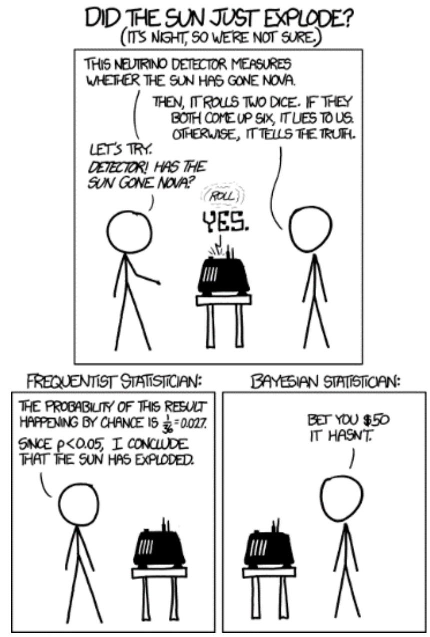
Discrete Probability (Chap 7)



Discrete Probability: Terminology

- **Experiment**: Procedure that yields an outcome
  - E.g., Tossing a coin three times
- **Outcome**: HHH in one trial, HTH in another trial, etc.
- **Sample space**: Set of all possible outcomes in the experiment
  - E.g., S = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}
- **Event**: subset of the sample space (i.e., event is a set consisting of individual outcomes)
  - E.g., Event that # of heads is an even number so E = {HHT, HTH, THH, TTT}
- If S is a sample space of equally likely outcomes, the probability of an event E, p(E) = |E|/|S|
  - Counting comes into play here!
  - E.g., probability of an event E of having even # of heads when tossing a coin three times = 4/8 = 0.5

Terminology example and some sample questions

1. Consider the question: "What is the probability of winning the lotto if there are 6 numbers drawn out of 42 numbers and you need to match them all?"
   - What is the experiment?
   - What is the sample space?
   - What is the event?
   - What is the answer to the question?

When life (or at least dice) isn't (aren't) fair.

So far we've assumed all events are equally likely. But what if they aren't? Traditionally this is discussed as having a coin (or die) that is "unfair" and the probability of each side coming up isn't the same. Let's consider a coin that has a head that is three times more likely to come up than a tail. That is, P(H)=3P(T).

- What is the probably of getting a head from a single coin toss?
- What is the probability of getting two heads when tossing two of these coins?
  A head and a tail?
  Two tails?

- What is the odds of getting an even number of heads with three flips of our biased coin?

Combinations of events and complements.

Sometimes we care about more than one event or an event not happening.

Probability of an event not happening?

Let E be an event in a sample space S. The probability of the event $\overline{E}$, the complementary event of E (i.e., a set of outcomes that E does not happen), is given by

$$p(\overline{E}) = \frac{|\overline{E}|}{|S|} = \frac{|S| - |E|}{|S|} = 1 - \frac{|E|}{|S|} = 1 - p(E)$$

- A sequence of 10 bits is randomly generated. What are the odds that at least one of those bits is a zero?

- What is the probability of having a hand that *doesn't* have a single ace (has either more or less than 1 ace)?

Union of two events.

Let E1 and E2 be events in the sample space S. Then $p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$ because $|E_1 \cup E_2| = |E_1| + |E_2| - |E_1 \cap E_2|$. We'd need to divide each term by |S| to get probabilities.

- Use this fact to find the probability of a a positive integer not exceeding 100 selected at random is divisible by 5 or 7.

Even with unfair/biased coins our combination rules still hold:

- $p(\overline{E}) = 1 - p(E)$
- $p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$

Conditional probability

Sometimes we want to know what the probability is of something given some other fact.

