Jeff Witmer
Oberlin College

Bayesian for Undergraduates:
A Prior is Modified

JSM almost in Philadelphia

4 August 2020

In 1985 about 10% of JASA Applications and Case Studies articles used Bayesian methods.

In 1996 33% of JASA Applications and Case Studies articles used Bayesian methods.

What was the percentage for 2016?

(a) 10%

(b) 27%

(c) 33%

(d) 46%

(e) 61%

Answer: 61%

David Moore 1996 JSM talk / 1997 paper

“Bayes for Beginners? Some Reasons to Hesitate.”

David’s reasons:

i) “Bayesian methods are rarely used in practice.”

ii) It isn’t clear what to teach (lack of standard methods and software).

iii) Conditional probability is hard.

iv) “Inference is only part of statistics in practice.”

Times have changed!

i) “Bayesian methods are rarely used in practice.”

See the JASA data above.

Sure, 10% was a small number in 1985.
Did David realize the 33% number in 1986? Maybe not.

In any event, 61% is not a small number.

ii) It isn't clear what to teach (lack of standard methods and software)

We have (good) books: McElreath; Kruschk; Reich and Ghosh; Johnson, Ott, Dogucu; Rundel et al.; Hu and Albert; others.

We have MCMC.

We have (good) software and it keeps getting better. We went from nothing, to BUGS, to jags and rjags, to Stan.

iii) Conditional probability is hard.

We don’t need to limit ourselves to hand calculation of Bayes’ Theorem.

(Also, fast computing means that we can conduct multiple analyses using different priors to conduct a sensitivity analysis.)

If your idea of teaching Bayesian statistics is to present examples of Bayes' Theorem

e.g., the Monte Hall problem:

Pr{ 🚗 behind Door #2| 🐐 behind Door #1}

or Pr{disease|test positive}

then yes, you will have students struggle with

Pr{A|B} = Pr{A and B}/Pr{B}.

But there is no need to do that.

(a) Such probability questions are optional;
(b) Set up a table – rather than present a formula – and life is easy.

Trump approval, late June 2020

	Dem	**Indep**	**Rep**	**TOTAL**
Approve	6	139	237	382
Don’t	314	281	23	618
TOTAL	320	420	260	1000

