Bayesian Statistics: An Introduction

Tim Frasier

Outline

1. Bayesian statistics, what is it?
2. Criticisms
3. Why haven't I heard of it before?

Bayesian Statistics,
What Is It?

Bayesian Statistics

- A way to reassign probabilities across possibilities (hypotheses)*

$$P(\mathrm{H}|\mathrm{D}) = \frac{P(\mathrm{H}) \text{ x } P(\mathrm{D}|\mathrm{H})}{P(\mathrm{D})}$$

* We'll get into a "richer" definition once you get some experience

Bayesian Statistics

- A way to reassign probabilities across possibilities (hypotheses)*

$$\boxed{P(\mathrm{H}|\mathrm{D})} = \frac{P(\mathrm{H}) \text{ x } P(\mathrm{D}|\mathrm{H})}{P(\mathrm{D})}$$

What we want!!
The "posterior probability"

Bayesian Statistics

- A way to reassign probabilities across possibilities (hypotheses)*

$$P(\mathrm{H}|\mathrm{D}) = \frac{P(\mathrm{H}) \text{ x } \boxed{P(\mathrm{D}|\mathrm{H})}}{P(\mathrm{D})}$$

Likelihood
(from before)

Bayesian Statistics

- A way to reassign probabilities across possibilities (hypotheses)*

$$P(\mathrm{H}|\mathrm{D}) = \frac{\boxed{P(\mathrm{H})} \text{ x } P(\mathrm{D}|\mathrm{H})}{P(\mathrm{D})}$$

"Prior" probability (probability of a hypothesis prior to collecting the data)

Bayesian Statistics

- A way to reassign probabilities across possibilities (hypotheses)*

$$P(\text{H}|\text{D}) = \frac{P(\text{H}) \text{ x } P(\text{D}|\text{H})}{\boxed{P(\text{D})}}$$

Normalizing constant.
Sum of all $P(\text{H})$ x $P(\text{D}|\text{H})$ combinations,
so that all $P(\text{H}|\text{D})$ sum to 1

Bayesian Statistics

- A way to reassign probabilities across possibilities (hypotheses)*

$$P(\mathrm{H}|\mathrm{D}) = \frac{P(\mathrm{H}) \text{ x } P(\mathrm{D}|\mathrm{H})}{P(\mathrm{D})}$$

Discrete Variables

$$P(\mathrm{H}|\mathrm{D}) = \frac{P(\mathrm{H}) \text{ x } P(\mathrm{D}|\mathrm{H})}{\Sigma P(\mathrm{H}) \text{ x } P(\mathrm{D}|\mathrm{H})}$$

Continuous Variables

$$P(\mathrm{H}|\mathrm{D}) = \frac{P(\mathrm{H}) \text{ x } P(\mathrm{D}|\mathrm{H})}{\int P(\mathrm{H}) \text{ x } P(\mathrm{D}|\mathrm{H})}$$

Bayesian Statistics

Example: marbles in jars

- Suppose there are 2 jars:
  - Jar 1: 100 red, 40 green
  - Jar 2: 40 red, 80 green

- You have a green marble. What's the probability that it came from each jar?

Bayesian Statistics
Example: marbles in jars

$$P(\mathrm{H}|\mathrm{D}) = \frac{P(\mathrm{H}) \times P(\mathrm{D}|\mathrm{H})}{P(\mathrm{D})}$$

- Suppose there are 2 jars:
  - Jar 1: 100 red, 40 green
  - Jar 2: 40 red, 80 green

- You have a green marble. What's the probability that it came from each jar?

Bayesian Statistics
Example: marbles in jars

$$P(\mathrm{H}|\mathrm{D}) = \frac{\boxed{P(\mathrm{H})} \times P(\mathrm{D}|\mathrm{H})}{P(\mathrm{D})}$$

- Suppose there are 2 jars:
  - Jar 1: 100 red, 40 green
  - Jar 2: 40 red, 80 green
- You have a green marble. What's the probability that it came from each jar?

