

Chapter 1

Probability Spaces

1.1 The sample space

The intuitive meaning of probability is related to some *experiment*, whether real or conceptual (e.g., playing the lottery, testing whether a newborn is a boy, measuring a person's height). We assign probabilities to *possible outcomes* of the experiment. Thus, we first need to develop an abstract model for an experiment. An experiment has two stages: (1) an action, for example, tossing a coin, selecting a passer-by or turning a roulette, and (2) the recording of an *outcome*, for example, the side shown by the coin, the age of the passer-by or the number indicated by the roulette. In probability theory an experiment (real or conceptual) is modeled by *all its possible outcomes*, i.e., by a *set*, which we call the *sample space* (מרחב מדיגם). It should be emphasized that a set is a mathematical entity, independent of any intuitive background.

We will usually denote the sample space by Ω and its elements by ω .

Examples:

- (a) The experiment is *tossing a coin* and recording the upper side of the coin: the sample space comprises two possible outcomes, $\Omega = \{H, T\}$.
- (b) The experiment is *tossing a coin three times* and recording, keeping track of the the order, each of the three results: The sample space comprises all possible triples of Heads and Tails,

$$\Omega = \{(a_1, a_2, a_3) : a_i \in \{H, T\}\} = \{H, T\}^3.$$

In more detail,

$$\Omega = \{(H, H, H), (H, H, T), (H, T, H), (H, T, T), \\ (T, H, H), (T, H, T), (T, T, H), (T, T, T)\}.$$

Note that for the same action, tossing a coin three times, we could record a different datum, for example, the number of Heads. In that case the probability space would be

$$\Omega = \{0, 1, 2, 3\}.$$

The main message from this example is that the sample space is not uniquely determined by the action of the experiment.

- (c) The experiment is *throwing two distinguishable dice* and recording the number shown by each die: The sample space comprises ordered pairs of numbers between 1 and 6,

$$\Omega = \{1, \dots, 6\}^2 = \{(1, 1), (1, 2), \dots, (5, 6), (6, 6)\}.$$

- (d) The experiment is *throwing two indistinguishable dice* and recording the pair of numbers shown by the dice: here we have no way distinguishing between (2, 3) and (3, 2). The sample space is

$$\Omega = \{(i, j) : 1 \leq i \leq j \leq 6\}.$$

- (e) The experiment is *throwing a dart into a unit circle* and recording the distance from the center. Then,

$$\Omega = [0, 1].$$

If we rather record the position of the dart, then we can take

$$\Omega = \{(r, \theta) : 0 \leq r \leq 1, 0 \leq \theta < 2\pi\} = [0, 1] \times [0, 2\pi),$$

but also

$$\Omega = \{(x, y) : x^2 + y^2 \leq 1\}.$$

- (f) The experiment is *an infinite sequence of coin tosses* recording the outcome of each toss: the sample space is

$$\Omega = \{H, T\}^{\mathbb{N}}.$$

This set is non-countable, and it can be identified with the segment (0, 1) via the binary representation of rational numbers.

- (g) The experiment is *watching a grain of pollen move in a fluid at rest during a unit time interval* (an experiment performed by botanist Robert Brown in 1827): it is common to take for sample space the set of all continuous functions from $[0, 1]$ to \mathbb{R}^3 ,

$$\Omega = C([0, 1]; \mathbb{R}^3).$$

- (h) The experiment is the following: a person throws a coin; if the result is Head he takes an exam in probability, which he either passes or fails; if the result is Tail he goes to sleep and we measure the duration of his sleep (in hours). This sample space is

$$\Omega = \{(H, P), (H, F)\} \cup \{(T, x) : x > 0\},$$

which we can write also as

$$\Omega = \{H\} \times \{P, F\} \cup \{T\} \times \mathbb{R}^+.$$

The sample space is the primitive notion of probability theory. It provides a model of an experiment in the sense that every thinkable outcome (even if extremely unlikely) is identified with one, and only one, sample point.

1.2 Events

Suppose that we throw a die and record the number on the upper face. The set of all possible outcomes (the sample space) is $\Omega = \{1, \dots, 6\}$. What about the result "the outcome is even"? Even outcome is *not* an element of Ω . It is a property shared by several elements in the sample space. It corresponds to the *subset* $\{2, 4, 6\}$ of Ω . "The outcome is even" is therefore not an *elementary* outcome of the experiment. It is an aggregate of elementary outcomes, which we call an *event* (מאורע).

Definition 1.1 Let Ω be a sample space. An event is a subset of the sample space.

Let $A \subset \Omega$ be an event. If the experiment yielded an outcome $\omega \in \Omega$, we say that A has occurred (המאורע התרחש) if $\omega \in A$; otherwise we say that A has not occurred.

The intuitive terms of "outcome" and "event" have been incorporated within an abstract framework of a set and its subsets. As such, we can perform on events

set-theoretic operations of unions, intersections and complementations. All set-theoretic relations apply as they are to events.

