CSCI 3022-002 Intro to Data Science
Bayes and pdfs

What is the probability of being dealt a flush in poker (five cards)?

Opening Example Sol'n

A, 2, 3, ... , 10, J, Q, K = len 13

Example: What is the probability of being dealt a flush in poker (five cards)?

2 ways. 1) $\frac{\text{Count flush hands}}{\text{all hands}} = \frac{\text{pick suit AND pick 5 values}}{\text{all hands}} = \frac{4 \cdot C(13,5)}{C(52,5)}$

c(4,1)

2) Conditionally

P(card 3 matches 1 & 2 GIVEN 1 & 2 matched)

$$\left(\left(\frac{1}{} \cdot \frac{12}{51}\right)\frac{11}{50}\right)\frac{10}{49} \cdot \frac{9}{48}$$

P(first card matches all prior cards)

P(flush w/ 2 cards) = P(card 2 matches card #1)

P(first 3 cards match)

recall: $C(n,k) = \frac{n!}{k!\,(n-k)!}$

Opening Example Sol'n

Example: What is the probability of being dealt a flush in poker (five cards)?

Solution: Two ways

1. Count all possible selections of five cards - $C(52,5)$ - then count all possible selections of flushes: $C(13,5)$ for the values on the flush and $C(4,1)$ for the possible suits. Then

$$P(\textit{flush}) = \frac{C(13,5)C(4,1)}{C(52,5)}$$

2. Do things *conditionally*:

$$\begin{aligned}
&P(\textit{all 5 cards same suit}) \\
&= P(\textit{cards 1-4 match suit AND card 5 matches that suit}) \\
&= P(\textit{cards 1-4 match suit})P(\textit{ card 5 matches that suit GIVEN cards 1-4 match suit}) \\
&= \cdots = \frac{52}{52}\frac{12}{51}\frac{11}{50}\frac{10}{49}\frac{9}{48}
\end{aligned}$$

$P(n,k)$ = # of ways to pick k distinct objects out of n distinct objects where order matters.

Second Opening Example

What is the probability of being dealt all 4 kings in poker (five cards)?

# of hands: 4 kings AND one other card: 48 options

Total prob: $\cancel{48/C(52,5)} = \dfrac{48}{\left(\dfrac{52!}{5!\,47!}\right)} = \dfrac{\cancel{48}\cdot 5\cdot 4\cdot 3\cdot 2}{52\cdot 51\cdot 50\cdot 49\cdot \cancel{48}}$

Conditionally: N-K-K-K-K vs. K-N-K-K-K

Same: equally likely 5 places for the "N"

$5\left(\dfrac{48}{52}\;\dfrac{4}{51}\;\dfrac{3}{50}\;\dfrac{2}{49}\;\dfrac{1}{48}\right)$

N K K K K

Second Opening Example

What is the probability of being dealt all 4 kings in poker (five cards)?
The 52 card deck has 48 "N" non-Kings and 4 "Ki" Kings. We are interested in 5 possible outcomes: that we are dealt NKiKiKiKi, KiNKiKiKi, KiKiNKiKi, KiKiKiNKi, or KiKiKiKiN. It turns out that these each have the same probability:

$$
\begin{aligned}
P(\{NKiKiKiKi\}) &= P(\#5 = N|KiKiKiKi) \cdot P(KiKiKiKi) \\
&= \frac{48}{48} \cdot P(KiKiKiKi) \\
&= \frac{48}{48} \cdot P(\#4 = K|KiKiKi) \cdot P(KiKiKi) \ldots \\
&= \frac{48}{48} \cdot \frac{1}{49} \cdot P(KiKiKi) \ldots \\
&\ldots \\
&= \frac{48}{48} \cdot \frac{1}{49} \cdot \frac{2}{50} \cdot \frac{3}{51} \cdot \frac{4}{52}
\end{aligned}
$$

Announcements and To-Dos

Announcements:

1. HW 2, due tonight (brief OH after class)!
2. Another nb day this Friday!

Last time we learned:

1. Basics of Probability in review.

To do:

1. Eyes open for next HW, posted soon.

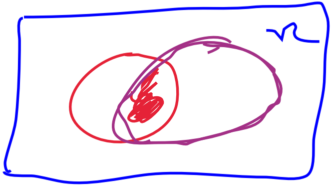
Last Time...

