Lecture 1 Notes: What is Bayes?

January 18, 2017

1 Conceptual overview and differences between frequentist and Bayesian statistics

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 frequentist")?

- Parameters are point estimates. Uncertainty in these estimates are in light of unobserved future data, e.g., if we were to repeat the experiment many times, how often would we expect the true parameter to fall within a certain interval around our sample estimate (usually 95%)?

- In frequentist statistics, the model is considered fixed or "true" while the data are random. Analyses essentially assess the probability of the data given the "true" model.

Bayesian statistics:

In contrast, Bayesian inference is the re-allocation of credibility (probability) from prior knowledge to posterior knowledge in light of 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 themselves 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 explanatory variables given the scientific question of interest.

  2. Create a descriptive mathamatical model for the question and data at hand. The parameterization should result in biologically meaningful estimates and theoretically sound.

  3. Specify prior distributions for the 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 that the model is reasonable and does a good job (step 4), interpret the posterior with respect to the greater biological and theoretical questions of interest.

2 Probability recap:

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

- Three properties:

  1. probabilities are nonnegative (p $\geq$ 0)

  2. Sum of probs across all events in 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 from 4.1 in the puppy book (Kruschke, 2015). Sorry I couldn'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 hair or eye color overall, regardless of the other variable. They are computed by summing the probabilities of each row/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 or $p(x|y)$. We call these *conditional probabilities*. $p(x|y)$ can be interpeted as among all joint outcomes with value y, what proportion share value x.

    - For example, using the example table above, given someone from this population is a brunette, what is the probability they have green eyes? It's easy to see that this is just the joint probability of brunette and green eyes divided by the the marginal probability of being a brunette: $0.05/0.48 = 0.104$.

- Formally, we can define conditional probabilities as:

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

and when x is continuous,

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

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

3 Bayes theorem

Derivation from conditional probabilities

From the definition of conditional probabilities:

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

which is the probability that x and y happen together relative to the probability that y occurs at all. Some algebraic wankery:

$$p(x|y)p(y) = p(x, y). \tag{4}$$

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

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

Because joint probabilities are communicative, $p(x, y) = p(y, x)$ and therefore

$$p(x|y)p(y) = p(y|x)p(x). \tag{6}$$

More algebra and we get Bayes theorem:

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

More formally, we can express Bayes' theorem as

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

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 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 the model parameters are discrete (rare for this class),

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

when θ is continuous (most of the time),

$$p(D) = \int p(D|\theta)p(\theta)d(\theta). \tag{10}$$

Simple example:

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?

It makes it easy if we create a contingency table based off of what we know:

	Disease	
Test result	Rickettsia	No Rickettsia
T = +	0.80	0.17
T = -	0.20	0.83

Adding the prior in, we can get the joint and marginal probabilities:

	Disease		Marginal
Test result	Rickettsia	No Rickettsia	Marginal
Test+	0.032	0.1632	0.1952
Test-	0.008	0.7968	0.848
Marginal	0.04	0.96	1.0

The posterior probability of a dog being infected with *Rickettsia* is the likelihood of a positive test result when the disease is present (0.8) * the prior probability of the disease prevalence, 0.04 (this is cell 1,1 of our matrix). We then divide that by the marginal probability of a positive test result (row 1).

```
lik <- 0.8
prior <- 0.04
falsePos <- 0.17

probData  <- lik*prior + falsePos*(1 - prior)
Posterior <- (lik * prior)/probData
return(Posterior)

## [1] 0.1639344
```