Let Ω be a sample space corresponding to a certain experiment. What is the collection of all possible events? The obvious answer is the power set 2^Ω of Ω . It turns out that in certain cases (and this is where measure-theoretical monsters are hiding), it is necessary to restrict the collection of events to a sub-collection of 2^Ω . While we leave the reasons to a more advanced course, there are certain requirements that we wish the collection of events to fulfill:

1. *Closure under complementation*: If $A \subseteq \Omega$ is an event so is A^c . We want this property to hold, because if we are allowed to ask whether A has occurred, we should also be allowed to ask whether A has not occurred.
2. *Closure under countable unions*: If $A_n \subseteq \Omega$ are events, so is their union $\bigcup_{n=1}^{\infty} A_n$. We want this property to hold, because if we are allowed to ask whether each A_n has occurred, we should be allowed to ask whether at least one of the A_n has occurred.
3. Ω is an event. This event stands for "some outcome among all possibilities".

In other words, if we restrict the collection of events to a sub-collection of the power set of Ω , we want this collection to be a σ -algebra.

(2 hrs)  (2 hrs)

Comment: You may wonder why not require the collection of event to be closed with respect to *any* (not necessarily countable) union. The reason is once again in the realm of measure theory. One may end up with an inconsistent theory.

Definition 1.2 A pair (Ω, \mathcal{F}) , where Ω is a set and \mathcal{F} is a σ -algebra of events is called a measurable space (מדיד מרחב).

Examples:

- (a) Consider the experiment of tossing three coins with $\Omega = \{H, T\}^3$. The event "second toss was Head" is


$$\{(H, H, H), (H, H, T), (T, H, H), (T, H, T)\}.$$

- (b) Consider the experiment “waiting for the fish to bite” (in hours). The sample space is

$$\Omega = \{x : x > 0\} = (0, \infty).$$

The event “waited more than an hour and less than two” is

$$\{x \in \Omega : 1 < x < 2\} = (1, 2).$$

 *Problem 1.1* Construct a sample space corresponding to filling a single column in the Toto. Define two events that are disjoint. Define three events that are mutually disjoint, and whose union is Ω .

1.3 Probability

Roughly speaking, a probability is an assignment of numbers to possible outcomes of experiments. As discussed, these numbers may represent a likelihood, or a measure of belief. Mathematically, however, a probability is an assignment rule, i.e., a function. The immediate issues, then are to define this function’s domain, range and other properties.

The most rudimentary construction of a probability is the following: to each point $\omega \in \Omega$ in the sample space, i.e., to each elementary outcome, assign a number, $p(\omega)$. We require these numbers to be non-negative—there is no probability less than zero—and as a normalization convention, we require them to add up to 1. Given an event $A \subset \Omega$, we define its probability as the sum of the probabilities of the sample points it comprises,

$$P(A) = \sum_{\omega \in A} p(\omega).$$

Thus, probability is a function $P : \mathcal{F} \rightarrow \mathbb{R}_+$.

The function P satisfies a number of properties that are readily derived. First, by construction, for every event A , $0 \leq P(A) \leq 1$. Moreover,

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

Second, if A and B are disjoint, then

$$P(A \cup B) = \sum_{\omega \in A \cup B} p(\omega) = \sum_{\omega \in A} p(\omega) + \sum_{\omega \in B} p(\omega) = P(A) + P(B),$$

where we needed A and B to be disjoint in order not to double count.

But there are also some subtleties. If the sample space is finite, then everything is well-defined. What about an infinite sample space? If Ω is infinite, yet countable, then the sum

$$\sum_{\omega \in \Omega} p(\omega)$$

is well defined provided that no matter how we arrange the points ω into a sequence, the resulting series converges absolutely. Thus, we may require that

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

The sum of $p(\omega)$ over every subset of Ω , whether finite or infinite, is well-defined. Problems arise when Ω is not countable. A sum comprising an uncountable number of addends is not even defined; as it turns out, this is not simply a lack of knowledge—an omission of the first year curriculum.

We define a notion of probability that applies to all cases, whether the probability space is countable or not. The idea is to attribute a probability to events, rather than to elements of the sample space. In the countable case, we can always resort to the rudimentary construction described above.

Definition 1.3 Let (Ω, \mathcal{F}) be a measurable space. A probability (הסתברות) on (Ω, \mathcal{F}) is a function P assigning a number to every event in \mathcal{F} (interpreted as the likelihood that this event will occur or has occurred). The function P has to satisfy the following properties:

1. For every event $A \in \mathcal{F}$, $0 \leq P(A)$ (there is no negative probability).
2. $P(\Omega) = 1$ (the probability that some outcome has occurred is one).
3. Let (A_n) be a sequence of mutually disjoint events. Then,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

This property is called countable additivity (אדיטיביות בת מנייה).