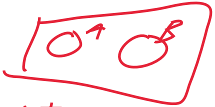
A few big takeaways from our second lecture on probability.

A

B

- *Conditional Probability*: $P(A|B) = \frac{P(A \cap B)}{P(B)}$,
- *Multiplication Rule*: $P(A \cap B) = P(A|B)P(B)$
- The following are equivalent:
  1. Two events A and B are said to be *independent*.
  2. $P(A|B) = P(A)$
  3. $P(B|A) = P(B)$
  4. $P(A \cap B) = P(A)P(B)$





not independent

- *Law of Total Probability*: Given disjoint $E_1, E_2, \ldots E_k$ such that $E_1 \cup E_2 \cup \cdots \cup E_k = \Omega$, for any A:
  $P(A) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \cdots + P(A|E_k)P(E_k)$

P(A) = P(A AND E1) + P(A AND E2) + ··· + P(A AND Ek)

$P(S) = \frac{250}{1200}$ $P(M) = \frac{150}{1200} =$ $P(M \cap S) = 40/1200$

Recall: Independence.

Example: In a school of 1200 students, 250 are seniors, 150 students take math, and 40 students are seniors and are also taking math. One student is selected at random. Let S represent the a senior is chosen, and M represent that a student taking math is chosen.

If the randomly chosen student is a senior, then what is the probability that they are taking math?

$$P(M \mid S) = \frac{P(\text{both})}{P(S)} = \frac{40/\cancel{1200}}{250/\cancel{1200}} = \frac{40}{250} = 0.16$$

Are these events independent?

$$P(M \mid S) \neq P(M) \Rightarrow \text{these are } \underline{\text{not}} \text{ independent.}$$

$$\rightarrow 3/24 = 1/8 = .125$$

Recall: Independence.
Example: In a school of 1200 students, 250 are seniors, 150 students take math, and 40 students are seniors and are also taking math. One student is selected at random. Let S represent the a senior is chosen, and M represent that a student taking math is chosen.

If the randomly chosen student is a senior, then what is the probability that they are taking math?
$P(S) = 250/1200$; $P(M) = 150/1200$; $P(M \cap S) = 40/1200/$
So, $P(M|S) = P(M \cap S)/P(S) = 250/1200 = 4/25$.

Are these events independent?
Does $P(M|S) = P(M)$? *No.*

Bayes' Theorem

The formula for $P(M|S)$ on the prior example is an example of *Bayes' Theoerem*.

multiplication rule

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

The proof follows directly from the multiplication rule, that

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

Bayes' theorem is most important mathematical way to describe *how much new information matters.*

$P(A)$ is called the *prior* information about A, and $P(A|B)$ is the *posterior* (post-data!) information about A. before B!

Bayes' Theorem

$P(1) = .7$ $\quad P(S|1) = .01$ $\quad P(S|3) = .05$

$P(2) = .2$ $\quad P(S|2) = .02$ $\quad P(N|3) = .95$

$P(3) = .1$

Example 1:
An individual has 3 different email accounts. Most of the messages, in fact 70%, come into account #1, whereas 20% come into account #2 and the remaining 10% into account #3. Of the messages into account #1, only 1% are spam, whereas the corresponding percentages for accounts #2 and #3 are 2% and 5%, respectively. What is the probability that a randomly selected message is spam?

