Course:	Introduction to Stochastic Processes
Term:	Fall 2019
Instructor:	Gordan Žitković

Lecture 4
Generating functions

The path-counting method used in the previous lecture only works for finite-horizon walks, where the of the horizon T is given in advance. We will see later that most of the interesting questions do *not* fall into this category. For example, the distribution of *the time it takes* for the random walk to hit the level $l \neq 0$ is like that. While we know that it will happen eventually, there is no way to give an a-priori upper bound on the number of steps it will take to get to l. To deal with a wider class of properties of random walks (and other processes), we need to develop some new mathematical tools. The central among those is the method of generating functions we describe in this lecture.

4.1 Definition and first properties

Generating functions provide a link between probability and analysis (calculus). The definition provided below is somewhat confusing at first, as it initially lacks any intuition. To give the proper treatment to the "why" behind it, we would need to introduce a lot more mathematics than we can at the moment[1]. The best way to gain understanding without the advanced mathematical background is to learn how to work with generating functions and appreciate what they can do for us.

Even though it can be used to deal with many other kinds of distributions, we will use generating functions to study distributions of random variables that take values in the set $\mathbb{N}_0$ of natural numbers (and zero). Such random variables often model *random times*, but can also be given other interpretations.

The distribution (table) of an $\mathbb{N}_0$-valued random variable looks like this

0	1	2	3	...
p_0	p_1	p_2	p_3	...

and so, the distribution is completely defined by its probability-mass function (pmf) $\{p_k\}_{k\in\mathbb{N}_0}$. Each such pmf is a **probability sequence**, i.e., a sequence

[1] if you are curious, generating functions are a special case of the Fourier/Laplace transform and the proper setting for their understanding is within the subfield of mathematics called harmonic analysis.

of numbers between 0 and 1 whose sum is 1. For example, the pmf of the random variable X which models a roll of a fair die is the sequence $p_0 = 0$, $p_1 = 1/6$, $p_2 = 1/6$, ..., $p_6 = 1/6$, $p_7 = 0$, $p_8 = 0$, ...

Definition 4.1.1. The **generating function** of a probability sequence $\{p_k\}_{k\in\mathbb{N}_0}$ is the function P defined by the following power series:

$$P(s) = \sum_{k=0}^{\infty} p_k s^k. \tag{4.1.1}$$

Given an $\mathbb{N}_0$-valued random variable X, we define its generating function simply as the generating function of its pmf, i.e., of the sequence $\{p_k\}_{k\in\mathbb{N}_0}$, with $p_k = \mathbb{P}[X = k]$, $k \in \mathbb{N}_0$. This generating function is denoted by P_X.

Since $\sum_k p_k = 1$ for each probability sequence the radius of convergence[2] of $\{p_k\}_{k\in\mathbb{N}_0}$ is at least equal to 1. Therefore, the function $P(s)$ given by (4.1.1) is well defined for $s \in [-1, 1]$, and, perhaps, other values of s, too.

Let us start by deriving expressions for the generating functions of some of the popular $\mathbb{N}_0$-valued distributions.

Example 4.1.2.

1. **Bernoulli** ($b(p)$). Here $p_0 = q$, $p_1 = p$, and $p_k = 0$, for $k \geq 2$. Therefore,
$$P_X(s) = ps + q.$$
2. **Binomial** ($b(n, p)$). Since $p_k = \binom{n}{k} p^k q^{n-k}$, $k = 0, \dots, n$, we have
$$P_X(s) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} s^k = (ps + q)^n,$$
by the binomial theorem.
3. **Geometric** ($g(p)$). For $k \in \mathbb{N}_0$, $p_k = q^k p$, so that
$$P_X(s) = \sum_{k=0}^{\infty} q^k s^k p = p \sum_{k=0}^{\infty} (qs)^k = \frac{p}{1 - qs}.$$
4. **Poisson** ($P(\lambda)$). Given that $p_k = e^{-\lambda} \frac{\lambda^k}{k!}$, $k \in \mathbb{N}_0$, we have
$$P_X(s) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} s^k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(s\lambda)^k}{k!} = e^{-\lambda} e^{s\lambda} = e^{\lambda(s-1)}.$$

[2] Remember, that the **radius of convergence** of a power series $\sum_{k=0}^{\infty} a_k x^k$ is the largest number $R \in [0, \infty]$ such that $\sum_{k=0}^{\infty} a_k x^k$ converges absolutely whenever $|x| < R$.

