Introducing Bayesian inference

Frank Edwards

Small group exercise: Define probability

In pairs, take five minutes to write a brief definition of what we mean when we say *probability*. Try not to consult the internet, tell me what you think it is!

Hint: there is more than one correct answer here!

Two definitions

- Frequentist: the probability of event A is the proportion of cases in which A occurs when an experiment is repeated many times (the long-run frequency)
- Bayesian: the probability of event A is our subjective assessment of whether we will observe A (degree of belief, support of evidence) in an experiment

Implications for science

- Frequentist: The truth is fixed. We can (kind of) approximate it with observed data by appealing to theories of repeated sampling and large number theorems

Implications for science

- Frequentist: The truth is fixed. We can (kind of) approximate it with observed data by appealing to theories of repeated sampling and large number theorems
- Bayesian: The truth is not necessarily fixed, we can describe our current knowledge by conditioning on both prior information and current data

Bayes' rule

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

Bayes' rule

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

Practice:

You take a test for a rare disease. The disease has a prevalence of 1 in 10000 in the population. The test accurately reports if a subject has the disease 99 percent of the time (true positive rate, 0.99). The test has a false positive rate of 1 percent.

- What is the probability that you have the disease if you test positive?
- What is the probability that you do not have the disease if you test negative?

Bayes' rule adapted for data analysis

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

Or in Bayesian vernacular:

$$Posterior \propto Likelihood \times Prior$$

Bayes' rule adapted for data analysis

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

Or in Bayesian vernacular:

$$Posterior \propto Likelihood \times Prior$$

The denominator doesn't depend on parameters, and is removed by taking integrals for most of our applications (it is a scaling factor to ensure results are within 0:1)

What proportion of marbles in the bag are blue?

- There's a bag of four marbles. Some are blue, some are red. For some reason, we can only draw one out at a time.
- We want to know θ: the proportion of marbles in the bag that are blue.
- As an experiment, we draw one marble, note its color, and put it back. Then we draw one more marble, note its color, and put it back. This is a sample of 2 with replacement.

```
# set the random seed to ensure we always get the same results
set.seed(1)
# mix up the bag
marble_bag <- sample(c("Red", "Blue"), size = 4, replace = T)
# Our experiment
our_sample <- sample(marble_bag, 2, replace = T)
```

Priors

The number of blue marbles could be any integer between 0 and 4 (5 possible values).

These counts correspond with proportions (θ) of 0, 0.25, 0.5, 0.75, and 1

We have no information here, so let's assume that all values of θ are equally likely as our *prior*.

Hypothesis	Prior: $P(\theta)$
$\theta = 0$	$\frac{1}{5}$
$\theta = 0.25$	$\frac{1}{5}$
$\theta = 0.5$	$\frac{1}{5}$
$\theta = 0.75$	$\frac{1}{5}$
$\theta = 1$	$\frac{1}{5}$

Observe the data!

Let's see what we got in our sample

```
our_sample
```

```
## [1] "Blue" "Red"
```

OK. So we find one blue, one red. Since we are treating 'blue' as a success in our calculation of θ, let's call this a 1 (out of possible values 0, 1, 2).

The likelihood

We'd like to know which value of θ is most likely to be correct given our observation.

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

The likelihood

We'd like to know which value of θ is most likely to be correct given our observation.

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

We've already established $P(\theta)$, our *priors*. Now we need to establish $P(data|\theta)$, the *likelihood* of a particular observation under different values of θ.

What outcomes are possible conditional on each parameter value?

For each hypothesized value of θ, let's describe the probability of observing what we actually observed (the data). We can use the binomial PDF for this.

```
# pr(1|theta = 0, 0.25, 0.5, 0.75, 1)
dbinom(1, size = 2, prob = c(0, 0.25, 0.5, 0.75, 1))
```

```
## [1] 0.000 0.375 0.500 0.375 0.000
```

Formalizing this as the likelihood

We will compute the likelihood of the data we *actually* observed (1) under each possible value of θ. We do this by counting the number of times our data could have occurred as a proportion of all possible occurrences.

Hypothesis	Prior $P(\theta)$	Likelihood $P(data\|\theta)$
$\theta = 0$	$\frac{1}{5}$	0
$\theta = 0.25$	$\frac{1}{5}$	0.375
$\theta = 0.5$	$\frac{1}{5}$	0.5
$\theta = 0.75$	$\frac{1}{5}$	0.375
$\theta = 1$	$\frac{1}{5}$	0

The denominator: the total probability of the data

How likely are we to observe what we did observe across all possible values of θ?

