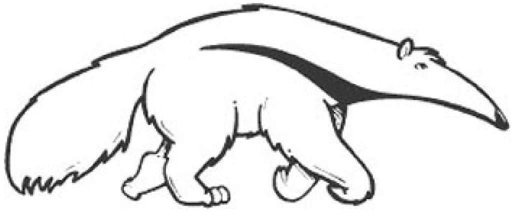
Bayesian benefits for the pragmatic researcher

Michael Lee
mdlee@uci.edu
Department of Cognitive Sciences,
University of California, Irvine



Recent paper

- Wagenmakers, E.-J., Morey, R.D., & Lee, M.D. (2016). Bayesian benefits for the pragmatic researcher. *Current Directions in Psychological Science, 25,* 169-176

EJ Wagenmakers

Richard Morey

Quentin Gronau

Outline

- Bayesian inference
- Bayesian parameter estimation
- Example: Bob's IQ
- Bayesian hypothesis testing
- Example: Adam Sandler

Bayesian inference

- In Bayesian inference, uncertainty or degree of belief is quantified by probability.
- **Prior** beliefs are updated by means of the data to yield **posterior** beliefs.

Outline

- Bayesian inference
- <u>Bayesian parameter estimation</u>
- Example: Bob's IQ
- Bayesian hypothesis testing
- Example: Adam Sandler

Bayesian parameter estimation: Example

- You take an exam in which you have to answer 10 factual questions of equal difficulty.
- You answer 9 out of 10 questions correctly.
- What is your latent probability θ of answering any one question correctly?

