1

Preliminaries

1.1 Probability Theory

This section summarizes the fundamental notions of probability theory and some results which we will need in the following chapters. In no way is it intended to serve as a substitute for a course in probability theory.

1.1.1 Definition. *A* **probability space**[1] *is a triple* $(\Omega, \Sigma, \mathrm{P})$*, where* Ω *is a set,* $\Sigma \subseteq 2^\Omega$ *is a* σ*-algebra on* Ω *(a collection of subsets containing* Ω *and closed on complements, countable unions and countable intersections), and* P *is a countably additive measure*[2] *on* Σ *with* $\mathrm{P}[\Omega] = 1$*. The elements of* Σ *are called* **events**[3] *and the elements of* Ω *are called* **elementary events***. For an event* A*,* $\mathrm{P}[A]$ *is called the* **probability** *of* A*.*

In this text, we will consider mostly *finite probability spaces* where the set of elementary events Ω is finite and $\Sigma = 2^\Omega$. Then the probability measure is determined by its values on elementary events; in other words, by specifying a function $p : \Omega \to [0, 1]$ with $\sum_{\omega\in\Omega} p(\omega) = 1$. Then the probability measure is given by $\mathrm{P}[A] = \sum_{\omega\in A} p(\omega)$.

The basic example of a probability measure is the *uniform distribution*[4] on Ω, where

$$\mathrm{P}[A] = \frac{|A|}{|\Omega|} \quad \text{for all } A \subseteq \Omega.$$

Such a distribution represents the situation where any outcome of an experiment (such as rolling a die)[5] is equally likely.

1.1.2 Definition (Random graphs). [6] *The probability space of random graphs* $G(n, p)$ *is a finite probability space whose elementary events are all graphs on a fixed set of* n *vertices, and where the probability of a graph with* m *edges is*

$$p(G) = p^m(1 - p)^{\binom{n}{2}-m}.$$

[1] probability space = pravděpodobnostní prostor
[2] measure = míra
[3] event = jev
[4] uniform distribution = rovnoměrné rozdělení
[5] rolling a die = hod kostkou
[6] random graph = náhodný graf

This corresponds to generating the random graph by including every potential edge independently with probability p. For $p = \frac{1}{2}$, we toss a fair coin[7] for each pair $\{u, v\}$ of vertices and connect them by an edge if the outcome is heads.[8] [9]

Here is an elementary fact which is used all the time:

1.1.3 Lemma. *For any collection of events $A_1, \ldots, A_n$,*

$$\mathrm{P}\Big[\bigcup_{i=1}^{n} A_i\Big] \leq \sum_{i=1}^{n} \mathrm{P}[A_i].$$

Proof. For $i = 1, \ldots, n$, we define

$$B_i = A_i \setminus (A_1 \cup A_2 \cup \ldots \cup A_{i-1}).$$

Then $\bigcup B_i = \bigcup A_i$, $\mathrm{P}[B_i] \leq \mathrm{P}[A_i]$, and the events $B_1, \ldots, B_n$ are disjoint. By additivity of the probability measure,

$$\mathrm{P}\Big[\bigcup_{i=1}^{n} A_i\Big] = \mathrm{P}\Big[\bigcup_{i=1}^{n} B_i\Big] = \sum_{i=1}^{n} \mathrm{P}[B_i] \leq \sum_{i=1}^{n} \mathrm{P}[A_i].$$

□

1.1.4 Definition. *Events A, B are* **independent**[10] *if*

$$\mathrm{P}[A \cap B] = \mathrm{P}[A]\,\mathrm{P}[B]\,.$$

More generally, events $A_1, A_2, \ldots, A_n$ are **independent** *if for any subset of indices $I \subseteq [n]$*

$$\mathrm{P}\Big[\bigcap_{i \in I} A_i\Big] = \prod_{i \in I} \mathrm{P}[A_i].$$

We use the convenient notation $[n]$ for the set $\{1, 2, \ldots, n\}$.

The independence of $A_1, A_2, \ldots, A_n$ is not equivalent to all the pairs A_i, A_j being independent. Exercise: find three events A_1, A_2 and A_3 that are pairwise independent but not mutually independent.

Intuitively, the property of independence means that the knowledge of whether some of the events $A_1, \ldots, A_n$ occurred does not provide any information regarding the remaining events.

