Probability-I

Spring 2024

Soumya Bhattacharya

Literature used in the preparation of these notes:

S. Bhattacharya, *The Probabilistic Pigeonhole Principle*, The American Mathematical Monthly, Vol. 130, No. 7, p. 678, 2023

G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, eds. D. R. Heath-Brown, J. H. Silverman, Oxford University Press, 2008

S. Ross, *A first course in Probability*, Pearson Education, 2002

`https://en.wikipedia.org/`

If you have any comments about these lecture notes, please send them to
soumya.bhattacharya@iiserkol.ac.in

Life is either a daring random experiment[*] *or nothing at all.*

[*]apologies to Helen Keller.

Contents

Probability Spaces

Probability is a measure of how likely an event is to occur. This subject has a wide range of applications in every aspect of human life. The first use of probabilistic notions occurred in Cardano's *Book on games of chance** (1663). Also, *The art of speculating*** (1713) by Bernoulli and *The doctrine of chances*† (1718) by De Moivre had immense impacts on the development of the Theory of Probability. Initially, people studying probability were concerned about the actions which could result in only finitely many outcomes. So, the probability of obtaining a particular set of outcomes was defined as

(0.1) *the proportion of the required outcomes among all the possible outcomes*

of the action. The first attempt to extend probabilistic ideas beyond finite sets was in Laplace's *Analytical Theory of Probability*‡ (1812). Much later, the modern axioms of Probability were introduced by Kolmogorov in his book *Foundations of the Theory of Probability*§ (1933).

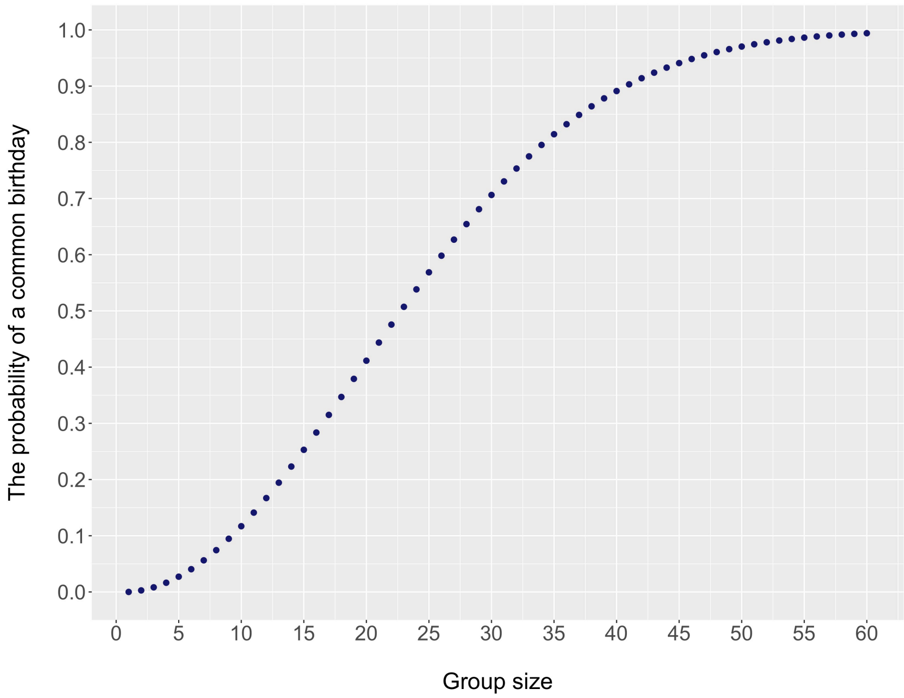
1 The probabilistic pigeonhole principle

The Pigeonhole principle asserts that if you put n pigeons into $m < n$ pigeonholes, then at least two pigeons would be in the same hole.

You may wonder what happens if there are more pigeonholes than pigeons. In any case, if the pigeons are allocated to the pigeonholes indiscriminately, regardless of the number of occupants of any pigeonhole, then it is still a possibility that two or more pigeons end up in the same hole. The *Birthday problem* is only a special

*in Italian
**in Latin
†in English
‡in French
§in German

case of this which asks for the probability* that in a group of n randomly chosen people, at least two shares a birthday. The apparent paradox is that the probability of two persons sharing a birthday exceeds 0.5 if the group size is at least 23. In fact, if a group has at least 57 members, the probability that at least two of the members have a common birthday is more than 0.99.



