Appendix A

A.1 Probability Theory Basics

A.1.1 Probability Spaces

Definition A.1. A *probability space* consists of a *sample space* Ω, which is the set of all possible *outcomes*, a set of subsets of Ω called *events* (which form a σ-algebra), and a function Pr, called the *probability measure*, assigning to every event a probability in $[0,1]$, such that $\mathsf{Pr}(\Omega) = 1$ and the probability of the union of disjoint events is the sum of their probabilities.[1]

A probability space is *discrete* if Ω is finite or countably infinite. One also uses the term *discrete random experiment*. For a discrete probability space one can consider all subsets of Ω as events, and it is therefore completely described by the probabilities of the elementary events $\omega \in \Omega$, i.e., by the so-called *probability mass function* $p : \Omega \to [0,1]$ satisfying

$$\sum_{\omega\in\Omega} p(\omega) = 1,$$

where $\mathsf{Pr}(\{\omega\}) = p(\omega)$ for all $\omega \in \Omega$ and for any event $\mathcal{A} \subseteq \Omega$ we have

$$\mathsf{Pr}(\mathcal{A}) = \sum_{\omega\in\mathcal{A}} p(\omega).$$

We will almost exclusively consider finite random experiments or infinite random experiments that can be understood as an infinite sequence of finite random experiments.

One sometimes considers more than one random experiment at the same time, and hence using Pr as the probability measure is ambiguous. In such a case we make the random experiment E explicit as a superscript: $\mathsf{Pr}^E(\cdot)$. If a random experiment E consists of selecting some random variables *independently*, then we write the list of random variables as superscript. For example, if the random variables X, Y, and Z are chosen independently, we write $\mathsf{Pr}^{XYZ}(\mathcal{A})$ for the probability of the event $\mathcal{A}$ in this random experiment. If for example Y and Z are not independent, but the pair (Y,Z) is independent of X, then we write $\mathsf{Pr}^{X(Y,Z)}(\mathcal{A})$, making explicit that the pair (Y,Z) is considered as a single (independent) random variable.

[1] For a good introduction to probability theory we refer to W. Feller, *An Introduction to Probability Theory*, Wiley, 3rd edition, 1968. or to John A. Rice, *Mathematical Statistics and Data Analysis*, Duxbury Press, 2nd edition, 1995.

Definition A.2. Two events $\mathcal{A}$ and $\mathcal{B}$ are called (statistically) *independent* if

$$\mathsf{Pr}(\mathcal{A}\cap\mathcal{B}) = \mathsf{Pr}(\mathcal{A})\cdot\mathsf{Pr}(\mathcal{B}).$$

Definition A.3. The *conditional probability* of an event $\mathcal{A}$, given an event $\mathcal{B}$, is defined as

$$\mathsf{Pr}(\mathcal{A}|\mathcal{B}) = \frac{\mathsf{Pr}(\mathcal{A}\cap\mathcal{B})}{\mathsf{Pr}(\mathcal{B})}$$

if $\mathsf{Pr}(\mathcal{B}) > 0$, and it is undefined if $\mathsf{Pr}(\mathcal{B}) = 0$.

A.1.2 Random Variables

Definition A.4. A *random variable* X is a function from Ω to some set $\mathcal{X}$. The function $\mathcal{X} \to [0,1]$ assigning to $x \in \mathcal{X}$ the probability of the event[2] $X = x$ is denoted as P_X and is called the *probability mass function* of X or also often the *probability distribution* of X:

$$\mathsf{P}_X(x) \;=\; \mathsf{Pr}(X = x) \;=\; \sum_{\omega\in\Omega: X(\omega)=x} p(\omega).$$

