Chapter 2

Evaluating Outcomes of a Random Process

An *random variable* is a very useful concept: It allows us to concentrate our attention to a numerical evaluation of a random process that we are interested in, and to compute properties of this evaluation, for example the expected value. When analyzing randomized algorithms, we are often interested in the running time, which can be a random variable, or in the expected quality (in case of an optimization problem). We define the basics for using random variables and computing expected values. An interesting special case are integer valued random variables, which we first discuss in general. Second, we see two types of integer valued random variables that are of special interest, geometrically and binomially distributed random variables.

2.1 Random Variables and Expected Values

A random variable is a mapping that assigns a value to every elementary event. We have already used random variables implicitly: the value of a die, the sum of multiple dice or the number of times that a coin flip shows heads.

Definition 2.1. *Let* $(\Omega, \mathbf{Pr})$ *be a discrete probability space. A* discrete random variable *is a mapping* $X : \Omega \to \mathbb{R}$ *that assigns a real number to every elementary event.*

We also already used random variables in the definition of events, for example that the sum of the values of two dice is eight. The event that a random variable X takes value a is usually abbreviated by $X = a$. The precise definition of this event is $\{\omega \in \Omega \mid X(\omega) = a\}$. We will also use this abbreviation, in particular we use

$$\mathbf{Pr}(X = a) = \mathbf{Pr}(\{\omega \in \Omega \mid X(\omega) = a\}).$$

Analogously, we use for $R \subset \mathbb{R}$:

$$\mathbf{Pr}(X \in R) = \mathbf{Pr}(\{\omega \in \Omega \mid X(\omega) \in R\}).$$

Example 2.2. *We have a look at three examples for random variables.*

1. *The sum of two dice, modeled by the probability space* $(\Omega, \mathbf{Pr})$ *with* $\Omega = \{1, \ldots, 6\}^2$ *and* $\mathbf{Pr}(\omega) = 1/36$ *for all* $\omega \in \Omega$*, is formally defined by* $X : \Omega \to \{1, \ldots, 12\}$ *with* $X(\omega) = \omega_1 + \omega_2$ *for all* $\omega = (\omega_1, \omega_2) \in \Omega$*. We observe that* $\mathbf{Pr}(X = 1) = 0$*,* $\mathbf{Pr}(X = 2) = 1/36$ *and* $\mathbf{Pr}(X = 7) = 1/6$*.*
2. *Let* $(\Omega, \mathbf{Pr})$ *be a probability space for modeling ten independent coin flips, i.e.* $\Omega = \{H, T\}^{10}$*,* $\mathbf{Pr}(\omega) = 1/1024$ *for all* $\omega \in \Omega$*. Let* $X : \Omega \to \{0, 1, 2, \ldots, 1024\}$ *be the number of times that the flip came up heads. We observe that* $\mathbf{Pr}(X = 0) = 1/1024$ *and* $\mathbf{Pr}(X = 1) = 10/1024$*. A little more consideration lets us observe that* $\mathbf{Pr}(X = 2) = 45/1024$*.*
3. *We model the following random experiment: A coin is tossed until the first time that we see tails. We have* $\Omega = \{T, HT, HHT, HHHT, \ldots\} = \{H^{i-1}T \mid i \geq 1\}$ *as a countably infinite discrete probability space. The probability of the elementary event to get* $i-1$ *times head and then tails for the first time is* $\mathbf{Pr}(H^{i-1}T) = 1/2^i$*. Observe that*
$$\sum_{i=1}^{\infty} \frac{1}{2^i} = -1 + \sum_{i=0}^{\infty} \frac{1}{2^i} = 2 - 1 = 1.$$
We define the random variable $X : \Omega \to \mathbb{N}$ *by setting* $X(H^{i-1}T) = i$ *for all* $H^{i-1}T \in \Omega$*. By definition, it is clear that* $\mathbf{Pr}(X = i) = 1/2^i$*. Assume that Alice and Bob bet whether the first tails will occur after an odd or even number of coin tosses. If the number is odd, then Alice wins i points, and if it is even, she loses i points; Bobs result is inverse. Thus Alice's gain is described by the random variable* $Y : \Omega \to \mathbb{N}$ *with*
$$Y(H^{i-1}T) = \begin{cases} i & \text{if } i \text{ is odd} \\ -i & \text{if } i \text{ is even} \end{cases},$$
and Bobs gain is described by $-Y$*.*

We call random variables independent if knowing the value of one random variable gives no information about the value of the other variable. Given Definition 1.7, it is straightforward to obtain the following definition.

