September 19, 2017

CS 361: Probability & Statistics

Independence & conditional probability

Recall the definition for independence

Useful Facts: 4.4 *Independent events*

Two events $\mathcal{A}$ and $\mathcal{B}$ are **independent** if and only if

$$P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A})P(\mathcal{B})$$

- So we can *suppose* events are independent and compute probabilities
- Or we can *test* to see if two events are independent

Example

- Suppose we send a DNA sample to a lab that has a database containing DNA information on 20,000 people
- Suppose the lab can compare two samples and tell if there is a match but the test will incorrectly report a match occasionally
- Suppose the probability that this happens for a particular pair of DNA samples independently with probability of 0.0001
- If we know for sure that our sample doesn't match any of the samples in their database, what's the probability that the lab will tell us there is a match anyway?

Example

- P(lab error) = 1 - P(no error)
- We know the probability of an error on a single sample is 0.0001
- So the probability of no error on a single sample is 0.9999
- And in order for the event "no error" to occur there needs to be no error on all 20,000 samples in the database
- So P(no error) = $(0.9999)^{20000}$ or approximately 0.14
- So the probability that the lab will tell us they've found a match even when we know there isn't one is 0.86

A quick review of counting

- How many different strings can we create by re-arranging the letters in the word "horse"?
- $5! = 120$
- How about the word "Illinois"?

$$\frac{8!}{3!2!} = 3360$$

Coin flips

- If I flip a coin *N* times, how many outcomes have exactly *k* heads?
- Think of this as a string of (*N*-*k*) Ts and and *k* Hs that is *N* long
- Every re-arrangement of such a string is a valid run of this experiment
- The number of such re-arrangements is "N choose k"

$$\binom{N}{k} = \frac{N!}{k!(N-k)!}$$

Overbooking 1

- An airline has a regular flight with 6 seats. They always sell 7 tickets for this flight. If passengers show up independently with probability *p* what is the probability that the flight is overbooked?
- Can think of each individual as making a biased coin-flip. With probability *p* the person comes up *S* which means they show and with probability *(1-p)* they come up *N* or no-show
- There's only one way to write a string of 7 *S*'s
- So our probability is going to just be p^7

Overbooking 2

- An airline has a flight with 6 seats and it sells 8 tickets. Ticket holders show up independently with probability p. What is the probability that exactly 6 passengers show up?
- The event "6 passengers show up" can be written as the union of disjoint events
- In each disjoint event 6 of the 8 ticket holders show up

Overbooking 2

- Each disjoint event where 6 passengers shows up occurs with what probability?

$$p^6(1-p)^2$$

- And how many such events are there?

$$\binom{8}{6} = \frac{8!}{6!2!}$$

- So the probability of the event that exactly six people show up is

$$\frac{8!}{6!2!}p^6(1-p)^2$$

Overbooking 3

- An airline has a flight with 6 seats and it sells 8 tickets. Ticket holders show up independently with probability *p*. What is the probability that more than 6 people show up?

$$P(\text{overbooked}) = P(7S's \cup 8S's)$$

$$P(\text{overbooked}) = P(7S's) + P(8S's)$$

$$P(\text{overbooked}) = 8p^7(1-p) + p^8$$

Overbooking 4

- An airline has a flight with s seats. They sell t tickets for this flight. Each person shows up independently with probability p. What is the probability that u passengers show up?
- How many disjoint events can we think of this event as consisting of?

$$\frac{t!}{u!(t-u)!}$$

- Each with probability

$$p^u(1-p)^{t-u}$$

- Giving a probability of $$\frac{t!}{u!(t-u)!}p^u(1-p)^{t-u}$$

Overbooking 5

- An airline has a flight with s seats. They sell t tickets for this flight. Each person shows up independently with probability p. What is the probability that too many passengers show up?
- We are looking for

$$P(\{s+1 \text{ show up}\} \cup \{s+2 \text{ show up}\} \cup \ldots \cup \{t \text{ show up}\})$$

$$P(\{s+1 \text{ show up}\}) + P(\{s+2 \text{ show up}\}) + \ldots + P(\{t \text{ show up}\})$$

Overbooking 5

- Or we could write this as

$$\sum_{i=s+1}^{t} P(\{i \text{ show up}\})$$

- Or if we use our formula from the last example, we get a probability of overbooking given by

$$\sum_{i=s+1}^{t} \frac{t!}{i!(t-i)!} p^i (1-p)^{t-i}$$

Conditional probability

- Suppose we roll two dice and are interested in the probability that the sum is less than 6
- The probability of this event is 10/36
- If someone tells us that one of the dice rolled was a 4, this probability goes down to 1/6
- If someone tells us that instead one of the dice rolled was a 1, the probability would increase to 2/3