1.1.5 Definition (Conditional probability). *For events A and B with $\mathrm{P}[B] > 0$, we define the conditional probability*[11] *of A, given that B occurs, as*

$$\mathrm{P}[A|B] = \frac{\mathrm{P}[A \cap B]}{\mathrm{P}[B]}.$$

[7] toss a fair coin = hodit spravedlivou mincí
[8] heads = líc (hlava)
[9] tails = rub (orel)
[10] independent events = nezávislé jevy
[11] conditional probability = podmíněná pravděpodobnost

Note that if A and B are independent, then $\mathrm{P}[A|B] = \mathrm{P}[A]$.

1.1.6 Definition (Random variable). *A real random variable*[12] *on a probability space* $(\Omega, \Sigma, \mathrm{P})$ *is a function* $X\colon \Omega \to \mathbf{R}$ *that is* P*-measurable. (That is, for any* $a \in \mathbf{R}$, $\{\omega \in \Omega\colon X(\omega) \le a\} \in \Sigma$.*)*

We can also consider random variables with other than real values; for example, a random variable can have complex numbers or n-component vectors of real numbers as values. In such cases, a random variable is a measurable function from the probability space into the appropriate space with measure (complex numbers or $\mathbf{R}^n$ in the examples mentioned above). In this text, we will mostly consider real random variables.

1.1.7 Definition. *The* **expectation**[13] *of a (real) random variable* X *is*

$$\mathbf{E}[X] = \int_\Omega X(\omega)\,\mathrm{dP}(\omega).$$

Any real function on a finite probability space is a random variable. Its expectation can be expressed as

$$\mathbf{E}[X] = \sum_{\omega\in\Omega} p(\omega)X(\omega).$$

1.1.8 Definition (Independence of variables). *Real random variables* X, Y *are independent if we have, for every two measurable sets* $A, B \subseteq \mathbf{R}$,

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

Note the shorthand notation for the events in the previous definition: For example, $\mathrm{P}[X \in A]$ stands for $\mathrm{P}[\{\omega \in \Omega\colon X(\omega) \in A\}]$.

Intuitively, the independence of X and Y means that the knowledge of the value attained by X gives us no information about Y, and vice versa. In order to check independence, one need not consider all measurable sets A and B; it is sufficient to look at $A = (-\infty, a]$ and $B = (-\infty, b]$. That is, if

$$\mathrm{P}[X \le a \text{ and } Y \le b] = \mathrm{P}[X \le a]\,\mathrm{P}[Y \le b]$$

for all $a, b \in \mathbf{R}$, then X and Y are independent.

As we will check in Chapter 3, $\mathbf{E}[X + Y] = \mathbf{E}[X] + \mathbf{E}[Y]$ holds for *any* two random variables (provided that the expectations exist). On the other hand, $\mathbf{E}[XY]$ is generally different from $\mathbf{E}[X]\,\mathbf{E}[Y]$. But we have

1.1.9 Lemma. *If* X *and* Y *are independent random variables, then*

$$\mathbf{E}[XY] = \mathbf{E}[X] \cdot \mathbf{E}[Y].$$

[12] random variable = náhodná proměnná
[13] expectation = **střední hodnota!!!**

Proof (for finite probability spaces). If X and Y are random variables on a finite probability space, the proof is especially simple. Let V_X, V_Y be the (finite) sets of values attained by X and by Y, respectively. By independence, we have $\mathrm{P}[X = a \text{ and } Y = b] = \mathrm{P}[X = a]\,\mathrm{P}[Y = b]$ for any $a \in V_X$ and $b \in V_Y$. We calculate

$$\begin{aligned}\mathbf{E}[XY] &= \sum_{a\in V_X, b\in V_Y} ab \cdot \mathrm{P}[X = a \text{ and } Y = b] \\ &= \sum_{a\in V_X, b\in V_Y} ab \cdot \mathrm{P}[X = a]\,\mathrm{P}[Y = b] \\ &= \Big(\sum_{a\in V_X} a\,\mathrm{P}[X = a]\Big)\Big(\sum_{b\in V_Y} b\,\mathrm{P}[Y = b]\Big) = \mathbf{E}[X]\,\mathbf{E}[Y]\,.\end{aligned}$$

For infinite probability spaces, the proof is formally a little more complicated but the idea is the same. □

1.2 Useful Estimates

In the probabilistic method, many problems are reduced to showing that certain probability is below 1, or even tends to 0. In the final stage of such proofs, we often need to estimate some complicated-looking expressions. The golden rule here is to start with the roughest estimates, and only if they don't work, one can try more refined ones. Here we describe the most often used estimates for basic combinatorial functions.

For the factorial function $n!$, we can often do with the obvious upper bound $n! \le n^n$. More refined bounds are

$$\left(\frac{n}{e}\right)^n \le n! \le en\left(\frac{n}{e}\right)^n$$

(where $e = 2.718281828\ldots$ is the basis of natural logarithms), which can be proved by induction. The well-known Stirling formula is very seldom needed in its full strength.

