CW 18

Outline of the Week

The following page summarizes the prerequisites, learning goals, and further readings for this week.

Work schedule for this week

- **Prerequisites**
  - ☐ Review and learn all required prerequisites
- **Preparation**
  - ☐ Read and understand the lecture notes
  - ☐ Listen and watch corresponding videos
  - ☐ Post your questions in the forum and actively answer questions in the forum
- **Guided deepening**
  - ☐ Actively participate in the deepening meeting
- **Self learning**
  - ☐ Work together in a group of up to 3 students and solve the problems
  - ☐ Present the solutions to your fellow students in class

Prerequisites

To follow the content of this week, we expect prerequisites in the following areas:

- Modular arithmetic
- Linear algebra: matrices
- Basic set theory
- Basic in algorithms: (probabilistic) algorithms, computability, complexity classes,

Learning goals

Even though you may have learned comparable content in other lectures to some extent, the goal of this week is to establish a common understanding and foundation for the following lecture. The weeks covers the following topics:

Groups The first gives an introduction to groups, which form the basis for most of the cryptographic schemes. Furthermore, we introduce elliptic curves used in the ECDSA signature process, which is part of the Bitcoin currency.

Probability Theory Most of the security analysis of cryptographic schemes and protocols require some basics in probability theory, such as discrete probability, probability set, conditional probability, random variables, expectation, and variance. We also review some important bounds, such as the union bound, Markov's inequality, and Chebyshev' inequality.

Further readings

Groups

- *An Introduction to Mathematical Cryptography*, Section 2.5, Jeffrey Hoffstein, Jill Pipher and Joseph H. Silverman, 2nd edition, Springer Publishing Company, Incorporated, 2014.
- *A Course on Group Theory*, John S. Rose, https://books.google.at/books?id=-NuQQgAACAAJ, Cambridge University Press, 1978.

Probability Theory

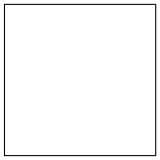
- *Lecture Notes on Topics of Mathematics in Cryptology: Probability Theory*, Dominique Unruh, University of Tartu, Estland, https://courses.cs.ut.ee/MTAT.07.025/2017_spring/uploads/Main/01-probability.pdf.

Background: Mathematics

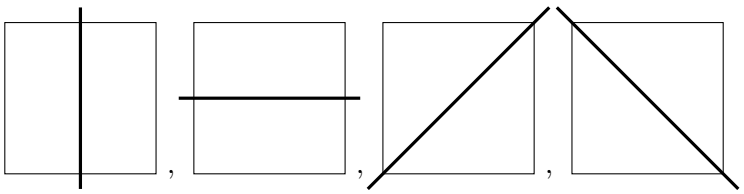
In this chapter, we introduce the mathematical concepts needed in the later chapters of these lecture notes. First, we introduce groups followed by a brief review of basic probability theory.

1 Groups

A group is a mathematical structure consisting of a set $\mathbb{G}$ and an operation $\star : \mathbb{G} \times \mathbb{G} \to \mathbb{G}$. The basic fact here is that, if we apply the group operation to two group elements, we will obtain another group element. Furthermore, the group must contain a neutral element id and, for each group element a an inverse element a^{-1} with $a \star a^{-1} = id$. An illustrating example is that of the symmetry group of a geometric object, e.g. a square.



The square has eight symmetries: Four rotational symmetries and four reflection symmetries. With respect to rotation, we can rotate the square by 90, 180, 270 or 360 degree without changing its appearance. With respect to reflection, we can reflect the square on each of the four axes depicted below.



It is obvious, that the concatenation of two of these symmetry mappings gives us another of the given 8 symmetry mappings. The neutral element is given by the rotation of 360 degree. For a rotation of i degrees ($i \in \{90, 180, 270, 360\}$), the inverse element is given by the rotation of $360 - i$ degrees. For a reflection ϕ, the inverse element of ϕ is ϕ itself. Altogether we find that the set of symmetry mappings for the square together with the composition of maps forms a group, the so called dihedral group D_4.

In the following we list some aspects why cryptographers are interested in (finite) groups.

- For many cryptographic schemes, the message space $\mathcal{M}$ and the ciphertext space $\mathcal{C}$ are chosen to be an identical finite set $\mathcal{S}$. The encryption function of the cryptosystem must be an invertible function from $\mathcal{S}$ to itself. Therefore, the encryption function must be a permutation, i.e., an element of the permutation group $\mathcal{S}_n$. Thus, even without defining a group structure on $\mathcal{M}$, groups are automatically involved in many cryptographic schemes.