A list of random variables can be viewed as a single random variable whose values are tuples. For example, we can consider the pair (X,Y) whose probability distribution[3]

$$\mathsf{P}_{XY} : \mathcal{X}\times\mathcal{Y} \to [0,1]$$

is a function of two arguments defined by[4]

$$\mathsf{P}_{XY}(x,y) \;=\; \mathsf{Pr}(X = x, Y = y).$$

The probability distributions of X and Y are determined uniquely by P_{XY} and are called *marginal distributions*:

$$\mathsf{P}_X(x) = \sum_{y:\mathsf{P}_{XY}(x,y)>0} \mathsf{P}_{XY}(x,y)$$

[2] Here $X = x$ is understood as an event, i.e., as $\{\omega \in \Omega : X(\omega) = x\}$.
[3] One writes P_{XY} instead of $\mathsf{P}_{(X,Y)}$.
[4] Here $\mathsf{Pr}(X = x, Y = y)$ stands for $\mathsf{Pr}(\{\omega \in \Omega : X(\omega) = x \text{ and } Y(\omega) = y\})$.

(and analogously for $\mathsf{P}_Y(y)$). More generally, for n random variables $X_1, \ldots, X_n$ and for $m < n$ we have:

$$\mathsf{P}_{X_1 \cdots X_m}(x_1, \ldots, x_m) = \sum_{(x_{m+1}, \ldots, x_n) : \mathsf{P}_{X_1 \cdots X_n}(x_1, \ldots, x_n) > 0} \mathsf{P}_{X_1 \cdots X_n}(x_1, \ldots, x_n).$$

The marginal distribution of any subset of the random variables is computed analogously.

Definition A.5. The *conditional probability distribution* of random variable X, given (or conditioned on) the event $\mathcal{A}$ with $\mathsf{Pr}(\mathcal{A}) > 0$, is defined as

$$\mathsf{P}_{X|\mathcal{A}}(x) = \mathsf{Pr}(X = x | \mathcal{A}).$$

The *conditional probability distribution of X, given Y* is the partial function

$$\mathsf{P}_{X|Y} : \mathcal{X} \times \mathcal{Y} \to [0, 1]$$

defined by

$$\mathsf{P}_{X|Y}(x, y) \;:=\; \mathsf{Pr}(X = x | Y = y) \;=\; \frac{\mathsf{Pr}(X = x, Y = y)}{\mathsf{Pr}(Y = y)}.$$

$\mathsf{P}_{X|Y}(x, y)$ is not defined for values y with $\mathsf{P}_Y(y) = 0$.

Note that $\mathsf{P}_{X|Y}$ is a real-valued partial function with *two* arguments. Note also that for every y with $\mathsf{P}_Y(y) > 0$ we have

$$\sum_x \mathsf{P}_{X|Y}(x, y) = 1,$$

i.e., $\mathsf{P}_{X|Y}(\cdot, y)$ is, when considered as a one-argument function, itself a probability distribution.

Definition A.6. Two random variables X and Y are *statistically independent* if (and only if)

$$\mathsf{P}_{XY}(x, y) \;=\; \mathsf{P}_X(x) \cdot \mathsf{P}_Y(y)$$

for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Analogously, a list of random variables $X_1, \ldots, X_n$ are independent if the probability distribution factors:

$$\mathsf{P}_{X_1 \cdots X_n}(x_1, \ldots, x_n) = \mathsf{P}_{X_1}(x_1) \cdots \mathsf{P}_{X_n}(x_n).$$

It follows from the definition that for statistically independent X and Y we have $\mathsf{P}_{X|Y}(x, y) = \mathsf{P}_X(x)$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ with $\mathsf{P}_Y(y) \neq 0$.

A.1.3 Expected Value and Variance

For a random variable X whose range is (a subset of) the real numbers, the *expected value* or *expectation* $E[X]$ and the *variance* $\mathrm{Var}[X]$ are defined as follows:

$$E[X] \;=\; \sum_x x \cdot \mathsf{P}_X(x)$$

and

$$\mathrm{Var}[X] \;=\; \sum_x (x - E[X])^2 \; \mathsf{P}_X(x) \;=\; E[(X - E[X])^2].$$

It is easy to show that

$$E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n].$$

If $X_1, \ldots, X_n$ are statistically independent (actually, pairwise statistically independent suffices), then

$$\mathrm{Var}[X_1 + \cdots + X_n] = \mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n].$$

