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

$$\Pr(\cup_{i=1}^k A_i) = \Sigma_i^k \Pr(A_i).$$

- 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*.

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).

The second-ace puzzle

Alice gets two cards from a deck with four cards:
A♠, 2♠, A♡, 2♡.

A♠ A♡	A♠ 2♠	A♠ 2♡
A♡ 2♠	A♡ 2♡	2♠ 2♡

Alice then tells Bob "I have an ace".

She then says "I have the ace of spades".

The situation is similar if if Alice says "I have the ace of hearts".

Puzzle: Why should finding out which particular ace it is raise the conditional probability of Alice having two aces?

The Monty Hall Puzzle

- You're on a game show and given a choice of three doors.
  - Behind one is a car; behind the others are goats.
- You pick door 1.
- Monty Hall opens door 2, which has a goat.
- He then asks you if you still want to take what's behind door 1, or to take what's behind door 3 instead.

Should you switch?