- In cryptographic schemes, we often want to choose elements uniformly at random. While this is easy for finite groups, it becomes infeasible if the group contains infinitely many elements.
- Today, most cryptographic algorithms involve the use of computers. The fact that we work over finite groups enables us to store every group element using a fixed number of bits. This also allows us to find easily an upper bound for the running time and the memory consumption of the algorithms.

We now give a formal definition of the term group.

Definition 1.1 (Group)**.** A *group* $(\mathbb{G}, \star)$ consists of a set $\mathbb{G}$ and an operation $\star : \mathbb{G} \times \mathbb{G} \to \mathbb{G}$ fulfilling the following properties

- Closure: $\forall\ a, b \in \mathbb{G}$ we have $a \star b \in \mathbb{G}$
- Associativity: $a \star (b \star c) = (a \star b) \star c\ \ \forall\ \ a, b, c \in \mathbb{G}$
- neutral element: $\exists\ 1 \in \mathbb{G}$ such that $1 \star a = a \star 1 = a\ \ \forall\ \ a \in \mathbb{G}$
- inverse elements: $\forall\ \ a \in \mathbb{G}\ \exists\ a^{-1} \in \mathbb{G}$ such that $a \star a^{-1} = a^{-1} \star a = 1$.

It can be shown that the neutral element 1 of the group $\mathbb{G}$ is unique. Furthermore, for every $a \in \mathbb{G}$, the inverse element a^{-1} is uniquely determined.

The following definition introduces a special type of groups that has the property the order in which we apply the group operation does not change the result, these groups are called abelian.

Definition 1.2 (Abelian group)**.** A group $\mathbb{G}$ is called *abelian*, if it also provides commutativity, i.e.,

$$a \star b = b \star a\ \ \forall\ \ a, b \in \mathbb{G}.$$

Definition 1.3 (Subgroup)**.** A subset $\mathbb{H} \subset \mathbb{G}$ is called a *subgroup* of $\mathbb{G}$ if it, together with the operation $\star$, is a group itself.

That means for a subset $\mathbb{H} \subset \mathbb{G}$ to be a subgroup, it must be closed under the operation $\star$ and contain the neutral element 1 of the group. Furthermore, for each $a \in \mathbb{H}$ it must contain the inverse element a^{-1}. Associativity (and commutativity in case $\mathbb{G}$ is abelian) is inherited from the corresponding property of the group $\mathbb{G}$. In the following examples we use $+$ and $\cdot$ to denote group operations. This is called additive $(+)$ and multiplicative $(\cdot)$ notation and will be mostly used in the remaining script.

Examples:

- $(\mathbb{N}, +)$ is not a group since e.g. 1 has no inverse (-1 is not contained in $\mathbb{N}$.)
- $(\mathbb{Z}, +)$ is an abelian group with neutral element 0 and $-a \in \mathbb{Z}$ being the inverse element of $a \in \mathbb{Z}$.
- $(\mathbb{Q}, \cdot)$ is not a group since 0 has no multiplicative inverse. However, $(\mathbb{Q} \setminus \{0\}, \cdot)$ is a group.

- $(GL_n(\mathbb{Q}), \cdot)$ is a non abelian group, consisting of all invertible $n \times n$ matrices with rational coefficients together with the matrix multiplication. The neutral element is the identity matrix $1_{n\times n}$, the inverse element of $A \in GL_n(\mathbb{Q})$ the inverse matrix A^{-1}. Since $\det(A) \neq 0$ for all $A \in GL_n(\mathbb{Q})$, the inverse matrix A^{-1} exists in $GL_n(\mathbb{Q})$.
- For any integer $N \in \mathbb{N}$, the set $\mathbb{Z}_N = \{0, 1, \ldots, N-1\}$ together with the operation $+_N$ (addition modulo N) is a finite abelian group.

We first prove a simple "cancellation lemma" needed for the results presented in the remainder of this chapter.

Lemma 1.1. *Let $\mathbb{G}$ be a group and $a, b, c \in \mathbb{G}$. If $ac = bc$ holds, then we have $a = b$.*

Proof. We have

$$ac = bc \Rightarrow ac \cdot c^{-1} = bc \cdot c^{-1} \Rightarrow a = b.$$

□

For cryptographic applications (use of computers), we are only interested in finite groups. For a multiplicative (or additive) group $\mathbb{G}$ and $g \in \mathbb{G}$ we define

$$g^m := \underbrace{g \cdot \ldots \cdot g}_{m-\text{times}}.$$

We have $g^m \cdot g^n = g^{m+n}$, $(g^m)^n = g^{mn}$ and $g^1 = g$. If $\mathbb{G}$ is abelian, we get $g^m \cdot h^m = (gh)^m$.

Definition 1.4 (Order of a group)**.** Let $\mathbb{G}$ be a finite group with m elements. Then we call m the *order* of $\mathbb{G}$.

