Conditional and Small Sample Probability

August 6, 2019

Bayes' Theorem

Bayes' Theorem will help us more easily calculate

$$P(\text{statement about variable 1} \mid \text{statement about variable 2})$$

when we have information about

$$P(\text{statement about variable 2} \mid \text{statement about variable 1}).$$

Example: Mammograms

- About 0.35% of women over 40 will develop breast cancer in any given year.
- In about 11% of patients with breast cancer, a mammogram test gives a **false negative**.
    - This means that the test indicates *no cancer* even though cancer is present.
- In about 7% of patients without breast cancer, the test gives a **false positive**.
    - This is when the test says that there *is cancer* when actually there is not.

Example: Mammograms

If we tested a random woman over 40 for breast cancer using a mammogram and the test came back positive for cancer, what is the probability that the patient actually has breast cancer?

Example: Mammograms

- We know that 11% of the time, a mammogram gives a false negative.
- We can use the complement to find the probability of testing positive for a woman with breast cancer:

$$1 - 0.11 = 0.89$$

- But we want the probability of cancer given a positive test result.

Example: Mammograms

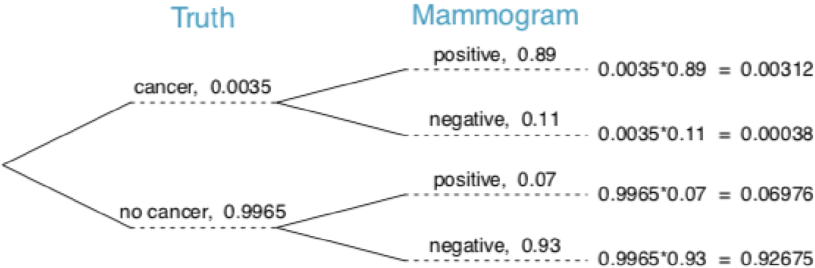
We can break this probability down into its component parts

$$P(\text{BC} \mid \text{mammogram+}) = \frac{P(\text{BC and mammogram+})}{P(\text{ mammogram+})}$$

where BC denotes breast cancer and mammogram+ denotes a positive breast cancer screening.

Example: Mammograms

We can construct a tree diagram from these probabilities:



Example: Mammograms

Returning to our desired probability,

$$P(\text{BC} \mid \text{mammogram+}) = \frac{P(\text{BC and mammogram+})}{P(\text{ mammogram+})},$$

the probability that a patient has cancer and the mammogram is positive is

$$\begin{aligned} P(\text{BC and mammogram+}) &= P(\text{mammogram+} \mid \text{BC}) \times P(\text{has BC}) \\ &= 0.89 \times 0.0035 = 0.00312 \end{aligned}$$

Example: Mammograms

The probability that the mammogram is positive is

$$
\begin{aligned}
&P(\text{mammogram+}) \\
&\quad = P(\text{mammogram+ and BC}) + P(\text{mammogram+ and no BC}) \\
&\quad = P(BC)P(\text{mammogram+} \mid \text{BC}) + P(\text{no BC})P(\text{mammogram+} \mid \text{no BC}) \\
&\quad = 0.0035 \times 0.89 + 0.9965 \times 0.07 = 0.07288
\end{aligned}
$$

Example: Mammograms

Plugging these back in,

$$\begin{aligned} P(\text{BC} \mid \text{mammogram+}) &= \frac{P(\text{BC and mammogram+})}{P(\text{ mammogram+})} \\ &= \frac{0.00312}{0.07288} = 0.0428 \end{aligned}$$

Even if a patient has a positive mammogram screening, there is still only a 4% chance of breast cancer!

This is why doctors usually run several tests before deciding that a person has a (relatively) rare disease or condition.

Law of Total Probability

Notice that the denominator of the previous equation was

$$\begin{aligned}&P(\text{mammogram+ and BC}) + P(\text{mammogram+ and no BC})\\&\quad = P(BC)P(\text{mammogram+} \mid \text{BC}) + P(\text{no BC})P(\text{mammogram+} \mid \text{no BC})\end{aligned}$$

This is the sum of the probabilities for each positive screening scenario.

Law of Total Probability

For two events A and B, the Law of Total Probability states

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_k)P(A_k)$$

where $A_1 \ldots A_k$ are the k possible outcomes for event A.

Bayes' Theorem

Consider the following conditional probability for variable 1 and variable 2:

$$P(\text{outcome } A_1 \text{ of variable 1} \mid \text{outcome } B \text{ of variable 2})$$

Bayes' Theorem states that this conditional probability can be identified as the following fraction

$$\frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_k)P(A_k)}$$

Bayes' Theorem

Bayes' Theorem is a generalization of what we've been doing with tree diagrams.

