

Lecture 1: What is Bayes?

Conceptual overview and differences between frequentist and Bayesian statistics

Statistics describe a sample (often called a parameter estimate). *Parameters* describe an entire population.

- Usually, parameters are unknowable quantities.

Frequentist statistics:

Frequentists interpret probability as a relative frequency of an event rather than a description of knowledge. In other words, frequentists approach data analysis by asking: how would the parameters change if the experiment were repeated many times. Hence “frequentism.”

- Parameters are considered point estimates. Uncertainty in these estimates are in light of unobserved future data
 - e.g., if we were to repeat the experiment many times with new samples each time, how often would we expect the true parameter to fall within a certain interval around our sample statistic (usually 95%)?
- In frequentist statistics, the model is considered fixed or “true” whereas the data are random. Analyses essentially assess the probability of the data given the “true” model parameters $P(D|\theta)$.

Bayesian statistics

In contrast, Bayesian inference is the re-allocation of credibility (probability) from prior knowledge to posterior knowledge in light of the data. The probabilistic question is flipped relative to frequentist statistics:

- What is the probability of the model (what we actually want) given the (usually) known and observed data.
 - Parameters are not point estimates but are instead random variables drawn from some distribution.
 - Emphasis is explicitly on parameter estimation (whether for inference or prediction) rather than null hypothesis testing.

Steps of analysis:

1. What is the question and what are the data? Identify the relevant response and exploratory variables given the question of interest.
2. Create a descriptive mathematical model for the question and data at hand. The parameterization should result in biologically meaningful estimates and be theoretically sound.

3. Specify prior distributions for parameters.
4. Use Bayes to reallocate credibility across parameters to get posterior parameter estimates.
5. Assess model validity using QC diagnostics and posterior predictive simulation. If model is a poor fit, go back to steps 1 & 2.
6. Assuming the model is reasonable and does a good job, interpret the posterior with respect to the greater biological questions of interest.

Probability recap:

As a refresher, probabilities are a way to assign numbers to possibilities.

Probabilities have three properties:

1. Probabilities are non-negative ($p \geq 0$).
2. The sum of probabilities across all events in the sample space must = 1.
3. For any two mutually exclusive events, the probability that one or the other happens is the sum of their probabilities.

Joint and conditional probabilities:

For the sake of expedience, I am going to use the example table 4.1 from the puppy book (Kruschke, 2015). Sorry I didn't come up with a better example; it's lame.

	Hair Color				
Eye Color	Black	Brunette	Red	Blond	Marginal (eye color)
Brown	0.11	0.20	0.04	0.01	0.37
Blue	0.03	0.14	0.03	0.16	0.36
Hazel	0.03	0.09	0.02	0.02	0.16
Green	0.01	0.05	0.02	0.03	0.11
Marginal (hair color)	0.18	0.48	0.12	0.21	1.00

Each main cell contains the *joint probability* of combinations of eye and hair color and is denoted $p(e, h)$.

- Joint probabilities are symmetric: $p(e, h) = p(h, e)$.

Marginal probabilities are the probabilities of eye or hair color overall, regardless of the other variable.

- They are computed by summing the probabilities of each row or column ($p(e) = \sum_h p(e, h)$ & $p(h) = \sum_e p(e, h)$).
- When variables are continuous, $p(e, h)$ is a probability density and the marginal probability requires integration rather than summation. Therefore, for an eye color row, the marginal probability would be $p(e) = \int p(e, h)dh$.
 - This is called *marginalizing over h* or *integrating out h* . Likewise, $p(h) = \int p(e, h)de$.

In both frequentist and Bayesian statistics we are ultimately after the probability of some event x *given* we know another outcome y happened or is true ($p(x|y)$). We call these *conditional probabilities*. $p(x|y)$ can be interpreted as among all the joint outcomes with value y , what proportion share value x .

- For example, from the table above: given that someone from this population is a brunette, what is the probability they have green eyes?
 - Easy! This is just the joint probability of brunette and green eyes divided by the marginal probability of being a brunette: $0.05/0.48 = 0.104$.

Formally, we can define conditional probabilities as:

$$p(x|y) = \frac{p(x, y)}{p(y)} = \frac{p(x, y)}{\sum_x p(x, y)} \quad (1)$$

and when x is continuous,

$$p(x|y) = \frac{p(x, y)}{p(y)} = \frac{p(x, y)}{\int p(x, y) dx}. \quad (2)$$

- It's important to note that generally $p(x|y) \neq p(y|x)$.

Bayes' theorem

Monty Hall problem. *Let's make a deal!*

This problem was made famous by a column “ask Marilyn” {Marilyn vos Savant} in Parade magazine, though other flavors have existed.

The game show Let's Make a Deal was hosted by Monty Hall. The protocol (for those not born in the 60's–80's) is that you are presented with three doors. One door has a big prize that you want like a new car, the others have goats.

1. You pick a door.
2. Monty picks one of the doors you haven't chosen—it's a goat. Monty always shows you a goat.
3. Monty asks you if you want to switch doors.

Do you switch? (this is where the audience starts shouting at you!)

Spoiler alert! you should switch. On average, you will win 66% of the time if you do.

Sound weird? It was weird to thousands of people who wrote in to Parade magazine to criticize Marilyn's suggestion, many of whom had Ph.D.s in mathematics and statistics. But she was correct.

From your perspective, when the game begins, each door has a $\frac{1}{3}$ probability of having a car, which you really want.

- This should make sense, given what we know about probabilities. Each door has an equal $\frac{1}{3}$ probability of having a car: $\frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 1$.

Given what we know, some might even say that our “*prior knowledge*” is that each door has a $\frac{1}{3}$ chance of having a car.

