typically the randomization will involve uniform sampling from a finite subset of the field; in such cases, we do not have to worry about whether the field is finite or not. The reader may find it helpful to think of $\mathbb{F}$ as the field $\mathbb{Q}$ of the rational numbers; when we restrict ourselves to finite fields, it may be useful to assume that $\mathbb{F}$ is $\mathbb{Z}_p$, the field of integers modulo some prime number $p$. We will use the unit-cost RAM model from Section 1.5.1 to measure the running time of an algorithm over the field $\mathbb{F}$. In this model each field operation (addition, subtraction, multiplication, division, comparison, or choosing a random element) takes unit time, provided the operand magnitude is polynomially related to the input size. For example, over the field of rationals we will assume that operations involving $O(\log n)$-bit numbers take unit time. This is not completely realistic as arithmetic operations are significantly more expensive in practice. However, in most applications described below this small additional factor in the running time is inconsequential, and we would get essentially the same result in the more expensive model.

7.1. Fingerprinting and Freivalds’ Technique

We illustrate fingerprinting by describing a technique for verifying matrix multiplication. The fastest known algorithm for matrix multiplication runs in time $O(n^{2.376})$, which improves significantly on the obvious $O(n^3)$ time algorithm but has the disadvantage of being extremely complicated. Suppose we are given an implementation of this algorithm and would like to verify its correctness. Since program verification is a difficult task, a reasonable goal might be to verify the correctness of the output produced on specific executions of the algorithm. (Such verification on specific inputs has been studied in the theory of program checking.) In other words, given $n \times n$ matrices $A$, $B$, and $C$ over the field $\mathbb{F}$, we would like to verify that $AB = C$. We cannot afford to use a simpler but slower algorithm for matrix multiplication to verify the output $C$, as this would defeat the purpose of using the fast matrix multiplication algorithm. Moreover, we would like to use the fact that we do not have to compute $C$; rather, our task is to verify that this product is indeed $C$. The following technique, known as Freivalds’ technique, provides an elegant solution. It gives an $O(n^2)$ time randomized algorithm with a bounded error probability.

The randomized algorithm first chooses a random vector $r \in \{0, 1\}^n$; each component of $r$ is chosen independently and uniformly at random from 0 and 1, the additive and multiplicative identities of the field $\mathbb{F}$. We can compute $x = Br$, $y = Ax = AB r$, and $z = Cr$ in $O(n^2)$ time; clearly, if $AB = C$ then $y = z$. We now show that for $AB \neq C$, the probability that $y \neq z$ is at least $1/2$. The algorithm errs only if $AB \neq C$ but $y$ and $z$ turn out to be equal.

**Theorem 7.1**: Let $A$, $B$, and $C$ be $n \times n$ matrices over $\mathbb{F}$ such that $AB \neq C$. Then for $r$ chosen uniformly at random from $\{0, 1\}^n$, $\Pr[ABr = Cr] \leq 1/2$. 

162
Proof: Let $D = AB - C$; we know that $D$ is not the all-zeroes matrix. We wish to bound the probability that $y = z$, or, equivalently, the probability that $Dr = 0$. Without loss of generality, we may assume that the first row in $D$ has a non-zero entry, and that all the non-zero entries in that row precede the zero entries. Let $d$ be the vector consisting of the entries from the first row in $D$, and assume that the first $k > 0$ entries in $d$ are non-zero. We concentrate on the probability that the inner product of $d$ and $r$ is non-zero; since the first entry in $Dr$ is exactly $d^T r$, this yields a lower bound on the probability that $y \neq z$.

Now, the inner product $d^T r = 0$ if and only if

$$r_1 = -\frac{\sum_{i=2}^{k} d_i r_i}{d_1}.$$  

(7.1)

We invoke the Principle of Deferred Decisions (Section 3.5) and assume that all the other random entries in $r$ are chosen before $r_1$. Then the right-hand side of (7.1) is fixed at some value $v \in \mathbb{F}$. Since $r_1$ is uniformly distributed over a set of size 2, the probability that it equals $v$ cannot exceed 1/2.

Exercise 7.1: Verify that there is nothing magical about choosing $r$ to have only entries drawn from $\{0, 1\}$. In fact, any two elements of $\mathbb{F}$ may be used instead.

Thus, in $O(n^2)$ time we have reduced the matrix product verification problem to that of verifying the equality of two vectors, and the latter can be done in $O(n)$ time. This gives an overall running time of $O(n^2)$ for this Monte Carlo procedure. The probability of error can be reduced to $1/2^k$ by performing $k$ independent iterations. The following exercise gives an alternative approach to reducing the probability of error.