THEOREM 1 (PROBABILISTIC PIGEONHOLE PRINCIPLE). *Given $m \in \mathbb{N}$ and $p \in [0, 1)$, let $n \in \mathbb{N}$ be larger than or equal to*

$$\frac{1}{2} + \sqrt{2m \log\left(\frac{1}{1-p}\right) + \frac{1}{4}}. \tag{1.1}$$

If n pigeons are placed in m pigeonholes randomly, then the probability that at least two pigeons would be in the same hole is greater than p.

PROOF. We require to to show that for an integer n which is larger than the quantity in (1.1), if n pigeons are placed in m pigeonholes randomly, then the probability of an overlap in the placement of the pigeons is greater than p. From (0.1), it follows that the probability of an overlap in the placement is greater than p if and only if the probability of no overlap in the placement of the pigeons is less than $1 - p$. Let us compute this probability: Since every pigeon could be placed in any of the m pigeonholes, the total number of ways in which n pigeons could be placed in m holes is m^n. Whereas, the number of ways in which n pigeons could be placed in m holes with no overlaps is ${}^m\mathrm{P}_n$. Now, (0.1) implies that the required probability is the proportion of the placements of the pigeons in distinct pigeonholes among all possible placements of n pigeons in m pigeonholes. This

*For now, we take (0.1) as the definition of Probability.

proportion is ${}^m\mathrm{P}_n/m^n$. Hence, it suffices to show that if the integer n is greater than or equal to the quantity in (1.1), then ${}^m\mathrm{P}_n/m^n$ is less than $1-p$.

Let n be greater than or equal to the quantity in (1.1). Then after subtracting $1/2$ from both sides, squaring and dividing by $2m$, we obtain

$$\frac{n(n-1)}{2m} > \log\Big(\frac{1}{1-p}\Big),$$

which implies that

$$-\sum_{j=1}^{n-1}\log\big(1-\frac{j}{m}\big) = \sum_{j=1}^{n-1}\sum_{\ell=1}^{\infty}\frac{j^\ell}{\ell m^\ell} > \sum_{j=1}^{n-1}\frac{j}{m} = \frac{n(n-1)}{2m} > \log\Big(\frac{1}{1-p}\Big).$$

After multiplying both sides by -1 and exponentiating, we obtain

$$\prod_{j=1}^{n-1}\big(1-\frac{j}{m}\big) < 1-p.$$

Since the left hand side of the above inequality is ${}^m\mathrm{P}_n/m^n$, the claim follows. □

Instead of using the logarithmic inequality as above, one may also use the fact that $e^{-x} > 1-x$ for all $x \in \mathbb{R}$ to prove Theorem 1.

Corollary 1. *If $3n+1$ or more pigeons are placed in n^2 pigeonholes randomly, then the probability that at least two pigeons would be in the same hole is greater than 0.98889.*

Proof. Putting $m = n^2$ and $p = 0.98889$ in (0.1), we obtain a quantity that is smaller than

$$\frac{1}{2} + \sqrt{9n^2 + \frac{1}{4}},$$

which is less than $3n+1$. □

2 The longest run in a sequence of tosses

During the last quarter of the twentieth century, it was a common practice* among probability teachers around world to demostrate a probabilistic method of lie detection: They used to give their students the homework of tossing a coin a hundred times** and noting down the sequence of outcomes. Quite predictably, it was also a common practice among the students to evade this boring homework and make an attempt to dupe the teacher by writing down a sequence of heads and tails which they thought are sufficiently *random*. As in general, human intuition of *randomness* is rather poor, in almost all cases these made-up sequences did not contain any run of heads or tails of length larger than five or six. In the next class,

*See this article: `https://www.maa.org/sites/default/files/images/upload_library/22/Polya/07468342.di020742.02p0021g.pdf`

**Or even a thousand times, as recollected by Prof. Bimal Roy from the B. Stat.(Hons.) batch of 1978 of ISI Kolkata. Their teacher Prof. Jogabrata Roy gave them this homework in their first class of Probability.

the teacher used to ask how many students obtained a run of eight* or more heads, which was often answered in the negative. Then he used to tell the students to compute the probability of their claim to see how unlikely it is! Let's demostrate how to compute such probabilities** with a small example:

Example 1. Find the probability of obtaining no two consecutive heads when a fair coin is tossed seven times.