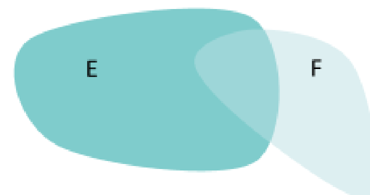
Let E and F be events with p(F) > 0. The conditional probability of E given F, defined by p(E | F), is defined as:

$$p(E \mid F) = \frac{p(E \cap F)}{p(F)}$$

Example

A bit string of length 4 is generated at random so that each of the 16 bit strings is equally likely. What is the probability that it contains at least two consecutive 0s, given that its first bit is a 0?

- E: event that the string contains at least two consecutive 0s
- F: event that the string has its first bit as 0.
- What is p(E|F)?



p(F)=___________________ p(E∩F)=_____________ (make a list)

So p(E|F)=___________

Independence

The events E and F are independent if and only if p(E∩F) = p(E)p(F). An equivalent statement is that

E and F are independent if and only if p(E | F) = p(E)

In the case above, we showed that p(E|F) and p(E) were not equal. So those two events are not independent. How about the following:

- Flip a coin three times
  - E: the first coin is a head
  - F: the second coin is a head
- Roll two dice
  - E: the sum of the two dice is 5
  - F: the first die is a 1
- Roll two dice
  - E: the sum of the two dice is 7
  - F: the first die is a 1
- Deal two five card poker hands
  - E: hand one has four of a kind
  - F: hand two has four of a kind

Another example (Example 4 page 457 of the text)

What is the conditional probability that a family with two children has two boys, given they have at least one boy? Assume that each of the possibilities BB, BG, GB, and GG is equally likely, where B represents a boy and G represents a girl. (Note that BG represents a family with an older boy and a younger girl while GB represents a family with an older girl and a younger boy.) Are those two events independent?

Bernoulli Trials and the Binomial Distribution

> Suppose that an experiment can have only two possible outcomes. For instance, when a bit is generated at random, the possible outcomes are 0 and 1. When a coin is flipped, the possible outcomes are heads and tails. Each performance of an experiment with two possible outcomes is called a Bernoulli trial, after James Bernoulli, who made important contributions to probability theory. In general, a possible outcome of a Bernoulli trial is called a success or a failure. If p is the probability of a success and q is the probability of a failure, it follows that p + q = 1.[i]

The probability of exactly k successes in n independent Bernoulli trials, with probability of success p and probability of failure $q = 1 - p$, is

$$C(n, k)p^k q^{n-k}.$$

So what is the probability of getting exactly 4 multiple choice questions out of 5 correct by guessing if each question has 3 options?

How about 3 correct? All 5?

Bayes' Theorem (7.3)

$$P(b \mid a) = \frac{P(a \mid b)P(b)}{P(a)}$$

- P(b) is the *prior* probability of b.
- P(b | a) is the *posterior* probability, after taking the evidence a into account.
- P(a | b) is the *likelihood* of the evidence, given the hypothesis.
- P(a) is the *prior* probability of the evidence
    - used as a normalizing constant

Why is this useful? Consider a medical diagnosis. Diagnostic evidence P(disease | symptom) is often hard to get. *But it's what you really want.* Causal evidence P(symptom | disease) is often easier to get. P(disease) is easy to get.

Diagnosing a Rare Disease

- Meningitis is rare: P(m) = 1/50000
- Meningitis causes stiff neck: P(s|m) = 0.5
- Stiff neck is not so rare: P(s) = 1/20
- You have a stiff neck. What is P(m|s)?

[i] Quote is from the text, page 458.

The point being that we can quickly figure out if a stiff neck is a good reason to guess someone has Meningitis.

Evaluating Public Policy

- Terrorists are rare: $P(t) = 1/10^6$
- Suppose we have a very accurate test:
  - Accuracy: $P(pos \mid t) = 0.99$
  - Specificity: $P(neg \mid \neg t) = 0.99$

	pos	neg	
t			
$\neg t$			
			300,000,000

So we want to know P(t|pos). P(pos|t) is known as is P(t). But how do we compute P(pos)?

