\[ P(\text{I'm near the ocean | I picked up a seashell}) = \]

\[
\frac{P(\text{I picked up a seashell | I'm near the ocean}) P(\text{I'm near the ocean})}{P(\text{I picked up a seashell})} 
\]

Statistically speaking, if you pick up a seashell and don't hold it to your ear, you can probably hear the ocean.
Probability
7\textsuperscript{th} grade classroom

\begin{itemize}
  \item Tall
  \item Short
\end{itemize}

\textbf{P(Tall) =} \frac{6}{12} = 0.5
Conditional Probability

7th Grade Classroom

We expect \( P(\text{Tall} \mid \text{Female}) > P(\text{Tall}) \) without taking any measurements of this particular class.

\[
P(\text{Tall} \mid \text{Female}) = \frac{4}{7}
\]

*Probability* that the student is tall *given that* the student is female (Conditional Probability)
Joint Probability

7th Grade Classroom

Probability that the student is tall and that the student is female (Joint Probability)

$$P(\text{Tall, Female}) = P(\text{Female}) \times P(\text{Tall | Female})$$

$$\frac{7}{12} \times \frac{4}{7} = \frac{4}{12} = \frac{1}{3}$$
Joint Probability

7th Grade Classroom

\[
P(\text{Tall}, \text{Female}) = P(\text{Female}) \cdot P(\text{Tall} \mid \text{Female}) = \frac{7}{12} \cdot \frac{4}{7} = \frac{4}{12} = \frac{1}{3}
\]

OR, equivalently

\[
P(\text{Female}, \text{Tall}) = P(\text{Tall}) \cdot P(\text{Female} \mid \text{Tall}) = \frac{6}{12} \cdot \frac{4}{6} = \frac{4}{12} = \frac{1}{3}
\]

\[
P(\text{Tall}, \text{Female}) = P(\text{Female}, \text{Tall})
\]
Deriving Bayes’ Rule

We have shown that:

\[
P(\text{Tall, Female}) = P(\text{Female}) \cdot P(\text{Tall} | \text{Female})
\]
\[
P(\text{Tall, Female}) = P(\text{Tall}) \cdot P(\text{Female} | \text{Tall})
\]

Therefore:

\[
P(\text{Female}) \cdot P(\text{Tall} | \text{Female}) = P(\text{Tall}) \cdot P(\text{Female} | \text{Tall})
\]

\[
P(\text{Tall} | \text{Female}) = \frac{P(\text{Female} | \text{Tall}) \cdot P(\text{Tall})}{P(\text{Female})}
\]

Or generally, for generic events \( A \) & \( B \), we have

\[
P(\text{A} | \text{B}) = \frac{P(\text{B} | \text{A}) \cdot P(\text{A})}{P(\text{B})}
\]
Bayes’ Rule: Terminology

\[
P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)}
\]
Applying Bayes’ Rule

Information:
• 1% of women in a given population have breast cancer
• If a woman has breast cancer, there is a 90% chance that a particular diagnostic test will return a positive result (10% false negative rate)
• If a woman does not have breast cancer, there is a 10% chance that this diagnostic test will return a positive result (10% false positive rate).

Question:
What is the probability that a woman with a positive test result actually has cancer?
Multiple Choice:

Which notation shows the probability that a woman with a positive test result actually has cancer?

a.) \( P(\text{Cancer} \mid \text{Positive Test}) \)

b.) \( P(\text{Cancer} \cap \text{Positive Test}) \)

c.) \( P(\text{Positive Test} \mid \text{Cancer}) \)

d.) \( P(\text{Positive Test} \cap \text{Cancer}) \)
Applying Bayes’ Rule

- Information:
  - 1% of women in a given population have breast cancer
  - If a woman has breast cancer, there is a 90% chance that a particular diagnostic test will return a positive result (10% false negative rate)
  - If a woman does not have breast cancer, there is a 10% chance that this diagnostic test will return a positive result (10% false positive rate).

\[
P(\text{Cancer} | \text{Positive}) = \frac{P(\text{Positive} | \text{Cancer}) \cdot P(\text{Cancer})}{P(\text{Positive})}
\]
Applying Bayes’ Rule

• Information:
  • 1% of women in a given population have breast cancer
  • If a woman has breast cancer, there is a 90% chance that a particular diagnostic test will return a positive result (10% false negative rate)
  • If a woman does not have breast cancer, there is a 10% chance that this diagnostic test will return a positive result (10% false positive rate).

\[
P(\text{Cancer} \mid \text{Positive}) = \frac{P(\text{Positive} \mid \text{Cancer}) \cdot P(\text{Cancer})}{P(\text{Positive})}
\]