The triple (Ω, \mathcal{F}, P) is called a probability space (מרחב הסתברות).

(3 hrs)  (3 hrs)

The following results are immediate:

Proposition 1.1 Let (Ω, \mathcal{F}, P) be a probability space. Then:

1. $P(\emptyset) = 0$.
2. For every finite sequence of N disjoint events (A_n)

$$P\left(\bigcup_{n=1}^N A_n\right) = \sum_{n=1}^N P(A_n).$$

3. For every event $A \in \mathcal{F}$,

$$P(A^c) = 1 - P(A).$$

4. For every $A \in \mathcal{F}$, $P(A) \leq 1$.

Proof:

1. Take a sequence (A_n) with $A_n = \emptyset$ for every n . Then,

$$\bigcup_{n=1}^{\infty} A_n = \emptyset.$$

By countable additivity,

$$\sum_{n=1}^{\infty} P(\emptyset) = P(\emptyset),$$

which implies that $P(\emptyset) = 0$.

2. Take $A_k = \emptyset$ for $k > n$ and apply the first assertion along with countable additivity.
3. Since $A \cup A^c = \Omega$, it follows from the second assertion that

$$P(A) + P(A^c) = P(\Omega) = 1.$$

4. The last assertion is immediate.



Examples:

- (a) The experiment is tossing a (fair) coin. The sample space is $\Omega = \{H, T\}$. The σ -algebra of events consists of all subsets of Ω ,

$$\mathcal{F} = 2^\Omega = \{\emptyset, \{H\}, \{T\}, \Omega\}.$$

We define a probability function by taking $P(\{H\}) = P(\{T\}) = 1/2$. Note that this defines indeed a unique probability function of \mathcal{F} . Choosing, for example $P(\{H\}) = 1/2$ and $P(\{T\}) = 1/4$ would have been inconsistent, as

$$1 = P(\Omega) = P(\{H\} \cup \{T\}) = P(\{H\}) + P(\{T\}) = \frac{3}{4}.$$

- (b) The experiment is throwing a fair die and recording the outcome, which is an integer between 1 and 6. The sample space is $\Omega = \{1, \dots, 6\}$. We take the σ -algebra of all subsets of Ω . This σ -algebra contains 2^6 events. We define a probability function by setting the probability of all singletons:

$$P(\{i\}) = \frac{1}{6}, \quad \forall i \in \{1, \dots, 6\}.$$

This defines uniquely a probability function since every event is a finite disjoint union of singletons; its probability is the sum of the probabilities of the singletons it contains.

Comment: You may justly ask yourself how did I know which probability space to assign to each experiment. What I did here is to *model* the likelihoods of outcomes of an experiment that hasn't yet been performed. A model is a model. It may relate to reality or not. In the present case, I applied symmetry considerations. Since we believe that all outcomes of the experiment are equally likely, i.e., all singletons should have the same probability, this probability could only assume one consistent value.

Having defined what a probability is, we can further derive properties satisfied by *all* probability function.

Proposition 1.2 (Monotonicity of probability) Let (Ω, \mathcal{F}, P) be a probability space. If A, B are events such that $A \subseteq B$, then

$$P(A) \leq P(B).$$

Proof: If $A \subseteq B$, then

$$B = A \cup (B \setminus A).$$

By additivity and the positivity of probability,

$$P(B) = P(A) + P(B \setminus A) \geq P(A).$$

■

Proposition 1.3 (Probability of a union) Let (Ω, \mathcal{F}, P) be a probability space. For every two events $A, B \in \mathcal{F}$,

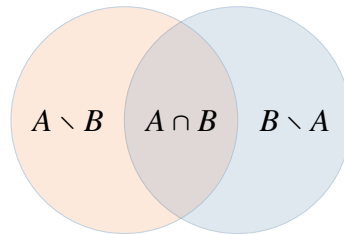
$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Proof: We have

$$A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$$

$$A = (A \setminus B) \cup (A \cap B)$$

$$B = (B \setminus A) \cup (A \cap B),$$



By additivity,

$$P(A \cup B) = P(A \setminus B) + P(B \setminus A) + P(A \cap B)$$

$$P(A) = P(A \setminus B) + P(A \cap B)$$

$$P(B) = P(B \setminus A) + P(A \cap B).$$

It remains to subtract the last two equations from the first. ■

Proposition 1.4 Let (Ω, \mathcal{F}, P) be a probability space. For every three events $A, B, C \in \mathcal{F}$,

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &\quad + P(A \cap B \cap C). \end{aligned}$$


Proof: Using the binary relation,


$$\begin{aligned} P(A \cup B \cup C) &= P(A \cup B) + P(C) - P((A \cup B) \cap C) \\ &= P(A) + P(B) - P(A \cap B) + P(C) - P((A \cap C) \cup (B \cap C)), \end{aligned}$$

and it remains to apply the binary relation for the last expression. ■

Proposition 1.5 (Inclusion-exclusion principle (עקרון הכללה הרחבה)) For any n events $(A_i)_{i=1}^n$,

$$\begin{aligned} P(A_1 \cup \dots \cup A_n) &= \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) \\ &\quad + \dots + (-1)^{n+1} P(A_1 \cap \dots \cap A_n). \end{aligned}$$

 *Problem 1.2* Prove the inclusion-exclusion principle.