Prior Probabilities

- Before collecting data, equally likely from either

```
P(jar1) = 0.5
P(jar2) = 0.5
```

Bayesian Statistics

Example: marbles in jars

$$P(\mathrm{H}|\mathrm{D}) = \frac{\boxed{P(\mathrm{H})} \times P(\mathrm{D}|\mathrm{H})}{P(\mathrm{D})}$$

- Suppose there are 2 jars:
  - Jar 1: 100 red, 40 green
  - Jar 2: 40 red, 80 green
- You have a green marble. What's the probability that it came from each jar?

```
P(jar1) = 0.5
P(jar2) = 0.5
```

Prior Probabilities

- Before collecting data, equally likely from either

```
P(jar1) = 0.5
P(jar2) = 0.5
```

Bayesian Statistics
Example: marbles in jars

$$P(\mathrm{H}|\mathrm{D}) = \frac{P(\mathrm{H}) \times \boxed{P(\mathrm{D}|\mathrm{H})}}{P(\mathrm{D})}$$

- Suppose there are 2 jars:
  - Jar 1: 100 red, 40 green
  - Jar 2: 40 red, 80 green

- You have a green marble. What's the probability that it came from each jar?

```
P(jar1) = 0.5
P(jar2) = 0.5
```

Likelihoods

- From the data

```
P(green | jar1) = (40 / (40 + 100)) = 0.286
P(green | jar2) = (80 / (80 + 40)) = 0.667
```

Bayesian Statistics

Example: marbles in jars

$$P(H|D) = \frac{P(H) \times \boxed{P(D|H)}}{P(D)}$$

- Suppose there are 2 jars:
  - Jar 1: 100 red, 40 green
  - Jar 2: 40 red, 80 green
- You have a green marble. What's the probability that it came from each jar?

```
P(jar1) = 0.5
P(jar2) = 0.5

P(green | jar1) = 0.286
P(green | jar2) = 0.667
```

Likelihoods

- From the data

```
P(green | jar1) = (40 / (40 + 100)) = 0.286
P(green | jar2) = (80 / (80 + 40)) = 0.667
```

Bayesian Statistics
Example: marbles in jars

$$\boxed{P(\mathrm{H}|\mathrm{D})} = \frac{P(\mathrm{H}) \text{ x } P(\mathrm{D}|\mathrm{H})}{P(\mathrm{D})}$$

- Suppose there are 2 jars:
  - Jar 1: 100 red, 40 green
  - Jar 2: 40 red, 80 green

- You have a green marble. What's the probability that it came from each jar?

```
P(jar1) = 0.5
P(jar2) = 0.5

P(green | jar1) = 0.286
P(green | jar2) = 0.667
```

Posterior Probabilities

$$P(\mathrm{jar1} \mid \mathrm{D}) = \frac{0.5 \text{ x } 0.286}{(0.5 \text{ x } 0.286) + (0.5 \text{ x } 0.667)} = 0.300$$

Bayesian Statistics
Example: marbles in jars

$$P(H|D) = \frac{P(H) \times P(D|H)}{P(D)}$$

- Suppose there are 2 jars:
  - Jar 1: 100 red, 40 green
  - Jar 2: 40 red, 80 green

- You have a green marble. What's the probability that it came from each jar?

$P(jar1) = 0.5$
$P(jar2) = 0.5$

$P(green \mid jar1) = 0.286$
$P(green \mid jar2) = 0.667$

Posterior Probabilities

$$P(jar2 \mid D) = \frac{0.5 \times 0.667}{(0.5 \times 0.286) + (0.5 \times 0.667)} = 0.700$$

$P(jar1 \mid D) = 0.300$
$P(jar2 \mid D) = 0.700$

Bayesian Statistics

Example: marbles in jars

$$P(\mathrm{H}|\mathrm{D}) = \frac{P(\mathrm{H}) \times P(\mathrm{D}|\mathrm{H})}{P(\mathrm{D})}$$

- Suppose there are 2 jars:
  - Jar 1: 100 red, 40 green
  - Jar 2: 40 red, 80 green

- You have a green marble. What's the probability that it came from each jar?

