p)z)^n} \\ &= \sum_{k=0}^{\infty} p^n \binom{-n}{k} (-1)^k (1-p)^k z^k \\ &= \sum_{k=0}^{\infty} b^-_{n,p}\big(\{k\}\big) z^k. \end{aligned}$$

Here, for $r \in (0, \infty)$ and $p \in (0, 1]$,

$$b^-_{r,p} = \sum_{k=0}^{\infty} \binom{-r}{k} (-1)^k p^r (1-p)^k \delta_k \tag{3.7}$$

is the negative binomial distribution with parameters r and p. By the uniqueness theorem for probability generating functions, we get $Y \sim b^-_{n,p}$; hence (see Definition 2.29 for the nth convolution power) $b^-_{n,p} = \gamma_p^{*n}$. ◊

Lemma 3.5 (Generalized binomial theorem) *For $\alpha \in \mathbb{R}$ and $k \in \mathbb{N}_0$, we define the binomial coefficient*

$$\binom{\alpha}{k} := \frac{\alpha \cdot (\alpha - 1) \ldots (\alpha - k + 1)}{k!}. \tag{3.8}$$

Then the generalized binomial theorem holds:

$$(1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k \quad \textit{for all } x \in \mathbb{C} \textit{ with } |x| < 1. \tag{3.9}$$

In particular, we have

$$\frac{1}{\sqrt{1-x}} = \sum_{n=0}^{\infty} \binom{2n}{n} 4^{-n} x^n \quad \textit{for all } x \in \mathbb{C} \textit{ with } |x| < 1. \tag{3.10}$$

Proof The map $f : x \mapsto (1+x)^\alpha$ is holomorphic up to possibly a singularity at $x = -1$. Hence it can be developed in a power series about 0 with radius of convergence at least 1:

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k \quad \text{for } |x| < 1.$$

For $k \in \mathbb{N}_0$, the kth derivative is $f^{(k)}(0) = \alpha(\alpha - 1) \ldots (\alpha - k + 1)$. Hence (3.9) holds.

The additional claim follows by the observation that (for $\alpha = -1/2$) we have $\binom{-1/2}{n} = \binom{2n}{n}(-4)^{-n}$. □

Exercise 3.1.1 Show that $b^-_{r,p} * b^-_{s,p} = b^-_{r+s,p}$ for $r, s \in (0, \infty)$ and $p \in (0, 1]$.

Exercise 3.1.2 Give an example for two different probability generating functions that coincide at countably many points $x_i \in (0, 1)$, $i \in \mathbb{N}$. (That is, in Theorem 3.2(iii), the assumption $\psi(z) < \infty$ for some $z > 1$ cannot be dropped.)

3.2 Poisson Approximation

Lemma 3.6 *Let μ and $(\mu_n)_{n \in \mathbb{N}}$ be probability measures on $(\mathbb{N}_0, 2^{\mathbb{N}_0})$ with generating functions ψ and ψ_n, $n \in \mathbb{N}$. Then the following statements are equivalent.*

(i) $\mu_n(\{k\}) \overset{n\to\infty}{\longrightarrow} \mu(\{k\})$ *for all* $k \in \mathbb{N}_0$.
(ii) $\mu_n(A) \overset{n\to\infty}{\longrightarrow} \mu(A)$ *for all* $A \subset \mathbb{N}_0$.
(iii) $\psi_n(z) \overset{n\to\infty}{\longrightarrow} \psi(z)$ *for all* $z \in [0, 1]$.
(iv) $\psi_n(z) \overset{n\to\infty}{\longrightarrow} \psi(z)$ *for all* $z \in [0, \eta)$ *for some* $\eta \in (0, 1)$.

We write $\mu_n \overset{n\to\infty}{\longrightarrow} \mu$ *if any of the four conditions holds and say that* $(\mu_n)_{n\in\mathbb{N}}$ *converges weakly to* μ.

Proof (i) $\Longrightarrow$ (ii) Fix $\varepsilon > 0$ and choose $N \in \mathbb{N}$ such that

$$\mu\big(\{N+1, N+2, \ldots\}\big) < \frac{\varepsilon}{4}.$$

For sufficiently large $n_0 \in \mathbb{N}$, we have

$$\sum_{k=0}^{N} \big|\mu_n\big(\{k\}\big) - \mu\big(\{k\}\big)\big| < \frac{\varepsilon}{4} \quad \text{for all } n \ge n_0.$$

In particular, for any $n \ge n_0$, we have $\mu_n(\{N+1, N+2, \ldots\}) < \frac{\varepsilon}{2}$. Hence, for $n \ge n_0$,

$$\begin{aligned}\big|\mu_n(A) - \mu(A)\big| &\le \mu_n\big(\{N+1, N+2, \ldots\}\big) + \mu\big(\{N+1, N+2, \ldots\}\big)\\ &\quad + \sum_{k\in A\cap\{0,\ldots,N\}} \big|\mu_n\big(\{k\}\big) - \mu\big(\{k\}\big)\big|\\ &< \varepsilon.\end{aligned}$$

(ii) $\Longrightarrow$ (i) This is trivial.

