3 Axioms of Probability

The question here is: how can we mathematically define a random experiment? What we have are *outcomes* (which tell you exactly what happens), *events* (sets containing certain outcomes), and *probability* (which attaches to every event the likelihood that it happens). We need to agree on which properties these objects must have in order to compute with them and develop a theory.

When we have finitely many equally likely outcomes all is clear and we have already seen many examples. However, as is common in mathematics, infinite sets are much harder to deal with. For example, we will soon see what it means to choose a random point within a unit circle. On the other hand, we will also see that there is no way to choose at random a positive integer — remember that "at random" means all choices are equally likely, unless otherwise specified.

Finally, a *probability space* is a triple $(\Omega, \mathcal{F}, P)$. The first object Ω is an arbitrary set, repesenting the set of outcomes, sometimes called the *sample space*.

The second object $\mathcal{F}$ is a collection of events, that is, a set of subsets of Ω. Therefore, an event $A \in \mathcal{F}$ is necessarily a subset of Ω. Can we just say that each $A \subset \Omega$ is an event? In this course *you can assume so without worry, although there are good reasons for not assuming so in general*! I will give the definition of what properties $\mathcal{F}$ needs to satisfy, but this is only for illustration and you should take a course in measure theory to understand what is really going on. Namely, $\mathcal{F}$ needs to be a *σ-algebra*, which means (1) $\emptyset \in \mathcal{F}$, (2) $A \in \mathcal{F} \implies A^c \in \mathcal{F}$, and (3) $A_1, A_2, \dots \in \mathcal{F} \implies \cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

What is important is that you can take the complement A^c of an event A (i.e., A^c happens when A does not happen), unions of two or more events (i.e., $A_1 \cup A_2$ happens when either A_1 or A_2 happens), and intersections of two or more events (i.e., $A_1 \cap A_2$ happens when both A_1 and A_2 happen). We call events $A_1, A_2, \dots$ *pairwise disjoint* if $A_i \cap A_j = \emptyset$ if $i \neq j$ — that is, at most one of these events can occur.

Finally, the probability P is a number attached to every event A and satisfies the following three axioms:

<u>Axiom 1</u>. For every event A, $P(A) \geq 0$.

<u>Axiom 2</u>. $P(\Omega) = 1$.

<u>Axiom 3</u>. If $A_1, A_2, \dots$ is a sequence of pairwise disjoint events, then

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

Whenever we have an abstract definition such as this one, the first thing to do is to look for examples. Here are some.

Example 3.1. $\Omega = \{1, 2, 3, 4, 5, 6\}$,

$$P(A) = \frac{\text{(number of elements in } A)}{6}.$$

The random experiment here is rolling a fair die. Clearly, this can be generalized to any finite set with equally likely outcomes.

Example 3.2. $\Omega = \{1, 2, \dots\}$ and you have numbers $p_1, p_2, \dots \geq 0$ with $p_1 + p_2 + \dots = 1$. For any $A \subset \Omega$,

$$P(A) = \sum_{i \in A} p_i.$$

For example, toss a fair coin repeatedly until the first Heads. Your outcome is the number of tosses. Here, $p_i = \frac{1}{2^i}$.

Note that p_i cannot be chosen to be equal, as you cannot make the sum of infinitely many equal numbers to be 1.

Example 3.3. Pick a point from inside the unit circle centered at the origin. Here, $\Omega = \{(x, y) : x^2 + y^2 < 1\}$ and

$$P(A) = \frac{\text{(area of A)}}{\pi}.$$

It is important to observe that if A is a singleton (a set whose element is a single point), then $P(A) = 0$. This means that we cannot attach the probability to outcomes — you hit a single point (or even a line) with probability 0, but a "fatter" set with positive area you hit with positive probability.

Another important theoretical remark: this is a case where A cannot be an arbitrary subset of the circle — for some sets area cannot be defined!

Consequences of the axioms

(C0) $P(\emptyset) = 0$.

Proof. In Axiom 3, take all sets to be $\emptyset$.

(C1) If $A_1 \cap A_2 = \emptyset$, then $P(A_1 \cup A_2) = P(A_1) + P(A_2)$.

Proof. In Axiom 3, take all sets other than first two to be $\emptyset$.

(C2)

$$\boxed{P(A^c) = 1 - P(A).}$$

Proof. Apply (C1) to $A_1 = A$, $A_2 = A^c$.

(C3) $0 \leq P(A) \leq 1$.

Proof. Use that $P(A^c) \geq 0$ in (C2).

(C4) If $A \subset B$, $P(B) = P(A) + P(B \setminus A) \geq P(A)$.