The following result is later used to show the correctness of cryptographic schemes such as RSA and DSA.

Theorem 1.1. *Let $\mathbb{G}$ be a finite abelian group of order m. Then, for any element $g \in \mathbb{G}$, we have $g^m = 1$.*

Proof. Let $g \in \mathbb{G}$ and let $g_1, \ldots, g_m$ be the elements of $\mathbb{G}$. Then we claim that

$$g_1 \cdot g_2 \cdot \ldots \cdot g_m = (gg_1) \cdot (gg_2) \cdot \ldots \cdot (gg_m).$$

To see this, note that $gg_i = gg_j$ implies $g_i = g_j$ (see Lemma 1.1). Therefore, the m factors on the right side of the equation are just the group elements of $\mathbb{G}$ in different order, as otherwise two group elements would be the same. This proves the correctness of the equation. Since $\mathbb{G}$ is an abelian group, we get

$$(gg_1) \cdot (gg_2) \cdot \ldots \cdot (gg_m) = g^m \cdot (g_1 \cdot \ldots \cdot g_m).$$

By Lemma 1.1, this implies $g^m = 1$. □

Thus, Theorem 1.1 also means is that if we multiply any group element by itself m times we get the neutral element. Therefore we have that $g^{m-1} = g^{-1}$. Note that this does not mean that m is the smallest number such that $g^{m-1} = g^{-1}$

The following proposition is an immediate consequence of Theorem 1.1.

Proposition 1.1. *Let $\mathbb{G}$ be a finite group with $m = |\mathbb{G}| > 1$. Then, for any $g \in \mathbb{G}$ and any integer x, we have $g^x = g^{x \mod m}$.*

Proof. Say $x = qm + r$, where q and r are integers and $r = x \mod m$. Then we have

$$g^x = g^{qm+r} = (g^m)^q \cdot g^r = 1^q \cdot g^r = g^r.$$

□

1.1 Cyclic Groups

Let $\mathbb{G}$ be a finite group of order m and $g \in \mathbb{G}$. We define

$$\langle g \rangle = \{g^0, g^1, \dots\}.$$

Since the group $\mathbb{G}$ has order m, we know from Theorem 1.1 that $g^m = 1$ holds. Therefore, the set $\langle g \rangle$ is finite. In many cases however, there exists an integer $i < m$ such that $g^i = 1$. We define

Definition 1.5 (Order of a (cyclic) group)**.** The smallest integer i such that $g^i = 1$ is called the **order** of g.

As already stated above, the order of any group element $g \in \mathbb{G}$ is smaller or equal to the order m of the group $\mathbb{G}$.

Let now i be the smallest integer such that $g^i = 1$. Then we have $g^{i+j} = g^i \cdot g^j = g^j$ which implies that the sequence

$$g^0, g^1, \dots$$

has period i. Therefore, the set $\langle g \rangle$ has i elements and is closed under multiplication. Since we also have $1 = g^0 \in \langle g \rangle$ and, for every $g^a \in \langle g \rangle$, the inverse $(g^a)^{-1} = g^{i-a}$ is contained in $\langle g \rangle$, $\langle g \rangle$ is a subgroup of $\mathbb{G}$.

Definition 1.6 (Cyclic group and generator)**.** A group $\mathbb{G}$ is called *cyclic*, if it can be generated by a single element $g \in \mathbb{G}$, i.e. $\mathbb{G} = \langle g \rangle$. We call g a *generator* of the group $\mathbb{G}$.

Obviously, every cyclic group is abelian. Figure 2 shows the structure of a cyclic group.

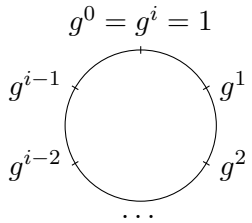


Figure 2: Structure of a cyclic group

The following two propositions are true for any finite group.

Proposition 1.2. *Let $\mathbb{G}$ be a finite group, and $g \in \mathbb{G}$ be an element of order i. Then $g^x = g^y$ if and only if $x = y \mod i$.*

Proof. "⇐": If $x = y \mod i$ holds, we can write $x = q \cdot i + y$ with $q \in \mathbb{Z}$. Therefore,

$$g^x = g^y \cdot (g^i)^q = g^y \cdot 1^q = g^y.$$

"⇒": Let $g^x = g^y$. Then, $1 = g^{x-y} = g^{(x-y) \mod i}$. Since $[(x-y) \mod i] < i$, but i is according to the definition the smallest positive integer such that $g^i = 1$, we must have $(x-y) \mod i = 0$. □

Proposition 1.3. *Let $\mathbb{G}$ be a finite group of order m, and say $g \in \mathbb{G}$ has order i. Then i divides m.*

Proof. We have $g^m = g^0 = 1$. From Proposition 1.2 it follows that $[m \mod i] = 0$ or $i|m$. □

In other words, Proposition 1.3 states that, in any finite group $\mathcal{G}$, the order of any group element $g \in \mathbb{G}$ divides the order of the group $\mathbb{G}$.