Conditional probability

- Knowing that an event has occurred might change the probability that we compute for some other event we haven't yet observed
- The probability of an event B given an event A, written $P(B \mid A)$ and called the **conditional probability** of B given A is how we capture this notion

Conditional probability

- Since event A is known to have occurred, the space of possible outcomes for the experiment, or the sample space, are only those in the event A
- The experiment outcome lies in A so P(B | A) is the probability that it also lies in $B \cap A$
- So we have

$$P(B|A) = cP(B \cap A)$$

Conditional probability

$P(B|A) = cP(B \cap A)$

Let's figure out what c is

For the event B, either it occurred or didn't
Even if we only consider the case where A occurred

$P(B|A) + P(B^c|A) = 1$

Rewriting

$cP(B \cap A) + cP(B^c \cap A) = cP(A)$

So we get

$$c = \frac{1}{P(A)}$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

Conditional probability

If we mess around with our original expression a little

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

We get

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

And

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

And this allows us to write our expression for conditional probability in the following useful way

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Car factories

- There are two car factories, A and B. Factory A produces 1000 cars, of which 10 are lemons. Factory B produces 2 cars, and both are lemons. They all go to your local car dealership
- If you buy a car, what is the probability that it is a lemon?
- $P(L) = 12/1002$
- What is the probability a car came from factory B?
- $P(B) = 2/1002$

Car factories

We had P(L) = 12/1002 and P(B) = 2/1002

Suppose you bought a car that was a lemon. What is the probability it came from factory B? I.e. what is P(B | L)?

$$P(B|L) = \frac{P(L|B)P(B)}{P(L)}$$

$$P(B|L) = \frac{(1)(2/1002)}{12/1002}$$

So P(B | L) = 1/6

Total probability

Notice that

$A = (A \cap B) \cup (A \cap B^c)$

And that $A \cap B$ and $A \cap B^c$

are disjoint events

Which means we can rewrite

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

Using the definition of conditional probability

$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$

More generally if some set of disjoint events "cover" A, e.g.

$A = A \cap (\cup_i B_i)$

Then

$$P(A) = \sum_i P(A|B_i)P(B_i)$$

False positives

- Suppose there is a blood test for a rare disease. The disease occurs in 1 in every 100,00 people. If you have the disease, the test will say so with probability 0.95. If you do not have the disease, the test will report a false positive with probability 0.001
- If you get a positive test result, what is the probability that you actually have the disease?

False positives

Suppose there is a blood test for a rare disease. The disease occurs in 1 in every 100,00 people. If you have the disease, the test will say so with probability 0.95. If you do not have the disease, the test will report a false positive with probability 0.001

We have a positive test result and want to know the probability we are actually sick

Let *S* be the event we are sick and *R* be the event we get a positive result. We want to know P(S | R)

$$P(S|R) = \frac{P(S \cap R)}{P(R)}$$

$$P(S|R) = \frac{P(R|S)P(S)}{P(R)}$$

$$P(S|R) = \frac{P(R|S)P(S)}{P(R|S)P(S) + P(R|S^c)P(S^c)}$$

False positives

Suppose there is a blood test for a rare disease. The disease occurs in 1 in every 100,00 people. If you have the disease, the test will say so with probability 0.95. If you do not have the disease, the test will report a false positive with probability 0.001

0.95 0.00001

$$P(S|R) = \frac{P(R|S)P(S)}{P(R|S)P(S) + P(R|S^c)P(S^c)}$$

0.95 0.00001 0.001 0.99999

$$P(S|R) \approx 0.0094$$

Prosecutor's fallacy

We've seen that conditional probability can easily mislead the intuition

In a trial, if a prosecutor has evidence E against a suspect, they may try to say that the probability of the evidence given that the person is innocent is very low

The quantity of relevance for justice to be served isn't how likely the evidence is, but how likely innocence is given the evidence

$$P(\mathcal{I}|\mathcal{E}) = \frac{P(\mathcal{E}|\mathcal{I})P(\mathcal{I})}{P(\mathcal{E})}$$

Quite possible for P(I | E) to be close to 1 even when P(E | I) is small

Independence and conditional probability

- Two independent events A and B have

$$P(A \cap B) = P(A)P(B)$$

- Think about what this interacts with the definition of conditional probability

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- If A and B are independent we will have

$$P(A|B) = P(A)$$

- Knowing that event B occurs tells us nothing about event A

Independence with more than two events

- If we have a set of events, there are a couple of notions of independence to be mindful of
- **Pairwise independence**: events $A_1, A_2, \ldots, A_n$ are pairwise independent if each pair of events is independent
- **Independence**: events $A_1, A_2, \ldots, A_n$ are independent if