For the binomial coefficient $\binom{n}{k}$, the basic bound is $\binom{n}{k} \le n^k$, and sharper ones are

$$\left(\frac{n}{k}\right)^k \le \binom{n}{k} \le \left(\frac{en}{k}\right)^k.$$

For all k, we also have $\binom{n}{k} \le 2^n$. Sometimes we need better estimates of the middle binomial coefficient $\binom{2m}{m}$; we have

$$\frac{2^{2m}}{2\sqrt{m}} \le \binom{2m}{m} \le \frac{2^{2m}}{\sqrt{2m}}$$

(also see Section 5.2 for a derivation of a slightly weaker lower bound).

Very often we need the inequality $1+x \le e^x$, valid for all real x. In particular, for bounding expressions of the form $(1-p)^m$ from above, with $p > 0$ small, one uses

$$(1-p)^m \le e^{-mp}$$

almost automatically. For estimating such expressions from below, which is usually more delicate, we can often use

$$1 - p \geq e^{-2p},$$

which is valid for $0 \leq p \leq \frac{1}{2}$.

2

The Probabilistic Method

The probabilistic method is a remarkable technique for proving the existence of combinatorial objects with specified properties. It is based on probability theory but, surprisingly, it can be used for proving theorems that have nothing to do with probability. The usual approach can be described as follows.

We would like to prove the existence of a combinatorial object with specified properties. Unfortunately, an explicit construction of such a "good" object does not seem feasible, and maybe we do not even need a specific example; we just want to prove that something "good" exists. Then we can consider a random object from a suitable probability space and calculate the probability that it satisfies our conditions. If we prove that this probability is strictly positive, then we conclude that a "good" object must exist; if all objects were "bad", the probability would be zero.

Let us start with an example illustrating how the probabilistic method works in its basic form.

2.1 Ramsey Numbers

The Ramsey theorem states that any sufficiently large graph contains either a clique or an independent set of a given size. (A *clique*[1] is a set of vertices inducing a complete subgraph and an *independent set*[2] is a set of vertices inducing an edgeless subgraph.)

2.1.1 Definition. *The* **Ramsey number** $R(k,\ell)$ *is*

$$R(k,\ell) = \min\{n\text{: any graph on } n \text{ vertices contains a clique of size } k \text{ or an independent set of size } \ell\}.$$

The Ramsey theorem guarantees that $R(k,\ell)$ is always finite. Still, the precise values of $R(k,\ell)$ are unknown but for a small number of cases, and it is desirable at least to estimate $R(k,\ell)$ for large k and ℓ. Here we use the probabilistic method to prove a lower bound on $R(k,k)$.

[1] clique = klika (úplný podgraf)

[2] independent set = nezávislá množina

2.1.2 Theorem. *For any $k \geq 3$,*

$$R(k,k) > 2^{k/2-1}.$$

Proof. Let us consider a random graph $G(n, 1/2)$ on n vertices where every pair of vertices forms an edge with probability $\frac{1}{2}$, independently of the other edges. (We can imagine flipping a coin for every potential edge to decide whether it should appear in the graph.) For any fixed set of k vertices, the probability that they form a clique is

$$p = 2^{-\binom{k}{2}}.$$

The same goes for the occurrence of an independent set, and there are $\binom{n}{k}$ k-tuples of vertices where a clique or an independent set might appear. Now we use the fact that the probability of a union of events is at most the sum of their respective probabilities (Lemma 1.1.3), and we get

$$\mathrm{P}\left[G(n,1/2) \text{ contains a clique or an indep. set of size } k\right] \leq 2\binom{n}{k}2^{-\binom{k}{2}}.$$

It remains to choose n so that the last expression is below 1. Using the simplest estimate $\binom{n}{k} \leq n^k$, we find that it is sufficient to have $2n^k < 2^{k(k-1)/2}$. This certainly holds whenever $n \leq 2^{k/2-1}$. Therefore, there are graphs on $\lfloor 2^{k/2-1} \rfloor$ vertices that contain neither a clique of size k nor an independent set of size k. This implies $R(k,k) > 2^{k/2-1}$. □

Let us remark that, by using finer estimates in the proof, the lower bound for $R(k,k)$ can be improved a little, say to $2^{k/2}$. But a result even slightly better than this seems to require a more powerful technique. In particular, no lower bound is known of the form c^k with a constant $c > \sqrt{2}$, although the best upper bound is about 4^k.