The next two theorems give some examples for cyclic groups.

Theorem 1.2. *If $\mathbb{G}$ is a group of prime order p, then $\mathbb{G}$ is cyclic. Moreover, all elements in the group except the neutral element are generators of $\mathbb{G}$.*

Proof. By Proposition 1.3, the only possible orders of elements in $\mathbb{G}$ are 1 and p. Only the neutral element has order 1, and so all the other elements have order p and therefore are generators of $\mathbb{G}$. □

In particular, for a prime p, the additive group $(\mathbb{Z}_p, +)$ is a cyclic group. Every element of $\mathbb{Z}_p \setminus \{0\}$ is a generator of the group.

Theorem 1.3. *For any prime p, the multiplicative group $\mathbb{Z}_p^\star$ is a cyclic group.*

Proof. See [23] for a proof in number theory. Note that the multiplicative group $\mathbb{Z}_p^\star$ is mentioned as a modulo multiplication group M_p therein. □

The generators of $\mathbb{Z}_p^\star$ are called the **primitive roots** modulo p. The order of the group $\mathbb{Z}_p^\star$ is $p-1$, which is, for $p > 3$, not a prime number. If $p-1$ contains small factors, the discrete logarithm problem (see **??**) in the group $\mathbb{G}$ can be solved more efficiently by the Pohlig-Hellman algorithm (see for example [14], Section 2.9). In cryptography, one therefore often works over subgroups of $\mathbb{Z}_p^\star$ of prime order q. It is easy to see that each of these subgroups is a cyclic group itself.

1.2 The Elliptic Curve Group

Another group often used in cryptographic applications is the group of points on an elliptic curve over the finite field $\mathbb{Z}_p$. An elliptic curve $E(\mathbb{Z}_p)$ consists of all (discrete) tuples $(x, y) \in \mathbb{Z}_p \times \mathbb{Z}_p$ (called points) such that

$$y^2 = x^3 + Ax + B \mod p$$

with $A, B \in \mathbb{Z}_p$ $(p \geq 5)$.
We can show: Each straight line intersects the elliptic curve $E(\mathbb{Z}_p)$ in exactly three points. Hereby,

- A tangent point counts twice
- If the line is vertical, we also count the point at infinity $\mathcal{O}$.

With this, we can define an addition law for the points on the elliptic curve $E(\mathbb{Z}_p)$ as follows (see also Figure 3):

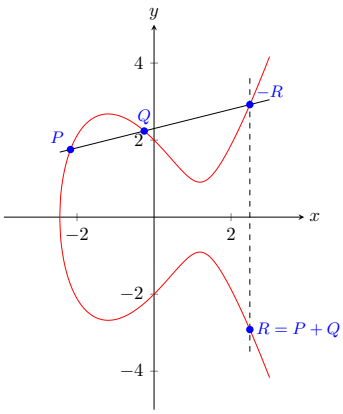


Figure 3: Addition of Points on an Elliptic Curve (over $\mathbb{R}$)

- The point at infinity $\mathcal{O}$ is defined to be the additive identity, i.e. we have

$$P + \mathcal{O} = \mathcal{O} + P = P \text{ for all } P \in E(\mathbb{Z}_p).$$

- For two points $P, Q \neq \mathcal{O}$ on $E(\mathbb{Z}_p)$, we evaluate the sum $P + Q$ by drawing the line through P and Q (if $P = Q$, then we draw the line tangent to the curve at P) and finding the third point of intersection $-R$ of this line with $E(\mathbb{Z}_p)$. If the line is vertical, the point $-R$ is defined to be $\mathcal{O}$. If $-R = (x, y) \neq \mathcal{O}$, we define $P + Q := R = (x, -y)$; if $-R = \mathcal{O}$, we set $P + Q = \mathcal{O}$.

Theorem 1.4. *The points on the elliptic curve $E(\mathbb{Z}_p)$ form, together with the above addition law, an abelian group.*

Proof. Closure, associativity and commutativity follow from the geometric interpretation of the addition law (see above). The neutral element is the point $\mathcal{O}$. The inverse element of $P = (x, y) \in E(\mathbb{Z}_p)$ is $-P = (x, -y) \in \mathbb{E}(\mathbb{Z}_p)$. Indeed, the line through P and $-P$ is vertical which means that the third point of intersection and therefore the sum of P and Q is the identity $\mathcal{O}$. □

From the geometric interpretation of the addition law we find the following algebraic formulation of the group law.