Proof. Use (C1) for $A_1 = A$ and $A_2 = B \setminus A$.

(C5) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof. Let $P(A \setminus B) = p_1$, $P(A \cap B) = p_{12}$ and $P(B \setminus A) = p_2$, and note that $A \setminus B$, $A \cap B$, and $B \setminus A$ are pairwise disjoint. Then $P(A) = p_1 + p_{12}$, $P(B) = p_2 + p_{12}$, and $P(A \cup B) = p_1 + p_2 + p_{12}$.

(C6)

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) = {} & P(A_1) + P(A_2) + P(A_3) \\ & - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_2) \\ & + P(A_1 \cap A_2 \cap A_3) \end{aligned}$$

and more generally

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

This is called the *inclusion-exclusion formula* and is commonly used when it is easier to compute probabilities of intersections than of unions.

Proof. We prove this only for $n = 3$. Let $p_1 = P(A_1 \cap A_2^c \cap A_3^c)$, $p_2 = P(A_1^c \cap A_2 \cap A_3^c)$, $p_3 = P(A_1^c \cap A_2^c \cap A_3)$, $p_{12} = P(A_1 \cap A_2 \cap A_3^c)$, $p_{13} = P(A_1 \cap A_2^c \cap A_3)$, $p_{23} = P(A_1^c \cap A_2 \cap A_3)$, and $p_{123} = P(A_1 \cap A_2 \cap A_3)$. Again, note that all sets are pairwise disjoint and that the right hand side of (6) is

$$\begin{aligned} & (p_1 + p_{12} + p_{13} + p_{123}) + (p_2 + p_{12} + p_{23} + p_{123}) + (p_3 + p_{13} + p_{23} + p_{123}) \\ & - (p_{12} + p_{123}) - (p_{13} + p_{123}) - (p_{23} + p_{123}) \\ & + p_{123} \\ & = p_1 + p_2 + p_3 + p_{12} + p_{13} + p_{23} + p_{123} = P(A_1 \cup A_2 \cup A_3). \end{aligned}$$

Example 3.4. Pick an integer in $[1, 1000]$ at random. Compute the probability that it is divisible neither by 12 nor by 15.

The sample space consists of the 1000 integers between 1 and 1000 and let A_r be the subset consisting of integers divisible by r. The cardinality of A_r is $\lfloor 1000/r \rfloor$. Another simple fact is that $A_r \cap A_s = A_{\text{lcm}(r,s)}$, where lcm stands for the least common multiple. Our probability equals

$$\begin{aligned} 1 - P(A_{12} \cup A_{15}) &= 1 - P(A_{12}) - P(A_{15}) + P(A_{12} \cap A_{15}) \\ &= 1 - P(A_{12}) - P(A_{15}) + P(A_{60}) \\ &= 1 - \frac{83}{1000} - \frac{66}{1000} + \frac{16}{1000} = 0.867. \end{aligned}$$

Example 3.5. Sit 3 men and 4 women at random in a row. What is the probability that either all the men or all the women end up sitting together?

Here, $A_1 = \{\text{all women sit together}\}$, $A_2 = \{\text{all men sit together}\}$, $A_1 \cap A_2 = \{\text{both women and men sit together}\}$, and so the answer is

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = \frac{4! \cdot 4!}{7!} + \frac{5! \cdot 3!}{7!} - \frac{2! \cdot 3! \cdot 4!}{7!}.$$

Example 3.6. A group of 3 Norwegians, 4 Swedes, and 5 Finns is seated at random around a table. Compute the probability that at least one of the three groups ends up sitting together.

Define $A_N = \{\text{Norwegians sit together}\}$ and similarly A_S, A_F. We have

$$\begin{aligned} &P(A_N) = \frac{3! \cdot 9!}{11!},\ P(A_S) = \frac{4! \cdot 8!}{11!},\ P(A_N) = \frac{5! \cdot 7!}{11!}, \\ &P(A_N \cap A_S) = \frac{3! \cdot 4! \cdot 6!}{11!},\ P(A_N \cap A_F) = \frac{3! \cdot 5! \cdot 5!}{11!},\ P(A_S \cap A_F) = \frac{4! \cdot 5! \cdot 4!}{11!}, \\ &P(A_N \cap A_S \cap A_F) = \frac{3! \cdot 4! \cdot 5! \cdot 2!}{11!}. \end{aligned}$$

Therefore,

$$P(A_N \cup A_S \cup A_F) = \frac{3! \cdot 9! + 4! \cdot 8! + 5! \cdot 7! - 3! \cdot 4! \cdot 6! - 3! \cdot 5! \cdot 5! - 4! \cdot 5! \cdot 4! + 3! \cdot 4! \cdot 5! \cdot 2!}{11!}.$$

Example 3.7. *Matching problem.* A large company with n employees has a scheme according to which each employee buys a Christmas gift and the gifts are then distributed at random to the employees. What is the probability that someone gets his or her own gift?