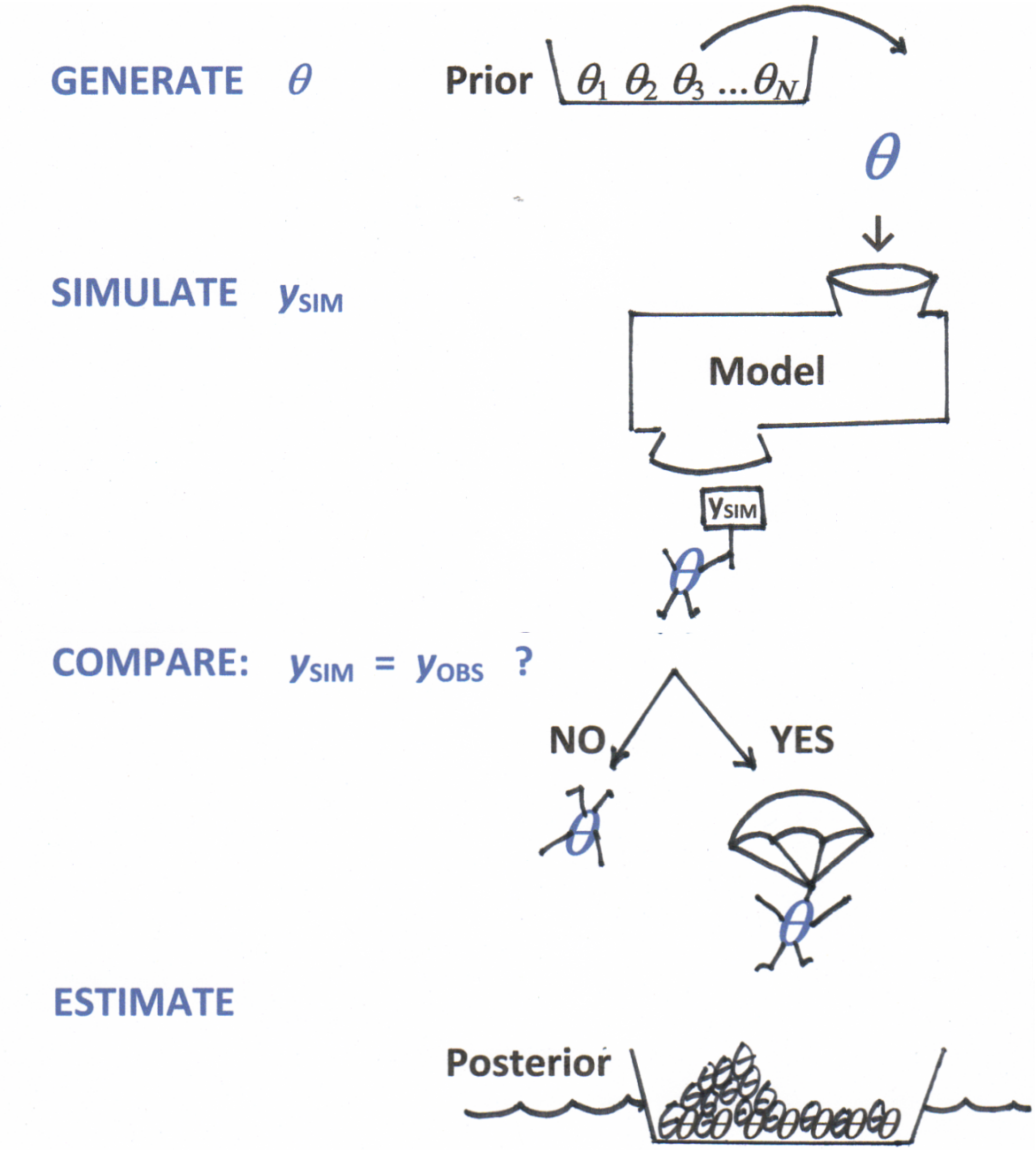
Pr{Rep} = 0.26

Pr{Rep | Approve} = 237/382 = 0.62

If you want to present Bayes' Theorem, then present this version of it:

$$\frac{\Pr\{\theta \mid \text{data}\}}{\Pr\{\theta\}} = \frac{\Pr\{\text{data} \mid \theta\}}{\Pr\{\text{data}\}}$$

The point is that Bayes' Theorem reallocates probability across possible values of the parameter and does so in proportion to how likely the data are given the parameter.



By George Cobb

iv) “Inference is only part of statistics in practice.”

Yes; no one disputes that. But when it comes to inference, we can improve on what p-values have given us.

Going from normal theory to the world of the bootstrap and randomization testing requires a change in perspective, but only a minor change.

Going to Bayesian thinking requires a larger change.

We naturally ask *"How likely is it that this drug works, given the data?"* but frequentist reasoning forces us to address the reverse question of *"How likely are these data if the drug doesn't work?"*

Floyd Bullard example

You are waiting on a subway platform for a train that is known to run on a regular schedule, only you don't know how much time is scheduled to pass between train arrivals, nor how long it's been since the last train departed.

(You're pretty ignorant.)

As more time passes, do you

(a) grow more confident that the train will arrive soon, since its eventual arrival can only be getting closer, not further away, or (b) grow less confident that the train will arrive soon, since the longer you wait, the more likely it seems that either the scheduled arrival times are far apart or else that you happened to arrive just after the last train left--or both.

If you choose (a), you're thinking like a frequentist. If you choose (b), you're thinking like a Bayesian.

