Lecture 8. Books on probability. Independent coin tosses? Comte de Buffon, J. Bertrand, P. Valtr, P. Erdős and A. Sah

Martin Klazar

November 24, 2020

PROBABILISTIC TECHNIQUES (NTIN022)
winter term 2020/21

Continuation of Lectures 1–7 by Martin Tancer.

The present lecturer (MK) views the *discipline of probability* as follows. Probability divides in (i) the *formalized probability* residing in mathematics and (ii) the *physical probability* residing in our heads and the world out there. The latter probability reflects our incomplete information, the unknown future, and the random nature of the world (quantum mechanics). The formalized probability can be viewed either *classically*, as a mathematical model of the physical probability, or *enumeratively* as a formal method to count in a generalized sense mathematical objects.

Now and later I will be mentioning interesting books and texts on probability.
• P. Diaconis and B. Skyrms, *Ten Great Ideas about Chance*, Princeton U. Press, 2018. The ideas are: Measurement, Judgment, Psychology, Frequency, Mathematics, Inverse Inference, Unification, Algorithmic Randomness, Physical Chance, and Induction. Not Counting :-(.
• G. J. Székely, *Paradoxes in Probability Theory and Mathematical Statistics*, Springer, 1987.
• J. Haigh, *Probability: A Very Short Introduction*, Oxford U. Press, 2012. A quote from p. 109: "The subject of probability is wholly free from real paradoxes." :-)
• E. T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge U. Press, 2003 (a preliminary version is available online).

• *Edwin T. Jaynes (1922–1998)* was an American physicist, for whom probability theory was an extension of logic. His book was published posthumously by efforts of the editor Larry Bretthorst.

• G. Shafer and V. Vovk, *Game-Theoretic Foundations for Probability and Finance*, J. Wiley, 2019 (a preliminary version is available online). The idea of Finance is definitely missing in the book of Diaconis and Skyrms. This book provides a practical advice in the title of Chapter 13.5: Getting Rich Quick with the Axiom of Choice.
• P. Billingsley, *Probability and Measure*, J. Wiley, 1995 (a version is available online).

• *Patrick Billingsley (1925–2011)* was an American mathematician and stage and screen actor.

We review notation. $\mathbb{N} = \{1, 2, \dots\}$ are the natural numbers, for $n \in \mathbb{N}$ we set $[n] := \{1, 2, \dots, n\}$, $\mathbb{N}_0 = \{0, 1, \dots\} = \mathbb{N} \cup \{0\}$ are the nonnegative integers, $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ are the integers, $\mathbb{Q}$ are the rational numbers (fractions), $\mathbb{R}$ are the real numbers, and $\mathbb{C}$ are the complex numbers. By $|X|$ we denote the cardinality of a set X, which for finite X means just the number of elements in X.

Although you heard it already in Lecture 1, I review definition of a probability space. For a set X we denote by $\mathcal{P}(X)$ the *power set of* X, the set of all subsets of X. A *probability space* is a triple

$$(\Omega,\, \Sigma,\, \Pr)$$

such that $\Omega \neq \emptyset$ is a set, $\Sigma \subset \mathcal{P}(\Omega)$ is a *σ-algebra* on Ω (which means that $\emptyset, \Omega \in \Sigma$, that Σ is closed to complements to Ω, and that Σ is closed to countable unions), and $\Pr\colon \Sigma \to [0, 1]$ is the *probability (function)* that is *σ-additive* (which means that for any pairwise disjoint sets $A_1, A_2, \dots$ in Σ one has that $\Pr(A_1 \cup A_2 \cup \dots) = \Pr(A_1) + \Pr(A_2) + \dots$) and has the value $\Pr(\Omega) = 1$. From these axioms one readily deduces the value $\Pr(\emptyset) = 0$, the *monotonicity rule* $A \subset B \Rightarrow \Pr(A) \leq \Pr(B)$ for any $A, B \in \Sigma$, and the *union bound*

$$\Pr(A_1 \cup A_2 \cup \dots) \leq \Pr(A_1) + \Pr(A_2) + \dots$$

for any $A_n \in \Sigma$. The elements of Σ are called *events*. The intersection of the empty set of events is always defined to be Ω.