 *Problem 1.3* Let A, B be two events in a probability space. Show that the probability that either A or B has occurred, but not both, is $P(A) + P(B) - 2P(A \cap B)$.

Lemma 1.1 (Boole's inequality) Let (Ω, \mathcal{F}, P) be a probability space. Probability is sub-additive (תת אדיטיביות): for every sequence (A_n) of events,

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} P(A_k),$$

where the right-hand side may be infinite, in which case the inequality holds trivially.

Proof: Define the following sequence of events,

$$B_1 = A_1 \quad B_2 = A_2 \setminus A_1, \quad B_3 = A_3 \setminus (A_1 \cup A_2), \quad \dots, \quad B_n = A_n \setminus \left(\bigcup_{k=1}^{n-1} A_k \right).$$

The B_n are disjoint and their union equals the union of the A_n . Indeed, if $\omega \in \bigcup_{k=1}^{\infty} A_k$, then the set

$$\{k \in \mathbb{N} : \omega \in A_k\}$$

is non-empty. This set has a minimum k^* and $\omega \in B_{k^*}$. It follows that $\omega \in \bigcup_{k=1}^{\infty} B_k$, hence

$$\bigcup_{k=1}^{\infty} A_k \subset \bigcup_{k=1}^{\infty} B_k.$$

The opposite direction is trivial as $B_n \subseteq A_n$ for all n . It follows that

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = P\left(\bigcup_{k=1}^{\infty} B_k\right) = \sum_{k=1}^{\infty} P(B_k) \leq \sum_{k=1}^{\infty} P(A_k).$$

■

1.4 Countable, or discrete probability spaces

The simplest probability spaces are those whose sample spaces include countably many points. In such case, we may always take the collection of events to be the power set,

$$\mathcal{F} = 2^{\Omega}.$$

Note that a countable set *can be* arranged into a sequence, however, generally, a countable set is not ordered.

For countable probability spaces, a probability function P on \mathcal{F} is fully determined by its value for every singleton $\{\omega\}$, i.e., by the probability assigned to every *elementary event*. Indeed, let $P(\{\omega\}) \equiv p(\omega)$ be given. Since every event A can be expressed as a finite, or countable union of disjoint singletons,

$$A = \bigcup_{\omega \in A} \{\omega\},$$

it follows from the additivity property that

$$P(A) = \sum_{\omega \in A} p(\omega).$$

A particular case often arising in applications is where the sample space is finite and all elementary events have the same probability; we denote by $|\Omega|$ the size of the sample space and by p the probability of a singleton, i.e.,

$$p(\omega) = p, \quad \forall \omega \in \Omega.$$

By the properties of the probability function,

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

i.e., $p = 1/|\Omega|$. The probability of every event A is then

$$P(A) = \sum_{\omega \in A} p(\omega) = p|A| = \frac{|A|}{|\Omega|}.$$

Comment: Recall that a probability space is a model for an experiment. There is no a priori reason why all outcomes should be equally probable. It is an assumption that should be made only when believed to be applicable.

Examples:

- (a) Two dice are rolled. What is the probability that the sum is 7? A sample space that is convenient for this problem is constructed by viewing the two dice as distinguishable (e.g. one blue and one red) and taking

$$\Omega = \{(i, j) : 1 \leq i, j \leq 6\}.$$

Due to symmetry considerations, it is natural to assume that each of the $|\Omega| = 36$ outcomes is equally likely. The event "sum equals 7" is

$$A = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\},$$

so that $P(A) = |A|/|\Omega| = 6/36$.

- (b) There are 11 balls in a jar, 6 white and 5 black. Two balls are drawn at random. What is the probability of drawing one white ball and one black ball?

To solve this problem, it is helpful to imagine that the balls are numbered from 1 to 11, the six white balls being numbered from 1 to 6 (note the use of a thought experiment). The sample space consists of all possible pairs,

$$\Omega = \{(i, j) : 1 \leq i < j \leq 11\},$$

and its size is $|\Omega| = \binom{11}{2}$. Invoking once again symmetry considerations, we assume that all pairs are equally likely. The event “one black ball and one white ball” is

$$A = \{(i, j) : 1 \leq i \leq 6, 7 \leq j \leq 11\}.$$

The size of A is the number of possibility to choose one white ball out of six, and one black ball out of five, i.e.,

$$P(A) = \frac{\binom{6}{1}\binom{5}{1}}{\binom{11}{2}} = \frac{6 \cdot 5}{10 \cdot 11 \div 2} = \frac{6}{11}.$$

- (c) A deck of 52 cards is dealt among four players. What is the probability that one of the players received all 13 spades?

