Lecture 29: Bayesian Statistics

Chapter 2.2.7

Recall from Lecture 3.3: Conditional Probability

The conditional probability of an event A given the event B, is defined by

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

This is read as "the probability of A given B" or "the probability of A conditional on B."

Back to Midterm: New Notation

Two possible outcomes for hypothesis test:

- "reject H_0 in favor of H_A" $= \oplus$'ve result
- "do not reject H_0" $= \ominus$'ve result.

Back to Midterm: New Notation

Two possible outcomes for hypothesis test:

- "reject H_0 in favor of H_A" $= \oplus$'ve result
- "do not reject H_0" $= \ominus$'ve result.

with performance measures:

- $\alpha = 0.05 = \text{Pr}(\text{Reject } H_0 \text{ when } H_0 \text{ true}) = \text{Pr}(\oplus | H_0)$
- Power
  $= 1 - \beta = 0.8 = \text{Pr}(\text{Reject } H_0 \text{ when } H_A \text{ true}) = \text{Pr}(\oplus | H_A)$

Back to Midterm

Say H_A is true 10% of the time.

So

- $\Pr(H_A) = 0.1$
- $\Pr(H_0) = 1 - \Pr(H_A) = 1 - 0.1 = 0.9$

Back to Midterm

Say H_A is true 10% of the time.

So

- $\Pr(H_A) = 0.1$
- $\Pr(H_0) = 1 - \Pr(H_A) = 1 - 0.1 = 0.9$

We conduct 1000 hypotheses of H_0 vs H_A, so

- H_A is true 100 times
- H_0 is true 900 times

Back to Midterm

So recall from the midterm we have the following 2×2 table of possible outcomes:

		Test conclusion	
		$\ominus$	$\oplus$
Truth	H_0 true	$(1 - 0.05) \times 900 = 855$	$0.05 \times 900 = 45$
	H_A true	$(1 - 0.8) \times 100 = 20$	$0.8 \times 100 = 80$

Back to Midterm

So recall from the midterm we have the following 2×2 table of possible outcomes:

		Test conclusion	
		$\ominus$	$\oplus$
Truth	H_0 true	$(1-0.05) \times 900 = 855$	$0.05 \times 900 = 45$
	H_A true	$(1-0.8) \times 100 = 20$	$0.8 \times 100 = 80$

- Of the $\oplus$'s, what prop'n was right?
  i.e. What is $\Pr(H_A|\oplus)$? $\frac{80}{80+45} = 64\%$?

Back to Midterm

So recall from the midterm we have the following 2×2 table of possible outcomes:

		Test conclusion	
		$\ominus$	$\oplus$
Truth	H_0 true	$(1-0.05) \times 900 = 855$	$0.05 \times 900 = 45$
	H_A true	$(1-0.8) \times 100 = 20$	$0.8 \times 100 = 80$

- Of the $\oplus$'s, what prop'n was right?
  i.e. What is $\Pr(H_A|\oplus)$? $\frac{80}{80+45} = 64\%$?
- Of the $\ominus$'s, what prop'n was right?
  i.e. What is $\Pr(H_0|\ominus)$? $\frac{855}{855+20} = 97.7\%$

Different Set-Up

Now say for the same machine H_A is true 40% of the time. i.e. $P(H_A) = 0.4$

Different Set-Up

Now say for the same machine H_A is true 40% of the time. i.e. $P(H_A) = 0.4$

		Test conclusion	
		$\ominus$	$\oplus$
Truth	H_0 true	$(1 - 0.05) \times 600 = 570$	$0.05 \times 600 = 30$
	H_A true	$(1 - 0.8) \times 400 = 80$	$0.8 \times 400 = 320$

Different Set-Up

Now say for the same machine H_A is true 40% of the time. i.e. $P(H_A) = 0.4$

		Test conclusion	
		$\ominus$	$\oplus$
Truth	H_0 true	$(1 - 0.05) \times 600 = 570$	$0.05 \times 600 = 30$
	H_A true	$(1 - 0.8) \times 400 = 80$	$0.8 \times 400 = 320$

- Of the $\oplus$'s, what prop'n was right?
  $\Pr(H_A|\oplus) = \frac{320}{320+30} = 91.4\%$?

