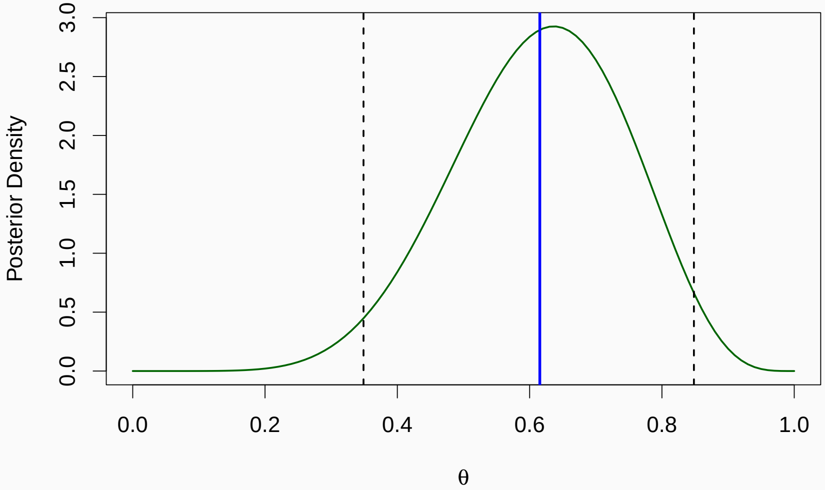
Lecture 3: Updating

Jeff Rouder

June, 2022

Previously... Posterior Was Goal



- Updating Orientation
- Bayes Rule Applied More Broadly

Up To Now...

Bayes' Rule:

$$\pi(\theta|Y) = \frac{p(Y|\theta)\pi(\theta)}{p(Y)}.$$

Proportional Form:

$$\pi(\theta|Y) \propto l(\theta; Y) \times \pi(\theta),$$

Critique

Problems:

- No sense of what $p(Y)$ means
- Stress posterior rather than the process of updating.
- Strive to minimize influence of priors.
- Separation of estimation and model comparison.

Ratio Form of Bayes Rule

Solution, restate Bayes Rule as follows:

$$\frac{\pi(\theta|Y)}{\pi(\theta)} = \frac{p(Y|\theta)}{p(Y)}.$$

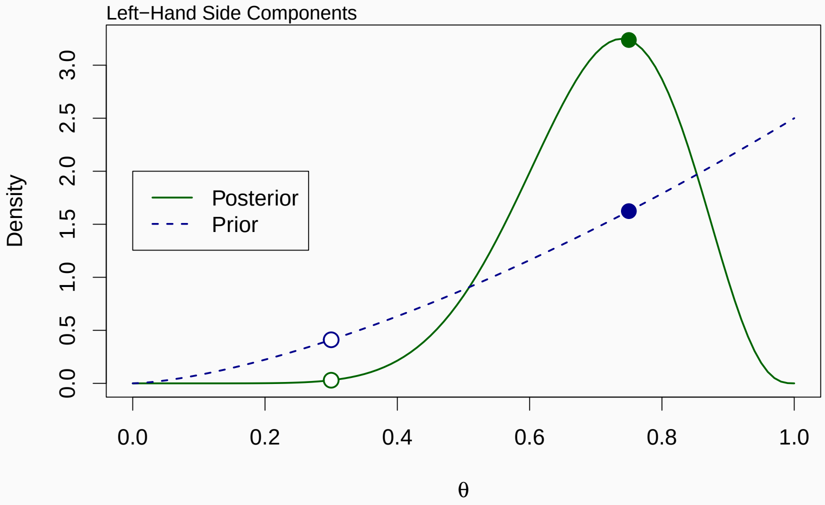
And study what this equation means!

