writes $b \geq a$. The strict linear order $<$ is defined by $a < b$ if and only if $a \leq b$ and $a \neq b$. The relation $<$ is transitive, irreflexive ($a < b$ implies $a \neq b$), and total in the sense that for all a and b either $a < b$ or $a = b$ or $a > b$. A typical example is the relation $\leq$ for real numbers.

linear preorder: (also linear quasi-order) a reflexive, transitive, and total relation. The symbols $\leq$ and $\geq$ are also used for linear preorders. Note that there can be distinct elements a and b with $a \leq b$ and $b \leq a$. The strict variant $<$ is defined as $a < b$ if $a \leq b$ and not $a \geq b$. An example is the relation $\mathrm{R} \subseteq \mathbb{R} \times \mathbb{R}$ defined by $x \, \mathrm{R} \, y$ if and only if $|x| \leq |y|$.

median: an element with rank $\lceil n/2 \rceil$ among n elements.

multiplicative inverse: if an object x is multiplied by a *multiplicative inverse* x^{-1} of x, we obtain $x \cdot x^{-1} = 1$ – the neutral element of multiplication. In particular, in a *field*, every element except zero (the neutral element of addition) has a unique multiplicative inverse.

prime number: an integer n, $n \geq 2$, is a prime iff there are no integers $a, b > 1$ such that $n = a \cdot b$.

rank: Let $\leq$ be a linear preorder on a set $S = \{e_1, \ldots, e_n\}$. A one-to-one mapping $r : S \to 1..n$ is a *ranking function* for the elements of S if $r(e_i) < r(e_j)$ whenever $e_i < e_j$. If $\leq$ is a linear order, there is exactly one ranking function.

reflexive: a relation $\mathrm{R} \subseteq A \times A$ is reflexive if $a \, \mathrm{R} \, a$ for all $a \in A$.

relation: a set of ordered pairs R over some set A. Often, we write relations as infix operators; for example, if $\mathrm{R} \subseteq A \times A$ is a relation, $a \, \mathrm{R} \, b$ means $(a, b) \in \mathrm{R}$.

symmetric relation: a relation $R \subseteq A \times A$ is *symmetric* if for all a and b in A, $a \, \mathrm{R} \, b$ implies $b \, \mathrm{R} \, a$.

total relation: a relation $\mathrm{R} \subseteq A \times A$ is *total* if for all a and b in A, either $a \, \mathrm{R} \, b$ or $b \, \mathrm{R} \, a$ or both. If a relation R is total and transitive, then the relation $\sim_{\mathrm{R}}$ defined by $a \sim_{\mathrm{R}} b$ if and only if $a \, \mathrm{R} \, b$ and $b \, \mathrm{R} \, a$ is an equivalence relation.

total order: a synonym for linear order.

transitive: a relation $\mathrm{R} \subseteq A \times A$ is *transitive* if for all a, b, and c in A, $a \, \mathrm{R} \, b$ and $b \, \mathrm{R} \, c$ imply $a \, \mathrm{R} \, c$.

1.3 Basic Probability Theory

Probability theory rests on the concept of a *sample space* $\mathscr{S}$. For example, to describe the rolls of two dice, we would use the 36-element sample space $\{1, \ldots, 6\} \times \{1, \ldots, 6\}$, i.e., the elements of the sample space (also called elementary events or events) are the pairs (x,y) with $1 \le x,y \le 6$ and $x,y \in \mathbb{N}$. Generally, a sample space is any nonempty set. In this book, all sample spaces are finite.[1] In a *random experiment*, any element of $s \in \mathscr{S}$ is chosen with some elementary *probability* p_s, where $\sum_{s\in\mathscr{S}} p_s = 1$. The function that assigns to each event s its probability p_s is called a *distribution*. A sample space together with a probability distribution is called a *probability space*. In this book, we use *uniform distributions* almost exclusively; in this case $p_s = p = 1/|\mathscr{S}|$. Subsets $\mathscr{E}$ of the sample space are called *events*. The probability of an *event* $\mathscr{E} \subseteq \mathscr{S}$ is the sum of the probabilities of its elements, i.e., $\operatorname{prob}(\mathscr{E}) = |\mathscr{E}|/|\mathscr{S}|$ in the uniform case. So the probability of the event $\{(x,y) : x+y = 7\} = \{(1,6),(2,5),\ldots,(6,1)\}$ is equal to $6/36 = 1/6$, and the probability of the event $\{(x,y) : x+y \ge 8\}$ is equal to $15/36 = 5/12$.