The denominator: the total probability of the data

How likely are we to observe what we did observe across all possible values of θ?

The *law of total probability* tells us that we can sum up joint probabilities to obtain a marginal probability

$$P(D) = \sum P(D, \theta) = \sum P(D|\theta)P(\theta)$$

The denominator: the total probability of the data

How likely are we to observe what we did observe across all possible values of θ?

The *law of total probability* tells us that we can sum up joint probabilities to obtain a marginal probability

$$P(D) = \sum P(D, \theta) = \sum P(D|\theta)P(\theta)$$

This is pretty easy here, but generally it is VERY difficult and involves tricky calculus.

Adding the denominator

The denominator for our Bayesian inference is $P(data)$, which we will compute by summing the products of the likelihood and the prior. How likely is our data under the sum of all values of θ?

```
sum(0.2 * c(0, 0.375, 0.5, 0.375, 0))
```

```
## [1] 0.25
```

Hypothesis	Prior $P(\theta)$	Likelihood $P(data\|\theta)$	$P(data)$
$\theta = 0$	$\frac{1}{5}$	0	0.25
$\theta = 0.25$	$\frac{1}{5}$	0.375	0.25
$\theta = 0.5$	$\frac{1}{5}$	0.5	0.25
$\theta = 0.75$	$\frac{1}{5}$	0.375	0.25
$\theta = 1$	$\frac{1}{5}$	0	0.25

Putting it all together

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

The portion we are interested in for inference is the *posterior probability* $P(\theta|data)$. That is the probability that θ takes on particular values *after* we observe the data.

H	$P(\theta)$	$P(data\|\theta)$	$P(data)$	$P(\theta\|data)$
$\theta = 0$	$\frac{1}{5}$	0	0.25	0
$\theta = 0.25$	$\frac{1}{5}$	0.375	0.25	0.3
$\theta = 0.5$	$\frac{1}{5}$	0.5	0.25	0.4
$\theta = 0.75$	$\frac{1}{5}$	0.375	0.25	0.3
$\theta = 1$	$\frac{1}{5}$	0	0.25	0

What did we learn?

Our *posterior* probabilities reflect the weighted average of our prior beliefs and the insights we've gained from the data.

$$P(\theta = 0|data) = 0$$

$$P(\theta = 0.25|data) = 0.3$$

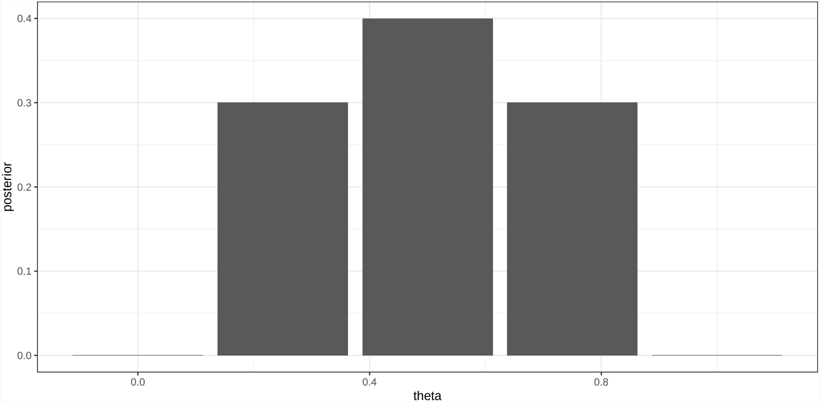
$$P(\theta = 0.5|data) = 0.4$$

$$P(\theta = 0.75|data) = 0.3$$

$$P(\theta = 0|data) = 0$$

The distribution of our posterior

Posterior distributions are probability distributions!



Now let's update!

Unlike frequentist analysis, we can *update* our beliefs about what we expect to observe. Let's fold our posteriors from the prior experiment in as *priors* for a new round of data collection.

H	$P(\theta)$
$\theta = 0$	0
$\theta = 0.25$	0.3
$\theta = 0.5$	0.4
$\theta = 0.75$	0.3
$\theta = 1$	0

Now let's update!

Unlike frequentist analysis, we can *update* our beliefs about what we expect to observe. Let's fold our posteriors from the prior experiment in as *priors* for a new round of data collection.