Exercise 7.2: Suppose that each component of $r$ is chosen uniformly and independently from some subset $S \subseteq \mathbb{F}$. Show that the probability of error in the verification procedure is no more than $1/|S|$. Compare the usefulness of the two different methods for reducing the error probability.

Freivalds' technique is applicable to verifying any matrix identity $X = Y$. Of course, if $X$ and $Y$ are explicitly provided, just comparing their entries takes only $O(n^2)$ time. But as in the case of matrix multiplication, there are situations where computing $X$ explicitly is expensive (or even infeasible, as we will see in Section 7.8), whereas computing $Xr$ is easy.

7.2. Verifying Polynomial Identities

Freivalds' technique is fairly general in that it can be applied to the verification of several different kinds of identities. In this section we show that it also applies
to the verification of identities involving polynomials. Two polynomials \( P(x) \) and \( Q(x) \) are said to be equal if they have the same coefficients for corresponding powers of \( x \). Verifying identities of integers, or, in general, strings over any fixed alphabet, is a special case since we can represent any string of length \( n \) as a polynomial of degree \( n \). This is achieved by treating the \( k \)th element in the string as the coefficient of the \( k \)th power of a symbolic variable.

We first consider the polynomial product verification problem: given polynomials \( P_1(x), P_2(x), P_3(x) \in \mathbb{F}[x] \), verify that \( P_1(x) \times P_2(x) = P_3(x) \). Assume that the polynomials \( P_1(x) \) and \( P_2(x) \) are of degree at most \( n \); then \( P_3(x) \) cannot have degree exceeding \( 2n \). Polynomials of degree \( n \) can be multiplied in \( O(n \log n) \) time using Fast Fourier Transforms, whereas the evaluation of a polynomial at a fixed point requires \( O(n) \) time.

The basic idea underlying the randomized algorithm for polynomial product verification is similar in spirit to the algorithm for matrices. Let \( \mathcal{S} \subseteq \mathbb{F} \) be a set of size at least \( 2n + 1 \). Pick \( r \in \mathcal{S} \) uniformly at random and evaluate \( P_1(r), P_2(r), \) and \( P_3(r) \) in \( O(n) \) time. The polynomial identity \( P_1(x)P_2(x) = P_3(x) \) is declared correct unless \( P_1(r)P_2(r) \neq P_3(r) \). This algorithm errs only when the polynomial identity is false but the evaluation of the polynomials at \( r \) fails to detect this.

Define the polynomial \( Q(x) = P_1(x)P_2(x) - P_3(x) \) of degree \( 2n \). We say that a polynomial \( P \) is identically zero, or \( P \equiv 0 \), if all of its coefficients are zero. Clearly, \( Q(x) \) is identically zero if and only if the polynomial product is correct.

We complete the analysis of the randomized verification algorithm by showing that if \( Q(x) \neq 0 \), then with high probability \( Q(r) = P_1(r)P_2(r) - P_3(r) \neq 0 \). Elementary algebra tells us that \( Q \) can have at most \( 2n \) distinct roots. Hence, unless \( Q \equiv 0 \), not more than \( 2n \) different choices of \( r \in \mathcal{S} \) will have \( Q(r) = 0 \). Thus, the probability of error is at most \( 2n/|\mathcal{S}| \). This probability can be reduced by either using independent iterations of the entire algorithm or by choosing a sufficiently large set \( \mathcal{S} \).

In the case where \( \mathbb{F} \) is an infinite field (such as the reals), the error probability can be reduced to \( 0 \) by choosing \( r \) uniformly from the entire field \( \mathbb{F} \). Unfortunately, this requires an infinite number of random bits! We could also use a deterministic version of this algorithm where each choice of \( r \in \mathcal{S} \) is tried once. But this requires \( 2n + 1 \) different evaluations of each polynomial, and the best algorithm for this requires \( \Theta(n \log^2 n) \) time, which is more than the time required to actually multiply \( P_1(x) \) and \( P_2(x) \).