(i) $\Longleftrightarrow$ (iii) $\Longleftrightarrow$ (iv) This follows from the elementary theory of power series. □

Let $(p_{n,k})_{n,k\in\mathbb{N}}$ be numbers with $p_{n,k} \in [0, 1]$ such that the limit

$$\lambda := \lim_{n\to\infty} \sum_{k=1}^{\infty} p_{n,k} \in (0, \infty) \tag{3.11}$$

exists and such that $\lim_{n\to\infty} \sum_{k=1}^{\infty} p_{n,k}^2 = 0$ (e.g., $p_{n,k} = \lambda/n$ for $k \le n$ and $p_{n,k} = 0$ for $k > n$). For each $n \in \mathbb{N}$, let $(X_{n,k})_{k\in\mathbb{N}}$ be an independent family of random variables with $X_{n,k} \sim \mathrm{Ber}_{p_{n,k}}$.

Define

$$S^n := \sum_{l=1}^{\infty} X_{n,l} \quad \text{and} \quad S_k^n := \sum_{l=1}^{k} X_{n,l} \quad \text{for } k \in \mathbb{N}.$$

Theorem 3.7 (Poisson approximation) *Under the above assumptions, the distributions* $(\mathbf{P}_{S^n})_{n\in\mathbb{N}}$ *converge weakly to the Poisson distribution* Poi_λ.

Proof The p.g.f. of the Poisson distribution is $\psi(z) = e^{\lambda(z-1)}$ (see (3.4)). On the other hand, $S^n - S^n_k$ and S^n_k are independent for any $k \in \mathbb{N}$; hence $\psi_{S^n} = \psi_{S^n_k} \cdot \psi_{S^n - S^n_k}$. Now, for any $z \in [0, 1]$,

$$1 \geq \frac{\psi_{S^n}(z)}{\psi_{S^n_k}(z)} = \psi_{S^n - S^n_k}(z) \geq 1 - \mathbf{P}\big[S^n - S^n_k \geq 1\big] \geq 1 - \sum_{l=k+1}^{\infty} p_{n,l} \overset{k\to\infty}{\longrightarrow} 1,$$

hence

$$\psi_{S^n}(z) = \lim_{k\to\infty} \psi_{S^n_k}(z) = \prod_{l=1}^{\infty}\big(p_{n,l}z + (1 - p_{n,l})\big)$$
$$= \exp\left(\sum_{l=1}^{\infty} \log\big(1 + p_{n,l}(z-1)\big)\right).$$

Note that $|\log(1 + x) - x| \leq x^2$ for $|x| < \frac{1}{2}$. By assumption, $\max_{l\in\mathbb{N}} p_{n,l} \to 0$ for $n \to \infty$; hence, for sufficiently large n,

$$\left|\left(\sum_{l=1}^{\infty} \log\big(1 + p_{n,l}(z-1)\big)\right) - \left((z-1)\sum_{l=1}^{\infty} p_{n,l}\right)\right|$$
$$\leq \sum_{l=1}^{\infty} p^2_{n,l} \leq \left(\sum_{l=1}^{\infty} p_{n,l}\right) \max_{l\in\mathbb{N}} p_{n,l} \overset{n\to\infty}{\longrightarrow} 0.$$

Together with (3.11), we infer

$$\lim_{n\to\infty} \psi_{S^n}(z) = \lim_{n\to\infty} \exp\left((z-1)\sum_{l=1}^{\infty} p_{n,l}\right) = e^{\lambda(z-1)}.$$

□

3.3 Branching Processes

Let $T, X_1, X_2, \ldots$ be independent $\mathbb{N}_0$-valued random variables. What is the distribution of $S := \sum_{n=1}^{T} X_n$? First of all, note that S is measurable since

$$\{S = k\} = \bigcup_{n=0}^{\infty}\{T = n\} \cap \{X_1 + \ldots + X_n = k\}.$$

Theorem 3.8 *If the random variables $X_1, X_2, \ldots$ are also identically distributed, then the probability generating function of S is given by $\psi_S(z) = \psi_T(\psi_{X_1}(z))$.*

Proof We compute

$$\begin{aligned}\psi_S(z) &= \sum_{k=0}^{\infty} \mathbf{P}[S=k]z^k \\ &= \sum_{k=0}^{\infty}\sum_{n=0}^{\infty} \mathbf{P}[T=n]\mathbf{P}[X_1+\ldots+X_n=k]z^k \\ &= \sum_{n=0}^{\infty} \mathbf{P}[T=n]\psi_{X_1}(z)^n = \psi_T\big(\psi_{X_1}(z)\big).\end{aligned}$$

□

Now assume that $p_0, p_1, p_2, \ldots \in [0,1]$ are such that $\sum_{k=0}^{\infty} p_k = 1$. Let $(X_{n,i})_{n,i\in\mathbb{N}_0}$ be an independent family of random variables with $\mathbf{P}[X_{n,i}=k]=p_k$ for all $i,k,n \in \mathbb{N}_0$.

