2 Axioms of Probability

2.1 Sample Spaces

In this chapter we more formally define the concept of probability.

Definition 2.1. The *sample space* S of an experiment is the set of possible outcomes. An *event* E is any *subset* of S; it *occurs* if the result of the experiment is any of the outcomes in E.

Examples 2.2. 1. When drawing a card from a standard 52 card pack, the sample space S is the set of cards. Let E be the set of hearts: this event occurs if we deal any heart from the pack.

2. Roll two standard dice and sum the total. The sample space has *eleven* elements;

$$S = \{2, 3, \dots, 12\}$$

The event $E = \{2, 3, 4, 5, 6\}$ occurs if we roll a total of 6 or less.

3. If the experiment involves rolling two standard dice and recording the *unordered* pair of values, then the sample space has $6 + 5 + 4 + 3 + 2 + 1 = 21$ elements;

$$S = \big\{\{x, y\} : 1 \le x \le y \le 6\big\}$$

If we care about order, then the sample space has 36 elements. The event "roll a double" is the subset $E = \big\{\{x, x\} : 1 \le x \le 6\big\}$.

4. Sample spaces can be infinite sets: the sample space of a human lifetime (in years) is (potentially!) the interval $\mathbb{R}_0^+ = [0, \infty)$. The event $E = [100, \infty)$ is 'reach your 100^{th} birthday."

5. The sample space for the experiment of throwing a dart at a dartboard might be represented as

$$S = \{(x, y) : x^2 + y^2 < 6.69^2\} \cup \{f\}$$

where (x, y) is the co-ordinate of a point on the board (disk radius 6.69 in), and f represents failing to hit the board. The inner bullseye has radius 0.5 in, whence the event "score 50 with a single dart" is the subset $E = \{(x, y) : x^2 + y^2 < 0.25\}$.

Since events are sets, they can be manipulated in the usual ways.

Definition 2.3. Suppose S is a sample space and E, F are events.

1. $E \subseteq F$ means that every outcome in E is also in F: equivalently, if E occurs then F occurs.

2. The *complement* of E is the event $E^{\mathsf{C}} = \{x \in S : x \notin E\}$, that is "$E$ doesn't occur."

3. The *union* $E \cup F = \{x \in S : x \in E \text{ or } x \in F\}$ is the event "E occurs or F occurs (or both)."

4. The *intersection* $E \cap F = \{x \in S : x \in E \text{ and } x \in F\}$ is the event "E and F both occur." It is permissible to write EF instead.

   We say that E, F are *mutually exclusive* if $E \cap F = \emptyset$.

These set theoretic constructions can be represented by Venn diagrams. The basic rules of set algebra also apply.

Lemma 2.4. *Suppose E, F, G are events (subsets of some sample space S). Then the have the basic set laws: observe the symmetry in these.*

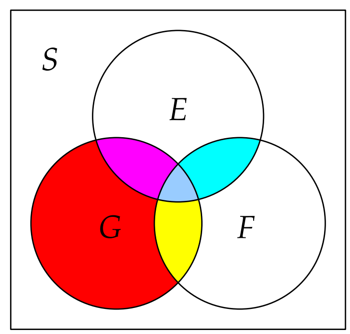
1. *Associativity:* $(E \cup F) \cup G = E \cup (F \cup G)$ *and* $(E \cap F) \cap G = E \cap (F \cap G)$
2. *Commutativity:* $E \cup F = F \cup E$ *and* $E \cap F = F \cap E$
3. *Distributivity:*[a] $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$ *and* $(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$
4. *De Morgan's Laws:* $(E \cap F)^{\mathsf{C}} = E^{\mathsf{C}} \cup F^{\mathsf{C}}$ *and* $(E \cup F)^{\mathsf{C}} = E^{\mathsf{C}} \cap F^{\mathsf{C}}$*. These extend to the general statements* $(\bigcup E_i)^{\mathsf{C}} = \bigcap E_i^{\mathsf{C}}$ *and* $(\bigcap E_i)^{\mathsf{C}} = \bigcup E_i^{\mathsf{C}}$ *even if there are infinitely many events* E_i*.*

[a] To remember these, pretend that $\cup$ is addition and $\cap$ multiplication, then it looks like $(E+F)G = EG + FG$. For the second law, switch the roles of addition and multiplication!