A *random variable* is a mapping from the sample space to the real numbers. Random variables are usually denoted by capital letters to distinguish them from plain values. For our example of rolling two dice, the random variable X could give the number shown by the first die, the random variable Y could give the number shown by the second die, and the random variable S could give the sum of the two numbers. Formally, if $(x,y) \in \mathscr{S}$, then $X((x,y)) = x$, $Y((x,y)) = y$, and $S((x,y)) = x+y = X((x,y)) + Y((x,y))$.

We can define new random variables as expressions involving other random variables and ordinary values. For example, if V and W are random variables, then $(V+W)(s) = V(s) + W(s)$, $(V \cdot W)(s) = V(s) \cdot W(s)$, and $(V+3)(s) = V(s) + 3$.

Events are often specified by predicates involving random variables. For example, $X \le 2$ denotes the event $\{(1,y),(2,y) : 1 \le y \le 6\}$, and hence $\operatorname{prob}(X \le 2) = 1/3$. Similarly, $\operatorname{prob}(X+Y = 11) = \operatorname{prob}(\{(5,6),(6,5)\}) = 1/18$.

Indicator random variables are random variables that take only the values zero and one. Indicator variables are an extremely useful tool for the probabilistic analysis of algorithms because they allow us to encode the behavior of complex algorithms into simple mathematical objects. We frequently use the letters I and J for indicator variables. Indicator variables and events are in a one-to-one correspondance. If $\mathscr{E}$ is an event, then $I_{\mathscr{E}}$ with $I_{\mathscr{E}}(s) = 1$ iff $s \in \mathscr{E}$ is the corresponding indicator variable. If an event is specificed by a predicate P, one sometimes writes $[P]$ for the corresponding indicator variable, i.e, $[P](s) = 1$, if $P(s)$, and $[P](s) = 0$, otherwise.

The *expected value* of a random variable $Z : \mathscr{S} \to \mathbb{R}$ is

$$\mathrm{E}[Z] = \sum_{s\in\mathscr{S}} p_s \cdot Z(s) = \sum_{z\in\mathbb{R}} z \cdot \operatorname{prob}(Z = z) \ , \tag{1.1}$$

i.e., every sample s contributes the value of Z at s times its probability. Alternatively, we can group all s with $Z(s) = z$ into the event $Z = z$ and then sum over the $z \in \mathbb{R}$.

In our example, $\mathrm{E}[X] = (1+2+3+4+5+6)/6 = 21/6 = 3.5$, i.e., the expected value of the first die is 3.5. Of course, the expected value of the second die is also 3.5. For an indicator random variable I, we have

[1] All statements made in this section also hold for countable infinite sets, essentially with the same proofs. Such sample spaces are for example needed to model the experiment “throw a dice repeatedly until the value six occurs”.

$$\mathrm{E}[I] = 0 \cdot \mathrm{prob}(I = 0) + 1 \cdot \mathrm{prob}(I = 1) = \mathrm{prob}(I = 1)\ .$$

Sometimes, we are more interested in a random variable Z and its behavior than in the underlying probability space. In such a situation, it suffices to know the set $Z[\mathscr{S}]$ and the induced probabilities $\mathrm{prob}(Z = z)$, $z \in Z[\mathscr{S}]$. We refer to the function $z \mapsto \mathrm{prob}(Z = z)$ defined on $Z[\mathscr{S}]$ as the *distribution of* Z. Two random variables X and Y with the same distribution, are called *identically distributed*.

For a random variable Z that takes only values in the natural numbers, there is a very useful formula for its expected value:

$$\mathrm{E}[Z] = \sum_{k \geq 1} \mathrm{prob}(Z \geq k),\ \text{if } Z[\mathscr{S}] \subseteq \mathbb{N}.$$

The formula is easy to prove. For $k, i \in \mathbb{N}$, let $p_k = \mathrm{prob}(Z \geq k)$ and $q_i = \mathrm{prob}(Z = i)$. Then $p_k = \sum_{i \geq k} q_i$ and hence

$$\mathrm{E}[Z] = \sum_{z \in Z[\mathscr{S}]} z \cdot \mathrm{prob}(Z = z) = \sum_{i \in \mathbb{N}} i \cdot \mathrm{prob}(Z = i) = \sum_{i \in \mathbb{N}} \sum_{1 \leq k \leq i} q_i = \sum_{k \geq 1} \sum_{i \geq k} q_i = \sum_{k \geq 1} p_k.$$

Here the next to last equality is a change of order of summation.