Proposition 1.4. *Let $p \geq 5$ be a prime and let E be the elliptic curve given by $y^2 = x^3 + Ax + B \mod p$ where $4A^3 + 27B^2 \neq 0 \mod p$. Let $P, Q \neq \mathcal{O}$ be points on E, with $P = (x_1, y_1)$ and $Q = (x_2, y_2)$. Then we have*

1. *If $x_1 \neq x_2$, then $P + Q = (x_3, y_3)$ with $x_3 = (m^2 - x_1 - x_2) \mod p$ and $y_3 = (m \cdot (x_1 - x_3) - y_1) \mod p$, where $m = \frac{y_2 - y_1}{x_2 - x_1} \mod p$.*

2. *If* $x_1 = x_2$, *but* $y_1 \neq y_2$, *then* $P + Q = \mathcal{O}$.
3. *If* $P = Q$ *and* $y_1 = 0$ *then* $P + Q = 2P = \mathcal{O}$
4. *If* $P = Q$ *and* $y_1 = 0$ *then* $P + Q = 2P = (x_3, y_3)$ *with* $x_3 = (m^2 - 2x_1) \mod p$ *and* $y_3 = (m \cdot (x_1 - x_3) - y_1) \mod p$, *where* $m = \frac{3x_1^2 + A}{2y_1} \mod p$.

The number of points on the elliptic curve $E(\mathbb{Z}_p)$ and therefore the group order is, in general, hard to find. However, we have

Theorem 1.5 (Hasse Bound)**.** *Let* p *be prime and let* E *be an elliptic curve over* $\mathbb{Z}_p$. *Then we have*

$$p + 1 - 2\sqrt{p} \leq |E(\mathbb{Z}_p)| \leq p + 1 + 2\sqrt{p}.$$

The group of $E(\mathbb{Z}_p)$ (resp. a cyclic subgroup of this group) is used as basis for the ECDSA signature scheme.

2 Discrete Probability

Analyzing the security of cryptographic schemes inherently requires some basic knowledge in probability theory. Probability theory is the scientific way of measuring uncertainty. In this section we review basic notations, concepts, and facts.

2.1 Sample space, events, probability

In the following, we provide the definitions of a probability set consisting of a sample space, the event set, and the probability distribution. Intuitively, we are considering a (random) experiment where the outcome of the experiment involves some uncertainty. For example, we might be interested in an experiment in which a we throw a (fair) coin. The outcome of this experiment is either either heads or tails. The definition of a probability set can be used to describe all components of this experiment:

Sample space The sample space Ω is the set of all possible *outcomes* or *elementary events.*

Probability distribution The probability distribution is a function

$$\Pr : \Omega \to \mathbb{R},$$

that assigns each element from the sample a value in $\mathbb{R}$, such that

$$\Pr[\omega] \geq 0 \text{ for all } \omega \in \Omega,$$

and the sum of all individual probabilities is 1:

$$\sum_{\omega \in \Omega} \Pr[\omega] = 1.$$

Event When we describe an experiment that might have different outcomes. We refer to the *set of possible outcomes* as an *event* and we denote this set by Σ.

Definition 2.1 (Probability Set)**.** A *probability set* consists of three components $(\Omega, \Sigma, \Pr)$ where

- Ω is the set of possible outcomes called *sample space*,
- Σ is a set of subsets of Ω called *event set* and
- Pr is a function $\Pr : \Sigma \to [0,1]$ with
  - $\Pr[\Omega]] = 1$ and
  - $\Pr[\bigcup_{i=1}^{\infty} \sigma_i] = \sum_{i=1}^{\infty} \Pr[\sigma_i] \quad \forall\, \sigma_i \in \Sigma,\ \sigma_i \cap \sigma_j = \emptyset\ \forall\, i \neq j$

  which is called *probability (distribution).*

Examples The following examples illustrate the usage of the definition of a probability set.

1. A coin has two sides called heads H and tail T. We are considering the game of throwing a fair, i.e., both sides have the same probability, and we are going to formalize the underlying random experiment. The sample space of a coin is $\Omega = \{H, T\}$ and the two possible events are H and T. We will identify H with 1 and T with 0. Since we are considering a fair coin, each of the two events have the same probability, i.e., $\Pr[H] = \Pr[T]] = 1/2$.
2. In the next example, we repeat the process of throwing a fair coin ℓ-times independently, i.e., each throw does not depend on the outcome of previous (throws) and has the same probability, then $\Omega = \{0,1\}^{\ell}$. For each element $x \in \Omega$ we have $\Pr[x] = 2^{\ell}$. Throwing a fair coin ℓ times has the same probability as sampling an ℓ bit string uniformly at random.