Sketch Proof. These are generally proved by offloading the propositions to formal statements involving *and, or, not*. For example:

$$x \in E \cup F \iff x \in E \text{ or } x \in F \iff x \in F \text{ or } x \in E \iff x \in F \cup E$$

For the first three propositions (and De Morgan when there are two events) it is useful instead to reason via a Venn diagram. For instance the shaded region in the the picture represents the distributive law



$$(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$$

Just think through what it means for an outcome x to lie in each side of this equality.

For a general proof of De Morgan:

$$\begin{aligned} x \in \left(\bigcup E_i\right)^{\mathsf{C}} &\iff x \notin \bigcup E_i \iff x \notin E_i \text{ for every listed event } E_i \\ &\iff x \in E_i^{\mathsf{C}} \text{ for every listed event } E_i \\ &\iff x \in \bigcap E_i^{\mathsf{C}} \end{aligned}$$

■

This first section is little more than the basic rules of set algebra interpreted in the language of *events*. For instance, it is more important to be able to recognize De Morgan's laws in English; the first being the equivalence of two sentences;

$$E \text{ and } F \text{ do not occur simultaneously} \iff \text{at least one of } E \text{ or } F \text{ does not occur.}$$

2.2 Axioms and Identities

Now that we understand a sample spaces, we want to assign measures of likelihood to various events. You've thought about many examples of this in your mathematical career:

Example 2.5. A roll of a fair die has equal likelihood of producing each of the six outcomes. We'd say, for instance, that the probability of rolling a multiple of three is $\frac{2}{6} = \frac{1}{3}$. This is merely counting: the number of ways to succeed (2: roll a 3 or 6), divided by the number of possible outcomes (6).

Since the probability is measuring the chance of an event occurring, we represented it as a fraction: in general the probability of an event will be a number $\mathbb{P}(E)$ between 0 and 1. Our example is straightforward for two reasons: there are *finitely many* possible outcomes and each is *equally likely*. We plainly cannot assume these properties for a general experiment! Instead we work backwards: what properties do we want a probability function to satisfy?

Definition 2.6. Suppose S is a sample space. A *probability measure*[1]is a function $\mathbb{P}$ which assigns a value $\mathbb{P}(E)$ to any event $E \subseteq S$ and additionally satisfies:

1. For every event $0 \le \mathbb{P}(E) \le 1$.
2. $\mathbb{P}(S) = 1$
3. If $E_1, E_2, \ldots$ are a *finite or countable* list of mutually exclusive events ($i \neq j \implies E_i \cap E_j = \emptyset$), then

$$\mathbb{P}\left(\bigcup E_i\right) = \sum \mathbb{P}(E_i)$$

We revisit the example in this language. The sample space for the experiment "roll a fair die" is

$$S = \{1, 2, 3, 4, 5, 6\}$$

This set has $2^6 = 64$ distinct subsets, each of which is an event. For instance $E = \{1, 3, 5\}$ is the event "roll an odd number." Since each event of the form "roll j" is equally likely, it seems reasonable to define the probability function

$$\mathbb{P}(E) = \frac{\text{number of elements in } E}{6} = \frac{1}{6}|E|$$

We easily verify our axioms:

1. Since $0 \le |E| \le 6$ we see that $0 \le \mathbb{P}(E) \le 6$. The zero corresponds to the event $E = \emptyset$; this is a legitimate event, even if it can never occur!
2. $|S| = 6 \implies \mathbb{P}(S) = \frac{6}{6} = 1$.
3. If $E_i \cap E_j = \emptyset$, then $\left|E_i \cup E_j\right| = |E_i| + \left|E_j\right|$, from which $\mathbb{P}(E_i \cup E_j) = \mathbb{P}(E_i) + \mathbb{P}(E_j)$. Since there are only finitely many possible events, verifying this for two mutually exclusive events is enough.

