Inclusion-Exclusion Rule

Remember the Sum Rule:

The Sum Rule: If there are $n(A)$ ways to do A and, distinct from them, $n(B)$ ways to do B, then the number of ways to do A *or* B is $n(A) + n(B)$.
What if the ways of doing A and B aren't distinct?

Example: If 112 students take CS280, 85 students take CS220, and 45 students take both, how many take either CS280 or CS220.

A = students taking CS280
B = students taking CS220

$$|A \cup B| = |A| + |B| - |A \cap B| = 112 + 85 - 45 = 152$$

This is best seen using a Venn diagram:

How many numbers ≤ 100 are multiples of either 2 or 5?

Let A = multiples of $2 \leq 100$
Let B = multiples of $5 \leq 100$

Then $A \cap B$ = multiples of $10 \leq 100$

$$|A \cup B| = |A| + |B| - |A \cap B| = 50 + 20 - 10 = 60.$$

What happens with three sets?

$$\begin{aligned}
&|A \cup B \cup C| = \\
&|A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| \\
&\qquad +|A \cap B \cap C|
\end{aligned}$$

Example: If there are 300 engineering majors, 112 take CS280, 85 takes CS 220, 95 take AEP 356, 45 take both CS280 and CS 220, 30 take both CS 280 and AEP 356, 25 take both CS 220 and AEP 356, and 5 take all 3, how many don't take any of these 3 courses?

A = students taking CS 280
B = students taking CS 220
C = students taking AEP 356

$$\begin{aligned}
& |A \cup B \cup C| \\
= \; & |A| + |B| + |C| - |A \cap B| - |B \cap C| - |A \cap C| \\
& \qquad +|A \cap B \cap C| \\
= \; & 112 + 85 + 95 - 45 - 30 - 25 + 5 \\
= \; & 197
\end{aligned}$$

We are interested in $\overline{A \cup B \cup C} = 300 - 197 = 103$.

The General Rule

More generally,

$$|\cup_{k=1}^n A_k| = \sum_{k=1}^{n} \sum_{\{I \mid I\subset\{1,\ldots,n\},\, |I|=k\}} (-1)^{k-1} |\cap_{i\in I} A_i|$$

Why is this true? Suppose $a \in \cup_{k=1}^n A_k$, and is in exactly m sets. a gets counted once on the LHS. How many times does it get counted on the RHS?

- a appears in m sets (1-way intersection)
- a appears in $C(m,2)$ 2-way intersections
- a appears in $C(m,3)$ 3-way intersections
- ...

Thus, on the RHS, a gets counted

$$\sum_{k=1}^{m} (-1)^{k-1} C(m,k) \text{ times.}$$

By the binomial theorem:

$$\begin{aligned} 0 = (-1+1)^m &= \Sigma_{k=0}^m (-1)^k 1^{m-k} C(m,k) \\ &= 1 + \Sigma_{k=1}^m (-1)^k C(m,k) \end{aligned}$$

Thus,

$$\sum_{k=1}^{m} (-1)^{k-1} C(m,k) = 1.$$

Each element in $\cup_{i=1}^k A_i$ gets counted once on both sides.

A Hard Example

Suppose $m \geq 10$. How many m-digit numbers have each of the digits 0–9 at least once? (View 00305 as a 5-digit number.)

We need a systematic way of tackling this.

Let A_j be the set of m-digit numbers that have at least one occurrence of j, for $j = 0, \ldots, 9$.

We are interested in $|A_0 \cap \ldots \cap A_9|$.

The inclusion-exclusion rule applies to unions. Can we use it?

$$\overline{A_0 \cap \ldots \cap A_9} = \overline{A_0} \cup \ldots \cup \overline{A_9}$$

$|\overline{A_i}| = 9^m$
$|\overline{A_i} \cap \overline{A_j}| = 8^m$
...

$$\begin{aligned} |\cup_{i=0}^9 \overline{A_i}| &= 10 \times 9^m - C(10,2) \times 8^m + \cdots \\ &= \Sigma_{k=1}^9 (-1)^{k-1} C(10,k) \times (10-k)^m \end{aligned}$$

Thus,

$$\begin{aligned} |\cap_{i=0}^9 A_i| &= 10^m - \Sigma_{k=1}^9 (-1)^{k-1} C(10,k) \times (10-k)^m \\ &= \Sigma_{k=0}^9 (-1)^k C(10,k)(10-k)^m \end{aligned}$$

The Pigeonhole Principle

The Pigeonhole Principle: If $n+1$ pigeons are put into n holes, at least two pigeons must be in the same hole.