For a real-valued function f whose domain includes $\mathcal{X}$, we have

$$E[f(X)] = \sum_x f(x) \mathsf{P}_X(x)$$

and

$$\mathrm{Var}[f(X)] = \sum_x (f(x) - E[f(X)])^2 \; \mathsf{P}_X(x).$$

A.2 Information Theory Basics

Information theory, invented by Claude Shannon in 1948, is a very powerful theory for analyzing many problems related to information transmission and storage. Information theory can be considered as a special branch of probability theory. Our purpose for using information theory is to prove results on information-theoretic security of cryptographic schemes, i.e., security against computationally unbounded adversaries.

The basic information-theoretic quantity is the entropy of a probability distribution or, equivalently, of a random variable (i.e., of its probability distribution). Entropy is a real-valued positive quantity that measures in some sense the uncertainty contained in a probability distribution. The larger the entropy, the larger the uncertainty about the outcome of a random experiment with the considered probability distribution.

Definition A.7. The *entropy* of a discrete random variable X (with probability distribution P_X) is defined as

$$H(X) \;:=\; -\sum_{x \in \mathcal{X} : \mathsf{P}_X(x) \neq 0} \mathsf{P}_X(x) \log_2 \mathsf{P}_X(x).$$

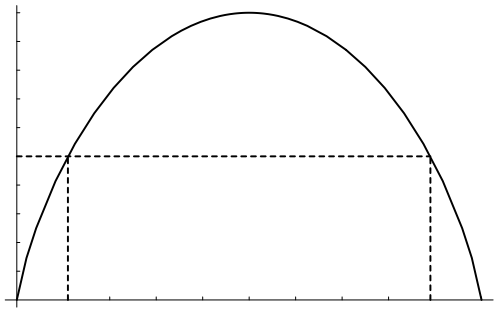

Figure A.1: The binary entropy function $h(p)$.

Entropy can also be defined as an expected value. Define the function $g\colon \mathcal{X} \to \mathbb{R}$ as $g(x) := -\log_2 \mathsf{P}_X(x)$. Then

$$H(X) \;=\; E[g(X)] \;=\; E[-\log_2 \mathsf{P}_X(X)]$$

We will not be interested in the particular formula, nor why this is the right measure of uncertainty for applications like data compression (source coding) or channel coding. We are only interested in certain properties of entropy which we discuss next. We do not prove the basic information-theoretic facts.

Example A.1. The probability distribution of a binary random variable X is characterized by the single parameter $\mathsf{P}_X(0) = p$ (since $\mathsf{P}_X(1) = 1 - p$). The entropy of X can be considered as a function of p; this is called the *binary entropy function* h:

$$h(p) \;=\; -p \log_2 p - (1-p)\log_2(1-p)$$

for $0 < p < 1$ and $h(0) = h(1) = 0$. The function $h(p)$ is strictly concave and takes its maximum at $p = 1/2$, where $h(1/2) = 1$ (see Figure A.1).

Theorem A.1.

$$0 \le H(X) \le \log_2 |\mathcal{X}|$$

with equality on the left side if and only if $\mathsf{P}_X(x) = 1$ *for one value* x *and with equality on the right side if and only if* P_X *is the uniform distribution over* $\mathcal{X}$.

A pair $[X, Y]$, a triple $[X, Y, Z]$, or a list $[X_1, \ldots, X_N]$ of random variables can be considered as a single random variable. Hence the entropy of several random variables is already defined. We write $H(XY)$ instead of $H([X, Y])$, and similarly for several random variables. For example, we have

$$H(XY) = -\sum_{(x,y)} \mathsf{P}_{XY}(x,y) \log_2 \mathsf{P}_{XY}(x,y) = E[-\log_2 \mathsf{P}_{XY}(X,Y)].$$