[1] It is not in fact necessary for $\mathbb{P}(E)$ to be defined for all subsets $E \subseteq S$. For instance, in Example 2.7 it is possible to consider subsets of the dartboard which do not have a measurable area! Such concerns are beyond the level of this course, though you should at least be aware of the imprecision of our definition. A more advanced study of probability requires a thorough grounding in analysis and typically starts by building the axioms within *measure theory*.

This is plainly a very simple situation. For something trickier, revisit our dartboard example.

Example 2.7. The experiment is to throw a single dart at the dartboard. We could define a suitable probability function as follows. For a given region of the dartboard, let E_R be the event "the dart lands inside R" and let

$$\mathbb{P}(E_R) = \frac{1}{2A}\,\mathrm{Area}(R)$$

where $A = \frac{1}{4}\pi 6.69^2\,\mathrm{in}^2$ is the area of the dartboard. In order to satisfy the axioms, the event "miss the board" would have probability $\frac{1}{2}$.

Lemma 2.8. *Suppose that $\mathbb{P}$ is a probability measure on S and E, F are events.*

1. $\mathbb{P}(\emptyset) = 0$
2. $\mathbb{P}(E^{\mathsf{C}}) = 1 - \mathbb{P}(E)$.
3. $E \subseteq F \implies \mathbb{P}(E) \le \mathbb{P}(F)$

Proof. 1. Plainly $S, \emptyset$ are mutually exclusive ($S \cap \emptyset = \emptyset$). By axiom 3,

$$\mathbb{P}(S) = \mathbb{P}(S \cup \emptyset) = \mathbb{P}(S) + \mathbb{P}(\emptyset) \implies \mathbb{P}(\emptyset) = 0$$

2. $E \cup E^{\mathsf{C}} = S$ and $E \cup E^{\mathsf{C}} = \emptyset$. Now use axioms 2 and 3 to see that $\mathbb{P}(E) + \mathbb{P}(E^{\mathsf{C}}) = 1$.

3. Write $F = E \cup (F \cap E^{\mathsf{C}})$. Since these two sets are mutually exclusive, axioms 1 and 3 tell us that

$$\mathbb{P}(F) = \mathbb{P}(E) + \mathbb{P}(F \cap E^{\mathsf{C}}) \ge \mathbb{P}(E)$$

■

Impossible and Certain Events? If light of the facts that $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(S) = 1$, it is reasonable to ask if the converses also hold. That is:

$$\mathbb{P}(E) = 0 \overset{???}{\implies} E = \emptyset, \qquad \mathbb{P}(E) = 1 \overset{???}{\implies} E = S$$

If there are finitely many distinct events and each singleton event $E = \{x\}$ has positive probability, then the above claims are indeed true. In such cases it common to say that an event with probability 1 is *certain* and one with probability 0 is *impossible*. A great many of our examples will be of this form. For instance,

> It is impossible to roll a three and a four simultaneously on a fair die.

When there are *infinitely many* possible events, things get stranger. Consider the example of the dartboard and the event "hit the point at the center of the board." Since a point has *zero area*, the probability of this event is zero, though it is clearly *possible*! It is more precise to say that a probability zero event happens *almost never* (a.n.) and a probability one event *almost always* (a.a.); thus

> A dart almost never hits the exact center of a dartboard.

The Inclusion–Exclusion Principle Recall Axiom 3. What should we do if we want to consider a union of *non-mutually exclusive* events.

Example 2.9. A fair die is rolled. Consider the events

- $E_1 = \{2,4,6\}$: roll and even number.
- $E_2 = \{3,6\}$: roll a number divisible by three.