One might object that the use of a probability space is artificial here and the same proof can be formulated in terms of counting objects. In effect, we are counting the number of bad objects and trying to prove that it is less than the number of all objects, so the set of good objects must be nonempty. In simple cases, it is indeed possible to phrase such proofs in terms of counting bad objects. However, in more sophisticated proofs, the probabilistic formalism becomes much simpler than counting arguments. Furthermore, the probabilistic framework allows us to use many results of probability theory—a mature mathematical discipline.

For many important problems, the probabilistic method has provided the only known solution, and for others, it has provided accessible proofs in cases where constructive proofs are extremely difficult.

3

Linearity of Expectation

3.1 Computing Expectation Using Indicators

The proofs in this chapter are based on the following lemma:

3.1.1 Lemma. *The expectation is a linear operator; i.e., for any two random variables X, Y and constants $\alpha, \beta \in \mathbf{R}$:*

$$\mathbf{E}[\alpha X + \beta Y] = \alpha \mathbf{E}[X] + \beta \mathbf{E}[Y].$$

Proof. $\mathbf{E}[\alpha X + \beta Y] = \int_\Omega (\alpha X + \beta Y)\, \mathrm{dP} = \alpha \int_\Omega X\, \mathrm{dP} + \beta \int_\Omega Y\, \mathrm{dP} = \alpha \mathbf{E}[X] + \beta \mathbf{E}[Y]$. □

This implies that the expectation of a sum of random variables $X = X_1 + X_2 + \cdots + X_n$ is equal to

$$\mathbf{E}[X] = \mathbf{E}[X_1] + \mathbf{E}[X_2] + \cdots + \mathbf{E}[X_n].$$

This fact is elementary, yet powerful, since there is no restriction whatsoever on the properties of X_i, their dependence or independence.

3.1.2 Definition (Indicator variables). *For an event A, we define the indicator variable I_A:*

- $I_A(\omega) = 1$ *if* $\omega \in A$, *and*
- $I_A(\omega) = 0$ *if* $\omega \notin A$.

3.1.3 Lemma. *For any event A, we have* $\mathbf{E}[I_A] = \mathrm{P}[A]$.

Proof.

$$\mathbf{E}[I_A] = \int_\Omega I_A(\omega)\, \mathrm{dP} = \int_A \mathrm{dP} = \mathrm{P}[A].$$

□

In many cases, the expectation of a variable can be calculated by expressing it as a sum of indicator variables

$$X = I_{A_1} + I_{A_2} + \cdots + I_{A_n}$$

of certain events with known probabilities. Then

$$\mathbf{E}[X] = \mathrm{P}[A_1] + \mathrm{P}[A_2] + \cdots + \mathrm{P}[A_n].$$

Example. Let us calculate the expected number of fixed points of a random permutation σ on $\{1, \ldots, n\}$. If

$$X(\sigma) = |\{i\colon \sigma(i) = i\}|,$$

we can express this as a sum of indicator variables:

$$X(\sigma) = \sum_{i=1}^{n} X_i(\sigma)$$

where $X_i(\sigma) = 1$ if $\sigma(i) = i$ and 0 otherwise. Then

$$\mathbf{E}[X_i] = \mathrm{P}[\sigma(i) = i] = \frac{1}{n}$$

and

$$\mathbf{E}[X] = \frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n} = 1.$$

So a random permutation has 1 fixed point (or "loop") on the average.

3.2 Hamiltonian Paths

We can use the expectation of X to estimate the minimum or maximum value of X, because there always exists an elementary event $\omega \in \Omega$ for which $X(\omega) \geq \mathbf{E}[X]$ and similarly, we have $X(\omega) \leq \mathbf{E}[X]$ for some $\omega \in \Omega$.

We recall that a *tournament* is an orientation of a complete graph (for any two vertices u, v, exactly one of the directed edges (u, v) and (v, u) is present). A *Hamiltonian path* in a tournament is a directed path passing through all vertices. The following result of Szele (1943) shows the existence of a tournament with very many Hamiltonian paths.

