9-16 Bayes

Monday, September 16, 2019 12:05 PM

9-16 Bayes

9-16: Bayes
CSCI 3022 Fall 19

"Counting" Example 3: What is the probability of being dealt all 4 kings in poker (five cards)?

Opening Example Sol'n

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

Deck: $\{4 \times K, 48 \times N\}$

$P(\{\{NKKKK\}, \{KNKKK\}, \ldots\}) \leftarrow 5$ "ways"

$P(\{NKKKK\}) = P(\#5 = K \mid \#1\text{-}4 = \{NKKK\}) \cdot P(\#1\text{-}4 = \{NKKK\})$

$= 1/48 \cdot P(\{NKKK\})$

$= 1/48 \cdot 2/49 \cdot 3/50 \cdot 4/51 \cdot \boxed{48}/52$

$P(\{N\})$

multiply this by 5

$P(\#2 K \mid \#1 K)$

K

$\frac{4}{52} \cdot \frac{3}{51} \cdot \frac{2}{50} \cdot \frac{1}{49} \cdot \frac{48}{48}$

Opening Example Sol'n

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:

$$P(\{NKiKiKiKi\}) = P(\#5 = N|KiKiKiKi) \cdot P(KiKiKiKi)$$
$$= \frac{48}{48} \cdot P(KiKiKiKi)$$

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

Announcements and Reminders

- Homework 2 posted
- No notebook days this week :(→ :)
- Last time in lecture: conditional probability

Last Time...

A few big takeaways from our second lecture on probability.

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

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.

$P(S) = \frac{250}{1200}$

$P(M) = \frac{150}{1200}$

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

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

find $P(M|S) = \frac{P(S \cap M)}{P(S)} = \frac{40/1200}{250/1200} = 4/25 = .16$

Are these events independent?

IS $P(M|S) = P(M)$? $= \frac{150}{1200} \neq \frac{4}{25}$ NOT

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.*

B is the data, A is the outcome studied.

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

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.

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?

$A = \{1, 2, 3\}$ $P(1) = .7$ $P(2) = .2$

$S = \{S\}$ $P(3) = .1$

$P(S|1) = .01$

$P(S|2) = .02$

$P(S|3) = .05$

$P(S) = P(S \cap 1) + P(S \cap 2) + P(S \cap 3)$ ← Law of total probability

$= P(S|1) \cdot P(1) + \text{''} + \text{''}$

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

$= .016$ or 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

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

$$P(S) = .007 + .004 + .005 = .018$$

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?

$$P(1 \mid S) = \frac{P(1 \cap S)}{P(S)}$$ → Law total probability

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

$$= \frac{.01 \cdot .7}{.016} \quad \frac{.007}{.016} = 7/16$$

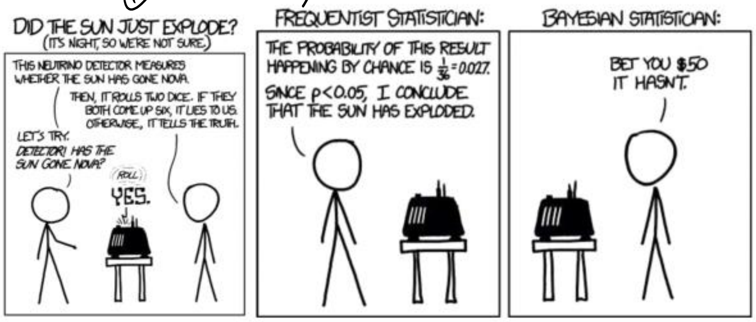
Bayes' Theorem

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'!

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

$$P(1|S) = \frac{.007}{.018} = \frac{7}{18}$$

'Data'. Says 35/36 prob. sun explodes. is there a prior?



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.

Dice 2

Die 1	1	2	3	4	5	6
1	$\boxed{2}$	3	4			
2	3	4				
3	4					
4						
5						
6						

. . . .

$X = \{2, 3, \ldots 12\}$

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 $P(X=x) = f(x)$. It is also called a probability mass function (pmf).

So for 2 fair dice..
$f(2) = P(X=2) = 1/36.$
$f(3) = P(X=3) = 2/36$

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:

"The probability X is between two 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 $f(x) = P(X = x)$. It is also called a probability mass function (pmf).

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:
"What is the probability that X takes on a value between a and b?"

Properties of pdfs

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

$$f(x) = P(X=x) \geq 0$$

non-negative

2. (For discrete distributions:) unity

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

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.

Properties of pdfs

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

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

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

$$P(\{TTTH\}) = P(T) \cdot P(T) \cdot P(T) \cdot P(H) = (1-p)(1-p)(1-p)(p)$$

$$P(X=i) = (1-p)^i \cdot p$$

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

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

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

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

$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) = .05 + .1 + .15 = .25$

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

$P(X \geq 3) = P(X=3) + P(X=4) + P(X=5) + P(X=6)$
$= 1 - P(\{X \geq 3\}^c) = 1 - P(X < 3) = 1 - P(X \leq 2) = 1 - .25$
$= .75$

3. Probability that there are 3 or 4 computers free

$1 - P($3 or 4 are not free$) = 1 - (P(X=3) + P(X=4)) = 1 - .45 = .55$

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
   $P(X \geq 3) = 1 - P(X3orX = 4) = 1 - (P(X = 3) + P(X = 4)) = 1 - (.25 + .2) = .55$

Cumulative Distribution Functions

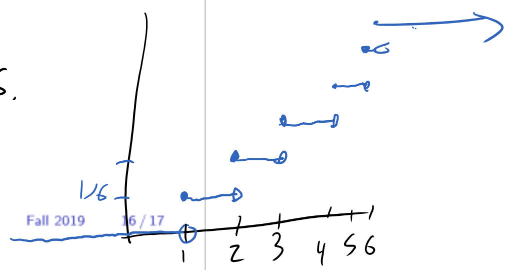
Definition: *Cumulative Density Function*
For a discrete r.v. X with pdf $f(x) = P(X = x)$, the *cumulative density function*, denoted $F(x)$, is defined for every real number x to be the probability that the observed value of X will be at most x.
Mathematically:

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

$P(X=1) = 1/6 = P(X=2) = P(X=3)$ - - - -

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

1. $F(0) = P(X \leq 0) = 0$
2. $F(1) = P(X \leq 1) = 1/6 = P(X=1)$
3. $F(2) = P(X \leq 2) = P(X=1) + P(X=2) = 2/6.$
4. $F(6) = P(X \leq 6) = 1$
   $= \boxed{P(X \leq 7) \; = P(X < 7)} = P(X \leq 42.4)$



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.