Doing Bayesian Inference

Beta-Binomial setting

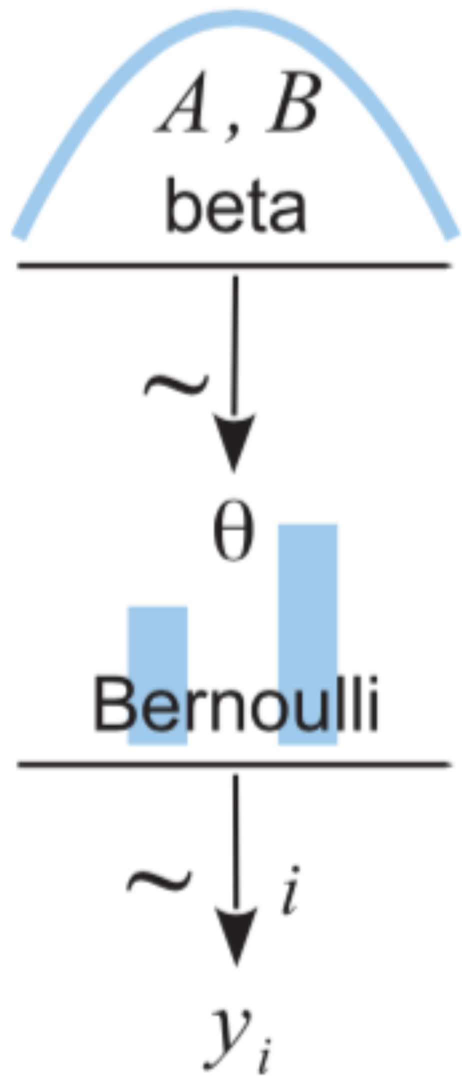
Three ways to analyze beta-binomial data:

(1) Theory – what we did before computers