- Big whoop, these are similar to what we would have come up with the old way!

```
P(jar1 | D) = 0.300
P(jar2 | D) = 0.700
```

Bayesian Statistics

Example: marbles in jars

$$P(\mathrm{H}|\mathrm{D}) = \frac{P(\mathrm{H}) \times P(\mathrm{D}|\mathrm{H})}{P(\mathrm{D})}$$

- Suppose there are 2 jars:
  - Jar 1: 100 red, 40 green
  - Jar 2: 40 red, 80 green
- You have a green marble. What's the probability that it came from each jar?
- Big whoop, these are similar to what we would have come up with the old way!
  - True, but now we have $P(\mathrm{H}|\mathrm{D})$ rather than $P(\mathrm{D}|\mathrm{H})$
  - Won't always be so similar
  - Can't tell yet, but opens up a whole new world

```
P(jar1 | D) = 0.300
P(jar2 | D) = 0.700
```

Bayesian Statistics

- Can think of it like a table

	Prior Prob. P(H)	Likelihood P(D|H)	Product Π
Jar 1			
Jar 2			
Sum (Σ)			

Bayesian Statistics

- Can think of it like a table

Hypotheses as rows

(need to include all reasonable hypotheses)

	Prior Prob. P(H)	Likelihood P(D|H)	Product Π
Jar 1			
Jar 2			
Sum (Σ)			

Bayesian Statistics

- Can think of it like a table

Needed data as columns

	Prior Prob. $P(H)$	Likelihood $P(D\|H)$	Product Π
Jar 1			
Jar 2			
Sum (Σ)			

Bayesian Statistics

- Can think of it like a table

	Prior Prob. P(H)	Likelihood P(D|H)	Product Π
Jar 1	0.5		
Jar 2	0.5		
Sum (Σ)	1		

Should sum to 1

Bayesian Statistics

- Can think of it like a table

	Prior Prob. P(H)	Likelihood P(D|H)	Product Π
Jar 1	0.5	0.286	
Jar 2	0.5	0.667	
Sum (Σ)	1	0.953	

Has no requirement to sum to 1, but sometimes does (depends on situation)

Bayesian Statistics

- Can think of it like a table

	Prior Prob. $P(H)$	Likelihood $P(D\|H)$	Product Π
Jar 1	0.5	0.286	0.143
Jar 2	0.5	0.667	0.334
Sum (Σ)	1	0.953	0.477

Calculate

Bayesian Statistics

- Can think of it like a table

	Prior Prob. P(H)	Likelihood P(D|H)	Product Π
Jar 1	0.5	0.286	0.143
Jar 2	0.5	0.667	0.334
Sum (Σ)	1	0.953	0.477

$$P(\mathrm{jar1} \mid \mathrm{D}) = \frac{0.143}{0.477} = 0.300$$

$$P(\mathrm{H}|\mathrm{D}) = \frac{P(\mathrm{H}) \times P(\mathrm{D}|\mathrm{H})}{P(\mathrm{D})}$$

$$P(\mathrm{jar2} \mid \mathrm{D}) = \frac{0.334}{0.477} = 0.700$$

Bayesian Statistics

Example: false positives in medical testing

Bayesian Statistics
Example: false positives in medical testing

- Suppose infection frequency in population is 1%
- False negative rate = 20% (80% success at detection)
- Rate of false positive is 10%

You’ve tested positive!

What’s the probability you’re infected?

Bayesian Statistics

Example: false positives in medical testing

$$P(\mathrm{H}|\mathrm{D}) = \frac{P(\mathrm{H}) \text{ x } P(\mathrm{D}|\mathrm{H})}{P(\mathrm{D})}$$

- Suppose infection frequency in population is 1%
- False negative rate = 20% (80% success at detection)
- Rate of false positive is 10%

Specify

	Prior Prob. P(H)	Likelihood P(D|H)	Product Π
Infected			
Not Infected			
Sum (Σ)			

Bayesian Statistics

Example: false positives in medical testing

$$P(\mathrm{H}|\mathrm{D}) = \frac{P(\mathrm{H}) \text{ x } P(\mathrm{D}|\mathrm{H})}{P(\mathrm{D})}$$

- Suppose infection frequency in population is 1%
- False negative rate = 20% (80% success at detection)
- Rate of false positive is 10%

	Prior Prob. P(H)	Likelihood P(D|H)	Product Π
Infected	`0.01`		
Not Infected	`1 - (0.01) = 0.90`		
Sum (Σ)	`1`		