Different Set-Up

Now say for the same machine H_A is true 40% of the time. i.e. $P(H_A) = 0.4$

		Test conclusion	
		$\ominus$	$\oplus$
Truth	H_0 true	$(1 - 0.05) \times 600 = 570$	$0.05 \times 600 = 30$
	H_A true	$(1 - 0.8) \times 400 = 80$	$0.8 \times 400 = 320$

- Of the $\oplus$'s, what prop'n was right?
  $\Pr(H_A|\oplus) = \frac{320}{320+30} = 91.4\%$?
- Of the $\ominus$'s, what prop'n was right?
  $\Pr(H_0|\ominus) = \frac{570}{570+80} = 87.7\%$

How Reliable Are Your Test Results?

For the exact same hypothesis testing machine we get

	$\Pr(H_A \mid \oplus)$	$\Pr(H_0 \mid \ominus)$
$P(H_A) = 10\%$	64%	97.7%
$P(H_A) = 40\%$	91.4%	87.7%

How Reliable Are Your Test Results?

For the exact same hypothesis testing machine we get

	$\Pr(H_A \mid \oplus)$	$\Pr(H_0 \mid \ominus)$
$P(H_A) = 10\%$	64%	97.7%
$P(H_A) = 40\%$	91.4%	87.7%

The probability that a positive result is right depends on how likely H_A is. Same goes for negative results.

How Reliable Are Your Test Results?

For the exact same hypothesis testing machine we get

	$\Pr(H_A \mid \oplus)$	$\Pr(H_0 \mid \ominus)$
$P(H_A) = 10\%$	64%	97.7%
$P(H_A) = 40\%$	91.4%	87.7%

The probability that a positive result is right depends on how likely H_A is. Same goes for negative results.

As quoted in the Economist article: *By and large, scientists want surprising results, and so they test hypotheses that are normally pretty unlikely and often very unlikely.*

Bayes Theorem

This brings us to Bayes Theorem:

Bayes Theorem

This brings us to Bayes Theorem:

- Let $A_1, \ldots, A_k$ be k events that are a partition of the sample space S.
- Let B be an event of interest

Bayes Theorem

This brings us to Bayes Theorem:

- Let $A_1, \ldots, A_k$ be k events that are a partition of the sample space S.
- Let B be an event of interest

The Bayes Theorem states:

$$\Pr(A_i|B) = \frac{\Pr(B|A_i) \times \Pr(A_i)}{\sum_{j=1}^{k} \Pr(B|A_j) \times \Pr(A_j)}$$

Bayes Theorem

This brings us to Bayes Theorem:

- Let $A_1, \ldots, A_k$ be k events that are a partition of the sample space S.
- Let B be an event of interest

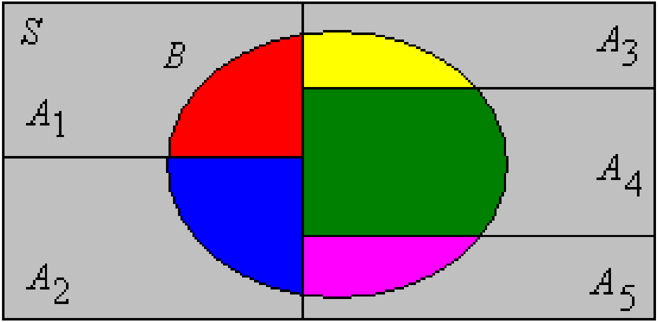
The Bayes Theorem states:

$$\Pr(A_i|B) = \frac{\Pr(B|A_i) \times \Pr(A_i)}{\sum_{j=1}^{k} \Pr(B|A_j) \times \Pr(A_j)}$$

Notice the flip between A_i and B.