Some of the most useful analytic properties of P_X are listed in the following proposition

Proposition 4.1.3. *Let X be an $\mathbb{N}_0$-valued random variable, $\{p_k\}_{k\in\mathbb{N}_0}$ its pmf, and P_X its generating function. Then*

1. $P_X(s) = \mathbb{E}[s^X]$, $s \in [-1,1]$,
2. *$P_X(s)$ is convex and non-decreasing with $0 \leq P_X(s) \leq 1$ for $s \in [0,1]$*
3. *$P_X(s)$ is infinitely differentiable on $(-1,1)$ with*

$$\frac{d^k}{ds^k} P_X(s) = \sum_{j=k}^{\infty} j(j-1)\dots(j-k+1)s^{j-k}p_j,\ k \in \mathbb{N}. \tag{4.1.2}$$

In particular,

$$p_k = \frac{1}{k!}\frac{d^k}{ds^k}P_X(s)\Big|_{s=0}$$

and so $s \mapsto P_X(s)$ uniquely determines the sequence $\{p_k\}_{k\in\mathbb{N}_0}$.

Proof. Statement 1. follows directly from the formula

$$\mathbb{E}[g(X)] = \sum_{k=0}^{\infty} g(k)p_k,$$

applied to $g(x) = s^x$. As far as (3) is concerned, we only note that the expression (4.1.2) is exactly what you would get if you differentiated the expression (4.1.1) term by term. The rigorous proof of the fact this is allowed is beyond the scope of these notes. With 3. at our disposal, 2. follows by the fact that the first two derivatives of the function P_X are non-negative and that $P_X(1) = 1$. □

Remark 4.1.4.

1. If you know about moment-generating functions, you will notice that $P_X(s) = M_X(\log(s))$, for $s \in (0,1)$, where $M_X(t) = \mathbb{E}[\exp(tX)]$ is the moment-generating function of X.
2. Generating functions can be used with sequences $\{a_k\}_{k\in\mathbb{N}_0}$ that are not necessarily pmf's of random variables. The method is useful for any sequence $\{a_k\}_{k\in\mathbb{N}_0}$ such that the power series $\sum_{k=0}^{\infty} a_k s^k$ has a positive (nonzero) radius of convergence (see the problem about Fibbonacci numbers in the Problems section).
3. The name *generating function* comes from the last part of the property (3). The knowledge of P_X implies the knowledge of the whole sequence $\{p_k\}_{k\in\mathbb{N}_0}$. Put differently, P_X generates the whole distribution of X.

4.2 Convolution and moments

The true power of generating functions comes from the fact that they behave very well under the usual operations in probability.

Definition 4.2.1. Let $\{p_k\}_{k\in\mathbb{N}_0}$ and $\{q_k\}_{k\in\mathbb{N}_0}$ be two sequences. The **convolution** $p * q$ of $\{p_k\}_{k\in\mathbb{N}_0}$ and $\{q_k\}_{k\in\mathbb{N}_0}$ is the sequence $\{r_k\}_{k\in\mathbb{N}_0}$, where

$$r_k = \sum_{j=0}^{k} p_j q_{k-j}, k \in \mathbb{N}_0.$$

This abstractly-defined operation will become much clearer once we prove the following proposition:

Proposition 4.2.2. *Let X, Y be two* independent *$\mathbb{N}_0$-valued random variables with pmfs $\{p_k\}_{k\in\mathbb{N}_0}$ and $\{q_k\}_{k\in\mathbb{N}_0}$. Then the sum $Z = X + Y$ is also $\mathbb{N}_0$-valued and its pmf is the convolution of $\{p_k\}_{k\in\mathbb{N}_0}$ and $\{q_k\}_{k\in\mathbb{N}_0}$ in the sense of Definition 4.2.1.*

Proof. Clearly, Z is $\mathbb{N}_0$-valued. To obtain an expression for its pmf, we use the law of total probability:

$$\mathbb{P}[Z = k] = \sum_{j=0}^{k} \mathbb{P}[X = j]\mathbb{P}[Z = k | X = j].$$