$P(pos) = P(pos \mid t)P(t) + P(pos \mid \neg t)P(\neg t)$ ← Does that seem reasonable? If so, we get:

$$P(t \mid pos) = \frac{P(pos \mid t)P(t)}{P(pos \mid t)P(t) + P(pos \mid \neg t)P(\neg t)}$$

That form is the one we find in the text.

BAYES' THEOREM Suppose that E and F are events from a sample space S such that $p(E) \neq 0$ and $p(F) \neq 0$. Then

$$p(F \mid E) = \frac{p(E \mid F)p(F)}{p(E \mid F)p(F) + p(E \mid \overline{F})p(\overline{F})}.$$

Applications

There are a huge number of applications, mainly in artificial intelligence, related to Bayes' theorem. One you interact with most every day is called Bayesian filtering, and it's (mostly) what is used to keep your inbox spam free.

Say you have a set of B messages known to be spam and a set of G messages known to not be spam (Google, for example, gets a good sense of this when you label things as spam…). We could then search for words (or addresses, or whatever) that tend to occur in B but are less common in G.

> Suppose that we have found that the word "Rolex" occurs in 250 of 2000 messages known to be spam and in 5 of 1000 messages known not to be spam. Estimate the probability that an incoming message containing the word "Rolex" is spam, assuming that it is equally likely that an

> incoming message is spam or not spam. If our threshold for rejecting a message as spam is 0.9, will we reject such messages? (Example 3, page 473 though we'll approach it a bit differently.)

We want to find the probability that a word with "Rolex" in it is spam. We don't know what percent of all messages are spam, so let's just assume 50% for now. For p(spam | Rolex) we get

$$P(s \mid r) = \frac{P(r \mid s)P(s)}{P(r \mid s)P(s) + P(r \mid \neg s)P(\neg s)} = \frac{P(r \mid s)}{P(r \mid s) + P(r \mid \neg s)}$$ (That last is true if P(s)=.5).

Of course, filtering based on one word isn't a very good idea—you get lots of false positives. Instead you use lots of different filters all together. Example 4 on page 474 does a nice job of giving an example of combining filters.

Monty Hall

What is the probability of winning with strategy 1? What would you have had to choose if strategy 1 is to work? If strategy 2?

This question was actually the subject of a big debate where serious mathematicians got the answer wrong.

Monty Hall Problem

- Monty: secretly choose one door to be the car (the other two are "empty").
- Contestant: tell Monty which door you choose.
- Monty: reveal an "empty" door, and offer to let contestant switch doors.

- Contestant: <u>decline the offer (not switch doors)</u> (strategy 1)
- Contestant: <u>accept the offer (switch doors)</u> (strategy 2)

- Monty: reveal the prize!

Note: In the original problem, the contestant can freely choose to decline or accept without committing to a fixed strategy. However, we will consider two fixed policies to analyze the problem.

Of course, we could also use Bayes' Theorem to work this out. I'm showing that mainly as a review, but also as a bit more complex of an example.

Monty Hall (using Bayes Theorem)

Consider 3 discrete random variables all taking on values in {1,2,3}

- C: the number of the door hiding the **C**ar
- S: the number of the door **S**elected by player
- H: the number of the door opened by **H**ost

$$P(C=2|H=3,S=1) = \frac{P(H=3|C=2,S=1)P(C=2|S=1)}{\sum_{i=1}^{3} P(H=3|C=i,S=1)P(C=i|S=1)}$$
$$= \frac{1 \times \frac{1}{3}}{\frac{1}{2} \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3}} = \frac{2}{3},$$

$$P(C=1|S=1) = P(C=2|S=1) = P(C=3|S=1) = \tfrac{1}{3}.$$
$$P(H=3|C=1,S=1) = \tfrac{1}{2},$$
$$P(H=3|C=2,S=1) = 1,$$
$$P(H=3|C=3,S=1) = 0.$$

Prob. of winning by switching