(2) Discrete/grid prior – with slow computers

(3) MCMC – with modern computers

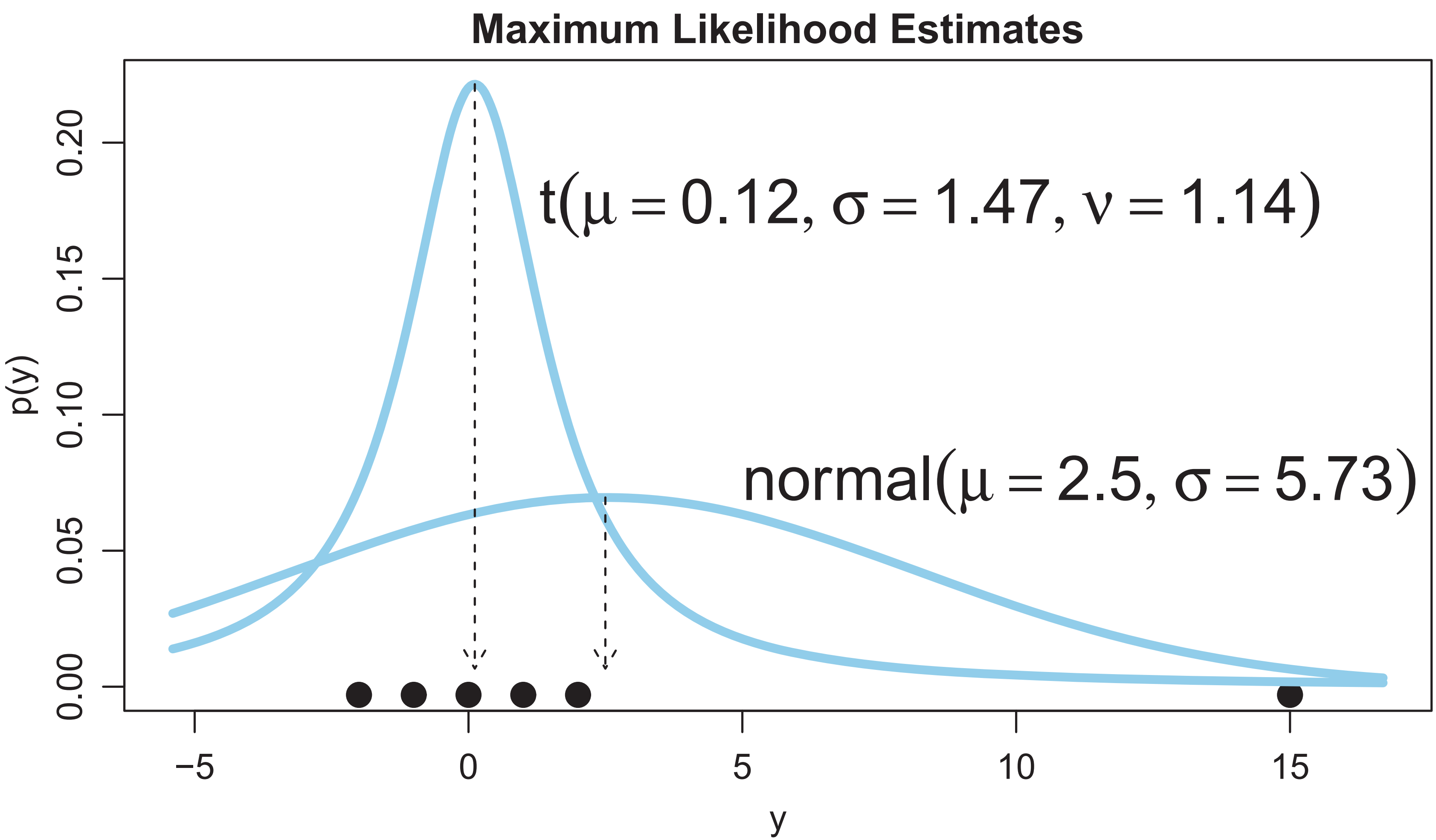
Krsuschke schematic – Figure 8.2 from DBDA 2ed