Finitely additive measures (probabilities), for which σ-additivity is relaxed and one requires that the sum identity only holds for finite collections of sets from Σ, is a large research topic which we mention here only by listing two books devoted to it. They are probably still available in the Academia bookstores in Prague.

- M. Paštéka, *On four approaches to density*, Veda, 2013.
- M. Paštéka, *Density and related topics*, Academia, 2017.

• *Milan Paštéka* is a Slovak mathematician working at the University of Trnava in the area of densities on integers. He should not be confused with the Slovak painter of the same name, who lived in 1931–1998.

In a *discrete probability space* $(\Omega, \Sigma, \Pr)$ the set Ω is at most countable (but Σ and $\Pr$ still may be uncountable sets); in our course we are interested most of the time only in such spaces. At this occasion we describe their structure.

Let (Ω, Σ) be a σ-algebra. A set $A \in \Sigma$ is an *atom* if $A \neq \emptyset$ and for any set $B \in \Sigma$ we have that if $B \subset A$ then $B = \emptyset$ or $B = A$. The σ-algebra (Ω, Σ) is *atomic* if every element of Ω lies in an atom. Atoms are disjoint and in every atomic σ-algebra the set $\mathcal{A}$ of atoms forms a *set partition* of Ω: the elements in $\mathcal{A}$ are nonempty and disjoint subsets of Ω such that $\bigcup \mathcal{A} = \Omega$. In fact, then every set $X \in \Sigma$ has a unique partition into atoms.

An *atomic probability space* $(\Omega, \Sigma, \Pr)$ has atomic σ-algebra (Ω, Σ) and in consequence has a simple structure. Namely, if $\mathcal{A}$ is the set of its atoms then $\Sigma = \{\bigcup \mathcal{B} \mid \mathcal{B} \subset \mathcal{A}\}$, for any nonempty $A \in \Sigma$ one has that

$$\Pr(A) = \sum_{B \in \mathcal{A},\, B \subset A} \Pr(B) \,,$$

and $\Pr(\emptyset) = 0$. Thus the function $\Pr$ is completely determined by its values on atoms. It is easy to see that if $\Pr\colon \mathcal{A} \to [0,1]$ is any function then its extension to Σ by the above displayed formula is a probability function iff $\sum_{A \in \mathcal{A}} \Pr(A) = 1$.

In the other way it is easy to see that for any at most countable set $\Omega \neq \emptyset$ and any function $f\colon \Omega \to [0,1]$ such that $\sum_{\omega \in \Omega} f(\omega) = 1$, the

triple

$$(\Omega,\, \mathcal{P}(\Omega),\, \Pr)\ ,$$

where $\Pr\colon \mathcal{P}(\Omega) \to [0,1]$ is defined by $\Pr(A) = \sum_{\omega\in A} f(\omega)$ (with $\Pr(\emptyset) = 0$), is an atomic discrete probability space with the set of atoms $\mathcal{A} = \{\{\omega\} \mid \omega \in \Omega\}$.

Proposition (on discrete probability spaces). *Every discrete probability space* $(\Omega, \Sigma, \Pr)$ *is atomic.*

Proof. We need to show that every element x in Ω is contained in an atom. For finite Ω this is clear as every strictly decreasing chain $\Omega = A_0 \supset A_1 \supset \ldots$ where $A_n \in \Sigma$ and $x \in A_n$, is finite and terminates in an atom containing x. We show that also for any countable set Ω each such chain terminates, after α steps for a countable ordinal α, in an atom containing the given element x.