Then $E_1 \cup E_2 = \{2,3,4,6\}$ from which

$$\mathbb{P}(E_1 \cup E_2) = \frac{4}{6} = \frac{2}{3} \neq \frac{1}{2} + \frac{1}{3} = \mathbb{P}(E_1) + \mathbb{P}(E_2)$$

The problem is that we have double-counted the event $E_1 \cap E_2 = \{6\}$ of rolling a six. We may therefore *subtract* the probability of this from the right side to balance the equation:

$$\mathbb{P}(E_1 \cup E_2) = \frac{2}{3} = \frac{1}{2} + \frac{1}{3} - \frac{1}{6} = \mathbb{P}(E_1) + \mathbb{P}(E_2) - \mathbb{P}(E_1 \cap E_2)$$

The is the simplest version of a general principle.

Theorem 2.10. *Suppose $E_1, E_2, \ldots, E_n$ are a finite list of events. Then*

$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} \mathbb{P}(E_i) - \sum_{i<j} \mathbb{P}(E_i \cap E_j) + \sum_{i<j<k} \mathbb{P}(E_i \cap E_j \cap E_k) - \cdots + (-1)^{n-1}\mathbb{P}(E_1 \cap \cdots \cap E_n)$$

Proof. We prove by induction. The base case ($n = 2$) is nothing but a Venn diagram generalizing the previous example:

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2) - \mathbb{P}(E_1 \cap E_2)$$

Now assume the principle holds for *any* list of n events and consider a list $E_1, \ldots, E_{n+1}$. Then

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^{n+1} E_i\right) &= \mathbb{P}\left(\bigcup_{i=1}^{n} E_i \cup E_{n+1}\right) = \mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) + \mathbb{P}(E_{n+1}) - \mathbb{P}\left(\left(\bigcup_{i=1}^{n} E_i\right) \cap E_{n+1}\right) \\ &= \mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) + \mathbb{P}(E_{n+1}) - \mathbb{P}\left(\bigcup_{i=1}^{n} (E_i \cap E_{n+1})\right)\end{aligned}$$

Read this carefully: expanding the first term using the induction hypothesis consists of all combinations of events in the inclusion-exclusion principle which *do not contain* E_{n+1}; expanding the remaining terms gives all combinations which *contain* E_{n+1}. ∎

Example 2.11. 100 math majors are each taking *at least one* of the classes: Probability, Analysis, Geometry. Multiple enrollment data is as follows;

Probability: 60 Analysis: 75 Geometry: 30

Probability & Analysis: 40 Probability & Geometry: 15 Analysis & Geometry: 20

1. How many students are taking all three classes?

   Label the events P, A, G. Since every student is in at least one class, then

$$1 = \mathbb{P}(P \cup A \cup G) = \frac{1}{100}(60 + 75 + 30 - 40 - 15 - 20) + \mathbb{P}(P \cap A \cap G) = \frac{90}{100} + \mathbb{P}(P \cap A \cap G)$$

   Thus 10 students are taking all three classes.

2. How many students are taking *only* geometry?

   Attacking this directly ($\mathbb{P}(P^C \cap A^C \cap G)$?) is messy. For an easier approach, start by observing that

$$\mathbb{P}(P \cup A) = \mathbb{P}(P) + \mathbb{P}(A) - \mathbb{P}(P \cap A) = \frac{95}{100}$$

   Since every student is taking at least one course, the remaining 5 students must be taking only geometry.

The inclusion–exclusion principle functions almost like a convergent alternating series, approximating the value $\mathbb{P}(E_1 \cup \cdots \cup E_n)$ alternately from above and below.

Corollary 2.12 (Bonferroni Inequalities). *If we truncate Theorem 2.10 after a sequence of additive terms, we have an upper bound; after a sequence of subtractions, a lower bound. In particular*

1. $\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(E_i) = \mathbb{P}(E_1) + \cdots + \mathbb{P}(E_n)$
2. $\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) \geq \sum_{i=1}^{n} \mathbb{P}(E_i) - \sum_{i<j} \mathbb{P}(E_i \cap E_j)$
3. $\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(E_i) - \sum_{i<j} \mathbb{P}(E_i \cap E_j) + \sum_{i<j<k} \mathbb{P}(E_i \cap E_j \cap E_k)$