Consider MCMC and ***using a t likelihood*** **versus a normal likelihood**.

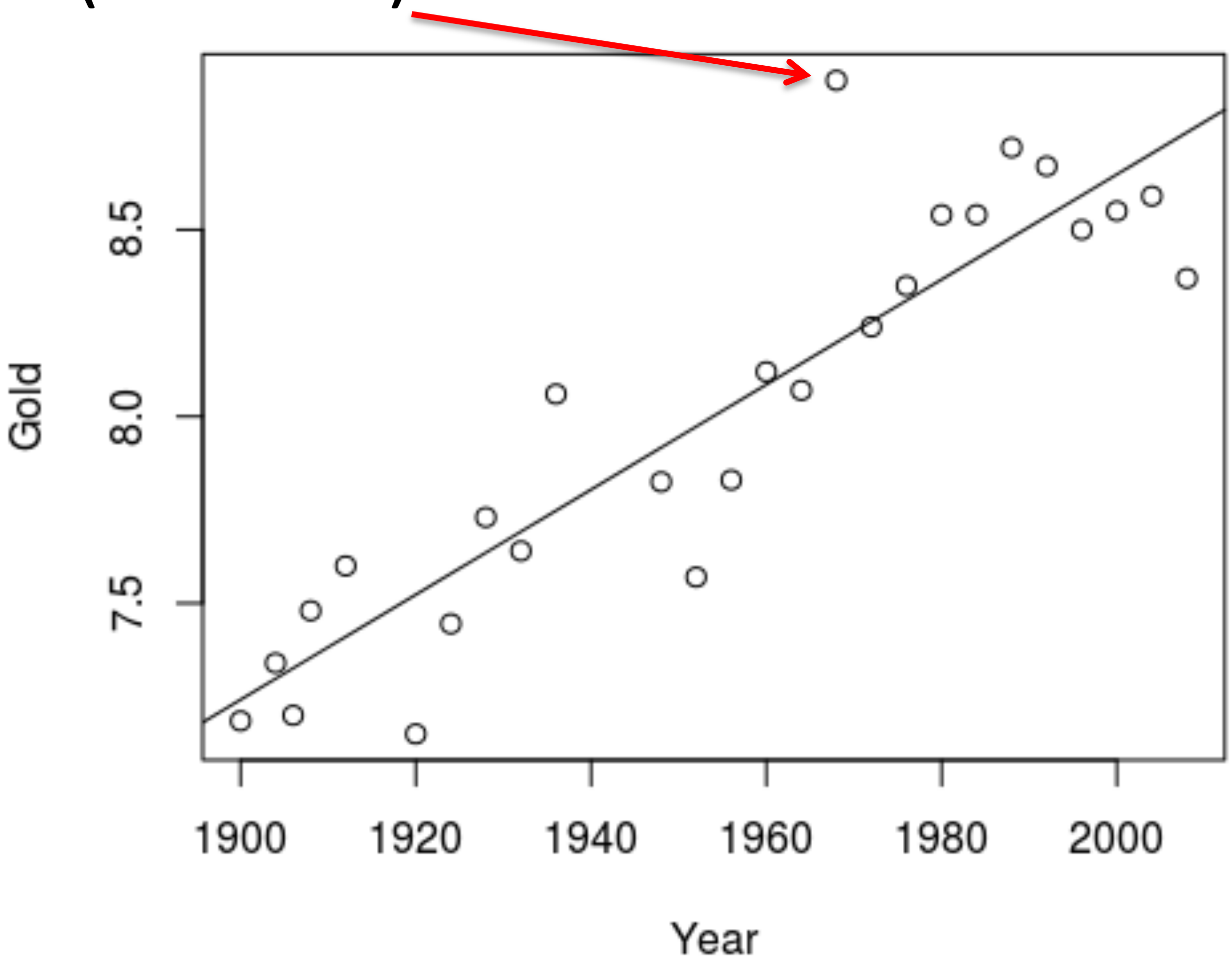
(Granted, a t likelihood could be used by a frequentist. I just don’t see that, whereas Bayesians using MCMC are naturally open to non-normal-theory models.)