This verification procedure is not restricted to polynomial product verification. It is a generic procedure for testing any polynomial identity of the form \( P_1(x) = P_2(x) \), by transforming it into the identity \( Q(x) = P_1(x) - P_2(x) \equiv 0 \). Obviously, if the polynomials \( P_1 \) and \( P_2 \) are explicitly provided, we can perform this task deterministically in \( O(n) \) time by comparing corresponding coefficients. The randomized algorithm will take as long to just evaluate the polynomials at a random point. However, the verification procedure pays off in situations where the polynomials are provided implicitly, such as when we have only a "black box" for computing the polynomial, with no means of accessing its coefficients. There are also situations where the polynomials are provided in
7.2 VERIFYING POLYNOMIAL IDENTITIES

A form where computing the actual coefficients is exceedingly expensive. One example is provided by the following problem concerning the determinant of a symbolic matrix; in fact, this problem will turn out to be the same as that of verifying a polynomial identity involving multivariate polynomials, necessitating a generalization of the idea used for univariate polynomials.

Let $M$ be an $n \times n$ matrix. The determinant of $M$ is defined by

$$
\det(M) = \sum_{\pi \in S_n} \text{sgn}(\pi) \prod_{i=1}^{n} M_{i,\pi(i)},
$$

(7.2)

where $S_n$ is the symmetric group of permutations of size $n$, and $\text{sgn}(\pi)$ is the sign of the permutation $\pi$. Recall that $\text{sgn}(\pi) = (-1)^t$, where $t$ is the number of pairwise element exchanges required to transform the identity permutation into $\pi$. Although the determinant has $n!$ terms, it can be evaluated in polynomial time given explicit values for the matrix entries $M_{ij}$.

**Definition 7.1:** The Vandermonde matrix $M(x_1, \ldots, x_n)$ is defined in terms of the indeterminates $x_1, \ldots, x_n$ such that $M_{ij} = x_i^{j-1}$, that is

$$
M = \begin{pmatrix}
1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\
1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_n & x_n^2 & \cdots & x_n^{n-1}
\end{pmatrix}
$$

Vandermonde's identity states that for this matrix $M$, $\det(M) = \prod_{i \neq j} (x_i - x_j)$. Suppose that we did not have a proof of this identity and would like to verify it efficiently. Computing the determinant of this symbolic matrix is prohibitively expensive since it has $n!$ terms. Instead, we will formulate this as the problem of verifying that the polynomial $Q(x_1, \ldots, x_n) = \det(M) - \prod_{i \neq j} (x_i - x_j)$ is identically zero. Drawing upon our experience with Freivalds' technique, it seems natural to substitute random values for each $x_i$ and check whether $Q \equiv 0$. The polynomial $Q$ is easy to evaluate at a specific point since the determinant can be computed in polynomial time for specified values of the variables $x_1, \ldots, x_n$.

We formalize this intuition by extending the analysis of Freivalds' technique for univariate polynomial identity verification to the multivariate case. In a multivariate polynomial $Q(x_1, \ldots, x_n)$, the degree of any term is the sum of the exponents of the variables, and the total degree of $Q$ is the maximum of the degrees of its terms.

**Theorem 7.2 (Schwartz-Zippel Theorem):** Let $Q(x_1, \ldots, x_n) \in \mathbb{F}[x_1, \ldots, x_n]$ be a multivariate polynomial of total degree $d$. Fix any finite set $\mathcal{S} \subseteq \mathbb{F}$, and let $r_1, \ldots, r_n$ be chosen independently and uniformly at random from $\mathcal{S}$. Then

$$
\Pr[Q(r_1, \ldots, r_n) = 0 \mid Q(x_1, \ldots, x_n) \neq 0] \leq \frac{d}{|\mathcal{S}|}.
$$

165
**Proof:** The proof is by induction on the number of variables $n$. The base case $n = 1$ involves a univariate polynomial $Q(x_1)$ of degree $d$, and by the preceding discussion we already know that for $Q(x_1) \neq 0$, the probability that $Q(r_1) = 0$ is at most $d/|S|$. Assume now that the induction hypothesis is true for a multivariate polynomial with up to $n - 1$ variables, for $n > 1$.

Consider the polynomial $Q(x_1, \ldots, x_n)$, and factor out the variable $x_1$ to obtain

$$Q(x_1, \ldots, x_n) = \sum_{i=0}^{k} x_1^i Q_i(x_2, \ldots, x_n),$$

where $k \leq d$ is the largest exponent of $x_1$ in $Q$. (Assume that $x_1$ affects $Q$, so that $k > 0$). The coefficient of $x_1^k$, $Q_k(x_2, \ldots, x_n)$, is not identically zero by our choice of $k$. Since the total degree of $Q_k$ is at most $d - k$, the induction hypothesis implies that the probability that $Q_k(r_2, \ldots, r_n) = 0$ is at most $(d - k)/|S|$.