Returning to the previous example, observe that

$$\mathbb{P}(P) + \mathbb{P}(A) + \mathbb{P}(G) - \mathbb{P}(P \cap A) - \mathbb{P}(P \cap G) - \mathbb{P}(A \cap G) = \frac{90}{100} \leq 1$$

Suppose we naively modified the problem so that only 5 students were taking both Analysis & Geometry. Then the second Bonferroni inequality fails

$$\begin{aligned}\mathbb{P}(P) + \mathbb{P}(A) + \mathbb{P}(G) - \mathbb{P}(P \cap A) - \mathbb{P}(P \cap G) - \mathbb{P}(A \cap G) &= \frac{60 + 75 + 30 - 40 - 15 - 5}{100} \\ &= \frac{105}{100} > 1 = \mathbb{P}(P \cap A \cap G)\end{aligned}$$

This indicates that the data is contradictory; there would be no solution to such a problem.

2.3 Counting Revisited: Equally Likely Outcomes

We will spend much of the course considering situations when outcomes have different probabilities. As we saw in Example 2.5, when a sample space S is *finite* and all outcomes are equally likely, computing probabilities is nothing more than a counting exercise.

Theorem 2.13. *Suppose $S = \{1, \ldots, n\}$ and that all outcomes have equal probability. Then, for any event E,*

$$\mathbb{P}(E) = \frac{1}{n} |E|$$

This *should* be obvious from Axiom 2: if p is the probability of any single outcome in S, then

$$1 = \mathbb{P}(S) = \sum_{k=1}^{n} \mathbb{P}(k) = np \implies p = \frac{1}{n}$$

The majority of this section consists of examples of this simple situation.

Examples 2.14.

1. Two cards are dealt from a standard pack. What is the probability that both are hearts?

   Assuming that the pack is well-shuffled so that any combination of cards is equally likely, the sample space has size $|S| = 52 \cdot 51$, while the event "deal two hearts" has size $|E| = 13 \cdot 12$. The probability is therefore

   $$\mathbb{P}(E) = \frac{13 \cdot 12}{52 \cdot 51} = \frac{1}{17}$$

   Our original approach treated the sample space as consisting of the pair of cards dealt *in order.* We could instead solve the problem using combinations. There are

   - $\left|\widetilde{S}\right| = \binom{52}{2} = \frac{52 \cdot 51}{2} = 1326$ ways to deal a pair of cards from the pack.
   - $\left|\widetilde{E}\right| = \binom{13}{2} = \frac{13 \cdot 12}{2} = 78$ ways to deal a pair of hearts.

   Of course, we obtain the same probability as before; $\mathbb{P}(\widetilde{E}) = \frac{\binom{13}{2}}{\binom{52}{2}} = \frac{1}{17}$

2. A pair of fair dice, one with four sides the other with six, are rolled and the total summed. Find the probability that the sum equals 7.

   The sample space is

   $$S = \{(x, y) : 1 \le x \le 4,\ 1 \le y \le 6\}$$

   where all pairs (x, y) are equally likely to occur. The event "sum equals 7" is the set

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

   from which $\mathbb{P}(E) = \frac{|E|}{|S|} = \frac{4}{4 \cdot 6} = \frac{1}{6}$.

3. A committee of six is to be formed from a group of ten women and eight men. What is the probability that the committee is equally balanced by sex?

   It is implicit in the question that all committees are equally likely. Plainly there are $|S| = \binom{18}{6}$ possible committees. The event E of having three people of each sex on the committee requires us to count how many ways to choose three women from ten and three men from eight: thus

$$\begin{aligned}\mathbb{P}(E) &= \frac{|E|}{|S|} = \frac{\binom{10}{3}\binom{8}{3}}{\binom{18}{6}} = \frac{10!8!6!12!}{3!7!3!5!18!} = \frac{10\cdot 9\cdot 8}{3\cdot 2}\cdot\frac{8\cdot 7\cdot 6}{3\cdot 2}\cdot\frac{6\cdot 5\cdot 4\cdot 3\cdot 2}{18\cdot 17\cdot 16\cdot 15\cdot 14\cdot 13}\\ &= \frac{10\cdot 8}{17\cdot 13} = \frac{80}{221}\end{aligned}$$

The last example is easily generalized, following our earlier counting discussions.