0.01
Applying Bayes’ Rule

• Information:
  • 1% of women in a given population have breast cancer
  • If a woman has breast cancer, there is a 90% chance that a particular diagnostic test will return a positive result (10% false negative rate)
  • If a woman does not have breast cancer, there is a 10% chance that this diagnostic test will return a positive result (10% false positive rate).

\[
P(\text{Cancer} \mid \text{Positive}) = \frac{P(\text{Positive} \mid \text{Cancer}) \cdot P(\text{Cancer})}{P(\text{Positive})}
\]
Applying Bayes’ Rule

**Information:**
- 1% of women in a given population have breast cancer.
- If a woman has breast cancer, there is a 90% chance that a particular diagnostic test will return a positive result (10% false negative rate).
- If a woman does not have breast cancer, there is a 10% chance that this diagnostic test will return a positive result (10% false positive rate).

\[
P(\text{Cancer} | \text{Positive}) = \frac{P(\text{Positive} | \text{Cancer}) \cdot P(\text{Cancer})}{P(\text{Positive})}
\]

\[
0.9 \quad 0.01
\]

\[
\frac{0.9 \cdot 0.01}{P(\text{Positive})}
\]
Applying Bayes’ Rule

**Information:**
- 1% of women in a given population have breast cancer
- If a woman has breast cancer, there is a 90% chance that a particular diagnostic test will return a positive result (10% false negative rate)
- If a woman does not have breast cancer, there is a 10% chance that this diagnostic test will return a positive result (10% false positive rate).

\[
P(\text{Positive}) = P(\text{True Positive}) + P(\text{False Positive})\]
\[
P(\text{Positive}) = P(\text{Positive} | \text{Cancer}) \cdot P(\text{Cancer}) + P(\text{Positive} | \text{Healthy}) \cdot P(\text{Healthy})\]
\[
P(\text{Positive}) = 0.9 \cdot 0.01 + 0.1 \cdot (1 - 0.01)\]
\[
P(\text{Positive}) = 0.108\]
Now we can complete Bayes’ Rule

\[
P(\text{Cancer} \mid \text{Positive}) = \frac{P(\text{Positive} \mid \text{Cancer}) \cdot P(\text{Cancer})}{P(\text{Positive})}
\]

\[
P(\text{Cancer} \mid \text{Positive}) = \frac{0.9 \cdot 0.01}{0.108} = 0.083
\]
How can we apply Bayes’ rule to estimating model parameters?
Frequentist Coin Flip: 20 Flips; 13 Heads

Objective: Estimate the Coin’s Bias with a 95% Confidence Interval

```
binom.test(13, 20)
```

```#
## Exact binomial test
##
## data:  13 and 20
## number of successes = 13, number of trials = 20, p-value =
##  0.2632
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.4078115 0.8460908
## sample estimates:
##  probability of success
##        0.65
```

Conclusion:
- Bias = 0.65
- 95% CI = (0.41, 0.85)
We know how to compute the probability of any particular data outcome

```r
dbinom(13, size = 20, prob = 0.5)
## [1] 0.07392883

dbinom(13, size = 20, prob = 0.25)
## [1] 0.0001541923
```
Computing the probability of getting the data that we observed at various values of the coin’s bias

```r
coin.bias <- seq(from = 0, to = 1, by = 0.01)
likelihood <- dbinom(13, 20, prob = coin.bias)
barplot(likelihood, names.arg = coin.bias, ylab = "likelihood")
```
Imagine we have a pool of 101 coins each with a different bias (0.00, 0.001, 0.002,...)

We can calculate the probability of each of the 101 coins being the one that we chose, given the data that we observed.

Probability of having chosen the fair coin:

\[
P(M_{0.50}|D_{13}) = \frac{P(D_{13}|M_{0.50}) \cdot P(M_{0.50})}{P(D_{13})}
\]
Imagine we have a pool of 101 coins each with a different bias (0.00, 0.001, 0.002,...)

We can calculate the probability of each of the 101 coins being the one that we chose, given the data that we observed.

Probability of having chosen the fair coin:

\[
P(M_{0.50} | D_{13}) = \frac{P(D_{13} | M_{0.50}) \cdot P(M_{0.50})}{P(D_{13})}
\]

\[
\frac{1}{1001} \approx 0.0099
\]
Bayesian Coin Flip: 20 Flips; 13 Heads

Objective: Identify the bias (x) that yields the highest posterior probability. Given 13 heads were observed out of 20 flips

\[
P(bias = x | 13 \text{ heads}) = \frac{P(13 \text{ heads} | bias = x) \cdot P(bias = x)}{P(13 \text{ heads})}
\]
Bayesian Coin Flip: Define Priors