Illustration

- The sample sample S is the overall grey box
- $A_1, \ldots, A_5$ are the five blocks that partition S.
- B is the oval



Tailored to our Situation

- The sample sample S is the 2×2 possible outcomes
- H_0 and H_A partition S. i.e. $k = 2$
- Let B be a $\oplus$ result

Tailored to our Situation

- The sample sample S is the 2×2 possible outcomes
- H_0 and H_A partition S. i.e. $k = 2$
- Let B be a $\oplus$ result

Then by Bayes Theorem, the probability that a $\oplus$ result is right is

$$\Pr(H_A|\oplus) = \frac{\Pr(\oplus|H_A)\Pr(H_A)}{\Pr(\oplus|H_A)\Pr(H_A) + \Pr(\oplus|H_0)\Pr(H_0)}$$

Tailored to our Situation

- The sample sample S is the 2×2 possible outcomes
- H_0 and H_A partition S. i.e. $k = 2$
- Let B be a $\oplus$ result

Then by Bayes Theorem, the probability that a $\oplus$ result is right is

$$\begin{aligned} \Pr(H_A|\oplus) &= \frac{\Pr(\oplus|H_A)\Pr(H_A)}{\Pr(\oplus|H_A)\Pr(H_A) + \Pr(\oplus|H_0)\Pr(H_0)} \\ &= \frac{(1-\beta) \times \Pr(H_A)}{(1-\beta) \times \Pr(H_A) + \alpha \times \Pr(H_0)} \end{aligned}$$

Notions of both type I error rate and power (AKA type II error rate) are included!

Tailored to our Situation

Back to initial example where $\alpha = 0.05$, $1 - \beta = 0.8$, $\Pr(H_A) = 0.10$

Tailored to our Situation

Back to initial example where $\alpha = 0.05$, $1 - \beta = 0.8$, $\Pr(H_A) = 0.10$

$$\Pr(H_A|\oplus) = \frac{0.8 \times 0.1}{0.8 \times 0.1 + 0.05 \times 0.9} = 0.64$$

Tailored to our Situation

Back to initial example where $\alpha = 0.05$, $1 - \beta = 0.8$, $\Pr(H_A) = 0.10$

$$\Pr(H_A|\oplus) = \frac{0.8 \times 0.1}{0.8 \times 0.1 + 0.05 \times 0.9} = 0.64$$

Similarly

$$\begin{aligned}
\Pr(H_0|\ominus) &= \frac{\Pr(\ominus|H_0)\Pr(H_0)}{\Pr(\ominus|H_A)\Pr(H_A) + \Pr(\ominus|H_0)\Pr(H_0)} \\
&= \frac{(1-\alpha) \times \Pr(H_0)}{\beta \times \Pr(H_A) + (1-\alpha) \times \Pr(H_0)} \\
&= \frac{0.95 \times 0.9}{0.2 \times 0.1 + 0.95 \times 0.9} = 0.977
\end{aligned}$$

The Debate

In the midterm you applied Bayes Theorem from scratch/intuition.

The Debate

In the midterm you applied Bayes Theorem from scratch/intuition.

Why isn't everybody taking into account $P(H_A)$ when testing H_0 vs H_A?

The Debate

In the midterm you applied Bayes Theorem from scratch/intuition.

Why isn't everybody taking into account $P(H_A)$ when testing H_0 vs H_A?

In this example, we assumed we knew the true $P(H_A)$. In real life however, we don't.

Statistics In General

Statistics is inferring about some unknown parameter θ.

Statistics In General

Statistics is inferring about some unknown parameter θ.

- Frequentist Statistics: the true θ is a single value that if we had an infinite sample size, we can compute it exactly.

Statistics In General

Statistics is inferring about some unknown parameter θ.

- Frequentist Statistics: the true θ is a single value that if we had an infinite sample size, we can compute it exactly.
- Bayesian Statistics: the true θ is a distribution of values that reflects our belief in the plausibility of different values.

Specific Example

Concrete example: the Probability of Flipping Heads

Specific Example

Concrete example: the Probability of Flipping Heads

- Frequentist Statistics: the true probability p is a single value

Specific Example

Concrete example: the Probability of Flipping Heads

- Frequentist Statistics: the true probability p is a single value
- Bayesian probability: the true probability p is a distribution of values

The Bayesian Procedure

To express our belief about θ from as a Bayesian, we have:

The Bayesian Procedure

To express our belief about θ from as a Bayesian, we have:

1. A prior distribution $\Pr(\theta)$. It reflects our prior belief about θ.

The Bayesian Procedure

To express our belief about θ from as a Bayesian, we have:

1. A prior distribution $\Pr(\theta)$. It reflects our prior belief about θ.
2. The likelihood function $\Pr(X|\theta)$. This is the mechanism that generates the data.