Lemma 2.15. *Suppose $S = A_1 \cup \cdots \cup A_r$ is a union of disjoint subsets of sizes $n_1, n_2, \ldots, n_r$ respectively. Suppose a subset $B \subseteq S$ of size k is chosen at random. Let E be the event, "B contains precisely k_j elements from A_j, for each j." Then,*

$$\mathbb{P}(E) = \frac{\binom{n_1}{k_1}\cdots\binom{n_r}{k_r}}{\binom{n}{k}}$$

Example 2.16. A lottery game has the following rules: balls numbered $1, \ldots, 20$ are placed in an urn. You choose *four* numbers, two odd and two even. At least one each of your even and odd numbers must be chosen for you to win. If six balls are chosen, with all balls equally likely, what is your probability of success?

This is not quite the lemma, since there are a few distinct cases to consider. First note that there are $\binom{20}{6}$ ways to choose the six balls. We have

- $\binom{2}{1}\binom{2}{1}\binom{16}{4} = 4\binom{16}{4} = 7280$ ways to select *one each* of the odd and even winning balls.
- $\binom{2}{2}\binom{2}{1}\binom{16}{3} = 2\binom{16}{3} = 1120$ ways to select *both odd* and *one even* winning balls; the same for *one odd* and *both even* balls.
- $\binom{2}{2}\binom{2}{2}\binom{16}{2} = \binom{16}{2} = 120$ ways to select *all four* winning balls.

Sum these possibilities and divide by $\binom{20}{6} = 38760$ for the answer

$$\mathbb{P}(\text{win}) = \frac{7280 + 1120 + 1120 + 120}{38760} = \frac{241}{969} \approx 24.9\%$$

Alternatively, you could count how many ways you could *lose*:

$$\mathbb{P}(\text{win}) = 1 - \mathbb{P}(\text{lose}) = 1 - \frac{\binom{2}{0}\binom{2}{0}\binom{16}{6} + \binom{2}{0}\binom{2}{1}\binom{16}{5} + \binom{2}{1}\binom{2}{0}\binom{16}{5} + \binom{2}{0}\binom{2}{2}\binom{16}{4} + \binom{2}{2}\binom{2}{0}\binom{16}{4}}{\binom{20}{6}}$$

Problems like these form the backbone of elementary probability. For much of the rest of the course we will consider special situations such as *conditional probabilities*, where the chance of an event depends on what has come before. Many of the problems we've seen can indeed be thought of in such a manner, though it is useful and often simpler to try to count your way to an answer by viewing a desired event as a list of equally possible outcomes of a given experiment.

Example 2.17. We revisit the motivational problem of points from the introduction. Recall that we solved this using a tree and that the various outcomes of the game were *not* equally likely. With a bit of sneakery however, we can view the problem differently.

Suppose player A needs a more points to win, and player B needs b more points. The maximum number of games that could be played is $a+b-1$. We may *pretend* that all these games are played: of the 2^{a+b-1} ordered outcomes, player A wins if he achieves at least a points. Since all outcomes are now equally likely, the question is simply to count how many outcomes result in player A scoring at least a. Plainly the answer is

$$N_A := \sum_{k=a}^{a+b-1} \binom{a+b-1}{k}$$

The probability of player A winning is therefore $\frac{N_A}{2^{a+b-1}}$, and the stakes should be split in the ratio $2^{a+b-1} - N_A : N_A$.

In the precise context of the introductory example, we had player A ahead 3–2 in a first-to-five game. Thus $a=2$, $b=3$ and we have

$$N_A = \sum_{k=2}^{4} \binom{4}{k} = \binom{4}{2} + \binom{4}{3} + \binom{4}{4} = 6+4+1 = 11$$

precisely in accordance with our previous approach.

Three famous problems

Simple counting problems often have counterintuitive answers. We finish by considering three famous examples where your gut feeling is likely wrong.