From John Kruschke's book: effect of an outlier on the MLE



E.g., Bob Beamon (1968) Olympic long jump gold medal (and world record) – an outlier. How to handle?

Outlier (Beamon)



Use a t distribution for the response (Gold medal distance) in a regression model.

This avoids the “keep it” or “toss it” question.

"keep it": 95% CI for slope is (**1.12, 1.69**), s = **0.238**

"toss it": 95% CI for slope is (**1.14, 1.60**), s = **0.191**

Bayes with t likelihood: 95% HDI is (**1.13, 1.67**), posterior mode for sigma is **0.212**.

(***) **Free Throw Shooting and Hierarchical structure.** I collected data on the centers and guards for the teams that made it to the "Elite Eight" in the 2015 NCAA basketball tournament.

The free throw shooting performance of

guards ranged from a low of **55%** to a high of **89%** success, compared to

centers who ranged from a low of **48%** to a high of **78%**.

A naïve analysis might pool together all hits and misses for each position and compare the aggregate guard percentage (78.6%) to the aggregate center percentage (61.3%), as if all guards were interchangeable and all centers were interchangeable.

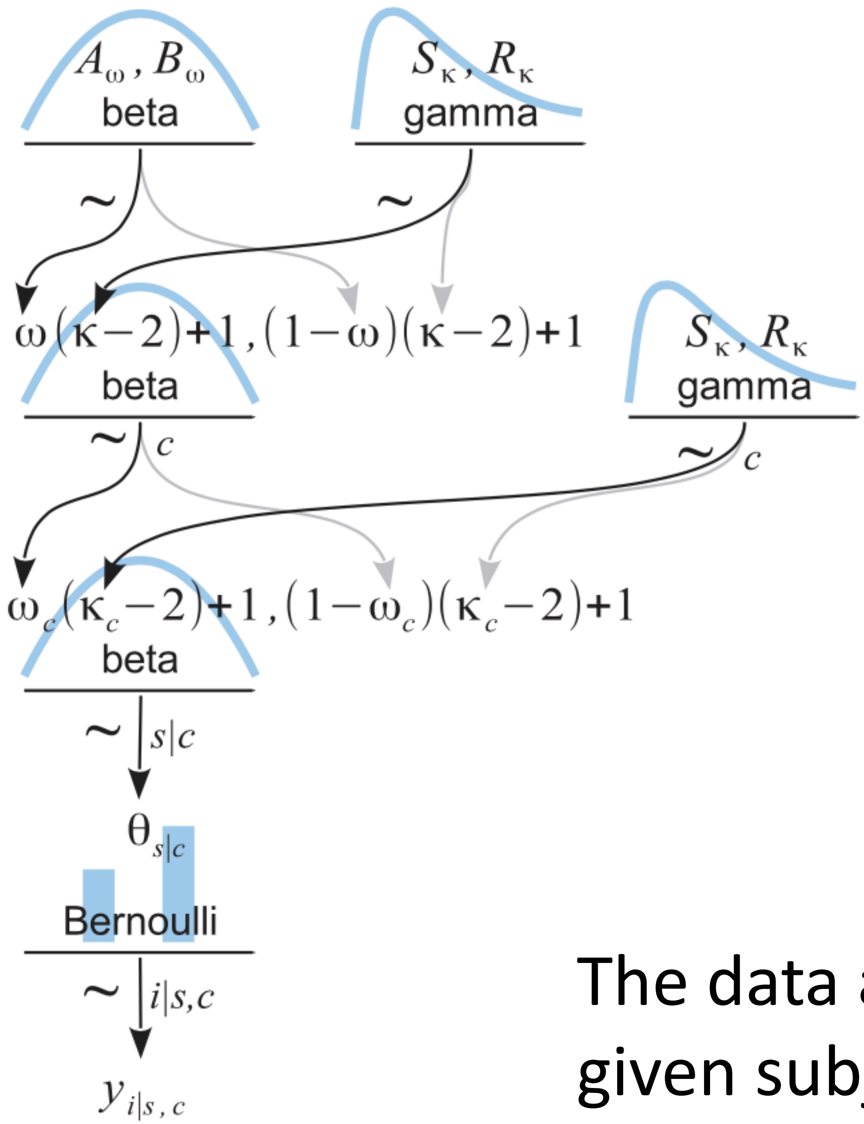
A conservative analysis might treat all players as separate, ignoring the fact that guards are similar to one another and centers are similar to one another.

A middle path: *Fit a hierarchical model*.

Denote each player's ability with a parameter θ_{Player}, let the θs for the centers come from a Beta(a_1,b_1) distribution and the θs for the guards come from a Beta(a_2,b_2) distribution, and let the parameters of the two Betas come from hyperpriors that describe prior belief about typical free throw shooting success.

E.g., use a Beta(30,10) for the mode and a diffuse gamma distribution for the concentration. (So we expect around 75% success.)