A natural sample space Ω is the set of all possible partitions of 52 cards among 4 players; its size is

$$|\Omega| = \frac{52!}{13! 13! 13! 13!}$$

(the number of possibilities to order 52 cards divided by the number of internal orders). Once again, symmetry considerations suggest that all partitions should have the same probability.

Let A_i be the event that the i -th player has all spades, and A be the event that *some* player has all spades; clearly,


$$A = A_1 \cup A_2 \cup A_3 \cup A_4.$$


For each i ,

$$|A_i| = \frac{39!}{13! 13! 13!},$$

hence,

$$P(A) = 4P(A_1) = 4 \times \frac{39! 13! 13! 13! 13!}{52! 13! 13! 13!} \approx 6.3 \times 10^{-12}.$$

 **Problem 1.4** Prove that it is not possible to construct a probability function on the sample space of integers, such that $P(\{i\}) = P(\{j\})$ for all i, j .

 **Problem 1.5** A fair coin is tossed five times. What is the probability that there was at least one instance of two Heads in a row? Start by building the probability space.

Next, we study a number of examples, all concerning finite probability spaces with equally likely outcomes. The importance of these examples stems from the fact that they are representatives of classes of problems recurring in many applications.

Example: (The birthday paradox) In a random assembly of n people, what is the probability that none of them share the same date of birth?

We assume that $n < 365$ and ignore leap years. We take for sample space Ω the set of all possible date-of-birth assignments to n people (order matters). For example, if $n = 2$, then

$$(\text{Jul-1, Feb-15}) \in \Omega.$$

The size of the sample space of size $|\Omega| = 365^n$. By symmetry considerations, all birthday assignments to n people are equally probably.

Let A_n be the event that all dates-of-birth are different. Then,

$$|A_n| = 365 \times 364 \times \cdots \times (365 - n + 1),$$

hence

$$P(A_n) = \frac{|A_n|}{|\Omega|} = 1 \times \frac{364}{365} \times \cdots \times \frac{365 - n + 1}{365}.$$

Values for various n are

$$P(A_{23}) < 0.5 \quad P(A_{50}) < 0.03 \quad P(A_{100}) < 3.3 \times 10^{-7}.$$

Comment: Some may have guessed, erroneously, that $P(A_{100}) \approx 1 - 100/365 > 2/3$.

Generalization Suppose that there were N days in a year and $n \leq N$ people; denote by A_n^N the event that all dates-of-birth are different. The same analysis yields that the probability of A_n^N is

$$P(A_n^N) = \frac{|A_n^N|}{|\Omega|} = \frac{N \times (N-1) \times \cdots \times (N - (n-1))}{N^n} = \prod_{i=0}^{n-1} \left(1 - \frac{i}{N}\right).$$

For any given N and n we can calculate this probability. However, it is more instructive to obtain an estimate that will allow us to “feel” when to expect this

probability to be large (close to 1) and when to expect it to be small (close to 0). Since we have a product, it is useful to evaluate its logarithm

$$\log P(A_n^N) = \sum_{i=0}^{n-1} \log \left(1 - \frac{i}{N} \right).$$

To bound $P(A_n^N)$ from above, we recall that $\log(1+x) \leq x$ for all real x (exercise: prove it), hence

$$P(A_n^N) \leq \exp \left(-\frac{n(n-1)}{2N} \right).$$

which is valid for any n . To get a lower bound we use the inequality $x - x^2 \leq \log(1+x)$ which is valid for any $-\frac{1}{4} \leq x$ (exercise: prove it). Thus, as long as $n \leq N/4$,

$$\begin{aligned} \log P(A_n^N) &\geq -\frac{\sum_{i=0}^{n-1} i}{N} - \frac{\sum_{i=0}^{n-1} i^2}{N^2} = -\frac{n(n-1)}{2N} - \frac{n(n-1)(2n-1)}{6N^2}, \\ &= -\frac{n(n-1)}{2N} \left(1 + \frac{2n-1}{3N} \right). \end{aligned}$$

Hence, for $n \leq N/4$,

$$\exp \left(-\frac{n(n-1)}{2N} \left(1 + \frac{2n-1}{3N} \right) \right) \leq P(A_n^N) \leq \exp \left(-\frac{n(n-1)}{2N} \right).$$

When n is small compared to N the upper and lower bounds are quite close to each other.

When n is small compared to \sqrt{N} , the probability of all dates-of-births being different is close to 1, whereas when n is large compared to \sqrt{N} , this probability is close to 0.

