1.6 General definition of probability

Unit 1: Probability and Counting

Adapted from Blitzstein-Hwang Chapter 1.

We have now seen several methods for counting outcomes in a sample space, allowing us to calculate probabilities if the naive definition applies. But the naive definition can only take us so far, since it requires equally likely outcomes and can't handle an infinite sample space. We now give the general definition of probability. It requires just two axioms, but from these axioms it is possible to prove a vast array of results about probability.

Definition 1.6.1 (General definition of probability).

A *probability space* consists of a sample space S and a *probability function* P which takes an event $A \subseteq S$ as input and returns $P(A)$, a real number between 0 and 1, as output. The function P must satisfy the following axioms:

1. $P(\emptyset) = 0$, $P(S) = 1$.
2. If $A_1, A_2, \ldots$ are disjoint events, then

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

(Saying that these events are *disjoint* means that they are *mutually exclusive*: $A_i \cap A_j = \emptyset$ for $i \neq j$.)

In Pebble World, the definition says that probability behaves like mass: the mass of an empty pile of pebbles is 0, the total mass of all the pebbles is 1, and if we have non-overlapping piles of pebbles, we can get their combined mass by adding the masses of the individual piles. Unlike in the naive case, we can now have pebbles of differing masses, and we can also have a countably infinite number of pebbles as long as their total mass is 1.

Any function P (mapping events to numbers in the interval $[0, 1]$) that satisfies the two axioms is considered a valid probability function. However, the axioms don't tell us how probability should be *interpreted*; different schools of thought exist.

The *frequentist* view of probability is that it represents a long-run frequency over a large number of repetitions of an experiment: if we say a coin has probability $1/2$ of Heads, that means the coin would land Heads 50% of the time if we tossed it over and over and over.

The *Bayesian* view of probability is that it represents a degree of belief about the event in question, so we can assign probabilities to hypotheses like "candidate A will win the election" or "the defendant is guilty" even if it isn't possible to repeat the same election or the same crime over and over again.

The Bayesian and frequentist perspectives are complementary, and both will be helpful for developing intuition in later chapters. Regardless of how we choose to interpret probability, we can use the two axioms to derive other properties of probability, and these results will hold for *any* valid probability function.

Theorem 1.6.2 (Properties of probability).

Probability has the following properties, for any events A and B.

1. $P(A^c) = 1 - P(A)$.
2. If $A \subseteq B$, then $P(A) \leq P(B)$.
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