The Next Card Problem A deck is shuffled and the cards dealt face up one at a time until the first ace appears. Is the next card more likely to be the ace of spades or the two of clubs?

The Birthday Problem Suppose n people are each equally likely to have any birthday form the 365 days of a non-leap year. How many people must be in the room before it is more likely than not that two share a birthday?

The Hat-Matching Problem n men throw their hats into the center of a room and the hats are distributed at random. What is the probability that none of the men get their own hat, and how does this depend on n?

Each problem has various interpretations; for instance the hat-matching problem could be rephrased to consider the outcomes of a game of musical chairs.

Have a think about these problems *before* looking at the answers on the next page. What does your gut say?

The Next Card Problem Your gut-feeling is likely that the two of clubs is more likely than the ace of spades, since the first ace might have been of spades. However...

A deck can be arranged in $|S| = 52!$ distinct ways. Remove *any card* C from the deck; the rest of the deck now has $51!$ possible orders.

Let E be the event "card C follows the first ace." For each order, there is *exactly one* place to put card C so that it follows the first ace. It follows that

$$\mathbb{P}(E) = \frac{51!}{52!} = \frac{1}{52}$$

is *independent* of the card C. The two events are therefore equally likely!

The Birthday Problem The naive approach is to round up $\frac{365}{2} = 182.5$ and assume the answer is $n = 183$. This is way off!

Instead observe that there are $|S| = 365^n$ possible arrangements of birthdays, where we include order. If these are distinct, then there are 365 choices for person one, 364 for person two, 363 for person three, etc. The result is

$$\mathbb{P}(n \text{ distinct birthdays}) = \frac{365 \cdot 364 \cdots (366-n)}{365^n} = \frac{n!}{365^n}\binom{365}{n}$$

As expected, the probabilities decrease as n increases;

$$\mathbb{P}(n+1) = \frac{366-n-1}{365}\mathbb{P}(n) < \mathbb{P}(n)$$

It doesn't take long to verify that $\mathbb{P}(22) \approx 52.4\%$, $\mathbb{P}(23) \approx 49.3\%$ so the answer is 23 people!

The Hat-Matching Problem As n grows, there are more chances for any given man to claim his own hat, so it is often assumed that the probability should increase to 1 as $n \to \infty$.

Break this down: let E_i ="man i claims his own hat." By simple counting, we see that

$$\mathbb{P}(E_i) = \frac{1}{n}, \quad \mathbb{P}(E_i \cap E_j) = \frac{1}{n(n-1)}, \quad \mathbb{P}(E_i \cap E_j \cap E_k) = \frac{1}{n(n-1)(n-2)}$$

since, after man i claims his hat there remain $n-1$ hats from which man j can claim his, etc. By the inclusion–exclusion principle,

$$\begin{aligned}
\mathbb{P}\left(\bigcup E_i\right) &= \sum_{i=1}^{n}\mathbb{P}(E_i) - \sum_{i<j}\mathbb{P}(E_i \cap E_j) - \cdots + (-1)^{n+1}\mathbb{P}(E_1 \cap \cdots \cap E_n) \\
&= \frac{n}{n} - \binom{n}{2}\frac{1}{n(n-1)} + \binom{n}{3}\frac{1}{n(n-1)(n-3)} - \cdots + (-1)^{n+1}\binom{n}{n}\frac{1}{n!} \\
&= \sum_{k=1}^{n}(-1)^{k+1}\binom{n}{k}\frac{(n-k)!}{n!} = \sum_{k=1}^{n}\frac{(-1)^{k+1}}{k!} = 1 - \sum_{k=0}^{n}\frac{(-1)^k}{k!}
\end{aligned}$$

The last sum is precisely the n^{th} Taylor polynomial for e^x evaluated at $x = -1$, whence

$$\mathbb{P}\left(\bigcap E_i^{\mathrm{C}}\right) = 1 - \mathbb{P}\left(\bigcup E_i\right) = \sum_{k=0}^{n}\frac{(-1)^k}{k!} \xrightarrow[n\to\infty]{} e^{-1} \approx 36.8\%$$