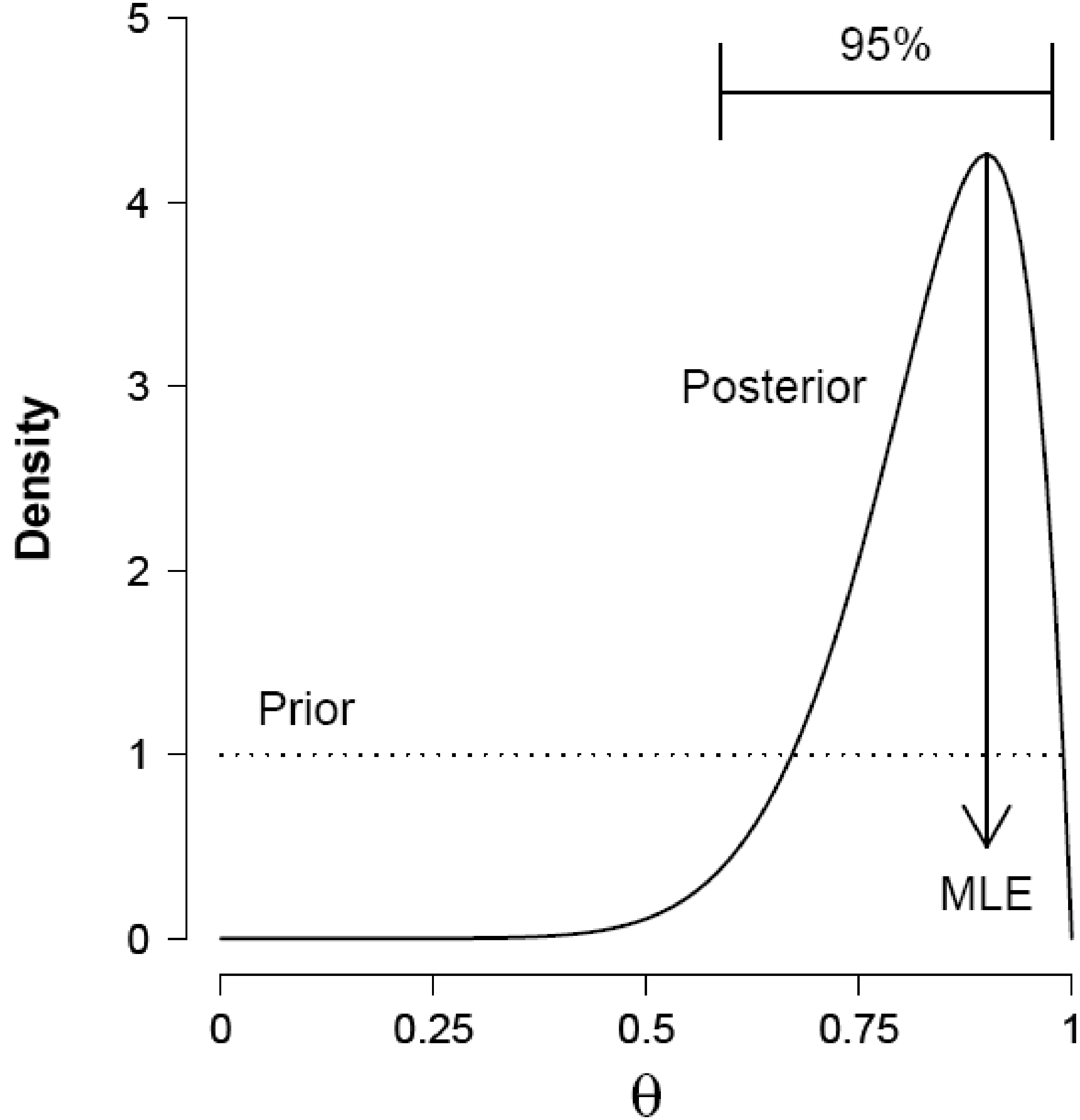
Bayesian parameter estimation: Example

- We start with a **prior distribution** for θ. This reflect all we know about θ prior to the experiment. Here we make a standard choice and assume that all values of θ are equally likely *a priori*.

Bayesian parameter estimation: Example

- We then update the prior distribution by means of the data (technically, the *likelihood*) to arrive at a **posterior distribution.**
- The posterior distribution is a compromise between what we knew before the experiment and what we have learned from the experiment
  - The posterior distribution reflects all that we now know about θ.

Prior and posterior distribution



What we knew to start

$$\underbrace{p(\theta)}_{\substack{\text{Prior beliefs} \\ \text{about parameters}}}$$

What we knew after the data

$$\underbrace{p(\theta \mid \text{data})}_{\substack{\text{Posterior beliefs} \\ \text{about parameters}}}$$

How the data updated our knowledge

$$\underbrace{\frac{p(\text{data} \mid \theta)}{p(\text{data})}}_{\substack{\text{Predictive} \\ \text{updating factor}}}$$

Bayesian parameter estimation

$$\underbrace{p(\theta \mid \text{data})}_{\substack{\text{Posterior beliefs}\\ \text{about parameters}}} = \underbrace{p(\theta)}_{\substack{\text{Prior beliefs}\\ \text{about parameters}}} \times \underbrace{\frac{p(\text{data} \mid \theta)}{p(\text{data})}}_{\substack{\text{Predictive}\\ \text{updating factor}}}$$

Advantages of Bayesian parameter estimation

- Knowledge about plausible values are reallocated according to relative predictive success
- Knowledge is updated coherently when new data arrive
- Inferences can incorporate important earlier information
- Researchers are able to address questions that are practically relevant

Outline

- Bayesian inference
- Bayesian parameter estimation
- <u>Example: Bob's IQ</u>
- Bayesian hypothesis testing
- Example: Adam Sandler