OR $\quad$ OR

$$P(S) = P(S \text{ AND } 1) + P(S \,\&\, 2) + P(S \,\&\, 3)$$

$$= P(S|1) \cdot P(1) + \cdots P(S|2)\, P(2) + \cdots$$

OR

$$\cancel{P(1|S) \cdot P(S)}$$

$$= .01 \cdot .7 + .02 \cdot .2 + .05 \cdot .1$$

$$= .007 + .004 + .005 = .016$$

1.6%

Bayes' Theorem

Example 1:
An individual has 3 different email accounts. Most of the messages, in fact 70%, come into account #1, whereas 20% come into account #2 and the remaining 10% into account #3. Of the messages into account #1, only 1% are spam, whereas the corresponding percentages for accounts #2 and #3 are 2% and 5%, respectively. What is the probability that a randomly selected message is spam?
We know:

$$P(1) = .7; P(2) = .2; P(3) = .1; P(S|1) = .01; P(S|2) = .02; P(S|3) = .05;$$

and by LTP

$$P(S) = P(S|1)P(1) + P(S|2)P(2) + P(S|3)P(3)$$

$$P(S) = .007 + .004 + .005 = .01\cancel{8}6$$

Bayes' Theorem

Example 2:
An individual has 3 different email accounts. Most of the messages, in fact 70%, come into account #1, whereas 20% come into account #2 and the remaining 10% into account #3. Of the messages into account #1, only 1% are spam, whereas the corresponding percentages for accounts #2 and #3 are 2% and 5%, respectively. Say she selects a message at random and it is spam. What is the probability that the message came from account #1?

$$P(1 \mid S) = \frac{P(\text{both})}{P(S)} = \frac{P(S|1)\cdot P(1)}{P(S)}$$

$P(S) = P(S \cap 1) + P(S \cap 2) + \cdots$

$$= \frac{P(S|1)\cdot P(1)}{P(S|1)\cdot P(1) + P(S|2)\cdot P(2) + P(S|3)\cdot P(3)}$$

Bayes' Theorem

Example 2:

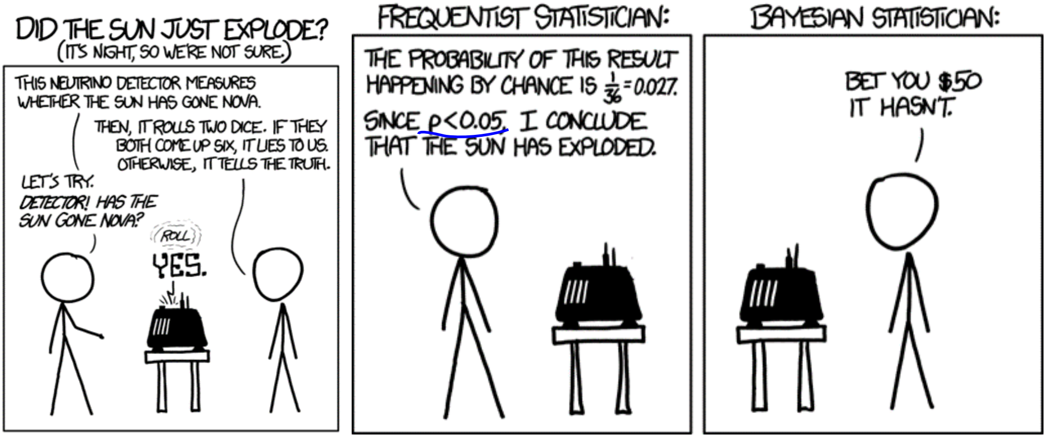
An individual has 3 different email accounts. Most of the messages, in fact 70%, come into account #1, whereas 20% come into account #2 and the remaining 10% into account #3. Of the messages into account #1, only 1% are spam, whereas the corresponding percentages for accounts #2 and #3 are 2% and 5%, respectively. Say she selects a message at random and it is spam. What is the probability that the message came from account #1?

Now we use Bayes'!

$$P(1|S) = \frac{P(S|1)P(1)}{P(S)}$$

$$P(1|S) = \frac{.007}{.01\cancel{8}\,6} = \frac{7}{1\cancel{8}\,6}$$

$P(\text{explode} \mid \text{machine says so})$



$= \underbrace{P(\text{machine says so} \mid \text{explode}) \cdot P(\text{explode})}$

small!

Definition: *Random Variable*
A *random variable* is a (measurable) function that maps elements or events in the sample space Ω to the real numbers $a_1, a_2, \ldots$ (or, more generally, to a measurable space. . . whatever that is!)
Example: Consider rolling two dice. The *Sample Space* is the full list of outcomes $\{\omega_1, \omega_2\}$.
But what if we only care about summing the two dice? We could skip the sample space and just count the *random variable:*

$X :=$ the sum of the two dice.