Answer. For $n \geq 1$, let F_n denote the number of outcomes with no two consecutive heads when a fair coin is tossed n times. Then we have

$$F_1 = |\{T, H\}| = 2,$$

$$F_2 = |\{TT, HT, TH\}| = 3$$

and for $n > 2$,

$$F_n = F_{n-1} + F_{n-2},$$

because, a sequence of $n \geq 2$ tosses without two consecutive heads ends either in T or TH. It follows from above that $F_7 = 34$. Since the total number of possible outcomes in 7 tosses (assuming that the coin does not land on its side) is 2^7. So, according to (0.1), the required probability is $F_7/2^7 = 17/64$.

3 The Riemann Hypothesis via probability

As a brief motivation for the students of probability, below we describe the Riemann Hypothesis† (1859) in probabilistic terms.

Definition 1 (Primorial). The k-th primorial is the product of the first k primes.

Definition 2 (Euler–Mascheroni constant). The limiting value γ of the difference between $\displaystyle\sum_{k=1}^{m} \frac{1}{k}$ and $\log m$ as m tends to infinity. Its first few digits are

$0.5772156649015328606065120900824024310421593359399235988057672\ldots.$

The Riemann Hypothesis is equivalent to the claim that if an integer n is chosen at random from $\{1, 2, \ldots, N_k\}$, where N_k denotes the k-th primorial, then the probability of n being coprime to N is less than

$$\frac{1}{e^{\gamma} \log\log N_k}$$

for all $k \in \mathbb{N}$.

*Of course, *eight* could be replaced by a larger number, depending on the class size.

**Here we compute the probability only for one sequence of a given length. In a class, the same experiment is repeated as many times as the number of students in the class. So, the probability that nobody obtains a run of heads of length eight or more in hundred tosses reduces significantly when the class size is big.

†A 165 years old open conjecture which claims that the real part of any the nontivial zero of the Riemann Zeta function is $1/2$.

DEFINITION 3 (EXPERIMENT). An experiment is an act that can be repeated under similar conditions.

DEFINITION 4 (SAMPLE SPACE). The set Ω of all possible outcomes of an experiment is called its sample space.

DEFINITION 5 (SET OF EVENTS AND THE AXIOMS OF PROBABILITY). The set of events is a subset $\mathcal{E}$ of the power set of the sample space Ω such that

(a) $\Omega \in \mathcal{E}$.
(b) If $A \in \mathcal{E}$, then $A^c \in \mathcal{E}$.
(c) The set $\mathcal{E}$ is closed under countable unions.
(d) There exists a function $P : \mathcal{E} \to [0, 1]$ such that $P(\Omega) = 1$. and for pairwise disjoint events $\{A_n\}_{n=1}^{\infty}$ with $A = \underset{n}{\cup} A_n$,

$$P(A) = \sum_{n=1}^{\infty} P(A_n).$$

In particular, for a finite sample space, we may always assume the set of events to be identical with the power set of the sample space.

DEFINITION 6 (PROBABILITY SPACE). The sample space Ω of an experiment together with the set of events $\mathcal{E}$ and the probability function $P : \mathcal{E} \to [0, 1]$, is called a probability space and it is denoted by the triplet $(\Omega, \mathcal{E}, P)$.

EXAMPLE 2 (GAMES OF CHANCES).

(i) Tossing a coin: $\Omega = \{H, T\}$.
(ii) Rolling a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$.
(iii) Guessing the first letter of a stranger's name: $\Omega = \{A, B, C, \ldots, Z\}$.
(iv) Guessing someone's palm temperature in Fahrenheit: $\Omega = [55, 115]$.
(v) Guessing the number of stars in the Milky Way: $\Omega := \{1, 2, 3, \ldots\}$.

