Statistical Rethinking 2024

Week 1 lecture notes

2024-01-05

Table of contents

Difference between Bayesian and frequentist

The frequentist approach to inference:

$P(data|model)$

The Bayesian approach to inference:

$P(model|data)$

What are each of these expressions saying?

How would you summarize the different approaches narratively?

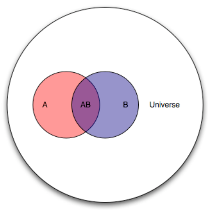
What estimation approach is used both by the frequentists and the Bayesians?

Bayes Rule

Let's derive the Bayes' Rule, drawing primarily from Oscar Bonilla.

Consider two overlapping events A and B occurring within a universe U.

```
include_graphics("venn-last.png")
```



Let's agree that the probability of something is how often it occurs. So in the Venn diagram above:

$P(A) = \frac{A}{U}$

$P(AB) = \frac{AB}{U}$

What if we shifted our question to ask, how much of A is in B? Our new universe is now B, and we're interested in the overlap of A and B as a proportion of B. In the language of probability, for "how much of A is in B" we say "what is the probability of A, given B?" In notation, we write that as $A|B$. This gives us:

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

Let's solve for $P(AB)$:

$P(AB) = P(A|B)P(B)$

Now let's ask, how much of B is in A? In probability language, what is the probability of B, given A?

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

Solve for $P(AB)$:

$P(AB) = P(B|A)P(A)$

The joint probability of A and B is the same whether you are asking about the proportion of A in B, or the proportion of B in A. Now as a simple algebraic trick, put the two identities together and solve for $A|B$:

$P(A|B)P(B) = P(B|A)P(A)$

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

And that's it. That's Bayes' Rule.

Bayes' Rule as an analytical tool

Now let's interpret Bayes' Rule in terms of a data analysis. We want to estimate the probability of a hypothesized cause, given its observed effect. The hypothesized cause is our model of what's happening. The observed effect is the set of data that we have collected.

$P(model|data) = \frac{P(data|model)P(model)}{P(data)}$

To express this narratively, we might say that we have some observed data and we want to estimate the probability of various causes of what we observed. That's $P(model|data)$. We look at our data and ask, how often would we see such data, if one of our hypothesized causes were true? That's $P(data|model)$, and is referred to as the *likelihood* of a proposed model, given the observed data. The likelihood of a proposed model given the observed data is scaled by how often that hypothesized cause occurs ($P(model)$). Note that $P(model)$ incorporates information that is outside of our statistical framework. We call $P(model)$ our *prior probability*. We multiply the likelihood of a model (given the observed data) by the overall incidence of that model ($P(model)$, our prior).

We then normalize the numerator of the expression to a probability. That normalizing constant is $P(data)$ and doesn't really have a substantive interpretation other than to say it's the average probability of your data which has the effect of converting your output into a probability.

Practice:

Consider the following table:

```
cp <- data.frame(x=c(1000,1500,2000),
                 y10=c(.802,.687, .456),
                 y20=c(.103, .172, .312),
                 y30=c(.034,.071, .124),
```

```
                         y40=c(.06,.071,.108),
                         p_x=c(.116, .099, .785))

flextable(cp)
```

x	y10	y20	y30	y40	p_x
1,000	0.802	0.103	0.034	0.060	0.116
1,500	0.687	0.172	0.071	0.071	0.099
2,000	0.456	0.312	0.124	0.108	0.785

The variable X takes on the values 1000, 1500, 2000

The variable Y takes on the values 10, 20, 30, 40

Of all elements with x=1000, 80 percent have a y value of 10. In probability notation:

$P(y = 10|x = 1000) = 80.2\%$

Note that this is a specific instantiation of $P(Y|X)$, where X and Y take on specific values. What expression did we derive for $P(Y|X)$?

Units with x=1000 make up 11.6 percent of the sample. $P(x = 1000) = 11.6\%$

Based on this information, what proportion of the total sample has an x value 1000 and y value of 10?

In probability language, this is the joint probability of X and Y, which is $P(XY)$ or, more formally, $P(X \cap Y)$. What expression did we derive for $P(XY)$?

Using the expression for $P(XY)$ and information in the table provided above, do the following calculations:

$P(x = 1000 \cap y = 40)$

$P(x = 2000 \cap y = 20)$

$P(x = 1500 \cap y = 30)$

Now consider another table:

```r
bp <- data.frame(x=c(1000,1500,2000, "p_y"),
                 "y:10"=c(.093, .068, .358, .519),
                 `y:20`=c(.012, .017, .245, .274),
                 `y:30`=c(.004, .007, .097, .108),
                 `y:40`=c(.007, .007, .085, .099),
                 `p_x`=c(.116, .099, .785, NA))
bp %>%
  flextable()
```

x	y.10	y.20	y.30	y.40	p_x
1000	0.093	0.012	0.004	0.007	0.116
1500	0.068	0.017	0.007	0.007	0.099
2000	0.358	0.245	0.097	0.085	0.785
p_y	0.519	0.274	0.108	0.099	

```r
sum(bp[1:3, 2:5])
```

```
[1] 1
```

```r
sum(bp[4,2:5])
```

```
[1] 1
```

```r
sum(bp[,6], na.rm=T)
```

```
[1] 1
```

The probability of x=1000 and y=20 is 1.2 percent. $P(y = 20 \cap x = 1000) = 1.2\%$

Among units with x=1000, what proportion have y=20? As mentioned above, this is a specific instantiation of $P(Y|X)$