Often, we are interested in the expectation of a random variable that is defined in terms of other random variables. This is particulary easy for sums of random variables: reason is the *linearity of expectations* of random variables: for any two random variables V and W,

$$\mathrm{E}[V + W] = \mathrm{E}[V] + \mathrm{E}[W]\ . \tag{1.2}$$

This equation is easy to prove and extremely useful. Let us prove it. It amounts essentially to an application of the distributive law of arithmetic. We have

$$\begin{aligned}\mathrm{E}[V + W] &= \sum_{s \in \mathscr{S}} p_s \cdot (V(s) + W(s)) \\ &= \sum_{s \in \mathscr{S}} p_s \cdot V(s) + \sum_{s \in \mathscr{S}} p_s \cdot W(s) \\ &= \mathrm{E}[V] + \mathrm{E}[W]\ .\end{aligned}$$

As our first application, let us compute the expected sum of two dice. We have

$$\mathrm{E}[S] = \mathrm{E}[X + Y] = \mathrm{E}[X] + \mathrm{E}[Y] = 3.5 + 3.5 = 7\ .$$

Observe that we obtain the result with almost no computation. Without knowing about the linearity of expectations, we would have to go through a tedious calculation:

$$\begin{aligned}\mathrm{E}[S] &= 2 \cdot \tfrac{1}{36} + 3 \cdot \tfrac{2}{36} + 4 \cdot \tfrac{3}{36} + 5 \cdot \tfrac{4}{36} + 6 \cdot \tfrac{5}{36} + 7 \cdot \tfrac{6}{36} + 8 \cdot \tfrac{5}{36} + 9 \cdot \tfrac{4}{36} + \ldots + 12 \cdot \tfrac{1}{36} \\ &= \frac{2 \cdot 1 + 3 \cdot 2 + 4 \cdot 3 + 5 \cdot 4 + 6 \cdot 5 + 7 \cdot 6 + 8 \cdot 5 + \ldots + 12 \cdot 1}{36} = 7\ .\end{aligned}$$

Exercise 1.1. What is the expected sum of three dice?

We shall now give another example with a more complex sample space. We consider the experiment of throwing n balls into m bins. The balls are thrown at random and distinct balls do not influence each other. Formally, our sample space is the set of all functions f from $1..n$ to $1..m$. This sample space has size m^n, and $f(i)$, $1 \leq i \leq n$, indicates the bin into which the ball i is thrown. All elements of the sample space are equally likely. How many balls should we expect in bin 1? We use W to denote the number of balls in bin 1. To determine $\mathrm{E}[W]$, we introduce indicator variables I_i, $1 \leq i \leq n$. The variable I_i is 1, if ball i is thrown into bin 1, and is 0 otherwise. Formally, $I_i(f) = 0$ iff $f(i) \neq 1$. Then $W = \sum_i I_i$. We have

$$\mathrm{E}[W] = \mathrm{E}[\sum_i I_i] = \sum_i \mathrm{E}[I_i] = \sum_i \mathrm{prob}(I_i = 1) \ ,$$

where the second equality is the linearity of expectations and the third equality follows from the I_i's being indicator variables. It remains to determine the probability that $I_i = 1$. Since the balls are thrown at random, ball i ends up in any bin[2] with the same probability. Thus $\mathrm{prob}(I_i = 1) = 1/m$, and hence

$$\mathrm{E}[I] = \sum_i \mathrm{prob}(I_i = 1) = \sum_i \frac{1}{m} = \frac{n}{m} \ .$$

Products of random variables behave differently. In general, we have $\mathrm{E}[X \cdot Y] \neq \mathrm{E}[X] \cdot \mathrm{E}[Y]$. There is one important exception: if X and Y are *independent*, equality holds. Random variables $X_1, \ldots, X_k$ are independent if and only if

$$\forall x_1, \ldots, x_k : \mathrm{prob}(X_1 = x_1 \wedge \cdots \wedge X_k = x_k) = \prod_{1 \leq i \leq k} \mathrm{prob}(X_i = x_i) \ . \tag{1.3}$$

As an example, when we roll two dice, the value of the first die and the value of the second die are independent random variables. However, the value of the first die and the sum of the two dice are not independent random variables.

Exercise 1.2. Let I and J be independent indicator variables and let $X = (I + J) \bmod 2$, i.e., X is one iff I and J are different. Show that I and X are independent, but that I, J, and X are dependent.