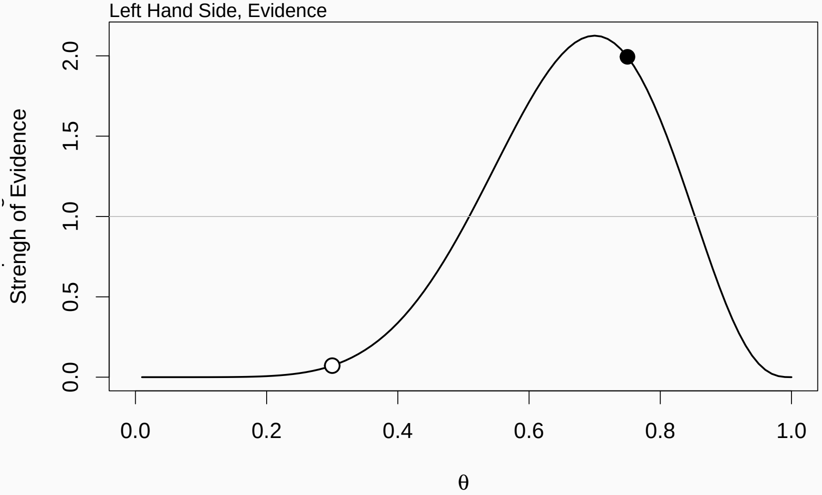
For Our Example

- 7 successes in 10 trials

Left-Hand Side

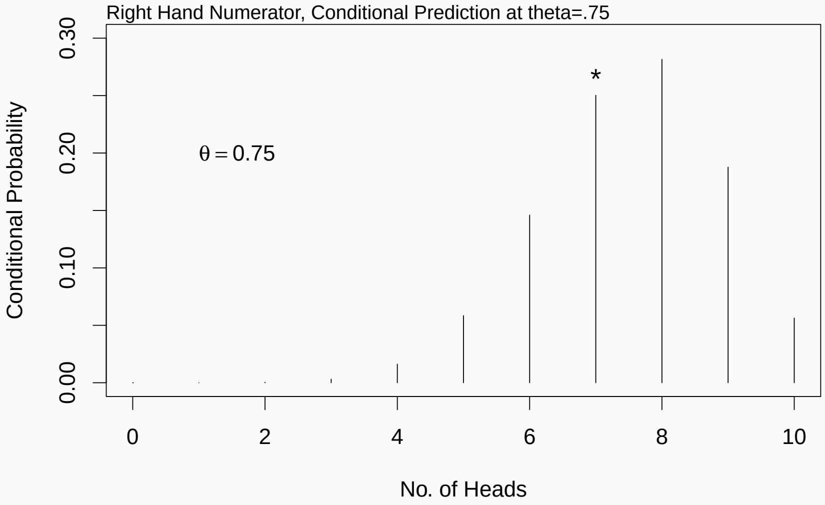
- Bayes Rule:
$$\frac{\pi(\theta|Y)}{\pi(\theta)} = \frac{p(Y|\theta)}{p(Y)}$$
- Left-Hand Side:
$$\frac{\pi(\theta|Y)}{\pi(\theta)}$$

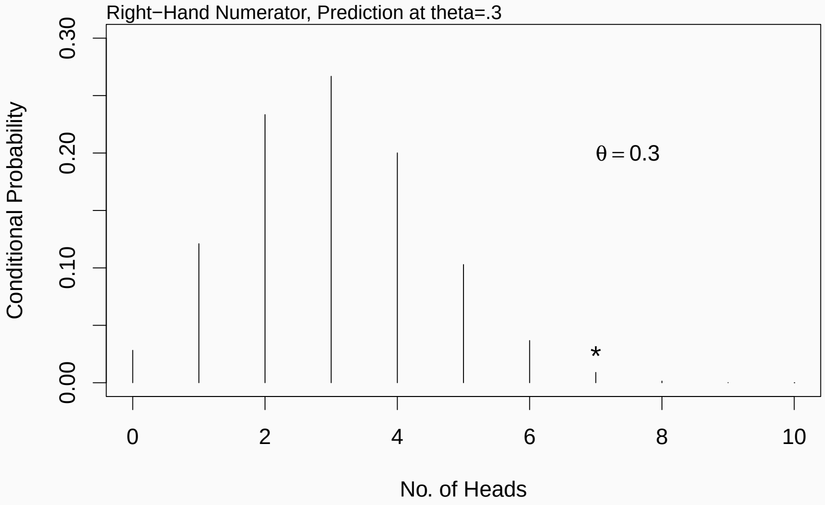

Left−Hand Side Components
Posterior
Prior
Density
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0.0
0.2
0.4
0.6
0.8
1.0
θ