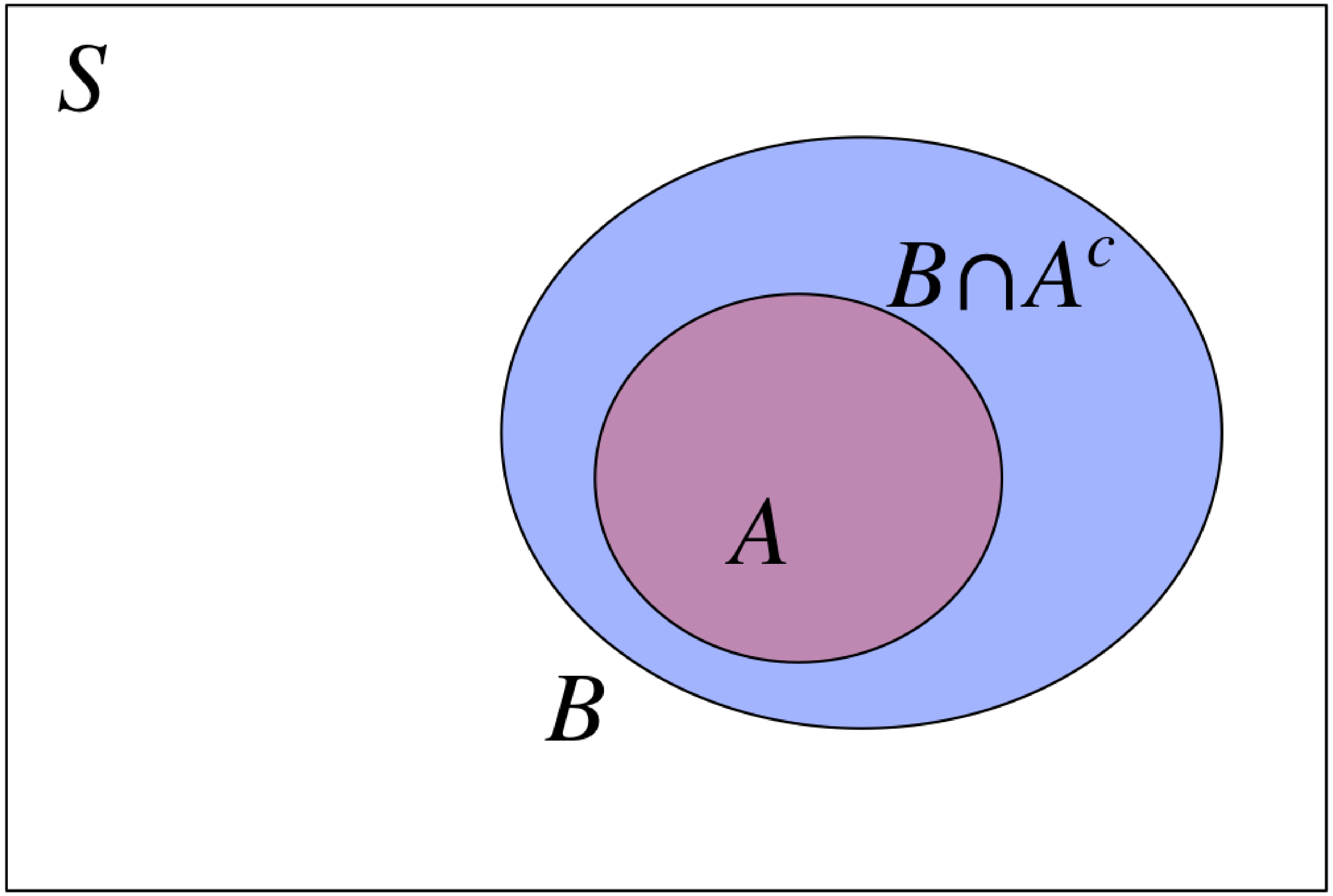
Proof

1. Since A and A^c are disjoint and their union is S, the second axiom gives

$$P(S) = P(A \cup A^c) = P(A) + P(A^c),$$

But $P(S) = 1$ by the first axiom. So $P(A) + P(A^c) = 1$.

2. If $A \subseteq B$, then we can write B as the union of A and $B \cap A^c$, where $B \cap A^c$ is the part of B not also in A. This is illustrated in the figure below.