Bayesian Statistics
Example: false positives in medical testing

$$P(\mathrm{H}|\mathrm{D}) = \frac{P(\mathrm{H}) \times P(\mathrm{D}|\mathrm{H})}{P(\mathrm{D})}$$

- Suppose infection frequency in population is 1%
- False negative rate = 20% (80% success at detection)
- Rate of false positive is 10%

	Prior Prob. P(H)	Likelihood P(D|H)	Product Π
Infected	0.01	0.80	
Not Infected	1 – (0.01) = 0.90	0.10	
Sum (Σ)	1		

Bayesian Statistics
Example: false positives in medical testing

$$P(\mathrm{H}|\mathrm{D}) = \frac{P(\mathrm{H}) \times P(\mathrm{D}|\mathrm{H})}{P(\mathrm{D})}$$

- Suppose infection frequency in population is 1%
- False negative rate = 20% (80% success at detection)
- Rate of false positive is 10%

	Prior Prob. P(H)	Likelihood P(D|H)	Product Π
Infected	0.01	0.80	0.008
Not Infected	1 – (0.01) = 0.99	0.10	0.099
Sum (Σ)	1		0.107

Bayesian Statistics

Example: false positives in medical testing

$$P(\mathrm{H}|\mathrm{D}) = \frac{P(\mathrm{H}) \times P(\mathrm{D}|\mathrm{H})}{P(\mathrm{D})}$$

	Prior Prob. P(H)	Likelihood P(D|H)	Product Π
Infected	0.01	0.80	0.008
Not Infected	1 - (0.01) = 0.99	0.10	0.099
Sum (Σ)	1		0.107

$$P(\text{infected} \mid \text{positive}) = \frac{0.008}{0.107} = 0.075$$

$$P(\text{not infected} \mid \text{positive}) = \frac{0.099}{0.107} = 0.925$$

Criticisms

Criticisms

Prior, prior pants on fire*

*From McElreath (2016) *Statistical Rethinking: A Bayesian Course with Example in R and Stan*. CRC Press.