Krsuschke schematic – Figure 9.13 from DBDA 2ed



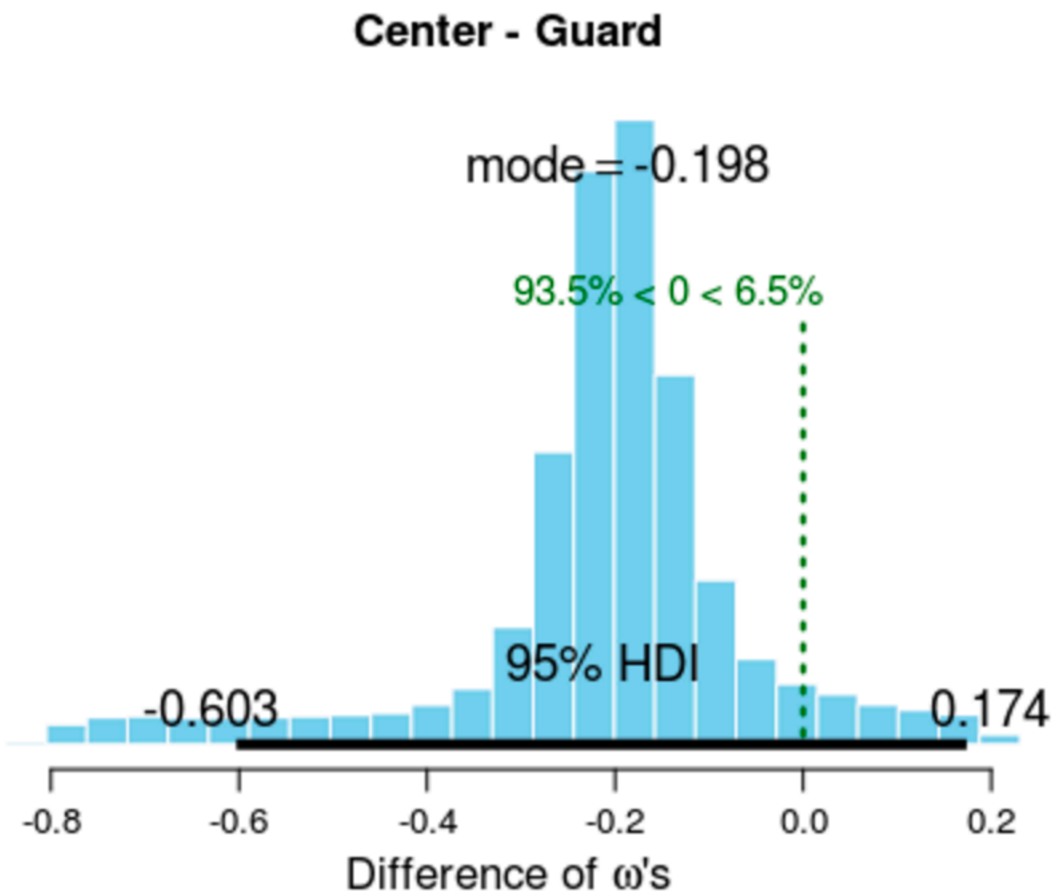
There are categories (c) – guard or center.

There are subjects (s) within categories – guard1, guard2, etc.

There is success or failure on a trial (i) – hit or miss a FT.

The data are Bernoulli, for a given subject in a given category.