- The numerator identifies the probability of getting both A_1 and B.
- The denominator is the marginal probability of getting B.
- This bottom component of the fraction looks complicated since we have to add up probabilities from all of the different ways to get B.

Bayes' Theorem

To apply Bayes' Theorem correctly, there are two preparatory steps:

1. Identify the marginal probabilities of each possible outcome of the first variable.
$$P(A_1), P(A_2), \ldots, P(A_k)$$
2. Identify the probability of the outcome B, conditioned on each possible scenario for the first variable.
$$P(B|A_1), P(B|A_2), \ldots, P(B|A_k)$$

When each of these has been identified, they can be plugged into Bayes' Theorem.

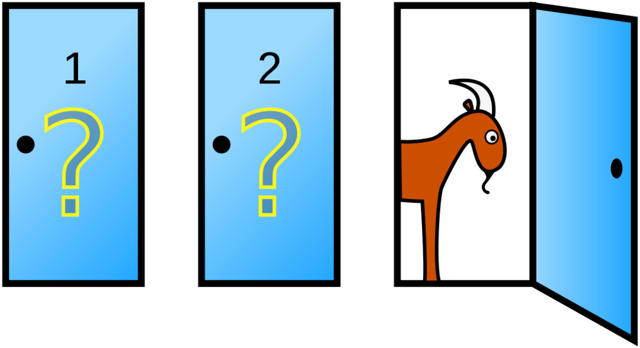
Bayes' Theorem

Bayes' Theorem tends to be a good option when there are so many scenarios that drawing a tree diagram would be very complex.

Each probability is found and identified in the same way as when creating a tree diagram.

Unless specifically asked to use either a tree diagram or Bayes' Theorem, you may use whichever method you prefer.

The Monty Hall problem comes from an old game show. There are three doors. Behind one of the doors is a car. Behind the other two doors there are goats. The goal is to win the car.



You begin by choosing a door. The host then opens one of the other two doors, always such that the opened door reveals a goat.

Monty Hall Problem

You then have the option to stay with your original choice or switch to the remaining unopened door.

Would you switch or stay? Does it matter?

Monty Hall Problem

Intuition suggests that there is a 50% chance of each of the remaining doors contain the car.

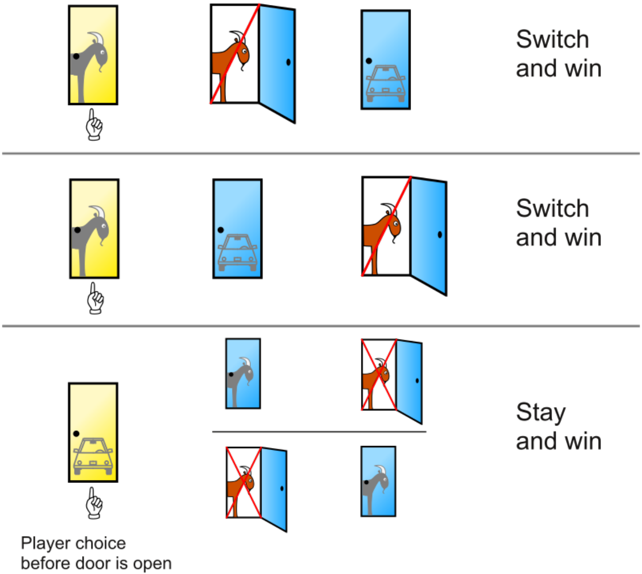
We will examine this using (1) a visual and (2) Bayes' Theorem.

Monty Hall Problem: Visual

The order of the doors doesn't matter, so for convenience we suppose that we start by choosing Door 1. The host always shows us a door with no goat. Let's see what happens in each scenario:

Door 1	Door 2	Door 3	Stay	Switch
Goat	Goat	**Car**	Lose	**Win**
Goat	**Car**	Goat	Lose	**Win**
Car	Goat	Goat	**Win**	Lose

2/3 of the time, switching leads to a win!


Switch
and win
Switch
and win
Stay
and win
Player choice
before door is open

Monty Hall Problem: Bayes' Theorem

Let D_A be the event that Door A has a car behind it, D_B the event that Door B has a car behind it, and D_C the event that Door C has a car behind it. Let H_B be the event that the host opens Door B.

Monty Hall Problem: Bayes' Theorem

Suppose we choose Door A. We want to know

$$P(D_A|H_B) = \frac{P(D_A \text{ and } H_B)}{P(H_B)}$$

or the probability that the car is behind Door A, our original choice, given that the host opened Door B. This is the probability that we **win** when we **stay**.