This seems obvious. How can it be used in combinatorial anlysis?

Q1: If you have only blue socks and brown socks in your drawer, how many do you have to pull out before you're sure to have a matching pair.

A: The socks are the pigeons and the holes are the colors. There are two holes. With three pigeons, there have to be at least two in one hole.

- What happens if we also have black socks?

Q2: Bob picks 10 numbers between 1 and 40. Alice wins if she can find two different sets of three of these numbers that have the same sum. Who wins?

A: The holes are the possible sums. The smallest sum is 6 (1+2+3), the largest is 117 (38+39+40). The pigeons are the possible ways for Alice to choose 3 numbers out of the 10 chosen by Bob.

$$\binom{10}{3} = \frac{10 \times 9 \times 8}{3 \times 2 \times 1} = 120.$$

There's always a way for Alice to win!

Probability

Life is full of uncertainty.
Probability is the best way we currently have to quantify it.

Applications of probability arise everywhere:

- Should you guess in a multiple-choice test with five choices?
  - What if you're not penalized for guessing?
  - What if you're penalized 1/4 for every wrong answer?
  - What if you can eliminate two of the five possibilities?

- Suppose that an AIDS test guarantees 99% accuracy:
  - of every 100 people who have AIDS, the test returns positive 99 times (very few *false negative*);
  - of every 100 people who don't have AIDS, the test returns negative 99 times (very few *false positives*)

  Suppose you test positive. How likely are you to have AIDS?
  - Hint: the probability is *not* .99
- How do you compute the average-case running time of an algorithm?
- Is it worth buying a $1 lottery ticket?
  - Probability isn't enough to answer this question

(I think) everybody ought to know something about probability.

Interpreting Probability

Probability can be a subtle.

The first (philosophical) question is "What does probability mean?"

- What does it mean to say that "The probability that the coin landed (will land) heads is 1/2"?

Two standard interpretations:

- Probability is *subjective*: This is a subjective statement describing an individual's feeling about the coin landing heads
  - This feeling can be quantified in terms of betting behavior
- Probability is an *objective* statement about frequency

Both interpretations lead to the same mathematical notion.

Formalizing Probability

What do we assign probability to?
Intuitively, we assign them to possible *events* (things that might happen, *outcomes* of an experiment)

Formally, we take a *sample space* to be a *set*.

- Intuitively, the sample space is the set of possible outcomes, or possible ways the world could be.

An *event* is a subset of a sample space.

We assign probability to events: that is, to subsets of a sample space.

Sometimes the hardest thing to do in a problem is to decide what the sample space should be.

- There's often more than one choice
- A good thing to do is to try to choose the sample space so that all outcomes (i.e., elements) are equally likely
  - This is not always possible or reasonable

Choosing the Sample Space

Example 1: We toss a coin. What's the sample space?

- Most obvious choice: {heads, tails}
- Should we bother to model the possibility that the coin lands on edge?
- What about the possibility that somebody snatches the coin before it lands?
- What if the coin is biased?

Example 2: We toss a die. What's the sample space?

Example 3: Two distinguishable dice are tossed together. What's the sample space?

- (1,1), (1,2), (1,3), …, (6,1), (6,2), …, (6,6)

What if the dice are indistinguishable?

Example 4: You're a doctor examining a seriously ill patient, trying to determine the probability that he has cancer. What's the sample space?

Example 5: You're an insurance company trying to insure a nuclear power plant. What's the sample space?

Probability Measures

A *probability measure* assigns a real number between 0 and 1 to every subset of (event in) a sample space.

- Intuitively, the number measures how likely that event is.
- Probability 1 says it's certain to happen; probability 0 says it's certain not to happen
- Probability acts like a *weight* or *measure*. The probability of separate things (i.e., disjoint sets) adds up.

Formally, a probability measure Pr on S is a function mapping subsets of S to real numbers such that:

1. For all $A \subseteq S$, we have $0 \leq \Pr(A) \leq 1$
2. $\Pr(\emptyset) = 0$; $\Pr(S) = 1$
3. If A and B are disjoint subsets of S (i.e., $A \cap B = \emptyset$), then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.

It follows by induction that if $A_1, \ldots, A_k$ are pairwise disjoint, then

$$\Pr(\cup_i^k A_i) = \Sigma_i^k \Pr(A_i).$$

- This is called *finite additivity*; it's actually more standard to assume a countable version of this, called *countable additivity*

In particular, this means that if $A = \{e_1, \ldots, e_k\}$, then

$$\Pr(A) = \sum_{i=1}^{k} \Pr(e_i).$$

In finite spaces, the probability of a set is determined by the probability of its elements.