DEFINITION 7 (MUTUALLY EXCLUSIVE EVENTS). If two events $A, B \in \mathcal{E}$ is such that $A \cap B = \emptyset$, then A and B are called mutually exclusive events.

Let $(\Omega, \mathcal{E}, P)$ be a probability space. If A and $B \in \mathcal{E}$ are mutually exclusive events, then Definition 5.(d) implies that $P(A \cup B) = P(A) + P(B)$.

However, if $A \cap B \neq \emptyset$, then we may write $A = (A \cap B) \cup (A \cap B^c)$, where $B^c := \Omega \smallsetminus B$. Since $A \cap B$ and $A \cap B^c$ are mutually exclusive, it follows from Definition 5.(d) that

$$P(A) = P(A \cap B) + P(A \cap B^c). \tag{4.1}$$

Since $A \cup B = (A \cap B^c) \cup B$ and since $A \cap B^c$ and B are mutually exclusive,

$$P(A \cup B) = P(A \cap B^c) + P(B). \tag{4.2}$$

From (4.1) and (4.2), it follows that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{4.3}$$

By induction, the above equality can be generalized to any finite collection of of events.

DEFINITION 8 (EXHAUSTIVE EVENTS). The events in a set $S \subseteq E$ are called exhaustive if $\underset{A \in S}{\cup} A = \Omega$, i.e. the entire sample space.

EXAMPLE 3 (COUNTABLE SAMPLE SPACES). Let $\Omega = \{\omega_n\}_{n=1}^{\infty}$. If $\mathcal{E} = 2^{\Omega}$, then in particular, $\{\omega_n\} \subseteq \Omega$ for all $n \in \mathbb{N}$. Since for $n \in \mathbb{N}$, the events $\{\omega_n\}$ are pairwise mutually exclusive and exhaustive, it follows from the axioms of probability that

$$\sum_{n=1}^{\infty} P(\omega_n) = P(\Omega) = 1.$$

EXAMPLE 4. Show that if you keep on tossing a coin where the probability of obtaining a head is $p > 0$, then the probability of eventual occurrence of at least one head is 1.

PROOF. Consider the random experiment, where you keep on tossing the coin till you get a head. The corresponding sample space is

$$\Omega := \{H, TH, TTH, TTTH, \ldots\} \cup \{TTTTT \ldots \infty\}$$

and the set of events is $\mathcal{E} = 2^{\Omega}$, where

$$P(H) = p,\ P(TH) = (1-p)p,\ P(TTH) = (1-p)^2 p, \ldots$$

Since probability of eventually obtaining a head is

$$P(\Omega) = \sum_{n=0}^{\infty} (1-p)^n p = \frac{p}{1-(1-p)} = 1,$$

the claim follows. □

Since the complement of the desired event (i.e. eventually obtaining a head) is $\{TTTTT \ldots \infty\}$, showing that

$$P(TTTTT \ldots \infty) = \lim_{n \to \infty} (1-p)^n = 0$$

would have also sufficed for Example 4 (see Exercise 10.i).

DEFINITION 9 (EQUALLY LIKELY). The events in a set $S \subseteq E$ are called equally likely if $P(A) = P(B)$ for all $A, B \in S$.

THEOREM 2. *All the outcomes in a countably infinite sample space can not be equally likely.*

PROOF. Let $\Omega := \{\omega_n\}_{n=1}^{\infty}$. Suppose, each of the outcomes ω_n has an equal probability p of occurrence. Since the events $\{\omega_n\}$ are mutually exclusive and exhaustive, it follows that

$$1 = P(\Omega) = \sum_{n=1}^{\infty} P(\omega_n) = \sum_{n=1}^{\infty} p,$$

which is absurd, since the right hand side does not converge if $p > 0$, whereas if $p = 0$, then the right hand side converges to zero! □

5 Earlier attempts at defining probability

Definition 10 (Random experiment). If the sample space of an experiment is known but none of the outcomes of the experiment occurs with certainty, then we call the experiment a random experiment.

In particular, all the games of chances mentioned in Example 2 are random experiments.