On the other hand,

$$\begin{aligned}\mathbb{P}[Z = k | X = j] &= \mathbb{P}[X + Y = k | X = j] = \mathbb{P}[Y = k - j | X = j] \\ &= \mathbb{P}[Y = k - j],\end{aligned}$$

where the last equality follows from independence of X and Y. Therefore,

$$\mathbb{P}[Z = k] = \sum_{j=0}^{k} \mathbb{P}[X = j]\mathbb{P}[Y = k - j] = \sum_{j=0}^{k} p_j q_{k-j}. \qquad \square$$

Corollary 4.2.3. *Let $\{p_k\}_{k\in\mathbb{N}_0}$ and $\{p_k\}_{k\in\mathbb{N}_0}$ be two sequences. Then*

1. *$p * q = q * p$, i.e., convolution is commutative.*
2. *The convolution $r = p * q$ of two* probability sequences *is a probability sequence itself, i.e., $r_k \geq 0$, for all $k \in \mathbb{N}_0$ and $\sum_{k=0}^{\infty} r_k = 1$.*

Corollary 4.2.4. *Let $\{p_k\}_{k\in\mathbb{N}_0}$ and $\{q_k\}_{k\in\mathbb{N}_0}$ be two probability sequences, and let*

$$P(s) = \sum_{k=0}^{\infty} p_k s^k \text{ and } Q(s) = \sum_{k=0}^{\infty} q_k s^k$$

be their generating functions. Then the generating function $R(s) = \sum_{k=0}^{\infty} r_k s^k$*, of the convolution* $r = p * q$ *is given by*

$$R(s) = P(s)Q(s).$$

Equivalently, the generating function P_{X+Y} *of the sum of two independent* $\mathbb{N}_0$*-valued random variables is equal to the product*

$$P_{X+Y}(s) = P_X(s)P_Y(s),$$

of the generating functions P_X *and* P_Y *of* X *and* Y*.*

Example 4.2.5.

1. The binomial $b(n, p)$ distribution is a sum of n independent Bernoullis $b(p)$. Therefore, if we apply Corrolary 4.2.4 n times to the generating function $(q + ps)$ of the Bernoulli $b(p)$ distribution we immediately get that the generating function of the binomial is

$$(q + ps) \dots (q + ps) = (q + ps)^n.$$

2. More generally, we can show that the sum of m *independent* random variables with the $b(n, p)$ distribution has a binomial distribution $b(mn, p)$. If you try to sum binomials with different values of the parameter p you will not get a binomial.

3. What is even more interesting, the following statement can be shown: Suppose that the sum Z of two independent $\mathbb{N}_0$-valued random variables X and Y is binomially distributed with parameters n and p. Then both X and Y must be binomial with parameters n_X, p and n_y, p where $n_X + n_Y = n$. In other words, the only way to get a binomial as a sum of independent random variables is if they are both binomial with the same p.

Another useful thing about generating functions is that they make the computation of moments easier.

Proposition 4.2.6. *Let* $\{p_k\}_{k\in\mathbb{N}_0}$ *be the pmf of the* $\mathbb{N}_0$*-valued random variable* X *and let* $P_X(s)$ *be its generating function. For* $n \in \mathbb{N}$ *the following two statements are equivalent*

1. $\mathbb{E}[X^k] < \infty$,
2. $\left.\frac{d^k P(s)}{ds^k}\right|_{s=1}$ *exists (in the sense that the left limit* $\lim_{s\nearrow 1} \frac{d^k P(s)}{ds^k}$ *exists)*

In either case, we have

$$\mathbb{E}[X(X-1)(X-2)\dots(X-k+1)] = \frac{d^k}{ds^k} P(s)\Big|_{s=1}.$$

The quantities

$$\mathbb{E}[X], \quad \mathbb{E}[X(X-1)], \quad \mathbb{E}[X(X-1)(X-2)], \dots$$

are called **factorial moments** of the random variable X. You can get the classical moments from the factorial moments by solving a system of linear equations. It is very simple for the first few:

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[X], \\
\mathbb{E}[X^2] &= \mathbb{E}[X(X-1)] + \mathbb{E}[X], \\
\mathbb{E}[X^3] &= \mathbb{E}[X(X-1)(X-2)]] + 3\mathbb{E}[X(X-1)] + \mathbb{E}[X], \dots
\end{aligned}$$

A useful identity which follows directly from the above results is the following:

$$\mathbb{E}[X] = P'(1) \text{ and } \operatorname{Var}[X] = P''(1) + P'(1) - (P'(1))^2,$$

and is valid if all the involved derivatives exist.