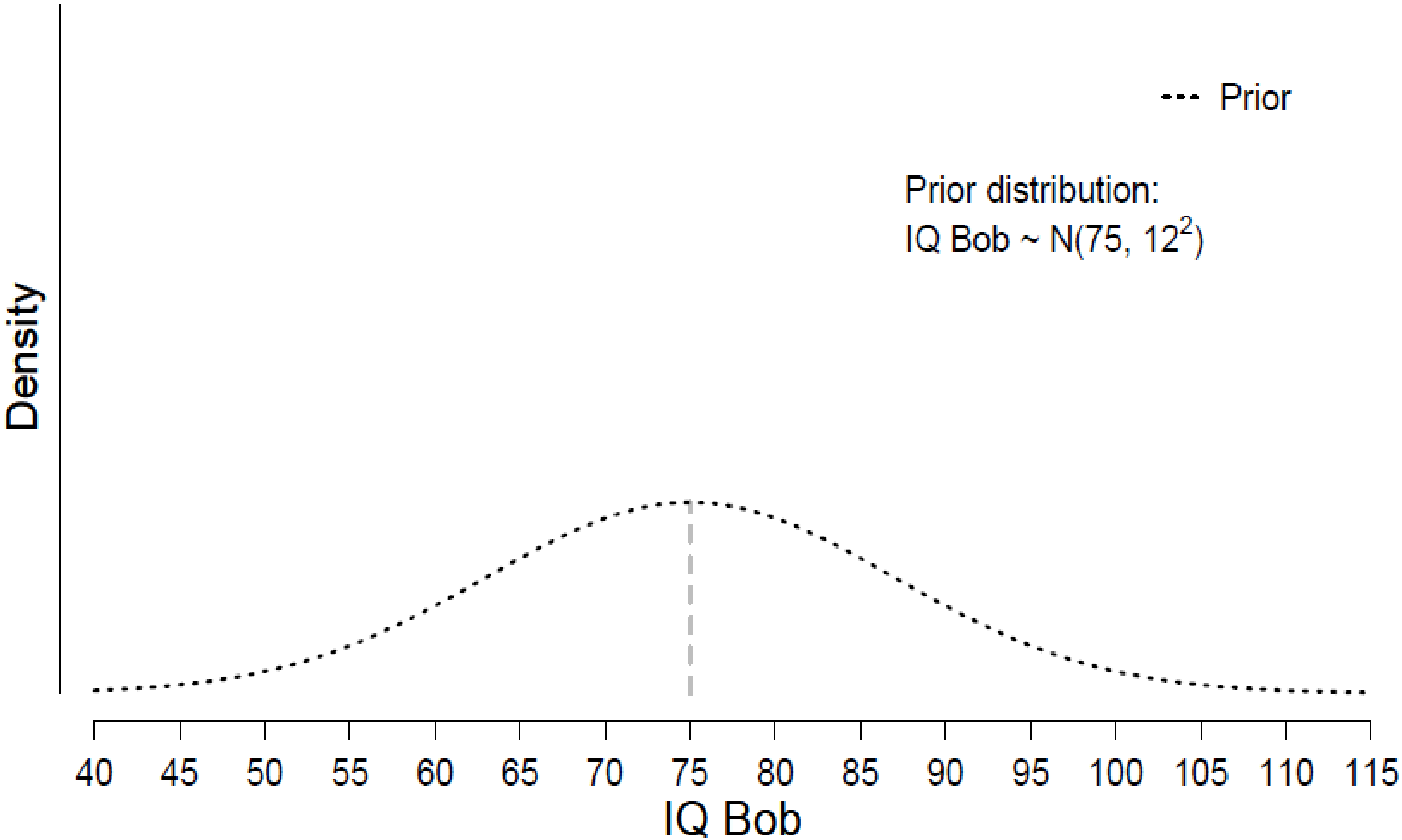
Bob is a killer

- “Florida Bob” has murdered his wife and faces the death sentence.
- The defence argues that Bob is cannot be held fully responsible because he is intellectually disabled
  - his IQ is presumably lower than 70.
- The judge rules that three IQ test be administered
  - The results:{73, 67, 79}.
- What can we conclude?