H	$P(\theta)$
$\theta = 0$	0
$\theta = 0.25$	0.3
$\theta = 0.5$	0.4
$\theta = 0.75$	0.3
$\theta = 1$	0

With these new priors in hand, let's draw a new sample

Sampling, updating the likelihood

```
our_sample <- sample(marble_bag, 2, replace = T)
our_sample
```

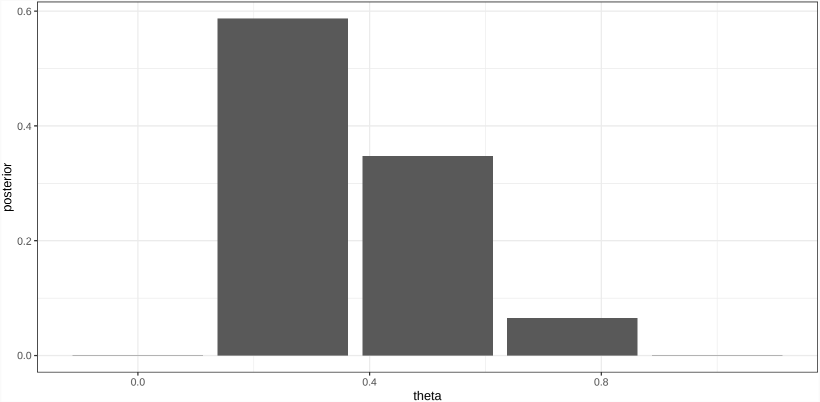
```
## [1] "Red" "Red"
```

We observed RR (0). Let's compute the probability of observing 0 under each value of θ

```
# let's make this easier with a data frame
round2 <- data.frame(theta = c(0, 0.25, 0.5, 0.75, 1), prior = c(0, 0.3, 0.4, 0.3,
    0))

round2 <- round2 %>%
    mutate(likelihood = dbinom(0, size = 2, prob = theta), p_d = sum(prior * likelihood),
        posterior = likelihood * prior/p_d)
```

Our updated findings

```r
ggplot(round2, aes(x = theta, y = posterior)) + geom_col()
```



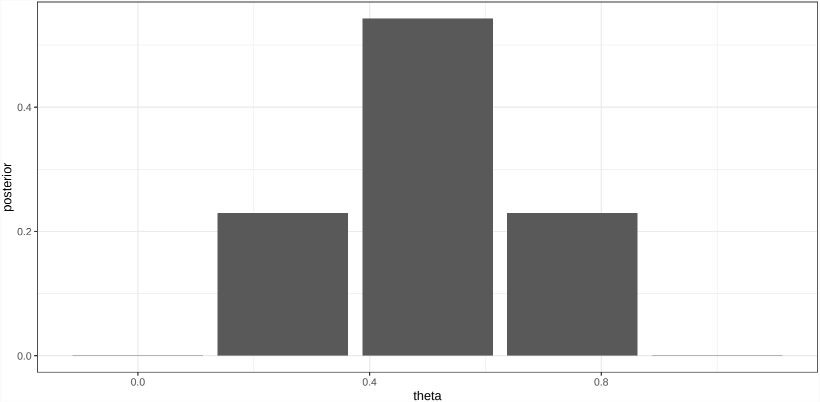
Can we go again?

```
new_sample <- sample(marble_bag, 2, replace = T)
n_blue <- sum(new_sample == "Blue")

round3 <- round2 %>%
    select(theta, posterior) %>%
    rename(prior = posterior) %>%
    mutate(likelihood = dbinom(n_blue, size = 2, prob = theta), p_d = sum(prior *
        likelihood), posterior = likelihood * prior/p_d)

round3
```

```
##   theta      prior likelihood       p_d posterior
## 1  0.00 0.00000000     0.0000 0.1603261 0.0000000
## 2  0.25 0.58695652     0.0625 0.1603261 0.2288136
## 3  0.50 0.34782609     0.2500 0.1603261 0.5423729
## 4  0.75 0.06521739     0.5625 0.1603261 0.2288136
## 5  1.00 0.00000000     1.0000 0.1603261 0.0000000
```

Our updated findings

```r
ggplot(round3, aes(x = theta, y = posterior)) + geom_col()
```



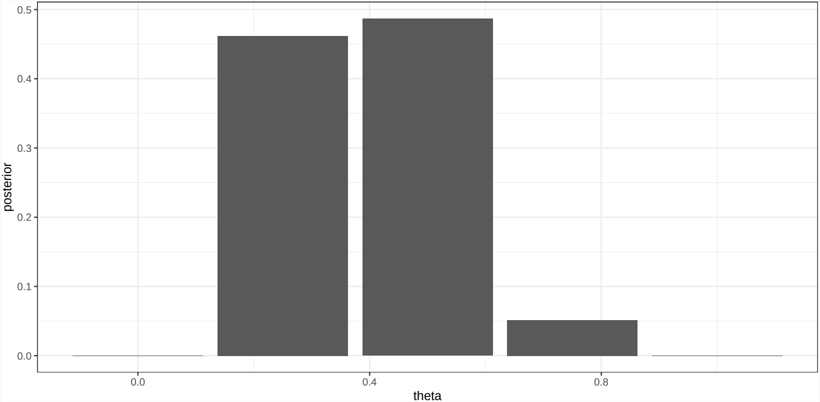
Ok, one more?!

```r
new_sample <- sample(marble_bag, 2, replace = T)
n_blue <- sum(new_sample == "Blue")

round4 <- round3 %>%
    select(theta, posterior) %>%
    rename(prior = posterior) %>%
    mutate(likelihood = dbinom(n_blue, size = 2, prob = theta), p_d = sum(prior *
        likelihood), posterior = likelihood * prior/p_d)

round4
```