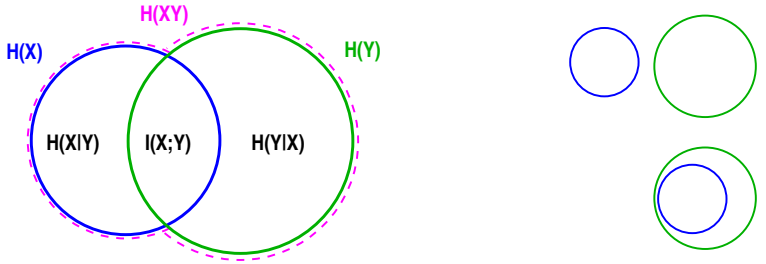

Figure A.2: Left: The relation between $H(X)$, $H(Y)$, and $H(XY)$. Right: Special cases of the diagram if X and Y are statistically independent (top) and if X is uniquely determined by Y (bottom).

One can prove:

Theorem A.2. *We have*

$$H(X) \le H(XY)$$

with equality if and only if Y *is uniquely determined by* X, *and*

$$H(XY) \;\le\; H(X) + H(Y)$$

with equality if and only if X *and* Y *are statistically independent.*

This motivates the definition of two derived quantities (see Figure A.2). It follows from the above theorem that both quantities are non-negative.

Definition A.8. The *conditional entropy of* Y, *given* X is defined as

$$H(Y|X) := H(XY) - H(X),$$

and the *mutual information between* X *and* Y is defined as

$$I(X;Y) := H(X) + H(Y) - H(XY).$$

Theorem A.2 can equivalently be stated as $H(Y|X) \ge 0$ and $I(X;Y) \ge 0$. The so-called chain rule follows by repeated application of this definition:

$$H(X_1 \cdots X_N) \;=\; \sum_{n=1}^{N} H(X_n | X_1 \cdots X_{n-1}).$$

There is one more basic quantity, now involving *three* random variables:

Definition A.9. The *conditional mutual information of* X *and* Y, *given* Z is defined as

$$I(X;Y|Z) := H(XZ) + H(YZ) - H(XYZ) - H(Z).$$

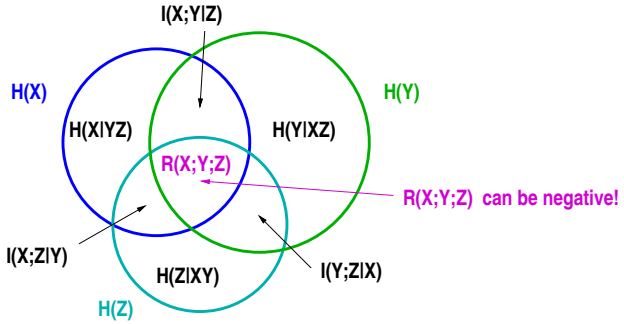

Figure A.3: Information-theoretic quantities for three random variables X, Y, and Z.

Theorem A.3. *We have*

$$I(X;Y|Z) \geq 0$$

with equality if and only if X and Y are statistically independent when given Z.

All these definitions and theorems hold when X, Y, and Z are replaced by lists of random variables.

Example A.2. Consider 5 random variables R, S, T, U, V. We have for example $H(RS) \leq H(RST)$ and $H(RSTUV) \leq H(RS) + H(TUV)$. Moreover, $H(RS|TUV) = H(RSTUV) - H(TUV)$ and and $I(RS;T|UV) = H(RSUV) + H(TUV) - H(RSTUV) - H(UV)$.

When considering a list $X_1, \ldots, X_n$ of n random variables, all information-theoretic quantities are determined by the $2^n - 1$ entropies of all (non-empty) subsets of the random variables. Figure A.3 shows the entropy diagram for three random variables X, Y, and Z. There are 7 regions, each of which corresponds to a certain combination (sum and differences) of the entropies $H(X)$, $H(Y)$, $H(Z)$, $H(XY)$, $H(XZ)$, $H(YZ)$, and $H(XYZ)$. The inner region, denoted $R(X;Y;Z)$, can be negative and does not seem to have a meaning in an application context. All other regions are non-negative.