Let Ω be countable, $x \in \Omega$ be any element, and $A_0 = A_\emptyset := \Omega \in \Sigma$. We suppose that $\alpha > 0$ is an ordinal and that A_β has been already defined for every ordinal $\beta < \alpha$ (i.e. $\beta \in \alpha$) and that always $x \in A_\beta$. Suppose that $\alpha = \beta + 1$ is a successor ordinal and let $X_\alpha := A_\beta \in \Sigma$. If there exists a set $X \in \Sigma$ such that $x \in X \subset X_\alpha$ and $X \neq X_\alpha$, we set $A_\alpha := X \in \Sigma$. Else we set $A_\alpha := \emptyset$ and terminate the construction at α. Suppose that α is a (countable) limit ordinal and let $X_\alpha := \bigcap_{\beta<\alpha} A_\beta \in \Sigma$. If there exists a set $X \in \Sigma$ such that $x \in X \subset X_\alpha$ and $X \neq X_\alpha$, we set $A_\alpha := X \in \Sigma$. Else we set $A_\alpha := \emptyset$ and terminate the construction at α. The ordinal sequence $r = (X_\beta)_{\beta>0} \subset \Sigma$ thus defined strictly decreases (in $\supset$) and each set in it contains x as an element. We associate with it the sequence $s = (a_\beta)_{\beta>0} \subset \Omega$ by selecting any elements $a_\beta \in X_\beta \setminus A_\beta$. Since s consists of mutually distinct elements of the countable set Ω, the sequence r has to terminate with a set X_α for a countable ordinal $\alpha < \omega_1$, before the first uncountable ordinal ω_1. It follows that $X_\alpha \in \Sigma$ is an atom containing x. □

Independence of events is a fundamental notion in probability, motivated by the physical probability. Events $A_1, A_2, \ldots, A_n$ in a probability space $(\Omega, \Sigma, \Pr)$ are *independent* if for every set $I \subset [n]$,

$$\Pr\left(\textstyle\bigcap_{i\in I} A_i\right) = \displaystyle\prod_{i\in I} \Pr(A_i)\ .$$

We say that these n events are *independent coin tosses* if they are independent and each has probability $\frac{1}{2}$. So then for any set $I \subset [n]$,

$$\Pr\left(\bigcap_{i\in I} A_i\right) = \left(\tfrac{1}{2}\right)^{|I|} \ .$$

Recall that $|I|$ is the number of elements in the set I. We think of the events A_i, $i \in [n]$, as of n mutually independent acts of flipping a fair coin, with its head and tail coming up with the same (physical) probability $\frac{1}{2}$. We show that for any discrete probability space there is an upper bound $N \in \mathbb{N}$ on the number of independent coin tosses in it. For the proof we need the next lemma.

Lemma. *If $A_1, A_2, \ldots, A_n$ are independent coin tosses in any probability space then for any set $I \subset [n]$,*

$$\Pr\left(\bigcap_{i\in I} A'_i\right) = \left(\tfrac{1}{2}\right)^{|I|}$$

where each A'_i means either the event A_i, or its complement $\overline{A_i} := \Omega \setminus A_i$.

Proof. We proceed by induction on the number $c \geq 0$ of complemented events in the collections $C = \{A'_i \mid i \in I\}$ with $I \subset [n]$. For $c = 0$ we are done by the assumption. Suppose that in C we have $c > 0$ and that the identity holds for any collection with fewer than c complemented events. Let $i_0 \in I$ be such that $A'_{i_0} = \overline{A_{i_0}}$. Then indeed

$$\begin{aligned} \Pr\left(\bigcap_{i\in I} A'_i\right) &= \Pr\Big(\bigcap_{i\in I\setminus\{i_0\}} A'_i\Big) - \Pr\Big(A_{i_0} \cap \bigcap_{i\in I\setminus\{i_0\}} A'_i\Big) \\ &= \left(\tfrac{1}{2}\right)^{|I|-1} - \left(\tfrac{1}{2}\right)^{|I|} = \left(\tfrac{1}{2}\right)^{|I|} \end{aligned}$$

where the second equality follows by induction. □

This is a generalization of the trivial fact that if $\Pr(A) = \frac{1}{2}$ then also $\Pr(\overline{A}) = \frac{1}{2}$. As an exercise you may generalize this lemma to general independent events. Later, for the Lovász Local Lemma (LLL), we will need another similar lemma.