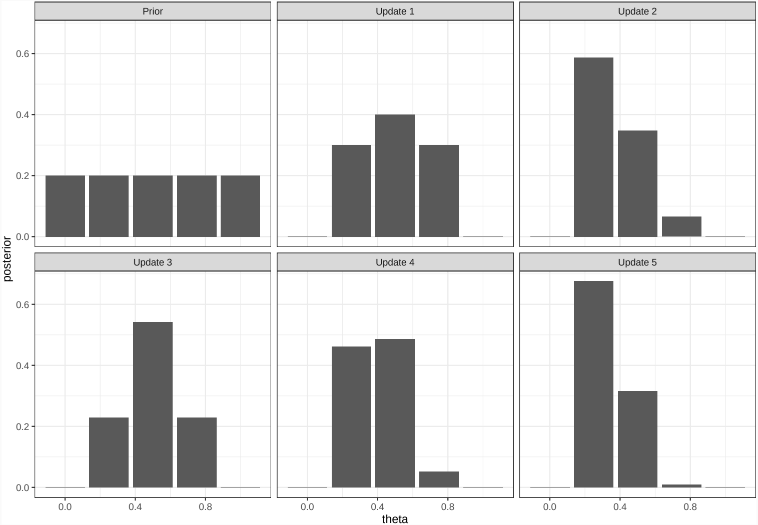
```
##   theta     prior likelihood       p_d posterior
## 1  0.00 0.0000000     1.0000 0.2786017 0.0000000
## 2  0.25 0.2288136     0.5625 0.2786017 0.4619772
## 3  0.50 0.5423729     0.2500 0.2786017 0.4866920
## 4  0.75 0.2288136     0.0625 0.2786017 0.0513308
## 5  1.00 0.0000000     0.0000 0.2786017 0.0000000
```

Our updated findings

```r
ggplot(round4, aes(x = theta, y = posterior)) + geom_col()
```



Let’s see what we learned



Now let's cheat and look in the bag

Our final posterior distribution for θ after 5 rounds was

```
##   theta   posterior
## 1  0.00 0.000000000
## 2  0.25 0.675416924
## 3  0.50 0.316244595
## 4  0.75 0.008338481
## 5  1.00 0.000000000
```

Now let's cheat and look in the bag

Our final posterior distribution for θ after 5 rounds was

```
##   theta   posterior
## 1  0.00 0.000000000
## 2  0.25 0.675416924
## 3  0.50 0.316244595
## 4  0.75 0.008338481
## 5  1.00 0.000000000
```

Here's the contents of the bag: Red, Blue, Red, Red, which means that $\theta = 0.25$

Now let's cheat and look in the bag

Our final posterior distribution for θ after 5 rounds was

```
##   theta   posterior
## 1  0.00 0.000000000
## 2  0.25 0.675416924
## 3  0.50 0.316244595
## 4  0.75 0.008338481
## 5  1.00 0.000000000
```

Here's the contents of the bag: Red, Blue, Red, Red, which means that $\theta = 0.25$

How did we do?

Nomenclature

- A hypothetical composition of the bag of marbles θ is a **parameter**, and is unknown

Nomenclature

- A hypothetical composition of the bag of marbles θ is a **parameter**, and is unknown
- The number of ways that a parameter could produce the data is a **likelihood**

Nomenclature

- A hypothetical composition of the bag of marbles θ is a **parameter**, and is unknown
- The number of ways that a parameter could produce the data is a **likelihood**
- The plausability of any value of θ before we conduct the experiment is a **prior probability**

Nomenclature

- A hypothetical composition of the bag of marbles θ is a **parameter**, and is unknown
- The number of ways that a parameter could produce the data is a **likelihood**
- The plausability of any value of θ before we conduct the experiment is a **prior probability**
- The new, updated plausability of any value of θ is a **posterior probability**