Example A.3. Let X and Y be independent, unbiased random bits, and let $Z := X \oplus Y$. The entropy diagram of X, Y, and Z is shown in Figure A.4. Note that $I(X;Y) = 0$ and $I(X;Y|Z) = 1$.

A.3 Number Theory and Algebra Basics

We summarize a number of facts from number theory and algebra. Most proofs are omitted. A more detailed version of much of this material can be found in

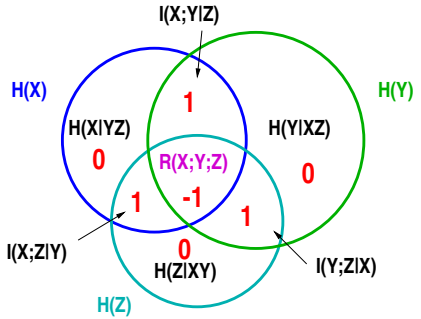

Figure A.4: Entropy diagram for Example A.3.

my lecture notes *Discrete Mathematics*.

A.3.1 Euclid's Extended GCD-Algorithm

Theorem A.4. *The algorithm described in Figure A.5 computes, for given nonnegative integers a and b with $a \geq b$ (not both 0), the integers $d = \gcd(a,b)$, as well as u and v satisfying $ua + vb = \gcd(a,b)$.*

$\sigma_1 := a;\ \sigma_2 := b;$
$u_1 := 1;\ u_2 := 0;$
$v_1 := 0;\ v_2 := 1;$
while $\sigma_2 > 0$ **do begin**
  $q := \sigma_1$ **div** σ_2;
  $r := \sigma_1 - q\sigma_2$;
  $\sigma_1 := \sigma_2;\ \sigma_2 := r;$
  $t := u_2;\ u_2 := u_1 - qu_2;\ u_1 := t;$
  $t := v_2;\ v_2 := v_1 - qv_2;\ v_1 := t;$
end;
$d := \sigma_1;\ u = u_1;\ v := v_1;$

Figure A.5: Extended gcd-algorithm (Euclid). For given nonnegative integers a and b with $a \geq b$, the algorithm computes $d = \gcd(a,b)$ as well as u and v satisfying $ua + vb = \gcd(a,b)$.

Proof. The algorithm terminates because the numbers (e.g. σ_1) decrease monotonically but remain nonnegative. The (well-known) proof that at the end of the algorithm we have $d = \gcd(a,b)$ is left as an exercise. We prove by induction

that at the beginning of the **while**-loop the two invariants

$$u_1 a + v_1 b = \sigma_1 \tag{A.1}$$

and

$$u_2 a + v_2 b = \sigma_2 \tag{A.2}$$

are always satisfied. After the initialization they are trivially satisfied. Equation (A.1) is satisfied at the end of the loop because the triple (σ_1, u_1, v_1) is assigned the value of the triple (σ_2, u_2, v_2) and therefore condition (A.1) at the end of the loop is equivalent to condition (A.2) at the beginning of the loop. Equation (A.2) holds at the end of the loop because

$$(u_1 - qu_2)a + (v_1 - qv_2)b = (u_1 a + v_1 b) - q(u_2 a + v_2 b) = r$$

holds at the beginning of the loop (by application of (A.1) and (A.2)) and because of the assignment of $(u_1 - qu_2)$, $(v_1 - qv_2)$ and r to v_1, v_2 and σ_2, respectively, within the loop. □

Modular Congruences

Definition A.10. For $a, b, m \in \mathbb{Z}$ with $m \geq 1$, we say that a *is congruent to* b *modulo* m if m divides $a - b$. We write $a \equiv b \pmod{m}$ or simply $a \equiv_m b$:

$$a \equiv_m b \iff m \mid (a - b).$$

Example A.4. We have $23 \equiv_7 44$ and $54321 \equiv_{10} 1$. Note that $a \equiv_2 b$ means that a and b are either both even or both odd. It is easy to see that $a \equiv_1 b$ is satisfied for all a and b.

The following lemma shows that modular congruences are compatible with the arithmetic operations on $\mathbb{Z}$.