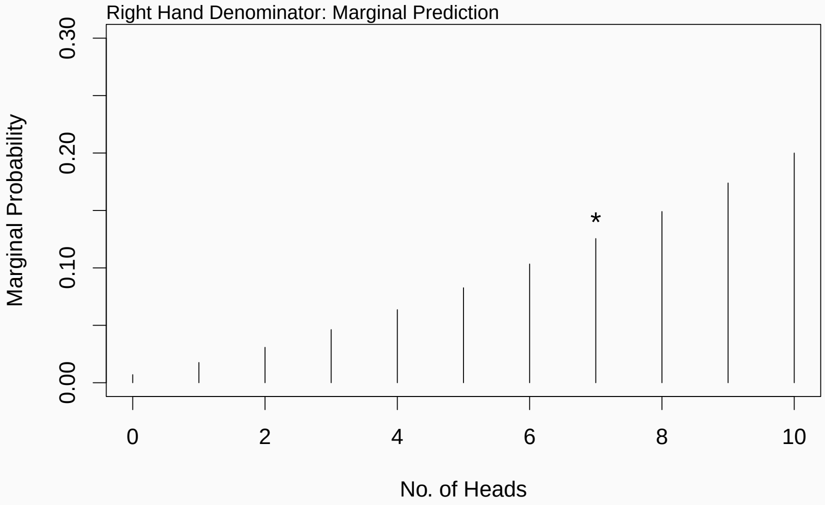

Left Hand Side, Evidence
Strengh of Evidence
0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0
θ

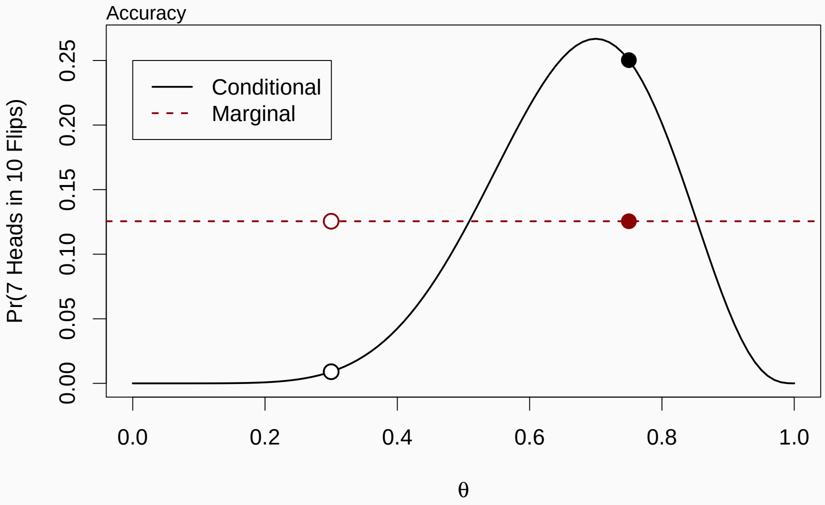
Right-Hand Side

- Bayes Rule:
$$\frac{\pi(\theta|Y)}{\pi(\theta)} = \frac{p(Y|\theta)}{p(Y)}$$
- Right-Hand Side:
$$\frac{p(Y|\theta)}{p(Y)}$$
- $p(Y|\theta)$: Conditional Predictive Accuracy
- $p(Y)$: Marginal Predictive Accuracy