Monty Hall Problem: Bayes' Theorem

First,

$$\begin{aligned} P(D_A \text{ and } H_B) &= P(H_B|D_A)P(D_A) \\ &= \frac{1}{2} \times \frac{1}{3} \\ &= \frac{1}{6} \end{aligned}$$

Why does $P(H_B|D_A) = 1/2$?

Monty Hall Problem: Bayes' Theorem

Then we need to find $P(H_B)$. Using the Law of Total Probability,

$$\begin{aligned} P(H_B) &= P(H_B|D_A)P(D_A) + P(H_B|D_B)P(D_B) + P(H_B|D_C)P(D_C) \\ &= \frac{1}{2} \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3} \\ &= \frac{1}{6} + 0 + \frac{1}{3} \\ &= \frac{1}{2} \end{aligned}$$

Monty Hall Problem: Bayes' Theorem

Plugging these back into our equation for Bayes' Theorem,

$$\begin{aligned} P(D_A|H_B) &= \frac{P(D_A \text{ and } H_B)}{P(H_B)} \\ &= \frac{1}{6}\Big/\frac{1}{2} \\ &= \frac{1}{3} \end{aligned}$$

So the probability of winning if we stay with our original door is 1/3!

Sampling From a Small Population

- Usually we sample only a very small fraction of the population.
- However, we may occasionally sample more than 10% of the population without replacement.
  - Without replacement means we do not have a chance of sampling the same cases twice.
  - Think back to the raffle drawing: without replacement is when we pull 10 raffle tickets without putting any of those tickets back.
- This can be important for how we analyze the sample.

Example: Sandwiches

Suppose we have

- Two types of bread.
- Four types of filling.
- Three different condiments.

Assume we use only one of each category.

How many different types of sandwiches can we make?

Example: Sandwiches

We can visualize this using a tree diagram. Let's do this on the board.

Example: Sandwiches

We can also calculate the number of different possible sandwiches directly.

- First, we choose one of two types of bread.
- For each bread choice, we can choose one of four filling types.
  - This makes $2 \times 4 = 8$ combinations.
- Then we choose one of three condiments.
  - Each of our 8 combinations can branch into 3 further options, for a total of $8 \times 3 = 24$ combinations.

Therefore, there are $2 * 4 * 3 = 24$ combinations.

Example: Sandwiches

Now that we know the possible number of sandwiches, we can calculate the probability of any particular sandwich.

If we grab bread, filling, and a condiment at random, what's the probability that we get a cheese sandwich on rye with mayonnaise?

This is one of 24 combinations, so $P(\text{rye and cheese and mayo}) = 1/24$.

Example: Sandwiches

If we chose a sour dough and then grabbed filling and a condiment at random, what's the probability that we put cheese and mustard on our sandwich?

Now we want to know P(cheese and mustard — sourdough).

$$P(\text{cheese and mustard} \mid \text{sourdough}) = \frac{P(\text{cheese and mustard and sourdough})}{P(\text{sourdough})}$$

Example: Sandwiches

Now, cheese and mustard and sourdough is one particular combination out of our eight possible combinations so

$$P(\text{cheese and mustard and sourdough}) = 1/24$$

and sourdough is one of two possible breads, so

$$P(\text{sourdough}) = 1/2.$$

Example: Sandwiches

If we chose a sour dough and then grabbed filling and a condiment at random, what's the probability that we put cheese and mustard on our sandwich?

Plugging in,

$$
\begin{aligned}
P(\text{cheese and mustard} \mid \text{sourdough}) &= \frac{P(\text{cheese and mustard and sourdough})}{P(\text{sourdough})} \\
&= \frac{1}{24}\Big/\frac{1}{2} \\
&= 1/12
\end{aligned}
$$

Example

Suppose your discussion TA asks 3 questions and calls on people at random to answer them. Assume that he will not call on the same person twice.

What is the probability that you will not be selected?

Example

Suppose there are 25 people in your discussion.

- For the first question, your TA will choose 1 of 25 students.
  - You have a $24/25 = 0.960$ chance of *not* being selected.
- For the second question, your TA will choose 1 of the 24 people who have not yet been called on.
  - You have a $23/24 = 0.0.958$ chance of *not* being selected.
- For the final question, your TA will choose 1 of the 23 people who have not yet been called on.
  - You have a $22/23 = 0.957$ chance of *not* being selected.