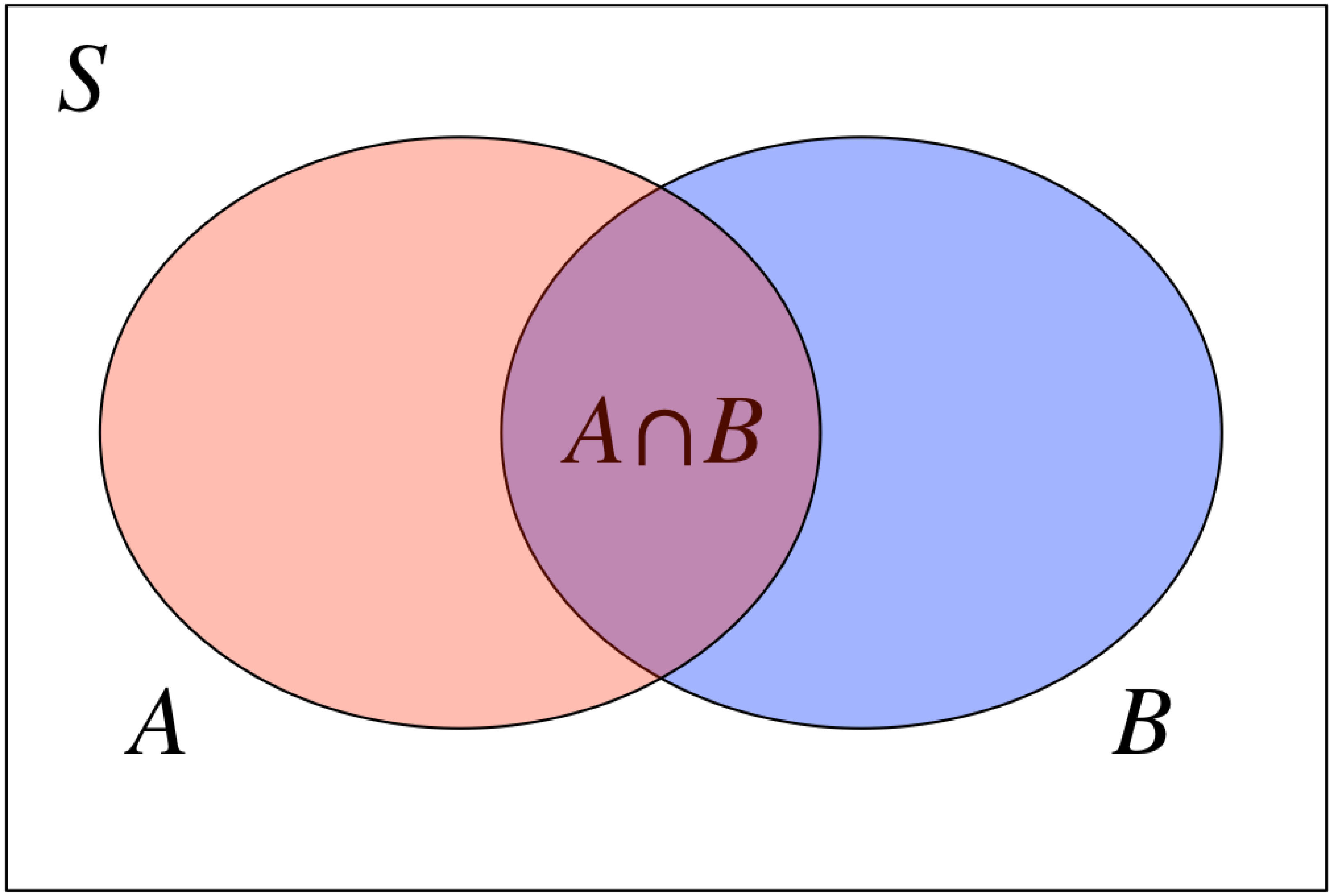
View Larger Image
Image Description

Since A and $B \cap A^c$ are disjoint, we can apply the second axiom:

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

Probability is nonnegative, so $P(B \cap A^c) \geq 0$, proving that $P(B) \geq P(A)$.

3. The intuition for this result can be seen using a Venn diagram like the one below.



View Larger Image
Image Description

The shaded region represents $A \cup B$, but the probability of this region is not $P(A) + P(B)$, because that would count the football-shaped intersection region $A \cap B$ twice. To correct for this, we subtract $P(A \cap B)$. This is a useful intuition, but not a proof.

For a proof using the axioms of probability, we can write $A \cup B$ as the union of the disjoint events A and $B \cap A^c$. Then by the second axiom,

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

So it suffices to show that $P(B \cap A^c) = P(B) - P(A \cap B)$. Since $A \cap B$ and $B \cap A^c$ are disjoint and their union is B, another application of the second axiom gives us

$$P(A \cap B) + P(B \cap A^c) = P(B).$$

So $P(B \cap A^c) = P(B) - P(A \cap B)$, as desired.