Note that this is different from asking, assuming that you are one of the employees, for the probability that *you* get your own gift, which is $\frac{1}{n}$.

Let $A_i = \{i\text{th employee gets his or her own gift}\}$. Then, what we are looking for is

$$P(\bigcup_{i=1}^{n} A_i).$$

We have

$$\begin{aligned} P(A_i) &= \frac{1}{n} \quad \text{(for all } i\text{)}, \\ P(A_i \cap A_j) &= \frac{(n-2)!}{n!} = \frac{1}{n(n-1)} \quad \text{(for all } i < j\text{)}, \\ P(A_i \cap A_j \cap A_k) &= \frac{(n-3)!}{n!} = \frac{1}{n(n-1)(n-2)} \quad \text{(for all } i < j < k\text{)}, \\ &\dots \\ P(A_1 \cap \dots \cap A_n) &= \frac{1}{n!}. \end{aligned}$$

Therefore, our answer is

$$\begin{aligned} & n \cdot \frac{1}{n} - \binom{n}{2} \cdot \frac{1}{n(n-1)} + \binom{n}{3} \cdot \frac{1}{n(n-1)(n-2)} - \ldots + (-1)^{n-1}\frac{1}{n!} \\ = \quad & 1 - \frac{1}{2!} + \frac{1}{3!} + \ldots + (-1)^{n-1}\frac{1}{n!} \\ \to \quad & 1 - \frac{1}{e} \approx 0.6321 \quad \text{(as } n \to \infty\text{)}. \end{aligned}$$

Example 3.8. *Birthday Problem.* Assume that there are k people in the room. What is the probability that there are two who share a birthday? We will ignore leap years, assume all birthdays are equally likely, and generalize the problem a little: from n possible birthdays, sample k times with replacement.

$$P(\text{a shared birthday}) = 1 - P(\text{no shared birthdays}) = 1 - \frac{n \cdot (n-1) \cdots (n-k+1)}{n^k}.$$

When $n = 365$, the lowest k for which the above exceeds 0.5 is, famously, $k = 23$. Some values are given in the following table.

k	prob. for $n = 365$
10	0.1169
23	0.5073
41	0.9032
57	0.9901
70	0.9992

Occurences of this problem are quite common in various contexts, so we give another example. Each day, the Massachusetts lottery chooses a four digit number at random, with leading 0's allowed. Thus, this is sampling with replacement from among $n = 10,000$ choices each day. On February 6, 1978, the *Boston Evening Globe* reported that

> "During [the lottery's] 22 months' existence [...], no winning number has ever been repeated. [David] Hughes, the expert [and a lottery official] doesn't expect to see duplicate winners until about half of the 10,000 possibilities have been exhausted."

What would an informed reader make of this? Assuming $k = 660$ days, the probability of no repetition works out to be about $2.19 \cdot 10^{-10}$, making it a remarkably improbable event. What happened was that Mr. Hughes, not understanding the Birthday Problem, did not check for repetitions, confident that there would not be any. Apologetic lottery officials announced later that there indeed were repetitions.

Example 3.9. *Coupon Collector Problem.* Within the context of the previous problem, assume that $k \geq n$ and compute P(all n birthdays are represented).

More often, this is described in terms of cereal boxes, each of which contains one of n different cards (coupons), chosen at random. If you buy k boxes, what is the probability that you have a complete collection?

When $k = n$, our probability is $\frac{n!}{n^n}$. More generally, let

$$A_i = \{i\text{th birthday is missing}\}.$$

Then, we need to compute

$$1 - P(\bigcup_{i=1}^{n} A_i).$$

Now,

$$\begin{aligned}
P(A_i) &= \frac{(n-1)^k}{n^k} \quad \text{(for all } i\text{)} \\
P(A_i \cap A_j) &= \frac{(n-2)^k}{n^k} \quad \text{(for all } i < j\text{)} \\
&\dots \\
P(A_1 \cap \dots \cap A_n) &= 0
\end{aligned}$$

and our answer is

$$\begin{aligned}
&1 - n\left(\frac{n-1}{n}\right)^k + \binom{n}{2}\left(\frac{n-2}{n}\right)^k - \ldots + (-1)^{n-1}\binom{n}{n-1}\left(\frac{1}{n}\right)^k \\
&= \sum_{i=0}^{n} \binom{n}{i}(-1)^i \left(1 - \frac{i}{n}\right)^k.
\end{aligned}$$