Definition 2.3. *Let* $(\Omega, \mathbf{Pr})$ *be a discrete probability space, let* $X, Y : \Omega \to \mathbb{R}$ *be discrete random variables. We say that* X *and* Y *are* independent *if*
$$\mathbf{Pr}((X = x) \cap (Y = y)) = \mathbf{Pr}(X = x) \cdot \mathbf{Pr}(Y = y)$$
holds for all $x, y \in \mathbb{R}$*. A Family* $X_1, \ldots, X_\ell : \Omega \to \mathbb{R}$ *of discrete random variables is* independent *if*
$$\mathbf{Pr}(\bigcap_{i \in I}(X_i = x_i) = \prod_{i \in I}(X_i = x_i)$$
holds for every $I \subseteq \{1, \ldots, n\}$ *and all* $x_i \in \mathbb{R}, i \in I$*.*

Since Ω is finite or countably infinite, the image of X is also finite or countably infinite. We name the image $R(X)$. When $R(X)$ is finite, then the expected value is the (weighted) average of the values in $R(X)$. When $R(X)$ is countably infinite, we can think of the expected value in a similar way. However, in this case it is possible that the expected value diverges.

Definition 2.4. *Let* $(\Omega, \mathbf{Pr})$ *be a discrete probability space and let* $X : \Omega \to \mathbb{R}$ *be a discrete random variable. We say that the expected value of* X exists *iff the series*

$$\sum_{\omega \in \Omega} |X(\omega)| \cdot \mathbf{Pr}(\{\omega\}) = \sum_{x \in R(X)} |x| \cdot \mathbf{Pr}(X = x)$$

converges. If the expected value exists, then it is defined as

$$\mathbf{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbf{Pr}(\{\omega\}) = \sum_{x \in R(X)} x \cdot \mathbf{Pr}(X = x).$$

Example 2.5. *We continue Example 2.2.*

1. *When* X *is the sum of two independent dice, then its expected value is*

$$\mathbf{E}[X] = \sum_{x=2}^{12} x \cdot \mathbf{Pr}(X = x) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + \ldots + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} = 7.$$

2. *For the random variable* X *that models the number of heads in ten independent coin flips, computing the expected value is somewhat tedious. Intuitively, we expect to see heads half of the time. Indeed, it turns out that*

$$\mathbf{E}[X] = 0 \cdot \frac{1}{1024} + 1 \cdot \frac{10}{1024} + 2 \cdot \frac{45}{1024} + \ldots + 10 \cdot \frac{1}{1024} = 5.$$

3. *What is the expected value of Alice's gain? Does the expected value exist? We can write* $Y(H^{i-1}T)$ *as* $(-1)^{i-1} \cdot \frac{1}{2^i}$. *To decide if the expected value exists, we have to decide whether*

$$\sum_{i=1}^{\infty} |(-1)^{i-1} i| \cdot \frac{1}{2^i} = \sum_{i=1}^{\infty} \frac{i}{2^i}$$

converges. We might know that this series converges or we can apply a convergence test to the series $\sum_{i=1}^{\infty} a_i$ *with* $a_i = \frac{i}{2^i}$ *to prove it (for example, to apply the ratio test, observe that* $\frac{i+1}{2^{i+1}} / (\frac{i}{2^i}) = 1/2 + 1/(2i)$ *and thus* $\lim_{i \to \infty} |a_{i+1}/a_i| = 1/2 < 1$*), and obtain that the expected value exists. When we actually compute the expected value, it might be surprising that it is not zero. We can compute some intermediate terms to get an intuition that the alternating series*

$$\mathbf{E}[Y] = \sum_{i=1}^{\infty} (-1)^{i-1} i \cdot \frac{1}{2^i}$$

converges to a value above zero:

$$\begin{aligned} & a_1 = 1/2, a_2 = 2/4, a_3 = 3/8, a_4 = 4/16, a_5 = 5/17, a_6 = 6/18 \\ \Rightarrow\ & a_1 = 1/2, a_1 - a_2 = 0, a_1 - a_2 + a_3 = 3/8, a_1 - a_2 + a_3 - a_4 = 1/8, \\ & a_1 - a_2 + a_3 - a_4 + a_5 = 9/32, a_1 - a_2 + a_3 - a_4 + a_5 - a_6 = 3/16. \end{aligned}$$