Theorem (few independent coin tosses). *For any discrete probability space $P = (\Omega, \Sigma, \Pr)$ there exists an $N \in \mathbb{N}$ such that any collection C of independent coin tosses in P has $|C| \leq N$ events.*

Proof. Let $P = (\Omega, \Sigma, \Pr)$ be an arbitrary discrete probability space. By the previous proposition it is atomic and we consider the set $\mathcal{A}$ of its atoms. We suppose for contradiction that for every $m \in \mathbb{N}$ there exist m independent coin tosses in P, and deduce that $\Pr(A) = 0$ for every atom $A \in \mathcal{A}$. Since there is the partition $\Omega = \bigcup_{A \in \mathcal{A}} A$, σ-additivity of Pr implies the contradiction that

$$1 = \Pr(\Omega) = \sum_{A \in \mathcal{A}} \Pr(A) = \sum_{A \in \mathcal{A}} 0 = 0 \,.$$

Let $A \in \mathcal{A}$ be an atom and $m \in \mathbb{N}$. We take m independent coin tosses $A_1, A_2, \ldots, A_m$ and define the events $B_1, B_2, \ldots, B_m$ by

$$B_k = \begin{cases} A_k & \ldots \quad A \subset A_k \,, \\ \Omega \setminus A_k & \ldots \quad A \subset \Omega \setminus A_k \,. \end{cases}$$

Then $A \subset B_k$ for every $k \in [m]$ and thus

$$\Pr(A) \leq \Pr\bigl(\textstyle\bigcap_{k=1}^{m} B_k\bigr) = \bigl(\frac{1}{2}\bigr)^m$$

where the last equality follows by the previous lemma and the assumption on the A_k. Since m may be arbitrarily large and then $(1/2)^m \to 0$, indeed $\Pr(A) = 0$. □

This theorem strengthens Problem 1.1 (a) in Billingsley's textbook which asserts nonexistence of infinitely many independent coin tosses in any discrete probability space. See Problem 1.1 (b) for a variant. To model infinitely many, or — as we have seen — even only arbitrarily many, independent coin tosses we therefore need a probability space $(\Omega, \Sigma, \Pr)$ with uncountable set Ω. Constructions of uncountable probability spaces, which can accommodate infinitely many independent coin tosses, are described in the initial part of Billingsley's textbook. This topic lies outside the scope of our course.

Despite it I will tell you three interesting results from the realm of continuous (uncountable) probability. Necessarily I will reason more or less informally. I apologize that this version of my lecture lacks pictures, see the handwritten version for them.

1. Buffon's needle (and noodle) problem. We draw in the Euclidean plane $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ the system $\{\ell_k \mid k \in \mathbb{Z}\}$ of parallel vertical

lines

$$\ell_k = \{(k,\, y) \mid y \in \mathbb{R}\}\ ,$$

and randomly drop in the plane a needle N of length $l \in (0, 1)$. That is, we randomly select in the plane a straight segment $N \subset \mathbb{R}^2$ with length l. What is the probability

$$P_{\mathrm{Bu}}(l) := \Pr(\exists\, k \in \mathbb{Z} :\ \ell_k \cap N \neq \emptyset)\ ?$$

In words, what is the probability that the needle intersects some of the lines? The lines delimit strips with the same width 1 and the needle is shorter, thus the needle intersects at most one line.

No probability space was specified — the next result shows that this *is* a problem — but as we said, we argue informally. To compute the probability $P_{\mathrm{Bu}}(l)$ we define $X_N \in \{0, 1\}$ to be the random variable counting intersections of the needle with the lines, formally

$$X_N := \left|N \cap \textstyle\bigcup_{k\in\mathbb{Z}} \ell_k\right|\ .$$

In the situation when $N \subset \ell_k$ for some $k \in \mathbb{Z}$ we agree to count just one intersection; this event has probability 0 and does not matter anyway. Then

$$P_{\mathrm{Bu}}(l) = \mathbb{E}\, X_N\ .$$

We make $k - 1$ marks on the needle, $k \in \mathbb{N}$, and divide it in effect in k subneedles N_i with lengths $l_i > 0$, $i = 1, 2, \ldots, k$, summing to l. To drop randomly in the plane the needle means to drop randomly in the plane each of the subneedles, and from $X_N = \sum_{i=1}^{k} X_{N_i}$ (the subneedles are disjoint) we get by the linearity of expectation that

$$\mathbb{E}\, X_N = \sum_{i=1}^{k} \mathbb{E}\, X_{N_i}\ .$$

We bend the needle in each of the marks and get a deformed needle N' with the shape of a broken line, but with the same length l. Now X_N becomes $X_{N'}$, each X_{N_i} becomes $X_{N_i'}$, but, crucially, $\mathbb{E}\, X_{N_i'} = \mathbb{E}\, X_{N_i}$. The reason is that, again, to drop randomly in the plane N' entails to drop randomly in the plane each N_i', and each N_i' differs from N_i insignificantly, it is still a straight segment with length l_i. (These are all plausible claims, but since no formal setup was given,