3.2.1 Theorem. *There is a tournament on n vertices that has at least $\frac{n!}{2^{n-1}}$ Hamiltonian paths.*

Proof. Let us calculate the expected number of Hamiltonian paths in a random tournament T (every edge has a random orientation, chosen independently with probability $\frac{1}{2}$). For a given permutation σ on $\{1, \ldots, n\}$, consider the sequence $\{\sigma(1), \sigma(2), \ldots, \sigma(n)\}$ and denote by X_σ the indicator of the event that all the edges $(\sigma(i), \sigma(i+1))$ appear in T with this orientation. Because the orientation of different edges is chosen independently,

$$\mathbf{E}[X_\sigma] = \mathrm{P}[(\sigma(i), \sigma(i+1)) \in T \text{ for } i = 1, 2, \ldots, n-1] = \frac{1}{2^{n-1}}.$$

The total number of Hamiltonian paths X equals the sum of these indicator variables over all potential Hamiltonian paths, i.e. permutations, and so

$$\mathbf{E}[X] = \sum_{\sigma} \mathbf{E}[X_\sigma] = \frac{n!}{2^{n-1}}.$$

So there is a tournament with at least $\frac{n!}{2^{n-1}}$ Hamiltonian paths. □

3.3 Splitting Graphs

The MAXCUT problem is the following important algorithmic problem: Given a graph $G = (V, E)$, divide the vertex set into two classes A and $B = V \setminus A$ so that the number of edges going between A and B is maximized. This problem is computationally hard (NP-complete). The following simple result tells us that it is always possible to achieve at least half of the edges going between A and B.

3.3.1 Theorem. *Any graph with m edges contains a bipartite subgraph with at least $\frac{m}{2}$ edges.*

Proof. Let $G = (V, E)$, and choose a random subset $T \subseteq V$ by inserting every vertex into T independently with probability $\frac{1}{2}$. For a given edge $e = \{u, v\}$, let X_e denote the indicator variable of the event that *exactly one* of the vertices of e is in T. Then we have

$$\mathbf{E}[X_e] = \mathrm{P}[(u \in T \ \&\ v \notin T) \text{ or } (u \notin T \ \&\ v \in T)] = \tfrac{1}{4} + \tfrac{1}{4} = \tfrac{1}{2}.$$

If X denotes the number of edges having exactly one vertex in T, then

$$\mathbf{E}[X] = \sum_{e \in E} \mathbf{E}[X_e] = \frac{m}{2}.$$

Thus for some $T \subseteq V$, there are at least $\frac{m}{2}$ edges crossing between T and $V \setminus T$, forming a bipartite graph. □

4

Alterations

Sometimes the first attempt to find a "good" object by random construction fails, but we prove that there exists an object which *almost* satisfies our conditions. Often it is possible to modify it in a deterministic way so that we get what we need.

Before we begin with examples, let us mention one simple tool which is useful when we need to estimate the probability that a random variable exceeds its expectation significantly.

4.0.2 Lemma (Markov's inequality). *If X is a non-negative random variable and $a > 0$, then*

$$\mathrm{P}[X \geq a] \leq \frac{\mathbf{E}[X]}{a}.$$

Proof. If X is non-negative, then

$$\mathbf{E}[X] \geq a \cdot \mathrm{P}[X \geq a].$$

□

4.1 Independent Sets

4.1.1 Definition (Independence number). *For a graph G, $\alpha(G)$ denotes the size of the largest independent set in G (a set of vertices such that no two of them are joined by an edge).*

The independence number of a graph is one of its basic parameters. We would like to know how it depends on the number of edges in the graph; specifically, how small the independence number can be for a given average degree.

4.1.2 Theorem (A weak Turán theorem). *If n is the number of vertices of G, m is the number of edges, and $d = \frac{2m}{n} \geq 1$ is the average degree, then*

$$\alpha(G) \geq \frac{n}{2d}.$$

Note. By Turán's theorem, we actually have $\alpha(G) \geq \frac{n}{d+1}$, and this is the best possible in general. For d integral, the extremal graph is a union of disjoint cliques of size $d+1$.

Proof. First, let us select a random subset of vertices $S \subseteq V$ in such a way that we insert every vertex into S independently with probability p (we will choose a suitable value of p later). If X denotes the size of S and Y denotes the number of edges in $G[S]$ (the subgraph induced by S), then

$$\mathbf{E}[X] = np$$

(this follows immediately by the method of indicators; see Section 3.1) and

$$\mathbf{E}[Y] = mp^2 = \tfrac{1}{2}ndp^2$$

(because the probability that both vertices of a given edge are in S is p^2).

We get

$$\mathbf{E}[X - Y] = np(1 - \tfrac{1}{2}dp),$$

so there exists $S \subseteq V$ where the difference of the number of vertices and edges is at least $A(p) = np(1 - \frac{1}{2}dp)$.

Now observe that we can modify S by removing one vertex from each edge inside S. We obtain an independent set with at least $A(p)$ vertices. It remains to choose the value of p so as to maximize $A(p)$; the optimal value is $p = \frac{1}{d}$, which yields

$$A(p) = \frac{n}{2d}.$$

□