Example 4.2.7. Let X be a Poisson random variable with parameter λ. Its generating function is given by

$$P_X(s) = e^{\lambda(s-1)}.$$

Therefore, $\frac{d^k}{ds^k} P_X(1) = \lambda^k$, and so, the sequence $(\mathbb{E}[X], \mathbb{E}[X(X-1)], \mathbb{E}[X(X-1)(X-2)], \dots)$ of factorial moments of X is just $(\lambda, \lambda^2, \lambda^3, \dots)$. It follows that

$$\begin{aligned}
\mathbb{E}[X] &= \lambda, \\
\mathbb{E}[X^2] &= \lambda^2 + \lambda, \ \operatorname{Var}[X] = \lambda \\
\mathbb{E}[X^3] &= \lambda^3 + 3\lambda^2 + \lambda, \dots
\end{aligned}$$

4.3 Random sums

Our next application of generating functions in the theory of stochastic processes deals with so-called *random sums*. Let $\{\xi_k\}_{k\in\mathbb{N}}$ be a sequence of random variables, and let N be a random time (we allow the value $+\infty$ here, too). We can define the random variable

$$Y = \sum_{k=1}^{N} \xi_k$$

for $\omega \in \Omega$ by

$$Y(\omega) = \begin{cases} 0, & N(\omega) = 0, \\ \sum_{k=1}^{N(\omega)} \xi_k(\omega), & 1 \leq N(\omega) < \infty. \\ \sum_{k=1}^{\infty} \xi_k(\omega), & N(\omega) = +\infty. \end{cases}$$

In the case when N does not take the value $+\infty$, we can define this more generally: for an arbitrary stochastic process $\{X_k\}_{k \in \mathbb{N}_0}$ we define the *random variable* X_N by $X_N(\omega) = X_{N(\omega)}(\omega)$, for $\omega \in \Omega$. When N is a constant ($N = n$), then X_N is simply equal to X_n. In general, think of X_N as a value of the stochastic process X taken at the time which is itself random. If $X_n = \sum_{k=1}^{n} \xi_k$, then $X_N = \sum_{k=1}^{N} \xi_k$.

Example 4.3.1. Let $\{\xi_k\}_{k \in \mathbb{N}}$ be the increments of a symmetric simple random walk; we denoted these by $\{\delta_k\}_{k \in \mathbb{N}}$ when we talked about random walks. Let N be a random variable *independent* of all $\{\xi_k\}_{k \in \mathbb{N}}$ with the following distribution

$$N \sim \begin{pmatrix} 0 & 1 & 2 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$$

Let us compute the distribution of $Y = \sum_{k=1}^{N} \xi_k$ in this case. The formula of total probability comes in very handy here:

$$\begin{aligned} \mathbb{P}[Y = m] &= \mathbb{P}[Y = m | N = 0]\, \mathbb{P}[N = 0] + \mathbb{P}[Y = m | N = 1]\, \mathbb{P}[N = 1] \\ &\quad + \mathbb{P}[Y = m | N = 2]\, \mathbb{P}[N = 2] \\ &= \mathbb{P}[\sum_{k=0}^{N} \xi_k = m | N = 0]\, \mathbb{P}[N = 0] \\ &\quad + \mathbb{P}[\sum_{k=0}^{N} \xi_k = m | N = 1]\, \mathbb{P}[N = 1] \\ &\quad + \mathbb{P}[\sum_{k=0}^{N} \xi_k = m | N = 2] \mathbb{P}[N = 2] \\ &= \frac{1}{3} \left(\mathbb{P}[0 = m] + \mathbb{P}[\xi_1 = m] + \mathbb{P}[\xi_1 + \xi_2 = m] \right). \end{aligned}$$

When $m = 1$ (for example), we get

$$\mathbb{P}[Y = 1] = \frac{0 + \frac{1}{2} + 0}{3} = 1/6.$$

Perform the computation for some other values of m for yourself.