Definition 2.2 (Uniform distribution)**.** A probability distribution Pr is *uniform* if each elementary event has the same probability.

Conjunction, disjunction, and the independence of events When analyzing probabilities, we often have the case the certain events occur together, or exclude each other, and sometimes the are independent of each other. If A and B are two events, then $A \wedge B$ is the conjunction of A and B, i.e., the event that *both* events occur. The disjunction of A and B is denoted by $A \vee B$ and it means the event that *either* A or B occurs.

The following proposition summarizes some basic facts that help to compute probabilities. In this proposition, we are also deal with the complement of an event. If E is an event, then $\bar{E}$ is the complement of E, which means the event that $\bar{E}$ does not occur.

Proposition 2.1. *Let Ω be a finite discrete probability space and A and B two events of Ω. Then*

1. $\Pr[A \wedge B] \leq \Pr[A]$
2. $\Pr[A \vee B] \geq \Pr[A]$
3. $\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B]$
4. $\Pr[A \cup B] = \Pr[A] + \Pr[B]$ *if A and B are disjoint, i.e., $A \cap B = \emptyset$*

5. $\Pr\left[\bar{A}\right] = 1 - \Pr[A]$
6. $\Pr[A] \leq \Pr[B]$ *if* $A \subseteq B$

The following definition describes the independence of two events.

Definition 2.3 (Independence of events)**.** Let $(\Omega, \Pr)$ be a finite discrete probability space and A and B two events of Ω. The events A and B are *independent* if

$$\Pr[A \wedge B] = \Pr[A] \cdot \Pr[B].$$

one of the most used bounds in cryptography to estimate probabilities is the union bound. It essentially says that the probability that at least one of two events A and B happens, is bounded by the sum of the two individual events.

Proposition 2.2 (Union bound)**.** *Let* $(\Omega, \Pr)$ *be a finite discrete probability space and* A *and* B *two events of* Ω*, then we have*

$$\Pr[A \vee B] \leq \Pr[A] + \Pr[B]$$

This bound can easily be generalized by repeating its application to an events $E_1, E_2, \ldots, E_\ell \in \Omega$ *as follows:*

$$\Pr\left[\bigvee_{i=1}^{\ell} E_i\right] \leq \sum_{i=1}^{\ell} \Pr\left[E_i\right].$$

2.2 Conditional probability

In many cases one event depends on another event that occurred before. Formally, we express this as conditional probability, which is defined as follows.

Definition 2.4 (Conditional probability)**.** Let $(\Omega, \Pr)$ be a finite probability space and A and B be two events. The *conditional probability* of A given B is defined as

$$\Pr[A \mid B] = \frac{\Pr[A \wedge B]}{\Pr[B]},$$

as long as $\Pr[B] \neq 0$ (for $\Pr[B] = 0$ the value $\Pr[A \mid B]$ is undefined).

By simple arithmetic operations, we obtain

$$\Pr[A \wedge B] = \Pr[A \mid B] \cdot \Pr[B],$$

which allows us to derive Bayes' theorem:

Theorem 2.1 (Bayes theorem)**.** *Let* $(\Omega, \Pr)$ *be a finite discrete probability space and* A *and* B *two events of* Ω*. If* $\Pr[B] > 0$*, then*

$$\Pr[A \mid B] = \frac{\Pr[B \mid A] \cdot \Pr[A]}{\Pr[B]}.$$

Proof. To see that the theorem holds, observe that:

$$\Pr[A \mid B] = \frac{\Pr[A \wedge B]}{\Pr[B]} = \frac{\Pr[B \wedge A]}{\Pr[B]} = \frac{\Pr[B \mid A] \cdot \Pr[A]}{\Pr[B]}.$$

□

The next theorem is known as the *law of total probability*. The basic idea is that if the probability of an event is unknown, then it can be calculated using the known probabilities of several *distinct* events.

Theorem 2.2 (Law of total probability)**.** *Let* $(\Omega, \Pr)$ *be a finite probability space and let* $E_1, E_2, \ldots, E_\ell$ *be events that* partition *the space of all possible events, i.e.,* $E_1 \cup E_2 \cup \cdots \cup E_n = \Omega$ *and* $E_i \cap E_j = 0$ *for all* $i \neq j$*. Then, for any event* F *it holds that*

$$\Pr[F] = \sum_{i=1}^{\ell} \Pr[F \wedge E_i] = \sum_{i=1}^{\ell} \Pr[F \mid E_i] \cdot \Pr[E_i].$$

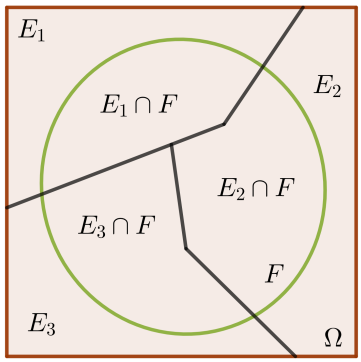


Figure 4: The law of total probability for $\ell = 3$.