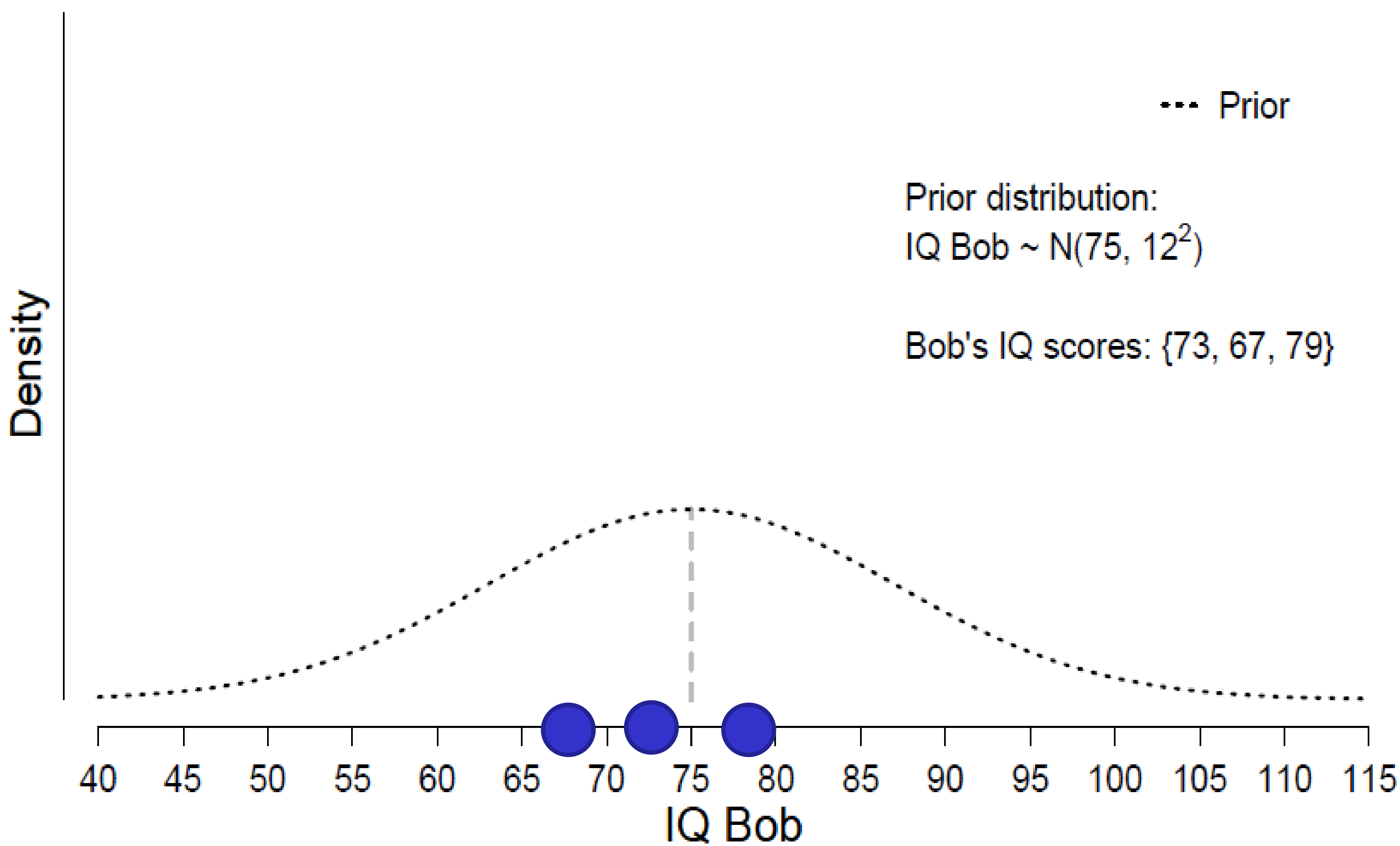
Background knowledge and assumptions

- Based on the literature, inmates classified as intellectually disabled have an IQ ~ N(75, σ=12).
  - We use this to quantify our prior uncertainty about Bob's IQ.
- Based on the literature, we assign the reliability of an IQ test
  - We give the standard deviation of the test a uniform prior from 5 to 15 points