Suppose that $Q_k(r_2, \ldots, r_n) \neq 0$. Consider the following univariate polynomial:

$$q(x_1) = Q(x_1, r_2, r_3, \ldots, r_n) = \sum_{i=0}^{k} x_1^i Q_i(r_2, \ldots, r_n).$$

The polynomial $q(x_1)$ has degree $k$, and it is not identically zero since the coefficient of $x_1^k$ is $Q_k(r_2, \ldots, r_n)$. The base case now implies that the probability that $q(r_1) = Q(r_1, r_2, \ldots, r_n)$ evaluates to 0 is at most $k/|S|$.

Thus, we have shown the following two inequalities.

$$\Pr[Q_k(r_2, \ldots, r_n) = 0] \leq \frac{d - k}{|S|};$$

$$\Pr[Q(r_1, r_2, \ldots, r_n) = 0 \mid Q_k(r_2, \ldots, r_n) \neq 0] \leq \frac{k}{|S|}.$$

Invoking the result in Exercise 7.3, we find that the probability that $Q(r_1, r_2, \ldots, r_n) = 0$ is no more than the sum of these two probabilities, which is $d/|S|$. This completes the induction.

Exercise 7.3: Show that for any two events $\mathcal{E}_1$ and $\mathcal{E}_2$,

$$\Pr[\mathcal{E}_1] \leq \Pr[\mathcal{E}_1 \mid \overline{\mathcal{E}_2}] + \Pr[\mathcal{E}_2].$$

The randomized verification procedure for polynomials has one potential problem. In the case of infinite fields, the intermediate results in the evaluation of the polynomial could involve enormous values. This problem can be avoided in the case of integers by performing all the computations modulo a small random prime number, without adversely affecting the error probability. We will return to this issue in Example 7.1.

As suggested in Problem 7.1, Theorem 7.2 can be viewed as a generalization of Freivalds' technique from Section 7.1. A generalized version of this theorem is described in Problem 7.6.
7.3. Perfect Matchings in Graphs

We illustrate the power of the techniques of Section 7.2 by giving a fascinating application. Consider a bipartite graph $G(U, V, E)$ with the independent sets of vertices $U = \{u_1, \ldots, u_n\}$ and $V = \{v_1, \ldots, v_n\}$. A matching is a collection of edges $M \subseteq E$ such that each vertex occurs at most once in $M$. A perfect matching is a matching of size $n$. Each perfect matching $M$ in $G$ can be viewed as a permutation from $U$ into $V$. More precisely, the perfect matchings in $G$ can be put into a one-to-one correspondence with the permutations in $S_n$, where the matching corresponding to a permutation $\pi \in S_n$ is given by the pairs $(u_i, v_{\pi(i)})$, for $1 \leq i \leq n$. The following theorem draws a connection between determinants and matchings.

**Theorem 7.3 (Edmonds' Theorem):** Let $A$ be the $n \times n$ matrix obtained from $G(U, V, E)$ as follows:

$$A_{ij} = \begin{cases} x_{ij} & (u_i, v_j) \in E \\ 0 & (u_i, v_j) \notin E \end{cases}.$$  

Define the multivariate polynomial $Q(x_{11}, x_{12}, \ldots, x_{nn})$ as being equal to $\det(A)$. Then, $G$ has a perfect matching if and only if $Q \neq 0$.

**Remark:** The matrix of indeterminates is sometimes referred to as the Edmonds matrix of a bipartite graph. We do not explicitly specify the underlying field because any field will do for the purposes of this theorem.

**Proof:** The determinant of $A$ is given by

$$\det(A) = \sum_{\pi \in S_n} \text{sgn}(\pi) A_{1, \pi(1)} A_{2, \pi(2)} \cdots A_{n, \pi(n)}.$$  

Since each indeterminate $x_{ij}$ occurs at most once in $A$, there can be no cancellation of the terms in the summation. Therefore the determinant is not identically zero if and only if there is a permutation $\pi$ for which the corresponding term in the summation is non-zero. The latter happens if and only if each of the entries $A_{i, \pi(i)}$, for $1 \leq i \leq n$, is non-zero. This is equivalent to having a perfect matching (the one corresponding to $\pi$) in $G$. \hfill $\Box$