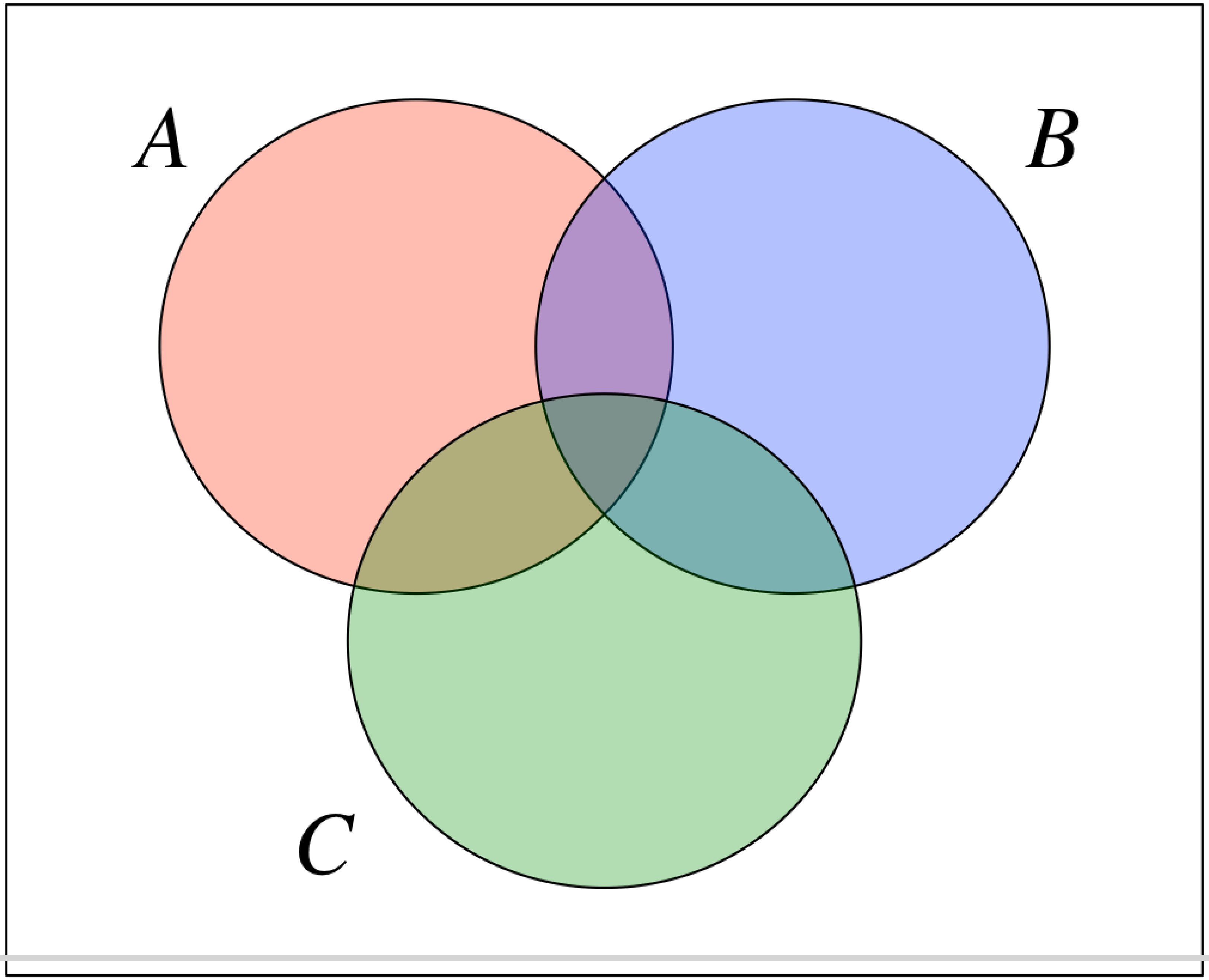
The third property is a special case of *inclusion-exclusion*, a formula for finding the probability of a union of events when the events are not necessarily disjoint. We showed above that for two events A and B,

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

For three events, inclusion-exclusion says

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

For intuition, consider a triple Venn diagram like the one below.