Example 5 (Examples of non-random experiments). All the experiments that verify certain physical/chemical/biological laws are not random experiments*.

Definition 11 (Classical definition of probability). If a random experiment results in m mutually exclusive, exhaustive and equally likely outcomes, of which exactly $n(A)$ outcomes are favourable for an event $A \in \mathcal{E}$, then the probability of the event A, denoted by $P(A)$ is given by

$$P(A) = \frac{n(A)}{m}.$$

Indeed, the last definition is meant only for finite sample spaces. However, an inherent flaw in the last definition of probability is that it is cyclic, since in order to define *equally likely* events, we have already used the notion of probability.

Definition 12 (Frequency definition of probability). Let a random experiment be repeated n times, in which the frequency of the event A was $f_n(A)$, i.e. the event A occurred exactly $f_n(A)$ times. The ratio $f_n(A)/n$ is called the relative frequency of the event A and the probability $P(A)$ is defined by the limit

$$P(A) = \lim_{n\to\infty} \frac{f_n(A)}{n}.$$

An inherent drawback of the above definition of probability is that the existence of the above limit can not be proved in any case. However, later we shall see that the law of large numbers has its roots in this idea.

6 Boole's and Bonferroni's inequalities

Theorem 3 (Boole's inequality). *Let $(\Omega, \mathcal{E}, P)$ be a probability space. Then for $A_1, A_2, \dots A_n \in \mathcal{E}$, we have*

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) \leq P(A_1) + P(A_2) + \cdots + P(A_n).$$

Proof. For $i \in \{1, 2, \dots, n\}$, define

$$B_i := \begin{cases} A_1 & \text{if } i = 1 \\ A_i \smallsetminus \bigcup_{j=1}^{i-1} A_j. & \text{if } i > 1. \end{cases}$$

*However, if the range of the possible errors in such an experiment is nonempty and is known a priori, then we may also view such an experiment as a random experiment, since no particular error occurs with certainty.

By construction, the events B_i for $i \in \{1, 2, \ldots, n\}$ are disjoint. Therefore,

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(B_1 \cup B_2 \cup \ldots B_n) = \sum_{i=1}^{n} P(B_i) \leq \sum_{i=1}^{n} P(A_i),$$

where the second equality follows from Definition 5.(d) and the last inequality follows from Exercise 10.ii. □

THEOREM 4 (BONFERRONI'S INEQUALITY). *Let* $(\Omega, \mathcal{E}, P)$ *be a probability space. Then for* $A_1, A_2, \ldots A_n \in \mathcal{E}$*, we have*

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) \geq P(A_1) + P(A_2) + \cdots + P(A_n) - (n-1).$$

PROOF. We proceed to prove the claim by induction. Note that since $P(A_1 \cup A_2) \leq 1$, for $n = 2$ the claim follows trivially from (4.3). Suppose, the claim holds for an integer $m \geq 2$. Then

$$\begin{aligned} P(A_1 \cup A_2 \cup \ldots \cup A_m \cup A_{m+1}) &\geq P(A_1 \cup A_2 \cup \ldots \cup A_m) + P(A_{m+1}) - 1 \\ &\geq \sum_{i=1}^{m} P(A_i) - (m-1) + P(A_{m+1}) - 1 \\ &= \sum_{i=1}^{m+1} P(A_i) - m, \end{aligned}$$

where the first inequality holds since the claim is true for two events, whereas the second inequality is implied by the induction hypothesis. Thus, the claim follows. □

Exercises

1. Show that the probability of at least two persons sharing a birthday among a group of n persons is
$$1 - \frac{{}^{365}\mathrm{P}_n}{365^n} - \frac{97 \times {}^{365}\mathrm{P}_{n-1}}{146097 \times 365^{n-1}}.$$
2. Show that for any sequence $\{a_n\}$ of positive reals which diverges to infinity, if $\lfloor a_n \sqrt{n} \rfloor$ pigeons are placed in n pigeonholes, then the probability that at least two pigeons are in the same hole tends to 1 as $n \to \infty$.
3. If you toss a fair coin 20 times, find the probability of obtaining 10 heads and 10 tails.
4. Compute the probability of obtaining a total of 10 points when two unbiased dice are rolled.
5. (i) Find the probability of obtaining no four consecutive heads when a fair coin is tossed ten times.
(ii) Find the probability of obtaining neither four consecutive heads nor four consecutive tails when a fair coin is tossed ten times.