Criticisms

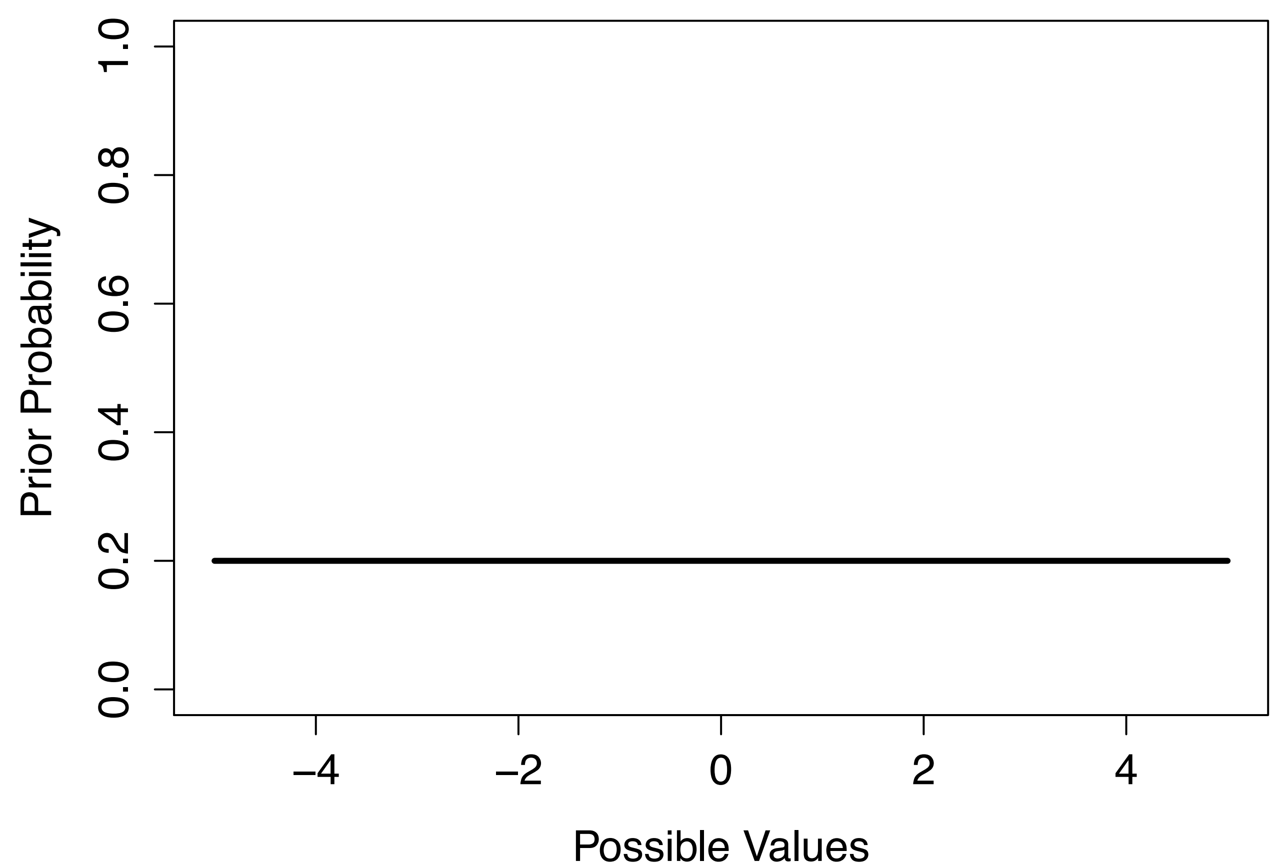
Prior, prior pants on fire

- Extremely useful when you have prior information…
  - Repeated experiments
  - Multiple sources if information
  - *Null hypothesis significance testing can't incorporate this*
- …but, most of the time we don't have any prior information
  - What do we do in these situations?

Criticisms

Prior, prior pants on fire

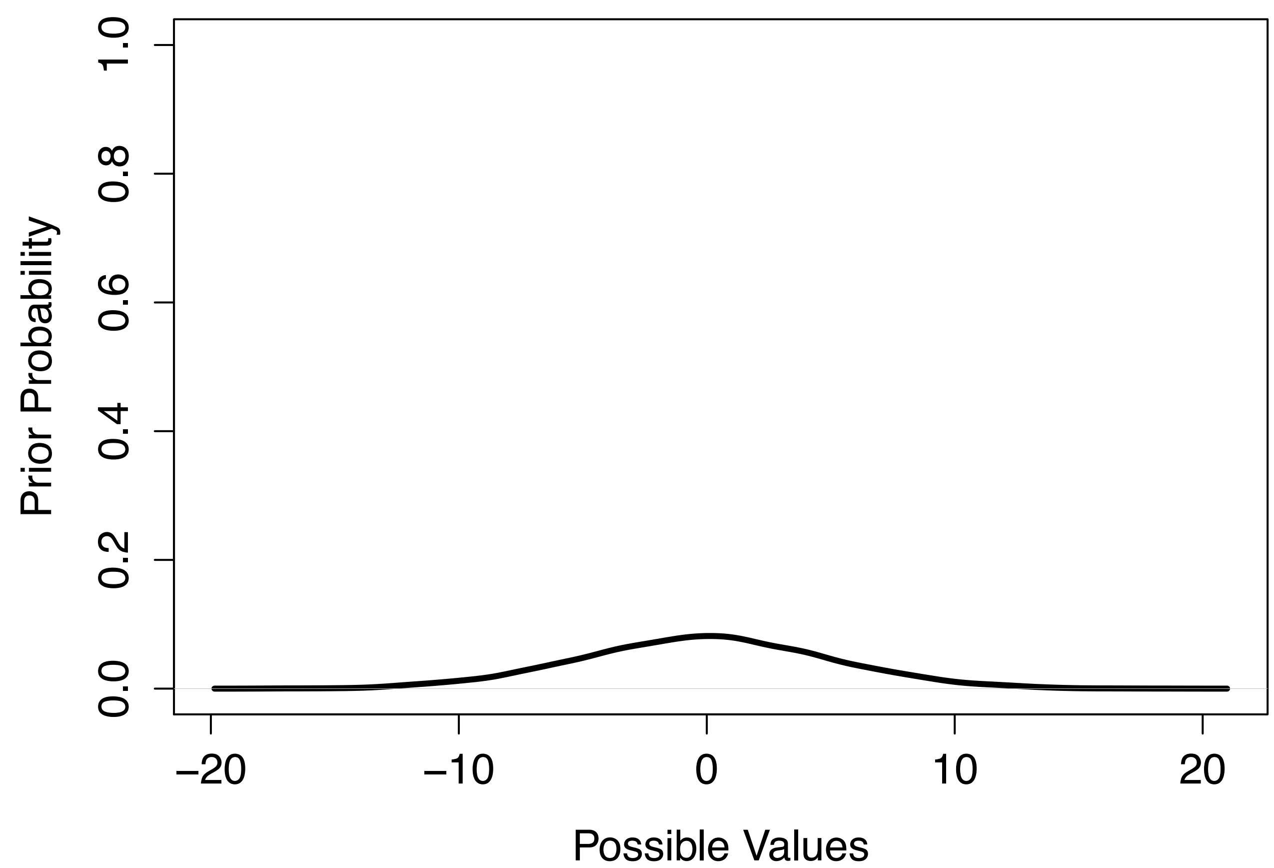
1. Choose “uninformative” priors
   - Equal probabilities across possible values