one cannot really prove or refute them. We argue informally, non-rigorously.) Again $X_{N'} = \sum_{i=1}^{k} X_{N'_i}$ and

$$\mathbb{E}\, X_{N'} = \sum_{i=1}^{k} \mathbb{E}\, X_{N'_i} = \sum_{i=1}^{k} \mathbb{E}\, X_{N_i} = \mathbb{E}\, X_N \ .$$

Note that for N' the random variable $X_{N'} \in \mathbb{N}_0$ may attain values larger than 1 because N' may have several intersections with one line.

The trick is to select the shape of N' so that $N' = C$, the circle with circumference l. We cannot get precisely C, of course, but taking k very large and $\max_{1\le i\le k} l_i$ very small we can certainly deform N in N' so that N' approximates C very closely. By the above displayed equations we get that

$$P_{\mathrm{Bu}}(l) = \mathbb{E}\, X_C \ .$$

Here, like above, X_C is the number of intersections of the random circle C with the lines.

It is easy to compute $\mathbb{E}\, X_C$. Let $c = (x, y) \in \mathbb{R}^2$ be the center of the random circle $C \subset \mathbb{R}^2$ with circumference l. Recall that any number $\alpha \in \mathbb{R}$ has its *integral part* $\lfloor \alpha \rfloor \in \mathbb{Z}$ and its *fractional part* $\{\alpha\} \in [0, 1)$ uniquely determined by the equation

$$\alpha = \lfloor \alpha \rfloor + \{\alpha\} \ .$$

Whether $C \cap \ell_k \neq \emptyset$ for some $k \in \mathbb{Z}$ depends only on the coordinate x of c. It happens if and only if

$$\{x\} \in [0,\, l/2\pi] \cup [1 - l/2\pi,\, 1) \subset [0,\, 1)$$

— C has radius $r = l/2\pi$ and the circle C intersects some line ℓ_k if and only if the center c is in distance r or less to ℓ_k, that is, if and only if x is in distance at most r to an integer. The union of the two subintervals of the unit interval has length l/π. Therefore

$$\Pr(\exists\, k \in \mathbb{Z} :\ C \cap \ell_k \neq \emptyset) = \frac{l/\pi}{1} = \frac{l}{\pi} \ .$$

For C we have $X_C \in \{0, 1, 2\}$ and $\Pr(X_C = 1) = 0$ as this is the event

that $\{x\} = l/2\pi$ or $1 - l/2\pi$. Thus we conclude that

$$\begin{aligned} P_{\mathrm{Bu}}(l) &= \mathbb{E}\, X_C \\ &= 0 \cdot \Pr(X_C = 0) + 1 \cdot \Pr(X_C = 1) + 2 \cdot \Pr(X_C = 2) \\ &= 0 + 0 + 2 \cdot \frac{l}{\pi} \\ &= \frac{2l}{\pi}\,. \end{aligned}$$

And where is the noodle? To shape the needle N in the form C, you may heat it up in the forge so that it softens and behaves like a noodle in soup. In conclusion I should mention that I learned the above argument in the book

• D. A. Klain and G.-C. Rota, *Introduction to Geometric Probability*, Cambridge U. Press, 1997.

• *Daniel A. Klain* is an American mathematician working at the University of Massachusetts Lowell.

• *Gian-Carlo Rota (1932–1999)* was an Italian–American mathematician and philosopher who spent most of his career at the Massachusetts Institute of Technology.

• *Georges-Louis Leclerc, Comte de Buffon (1707–1788)* was, by the Wikipedia, a French naturalist, mathematician, cosmologist, and encyclopédiste.

2. Bertrand's paradox. This paradox was published by J. Bertrand in 1889 in his book *Calcul des probabilités*. It consists in three numerically different solutions to the same problem in geometric probability, each of which is clearly correct.

• *Joseph Bertrand (1822–1900)* was a French mathematician who worked also in theoretical economy. In fact, there he authored another paradox.