What happens when N and $\{\xi_k\}_{k \in \mathbb{N}}$ are dependent? This will usually be the case in practice, as the value of the time N when we stop adding increments will typically depend on the behaviour of the sum itself.

Example 4.3.2. Let $\{\xi_k\}_{k\in\mathbb{N}}$ be as in Example 4.3.1 above - we can think of a situation where a gambler is repeatedly playing the same game in which a coin is tossed and the gambler wins a dollar if the outcome is *heads* and loses a dollar otherwise. A "smart" gambler enters the game and decides on the following tactic: *Let's see how the first game goes. If I lose, I'll play another 2 games and hopefully cover my losses, and If I win, I'll quit then and there.* The described strategy amounts to the choice of the random time N as follows:

$$N = \begin{cases} 1, & \xi_1 = 1, \\ 3, & \xi_1 = -1. \end{cases}$$

Then

$$Y = \begin{cases} 1, & \xi_1 = 1, \\ -1 + \xi_2 + \xi_3, & \xi_1 = -1. \end{cases}$$

Therefore,

$$\begin{aligned} \mathbb{P}[Y = 1] &= \mathbb{P}[Y = 1|\xi_1 = 1]\mathbb{P}[\xi_1 = 1] + \mathbb{P}[Y = 1|\xi_1 = -1]\mathbb{P}[\xi_1 = -1] \\ &= 1 \cdot \mathbb{P}[\xi_1 = 1] + \mathbb{P}[\xi_2 + \xi_3 = 2]\mathbb{P}[\xi_1 = -1] \\ &= \tfrac{1}{2}(1 + \tfrac{1}{4}) = \tfrac{5}{8}. \end{aligned}$$

Similarly, we get $\mathbb{P}[Y = -1] = \frac{1}{4}$ and $\mathbb{P}[Y = -3] = \frac{1}{8}$. The expectation $\mathbb{E}[Y]$ is equal to $1 \cdot \frac{5}{8} + (-1) \cdot \frac{1}{4} + (-3) \cdot \frac{1}{8} = 0$. This is not an accident. One of the first powerful results of the beautiful *martingale theory* states that no matter how smart a strategy you employ, you cannot beat a fair gamble.

We will return to the general (non-independent) case in the next lecture. Let us use generating functions to give a full description of the distribution of $Y = \sum_{k=1}^{N} \xi_k$ in this case.

Proposition 4.3.3. *Let $\{\xi_n\}_{n\in\mathbb{N}}$ be a sequence of independent $\mathbb{N}_0$-valued random variables, all of which share the same distribution with pmf $\{p_k\}_{k\in\mathbb{N}_0}$ and generating function $P_\xi(s)$. Let N be a random time independent of $\{\xi_n\}_{n\in\mathbb{N}}$. Then the generating function P_Y of the random sum $Y = \sum_{k=1}^{N} \xi_k$ is given by*

$$P_Y(s) = P_N(P_\xi(s)).$$

Proof. (*) We use the idea from Example 4.3.1 and condition on possible values of N. We also use the following fact (Tonelli's theorem) without proof:

$$\text{If } a_{ij} \geq 0, \text{ for all } i, j, \text{ then } \sum_{j=0}^{\infty}\sum_{i=0}^{\infty} a_{ij} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} a_{ij}. \tag{4.3.1}$$

$$\begin{aligned}
P_Y(s) &= \sum_{k=0}^{\infty} s^k \mathbb{P}[Y = k] \\
&= \sum_{k=0}^{\infty} s^k \Big(\sum_{i=0}^{\infty} \mathbb{P}[Y = k | N = i] \mathbb{P}[N = i] \Big) \\
&= \sum_{k=0}^{\infty} s^k \Big(\sum_{i=0}^{\infty} \mathbb{P}[\sum_{j=0}^{i} \xi_j = k] \mathbb{P}[N = i] \Big) \qquad \text{(by independence)} \\
&= \sum_{i=0}^{\infty}\sum_{k=0}^{\infty} s^k \mathbb{P}[\sum_{j=0}^{i} \xi_j = k] \mathbb{P}[N = i] \qquad \text{(by Tonelli)} \\
&= \sum_{i=0}^{\infty} \mathbb{P}[N = i] \sum_{k=0}^{\infty} s^k \mathbb{P}[\sum_{j=0}^{i} \xi_j = k] \qquad \text{(by (4.3.1))}
\end{aligned}$$