Criticisms

Prior, prior pants on fire

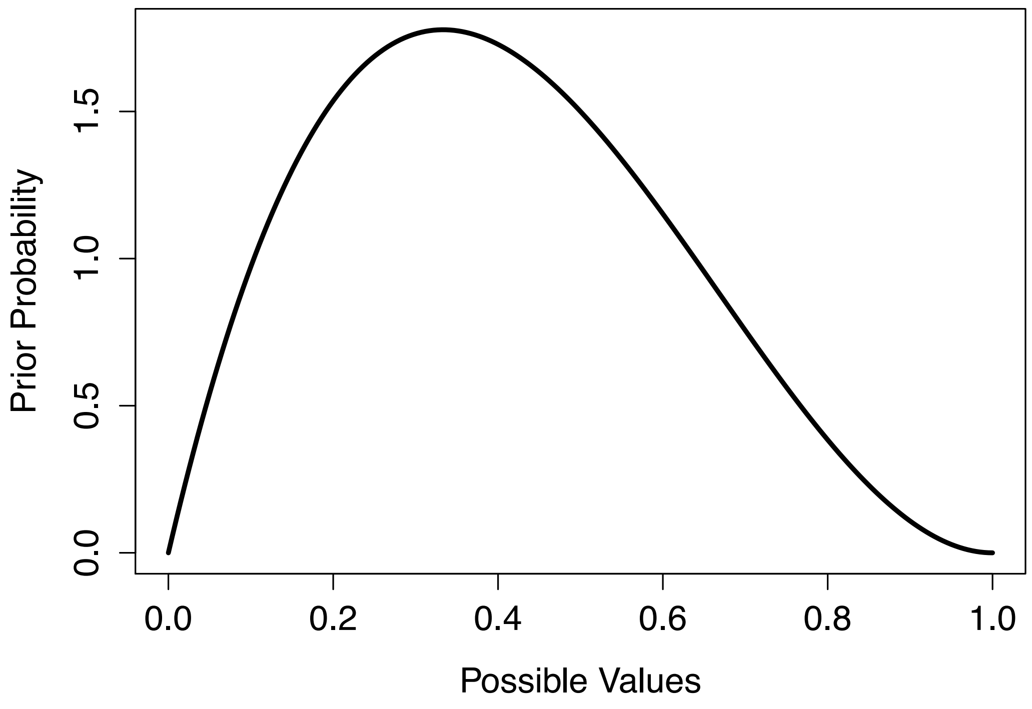
2. Choose "weakly informed" priors
   - Very loosely based on the data