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2)\ldots P(A_n)$$

- Independence is a much stronger assumption

Cards and independence

- Draw a card from a shuffled deck, replace it, shuffle again, draw again, shuffle again draw again. So we have three cards drawn "with replacement"
- Let A be the event that card 1 and card 2 have the same suit, B be the event that card 2 and card 3 have the same suit, and C be the event that card 1 and card 3 have the same suit

Cards and independence

3 cards, drawn with replacement

Event A: card 1 and 2 are the same suit

Event B: card 2 and 3 are the same suit

Event C: card 1 and 3 are the same suit

We have

P(A)=P(B)=P(C) $\frac{4(13)^2}{52^2} = 1/4$

And $P(A \cap B) = P(B \cap C) = P(A \cap C)$

$$\frac{4(13)^3}{52^3} = 1/16$$

So A, B, and C are pairwise independent

However, if any two of the events occurred, the third has as well, so

$$P(A \cap B \cap C) = 1/16$$

But

$$P(A)P(B)P(C) = \frac{1}{4^3}$$

So A, B, and C are only pairwise independent but not simply independent

Conditional independence

- Another notion we will use is that of conditional independence
- We say that events $A_1, A_2, \ldots, A_n$ are conditionally independent given event B if

$$P(A_1 \cap A_2 \cap \ldots \cap A_n | B) = P(A_1|B)P(A_2|B)\ldots P(A_n|B)$$

Monty hall problem

- Recall the setup, there are 3 doors, behind two of them are indistinguishable goats, behind one is a car. You pick a door and win what's behind it. You prefer to win a car to a goat
- Let's suppose you pick a door at random and before you open it, Monty announces that he will now open a door and show you a goat from among the doors you didn't pick.
- After the does this, should you switch doors from your original pick to the one that you didn't pick that is still closed?

Monty hall

- Let's call the door you picked door #1, the one the host opened door #2, and the one that you didn't pick that is still closed door #3
- Let C_i be the event that the car is behind door i and H_j be the event that the host opened door j
- We want to compute $P(C_1 | H_2)$ and compare it to $P(C_3 | H_2)$ to see if we should switch

Monty hall

First we compute P(C_1 | H_2)

1/2 1/3

$$P(C_1|H_2) = \frac{P(H_2|C_1)P(C_1)}{P(H_2|C_1)P(C_1) + P(H_2|C_2)P(C_2) + P(H_2|C_3)P(C_3)}$$

=1/3

1/2 1/3 0 1/3 1 1/3

Now let's compute P(C_3 | H_2)

1 1/3

$$P(C_3|H_2) = \frac{P(H_2|C_3)P(C_3)}{P(H_2|C_1)P(C_1) + P(H_2|C_2)P(C_2) + P(H_2|C_3)P(C_3)}$$

=2/3

1/2 1/3 0 1/3 1 1/3

Takeaway

- See the text for other way to set up Monty Hall and why it matters
- Conditional probabilities can be quite counterintuitive

Random Variables

Random variables

- We have figured out how to talk about assigning probabilities to outcomes and events
- We now look at a way of associating numbers with experiments, numbers that change as a function of the outcome of the experiment

Random variables

Definition: 5.1 *Discrete random variable*

Given a sample space Ω, a set of events $\mathcal{F}$, and a probability function P, and a countable set of of real numbers D, a discrete random variable is a function with domain Ω and range D.

Thus, for every outcome ω a random variable X associates to that outcome a real number $X(\omega)$

Example

- Flip a coin and observe the result. If it is heads, we report 1, if it is tails, we report 0. This is a random variable

Example

- We flip a coin 32 times, recording a 1 when we see heads and a 0 when we see tails. This produces a 32 bit random number which is a random variable

Example

- We flip a coin 32 times, reporting 1 for heads and 0 for tails. The parity of this 32 bit number is a random variable

Example

- We draw a hand of 5 cards. The number of pairs in the hand is a random variable (0, 1, or 2)

The relationship to events

Any value x of a random variable determines
a set of outcomes, e.g. an event

$$\{\omega : X(\omega) = x\}$$

So we will make reference to the probability
that the random variable X is equal to x

$$P(\{\omega : X(\omega) = x\})$$

And we will use shorthand P(X=x) or just P(x) to express this

Relationship to events

Another event frequently of interest is

$$\{\omega : X(\omega) \leq x\}$$

We usually write the probability of this event

$$P(\{\omega : X(\omega) \leq x\})$$

With the following shorthand

$$P(X \leq x)$$

Distributions

The function of x given by

$$P(X = x)$$

is called the **probability distribution**
of the discrete random variable X

It is defined for each value that X
can take and is 0 everywhere else

This function might also be written as p(x)

Distributions

We also give a special name to the function of x given by

$$P(X \leq x)$$

We call this the **cumulative distribution function** of the discrete random variable X

Note that this is a non-decreasing function of x

Cumulative distributions are often written with an *f*. So we may have

$$f(x) = P(\{X \leq x\})$$