The Bayesian Procedure

To express our belief about θ from as a Bayesian, we have:

1. A prior distribution $\Pr(\theta)$. It reflects our prior belief about θ.
2. The likelihood function $\Pr(X|\theta)$. This is the mechanism that generates the data.
3. A posterior distribution $\Pr(\theta|X)$. We update our belief about θ after observing data X.

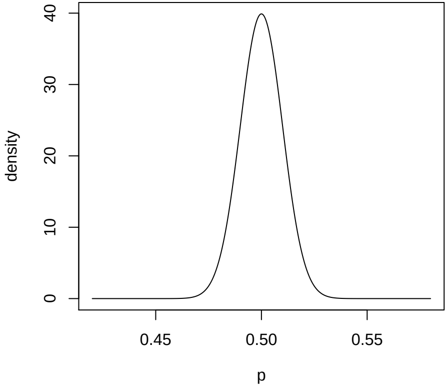
$$\Pr(\theta|X) = \frac{\Pr(X|\theta) \cdot \Pr(\theta)}{\Pr(X)}$$

The Issue: The Bayesian Procedure

Where do you come up with $\Pr(\theta)$? It's completely subjective! You decide!

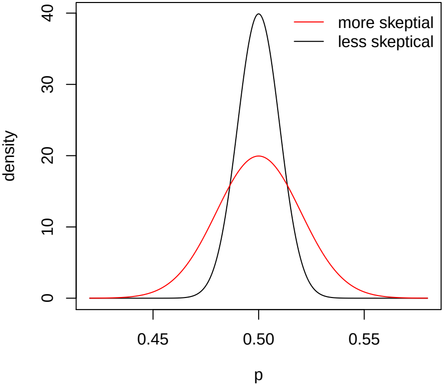
Prior Distribution

This distribution can reflect someone's prior belief of p.



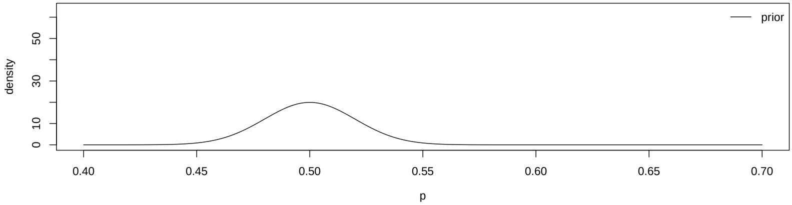
Prior Distribution

Say someone is more skeptical that $p = 0.5$, we can lower it.