By (iteration of) Corollary 4.2.4, we know that the generating function of the random variable $\sum_{j=0}^{i} \xi_j$ - which is exactly what the second sum above represents - is $(P_\xi(s))^i$. Therefore, the chain of equalities above can be continued as

$$= \sum_{i=0}^{\infty} \mathbb{P}[N = i] (P_\xi(s))^i = P_N(P_\xi(s)). \qquad \square$$

Corollary 4.3.4 (Wald's Identity I). *Let $\{\xi_n\}_{n\in\mathbb{N}}$ and N be as in Proposition 4.3.3. Suppose, also, that $\mathbb{E}[N] < \infty$ and $\mathbb{E}[\xi_1] < \infty$. Then*

$$\mathbb{E}[\sum_{k=1}^{N} \xi_k] = \mathbb{E}[N]\,\mathbb{E}[\xi_1].$$

Proof. We just apply the composition rule for derivatives to the equality $P_Y = P_N \circ P_\xi$ to get

$$P_Y'(s) = P_N'(P_\xi(s)) P_\xi'(s).$$

After we let $s \nearrow 1$, we get

$$\mathbb{E}[Y] = P_Y'(1) = P_N'(P_\xi(1)) P_\xi'(1) = P_N'(1) P_\xi'(1) = \mathbb{E}[N]\,\mathbb{E}[\xi_1]. \qquad \square$$

Example 4.3.5. Every time Springfield Wildcats play in the Superbowl, their chance of winning is $p \in (0, 1)$. The number of years between two Superbowls they get to play in has the Poisson distribution $P(\lambda)$, $\lambda >$

0. What is the expected number of years Y between two consecutive Superbowl wins?

Let $\{\xi_k\}_{k\in\mathbb{N}}$ be the sequence of independent $P(\lambda)$-random variables modeling the number of years between two consecutive Superbowl appearances by the Wildcats. Moreover, let $\tilde{N}$ be a geometric $g(p)$ random variable with success probability p, and let $N = \tilde{N} + 1$ be its "shifted-by-one" version[a]. Every time the Wildcats lose the Superbowl, another $\xi_\cdot$ years have to pass before they get another chance and the whole thing stops when they finally win. Therefore,

$$Y = \sum_{k=1}^{N} \xi_k.$$

To compute the expectation of Y we use Corollary 4.3.4

$$\mathbb{E}[Y] = \mathbb{E}[N]\,\mathbb{E}[\xi_k] = \frac{\lambda}{p}.$$

[a] this is one of the examples where the version of the geometric that starts from 1, and not from 0, is better suited. Indeed, you cannot win two Superbowls in the same year.

Problems

Problem 4.3.1. If $P(s)$ is the generating function of the random variable X, then the generating function of $2X+1$ is

(a) $P(2s+1)$

(b) $2P(s)+1$

(c) $P(s^2+1)$

(d) $sP(s^2)$

(e) none of the above

Problem 4.3.2. Let $(p_0, p_1, p_2, \dots)$ be a sequence, and let $P(s)$ be its generating function. Then $(1-s)P(s^2)$ is the generating function of the sequence:

(a) $(p_0, p_0 + p_2, p_0 + p_2 + p_4, p_0 + p_2 + p_4 + p_6, \dots)$

(b) $(p_0, p_0 + p_1, p_0 + p_1 + p_2, p_0 + p_1 + p_2, \dots)$

(c) $(p_0, -p_0, p_1, -p_1, p_2, -p_2, p_3, -p_3, \dots)$

(d) $(p_0, p_1 - p_0, p_2 - p_1, p_3 - p_2, \dots)$

(e) none of the above

Problem 4.3.3. Let X be a random variable with the generating function P_X. The generating function of the random variable X^2 is

(a) $2P_X(s)$

(b) $P_X(2s)$

(c) $P_X(s^2)$

(d) $P_X(s)^2$

(e) none of the above

Problem 4.3.4. Let X be a random variable whose generating function P_X is given by

$$P_X(s) = \tfrac{1}{2}(1+s)/(2-s)$$

Compute the following:

1. $\mathbb{E}[X]$, $\mathbb{E}[X^2]$ and $\mathbb{E}[X^3]$.
2. $\mathbb{P}[X > 2]$.
3. The generating function of the random variable $Y = 3X + 2$
4. The generating function of the random variable Z obtained as follows. A coin is tossed and the value of X is drawn (independently). If the outcome of the coin is H, we set $Z = X$. Otherwise, $Z = 2X$.