Lemma A.5. *If $a \equiv_m b$ and $c \equiv_m d$, then $a + c \equiv_m b + d$ and $ac \equiv_m bd$.*

Corollary A.6. *Let $f(x_1, \ldots, x_k)$ be a multi-variate polynomial in k variables with integer coefficients, and let $m \geq 1$. If $a_i \equiv_m b_i$ for $1 \leq i \leq k$, then $f(a_1, \ldots, a_k) \equiv_m f(b_1, \ldots, b_k)$.*

There are m equivalence classes of the equivalence relation $\equiv_m$, namely $[0], [1], \ldots, [m-1]$. Each equivalence class $[a]$ has a natural representative $R_m(a) \in [a]$ in the set

$$\mathbb{Z}_m := \{0, \ldots, m-1\}$$

of remainders modulo m. Here $R_m(a)$ denotes the remainder when a is divided by m.

One is often interested only in the remainder of an integer (e.g. the result of a computation) modulo some modulus m. The following lemma establishes the simple connection between congruence modulo m and remainders modulo m. The proof is easy and left as an exercise.

Lemma A.7. *For any $a, b, m \in \mathbb{Z}$ with $m \geq 1$,*

(i) $a \equiv_m R_m(a)$.

(ii) $a \equiv_m b \iff R_m(a) = R_m(b)$.

The above lemma together with Lemma A.5 implies the following well-known fact: If in a computation involving addition and multiplication one is interested only in the remainder of the result modulo m, then one can compute remainders modulo m at any intermediate step (thus keeping the numbers small), without changing the result. This is referred to as *modular arithmetic*.

Lemma A.8. *For any $a, b, m \in \mathbb{Z}$ with $m \geq 1$,*

(i) $R_m(a + b) = R_m(R_m(a) + R_m(b))$.

(ii) $R_m(ab) = R_m(R_m(a) \cdot R_m(b))$.

A.3.2 Inverses Modulo m

Lemma A.9. *If $\gcd(a, m) = 1$, there is a unique solution $x < m$ to the congruence equation*

$$ax \equiv_m 1.$$

Proof. According to Theorem A.4 there exist integers u and v such that $ua + vm = \gcd(a, m) = 1$. Since $vm \equiv_m 0$ we have $ua \equiv_m 1$. Hence $R_m(u)$ is a solution. Suppose there is another solution u'. Then $ua - u'a \equiv_m 0$, thus $(u - u')a \equiv_m 0$ and hence m divides $(u - u')a$. Since $\gcd(a, m) = 1$, m must divide $(u - u')$, i.e., $R_m(u)$ is the unique solution in $\mathbb{Z}_m$. □

Definition A.11. If $\gcd(a, m) = 1$, the unique solution $x \in \mathbb{Z}_m$ to the congruence equation $ax \equiv_m 1$ is called the *multiplicative inverse of a modulo m*. One also uses the notation $x \equiv_m a^{-1}$.

Example A.5. The multiplicative inverse of 5 modulo 13 is 8 since $5 \cdot 8 = 40 \equiv_{13} 1$.

The multiplicative inverse can efficiently be computed using the extended Euclidean algorithm of Section A.3.1 (setting $b = m$) which yields u (and v) such that $ua + vm = \gcd(a, m) = 1$. Since $vm \equiv_m 0$ we have $ua \equiv_m 1$. Hence the multiplicative inverse of a modulo m is $R_m(u)$.

A.3.3 The Chinese Remainder Theorem

We now consider a system of congruences for an integer x. The following theorem is known as the Chinese Remainder Theorem (CRT).

Theorem A.10. *Let $m_1, m_2, \ldots, m_r$ be pairwise relatively prime integers and let $M = \prod_{i=1}^{r} m_i$. For every list $a_1, \ldots, a_r$ with $0 \le a_i < m_i$ for $1 \le i \le r$, the system of congruence equations*

$$\begin{aligned} x &\equiv_{m_1} a_1 \\ x &\equiv_{m_2} a_2 \\ &\ldots \\ x &\equiv_{m_r} a_r \end{aligned}$$

for x has a unique solution x satisfying $0 \le x < M$.