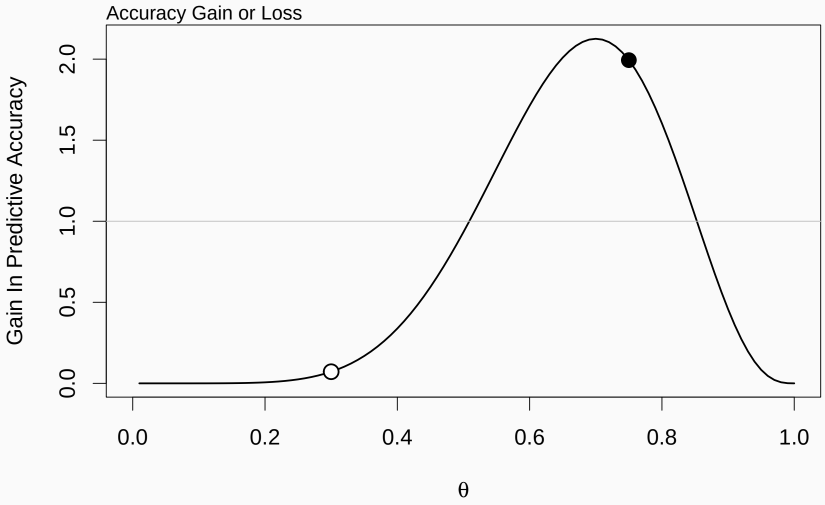

Right Hand Numerator, Conditional Prediction at theta=.75
θ = 0.75
*
Conditional Probability
0.00
0.10
0.20
0.30
No. of Heads
0
2
4
6
8
10


Right−Hand Numerator, Prediction at theta=.3
Conditional Probability
0.00
0.10
0.20
0.30
θ = 0.3
*
0
2
4
6
8
10
No. of Heads


Right Hand Denominator: Marginal Prediction
Marginal Probability
0.00
0.10
0.20
0.30
*
0
2
4
6
8
10
No. of Heads



Accuracy
Conditional
Marginal
Pr(7 Heads in 10 Flips)
0.00
0.05
0.10
0.15
0.20
0.25
0.0
0.2
0.4
0.6
0.8
1.0
θ


Accuracy Gain or Loss
Gain In Predictive Accuracy
0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0
θ

Evidence for a point is the degree to which it improves the prediction for the observed data.

- “Evidence = Prediction”
- way catchier than “Posterior is Proportional To The Likelihood Times The Prior”

Your Turn: Which is the better parameter value?

- A coin has probability either $\theta = 1/3$ or $\theta = 2/3$.
- We observe Y out of N successes.
- Q: What is the relative evidence of $\theta = 1/3$ vs. $\theta = 2/3$

Need a Hint?

Compute

$$\frac{\frac{\pi(\theta_1|Y)}{\pi(\theta_1)}}{\frac{\pi(\theta_2|Y)}{\pi(\theta_2)}}$$

- where $\theta_1 = 1/3$ and $\theta_2 = 2/3$.
- derive generally and then put in values.

Fair Coin?

We have observed 7 of 10 successes. What is the evidence for a fair coin?

General Model, Model M_a

$$Y \mid \theta \sim \text{Binomial}(\theta, N),$$
$$\theta \sim \text{Uniform}(0, 1).$$

Fair-Coin Model, Model M_b:

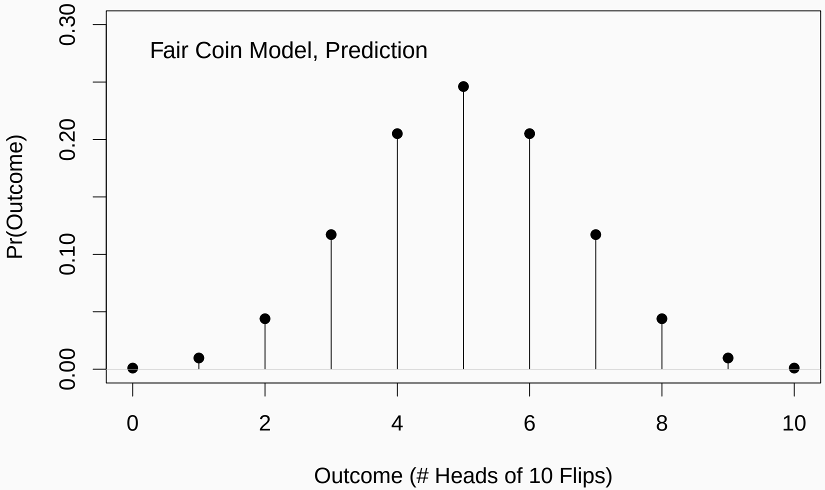
$$Y \sim \text{Binomial}(.5, N).$$

Bayes Rule:

$$\frac{\frac{\pi(M_a|Y)}{\pi(M_a)}}{\frac{\pi(M_b|Y)}{\pi(M_b)}} = \frac{\frac{p(Y|M_a)}{p(Y)}}{\frac{p(Y|M_b)}{p(Y)}} = \frac{p(Y|M_a)}{p(Y|M_b)} = B_{ab}.$$

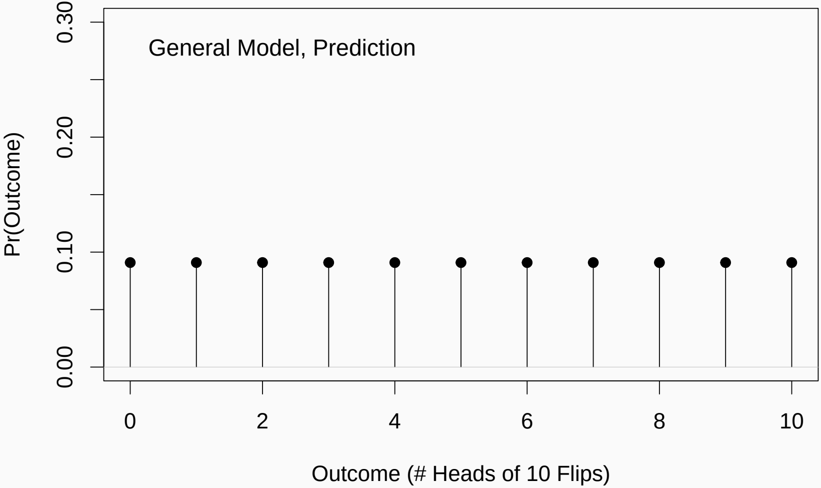
Fair Coin Prediction

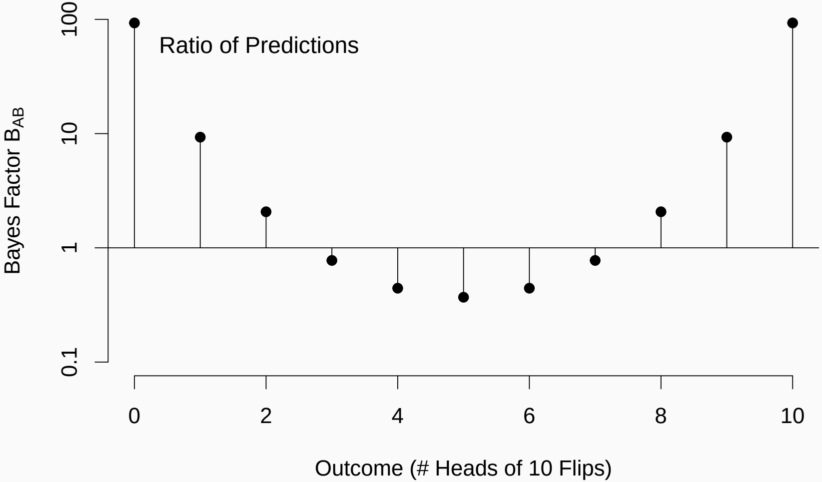
$$Pr(Y = y) = \binom{N}{y} .5^y , 5^{(N-y)}$$