Problem 4.3.5. 1. Use generating functions to compute the probability that the sum on two independent fair dice is 9.

2. Determine the distribution of the sum of two independent Poisson random variables with parameters $\lambda_1 > 0$ and λ_2.
3. Determine the distribution of the sum of two independent geometric random variables with (the same) parameter $p > 0$.

Problem 4.3.6. Let X and Y be two N_0-valued random variables, let $P_X(s)$ and $P_Y(s)$ be their generating functions and let $Z = X - Y$, $V = X + Y$, $W = XY$. Then

(a) $P_X(s) = P_Z(s)P_Y(s)$

(b) $P_X(s)P_Y(s) = P_Z(s)$

(c) $P_W(s) = P_X(P_Y(s))$,

(d) $P_Z(s)P_V(s) = P_X(s)P_Y(s)$

(e) none of the above.

Problem 4.3.7. Let X be an $\mathbb{N}_0$-valued random variable and $P(s)$ its generating function. If $Q(s) = P(s)/(1-s)$, then

(a) $Q(s)$ is a generating function of a random variable,

(b) $Q(s)$ is a generating function of a non-decreasing sequence of non-negative numbers,

(c) $Q(s)$ is a concave function on $(0,1)$,

(d) $Q(0) = 1$,

(e) none of the above.

Problem 4.3.8. The generating function of the $\mathbb{N}_0$-valued random variable X is given by

$$P_X(s) = \frac{s}{1+\sqrt{1-s^2}}.$$

1. Compute $p_0 = \mathbb{P}[X=0]$.
2. Compute $p_1 = \mathbb{P}[X=1]$.
3. Does $\mathbb{E}[X]$ exist? If so, find its value; if not, explain why not.

Problem 4.3.9. Let $P(s)$ be the generating function of the sequence $(p_0, p_1, \dots)$ and $Q(s)$ the generating function of the sequence $(q_0, q_1, \dots)$. If the sequence $\{r_n\}_{n\in\mathbb{N}_0}$ is defined by

$$r_n = \begin{cases} 0, & n \leq 1 \\ \sum_{k=1}^{n-1} p_k q_{n-1-k}, & n > 1, \end{cases}$$

then its generating function is given by (*Note:* Careful! $\{r_n\}_{n\in\mathbb{N}_0}$ is not exactly the convolution of $\{p_n\}_{n\in\mathbb{N}_0}$ and $\{q_n\}_{n\in\mathbb{N}_0}$.)

(a) $P(s)Q(s) - p_0q_0$

(b) $(P(s) - p_0)(Q(s) - q_0)$

(c) $\frac{1}{s}(P(s) - p_0)Q(s)$

(d) $\frac{1}{s}P(s)(Q(s) - q_0)$

(e) $s(P(s) - p_0)Q(s)$

Problem 4.3.10. A fair coin and a fair 6-sided die are thrown repeatedly until the the first time 6 appears on the die. Let X be the number of *heads* obtained (we are including the *heads* that may have occurred together with the first 6) in the count. Find the generating function of X.

Problem 4.3.11. Let N be geometrically distributed with parameter $p = \frac{1}{2}$, and let $\{\xi_n\}_{n\in\mathbb{N}}$ be iid with

$$\xi_1 \sim \begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{array}.$$

Then the generating function of the random sum $Y = \sum_{k=0}^{N} \xi_k$ is

(a) $(\frac{7}{4} - \frac{s}{2} - \frac{s^2}{4})^{-1}$

(b) $\frac{1}{4} + \frac{s}{2} + \frac{s^2}{4}$

(c) $\frac{(s-3)^2}{4(s-2)^2}$

(d) $\frac{\frac{1}{4}+\frac{s}{2}+\frac{s^2}{4}}{\frac{1}{2}-\frac{1}{2}s}$

(e) none of the above

Problem 4.3.12. A fair coin is tossed 100 times, and the number of H (heads) is denoted by N_1.