In fact, the sum of the series is $2/9$! *Thus, Alice wins* $2/9$ *points on expectation, and Bob loses* $2/9$ *points.*

Assume that we have two random variables $X, Y : \Omega \to \mathbb{R}$ in $(\Omega, \mathbf{Pr}())$. By stating $X \le Y$ we mean that $X(\{\omega\}) \le Y(\{\omega\})$ for all $\omega \in \Omega$.

Observation 2.6. *If $X, Y : \Omega \to \mathbb{R}$ satisfy $X \le Y$, then $\mathbf{E}[X] \le \mathbf{E}[Y]$.*

Proof. The observation follows by the definition of the expected value since

$$\mathbf{E}[X] = \sum_{\omega\in\Omega} X(\omega) \cdot \mathbf{Pr}(\{\omega\}) \le \sum_{\omega\in\Omega} Y(\omega) \cdot \mathbf{Pr}(\{\omega\}) = \mathbf{E}[Y]. \qquad \square$$

Computations with the expected value often become a lot easier when we use the fact that the expected value is linear. We show this and more important rules for the expected value below. Intuitively, we know what the sum or product of a random variable is, or how we multiply constants or add constants. Formally, we define three types of composed random variables. Let $X, Y : \Omega \to \mathbb{R}$ be two random variables in a discrete probability space $(\Omega, \mathbf{Pr})$.

- For all $c\mathbb{R}$, the random variable $(cX) : \Omega \to \mathbb{R}$ is defined by $(cX)(\omega) = c \cdot X(\omega)$ for all $\omega \in \Omega$.
- The random variable $X + Y : \Omega \to \mathbb{R}$ is defined by $(X + Y)(\omega) = X(\omega) + Y(\omega)$ for all $\omega \in \Omega$.
- The random variable $X \cdot Y : \Omega \to \mathbb{R}$ is defined by $(X \cdot Y)(\omega) = X(\omega) \cdot Y(\omega)$ for all $\omega \in \Omega$.

In all three cases, we usually omit the brackets for a shorter notation when possible.

Theorem 2.7. *Let $(\Omega, \mathbf{Pr})$ be a discrete probability space, let $X, Y : \Omega \to \mathbb{R}$ be discrete random variables. If the expected values of X and Y exist, then it also holds that:*

1. *The expected value of cX exists for all $c \in \mathbb{R}$ and $\mathbf{E}[cX] = c\mathbf{E}[X]$.*
2. *The expected value of $X + Y$ exists and $\mathbf{E}[X + Y] = \mathbf{E}[X] + \mathbf{E}[Y]$.*
3. *If X and Y are independent, then the expected value of $X \cdot Y$ exists and $\mathbf{E}[X \cdot Y] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$.*

Proof. Statements a) and b) are proven similarly, we prove a) as an example. First we observe that by the definition of cX, we have

$$\begin{aligned}\sum_{\omega\in\Omega} |(cX)(\omega)| \cdot \mathbf{Pr}(\{\omega\}) &= \sum_{\omega\in\Omega} |c \cdot X(\omega)| \cdot \mathbf{Pr}(\{\omega\}) \\ &\le \sum_{\omega\in\Omega} (|c| \cdot |X(\omega)|) \cdot \mathbf{Pr}(\{\omega\}) \\ &= |c| \cdot \sum_{\omega\in\Omega} |X(\omega)| \cdot \mathbf{Pr}(\{\omega\}).\end{aligned}$$

The expected value of X exists, so $\sum_{\omega\in\Omega} |X(\omega)| \cdot \mathbf{Pr}(\{\omega\})$ converges. Since c is a constant, this means that $\sum_{\omega\in\Omega} |(cX)(\omega)| \cdot \mathbf{Pr}(\{\omega\})$ converges as well, so the expected value of cX exists. We can compute it in a similar fashion and see that

$$\sum_{\omega\in\Omega} (cX)(\omega) \cdot \mathbf{Pr}(\{\omega\}) = c \cdot \sum_{\omega\in\Omega} X(\omega) \cdot \mathbf{Pr}(\{\omega\}) = c\mathbf{E}[X].$$