Assume now that X and Y are independent. Then

[2] Formally, there are exactly m^{n-1} functions f with $f(i) = 1$.

$$
\begin{aligned}
\mathrm{E}[X]\cdot\mathrm{E}[Y] &= \left(\sum_x x\cdot\operatorname{prob}(X=x)\right)\cdot\left(\sum_y y\cdot\operatorname{prob}(X=y)\right)\\
&= \sum_{x,y} x\cdot y\cdot\operatorname{prob}(X=x)\cdot\operatorname{prob}(X=y)\\
&= \sum_{x,y} x\cdot y\cdot\operatorname{prob}(X=x\wedge Y=y)\\
&= \sum_z \sum_{x,y \text{ with } z=x\cdot y} z\cdot\operatorname{prob}(X=x\wedge Y=y)\\
&= \sum_z z\cdot \sum_{x,y \text{ with } z=x\cdot y} \operatorname{prob}(X=x\wedge Y=y)\\
&= \sum_z z\cdot\operatorname{prob}(X\cdot Y=z)\\
&= \mathrm{E}[X\cdot Y]\ .
\end{aligned}
$$

How likely is it that a random variable will deviate substantially from its expected value? *Markov's inequality* gives a useful bound. Let X be a nonnegative random variable and let c be any constant. Then

$$\operatorname{prob}(X \ge c\cdot\mathrm{E}[X]) \le \frac{1}{c}\ . \tag{1.4}$$

The proof is simple. We have

$$
\begin{aligned}
\mathrm{E}[X] &= \sum_{z\in\mathbb{R}} z\cdot\operatorname{prob}(X=z)\\
&\ge \sum_{z\ge c\cdot\mathrm{E}[X]} z\cdot\operatorname{prob}(X=z)\\
&\ge c\cdot\mathrm{E}[X]\cdot\operatorname{prob}(X\ge c\cdot\mathrm{E}[X])\ ,
\end{aligned}
$$

where the first inequality follows from the fact that we sum over a subset of the possible values and X is nonnegative, and the second inequality follows from the fact that the sum in the second line ranges only over z such that $z \ge c\mathrm{E}[X]$.

Much tighter bounds are possible for some special cases of random variables. The following situation arises several times, in the book. We have a sum $X = X_1 + \cdots + X_n$ of n independent indicator random variables $X_1,\ldots, X_n$ and want to bound the probability that X deviates substantially from its expected value. In this situation, the following variant of the *Chernoff bound* is useful. For any $\varepsilon > 0$, we have

$$\operatorname{prob}(X < (1-\varepsilon)\mathrm{E}[X]) \le e^{-\varepsilon^2\mathrm{E}[X]/2}\ , \tag{1.5}$$

$$\operatorname{prob}(X > (1+\varepsilon)\mathrm{E}[X]) \le \left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{(1+\varepsilon)}}\right)^{\mathrm{E}[X]}\ . \tag{1.6}$$

A bound of the form above is called a *tail bound* because it estimates the “tail” of the probability distribution, i.e., the part for which X deviates considerably from its expected value.

Let us see an example. If we throw n coins and let X_i be the indicator variable for the i-th coin coming up heads, $X = X_1 + \cdots + X_n$ is the total number of heads. Clearly, $\mathrm{E}[X] = n/2$. The bound above tells us that $\mathrm{prob}(X \le (1-\varepsilon)n/2) \le e^{-\varepsilon^2 n/4}$. In particular, for $\varepsilon = 0.1$, we have $\mathrm{prob}(X \le 0.9 \cdot n/2) \le e^{-0.01 \cdot n/4}$. So, for $n = 10\,000$, the expected number of heads is 5 000 and the probability that the sum is less than 4 500 is smaller than e^{-25}, a very small number.

Exercise 1.3. Estimate the probability that X in the above example is larger than 5 050.

If the indicator random variables are independent and identically distributed with $\mathrm{prob}(X_i = 1) = p$, X is *binomially distributed*, i.e.,

$$\mathrm{prob}(X = k) = \binom{n}{k} p^k (1-p)^{(n-k)} \,. \tag{1.7}$$

Exercise 1.4 (balls and bins continued). Let, as above, W denote the number of balls in bin 1. Show

$$\mathrm{prob}(W = k) = \binom{n}{k} \left(\frac{1}{m}\right)^k \left(1 - \frac{1}{m}\right)^{(n-k)} \,,$$

and then attempt to compute $\mathrm{E}[W]$ as $\sum_k \mathrm{prob}(W = k)k$.

1.4 Useful Formulae

We shall first list some useful formulae and then prove some of them.

- A simple approximation to the factorial:

$$\left(\frac{n}{e}\right)^n \le n! \le n^n \text{ or, more precisely } e\left(\frac{n}{e}\right)^n \le n! \le (en)\left(\frac{n}{e}\right)^n \,. \tag{1.8}$$

- Stirling's approximation to the factorial:

$$n! = \left(1 + \mathrm{O}\left(\frac{1}{n}\right)\right) \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \,. \tag{1.9}$$

- An approximation to the binomial coefficients:

$$\binom{n}{k} \le \left(\frac{n \cdot e}{k}\right)^k \,. \tag{1.10}$$

- The sum of the first n integers:

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2} \,. \tag{1.11}$$