- By extension, we could also say that our door has a $\frac{2}{3}$ chance of not having a car. More important, there is a $\frac{2}{3}$ chance that the car is in one of the other two doors.

Then Monty opens one of the other doors. He never opens a door that has a car and doesn’t open the door you have chosen. The door he opens will always have a stinkin goat (I assume it stinks because goats are smelly animals).

- In other words, Monty does not open a door at random. He knows what he’s doing because he knows where the car is. But he has done more than occupy time between commercials! He has provided us with useful information.

Here is why you switch. You have learned something new. You started with a $\frac{1}{3}$ chance of picking the correct door, and still have that $\frac{1}{3}$ if you stay with your initial choice.

But, the remaining door has a $\frac{2}{3}$ chance of having the car because you know that the door he has already opened has a $\frac{0}{3}$ chance of having the car.

Another way to think about it is to imagine if Monty asked you if you want to pick one door or two doors. Obviously you want to pick two doors.

The trick to this question is that you make a choice based on some prior information and then you have to rethink your probabilities once you get more information.

More explicitly:

Your prior probability of choosing the correct door is $\frac{1}{3}$.

Then Monty opens a door...and gives you some information.

Now the probability that it is in either of the two doors remains $\frac{1}{3}$. This probability is the likelihood of each remaining door hiding the car.

Let’s say you choose door A. Monty opened door B to expose his goat. Now we can reframe what we know.

Our prior probability of winning the car is $\frac{1}{3}$. $P(A)$ —the probability it is behind door A.

If we know the goat is behind door B, because Monty showed us, then the likelihood that it is behind door A is $\frac{1}{2}$. Or, $P(B_{goat}|A_{car}) = \frac{1}{2}$.

Now we have to break down all the possible ways Monty might have chosen door B if we had guessed correctly with door A:

1. The probability of Monty choosing door B given that the car is behind door A. In this case, Monty could have chosen either of the other doors. So the probability in this scenario is the $\frac{1}{3}$ chance we guessed correct, times the probability he would choose door B, $\frac{1}{2}$.
 - In math speak, $\frac{1}{3} \times \frac{1}{2}$.
2. The probability of Monty choosing door B given the car is behind door B. Monty would never do this!. So the probability in this scenario is the $\frac{1}{3}$ chance we guessed correct, times the probability he would choose door B, in this case 0.
 - In math speak, $\frac{1}{3} \times 0$.

3. The probability of Monty choosing door B given that the car is behind door C. Monty will never open your door, and will never show you the car. If the car is indeed behind door B, he will always show you door C. So, the probability in this scenario is the $\frac{1}{3}$ chance we guessed correct times the probability he would choose door B, in this case 1.
 - In math speak, $\frac{1}{3} \times 1$.

We can sum all these things up as a way of describing the probability of the data, in particular the probability of him choosing door B given that you choose door A.

$$\frac{1}{3} \times \frac{1}{2} + \frac{1}{3} \times 0 + \frac{1}{3} \times 1 = \frac{1}{6} + \frac{0}{6} + \frac{1}{3} = \frac{3}{6} = \frac{1}{2}$$

Now here is what we know. The prior probability of choosing the car $P(A_{car}) = \frac{1}{3}$. The likelihood Monty opened door B given it was actually behind door A is $\frac{1}{2}$.

So, the probability that we picked the winning door is the likelihood that Monty chose door B given that the car is behind door A times the prior probability we chose the winning door divided by the probability that Monty chose door B.

This probability is the *Posterior probability*.

$$P(A_{car}|B_{opened}) = \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2}}$$

$$P(A_{car}|B_{opened}) = \frac{1}{3}$$

So, given what we know, the probability that the car is actually behind door A is $\frac{1}{3}$. Thus, the probability the car is behind door C must be $1 - \frac{1}{3} = \frac{2}{3}$. We should totally switch cars!

Welcome to Bayes' theorem!

Derivation from conditional probabilities

We are about to do a bit of math. Relax and calm down. It will be ok, I promise!

From the definition of conditional probabilities:

$$p(x|y) = \frac{p(x, y)}{p(y)}. \quad (3)$$

Again, this is the probability that x and y happen together relative to the probability that y occurs at all. Now some algebra:

$$p(x|y)p(y) = p(x, y). \quad (4)$$

The same goes for $p(y|x)$:

$$p(y|x) = \frac{p(y, x)}{p(x)} \Rightarrow p(y|x)p(x) = p(y, x). \quad (5)$$

Because joint probabilities are communicative, $p(x, y) = p(y, x)$. Therefore

$$p(x|y)p(y) = p(y|x)p(x). \quad (6)$$

We are getting close! More algebra and

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}. \quad (7)$$

This is Bayes' theorem! More formally, we can express Bayes' theorem as

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}. \quad (8)$$

In words, we would say the *posterior* probability of the model (or model parameters θ) given the data D equals the probability of the data given the model (*likelihood*) times the *prior* probability of the model divided by the overall probability of the data.

The probability of the data $p(D)$ is the *marginal likelihood* or evidence for the model. It is the overall probability of the data averaged across all possible parameter values θ weighted by the prior probability of θ . When model parameters are discrete (rare for this class),

$$p(D) = \sum_{\theta} p(D|\theta)p(\theta); \quad (9)$$

when θ is continuous (most of the time),

$$p(D) = \int p(D|\theta)p(\theta)d(\theta). \quad (10)$$

Homework

In urban areas of Monte Negro Municipality, Western Amazon, Brazil, 4% of dogs are infected with *Rickettsia*. 80% of serological tests detect *Rickettsia* when present. The test's false positive rate is 17% (i.e., *Rickettsia* is detected but not present). A randomly sampled dog has tested positive for *Rickettsia*. What is the probability that the dog is indeed infected?