In the problem one should compute the probability

$$P_{\mathrm{Be}} := \Pr(\text{length of } T > \sqrt{3})$$

that a random chord T in a unit circle C, which is a circle with

radius 1, is longer than the side of the inscribed equilateral triangle. We present three solutions and in each the chord T is determined by two one-dimensional parameters.

Solution 1. The chord $T = AB$ is determined by its endpoints A and B lying on C. We randomly select first A and then B. Let $D, E \in C$ be the other two vertices of the equilateral triangle with vertex A inscribed in C. The points A, D, E divide C in three arcs with equal lengths, and T is longer than the side of the triangle iff B falls in the arc between D and E. Thus

$$P_{\mathrm{Be}} = \frac{\text{the length of the arc } DE}{\text{the circumference of } C} = \frac{1}{3}\,.$$

Solution 2. The chord T is determined by its direction and distance from the center A of C. First we randomly select a radius B of C, a segment $B = AD$ with the other endpoint $D \in C$, which determines the direction of T in the way that T intersects B and is perpendicular to it. Then we randomly select the intersection point E of B and T. We consider the equilateral triangle inscribed in C so that one its side intersects the radius B in a point F and is perpendicular to B. It is easy to see that F is the midpoint of B and that T is longer than the side of the triangle iff E falls in the subsegment AF of the radius B. Thus

$$P_{\mathrm{Be}} = \frac{\text{the length of } AF}{\text{the length of } AD} = \frac{1}{2}\,.$$

Solution 3. The chord T is determined by the position of its central point A. As we know from the previous solution, T is longer than the side of the triangle iff the distance of A from the center of C is less than $\frac{1}{2}$, that is, iff A lies inside the circle B that is concentric with C and has radius $\frac{1}{2}$. Thus

$$P_{\mathrm{Be}} = \frac{\text{the area inside } B}{\text{the area inside } C} = \frac{1}{4}\,.$$

;-) Some texts on probability explain Bertrand's paradox away by saying: these are three different probability spaces, thus it is absolutely no wonder that one gets three different solutions. But we do not get rid of the paradox so cheaply, one feels that the real problem lies elsewhere. The selection of random chords T in C should model

some real physical experiment in which real, physical chords are being randomly selected, in some way. That experiment can have only one result, at least in the macro-world. What is the experiment and what is the result? As recently as 2014 people were solving Bertrand's paradox, see for example the preprint D. Aerts and M. Sassoli de Bianchi, Solving the Hard Problem of Bertrand's Paradox, *arXiv:1403.4139*, 15 pp.

3. Valtr's convex chains. This is also a result from geometric probability, as the previous two, but it is undeservedly not so well known. Let $C = [0,1] \times [0,1]$ be the unit square and $p_i \in C$, $i = 1, 2, \ldots, n$, be n random points in it. Here the probability space is without dispute the square of the Lebesgue measure on the unit interval $[0,1]$. It should be clear what it means that $p_1, \ldots, p_n$ form a convex n-gon. We say that $p_1, \ldots, p_n$ form a *convex n-chain*, which is a particular case of a convex n-gon, if these points lie on a graph of a convex function. Said more explicitly, when the points are relabeled so that in the sequence $p_1, p_2, \ldots, p_n$ the x-coordinate increases (two x-coordinates coincide with zero probability), then the direction vectors

$$p_2 - p_1,\ p_3 - p_2,\ \ldots,\ p_n - p_{n-1}$$

rotate counter-clockwisely. What is the conditional probability

$$P_{\mathrm{Va}}(n) := \Pr(p_i \text{ form a convex } n\text{-chain} \,|\, p_i \text{ form a convex } n\text{-gon}) \ ?$$

Trivially, $P_{\mathrm{Va}}(1) = P_{\mathrm{Va}}(2) = 1$, and easily $P_{\mathrm{Va}}(3) = \frac{1}{2}$. The answer for general $n \in \mathbb{N}$ involves the *Catalan numbers*

$$C_n := \frac{1}{n}\binom{2n-2}{n-1} = \frac{1}{n} \cdot \frac{(2n-2)!}{(n-1)!^2} \ .$$

Their sequence begins as $(C_n)_{n\geq 1} = (1, 1, 2, 5, 14, 42, 130, \ldots)$. In 1997 P. Valtr proved in his article Catalan numbers via random planar point sets, *Intuitive geometry (Budapest, 1995)*, János Bolyai Math. Soc., 441–443 that

$$P_{\mathrm{Va}}(n) = \frac{1}{C_n} \ .$$

The result comes out from the proofs as a ratio of two probabilities that are computed separately. One feels that there might be a more

direct argument based on some symmetry extending the case $n = 3$. We state it as a research problem.