For statement c), we apply the definition of $X \cdot Y$ and the fact that X and Y are independent to obtain that

$$\begin{aligned}\sum_{\omega\in\Omega} |(X \cdot Y)(\omega)| \cdot \mathbf{Pr}(\{\omega\}) &= \sum_{\omega\in\Omega} |X(\omega) \cdot Y(\omega)| \cdot \mathbf{Pr}(\{\omega\}) \\ &= \sum_{x\in R(X)} \sum_{y\in R(Y)} |xy| \cdot \mathbf{Pr}((X = x) \cap (Y = y)) \\ &\le \sum_{x\in R(X)} \sum_{y\in R(Y)} |x| \cdot |y| \cdot \mathbf{Pr}((X = x)) \cdot \mathbf{Pr}((Y = y)) \\ &= \sum_{x\in R(X)} |x| \cdot \mathbf{Pr}((X = x)) \sum_{y\in R(Y)} |y| \cdot \mathbf{Pr}((Y = y)).\end{aligned}$$

We know that the expected value of Y exists, thus $\sum_{y\in R(Y)} |y| \cdot \mathbf{Pr}((Y = y))$ converges to a constant g. Thus,

$$\begin{aligned}\sum_{\omega\in\Omega} |(X \cdot Y)(\omega)| \cdot \mathbf{Pr}(\{\omega\}) &\le \sum_{x\in R(X)} |x|\mathbf{Pr}((X = x)) \cdot g \\ &= g \sum_{x\in R(X)} |x|\mathbf{Pr}((X = x)).\end{aligned}$$

The expected value of X exists, so $\sum_{x\in R(X)} |x|\mathbf{Pr}((X = x))$ converges. Thus, $\sum_{\omega\in\Omega} |(X\cdot Y)(\omega)| \cdot \mathbf{Pr}(\{\omega\})$ converges and the expected value of $X \cdot Y$ exists. We can compute it in a similar fashion and obtain that $\mathbf{E}[X \cdot Y] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$. Notice that we crucially needed the fact that X and Y are independent. □

Statement b) from Theorem 2.7 says that the expected value is additive, the rule by itself is often referred to as *linearity of expectation*. Statement a) says that the expected value is also homogeneous , so the expected value is indeed a linear function. Notice that rule b) holds for all random variables over the same probability space, even for dependent random variables! Statement c) is very useful when random variables are independent. Observe that for dependent random variables X and Y, it is possible that $\mathbf{E}[X]$ and $\mathbf{E}[Y]$ exist, but $\mathbf{E}[X \cdot Y]$ does not exist.

Example 2.8. *Let us construct an example where $\mathbf{E}[X]$ exists, but $\mathbf{E}[X \cdot X]$ does not exist. We need a probability space with a countably infinite set Ω and choose $\Omega = \mathbb{N}$. Our random experiment will be to draw a random number from $\mathbb{N}$. Over the infinite set of all positive integers, we cannot choose uniformly at random. If we choose i with probability $1/(2i)$, then we know that this defines a probability measure (this space naturally arises in our coin flip examples). An alternative is to pick i with a probability proportional to $1/i^4$. In the following, we will use some facts from mathematical analysis without worrying about how to obtain them. The series*

$$\sum_{i=1}^{\infty} \frac{1}{i^4}$$

converges to $\pi^4/90$, so in order to get a probability measure, we choose number i with probability $90/(\pi^4 i^4)$, because then we have $\mathbf{Pr}(\Omega) = \sum_{i=1}^{\infty} \frac{90}{\pi^4 i^4} = 1$. Now consider the random variable $X(i) = i^2$. The expected value of $X(i)$ exists because

$$\mathbf{Pr}(\Omega) = \sum_{i=1}^{\infty} \frac{90 i^2}{\pi^4 i^4} = \frac{90}{\pi^4} \sum_{i=1}^{\infty} \frac{1}{i^2}$$

converges (to $15/\pi^2$*). However, the expected random variable* $X \cdot X$ *does not exist because the series*

$$\mathbf{Pr}(\Omega) = \sum_{i=1}^{\infty} \frac{90i^4}{\pi^4 i^4} = \mathbf{Pr}(\Omega) = \sum_{i=1}^{\infty} \frac{90}{\pi^4}$$

obviously diverges.