Prior for each potential bias = 1/101 ≅ 0.0099

```
prior.probability <- numeric(101)
prior.probability[0:101] <- 1
# Normalize; since it is a PDF, sum must be 1.0
prior.probability <- prior.probability / (sum(prior.probability))
barplot(prior.probability, names.arg = coin.bias, xlab = "Coin Bias (x)", ylab = "P(bias = x)", ylim = c(0,0.02), main = "Prior Probability Density Function: Biases Equally Likely", col = "#85C0F9")
```
Marginal Likelihood

\[ P(13 \text{ heads}) \]
\[ = P(13 \text{ heads} | \text{ bias} = 0.00) \cdot P(\text{ bias} = 0.00) \]
\[ + P(13 \text{ heads} | \text{ bias} = 0.01) \cdot P(\text{ bias} = 0.01) \]
\[ + P(13 \text{ heads} | \text{ bias} = 0.02) \cdot P(\text{ bias} = 0.02) \]
\[ \ldots \]
\[ + P(13 \text{ heads} | \text{ bias} = 0.50) \cdot P(\text{ bias} = 0.50) \]
\[ \ldots \]
\[ + P(13 \text{ heads} | \text{ bias} = 0.99) \cdot P(\text{ bias} = 0.99) \]
\[ + P(13 \text{ heads} | \text{ bias} = 1.00) \cdot P(\text{ bias} = 1.00) \]
Marginal Likelihood

\[ P(13 \text{ heads}) = P(13 \text{ heads} | \text{bias} = 0.00) \cdot P(\text{bias} = 0.00) + P(13 \text{ heads} | \text{bias} = 0.01) \cdot P(\text{bias} = 0.01) + P(13 \text{ heads} | \text{bias} = 0.02) \cdot P(\text{bias} = 0.02) + \ldots + P(13 \text{ heads} | \text{bias} = 0.50) \cdot P(\text{bias} = 0.50) + \ldots + P(13 \text{ heads} | \text{bias} = 0.99) \cdot P(\text{bias} = 0.99) + P(13 \text{ heads} | \text{bias} = 1.00) \cdot P(\text{bias} = 1.00) \]
Marginal Likelihood

\[ P(13 \text{ heads}) = P(13 \text{ heads} \mid \text{bias} = 0.00) \cdot 0.0099 + P(13 \text{ heads} \mid \text{bias} = 0.01) \cdot 0.0099 + P(13 \text{ heads} \mid \text{bias} = 0.02) \cdot 0.0099 + \cdots + P(13 \text{ heads} \mid \text{bias} = 0.50) \cdot 0.0099 + \cdots + P(13 \text{ heads} \mid \text{bias} = 0.99) \cdot 0.0099 + P(13 \text{ heads} \mid \text{bias} = 1.00) \cdot 0.0099 \]
Marginal Likelihood

\[ P(13 \text{ heads}) = P(13 \text{ heads} | \text{ bias} = 0.00) \cdot 0.0099 + P(13 \text{ heads} | \text{ bias} = 0.01) \cdot 0.0099 + P(13 \text{ heads} | \text{ bias} = 0.02) \cdot 0.0099 + \ldots + P(13 \text{ heads} | \text{ bias} = 0.50) \cdot 0.0099 + \ldots + P(13 \text{ heads} | \text{ bias} = 0.99) \cdot 0.0099 + P(13 \text{ heads} | \text{ bias} = 1.00) \cdot 0.0099 \]
Marginal Likelihood

\[ P(13 \text{ heads}) = 0.0 \cdot 0.0099 + 7.2 \cdot 10^{-22} \cdot 0.0099 + 5.5 \cdot 10^{-18} \cdot 0.0099 + \ldots + 0.07392883 \cdot 0.0099 + \ldots + 6.8 \cdot 10^{-10} \cdot 0.0099 + 0.0 \cdot 0.0099 \]

= 0.04714757
Marginal Likelihood

\[
\text{coin.bias} \leftarrow \text{seq}(\text{from} = 0, \text{to} = 1, \text{by} = 0.01)
\]

\[
\text{p.d13} \leftarrow \text{sum(dbinom(13, 20, \text{coin.bias}) * (1 / 101))}
\]

\[
\# [1] 0.04714757
\]

\[
= 0.04714757
\]
Bayesian Coin Flip: Likelihood

```r
coin.bias <- seq(from = 0, to = 1, by = 0.01)
likelihood <- dbinom(13, 20, prob = coin.bias)
barplot(likelihood, names.arg = coin.bias, ylab = "Likelihood: P(13 Heads | bias = x)", xlab = "Coin Bias (x)", col = "#A95AA1") #Color-blindess friendly purple
```
Posterior Probability

Recall Frequentist Conclusion:
- Bias = 0.65
- 95% CI = (0.41, 0.85)
Summary: Flipping a Coin with No expectations of fairness

\[ P(\text{bias} = x \mid 13 \text{ heads}) = \frac{P(13 \text{ heads} \mid \text{bias} = x) \cdot P(\text{bias} = x)}{P(13 \text{ heads})} \]

\[ P(13 \text{ Heads}) = 0.04714757 \]
Summary of Bayes’ method

Prior
probability distribution

Observe Data
Re-evaluate model / prior distribution

Posterior
probability distribution

Data: 13 heads in 20 coin flips
What if I assume there is a good chance of the coin having a certain “bias”?