We can now construct a simple randomized test for the existence of perfect matchings. Using the algorithm from Section 7.2, we can determine whether the determinant is identically zero or not. The time required is dominated by the cost of computing a determinant, which is essentially that of multiplying two matrices. As it turns out, there are algorithms for constructing a maximum matching in a graph in time $O(m\sqrt{n})$, where $m = |E|$. Since the time to compute the determinant exceeds $m\sqrt{n}$ for small $m$, the payoff in using this randomized decision procedure is marginal at best. However, we will see later (in Section 12.4) that this decision procedure is essential for devising a fast parallel algorithm for computing a maximum matching in a graph. In Problem 7.8 we will see that this technique also applies to the case of non-bipartite graphs.
MATHEMATICAL BACKGROUND

Note that the equality $f(n) = O(g(n))$ does not use "=" in a symmetric fashion.

Combinatorial Inequalities

We now turn our attention to the binomial coefficients, defined as follows. Let $n \geq k \geq 0$.

$$\binom{n}{k} = \binom{n}{n-k} = \frac{n!}{k!(n-k)!}$$

If $k > n \geq 0$ we define $\binom{n}{k} = 0$. The reason for the name "binomial coefficients" is their appearance in the binomial expansion:

$$(p + q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}.$$

Proposition B.1 (Stirling's Formula):

$$n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \left(1 + \frac{1}{12n} + O\left(\frac{1}{n^2}\right)\right)$$

From this one obtains the following inequalities involving binomial coefficients.

Proposition B.2: Let $n \geq k \geq 0$.

1. $\binom{n}{k} \leq \frac{n^k}{k!}$.
2. For large $n$, $\binom{n}{k} \sim \frac{n^k}{k!}$.
3. $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$.
4. $\binom{n}{k} \geq \left(\frac{e}{k}\right)^k$.

The following power series expansions sometimes allow us to obtain useful inequalities.

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots$$

We list below several inequalities involving the exponential function. The reader may refer to Mitrinović [311] for the derivations and other variants.

Proposition B.3:

1. For all $t \in \mathbb{R}$, $e^t \geq 1 + t$ with equality holding only at $t = 0$. 

434
MATHEMATICAL BACKGROUND

2. For all \( t, n \in \mathbb{R} \), such that \( n \geq 1 \) and \( |t| \leq n \),
   \[
   e^t \left( 1 - \frac{t^2}{2n} \right) \leq \left( 1 + \frac{t}{n} \right)^n \leq e^t.
   \]
   Note that this holds even for negative values of \( t \).

3. For all \( t, n \in \mathbb{R}^+ \),
   \[
   \left( 1 + \frac{t}{n} \right)^n \leq e^t \leq \left( 1 + \frac{t}{n} \right)^{n+1/2}.
   \]

For any \( n \in \mathbb{N} \), we define the \( n \)th Harmonic number \( H_n \) as follows:
   \[
   H_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}.
   \]

**Proposition B.4:** For any \( n \in \mathbb{N} \), the \( n \)th Harmonic number is
   \[
   H_n = \ln n + \Theta(1).
   \]

The \( n \)th Fibonacci number is defined as follows:
   \[
   F_0 = F_1 = 1,
   \]
   and for \( n \geq 2 \),
   \[
   F_n = F_{n-1} + F_{n-2}.
   \]

**Proposition B.5:** For all \( n \in \mathbb{N} \), \( F_n = \Theta(\phi^n) \), where \( \phi = (1 + \sqrt{5})/2 \) is the golden ratio.

Linear Algebra

Consider the field \( \mathbb{R} \) of real numbers under addition and multiplication, and
the real vector space \( \mathbb{R}^n \) of \( n \)-dimensional vectors over \( \mathbb{R} \). This vector space is
an inner product vector space, where we define the inner product of two vectors
\( \mathbf{v}, \mathbf{w} \in \mathbb{R}^n \) as
   \[
   \mathbf{v}^T \mathbf{w} = \sum_{i=1}^{n} v_i w_i,
   \]
where \( v_i \) and \( w_i \) are the \( i \)th components of the vectors \( \mathbf{v} \) and \( \mathbf{w} \). The vectors
\( \mathbf{v} \) and \( \mathbf{w} \) are said to be orthogonal, denoted \( \mathbf{v} \perp \mathbf{w} \), if their inner product \( \mathbf{v}^T \mathbf{w} \)
equals 0. A subspace \( \mathcal{W} \) of a vector space \( \mathcal{V} \) is a subset \( \mathcal{W} \subseteq \mathcal{V} \), which forms a
vector space; its orthogonal subspace is \( \mathcal{W}^\perp = \{ \mathbf{v} \in \mathcal{V} \mid \forall \mathbf{w} \in \mathcal{W}, \mathbf{v} \perp \mathbf{w} \} \). The
vector space \( \mathcal{V} \) is a direct sum of the orthogonal subspaces \( \mathcal{W} \) and \( \mathcal{W}^\perp \). In other
words, every vector \( \mathbf{v} \in \mathcal{V} \) can be uniquely expressed as \( \mathbf{v} = \mathbf{w} + \mathbf{w}' \), where \( \mathbf{w} \in \mathcal{W} \)
and \( \mathbf{w}' \in \mathcal{W}^\perp \).