Problem. Derive by a direct symmetry argument, by means of some combinatorial properties of the Catalan numbers, the formula for $P_{\mathrm{Va}}(n)$.

• *Eugène Catalan (1814–1894)*, after whom the numbers C_n are named, was a French and Belgian mathematician whose areas of interest were number theory and combinatorics. The Catalan conjecture in Diophantine analysis, that the only solution to the equation $x^a - y^b = 1$ in numbers $x, y, a, b \in \mathbb{N} \setminus \{1\}$ is $3^2 - 2^3 = 1$, is also named after him. The conjecture was proved in 2004 by P. Mihăilescu.

• *Pavel Valtr* is lecturer's colleague in the Department of Applied Mathematics of MFF UK, indeed he was the chairman of the department, and is a world-famous researcher in the area of computational and discrete geometry.

We turn to the last topic of this lecture, bounds on the *Ramsey numbers* $R(k)$, $k \in \mathbb{N}$. You saw a proof of a lower bound in Lecture 1 but now we look at this topic from a bit different angle. $R(k)$ is the minimum natural number n such that every graph on n vertices has a clique or an independent set of size k. Formally,

$$\begin{aligned} R(k) := \min \big(\{n \in \mathbb{N} \mid \chi\colon \tbinom{[n]}{2} \to [2] \Rightarrow \\ \exists\, X \subset [n]\colon\ |X| = k \wedge \chi \,|\, \tbinom{X}{2} = \chi_c\}\big)\ . \end{aligned}$$

Here $\binom{X}{2}$ denotes the set of all two-element subsets of the set X, $\cdot\,|\,\cdot$ is the restriction operation for functions, and χ_c denotes any constant function.

• *Frank P. Ramsey (1903–1930)*, after whom the numbers $R(k)$ are named, was a British philosopher, mathematician, economist, and truly a genius. Unfortunately, he conformed to the image of a romantic genius also by dying very young (probably because of liver infection he contracted when he was swimming in the river Cam). For his life see the biography

• Ch. Misak, *Frank Ramsey. A Sheer Excess of Powers*, Oxford Uni-

versity Press, 2020.

We upper bound $R(k)$.

Theorem (an upper bound on $R(k)$). For every $k \in \mathbb{N}$,

$$R(k) \leq 4^{k-1} .$$

Proof. Let $k \in \mathbb{N}$ and $n = 4^{k-1} = 2^{2k-2}$. For $k = 1$ the bound holds and we assume that $k > 1$. Let $\chi\colon \binom{[n]}{2} \to [2]$ be any coloring. We set $x_1 = n$ and select a subset $X_1 \subset [n-1]$ such that $|X_1| \geq 2^{2k-3}$ and the colors $\chi(\{x, x_1\})$, $x \in X_1$, are the same. We set $x_2 = \max(X_1)$ and select a subset $X_2 \subset X_1 \setminus \{x_2\}$ such that $|X_2| \geq 2^{2k-4}$ and the colors $\chi(\{x, x_2\})$, $x \in X_2$, are the same. We continue this way and define elements $n = x_1 > x_2 > \cdots > x_{2k-1} \geq 1$ (x_{2k-1} is any element of X_{2k-2}) such that for any $j < i, i'$ one has that $\chi(\{x_i, x_j\}) = \chi(\{x_{i'}, x_j\})$. Thus the color of each pair from the set $X = \{x_1, x_2, \ldots, x_{2k-1}\}$ depends only on the larger element of the pair, and we can select a set $Y \subset X$ such that $|Y| = k$ and $\chi \mid \binom{Y}{2}$ is a constant coloring. □

It was necessary to give an upper bound on $R(k)$ to make sure that $R(k)$ is a well defined thing. Think of the following example. We define the number $R'(k)$ by omitting in the above definition of $R(k)$ the clause "or an independent set", that is, we replace at the end of the formal definition χ_c with χ_1, the constant map $\chi \equiv 1$. The proof below is easily adapted with almost no change to prove that $R'(k) > 2^{k/2}$, it even gets simpler. But the whole thing is a nonsense because $R'(k)$ is not defined for any $k > 1$, the set $\{n \in \mathbb{N} \mid \ldots\}$ in its formal definition is empty for any $k > 1$.