Prior probability distribution

Observe Data
Re-evaluate model / prior distribution

Data: 13 heads in 20 coin flips

Posterior probability distribution

?
Our prior will reflect our assumption that our friend is honest

```r
prior.probability <- numeric(101)
prior.probability[0:101] <- 1
prior.probability[48:54] <- 3
prior.probability[50:52] <- 5
prior.probability[51] <- 7
# Normalize; since it is a PDF, sum must be 1.0
prior.probability <- prior.probability / (sum(prior.probability))
barplot(prior.probability, names.arg = coin.bias)
```
Our posterior probability distribution reflects a complex interplay between the prior and the data

```r
posterior.probability <-
    dbinom(13, 20, coin.bias) * prior.probability / p.d13
barplot(posterior.probability, names.arg = coin.bias)
```
What if I assume there is a good chance of the coin having a certain “bias”?

Prior
probability distribution

Observe Data
Re-evaluate model / prior distribution

Posterior
probability distribution

Data: 13 heads in 20 coin flips
What if we collect more data?

Prior probability distribution

Observe Data
Re-evaluate model / prior distribution

Posterior probability distribution

Data: 130 heads in 200 coin flips

?
Now in our **posterior** probability, the data “overwrites” our prior

```r
p.d130 <- sum(dbinom(130, 200, coin.bias) * prior.probability)
posterior.probability <-
    dbinom(130, 200, coin.bias) * prior.probability / p.d130
barplot(posterior.probability, names.arg = coin.bias)
```
When we collect more data, the data “overwrites” our prior

Prior
probability distribution

Observe Data
Re-evaluate model / prior distribution

Posterior
probability distribution

Data: 130 heads in 200 coin flips
Another example – imagine there is a magic shop around the corner...

```r
prior.probability <- numeric(101)
prior.probability[0:101] <- 1
prior.probability[48:54] <- 3
prior.probability[73:78] <- 3

# Normalize; since it is a PDF, sum must be 1.0
prior.probability <- prior.probability / (sum(prior.probability))
barplot(prior.probability, names.arg = coin.bias)
```
If we assume we are talking to a swindler, our posterior will reflect that

```r
(p.d6 <- sum(dbinom(6, 10, coin.bias) * prior.probability))
## [1] 0.1083962
posterior.probability <- dbinom(6, 10, coin.bias) * prior.probability / p.d6
barplot(posterior.probability, names.arg = coin.bias)
```
If there is a magic shop around the corner, we conclude the coin may be biased.

Prior
- probability distribution

Observe Data
- Re-evaluate model / prior distribution

Posterior
- probability distribution

Data: 6 heads in 10 coin flips
Conclusion

Bayesian statistics
• Start with our understanding of how something works/what is likely to happen
• We then update our belief based on our data
• Possible to perform multiple rounds of formulation of prior, evaluation of prior based on data and formulation of posterior.
• Does not rely on the notion of a finding “as or more inconsistent with our $H_0$”

Frequentist approaches
• Do not assign probabilities to a hypothesis (no prior, posterior)
• Usually less computationally intensive
• Lower risk of bias
References

• Banfelder, J. Quantitative Understanding in Biology 1.7 Bayesian Methods (https://physiology.med.cornell.edu/people/banfelder/qbio/lecture_notes/1.7_bayesian.pdf)


Further interesting materials on this topic:
• Kruschke, J. “Doing Bayesian Data Analysis”
• https://boyangzhao.github.io/posts/vaccine_efficacy_bayesian (advanced blog post about how Bayesian statistics were used to determine COVID-19 vaccine efficacy)
• https://youtu.be/9TDjifpGj-k (fun crash course on the basics of Bayesian statistics)
DID THE SUN JUST EXPLODE?
(IT'S NIGHT, SO WE'RE NOT SURE.)

This neutrino detector measures whether the Sun has gone nova.
Then, it rolls two dice. If they both come up six, it lies to us. Otherwise, it tells the truth.

Let's try.
Detector! Has the Sun gone nova?

ROLL

YES.

FREQUENTIST STATISTICIAN:

The probability of this result happening by chance is \( \frac{1}{36} = 0.027 \).
Since \( p < 0.05 \), I conclude that the Sun has exploded.

BAYESIAN STATISTICIAN:

BET YOU $50 IT HASN'T.