Example

Then, based on the General Multiplication Rule

$$P(\texttt{Q1} = \texttt{not selected} \text{ and } \texttt{Q2} = \texttt{not selected} \text{ and } \texttt{Q3} = \texttt{not selected})$$
$$= \frac{24}{25} \times \frac{23}{24} \times \frac{22}{23}$$
$$= \frac{22}{25} = 0.88$$

Example

The three probabilities we computed were actually one marginal probability:

$$P(\texttt{Q1} = \texttt{not selected})$$

and two conditional probabilities:

$$P(\texttt{Q2} = \texttt{not selected} \mid \texttt{Q1} = \texttt{not selected})$$
$$P(\texttt{Q3} = \texttt{not selected} \mid \texttt{Q1} = \texttt{not selected},\ \texttt{Q2} = \texttt{not selected}).$$

Using the General Multiplication Rule, the product of these three probabilities is the probability of not being picked in 3 questions.

Small Sample Probabilities

When it comes to small samples...

- If we sample from a small population **without replacement**, we no longer have independence between our observations.
- If we sample from a small population **with replacement**, we have independent observations.

The key to working with small sample probabilities is to determine which sampling method was used.

Example: Socks

In your sock drawer you have 4 blue, 5 grey, and 3 black socks. You grab 2 socks at random and put them on.

Find the probability you end up wearing matching socks.

Example: Socks

Find the probability you end up wearing matching socks.

There are three ways to get matching socks:

1. $P(\text{blue and blue}) = 4/12 \times 3/11 = 0.0909$
2. $P(\text{grey and grey}) = 5/12 \times 4/11 = 0.1515$
3. $P(\text{black and black}) = 3/12 \times 2/11 = 0.0455$

Find the probability you end up wearing matching socks.

We want to find

$$\begin{aligned}
&P(\text{matching socks})\\
&\quad= P(\text{blue and blue OR grey and grey OR black and black})\\
&\quad= P(\text{blue and blue}) + P(\text{grey and grey}) + P(\text{black and black})\\
&\quad= 0.0909 + 0.1515 + 0.0455\\
&\quad= 0.2879.
\end{aligned}$$

Random Variables

- We often model processes using what's called random variables.
- Random variables give us a mathematical framework for working with real-world variables.
- This allows us to make predictions and statistical inference.

Example: Textbooks

Two books are assigned for a statistics class: a textbook and its corresponding study guide. The university bookstore determined that

- 20% of enrolled students do not buy either book
- 55% buy the textbook only
- 25% buy both books

If there are 100 students enrolled, how many books should the bookstore expect to sell to this class?

Example: Textbooks

If there are 100 students enrolled, how many books should the bookstore expect to sell to this class?

- Around $100 \times 0.20 = 20$ students will buy neither book (0 books sold).
- Around $100 \times 0.55 = 55$ students will buy the textbook only (55 books sold).
- Around 100×0.25 students will buy both books (50 books sold).

The bookstore should expect to sell about $55 + 50 = 105$ books for this class.

Example: Textbook

Now suppose the textbook costs \$137 and the study guide \$33. How much revenue should the bookstore expect from this class of 100 students?

- A student who buys only the textbook spends \$137.
  - We expected about 55 students to buy the textbook only, for a total of $\$137 \times 55 = \$7535$
- A student who buys both books spends $\$137 + \$33 = \$170$
  - We expected about 25 students to buy both books, for a total of $\$170 \times 25 = \$4250$

Example: Textbook

Now suppose the textbook costs \$137 and the study guide \$33. How much revenue should the bookstore expect from this class of 100 students?

- In total, the bookstore can expect $\$7535 + \$4250 = \$11785$ from this class each term.
- However, some *sampling variability* will cause this number to differ slightly each term.

Expectation

- We call a variable or process with a numerical outcome a **random variable**.
- We usually represent random variables with capital letters such as X, Y, or Z.
- The amount of money a single student will spend on her statistics books is a random variable. We might represent it by X.

Expectation

- The possible outcomes of X are labeled with a corresponding lower case letter x and subscripts.
- For our textbook example, we would write
  - $x_1 = \$0$
  - $x_2 = \$137$
  - $x_3 = \$170$

Expectation

- The corresponding probabilities may be written as
  - $P(X = x_1) = P(X = \$0) = 0.20$
  - $P(X = x_2) = P(X = \$137) = 0.55$
  - $P(X = x_3) = P(X = \$170) = 0.25$

Expectation

The probability distribution for X looks like

i	1	2	3	Total
x_i	\$0	\$137	\$170	-
$P(X = x_i)$	0.20	0.55	0.25	1.00

Expectation

Previously, we computed the average outcome of X as \$117.85.