The proof shows that the solution x can actually be computed efficiently.

Proof. Let $M_i = M/m_i$. Hence $\gcd(M_i, m_i) = 1$ and thus there exists an N_i satisfying

$$M_i N_i \equiv_{m_i} 1.$$

Note that

$$M_i N_i \equiv_{m_j} 0$$

for all $j \neq i$. Hence the integer x defined by

$$x = R_M \left(\sum_{i=1}^{r} a_i M_i N_i \right)$$

satisfies all the congruences. In order to prove uniqueness, observe that for two solutions x' and x'', $x' - x'' \equiv_{m_i} 0$ for all i, i.e., $x' - x''$ is a multiple of all the m_i and hence of M. Thus $x' \equiv_M x''$. □

A.3.4 Groups

Definition A.12. A *group* $\langle G; *, e \rangle$ is a set G together with an operation $*$ on G and a special (neutral) element e such that

(1) $*$ is associative.

(2) e is a neutral element: $a * e = e * a = a$ for all $a \in G$.

(3) Every $a \in G$ has an inverse element, denoted $\widehat{a}$, i.e., $a * \widehat{a} = \widehat{a} * a = e$.

A group $\langle G; *, e \rangle$ is called *commutative* or *abelian* if $*$ commutes, i.e., if $a * b = b * a$ for all $a, b \in G$.

Some standard examples of groups are $\langle \mathbb{Z}; +, 0 \rangle$, $\langle \mathbb{Q}; +, 0 \rangle$, $\langle \mathbb{Q} \setminus \{0\}; \cdot, 1 \rangle$, $\langle \mathbb{R}; +, 0 \rangle$, $\langle \mathbb{R} \setminus \{0\}; \cdot, 1 \rangle$, and $\langle \mathbb{Z}_m; \oplus, 0 \rangle$.

Definition A.13. For a group $\langle G; *, e \rangle$ and a subset H of G, H (or, more precisely $\langle H; *, e \rangle$) is a *subgroup* of $\langle G; *, e \rangle$ if it is itself a group for the operation $*$ restricted to H.

In the remainder of this section we will use a multiplicative notation for groups, i.e., we denote the group operation as "$\cdot$" (which can also be omitted). But it is important to point out that this is only a notational convention that entails no loss of generality of the kind of group operation. In many cases (but not always) we denote the neutral element of a multiplicatively written group as 1. The inverse of a is denoted a^{-1} or $1/a$, and a/b stands for ab^{-1}. Furthermore, we use the common notation a^n for the nth power of an element a. For all $m, n \in \mathbb{Z}$ we have

$$a^m \cdot a^n = a^{m+n} \qquad \text{and} \qquad (a^m)^n = a^{mn}.$$

Definition A.14. Let G be a group and let a be an element of G. The *order* of a, denoted $\mathrm{ord}(a)$, is the least $m \ge 1$ such that $a^m = 1$, if such an m exists, and $\mathrm{ord}(a) = \infty$ otherwise.

Example A.6. By definition, $\mathrm{ord}(1) = 1$. If $\mathrm{ord}(a) = 2$ for some a, then $a^{-1} = a$. The order of 6 in $\langle \mathbb{Z}_{20}; \oplus \rangle$ is 10.

Lemma A.11. *In a finite group G, every element has a finite order.*

Definition A.15. For a finite group G and $a \in G$, the *subgroup generated by a* is[5]

$$\langle a \rangle := \{1, a, a^2, \ldots, a^{\mathrm{ord}(a)-1}\}.$$

Definition A.16. A finite group $G = \langle g \rangle$ generated by an element $g \in G$ is called *cyclic*, and g is called a *generator* of G.

Example A.7. The group $\langle \mathbb{Z}_n, \oplus \rangle$ is cyclic for every n, where 1 is a generator. The generators of $\langle \mathbb{Z}_n, \oplus \rangle$ are all $g \in \mathbb{Z}_n$ for which $\gcd(g, n) = 1$.