We proceed to the lower bound on $R(k)$ and let speak its autor P. Erdős himself. Below is a proof of the lower bound $R(k) > 2^{k/2}$, quoted verbatim from p. 292 of the 1947 article P. Erdös, Some remarks on the theory of graphs, *Bull. Amer. Math. Soc.* **53**, 292–294. By the canonical textbook of the probabilistic method

• N. Alon and J. H. Spencer, *The Probabilistic Method*, John Wiley & Sons, 2000,

it is "The three-page paper that 'started' the probabilistic method, giving an exponential lower bound on Ramsey $R(k, k)$" (p. 276).

• *Paul (Pál) Erdős (1913–1996)* was a prolific Hungarian mathematician working in combinatorics, number theory, mathematical analysis, probability theory, and set theory.

• *Noga Alon* is an Israeli mathematician and computer scientist. He is the greatest living combinatorialist.

• *Joel H. Spencer* is an American mathematician, working in combinatorics. His book *The Strange Logic of Random Graphs*, Springer, 2001, is quite interesting.

Erdős's notation: $f(k, k) := R(k)$ and $C_{a,b} := \binom{a}{b}$. Now PE speaks:

"Theorem I. *Let $k \geq 3$. Then*

$$2^{k/2} < f(k,\, k) \leq C_{2k-2,k-1} < 4^{k-1}.$$

The second inequality of Theorem I was proved by Szekeres, thus we only consider the first one. Let $N \leq 2^{k/2}$. Clearly the number of different graphs of N vertices equals $2^{N(N-1)/2}$. (We consider the vertices of the graph as distinguishable.) The number of different graphs containing a given complete graph of order k is clearly $2^{N(N-1)/2}/2^{k(k-1)/2}$. Thus the number of graphs of $N \leq 2^{k/2}$ vertices containing a complete graph of order k is less than

$$C_{N,k}\frac{2^{N(N-1)/2}}{2^{k(k-1)/2}} < \frac{N^k}{k!}\frac{2^{N(N-1)/2}}{2^{k(k-1)/2}} < \frac{2^{N(N-1)/2}}{2} \tag{1}$$

since by a simple calculation for $N \leq 2^{k/2}$ and $k \geq 3$

$$2N^k < k!2^{k(k-1)/2}.$$

But it follows immediately from (1) that there exists a graph such that neither it nor its complementary graph contains a complete subgraph of order k, which completes the proof of Theorem I."

This reminds me that at the beginning of the lecture I wrote and said that formalized probability can be viewed also as a method of

enumeration. P. Erdős wisely saw that his simple and beautiful proof should not be obfuscated by unnecessary probabilistic jargon, and in the whole article there is no single mention of probability theory or randomness. Alon and Spencer gloss over this fact.

We conclude with the best currently known lower and upper bounds on $R(k)$ when $k \to \infty$ ($c > 0$ is a constant):

$$(1 + o(1))\frac{\sqrt{2}}{e} k \big(\sqrt{2}\big)^k < R(k) < e^{-c \log^2 k} \binom{2k}{k} \sim e^{-c \log^2 k} 4^k \ .$$

The lower bound is due to J. Spencer in 1975, by means of the LLL which is the topic of the next lecture. The upper bound was proved this year 2020 in the preprint A. Sah, Diagonal Ramsey via effective quasi-randomness, *arXiv:2005.09251*, 14 pp. This proof also uses the probabilistic method, the mentioned quasi-randomness. A tantalizing and no doubt very difficult problem is to close or at least narrow the gap between $\sqrt{2}$ and 4 in the above displayed exponential bounds on $R(k)$.

• *Ashwin Sah*— "Hello! Welcome to my homepage. I am a mathematics graduate student at MIT (since Fall 2020). My research interests include combinatorics, probability, and number theory."

Thank you!

(final version of January 13, 2021)