Fair Coin Model, Prediction
Pr(Outcome)
0.00
0.10
0.20
0.30
0
2
4
6
8
10
Outcome (# Heads of 10 Flips)

General Model Prediction

$$Pr(Y = y) = \int_0^1 \binom{N}{y} \theta^y, (1-\theta)^{(N-y)} \, d\theta$$


General Model, Prediction
Pr(Outcome)
0.00
0.10
0.20
0.30
0
2
4
6
8
10
Outcome (# Heads of 10 Flips)


Ratio of Predictions
Bayes Factor B_{AB}
100
10
1
0.1
0
2
4
6
8
10
Outcome (# Heads of 10 Flips)

Comparison of Updating vs. Posterior Orientation

General Model (M_1):

$$\begin{aligned} Y_i|\mu \sim & \mathrm{N}(\mu, \sigma^2) \\ \mu \sim & \mathrm{N}(0, b) \end{aligned}$$

vs.

Null Model (M_0):

$$Y_i \sim \mathrm{N}(0, \sigma^2)$$

To make the lecture easier to follow, I am going to assume we know σ^2 and set it to $\sigma^2 = 1$.

Data

```r
#make data
set.seed(456)
N=20
t.mean=-1
y=rnorm(N,t.mean,1)
print(mean(y))
```

```
## [1] -0.5169787
```

Conventional (Frequentist) Approach

When the variance is known, the conventional approach to this problem is to compute a z-value and compare it to a z distribution. The z-score here is

$$z = \frac{\sqrt{N}\bar{y}}{\sigma}$$

For our case, $z = \sqrt{N}\bar{y}$ (so easy).

Frequentist Analysis

```
z=mean(y)*sqrt(N)
#one-tail and two-tailed pvalues
print(c(pnorm(z),2*(pnorm(z))))
```

```
## [1] 0.01038887 0.02077774
```

```
#95% CI
ci=c(mean(y)-1.96/sqrt(N),mean(y)+1.96/sqrt(N))
print(ci)
```

```
## [1] -0.95524801 -0.07870936
```

Posterior Estimation

For the posterior orientation, we compute a posterior on μ for the general model.

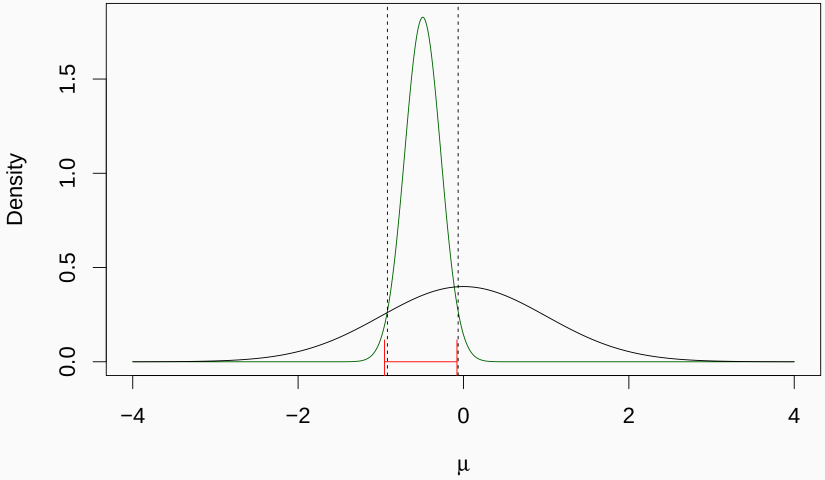
The resulting posterior is:

$$\mu|Y \sim \text{Norma}(vc, v),$$

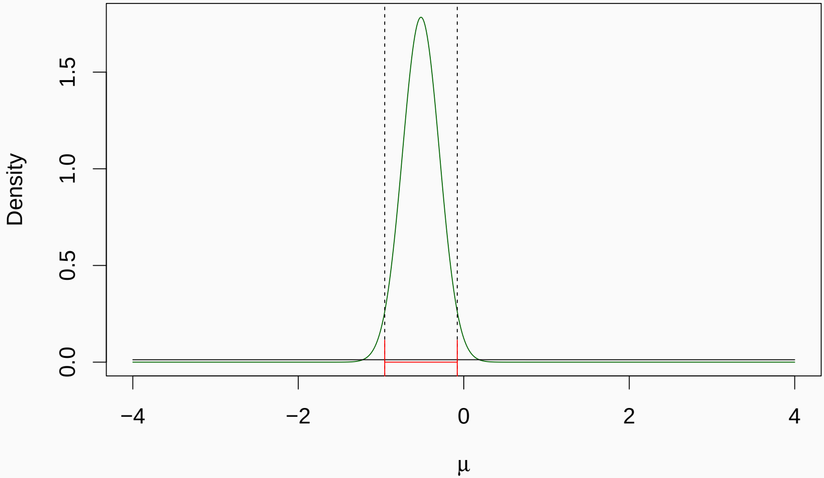
where

$$v = (N + 1/b)^{-1}, \quad c = N\bar{y}.$$

Example, $b = 1$



Example, $b = 1000$



Upshot

Inference by intervals are very similar. Posterior Bayesian analysis has the same calibrations as conventional analysis.

Questions

Questions for the back of your mind:

- Should we just use flat, noninformative priors for inference?
- Why be Bayesian then?
- Is inference by interval (or p-value) principled? Is it a good idea?
- Is the insensitivity to the prior reasonable?

Prior Is Part of The Model

$$\begin{aligned} Y_i|\mu \sim& \mathrm{N}(\mu, \sigma^2) \\ \mu \sim& \mathrm{N}(0, b) \end{aligned}$$

$$\begin{aligned} \bar{Y}|\mu \sim& \mathrm{N}(\mu, \sigma^2/N) \\ \mu \sim& \mathrm{N}(0, b) \end{aligned}$$

Parameter-free version

$$\bar{Y} \sim \mathrm{N}(0, b + \sigma^2/N)$$

Prior Is Part of The Model

General Model:

$$\bar{Y} \mid M_1 \sim \mathrm{N}(0, b + \sigma^2/N)$$

Null Model

$$\bar{Y} \mid M_0 \sim \mathrm{N}(0, \sigma^2/N)$$

Prior defines difference in models!

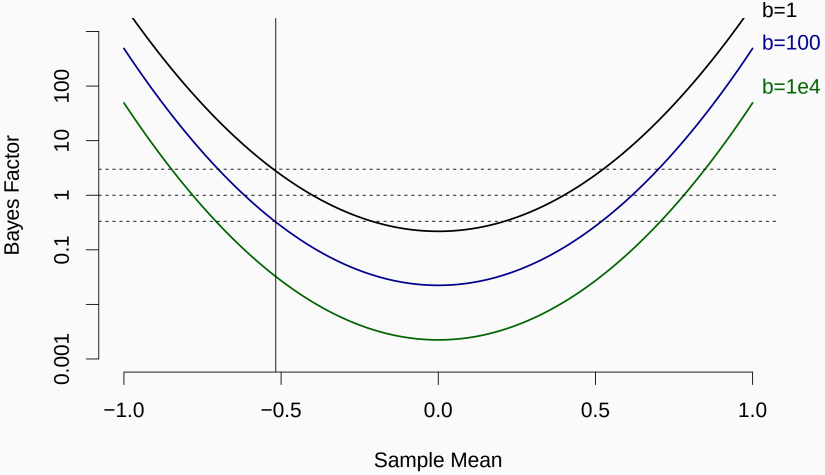
Relative Strength of Models:

$$\frac{\frac{\pi(M_1|\bar{Y})}{\pi(M_1)}}{\frac{\pi(M_0|\bar{Y})}{\pi(M_0)}} = \frac{\frac{p(\bar{Y}|M_1)}{p(Y)}}{\frac{p(\bar{Y}|M_0)}{p(\bar{Y})}} = \frac{p(\bar{Y}|M_1)}{p(\bar{Y}|M_0)} = B_{10}.$$