We define three norms for vectors in an inner product vector space.
**Appendix C**

**Basic Probability Theory**

**Proposition C.1 (Principle of Inclusion–Exclusion):** Let $E_1, E_2, \ldots, E_n$ be arbitrary events. Then

$$
\Pr[\bigcup_{i=1}^{n} E_i] = \sum_{i} \Pr[E_i] - \sum_{i<j} \Pr[E_i \cap E_j] + \sum_{i<j<k} \Pr[E_i \cap E_j \cap E_k] - \cdots + (-1)^{j+1} \sum_{i_1<i_2<\cdots<i_j} \Pr[\bigcap_{k=1}^{j} E_{i_k}] + \cdots
$$

**Proposition C.2 (Boole-Bonferroni Inequalities):** Let $E_1, E_2, \ldots, E_n$ be arbitrary events. Then, for even $k$

$$
\Pr[\bigcup_{i=1}^{n} E_i] \geq \sum_{j=1}^{k} (-1)^{j+1} \sum_{i_1<i_2<\cdots<i_j} \Pr[\bigcap_{i=1}^{j} E_{i_k}],
$$

and for odd $k$

$$
\Pr[\bigcup_{i=1}^{n} E_i] \leq \sum_{j=1}^{k} (-1)^{j+1} \sum_{i_1<i_2<\cdots<i_j} \Pr[\bigcap_{i=1}^{j} E_{i_k}].
$$

**Definition C.4:** The **conditional probability** of $E_1$ given $E_2$ is denoted by $\Pr[E_1 \mid E_2]$ and is given by

$$
\frac{\Pr[E_1 \cap E_2]}{\Pr[E_2]}
$$

assuming that $\Pr[E_2] > 0$.

This corresponds to the probability that an experiment has an outcome in the set $E_1$ when we already know that it is in the set $E_2$.

**Proposition C.3:** Let $E_1, E_2, \ldots, E_k$ be a partition of the sample space $\Omega$. Then for any event $E$

$$
\Pr[E] = \sum_{i=1}^{k} \Pr[E \mid E_i] \Pr[E_i].
$$

Since $\Pr[E_1 \cap E_2] = \Pr[E_1 | E_2] \Pr[E_2] = \Pr[E_2 | E_1] \Pr[E_1]$, we obtain Bayes' rule from the previous proposition.

**Proposition C.4 (Bayes' Rule):** Let $E_1, E_2, \ldots, E_k$ be a partition of the sample space $\Omega$. Then for any event $E$

$$
\Pr[E_1 \mid E] = \frac{\Pr[E_1 \cap E]}{\Pr[E]} = \frac{\Pr[E \mid E_1] \Pr[E_1]}{\sum_{j=1}^{k} \Pr[E \mid E_j] \Pr[E_j]}.
$$
Definition C.5: A collection of events \{E_i \mid i \in I\} is independent if for all subsets \(S \subseteq I\),

\[
\Pr[\bigcap_{i \in S} E_i] = \prod_{i \in S} \Pr[E_i].
\]

These events are said to be \(k\)-wise independent if every subcollection consisting of \(k\) events is independent. The special case of 2-wise independence is often referred to as pairwise independence.

Equivalently, using the definition of conditional expectations, we can say that a collection of events \(\{E_i \mid i \in I\}\) is independent if for any \(j \in I\) and all subsets \(S \subseteq I \setminus \{E_j\}\),

\[
\Pr[E_j \mid \bigcap_{i \in S} E_i] = \Pr[E_j].
\]

In particular, if the events are pairwise independent then \(\Pr[E_i \mid E_j] = \Pr[E_i],\) for all \(i \neq j\).