6. Let x be a point inside a convex quadrilateral $\mathcal{Q}$. Find the probability that x is neither on the boundary nor inside any of the circles drawn with the sides of the quadrilateral Q as their diameters.

7. (Bertrand's Paradox) Explain why there can't be a single answer to the question that asks for the probability of choosing a chord randomly of length less than r in a circle of radius r.

8. Let $(\Omega, \mathcal{E}, P)$ be a probability space. Show that $\mathcal{E}$ is closed under countable intersections.

9. Let $(\Omega, \mathcal{E}, P)$ be a probability space. Let $A, B \in \mathcal{E}$. Show that
$$P(B \cap A^c) = P(B) - P(B \cap A).$$

10. Let Suppose $(\Omega, \mathcal{E}, P)$ be a probability space and $A, B \in \mathcal{E}$. Using the axiomatic definition of probability, prove the following statements:
    i) $P(A^c) = 1 - P(A)$.
    ii) If $A \subseteq B$, then $P(A) \leq P(B)$.

11. Let $(\Omega, \mathcal{E}, P)$ be a probability space and $A_1, A_2, \ldots, A_n \in \mathcal{E}$. Show that $P(A_1 \cup A_2 \cup \ldots \cup A_n)$ is equal to
$$\sum_{i=1}^{n} P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) - \\ \ldots + (-1)^{n-1} P(A_1 \cap \ldots \cap A_n).$$

12. Show that if you keep on throwing a fair die, the probability of eventual occurence of the outcome 6 is 1.

13. Find the probability of obtaining no two consecutive heads when a fair coin is tossed n times.

14. Find the probability of obtaining a strictly increasing sequence of integers if you pick 3 integers (one at a time) (a) with replacement or (b) without replacement from $\{1, 2, 3, \ldots, 1000\}$.

15. To the choice of each $n \in \mathbb{N}$, could you assign a probability $P(n) > 0$ such that the following conditions hold?
    (1) $P(m) \neq P(n)$ for all $m, n \in \mathbb{N}$, $m \neq n$.
    (2) The probability of choosing an odd positive integer is the same as the probability of choosing an even positive integer.

16. Seven students of IISER Kolkata went to participate in an event at IISER Mohali. They booked AC 3-tier tickets from Howrah to Chandigarh in

Netaji Express, which has three AC 3-tier coaches. Every such coach has eight coupes (i.e. compartments), each coupe containing eight berths. If the berths were allocated randomly, find the probability of at least two among the seven students being allocated berths in the same coupe.

17. A point is selected at random inside an equilateral triangle whose side length is 3. Find the probability that the distance of the chosen point from any vertex of the traingle is greater than 1.

18. On a rainy day, 4 students went for their coaching classes with nearly identical raincoats. They put their raincoats at the same place before going to the class. Find the probability that none of the students select their own raincoat after the class.

19. Suppose, the planetary system of a star contains ten planets with coplanar orbits and suppose the distance from the star to the farthest point(s) on the orbit of the outermost planet is 1.5 astronomical units. Find the probability that at an arbitrary instant of time, at least two of these planets are less than or equal to $\sqrt{2}$ astronomical units apart from each other.

20. Suppose, the planetary system of a star contains twenty-five planets with at least two planets having non-coplanar orbits and suppose the distance from the star to the farthest point(s) on the orbit of the outermost planet is 2 astronomical units. Show that the probability of at least two of these planets being less than or equal to $\sqrt{3}$ astronomical units apart from each other at an arbitrary instant of time is greater than 0.99.

21. Carrom is played with a red, nine black and nine white coins (and a striker) on a square board with a pocket in each corner. If all these coins are scattered randomly (none of them being in any pocket) on a 29 inch $\times$ 29 inch carrom board, show that the probability of at least two of the coins being less than three inches apart is greater than $1/2$.