- We call this average outcome the **expected value** of X, denoted $E(X)$.
- The expected value of a random variable is computed by adding each outcome weighted by its probability.

$$\begin{aligned} E(X) &= 0 \times P(X = 0) + 137 \times P(X = 137) + 170 \times P(X = 170) \\ &= 0 \times 0.20 + 137 \times 0.55 + 170 \times 0.25 \\ &= 117.85 \end{aligned}$$

Expected Value of a Discrete Random Variable

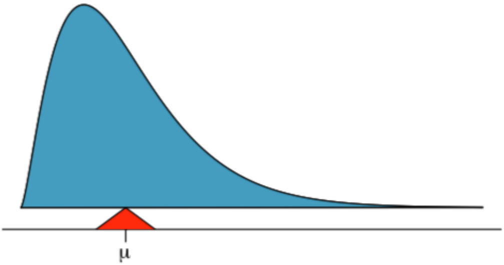
If X takes outcomes $x_1, \ldots, x_k$ with probabilities $P(X = x_1), \ldots, P(X = x_k)$, the expected value of X is the sum of each outcome multiplied by its corresponding probability:

$$\begin{aligned} E(X) &= x_1 \times P(X = x_1) + \cdots + x_k \times P(X = x_k) \\ &= \sum_{i=1}^{k} x_i P(X = x_i) \end{aligned}$$

Expected Values

- The expected value for a random variable represents the average outcome.
- For example, $E(X) = 117.85$ represents the average amount the bookstore expects to make from a single student.
- You will occasionally see the expected value denoted as μ. We will explore how this relates to the true/population mean as we go.

Expected Value of a Continuous Random Variable



- We can also calculate the expected value for a continuous random variable.
- This requires a little bit of calculus, so we won't require it for this course.
- If you are familiar with Riemann sums and integrals, this is a similar transition from discrete to continuous.

Variability in Random Variables

For the bookstore looking at textbook revenues, it might also be of interest to know about the variability in revenue.

- The variance and standard deviation can be used to describe the variability of a random variable.
- We talked about calculating variance as the sum of the squared deviances from the mean.

Variability in Random Variables

For the bookstore looking at textbook revenues, it might also be of interest to know about the variability in revenue.

- Calculating a variance for a random variable is similar, but now we weight each squared deviance by its corresponding probability.
- This is somewhere in between the variance formula we talked about in Chapter 2 and the weighting we used for the expected value.
- We again calculate the standard deviation as the square root of the variance.

Variance Formula

If X takes outcomes $x_1, \ldots, x_k$ with probabilities $P(X = x_1), \ldots, P(X = x_k)$ and expected value $\mu = E(X)$, then the variance of X, denoted by $Var(X)$ or σ^2, is

$$\begin{aligned} Var(X) &= (x_1 - \mu)^2 \times P(X = x_1) + \cdots + (x_k - \mu)^2 \times P(X = x_k) \\ &= \sum_{j=1}^{k} (x_j - \mu)^2 P(X = x_j) \end{aligned}$$

The standard deviation of X, labeled $sd(X)$ or σ, is the square root of the variance.

Example: Textbooks

Compute the expected value, variance, and standard deviation of X, the revenue of a single statistics student for the bookstore.

Example: Textbooks

Compute the expected value of X, the revenue of a single statistics student for the bookstore.

It may be helpful to modify our probability distribution table to include additional calculations:

i	1	2	3	Total
x_i	\$0	\$137	\$170	-
$P(X = x_i)$	0.20	0.55	0.25	1.00
$x_i \times P(X = x_i)$	0	75.35	42.50	117.85

This total is our expected value, $E(X) = \$117.85$.

Example: Textbooks

Compute the variance and standard deviation of X.

We will continue to modify our probability distribution table to include other calculations:

i	1	2	3	Total
x_i	\$0	\$137	\$170	
$P(X = x_i)$	0.20	0.55	0.25	
$x_i \times P(X = x_i)$	0	75.35	42.50	117.85
$x_i - \mu$	-117.85	19.15	52.15	
$(x_i - \mu)^2$	13888.62	366.72	2719.62	
$(x_i - \mu)^2 \times P(X = x_i)$	2777.7	201.7	679.9	3659.3

The second total is our variance, $Var(X) = 3659.3$. The standard deviation is $sd(X) = \sqrt{3659.3} = \$60.49$

Linear Combinations of Random Variables

So far, we've considered each variable individually, but sometimes we may be more interested in a combination of variables.

For example, the amount of time a person spends commuting to work each week may be broken down into daily commutes.