Here we will discuss some basics of probability theory that will enable us analyze some randomized algorithms. We start with some basic definitions. We let $\mathbb{N} = \{0, 1, 2, \ldots \}$ be the set of all natural numbers.

A pair $(\Omega, P)$ is called a discrete probability space if $\Omega$ is a countable set and $P : \Omega \to [0, 1]$ is a function such that $\sum_{\omega \in \Omega} P(\omega) = 1$.

We call the elements of $\Omega$ elementary events; for each $\omega \in \Omega$ we call $P(\omega)$ the probability of $\omega$. We call subsets of $\Omega$ events. For any $A \subseteq \Omega$ we let $P(A) = \sum_{a \in A} P(a)$; thus, we extended $P$ to arbitrary events.

As an example, consider $\Omega = \{1, 2, 3, 4, 5, 6\}$ and, for all $\omega \in \Omega$, $P(\omega) = 1/6$. This models rolling a die, where each outcome (1 through 6) has the same property of appearing. In general, when $\Omega$ is finite, we can consider the uniform distribution which assigns each elementary event a probability of $1/|\Omega|$.

As another example, consider $\Omega = \mathbb{N}$ and, for all $n \in \mathbb{N}$, $P(n) = 2^{-n-1}$. Note that $\sum_{n \in \mathbb{N}} 2^{-n-1} = 1$ (geometric sum).

For all events $A, B \subseteq \Omega$, we have the following laws for dealing with probabilities.

(a) If $A \subseteq B$, then $P(A) \leq P(B)$.

(b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

(c) $P(\Omega \setminus A) = 1 - P(A)$.

(d) For all sequences $(A_i)$ of events, $P(\bigcup_i A_i) \leq \sum_i P(A_i)$.

A random variable is a mapping $X : \Omega \to \mathbb{R}$. As an example, consider $\Omega = \{1, 2, 3, 4, 5, 6\}^2$ and $P$ as the uniform distribution on $\Omega$, the result of rolling two dice. Let $X$ be a random variable such that, for all $a, b \in \{1, 2, 3, 4, 5, 6\}$, $X(a, b) = a + b$. In other words, $X$ is the sum of the results of two dice rolls. We can now consider such events as “$X = 12$” (this is the event consisting of all $\omega \in \Omega$ such that $X(\omega) = 12$). As an exercise, how much is $P(X = 12)$? What about $P(X = 0)$? What is $P(X \geq 7)$?

For the same $(\Omega, P)$ based on rolling two dice, let $X$ and $Y$ be random variables such that, for all $a, b \in \{1, 2, 3, 4, 5, 6\}$, $X(a, b) = a$ and $Y(a, b) = b$. In other words, $X$ is the result of the first die and $Y$ of the second. We can now consider such events as “$X = Y$” (this is the event consisting of all $\omega \in \Omega$ such that $X(\omega) = Y(\omega)$). As an exercise, how much is $P(X = Y)$? What about $P(X = Y + 2)$? What is $P(X \geq Y)$?

We call two random variables $X, Y$ identically distributed if, for all $r \in \mathbb{R}$, $P(X = r) = P(Y = r)$. We then write $X \sim Y$. Note that the two random variables $X, Y$ just
above are identically distributed, but not identical (if they were identical, we would have \( P(X = Y) = 1 \)).

We call two random variables \( X, Y \) independent if for all sets \( A, B \subseteq \mathbb{R} \) we have

\[
P(X \in A \text{ and } Y \in B) = P(X \in A) \cdot P(Y \in B).
\]

Similarly, we call a sequence of random variables \((X_i)_i\) independent if for all sequences \((A_i)_i\) of subsets of real numbers we have

\[
P(\bigwedge_i X_i \in A_i) = \prod_i P(X_i \in A_i).
\]

We call two random variables independently identically distributed (i.i.d.) if they are identically distributed and independent. We extend this naturally to sequences of random variables.

The expected value of a random variable \( X \) is

\[
E(X) = \sum_{\omega \in \Omega} P(\omega) \cdot X(\omega).
\]

We note that

\[
E(X) = \sum_{\omega \in \Omega} P(\omega) \cdot X(\omega)
= \sum_{r \in \mathbb{R}} \sum_{\omega : X(\omega) = r} P(\omega) \cdot r
= \sum_{r \in \mathbb{R}} r \cdot P(X = r).
\]

Whenever \( X, Y \) are random variables, we define \( X + Y \) to be the random variable such that, for all \( \omega \in \Omega \), \((X + Y)(\omega) = X(\omega) + Y(\omega)\). Similarly we can define all kinds of other operations on random variables, for example, for \( r \in \mathbb{R} \), \( rX \) is the random variable such that \((rX)(\omega) = rX(\omega)\).

We have the following rules for working with random variables \( X, Y \) and \( r \in \mathbb{R} \).

(a) \( E(X + Y) = E(X) + E(Y) \);

(b) \( E(rX) = rE(X) \).

In other words, \( E \) is linear.