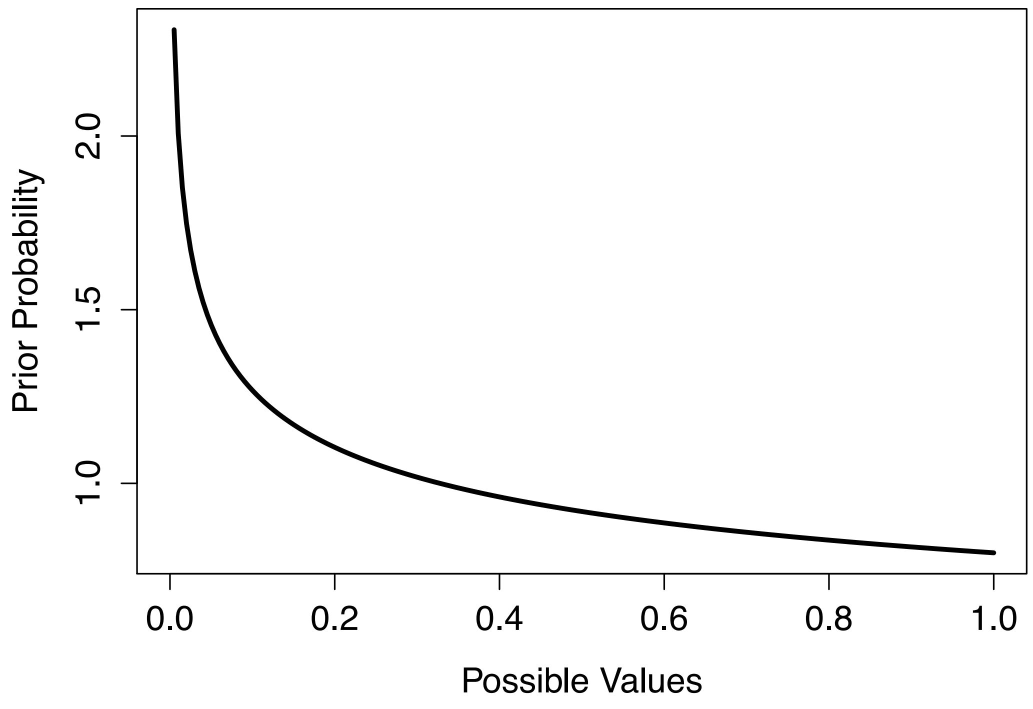


Criticisms

Prior, prior pants on fire

3. Be explicit about what you really think!





Criticisms

Prior, prior pants on fire

4. Choose multiple types of priors and ensure results are robust to these changes
    - If you get different posterior probabilities when you use different realistic priors, then there is not enough information in your data to obtain a good estimate

Criticisms

Prior, prior pants on fire

- Some people think this requirement makes the analyses too subjective

Criticisms

Prior, prior pants on fire

- Some people think this requirement makes the analyses too subjective
  - There are many subjective aspects to NHST too!!!
    - Choice of critical p-value
    - Requirement of normal distribution (which Bayesian analyses don't have)
    - Intentions of the researcher
    - What test to use
    - etc…

Criticisms
Prior, prior pants on fire

- Some people think this requirement makes the analyses too subjective
  - There are many subjective aspects to NHST too!!!

Statistical Analysis and the Illusion of Objectivity

James O. Berger
Donald A. Berry

Berger & Berry (1988) *American Scientist* **76**: 159-165.

Criticisms

Prior, prior pants on fire

- Some people think this requirement makes the analyses too subjective
  - I think this is a strength
    - Forces us to think about, and be explicit about, our assumptions (many of these are hidden in NHST)
    - If there is debate, use multiple priors (that make everyone happy), and ensure the same results are obtained
    - What could be wrong with that?

If Bayesian analysis is so great, why haven’t I heard of it before?

Why Haven’t I Heard Of It?

- Lack of possibilities prior to good computing power

Why Haven't I Heard Of It?

- Lack of possibilities prior to good computing power
- Problem
  - Could only apply Bayesian analysis to "simple" problems

$$P(\mathrm{H}|\mathrm{D}) = \frac{P(\mathrm{H}) \text{ x } P(\mathrm{D}|\mathrm{H})}{\boxed{\int P(\mathrm{H}) \text{ x } P(\mathrm{D}|\mathrm{H})}}$$

Unsolvable for most problems

Why Haven't I Heard Of It?

- Lack of possibilities prior to good computing power
- Solution
  - Instead of calculating it exactly, can estimate it with Markov Chain Monte Carlo (MCMC) methods
  - Require substantial computing power not available until recently (now common)

Why Haven't I Heard Of It?

- Lack of possibilities prior to good computing power
- Unfortunate situation
  - Our current state of statistics would likely be very different were it not for this limitation/requirement

Questions?