Usually the events we deal with can be expressed in terms of real-valued functions called random variables. The argument of such a function is generally omitted as it always corresponds to a single experiment from the underlying probability space.

Definition C.6: A random variable \(X\) is a real-valued function over the sample space, \(X : \Omega \rightarrow \mathbb{R}\), such that for all \(x \in \mathbb{R}\),

\[
\{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F}.
\]

This gives us a compact representation of complex events since \(\Pr[X \leq x]\) is just another way of denoting \(\Pr[\{\omega \in \Omega \mid X(\omega) \leq x\}]\).

Definition C.7: The distribution function \(F : \mathbb{R} \rightarrow [0, 1]\) for a random variable \(X\) is defined as \(F_X(x) = \Pr[X \leq x]\).

A discrete random variable is a function over the sample space whose range is either a finite or countably infinite subset of \(\mathbb{R}\). Typically, we will be interested in discrete random variables that are integer-valued. An indicator variable is a discrete random variable that takes on only the values 0 or 1. An indicator variable \(E\) is used to denote the occurrence or non-occurrence of an event \(E\), where \(E = \{\omega \in \Omega \mid X(\omega) = 1\}\) and \(\overline{E} = \{\omega \in \Omega \mid X(\omega) = 0\}\). Observe that the notions of conditional probability and independence carry over to random variables, since they are just another way of denoting events. More precisely, two random variables \(X\) and \(Y\) are said to be independent if for each \(x, y \in \mathbb{R}\), the events \(\{X = x\}\) and \(\{Y = y\}\) are independent.

A random variable \(X\) is said to be continuous if it has a distribution function \(F\) whose derivative \(F'\) is a positive, integrable function. (In other words, \(F\) is absolutely continuous.) The function \(F'\) is referred to as the density function of the random variable \(X\). From here on all random variables are assumed to be
discrete, although with some care\(^1\) the following definitions can be extended to continuous random variables.

\textbf{Definition C.8:} The \textit{density function} \( p : \mathbb{R} \rightarrow [0, 1] \) for a random variable \( X \) is defined as \( p_X(x) = \Pr[X = x] \).

It is sometimes useful to combine the density or distribution functions for dependent random variables.

\textbf{Definition C.9:} The \textit{joint distribution function} \( F_{X,Y} : \mathbb{R} \times \mathbb{R} \rightarrow [0, 1] \) for random variables \( X \) and \( Y \) is defined as

\[ F_{X,Y}(x,y) = \Pr[(X \leq x) \cap (Y \leq y)]. \]

The \textit{joint density function} \( p_{X,Y} : \mathbb{R} \times \mathbb{R} \rightarrow [0, 1] \) for random variables \( X \) and \( Y \) is defined as

\[ p_{X,Y}(x,y) = \Pr[(X = x) \cap (Y = y)]. \]

Thus, \( \Pr[Y = y] = \sum_x p(x,y) \), and

\[ \Pr[X = x \mid Y = y] = \frac{p(x,y)}{\Pr[Y = y]}. \]

We can now restate the independence of \( X \) and \( Y \) as requiring that the joint density function be the product of the individual density functions of the two random variables.

\textbf{Definition C.10:} Random variables \( X \) and \( Y \) are said to be \textit{independent} if for all \( x, y \in \mathbb{R} \),

\[ p(x,y) = \Pr[X = x] \Pr[Y = y] \]

or, equivalently,

\[ \Pr[X = x \mid Y = y] = \Pr[X = x]. \]

These definitions extend to a set \( X_1, X_2, \ldots \) of more than two random variables, and the notion of \( k \)-wise independence can be defined as in Definition C.5.

The following discussion of expectations is in terms of single random variables, but they have the obvious generalizations to functions of multiple random variables using their joint density function.

\textbf{Definition C.11:} The \textit{expectation} of a random variable \( X \) with density function \( p \) is defined as \( \mathbb{E}[X] = \sum_x xp(x) \), where the summation is over the range of \( X \).

\(^1\) Basically all the definitions can be made in terms of the distribution function for the discrete random variable, and then carried over to the continuous case. For example, we say that two continuous random variables \( X \) and \( Y \) are independent if for each \( x, y \in \mathbb{R} \), the events \( \{X \leq x\} \) and \( \{Y \leq y\} \) are independent.
Note that the expectation may not be well-defined if the summation does not converge absolutely. For any real-valued function $g(X)$, we extend the definition of expectation to $E[g(X)] = \sum_x g(x)p(x)$. For any two random variables $X$ and $Y$, $E[X + Y] = E[X] + E[Y]$. The remarkable thing about this property is that it does not assume anything about the independence of the two random variables. In fact, this can be generalized as follows.