▲ ▲ ▲

Comment: You may ask yourself what is so “generic” about the birthday paradox. Here is an application: suppose that a computer assigns IDs to users by randomly choosing a 16-bit words. As long as the number of users is less than the square-root of $2^{16} = 256$, there is a low probability that the same ID will be assigned to two users.

Example: (The inattentive secretary, or the matching problem) A secretary places randomly n letters into n envelopes. What is the probability that no letter reaches

its destination? What is the probability that exactly k letters reach their destination?

We start by constructing the sample space. Assume that the letters and the envelopes are numbered from one to n . The sample space Ω consists of all possible assignments of letters into envelopes, i.e., the set of all *permutations* of the numbers 1-to- n . If for example $n = 5$, then

$$(1, 3, 5, 4, 2) \in \Omega$$

(Letter 1 in Envelope 1, Letter 3 in Envelope 2, etc.). The first question can be reformulated as follows: take a random one-to-one function from $\{1, \dots, n\}$ to itself; what is the probability that it has no fixed points?

If A is the event that no letter has reached its destination, then its complement, A^c is the event that at least one letter has reached its destination (at least one fixed point). Let B_i be the event that the i -th letter reached its destination, then

$$A^c = \bigcup_{i=1}^n B_i.$$

We apply the inclusion-exclusion principle:

$$\begin{aligned} P(A^c) &= \sum_{i=1}^n P(B_i) - \sum_{i < j} P(B_i \cap B_j) + \sum_{i < j < k} P(B_i \cap B_j \cap B_k) - \dots \\ &= \binom{n}{1} P(B_1) - \binom{n}{2} P(B_1 \cap B_2) + \binom{n}{3} P(B_1 \cap B_2 \cap B_3) - \dots \\ &\quad + (-1)^{n+1} \binom{n}{n} P(B_1 \cap \dots \cap B_n), \end{aligned}$$

where we used the symmetry of the problem to determine, for example, that $P(B_i)$ is independent of i . Now,

$$\begin{aligned} P(B_1) &= \frac{|B_1|}{|\Omega|} = \frac{(n-1)!}{n!} \\ P(B_1 \cap B_2) &= \frac{|B_1 \cap B_2|}{|\Omega|} = \frac{(n-2)!}{n!}, \end{aligned}$$

etc. It follows that

$$\begin{aligned} P(A^c) &= n \frac{1}{n} - \binom{n}{2} \frac{(n-2)!}{n!} + \binom{n}{3} \frac{(n-3)!}{n!} - \dots + (-1)^{n+1} \binom{n}{n} \frac{0!}{n!} \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^{n+1} \frac{1}{n!} \\ &= \sum_{k=1}^n \frac{(-1)^{k+1}}{k!}. \end{aligned}$$

(5 hrs)  (5 hrs)

For large n ,

$$P(A) = 1 - P(A^c) = \sum_{k=0}^n \frac{(-1)^k}{k!} \approx e^{-1},$$

from which we deduce that for large n , the probability that no letter has reached its destination is about 0.37. The fact that the limit is neither 0 nor 1 may be surprising (one could have argued equally well that the limit should be both 0 and 1...). Note that as a side result, the number of permutations that have no fixed points is

$$|A| = |\Omega| P(A) = n! \sum_{\ell=0}^n \frac{(-1)^\ell}{\ell!}.$$

Now to the second part of this question. Before we answer what is the number of permutations that have exactly k fixed points, let's compute the number of permutations that have exactly k *specific* fixed points; for example

$$\begin{aligned} 1 &\mapsto 1, \dots, k \mapsto k \\ k+1 &\not\mapsto k+1, \dots, n \not\mapsto n. \end{aligned}$$

This number equals the number of permutations of $n - k$ elements not having a fixed point,

$$(n-k)! \sum_{\ell=0}^{n-k} \frac{(-1)^\ell}{\ell!}.$$

The choice of k fixed points is exclusive, e.g., the events

$$B_{1,3} = \{1 \text{ and } 3 \text{ are the only fixed points}\}$$

and

$$B_{1,4} = \{1 \text{ and } 4 \text{ are the only fixed points}\}$$

are disjoint. To find the total number of permutations that have exactly k fixed points, we need to multiply the number of permutations having k specific fixed points by the number of ways to choose k elements out of n . Thus, if C denotes the event that there are exactly k fixed points, then

$$|C| = \binom{n}{k} (n-k)! \sum_{\ell=0}^{n-k} \frac{(-1)^\ell}{\ell!} = \frac{n!}{k!} \sum_{\ell=0}^{n-k} \frac{(-1)^\ell}{\ell!},$$


and

$$P(C) = \frac{|C|}{n!} = \frac{1}{k!} \sum_{\ell=0}^{n-k} \frac{(-1)^\ell}{\ell!}.$$

For large n and fixed k we have


$$P(C) \approx \frac{e^{-1}}{k!}.$$

We will return to such expressions later on, in the context of the *Poisson distribution*. ▲▲▲

 **Problem 1.6** Seventeen people attend a party. At the end of it, the drunk people collect at random a hat from the hat hanger. What is the probability that

1. At least someone got their own hat.
2. John Doe got his own hat.
3. Exactly 3 people got their own hats.
4. Exactly 3 people got their own hats, one of which is John Doe.