This must be $\frac{n!}{n^n}$ for $k = n$, and 0 when $k < n$, neither of which is obvious from the formula. Neither will you, for large n, get anything close to the correct numbers when $k \leq n$ if you try to compute the probabilities by computer, due to the very large size of summands with alternating signs and the resulting rounding errors. We will return to this problem later for a much more efficient computational method, but some numbers are in the two tables below. Another remark for those who know a lot of combinatorics: you will perhaps notice that the above probability is $\frac{n!}{n^k} S_{k,n}$, where $S_{k,n}$ is the Stirling number of the second kind.

k	prob. for $n = 6$
13	0.5139
23	0.9108
36	0.9915

k	prob. for $n = 365$
1607	0.0101
1854	0.1003
2287	0.5004
2972	0.9002
3828	0.9900
4669	0.9990

More examples with combinatorial flavor

We will now do more problems which would rather belong to the previous chapter, but are a little harder, so we do them here instead.

Example 3.10. Roll a die 12 times. Compute the probability that a number occurs 6 times and two other numbers occur three times each.

The number of outcomes is 6^{12}. To count the number of good outcomes:

1. Pick the number that occurs 6 times: $\binom{6}{1} = 6$ choices.
2. Pick the two numbers that occur 3 times each: $\binom{5}{2}$ choices.
3. Pick slots (rolls) for the number that occurs 6 times: $\binom{12}{6}$ choices.
4. Pick slots for one of the numbers that occur 3 times: $\binom{6}{3}$ choices.

Therefore, our probability is $\frac{6\binom{5}{2}\binom{12}{6}\binom{6}{3}}{6^{12}}$.

Example 3.11. You have 10 pairs of socks in the closet. Pick 8 socks at random. For every i, compute the probability that you get i complete pairs of socks.

There are $\binom{20}{8}$ outcomes. To count the number of good outcomes:

1. Pick i pairs of socks from the 10: $\binom{10}{i}$ choices.
2. Pick pairs which are represented by a single sock: $\binom{10-i}{8-2i}$ choices.
3. Pick a sock from each of the latter pairs: 2^{8-2i} choices.

Therefore, our probability is $\frac{2^{8-2i}\binom{10-i}{8-2i}\binom{10}{i}}{\binom{20}{8}}$.

Example 3.12. *Poker Hands*. In the definitions, the word *value* refers to A, K, Q, J, 10, 9, 8, 7, 6, 5, 4, 3, 2. This sequence orders the cards in descending consecutive values, with one exception: an Ace may be regarded as 1 for the purposes of making a straight (but note that, for example, K, A, 1, 2, 3 is *not* a valid straight sequence — A can only begin or end a straight). From the lowest to the highest, here are the hands:
(a) *one pair*: two cards of the same value plus 3 cards with different values

$$J\spadesuit\ J\clubsuit\ 9\heartsuit\ Q\clubsuit\ 4\spadesuit$$

(b) *two pairs*: two pairs plus another card of different value

$$J\spadesuit\ J\clubsuit\ 9\heartsuit\ 9\clubsuit\ 3\spadesuit$$

(c) *three of a kind*: three cards of the same value plus two with different values

$$Q\spadesuit\ Q\clubsuit\ Q\heartsuit\ 9\clubsuit\ 3\spadesuit$$

(d) *straight*: five cards with consecutive values

$$5\heartsuit\ 4\clubsuit\ 3\clubsuit\ 2\heartsuit\ A\spadesuit$$

(e) *flush*: five cards of the same suit

$$K\clubsuit\ 9\clubsuit\ 7\clubsuit\ 6\clubsuit\ 3\clubsuit$$

(f) *full house*: a three of a kind and a pair

$$J\clubsuit\ J\diamondsuit\ J\heartsuit\ 3\clubsuit\ 3\spadesuit$$

(g) *four of a kind*: four cards of the same value

$$K\clubsuit\ K\diamondsuit\ K\heartsuit\ K\clubsuit\ 10\spadesuit$$

(e) *straight flush*: five cards of the same suit with consecutive values

$$A\spadesuit\ K\spadesuit\ Q\spadesuit\ J\spadesuit\ 10\spadesuit$$

Here are the probabilities:

hand	no. combinations	approx. prob.
one pair	$13 \cdot \binom{12}{3} \cdot \binom{4}{2} \cdot 4^3$	0.422569
two pairs	$\binom{13}{2} \cdot 11 \cdot \binom{4}{2} \cdot \binom{4}{2} \cdot 4$	0.047539
three of a kind	$13 \cdot \binom{12}{2} \cdot \binom{4}{3} \cdot 4^2$	0.021128
straight	$10 \cdot 4^5$	0.003940
flush	$4 \cdot \binom{13}{5}$	0.001981
full house	$13 \cdot 12 \cdot \binom{4}{3} \cdot \binom{4}{2}$	0.001441
four of a kind	$13 \cdot 12 \cdot 4$	0.000240
straight flush	$10 \cdot 4$	0.000015

Note that the probabilities of a straight and a flush above include the possibility of a straight flush.