For any random variable \( X \) we let \( \text{Var}(X) = E((X - E(X))^2) \) be the variance of the random variable \( X \).
Some Theorems about Random Variables

**Theorem 1** Let $X, Y$ be independent random variables. We have $E(XY) = E(X)E(Y)$.

*Proof.* We have the following chain of equalities.

\[
E(XY) = \sum_{\omega \in \Omega} P(\omega)(XY)(\omega)
\]
\[
= \sum_{\omega \in \Omega} P(\omega)X(\omega)Y(\omega)
\]
\[
= \sum_{(a,b) \in \mathbb{R}} P(X = a, Y = b)ab
\]
\[
= \sum_{(a,b) \in \mathbb{R}} P(X = a)P(Y = b)ab
\]
\[
= \sum_{a \in \mathbb{R}} \sum_{b \in \mathbb{R}} (P(X = a)a)(P(Y = b)b)
\]
\[
= \sum_{a \in \mathbb{R}} \left( (P(X = a)a) \sum_{b \in \mathbb{R}} (P(Y = b)b) \right)
\]
\[
= \left( \sum_{a \in \mathbb{R}} (P(X = a)a) \right) \left( \sum_{b \in \mathbb{R}} P(Y = b)b \right)
\]
\[
= E(X)E(Y).
\]

This concludes the proof. \qed

**Theorem 2** Let $X$ be a random variable. We have $\text{Var}(X) = E(X^2) - E(X)^2$.

*Proof.* We have the following chain of equalities.

\[
\text{Var}(X) = E((X - E(X))^2)
\]
\[
= E(X^2 - 2XE(X) + E(X)^2)
\]
\[
= E(X^2) - 2XE(X) + E(X)^2
\]
\[
= E(X^2) - 2E(X)E(X) + E(X)^2
\]
\[
= E(X^2) - E(X)^2.
\]

This concludes the proof. \qed

**Theorem 3** Let $X, Y$ be independent random variables. We have $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.
Proof. We have the following chain of equalities.

\[
\begin{align*}
\text{Var}(X + Y) &= E((X + Y)^2) - E(X + Y)^2 \\
&= E(X^2 + 2XY + Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y)^2 \\
&= E(X^2) + 2E(X)E(Y) + E(Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y)^2 \\
&= E(X^2) - E(X)^2 + E(Y^2) - E(Y)^2 \\
&= \text{Var}(X) + \text{Var}(Y).
\end{align*}
\]

This concludes the proof. \qed

Theorem 4 (Markov’s Inequality) Let $X$ be a random variable with $P(X < 0) = 0$. For all $a > 0$ we have

\[
P(X \geq a) \leq \frac{E(X)}{a}.
\]

Proof. We have

\[
E(X) = \sum_{b \geq 0} bP(X = b)
\]

\[
= \sum_{0 \leq b < a} bP(X = b) + \sum_{b \geq a} bP(X = b)
\]

\[
\geq \sum_{0 \leq b < a} 0P(X = b) + \sum_{b \geq a} aP(X = b)
\]

\[
= a \sum_{b \geq a} P(X = b)
\]

\[
= aP(X \geq a).
\]

Dividing both sides by $a$ concludes the proof. \qed

Theorem 5 Let $X$ be a random variable which only takes values in the natural numbers. Then

\[
E(X) = \sum_{a=1}^{\infty} P(X \geq a).
\]
Proof. We have

\[ \sum_{a=1}^{\infty} P(X \geq a) = \sum_{a=1}^{\infty} \sum_{b=a}^{\infty} P(X = b) \]
\[ = \sum_{b=1}^{\infty} \sum_{a=1}^{b} P(X = b) \]
\[ = \sum_{b=1}^{\infty} bP(X = b) \]
\[ = E(X). \]

This concludes the proof. \(\square\)
Some Example Probability Distributions

We will need some typical probability distributions. The simplest distribution is the Bernoulli distribution. We say that a random variable \( X \) has Bernoulli distribution with parameter \( p \in [0, 1] \) if \( P(X = 1) = p \) and \( P(X = 0) = 1 - p \). Thus, the random variables takes on (at most) two values.

If we have \( n \) i.i.d. Bernoulli-distributed random variables \( (X_i)_{i \leq n} \) with parameter \( p \), then \( \sum_{i=1}^{n} X_i \) is a Binomial distribution with parameters \( n \) and \( p \). We write a Binomial distribution with parameters \( n \) and \( p \) as \( B(n, p) \). We have \( E(B(n, p)) = np \).

We say that a random variable \( X \) has geometric distribution with parameter \( p \in (0, 1] \) if, for all natural numbers \( k \),

\[
P(X = k) = (1 - p)^k p.
\]

We can imagine \( X \) as the number of times we need to be unsuccessful before being successful, if we are successful each time with probability \( p \). We have

\[
\sum_{k=0}^{\infty} (1 - p)^k p = p \sum_{k=0}^{\infty} (1 - p)^k = p \frac{1}{p} = 1.
\]

This uses the formula for geometric series. Note that, for all \( k \), \( P(X \geq k) = (1 - p)^k \). Thus, we can easily compute \( E(X) = 1/p \), using Theorem 5 (and the formula for geometric series).