As usual, start by specifying the probability space.

 **Problem 1.7** A deck of cards is dealt out. What is the probability that the fourteenth card dealt is an ace? What is the probability that the first ace occurs on the fourteenth card?

1.5 Continuity of probability

An important property of the probability function is its *continuity*, in the sense that if a sequence of events (A_n) has a limit, then $P(\lim_{n \rightarrow \infty} A_n) = \lim_{n \rightarrow \infty} P(A_n)$.

We start with a “soft” version:

Theorem 1.1 (Continuity for increasing sequences) Let (A_n) be an increasing sequence of events, then

$$P(\lim_{n \rightarrow \infty} A_n) = \lim_{n \rightarrow \infty} P(A_n).$$

Comment: We have already seen that increasing sequence of event converge, with

$$\lim_{n \rightarrow \infty} A_n = \bigcup_{k=1}^{\infty} A_k.$$

Note that for every n ,

$$\bigcup_{k=1}^n A_k = A_n,$$

hence

$$P\left(\bigcup_{k=1}^n A_k\right) = P(A_n),$$

but we can't just replace n by ∞ .

Proof: For an increasing sequence of events, $\lim_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} A_n$. Construct the following sequence of disjoint events,

$$\begin{aligned} B_1 &= A_1 \\ B_2 &= A_2 \setminus A_1 \\ B_3 &= A_3 \setminus A_2, \end{aligned}$$

etc. Clearly,

$$\bigcup_{i=1}^n B_i = \bigcup_{i=1}^n A_i = A_n \quad \text{and} \quad \bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$$


(every ω contained in at least one A_n is also contained in at least one B_n). Now,

$$\begin{aligned} P(\lim_{n \rightarrow \infty} A_n) &= P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} B_i\right) \\ &= \sum_{i=1}^{\infty} P(B_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^n P(B_i) \\ &= \lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n B_i\right) = \lim_{n \rightarrow \infty} P(A_n). \end{aligned}$$

■

Comment: Since $P(A_n)$ is an increasing function,

$$P(A_n) \nearrow P(\lim_{n \rightarrow \infty} A_n).$$

 *Problem 1.8 (Continuity for decreasing sequences)* Prove that if (A_n) is a decreasing sequence of events, then

$$P(\lim_{n \rightarrow \infty} A_n) = \lim_{n \rightarrow \infty} P(A_n).$$

More precisely, $P(A_n) \searrow P(\lim_{n \rightarrow \infty} A_n)$.

The next two lemmas hold for arbitrary sequences of events, without assuming the existence of a limit.

Lemma 1.2 (Fatou) Let (A_n) be a sequence of events, then

$$P(\liminf_{n \rightarrow \infty} A_n) \leq \liminf_{n \rightarrow \infty} P(A_n).$$

Proof: First, note that the lim infs on both side represent different notions. The lim inf on the left-hand side refers to sets,

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k \equiv \bigcup_{n=1}^{\infty} G_n.$$

The lim inf of the right-hand side refers to numbers. Recall that if (a_n) is a sequence of numbers, then

$$\liminf_{n \rightarrow \infty} a_n = \lim_{k \rightarrow \infty} \inf_{k \geq n} a_n,$$

which is also the smallest partial limit (it exists because the sequence $P(A_n)$ is bounded).

The sequence of events (G_n) is increasing, and therefore by the continuity theorem for increasing sequences,

$$\lim_{n \rightarrow \infty} P(G_n) = P(\lim_{n \rightarrow \infty} G_n) = P\left(\bigcup_{n=1}^{\infty} G_n\right) = P(\liminf_{n \rightarrow \infty} A_n).$$

On the other hand, fix n . Since $G_n = \bigcap_{k=n}^{\infty} A_k$, it follows that

$$G_n \subseteq A_k \quad \text{for all } k \geq n,$$

which in turn implies that

$$P(G_n) \leq P(A_k) \quad \text{for all } k \geq n,$$

and hence

$$P(G_n) \leq \inf_{k \geq n} P(A_k).$$

The left hand side converges to $P(\liminf_{n \rightarrow \infty} A_n)$ whereas the right hand side converges to $\liminf_{n \rightarrow \infty} P(A_n)$, which concludes the proof. ■

Lemma 1.3 (Reverse Fatou) Let (A_n) be a sequence of events, then

$$\limsup_{n \rightarrow \infty} P(A_n) \leq P(\limsup_{n \rightarrow \infty} A_n).$$

