2.1.1 Non-negative Integer Valued Random Variables

We are often interested whether a specific event occurs or not. Is my algorithm successful? Did the die roll give a six? It can be very useful to express this by an *indicator random variable*. Such a variable can only take the values 0 and 1 depending on whether the event in question occurred or not – it 'indicates' whether the event happened. For example, we could use a random variable that is 0 if one run of the algorithm at hand fails, and 1 if it succeeds. We will often define indicator random variables in the following shorter notation instead of referring to the actual elementary events:

$$X = \begin{cases} 1 & \text{if event } A \text{ occurrs} \\ 0 & \text{else} \end{cases} .$$

We observe that the expected value of an indicator variable X is equal to the probability that we get the value 1:

$$\mathbf{E}[X] = 0 \cdot \mathbf{Pr}(X = 0) + 1 \cdot \mathbf{Pr}(X = 1) = \mathbf{Pr}(X = 1). \tag{2.1}$$

A random variable X with values 0 and 1 and $p = \mathbf{Pr}(X = 1)$ is also called a *Bernoulli random variable with parameter p*.

Example 2.9. *We continue 2. from Example 2.5 where our random experiment consists of ten independent coin flips. We define the indicator variable X_i*

$$X_i = \begin{cases} 1 & \textit{if the } i\textit{th coin flip comes up heads} \\ 0 & \textit{else} \end{cases} .$$

for every $i \in \{1, \dots, 10\}$. We immediately observe that $\mathbf{E}[X_i] = \mathbf{Pr}(X_i = 1) = 1/2$. Before, we considered the random variable X which is the number of heads during the ten coin flips. This is just the sum of the indicator variables. By using linearity of expectation, the computation of the expected value of X now becomes a much easier task:

$$\mathbf{E}[X] = \mathbf{E}\left[\sum_{i=1}^{10} X_i\right] = \sum_{i=1}^{10} \mathbf{E}[X_i] = \sum_{1=1}^{10} \frac{1}{2} = 5.$$

For random variables that map to non-negative integers $\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$, there is a useful generalization of (2.1):

Lemma 2.10. *Let $(\Omega, \mathbf{Pr})$ be a discrete probability space and let $X : \Omega \to \mathbb{N}_0$ be a discrete random variable. Assume that $\mathbf{E}[X]$ exists. Then it holds that*

$$\mathbf{E}[X] = \sum_{i=1}^{\infty} \mathbf{Pr}(X \geq i).$$

Proof. The statement is true because

$$\mathbf{E}[X] = \sum_{i=0}^{\infty} i \cdot \mathbf{Pr}(X = i) = \sum_{i=1}^{\infty} i \cdot \mathbf{Pr}(X = i)$$

$$= \sum_{i=1}^{\infty} \sum_{j=1}^{i} \mathbf{Pr}(X = i).$$

Now we use that the existence of $\mathbf{E}[X]$ means that all sums in the equation converge and even converge absolutely. For absolutely convergent series, we can reorder the terms arbitrarily. The term $\mathbf{Pr}(X = 1)$ appears one time, the term $\mathbf{Pr}(X = 2)$ twice, the term $\mathbf{Pr}(X = 3)$ three times, and so on. We get the same terms in different order for

$$\sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \mathbf{Pr}(X = i) = \sum_{i=1}^{\infty} \mathbf{Pr}(X \geq i).$$

□

2.1.2 Conditional Expected Values

We know from Example 2.5 that the expected value of the sum of two dice is 7. Assume that we observe the outcome of the first die roll. Then the expected value of the sum will change. We already know the concept of conditional probabilities and now define conditional expectations.