Relative Strength of Models ($sigma^2 = 1$):

$$\frac{p(\bar{Y}|M_1)}{p(\bar{Y}|M_0)} = \frac{\phi\left(\frac{\bar{Y}}{\sqrt{b+1/N}}\right)}{\phi\left(\frac{\bar{Y}}{\sqrt{1/N}}\right)}$$

Relative Strength of Models ($sigma^2 = 1$):



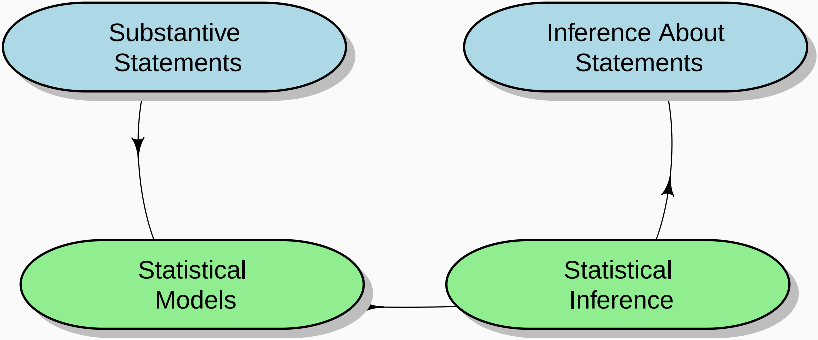
Your Turn

- The above plot is for fixed sample size as a function of sample mean for a few values of b.
- Let's fix $\bar{Y} = -.5$, $\sigma^2 = 1$.
- Make plots showing the effects of n (sample size) for a few values of b.

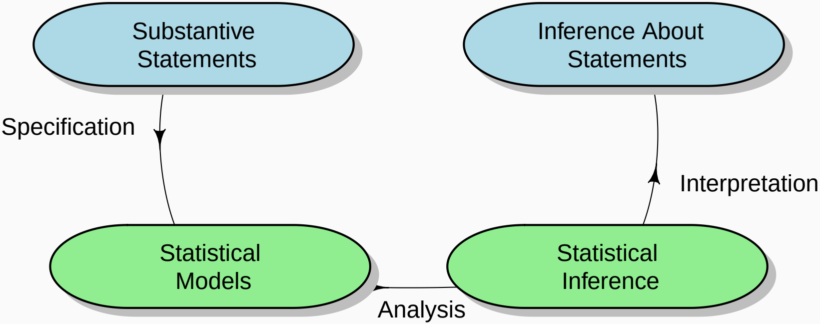
Coda, Why Model?

1. Measurement (applied setting)
2. Theory development

Coda, What Are Models?



Coda, What Are Models



Coda, Substantive Statements

Examples:

- Working-Memory Training Transfer: *"Practicing one working-memory task will increase performance in other related tasks."*
- Weber's Law: $\Delta I = kI$.
- Stroop: *Responses to congruent items is better than to incongruent ones*
- Medical: *A certain drug is beneficial to all but benefits women more than men*

Coda, Substantive Statments

- Each statement describes constraint in the world
- None of these statements may be tested with data
- None has specification of sampling noise or errors in measurements.
- And that is appropriate

Coda, What Are Models

Models Are Instantiations of Substantive Constraint That *Connect* To Data

1. Models are proxies that capture theoretically interesting constraint in the substantive world.
2. Models result in probability statements on observables

Coda, Inference

- IF models are:
  - Proxies that capture theoretically interesting constraint in the substantive world.
  - Probability statements on observables (predictions)
- THEN
  - Strength of evidence from data for models are the predictive accuracy of each.
  - by proxy, strength of evidence from data for theoretical positions.
- CONCLUSION
  - Inference in science is done through Bayes factor.