Proof: Recall that

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \equiv \bigcap_{n=1}^{\infty} H_n$$

is the set of outcomes that occur “infinitely often”. The sequence (H_n) is decreasing, and therefore

$$\lim_{n \rightarrow \infty} P(H_n) = P(\lim_{n \rightarrow \infty} H_n) = P\left(\bigcap_{n=1}^{\infty} H_n\right) = P(\limsup_{n \rightarrow \infty} A_n).$$

On the other hand, since $H_n = \bigcup_{k=n}^{\infty} A_k$, it follows that

$$P(H_n) \geq \sup_{k \geq n} P(A_k).$$

The left hand side converges to $P(\limsup_{n \rightarrow \infty} A_n)$ whereas the right hand side converges to $\limsup_{n \rightarrow \infty} P(A_n)$, which concludes the proof. ■

Theorem 1.2 (Continuity of probability) If a sequence of events (A_n) has a limit, then

$$P(\lim_{n \rightarrow \infty} A_n) = \lim_{n \rightarrow \infty} P(A_n).$$

Proof: This is an immediate consequence of the two Fatou lemmas, for

$$\limsup_{n \rightarrow \infty} P(A_n) \leq P(\limsup_{n \rightarrow \infty} A_n) = P(\liminf_{n \rightarrow \infty} A_n) \leq \liminf_{n \rightarrow \infty} P(A_n).$$

■

Lemma 1.4 (First Borel-Cantelli) Let (A_n) be a sequence of events such that $\sum_n P(A_n) < \infty$. Then,

$$P(\limsup_{n \rightarrow \infty} A_n) = P(\{x : x \in A_n \text{ infinitely often}\}) = 0.$$

Proof: Let as above $H_n = \cup_{k=n}^{\infty} A_k$. Since $P(H_n) \searrow P(\limsup_{n \rightarrow \infty} A_n)$, then for all m

$$P(\limsup_{n \rightarrow \infty} A_n) \leq P(H_m) \leq \sum_{k \geq m} P(A_k).$$

Letting $m \rightarrow \infty$ and using the fact that the right hand side is the tail of a converging series we get the desired result. ■

🔗 *Problem 1.9* Does the “reverse Borel-Cantelli” hold? Namely, is it true that if $\sum_n P(A_n) = \infty$ then

$$P(\limsup_{n \rightarrow \infty} A_n) > 0?$$

No. Construct a counter example.

Example: Consider the following scenario. At a minute to noon we insert into an urn balls numbered 1-to-10 and remove the ball numbered “10”. At half a minute to noon we insert balls numbered 11-to-20 and remove the ball numbered “20”, and so on. Which balls are inside the urn at noon? Clearly all integers except for the “10n”.

Now we vary the situation: the first time we remove the ball numbered “1”, next time the ball numbered “2”, etc. Which balls are inside the urn at noon? none.

In the third variation we remove each time a ball at random (from those inside the urn). Are there any balls left at noon? If this question is too bizarre, here is a more sensible picture. Our sample space consists of random sequences of numbers, whose elements are distinct, and whose first element is in the range 1-to-10, its second element is in the range 1-to-20, and so on. We are asking what is the probability that such a sequence contains all integers?

Let’s focus on ball number “1” and denote by E_n then event that it is still inside the urn after n steps. We have

$$P(E_n) = \frac{9}{10} \times \frac{18}{19} \times \cdots \times \frac{9n}{9n+1} = \prod_{k=1}^n \frac{9k}{9k+1}.$$

The events (E_n) form a decreasing sequence, whose countable intersection corresponds to the event that the first ball was not ever removed. Now,

$$\begin{aligned} P(\lim_{n \rightarrow \infty} E_n) &= \lim_{n \rightarrow \infty} P(E_n) = \lim_{n \rightarrow \infty} \prod_{k=1}^n \frac{9k}{9k+1} \\ &= \lim_{n \rightarrow \infty} \left[\prod_{k=1}^n \frac{9k+1}{9k} \right]^{-1} = \lim_{n \rightarrow \infty} \left[\prod_{k=1}^n \left(1 + \frac{1}{9k} \right) \right]^{-1} \\ &= \lim_{n \rightarrow \infty} \left[\left(1 + \frac{1}{9} \right) \left(1 + \frac{1}{18} \right) \cdots \right]^{-1} \\ &\leq \lim_{n \rightarrow \infty} \left(1 + \frac{1}{9} + \frac{1}{18} + \cdots \right)^{-1} = 0. \end{aligned}$$

Thus, there is zero probability that the ball numbered “1” is inside the urn after infinitely many steps. The same holds ball number “2”, “3”, etc. If F_n denotes the event that the n -th ball has remained inside the box at noon, then

$$P\left(\bigcup_{n=1}^{\infty} F_n\right) \leq \sum_{n=1}^{\infty} P(F_n) = 0.$$

▲ ▲ ▲