Definition 2.11. *Let* $(\Omega, \mathbf{Pr})$ *be a discrete probability space, let* $A \in 2^{\Omega}$ *be an event and let* $X : \Omega \to \mathbb{R}$ *be a discrete random variable. The* conditional expected value of X under A exists *if* $\sum_{x \in R(X)} |x| \cdot \mathbf{Pr}(X = x \mid A)$ *converges and is then defined as*

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

In particular, if $Y : \Omega \to \mathbb{R}$ *is another discrete random variable and* $\mathbf{E}[X \mid Y = y]$ *for* $y \in \mathbb{R}$ *exists, then we get:*

$$\mathbf{E}[X \mid Y = y] = \sum_{x \in R(X)} x \cdot \mathbf{Pr}(X = x \mid Y = y).$$

Example 2.12. *We continue 1. from Example 2.5. Let* A *be the event that the first die roll is* 2, *let* X *be the random variable for the sum of the two dice. Then* $\mathbf{E}[X \mid A] = \sum_{x=3}^{8} x \cdot \frac{1}{6} = \frac{11}{2}$. *We can also say that* X_1 *is the random variable for the first roll,* X_2 *for the second roll, and then observe that*

$$\mathbf{E}[X \mid X_1 = 2] = \mathbf{E}[X_1 + X_2 \mid X_1 = 2] = \sum_{x=3}^{8} x \cdot \frac{1}{6} = \frac{11}{2}.$$

We can also use information about X *to obtain information about the first roll:*

$$\begin{aligned} \mathbf{E}[X_1 \mid X = 5] &= \sum_{x=1}^{4} x \cdot \mathbf{Pr}(X_1 = x \mid X = 5) \\ &= \sum_{x=1}^{4} x \cdot \frac{\mathbf{Pr}(X_1 = x \cap X = 5)}{\mathbf{Pr}(X = 5)} \end{aligned}$$

$$= \sum_{x=1}^{4} x \cdot \frac{\mathbf{Pr}((X_1 = x) \cap (X_2 = 5 - x))}{\mathbf{Pr}(X = 5)}$$
$$= \sum_{x=1}^{4} x \cdot \frac{1/36}{4/36} = \frac{1}{4} \cdot 10 = \frac{5}{2}$$

We do not want to miss the following rules when computing conditional expected values. They are proven similarly to Theorem 2.7.

Lemma 2.13. *Let $(\Omega, \mathbf{Pr})$ be a discrete probability space, let $X, Y : \Omega \to \mathbb{R}$ be discrete random variables and let $A \in 2^{\Omega}$ be an event. If the expected values of X and Y exist, then it also holds that:*

1. *The conditional expected value of cX under A exists for all $c \in \mathbb{R}$ and it is $\mathbf{E}[cX \mid A] = c\mathbf{E}[X \mid A]$.*
2. *The conditional expected value of $X + Y$ under A exists and $\mathbf{E}[X + Y \mid A] = \mathbf{E}[X \mid A] + \mathbf{E}[Y \mid A]$.*

Also observe that if X and Y are independent, then

$$\mathbf{E}[X \mid Y = y] = \sum_{x \in R(X)} x \cdot \mathbf{Pr}(X = x \mid Y = y) = \sum_{x \in R(X)} x \cdot \frac{\mathbf{Pr}((X = x) \cap (Y = y))}{\mathbf{Pr}((Y = y))}$$
$$= \sum_{x \in R(X)} x \cdot \frac{\mathbf{Pr}((X = x)) \cdot \mathbf{Pr}((Y = y))}{\mathbf{Pr}((Y = y))} = \sum_{x \in R(X)} x \cdot \mathbf{Pr}(X = x) = \mathbf{E}[X].$$

Example 2.14. *We continue 2. from Example 2.5. Let Y be the random variable that is equal to the number of heads we see in the first six of ten independent coin flips, let Z be the random variable that is equal to the number of heads we see in the last four of ten independent coin flips and let $X = Y + Z$ be the number of heads in all ten coin flips. Then $\mathbf{E}[X \mid Y = 4] = \mathbf{E}[Y + Z \mid Y = 4] = \mathbf{E}[Y \mid Y = 4] + \mathbf{E}[Z \mid Y = 4] = 4 + 2 = 6$.*