Equiprobable Measures

Suppose S has n elements, and we want Pr to make each element equally likely.

- Then each element gets probability $1/n$
- $\Pr(A) = |A|/n$

In this case, Pr is called an *equiprobable measure*.

Examples

Example 1: In the coin example, if you think the coin is fair, and the only outcomes are heads and tails, then we can take $S = \{\text{heads,tails}\}$, and
Pr(heads) = Pr(tails) = 1/2.

Example 2: In the two-dice example where the dice are distinguishable, if you think both dice are fair, then we can take $\Pr((i, j)) = 1/36$.

- Should it make a difference if the dice are indistinguishable?

Equiprobable measures on infinite sets

Defining an equiprobable measure on an infinite set can be tricky.

Theorem: There is no equiprobable measure on the positive integers.

Proof: By contradiction. Suppose Pr is an equiprobable measure on the positive integers, and $\Pr(1) = \epsilon > 0$.

There must be some N such that $\epsilon > 1/N$.
Since $\Pr(1) = \cdots = \Pr(N) = \epsilon$, we have

$$\Pr(\{1, \ldots, N\}) = N\epsilon > 1 \text{ — a contradiction}$$

But if $\Pr(1) = 0$, then $\Pr(S) = \Pr(1) + \Pr(2) + \cdots = 0$.

Some basic results

How are the probability of E and $\overline{E}$ related?

- How does the probability that the dice lands either 2 or 4 (i.e., $E = \{2, 4\}$) compare to the probability that the dice lands 1, 3, 5, or 6 ($\overline{E} = \{1, 3, 5, 6\}$)

Theorem 1: $\Pr(\overline{E}) = 1 - \Pr(E)$.

Proof: E and $\overline{E}$ are disjoint, so that

$$\Pr(E \cup \overline{E}) = \Pr(E) + \Pr(\overline{E}).$$

But $E \cup \overline{E} = S$, so $\Pr(E \cup \overline{E}) = 1$.
Thus $\Pr(E) + \Pr(\overline{E}) = 1$, so

$$\Pr(\overline{E}) = 1 - \Pr(E).$$

Theorem 2: $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.

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

So

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

The result now follows.

Remember the Inclusion-Exclusion Rule?

$$|A \cup B| = |A| + |B| - |A \cap B|$$

This follows easily from Theorem 2, if we take Pr to be an equiprobable measure. We can also generalize to arbitrary unions.

Disclaimer

- Probability is a well defined mathematical theory.
- Applications of probability theory to "real world" problems is not.
- Choosing the sample space, the events and the probability function requires a "leap of faith".
- We cannot prove that we chose the right model but we can argue for that.
- Some examples are easy some are not:
  - Flipping a coin or rolling a die.
  - Playing a lottery game.
  - Guessing in a multiple choice test.
  - Determining whether or not the patient has AIDS based on a test.
  - Does the patient have cancer?

Conditional Probability

One of the most important features of probability is that there is a natural way to *update* it.

Example: Bob draws a card from a 52-card deck. Initially, Alice considers all cards equally likely, so her probability that the ace of spades was drawn is 1/52. Her probability that the card drawn was a spade is 1/4.

Then she sees that the card is black. What should her probability now be that

- the card is the ace of spades?
- the card is a spade?

A reasonable approach:

- Start with the original sample space
- Eliminate all outcomes (elements) that you now consider impossible, based on the observation (i.e., assign them probability 0).
- Keep the relative probability of everything else the same.
    - Renormalize to get the probabilities to sum to 1.

What should the probability of B be, given that you've observed A? According to this recipe, it's

$$\Pr(B \mid A) = \frac{\Pr(A \cap B)}{\Pr(A)}$$

Pr(A♠ | black) = (1/52)/(1/2) = 1/26
Pr(spade | black) = (1/4)/(1/2) = 1/2.

A subtlety:

- What if Alice doesn't completely trust Bob? How do you take this into account? Two approaches:

  (1) Enlarge sample space to allow more observations.

  (2) Jeffrey's rule:

  Pr(A♠ | black) · Pr(Bob telling the truth)+
  Pr(A♠ | red) · Pr(Bob lying).

Independence

Intuitively, events A and B are independent if they have no effect on each other.

This means that observing A should have no effect on the likelihood we ascribe to B, and similarly, observing B should have no effect on the likelihood we ascribe to A.