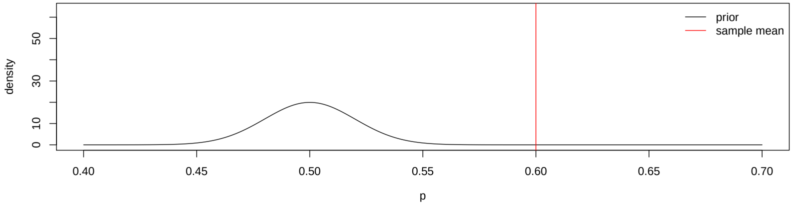
The Bayesian Procedure

Say we have the following prior belief centered at $p = 0.5$



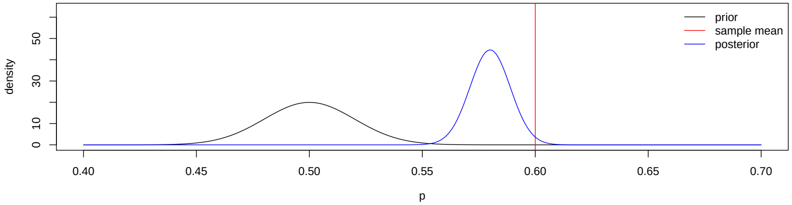
The Bayesian Procedure

Say we collect data, represented by the red line, suggesting $p = 0.6$



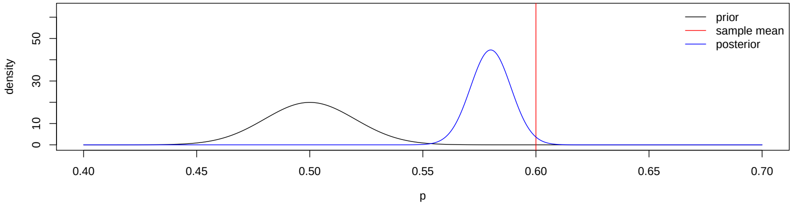
The Bayesian Procedure

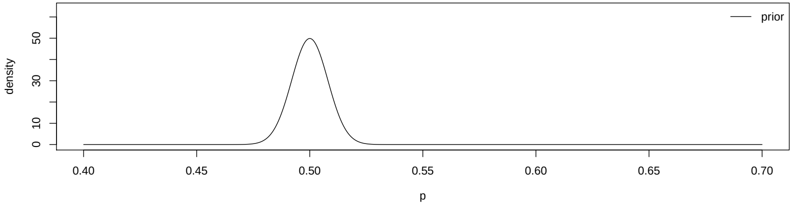
We then update our belief, as reflected in the posterior distribution!



The Bayesian Procedure

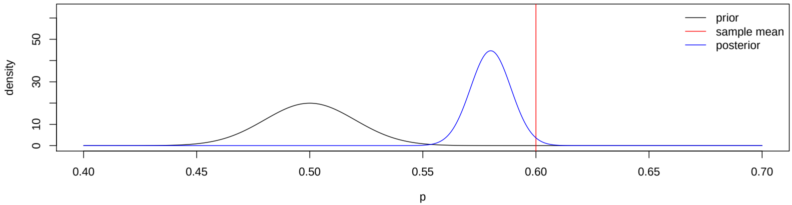
Now say we have a stronger prior belief that $p = 0.5$

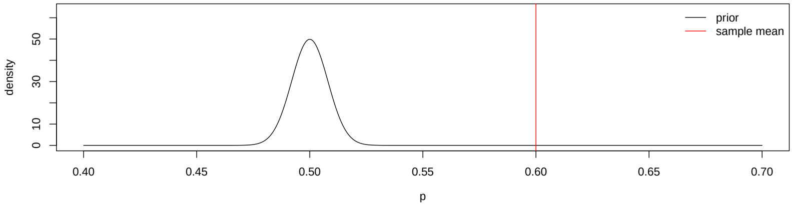




The Bayesian Procedure

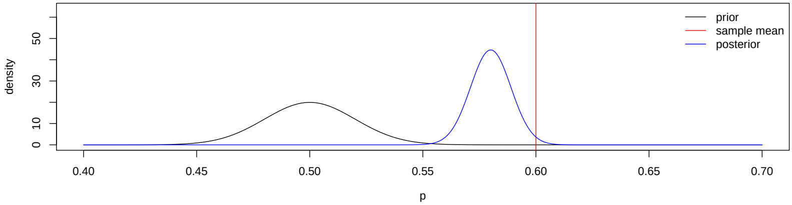
Say we observed the same data (as represented in red).

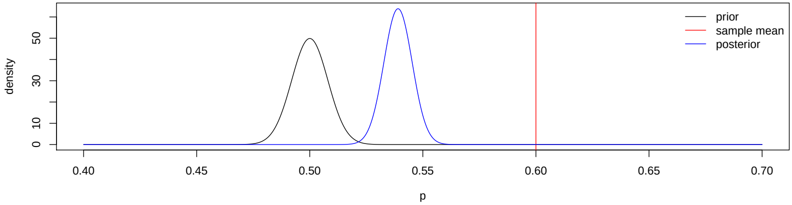




The Bayesian Procedure

The posterior in this case is pulled left due to the sharper prior.





Back to Debate

Frequentists believe statistics should be completely objective and therefore do not accept the premise of a subjective prior.

Back to Debate

Frequentists believe statistics should be completely objective and therefore do not accept the premise of a subjective prior.

Back to Hypothesis Testing Machine example. In order to use such a procedure in real life, we would need to specify some prior belief $\Pr(H_A)$ that H_A is true.