Lemma A.12. *A cyclic group of order n is isomorphic to $\langle \mathbb{Z}_n, \oplus \rangle$ (and hence abelian).*

Definition A.17. For a finite cyclic group $G = \langle g \rangle$ and an element $a \in G$, the unique integer $x \in \{0, \ldots, |G|-1\}$ satisfying $a = g^x$ is called the *discrete logarithm* of a to the base g.

Definition A.18. For a finite group G, the cardinality $|G|$ is called the *order* of G.

Theorem A.13 (Lagrange)**.** *Let G be a finite group and let H be a subgroup of G. Then $|H|$ divides $|G|$.*

The following corollaries are direct applications of Lagrange's theorem.

Corollary A.14. *Let G be a finite group. Then $\mathrm{ord}(a)$ divides $|G|$ for every $a \in G$.*

Corollary A.15. *Let G be a finite group. Then $a^{|G|} = e$ for every $a \in G$.*

[5] The concept of being generated by an element can naturally be generalized to infinite groups. For example, the group $\langle \mathbb{Z}; +, 0 \rangle$, is generated by 1.

Corollary A.16. *Every group of prime order*[6] *is cyclic, and every element except the neutral element is a generator.*

Definition A.19.

$$\mathbb{Z}_m^* := \{a \in \mathbb{Z}_m \mid \gcd(a, m) = 1\}.$$

The *Euler function* $\varphi : \mathbb{Z}^+ \to \mathbb{Z}^+$ is defined as the cardinality of $\mathbb{Z}_m^*$:

$$\varphi(m) = |\mathbb{Z}_m^*|.$$

Example A.8. $\mathbb{Z}_{18}^* = \{1, 5, 7, 11, 13, 17\}$. Hence $\varphi(18) = 6$.

Example A.9. If p is a prime, then $\mathbb{Z}_p^* = \{1, \ldots, p-1\} = \mathbb{Z}_p - \{0\}$.

Lemma A.17. *If* $m = \prod_{i=1}^{r} p_i^{e_i}$, *then*

$$\varphi(m) = \prod_{i=1}^{r} (p_i - 1) p_i^{e_i - 1}.$$

Lemma A.18. $\langle \mathbb{Z}_m^*; \odot \rangle$ *is a group, where* $\odot$ *denotes multiplication modulo* m.

Now we obtain the following simple but powerful corollary to Lagrange's theorem:

Corollary A.19 (Fermat, Euler). *For all* $m \geq 2$ *and all* a *with* $\gcd(a, m) = 1$,

$$a^{\varphi(m)} \equiv_m 1.$$

In particular, for every prime p *and every* a *not divisible by* p,

$$a^{p-1} \equiv_p 1.$$

A.3.5 Fast Exponentiation

Consider a general group G, an element $g \in G$, and an exponent $e \in \mathbb{Z}$. We show how g^e can be computed efficiently. Let the binary representation of the n-bit exponent e be $e_{n-1} \ldots e_1 e_0$, i.e.

$$e = \sum_{i=0}^{n-1} e_i 2^i.$$

Then g^e can be computed by the following simple algorithm often referred to as "square and multiply" algorithm or the "binary method". The result is accumulated in the variable a.

[6] i.e., $|G| = p$ for some prime p. Groups of prime order play a very important role in cryptography.


```
a := 1;
for i := n − 1 downto 0 do begin
    a := a ∗ a;
    if e_i = 1 then a := a ∗ g;
end;
```

The number of actual[7] multiplications is

$$n + w(e) - 2 \;\leq\; 2n - 2,$$

where $w(e)$ is the number of 1's in the binary representation of e. For most exponents there exist methods for exponentiation requiring fewer multiplications. For example, g^{23} is in the above procedure obtained by computing $g^2, g^4, g^5, g^{10}, g^{11}, g^{22}$ and g^{23}, but an alternative sequence of powers requiring one multiplication less is $g^2, g^3, g^5, g^{10}, g^{20}, g^{23}$.

[7] The algorithm, as described, performs also trivial multiplications (by 1).