Thus, if $\Pr(A) \neq 0$ and $\Pr(B) \neq 0$ and A is independent of B, we would expect

$$\Pr(B|A) = \Pr(B) \text{ and } \Pr(A|B) = \Pr(A).$$

Interestingly, one implies the other.

$\Pr(B|A) = \Pr(B)$ iff $\Pr(A \cap B)/\Pr(A) = \Pr(B)$ iff

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B).$$

Formally, we say A and B are *(probablistically) independent* if

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B).$$

This definition makes sense even if $\Pr(A) = 0$ or $\Pr(B) = 0$.

Example

Alice has two coins, a fair one f and a loaded one l.

- l's probability of landing H is $p > 1/2$.

Alice picks f and flips it twice.

- What is the sample space?

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

- What is Pr?
- By symmetry this should be an equiprobable space.

Let $H_1 = \{(H,H),(H,T)\}$ and let $H_2 = \{(H,H),(T,H)\}$.

H_1 and H_2 are independent:

- $H_1 = \{(H,H),(H,T)\} \Rightarrow \Pr(H_1) = 2/4 = 1/2$.
- Similarly, $P(H_2) = 1/2$.
- $H_1 \cap H_2 = \{(H,H)\} \Rightarrow \Pr(H_1 \cap H_2) = 1/4$.
- So, $\Pr(H_1 \cap H_2) = \Pr(H_1) \cdot \Pr(H_2)$.

Alice next picks l and flips it twice.

- The sample space is the same as before:

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

- We now define Pr by *assuming* the flips are independent:
  - $\Pr\{(H,H)\} = \Pr(H_1 \cap H_2) := p^2$
  - $\Pr\{(H,T)\} = \Pr(H_1 \cap \bar{H}_2) := p(1-p)$
  - $\Pr\{(T,H)\} = \Pr(\bar{H}_1 \cap H_2) := (1-p)p$
  - $\Pr\{(T,T)\} = \Pr(\bar{H}_1 \cap \bar{H}_2) := (1-p)^2$.
- H_1 and H_2 are now independent by construction:

$$\begin{aligned}\Pr(H_1) &= \Pr\{(H,H),(H,T)\} = \\ &= p^2 + p(1-p) = p[p+(1-p)] = p.\end{aligned}$$

  Similarly, $\Pr(H_2) = \Pr\{(H,H),(T,H)\} = p$
  $\Pr(H_1 \cap H_2) = \Pr(H,H) = p^2$.
- For either coin, the two flips are independent.

What if Alice randomly picks a coin and flips it twice?

- What is the sample space?

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

  The sample space has to specify which coin is picked!

- How do we construct Pr?
- E.g.: $\Pr(f,H,H)$ should be probability of getting the fair times the probability of getting heads with the fair coin: $1/2 \times 1/4$
- Follows from the following general result:

$$\Pr(A \cap B) = \Pr(B|A)\Pr(A)$$

- So with F, L denoting the events f (respectively, l) was picked,

$$\begin{aligned}\Pr\{(f,(H,H))\} &= \Pr(F \cap (H_1 \cap H_2)) \\ &= \Pr(H_1 \cap H_2|F)\Pr(F) \\ &= 1/2 \cdot 1/2 \cdot 1/2.\end{aligned}$$

  Analogously, we have for example

$$\Pr\{(l,(H,T))\} = p(1-p)\cdot 1/2.$$

Are H_1 and H_2 independent now?

Claim. $\Pr(A) = \Pr(A|E)\Pr(E) + \Pr(A|\bar{E})\Pr(\bar{E})$

Proof. $A = (A \cap E) \cup (A \cap \bar{E})$, so

$$\Pr(A) = \Pr(A \cap E) + \Pr(A \cap \bar{E}).$$

$\Pr(H_1) = \Pr(H_1|F)\Pr(F) + \Pr(H_1|L)\Pr(L) = p/2 + 1/4$.

Similarly, $\Pr(H_2) = p/2 + 1/4$.

However,

$$\begin{aligned}&\Pr(H_1 \cap H_2) \\ =\ &\Pr(H_1 \cap H_2|F)\Pr(F) + \Pr(H_1 \cap H_2|L)\Pr(L) \\ =\ &p^2/2 + 1/4 \cdot 1/2 \\ \neq\ &(p/2 + 1/4)^2 \\ =\ &\Pr(H_1)\cdot\Pr(H_2).\end{aligned}$$

So H_1 and H_2 are dependent events.