Examples:

Your team is on the way to an E-Sport tournament playing your favorite game. An analysis tool tells that your team can win top 20% teams (class 1) with probability 0.3, and win the next 30% next teams (class 2) with probability 0.4, and win the rest half teams with probability 0.55 (class 3). Let E_1, E_2 and E_3 be the events for you to meet the team class 1, 2 and 3, respectively and W be the event that you win. Then

$$\Pr[W] = \sum_{i=1}^{3} \Pr[W \mid E_i] \cdot \Pr[E_i] = 0.3 \cdot 0.2 + 0.4 \cdot 0.3 + 0.55 \cdot 0.5 = 0.455.$$

2.3 Random variable

Intuitively, a random variable assigns a random process to some value. In cryptography these processes are often security games in which the adversary interacts with a challenger and the adversary's success is measured and mapped to some value.

Definition 2.5 (Random variable)**.** Let $(\Omega, \Pr)$ be a finite probability space. A function $X : \Omega \to \mathbb{R}$ is called a *random variable*.

We are often interested in the set of elementary events that map to a certain value r and we use the notation $\Pr[X = r]$ to refer to this set (instead of $X^{-1}(r)$). Analogously, $\Pr[X \leq r]$ refers to the probability of the event $X^{-1}((\infty, r])$. In the following we introduces the notion of a probability mass function $f(r)$, which can be seen as a list of all possibility for each state.

Definition 2.6 (Probability mass function)**.** Let X be a random variable on a finite probability space. The function

$$f : \mathbb{R} \to [0, 1] \text{ defined by } f(r) = \Pr[X = r]$$

is called the *probability mass function* of X.

Similar, if we are interested in a bounded interval of possible values, i.e., $\Pr[X \leq r]$, then we introduce the notion of a cumulative distribution function of X.

Definition 2.7 (cumulative distribution function)**.** Let X be a random variable on a finite probability space. The function

$$F : \mathbb{R} \to [0, 1] \text{ defined by } F(r) = \Pr[X \leq r]$$

is called the *cumulative distribution function* of X.

2.3.1 Expected value

The expected value of a random variable can be seen as the weighted average of possible outcomes and is defined as follows:

Definition 2.8 (Expected value)**.** Let $(\Omega, \Pr)$ be a finite discrete probability space and X be a random variable that space, which takes values in a set S. Then the *expected value* or *expectation* of a random variable X is

$$\mathsf{Exp}[X] = \sum_{s \in S} s \cdot \Pr[X = s].$$

One should not confuse the notation $\mathsf{Exp}[X]$ of the expected value here with the exponent function e^X. The following theorem shows that the linearity of the expectation.

Theorem 2.3. *Let $(\Omega, \Pr)$ be a finite discrete probability space and X_1 and X_2 be two random variable on that space. If $\lambda \in \mathbb{R}$ is a scalar, then*

$$\begin{aligned} \mathsf{Exp}\left[X_1 + X_2\right] &= \mathsf{Exp}\left[X_1\right] + \mathsf{Exp}\left[X_2\right]. \\ \mathsf{Exp}\left[\lambda X_1\right] &= \lambda \mathsf{Exp}\left[X_1\right]. \end{aligned} \tag{1}$$

If two random variables X_1 and X_2 are independent, then we have

$$\mathsf{Exp}[X_1 \cdot X_2] = \mathsf{Exp}[X_1] \cdot \mathsf{Exp}[X_2].$$

2.3.2 Variance

The variance is a measurement that describes how much the random variable X deviates from its expectation $\mathsf{Exp}[X]$.

Definition 2.9 (Variance)**.** Let $(\Omega, \Pr)$ be a finite discrete probability space and X be a random variable in that space. The *variance* $\mathsf{Var}[X]$ is defined as

$$\mathsf{Var}[x] = \mathsf{Exp}[(X - \mathsf{Exp}(X))^2] = \mathsf{Exp}[X^2] - \mathsf{Exp}[X]^2.$$

Examples:

- We consider the Binomial distribution. Let $(\Omega, \Pr)$ be a finite discrete probability space and $X_1, \ldots, X_n$ be mutually independent Bernoulli $B(1, p)$ random variables in that space (i.e., $\Omega = \{0, 1\}$ and $\Pr[X_i = 1] = p$ for all i). Let $X = X_1 + \cdots + X_n$, then X is called the Binomial random variable. Intuitively, we treat each X_i as an identical Boolean trial with two outcome, 0/failure and 1/success, then X is the number of success times after n trials.