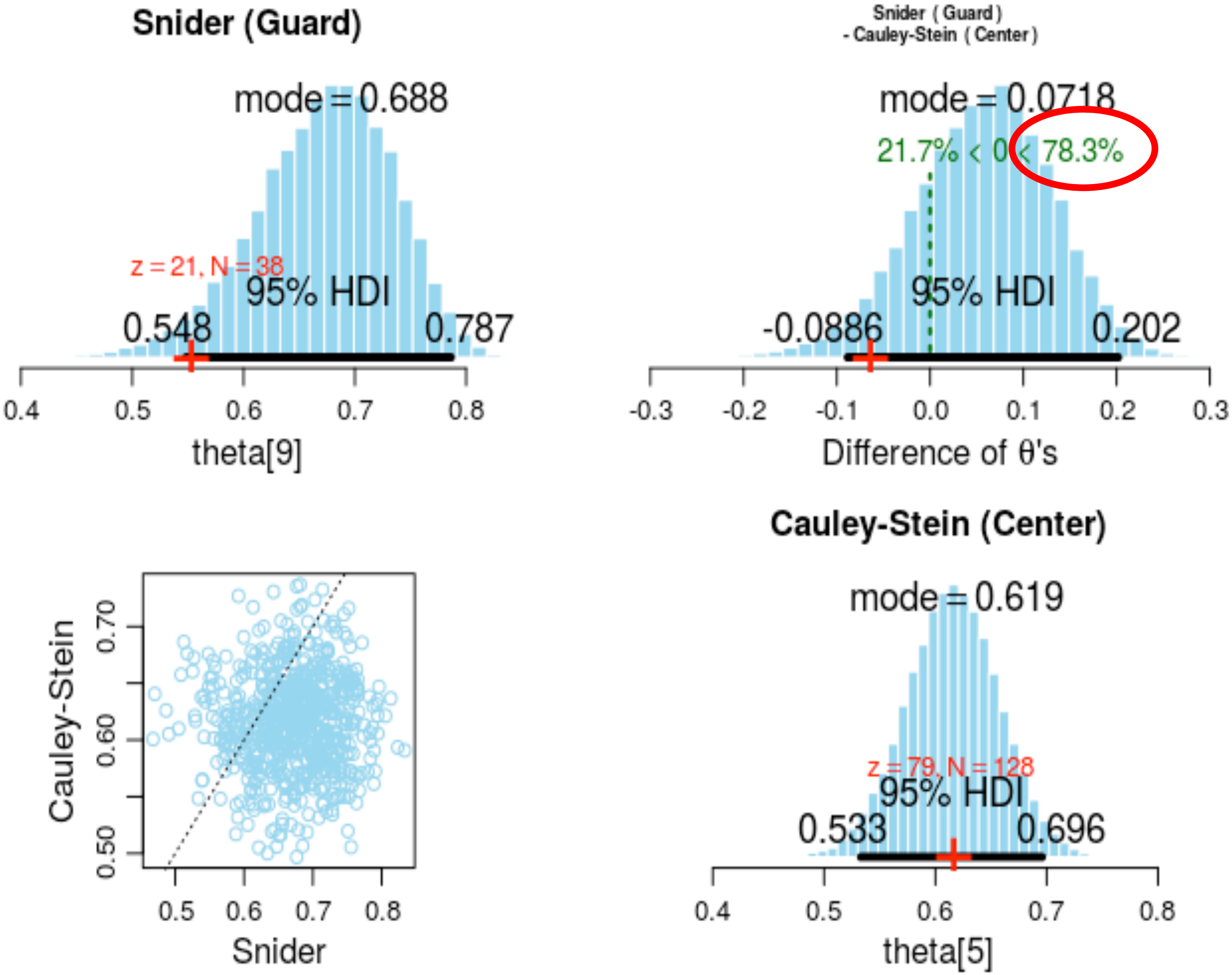
The posterior distributions tell us that there is a 93.5% chance that in general guards are better free throw shooters than centers.



Willie Cauley-Stein, a center, made 79 of 128 free throws, for a **61.7%** success rate.

Quentin Snider, a guard, made 21 of 38 free throws, for a **55.3%** success rate.

The small number of attempts by Snider combined with the fact that he is a guard suggests that in the long run he will do quite a bit better than 55.3% and we find that *there is a* ***78% posterior probability*** *that Snider is more skilled than Cauley-Stein at shooting free throws*.



Snider (Guard)
mode = 0.688
z = 21, N = 38
95% HDI
0.548
0.787
0.4
0.5
0.6
0.7
0.8
theta[9]
Snider (Guard)
- Cauley-Stein (Center)
mode = 0.0718
21.7% < 0 < 78.3%
95% HDI
-0.0886
0.202
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
Difference of θ's
Cauley-Stein
0.50
0.60
0.70
0.5
0.6
0.7
0.8
Snider
Cauley-Stein (Center)
mode = 0.619
z = 79, N = 128
95% HDI
0.533
0.696
0.4
0.5
0.6
0.7
0.8
theta[5]

During the 2015-16 season, Snider made 45 of 60 free throws (**75%**) while Cauley-Stein made 81 of 125 free throws (**64.8%**). Given the Bayesian analysis above, this Snider advantage comes as no surprise.

For the following four seasons Snider shot **72.0%**, **88.6%**, **75.9%**, and **84.0**% on free throws while Cauley-Stein shot **66.9%**, **61.9%**, **55.1%**, and **60.6%**.