Probability Distributions

Definition: *Probability Density Function*
A *Probability density function* (pdf) is a function f that describes the probability distribution of a random variable X.

(outcome a)

If X is discrete, the pdf provides answers to questions like $P(X=a)$. It is also called a probability mass function (pmf).

$P(X=x)$ ← lower-case is a specific outcome

↑ Capital: random

If X is continuous, then $P(X=x)$ =0 for all x. Why?! In this case, the distribution function is called a probability density function. In the continuous case, the pdf provides answers to questions like:

$P(a \leq X \leq b)$: a region of values.

Probability Distributions

Definition: *Probability Density Function*
A *Probability density function* (pdf) is a function f that describes the probability distribution of a random variable X.

If X is discrete, the pdf provides answers to questions like $\underline{f(x) = P(X = x)}$. It is also called a probability mass function (pmf).

If X is continuous, then $\underline{P(X = x){=}0}$ for all x. Why?! In this case, the distribution function is called a probability density function. In the continuous case, the pdf provides answers to questions like:
"What is the probability that X takes on a value between a and b?"

recall: Prob is a fn: $0 \le P(w) \le 1$

$P(\Omega) = 1$

Properties of pdfs

For $f(x)$ to be a legitimate pdf, it must satisfy the following two conditions:

non-negative!

$f(x) \geq 0$ for x in the (with events) sample space

2. (For discrete distributions:)

$$\sum_{x \in \Omega} f(x) = 1$$

f is called a *probability mass function* because it describes how all of the possible outcomes in Ω have some probability or "mass" associated with them.

Properties of pdfs

For $f(x)$ to be a legitimate pdf, it must satisfy the following two conditions:

1.
$$f(x) = P(X = x) \geq 0 \qquad \forall x \text{ (with events in } \Omega\text{)}$$

2. (For discrete distributions:)
$$\sum_{x \in \Omega} f(x) = \sum_{x \in \Omega} P(X = x) = 1$$

f is called a *probability mass function* because it describes how all of the possible outcomes in Ω have some probability or "mass" associated with them.

Making a pdf

Recall; last time our opening **example**: Suppose we flip a coin with a p chance per flip of landing on heads. Define $X =$ the number of tails flips before we see a heads. What is $P(X = 0)$? $P(X = 1)$? $P(X = i)$? Verify that $P(X) = 1$ over all of Ω.

- State space: $\{H, TH, TTH, \ldots\}$
- Associated r.v. possible values or *support:* $X = \{0, 1, 2, 3, \ldots\}$
- pdf $P(X = x) =$ probability of x tails before a heads:

$$P(X=x) = P(T)^x \, P(H)$$

$$f(x) = (1-p)^x \cdot p$$

Making a pdf

Recall; last time our opening **example**: Suppose we flip a coin with a p chance per flip of landing on heads. Define $X =$ the number of tails flips before we see a heads. What is $P(X = 0)$? $P(X = 1)$? $P(X = i)$? Verify that $P(X) = 1$ over all of Ω.

- State space: $\{H, TH, TTH, TTTH, \dots\}$
- Associated r.v. possible values or *support*: $\{0, 1, 2, 3, \dots\}$
- pdf $P(X = x) =$ probability of x tails before a heads:

$$P(X = x) = P(\{T \dots TH\}) = P(\{T\})^x P(\{H\}) = (1 - p)^x \cdot p$$

So we report $f(x) = (1 - p)^x \cdot p$

Discrete pdfs

P(exactly 0 are in use) = .05

Example:
A lab has 6 computers. Let X denote the number of these computers that are in use during lunch hour, so

$$\Omega = \{0, 1, 2, \ldots, 6\}.$$

Suppose that the probability distribution of X is as given in the following table:

x	0	1	2	3	4	5	6
$P(X = x)$	.05,	.1	.15	.25	.2	.15	.1

← 7 outcomes for X

← 7 "masses"

Discrete pdfs

Example, cont'd:

x	0	1	2	3	4	5	6
$P(X = x)$	.05,	.1	.15	.25	.2	.15	.1

From here, we can find almost anything we might want to know about X.