  The probability mass function of X is $f_X(r) = \binom{n}{r} p^r (1-p)^{n-r}$.

  The expected value of X is

$$\begin{aligned}\mathsf{Exp}[X] = \sum_{r=1}^{n} r\binom{n}{r} p^r (1-p)^{n-r} &= \sum_{r=1}^{n} n\binom{n-1}{r-1} p^r (1-p)^{n-r}. \\ &= np \sum_{r=1}^{n} \binom{n-1}{r-1} p^{r-1} (1-p)^{(n-1)-(r-1)} \\ &= np(p + 1 - p)^{n-1} = np.\end{aligned}$$

  Another way to compute the expected value is that $\mathsf{Exp}[X_i] = 0 \cdot (1-p) + 1 \cdot p = p$ and then by linearity of expected value, $\mathsf{Exp}[X] = \mathsf{Exp}[X_1] + \cdots + \mathsf{Exp}[X_n] = np$.

  We can further compute $\mathsf{Exp}[X^2]$ from the probability mass function

$$\mathsf{Exp}[X] = \sum_{r=0}^{n} r^2 C_r^n p^r (1-p)^{n-r} = np(1-p) + n^2 p^2.$$

  The variance of X is $\mathsf{Var}[X] = \mathsf{Exp}[X^2] - \mathsf{Exp}[X]^2 = np(1-p)$.

- Another example we consider is the geometrical distribution. Extend the sequence $X_1, X_2, \ldots$ to a finite sequence and let Y be the number of failures until the first success, i.e., $Y = i$ such that $X_j = 0$ for all $j < i$ and $X_i = 1$.

  The probability mass function of Y is $f_Y(r) = p(1-p)^{r-1}$ for $r \in \mathbb{Z}^+$.

  The expected value and the variance of Y is

$$\mathsf{Exp}[Y] = \sum_{r=1}^{\infty} rp(1-p)^{r-1} = \frac{1}{p},$$

$$\mathsf{Var}[Y] = \frac{1-p}{p^2}$$

  by computing the generating functions. See [13] for full proofs.

2.4 Useful bounds

In this section, we review some bounds that we need in this lecture. They are called concentration inequalities or tail bounds in the probability theory. The first one is Markov's inequality, which is useful when not many information is known about a random variable X. Markov's inequality gives an upper bound for the probability that a random variable is greater than or equal to some positive value.

Proposition 2.3 (Markov's inequality)**.** *Let $(\Omega, \Pr)$ be a finite discrete probability space and X be a non-negative random variable in that space and let $v > 0$. Then,*

$$\Pr[X \geq v] \leq \frac{\mathsf{Exp}[X]}{v}.$$

The next inequality is known as Chebyshev's inequality. The inequality gives an estimate for the deviation of a random variable by more than a given threshold value from its expected value.

Proposition 2.4 (Chebyshev's inequality)**.** *Let $(\Omega, \Pr)$ be a finite discrete probability space and X be a random variable in that space and let $\delta > 0$. Then,*

$$\Pr[X - \mathsf{Exp}[X] \geq \delta] \leq \frac{\mathsf{Var}[X]}{\delta^2}.$$

For some more specific probability distribution, we obtain some better bounds on the tail probability. The following inequality gives a tail bound for the Binomial distribution.

Proposition 2.5 (Chernoff's inequality)**.** *Let $(\Omega, \Pr)$ be a finite discrete probability space and $X_1, \ldots, X_n$ be mutually independent Bernoulli $B(1, p)$ random variables in that space (i.e., $\Omega = \{0, 1\}$ and $\Pr[X_i = 1] = p$ for all i) and let $\delta > 0$. Let $X = X_1 + \cdots + X_n$, then $\mu = \mathsf{Exp}[X] = pn$ (X is called the Binomial distribution). Then we have*

$$\Pr[X \geq (1-\delta)\mu] \leq e^{\frac{-\delta^2\mu}{2}} \text{ and } \Pr[X \leq (1+\delta)\mu] \leq e^{\frac{-\delta^2\mu}{3}}.$$

The following is a tail bound specific for a random variable which is a Lipschitz function of independent random variables.

Definition 2.10. A n-variable function f is called k-Lipschitz if $|f(x) - f(x')| \leq k$, whenever x and x' differ in at most one coordinate.

Proposition 2.6 (Azuma's inequality)**.** *Let f be an n-variable Lipschitz function and $X_1, \cdot, X_n$ be independent random variables. Then we have*

$$\Pr[f \geq \mathsf{Exp}(f) + t] \leq e^{\frac{-2t^2}{nk^2}} \text{ and } \Pr[f \leq \mathsf{Exp}(f) - t] \leq e^{\frac{-2t^2}{nk^2}}.$$