Let us see how some of these are computed. First, the number of all outcomes is $\binom{52}{5} = 2,598,960$. Then, for example, for the *three of a kind*, the number of good outcomes may be obtained by listing the number of choices:

1. Choose a value for the triple: 13.
2. Choose the values of other two cards: $\binom{12}{2}$.
3. Pick three cards from the four of the same chosen value: $\binom{4}{3}$.
4. Pick a card (i.e., the suit) from each of the two remaining values: 4^2.

One could do the same for *one pair*:

1. Pick a number for the pair: 13.
2. Pick the other three numbers: $\binom{12}{3}$
3. Pick two cards from the value of the pair: $\binom{4}{2}$.
4. Pick the suits of the other three values: 4^3

And for the *flush*:

1. Pick a suit: 4.
2. Pick five numbers: $\binom{13}{5}$

Our final worked out case is *straight flush*:

1. Pick a suit: 4.
2. Pick the beginning number: 10.

We end this example by computing the probability of not getting any hand listed above, that is,

$$\begin{aligned}
P(\text{nothing}) &= P(\text{all cards with different values}) - P(\text{straight or flush}) \\
&= \frac{\binom{13}{5} \cdot 4^5}{\binom{52}{5}} - (P(\text{straight}) + P(\text{flush}) - P(\text{straight flush})) \\
&= \frac{\binom{13}{5} \cdot 4^5 - 10 \cdot 4^5 - 4 \cdot \binom{13}{5} + 40}{\binom{52}{5}} \\
&\approx 0.5012.
\end{aligned}$$

Example 3.13. Assume that 20 Scandinavians, 10 Finns, and 10 Danes, are to be distributed at random into 10 rooms, 2 per room. What is the probability that exactly $2i$ rooms are mixed, $i = 0, \dots 5$?

This is an example when careful thinking about what the outcomes should be really pays off. Consider the following model for distributing the Scandinavians into rooms. First arrange them at random into a row of 20 slots $S1, S2, \dots, S20$. Assume that room 1 takes people in slots $S1, S2$, so let us call these two slots $R1$. Similarly, room 2 takes people in slots $S3, S4$, so let us call these two slots $R2$, etc.

Now, it is clear that we only need to keep track of the distribution of 10 D's into the 20 slots, corresponding to the positions of the 10 Danes. Any such distribution constitutes an outcome and they are equally likely. Their number is $\binom{20}{10}$.

To get $2i$ (for example, 4) mixed rooms, start by choosing $2i$ (ex., 4) out of the 10 rooms which are going to be mixed; there are $\binom{10}{2i}$ choices. You also need to decide into which slot in each of the $2i$ chosen mixed rooms the D goes, for 2^{2i} choices.

Once these two choices are made, you still have $10 - 2i$ (ex., 6) D's to distribute into $5 - i$ (ex., 3) rooms, as there are two D's in each of these rooms. For this, you need to choose $5 - i$ (ex., 3) rooms from the remaining $10 - 2i$ (ex., 6), for $\binom{10-2i}{5-i}$ choices, and this choice fixes a good outcome.

The final answer, therefore, is

$$\frac{\binom{10}{2i} 2^{2i} \binom{10-2i}{5-i}}{\binom{20}{10}}.$$

Problems

1. Roll a single die 10 times. Compute the following probabilities: (a) that you get at least one 6; (b) that you get at least one 6 *and* at least one 5; (c) that you get three 1's, two 2's, and five 3's.

2. Three married couples take seats around a table at random. Compute P(no wife sits next to her husband).

3. A group of 20 Scandinavians consists of 7 Swedes, 3 Finns, and 10 Norwegians. A committee of five people is chosen at random from this group. What is the probability that at least one of the three nations is *not* represented on the committee?

4. Choose each digit of a 5 digit number at random from digits $1, \dots 9$. Compute the probability that no digit appears more than twice.

5. Roll a fair die 10 times. (a) Compute the probability that at least one number occurs exactly 6 times. (b) Compute the probability that at least one number occurs exactly once.