**Proposition C.5 (Linearity of Expectation):** Let $X_1, \ldots, X_k$ be arbitrary random variables, and $h(X_1, \ldots, X_k)$ a linear function. Then

$$E[h(X_1, \ldots, X_k)] = h(E[X_1], \ldots, E[X_k]).$$

This does not generalize to nonlinear functions, although with the assumption of independence we can prove a similar result for any polynomial $h$ using the following.

**Proposition C.6:** For independent random variables $X$ and $Y$,

$$E[XY] = E[X]E[Y].$$

Here are some other useful properties of expectations. We say that random variable $X$ stochastically dominates random variable $Y$ if, for all $z \in \mathbb{R}$, $\Pr[X > z] \geq \Pr[Y > z]$.

**Proposition C.7:** Let $X$ and $Y$ be random variables with finite expectations.

1. If $X$ stochastically dominates $Y$, then $E[X] \geq E[Y]$; equality holds if and only if $X$, $Y$ are identically distributed.
2. $|E[X]| \leq E[|X|]$.
3. For a non-negative integer-valued random variable $X$, $E[X] = \sum_{x=0}^{\infty} \Pr[X > x]$.

The density function of a random variable can be characterized in terms of the following expectations.

**Definition C.12:** For $k \in \mathbb{N}$, the $k$th moment $m_X^k$ and the $k$th central moment $\mu_X^k$ of a random variable $X$ are defined as follows:

$$m_X^k = E[X^k]$$

$$\mu_X^k = E[(X - E[X])^k].$$

The expected value of $X$ is sometimes denoted by $\mu_X = m_X^1$. The variance of $X$ is denoted $\text{var}[X]$ or $\sigma_X^2$ and this is $\mu_X^2$; the standard deviation $\sigma_X$ is the positive square root of the variance.

**Proposition C.8:** $\text{var}[X] = m_X^2 - (E[X])^2 = E[X^2] - E[X]^2$. 

443
Note that the next proposition does not generalize to arbitrary linear functions since \( \text{var}(cX) = c^2 \text{var}(X) \).

**Proposition C.9:** For independent random variables \( X \) and \( Y \),

\[
\text{var}(X + Y) = \text{var}(X) + \text{var}(Y).
\]

The use of generating functions to describe a density function often leads to simplified analysis of the moments.

**Definition C.13:** Let \( X \) be a non-negative integer-valued random variable with the density function \( p \). Then the probability generating function (pgf) of \( X \) is

\[
G_X(z) = E[z^X] = \sum_{i=0}^{\infty} p(i)z^i.
\]

The summation in the definition of \( G_X(z) \) always converges for \(|z| \leq 1\), and so we assume that the symbolic variable \( z \) lies in the interval \([-1, 1]\). The following results can be obtained by suitable differentiation and algebraic manipulation of \( G_X(z) \). The reader should keep in mind that the pgf \( G_X \) may not be well-defined for all real values of \( z \), but whenever \( G_X \) is well-defined, so are its derivatives with respect to \( z \). In the sequel, for a function \( f \), we denote by \( f' \) its derivative and by \( f^{(k)} \) its \( k \)th derivative.

**Proposition C.10:** Let \( X \) be a non-negative integer-valued random variable with the pgf \( G(z) \).

1. \( G(1) = 1 \).
2. \( E[X] = G'(1) \).
3. \( E[X^2] = G''(1) + G'(1) \).
4. \( \text{var}(X) = G''(1) + G'(1) - G'(1)^2 \).

**Proposition C.11:** Let \( X_1, \ldots, X_k \) be independent random variables with the pgf’s \( G_1(z), \ldots, G_k(z) \). Then the pgf of the random variable \( Y = \sum_{i=1}^{k} X_i \) is given by

\[
G(z) = \prod_{i=1}^{k} G_i(z).
\]

**Proposition C.12:** Let \( X_1, X_2, \ldots \) be a sequence of independent and identically distributed (i.i.d.) random variables with the common pgf \( G_X(z) \). If \( Y \) is a random variable with the pgf \( G_Y(z) \) and \( Y \) is independent of all the \( X_i \), then \( S = X_1 + X_2 + \cdots + X_Y \) has the pgf

\[ G_S(z) = G_Y(G_X(z)). \]