View Larger Image
Image Description

To get the total area of the shaded region $A \cup B \cup C$, we start by adding the areas of the three circles, $P(A) + P(B) + P(C)$. The three football-shaped regions have each been counted twice, so we then subtract $P(A \cap B) + P(A \cap C) + P(B \cap C)$. Finally, the region in the center has been added three times and subtracted three times, so in order to count it exactly once, we must add it back again. This ensures that each region of the diagram has been counted once and exactly once.

Now we can write inclusion-exclusion for n events.

Theorem 1.6.6 (Inclusion-exclusion).

For any events $A_1, \dots, A_n$,

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} P(A_1 \cap \dots \cap A_n).$$

The next example, *de Montmort's matching problem*, is a famous application of inclusion-exclusion. Pierre Rémond de Montmort was a French mathematician who studied probability in the context of gambling and wrote a treatise devoted to the analysis of various card games. He posed the following problem in 1708, based on a card game called Treize.

Example 1.6.7 (de Montmort's matching problem).

Consider a well-shuffled deck of n cards, labeled 1 through n. You flip over the cards one by one, saying the numbers 1 through n as you do so. You win the game if, at some point, the number you say aloud is the same as the number on the card being flipped over (for example, if the 7th card in the deck has the label 7). What is the probability of winning?

Solution

Let A_i be the event that the ith card in the deck has the number i written on it. We are interested in the probability of the union $A_1 \cup \dots \cup A_n$: as long as at least one of the cards has a number matching its position in the deck, you will win the game. (An ordering for which you lose is called a *derangement*, though hopefully no one has ever become deranged due to losing at this game.)

To find the probability of the union, we'll use inclusion-exclusion. First,

$$P(A_i) = \frac{1}{n}$$

for all i. One way to see this is with the naive definition of probability, using the full sample space: there are $n!$ possible orderings of the deck, all equally likely, and $(n-1)!$ of these are favorable to A_i (fix the card numbered i to be in the ith position in the deck, and then the remaining $n-1$ cards can be in any order). Another way to see this is by symmetry: the card numbered i is equally likely to be in any of the n positions in the deck, so it has probability $1/n$ of being in the ith spot. Second,

$$P(A_i \cap A_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)},$$

since we require the cards numbered i and j to be in the ith and jth spots in the deck and allow the remaining $n-2$ cards to be in any order, so $(n-2)!$ out of $n!$ possibilities are favorable to $A_i \cap A_j$. Similarly,

$$P(A_i \cap A_j \cap A_k) = \frac{1}{n(n-1)(n-2)},$$

and the pattern continues for intersections of 4 events, etc.

In the inclusion-exclusion formula, there are n terms involving one event, $\binom{n}{2}$ terms involving two events, $\binom{n}{3}$ terms involving three events, and so forth. By the symmetry of the problem, all n terms of the form $P(A_i)$ are equal, all $\binom{n}{2}$ terms of the form $P(A_i \cap A_j)$ are equal, and the whole expression simplifies considerably:

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) &= \frac{n}{n} - \frac{\binom{n}{2}}{n(n-1)} + \frac{\binom{n}{3}}{n(n-1)(n-2)} - \cdots + (-1)^{n+1} \cdot \frac{1}{n!} \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n+1} \cdot \frac{1}{n!}. \end{aligned}$$

Comparing this to the Taylor series for $1/e$, which says that,

$$e^{-1} = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \ldots,$$

we see that for large n, the probability of winning the game is extremely close to $1 - 1/e$, or about 0.63. Interestingly, as n grows, the probability of winning approaches $1 - 1/e$ instead of going to 0 or 1. With a lot of cards in the deck, the number of possible locations for matching cards increases while the probability of any particular match decreases, and these two forces offset each other and balance to give a probability of about $1 - 1/e$.

Inclusion-exclusion is a very general formula for the probability of a union of events, but it helps us the most when there is symmetry among the events A_j; otherwise the sum can be extremely tedious. In general, when symmetry is lacking, we should try to use other tools before turning to inclusion-exclusion as a last resort.