1. Probability that at most 2 computers are in use

$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) = .3$

2. Probability that at least half of the computers are in use

$P(X \geq 3) = 1 - P(X < 3) = 1 - P(X \leq 2)$

3. Probability that there are 3 or 4 computers free

Discrete pdfs

Example, cont'd:

x	0	1	2	3	4	5	6
$P(X = x)$	.05,	.1	.15	.25	.2	.15	.1

From here, we can find almost anything we might want to know about X.

1. Probability that at most 2 computers are in use
   $P(X = 0) + P(X = 1) + P(X = 2) = .3$

2. Probability that at least half of the computers are in use
   $P(X \geq 3) = 1 - P(X < 3) = 1 - (P(X = 0) + P(X = 1) + P(X = 2)) = 1 - .3 = .7$

3. Probability that there are 3 or 4 computers free
   $P(X \cancel{\geq} 3) \cancel{= 1 - P(X = 3 \; or \; X = 4)} = 1 - (P(X = 3) + P(X = 4)) = 1 - (.25 + .2) = .55$

Cumulative Distribution Functions

Definition: *Cumulative Density Function*
For a discrete r.v. X with pdf $f(x) = P(X = x)$, the *cumulative density function*, denoted $F(x)$, is defined for every real number x to be the probability that the observed value of X will be at most x.
Mathematically:

$\{1,2,3,4,5,6\}$

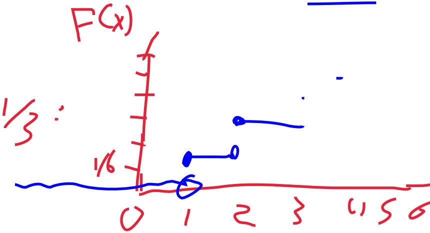
$$F(x) = P(X \leq x)$$

$f(x) = P(X = x)$

$F(x) = P(X \leq x)$

Example: If I roll a single fair die, what is the cdf?

1. $F(0) = P(X \leq 0) = 0$
2. $F(1) = P(X \leq 1) = P(X = 1) = 1/6$
3. $F(2) = P(X \leq 2) = P(X=1) + P(X=2) = 1/3$
4. $F(6)$



Cumulative Distribution Functions

Definition: *Cumulative Density Function*
For a discrete r.v. X with pdf $f(x) = P(X = x)$, the *cumulative density function*, denoted $F(x)$, is defined for every real number x to be the probability that the observed value of X will be at most x.
Mathematically:

$$F(x) = P(X \leq x)$$

Example: If I roll a single fair die, what is the cdf?

1. $F(0) = 0$
2. $F(1) = 1/6$
3. $F(2) = 2/6$
4. $F(6) = 1$: with probability 1, our roll will be $\leq$ 6.

pdf to cdf

The relationship between pdf and cdf is very important!

$$F(a) = P(X \leq a) = \sum_{x \leq a} P(X = x)$$

Example: What is the probability that if I roll two dice, they add up to at least 9. Write in terms of $F(x)$, then compute.

pdf to cdf

The relationship between pdf and cdf is very important!

$$F(a) = P(X \leq a) = \sum_{x \leq a} P(X = x)$$

Example: What is the probability that if I roll two dice, they add up to at least 9. Write in terms of $F(x)$, then compute.
$X :=$ the sum of the two dice, we want

$$P(X \geq 9) = 1 - P(X < 9) = 1 - P(X \leq 8) = 1 - F(8).$$

The easier probability is probably the
$P(X \geq 9) = P(X = 9) + P(X = 10) + P(X = 11) + P(X = 12) = 10/36.$

Daily Recap

Today we learned

1. Bayes Review... and pdfs and cdfs!

Moving forward:

- nb day Friday!
- Tonight: HW 2 due: make sure you have current version (end of problem 2)

Next time in lecture:

- We start giving special and common pdfs names!