1. After that, N_1 fair coins are tossed and the number of H is denoted by N_2. Compute $\mathbb{P}[N_2 = 1]$.
2. We continue by tossing N_2 fair coins, count the number of heads obtained and denote the result by N_3. Compute $\mathbb{P}[N_3 = 1]$.

Problem 4.3.13. Let N be a random time, independent of $\{\xi_k\}_{k\in\mathbb{N}}$, where $\{\xi_k\}_{k\in\mathbb{N}}$ is a sequence of mutually independent Bernoulli ($\{0,1\}$-valued) random variables with parameter $p_B \in (0,1)$. Suppose that N has a geometric distribution $g(p_g)$ with parameter $p_g \in (0,1)$. Compute the distribution of the random sum

$$Y = \sum_{k=1}^{N} \xi_k,$$

i.e., find $\mathbb{P}[Y = k]$, for $k \in \mathbb{N}_0$. (*Note:* You can think of Y as a binomial random variable $b(n,p)$ with "random n".)

Problem 4.3.14. Six fair gold coins are tossed, and the total number of *tails* is recorded; let's call this number N. Then, a set of three fair silver coins is tossed N times. Let X be the total number of times at least two *heads* are observed (among the N tosses of the set of silver coins).

(*Note:* A typical outcome of such a procedure would be the following: out of the six gold coins 4 were *tails* and 2 were *heads*. Therefore $N = 4$ and the 4 tosses of the set of three silver coins may look something like $\{HHT, THT, TTT, HTH\}$, so that $X = 2$ in this state of the world.)

Find the generating function *and* the pmf of X. You don't have to evaluate binomial coefficients.

Problem 4.3.15. Tony Soprano collects his cut from the local garbage management companies. During a typical day he can visit a geometrically distributed number of companies with parameter $p = 0.1$. According to many years' worth of statistics gathered by his consigliere Silvio Dante, the amount he collects from the i^{th} company is random with the following distribution

$$X_i \sim \begin{pmatrix} \$1000 & \$2000 & \$3000 \\ 0.2 & 0.4 & 0.4 \end{pmatrix}$$

The amounts collected from different companies are independent of each other, and of the number of companies visited.

1. Find the (generating function of) the distribution of the amount of money S that Tony will collect on a given day.
2. Compute $\mathbb{E}[S]$ and $\mathbb{P}[S > 0]$.

Problem 4.3.16. (*) The **Fibonacci sequence** $\{F_n\}_{n\in\mathbb{N}_0}$ is defined recursively by

$$F_0 = F_1 = 1, F_{n+2} = F_{n+1} + F_n, n \in \mathbb{N}_0.$$

1. Find the generating function $P(s) = \sum_{k=0}^{\infty} F_k s^k$ of the sequence $\{F_n\}_{n\in\mathbb{N}_0}$. (*Hint:* What is $(1 - s - s^2)P(s)$?)

2. Use P to derive an explicit expression for F_n, $n \in \mathbb{N}_0$. (*Hint:* use partial fractions)

Note: the purpose of this problem is to show that one can use generating functions to do other things, as well. Indeed $\{F_n\}_{n\in\mathbb{N}_0}$ is not a probability distribution, but the generating function techniques still apply.

Problem 4.3.17. (*) When two fair dice are thrown, and their sum is computed, different values come with different probabilities (getting a sum of 2 is less likely than getting the sum of 3, etc.). How about if we start with two *loaded* dice? Suppose that you can build a die with face probabilities of $p_1, \dots, p_6$, for any 6-tuple of positive numbers that sum to 1. Can you find "loadings" $p_1, \dots, p_6$ for the first die and $q_1, \dots, q_6$ for the second one, so that, when we toss them together, all possible sums (namely 2, 3, ... 12) have the same probability (namely $1/11$)?

(*Hint:* Argue that no matter how you pick $p_1, \dots, p_6$ and $q_1, \dots, q_6$, the product of their generating functions must be of the form $s^2F(s)$, where F has a real zero.)