Let $Z_0 = 1$ and

$$Z_n = \sum_{i=1}^{Z_{n-1}} X_{n-1,i} \quad \text{for } n \in \mathbb{N}.$$

Z_n can be interpreted as the number of individuals in the nth generation of a randomly developing population. The ith individual in the nth generation has $X_{n,i}$ offspring (in the $(n+1)$th generation).

Definition 3.9 $(Z_n)_{n\in\mathbb{N}_0}$ is called a *Galton–Watson process* or *branching process* with offspring distribution $(p_k)_{k\in\mathbb{N}_0}$.

Probability generating functions are an important tool for the investigation of branching processes. Hence, let

$$\psi(z) = \sum_{k=0}^{\infty} p_k z^k$$

be the p.g.f. of the offspring distribution and let ψ' be its derivative. Recursively, define the nth iterate of ψ by

$$\psi_1 := \psi \quad \text{and} \quad \psi_n := \psi \circ \psi_{n-1} \quad \text{for } n = 2, 3, \ldots.$$

Finally, let ψ_{Z_n} be the p.g.f. of Z_n.

Lemma 3.10 $\psi_n = \psi_{Z_n}$ *for all* $n \in \mathbb{N}$.

Proof For $n=1$, the statement is true by definition. For $n \in \mathbb{N}$, we conclude inductively by Theorem 3.8 that $\psi_{Z_{n+1}} = \psi \circ \psi_{Z_n} = \psi \circ \psi_n = \psi_{n+1}$. □

Clearly, the probability $q_n := \mathbf{P}[Z_n = 0]$ that Z is extinct by time n is monotone increasing in n. We denote by

$$q := \lim_{n\to\infty} \mathbf{P}[Z_n = 0]$$

the *extinction probability*; that is, the probability that the population will *eventually* die out.

Under what conditions do we have $q = 0$, $q = 1$, or $q \in (0, 1)$? Clearly, $q \geq p_0$. On the other hand, if $p_0 = 0$, then Z_n is monotone in n; hence $q = 0$.

Theorem 3.11 (Extinction probability of the Galton–Watson process) *Assume* $p_1 \neq 1$. *Then*:

(i) $F := \{r \in [0, 1] : \psi(r) = r\} = \{q, 1\}$.
(ii) *The following equivalences hold*:

$$q < 1 \quad\Longleftrightarrow\quad \lim_{z\uparrow 1} \psi'(z) > 1 \quad\Longleftrightarrow\quad \sum_{k=1}^{\infty} k p_k > 1.$$

Proof (i) We have $\psi(1) = 1$; hence $1 \in F$. Note that

$$q_n = \psi_n(0) = \psi(q_{n-1}) \quad \text{for all } n \in \mathbb{N}$$

and $q_n \uparrow q$. Since ψ is continuous, we infer

$$\psi(q) = \psi\Big(\lim_{n\to\infty} q_n\Big) = \lim_{n\to\infty} \psi(q_n) = \lim_{n\to\infty} q_{n+1} = q.$$

Thus $q \in F$. If $r \in F$ is an arbitrary fixed point of ψ, then $r \geq 0 = q_0$. Since ψ is monotone increasing, it follows that $r = \psi(r) \geq \psi(q_0) = q_1$. Inductively, we get $r \geq q_n$ for all $n \in \mathbb{N}_0$; that is, $r \geq q$. We conclude $q = \min F$.

(ii) For the first equivalence, we distinguish two cases.

Case 1: $\lim_{z\uparrow 1} \psi'(z) \leq 1$. Since ψ is strictly convex, in this case, we have $\psi(z) > z$ for all $z \in [0, 1)$; hence $F = \{1\}$. We conclude $q = 1$.

Case 2: $\lim_{z\uparrow 1} \psi'(z) > 1$. As ψ is strictly convex and since $\psi(0) \geq 0$, there is a unique $r \in [0, 1)$ such that $\psi(r) = r$. Hence $F = \{r, 1\}$ and $q = \min F = r$.

The second equivalence in (ii) follows by (3.2). □

For further reading, we refer to [5].