Prior distribution

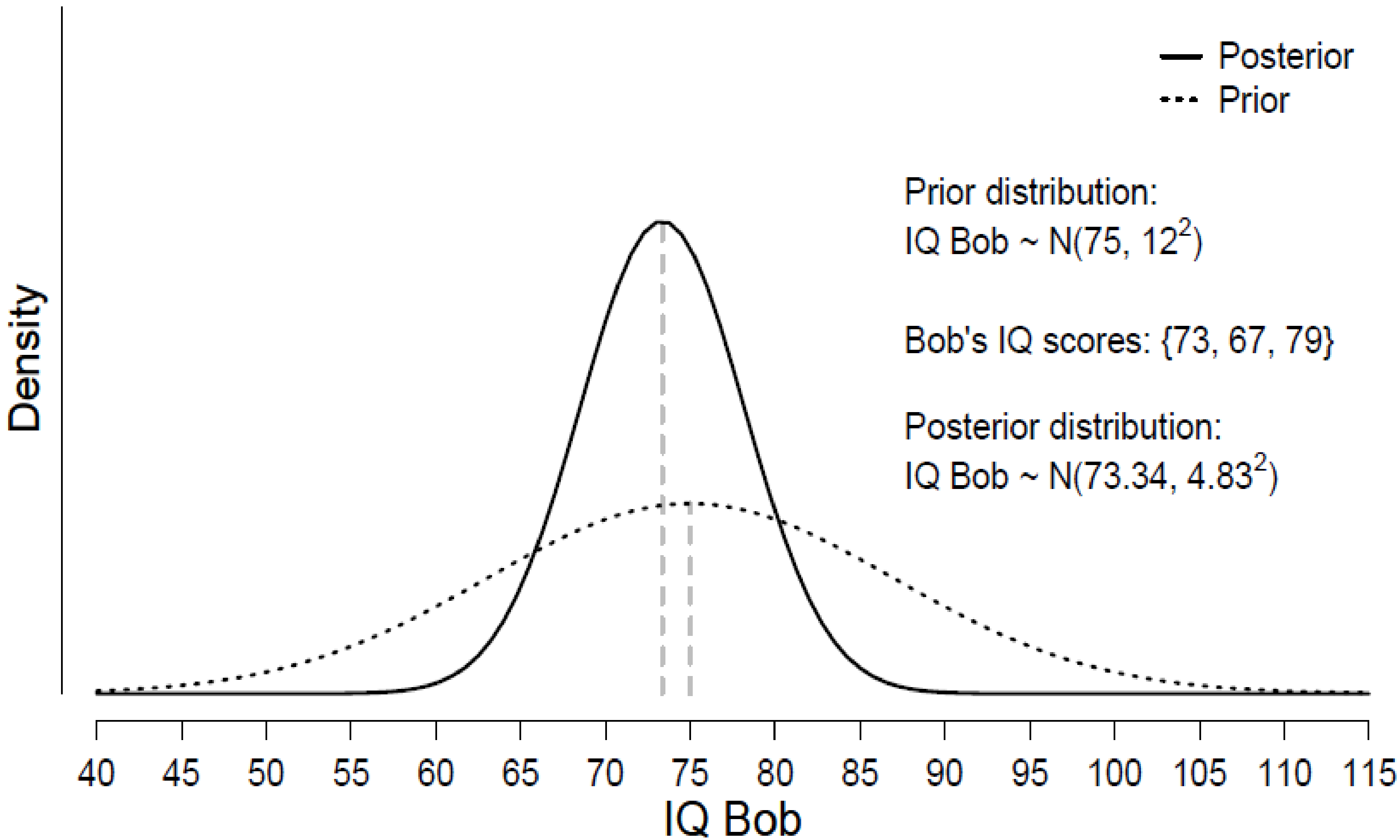


New data

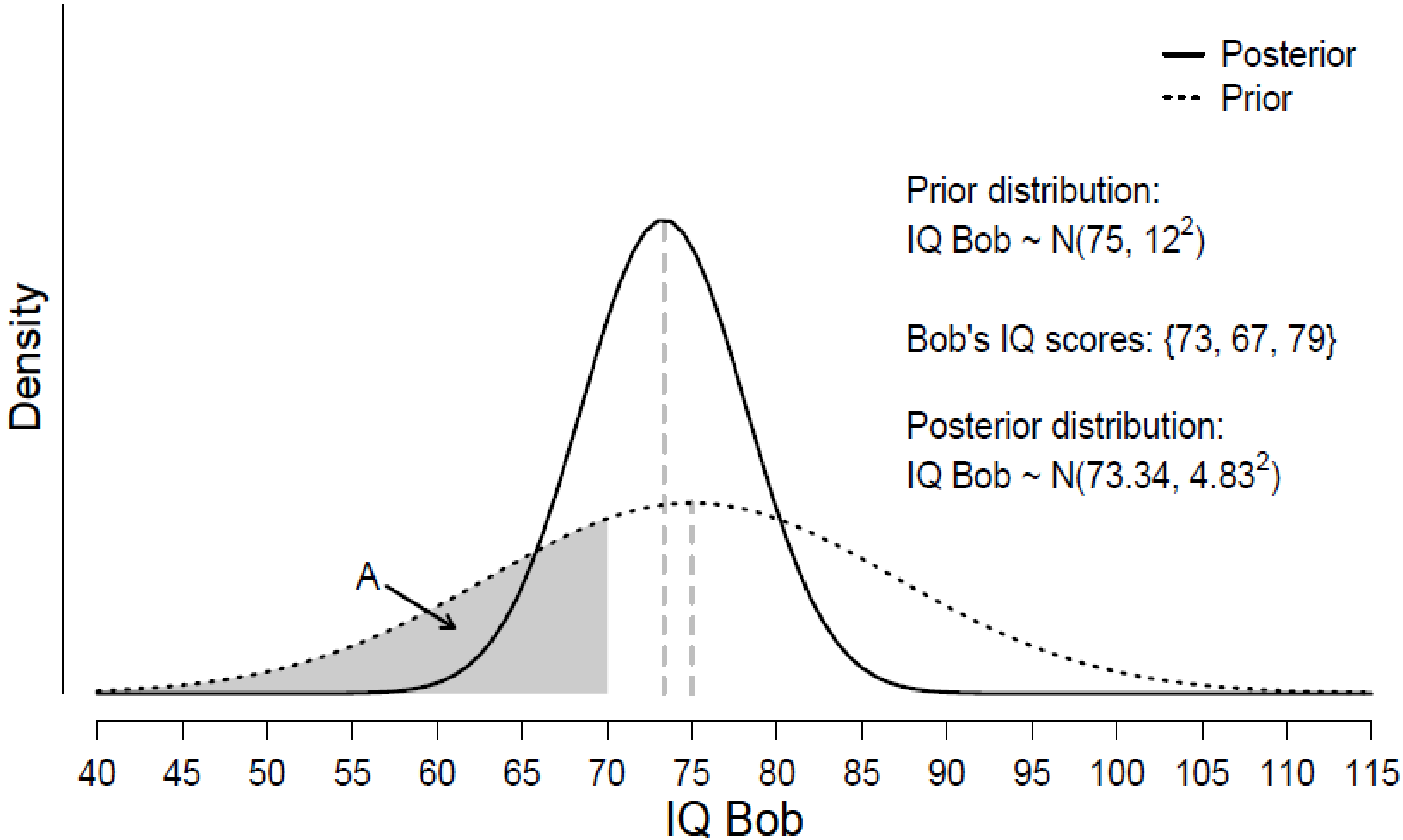


Prior
Prior distribution:
IQ Bob ~ N(75, 12²)
Bob's IQ scores: {73, 67, 79}
Density
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
IQ Bob

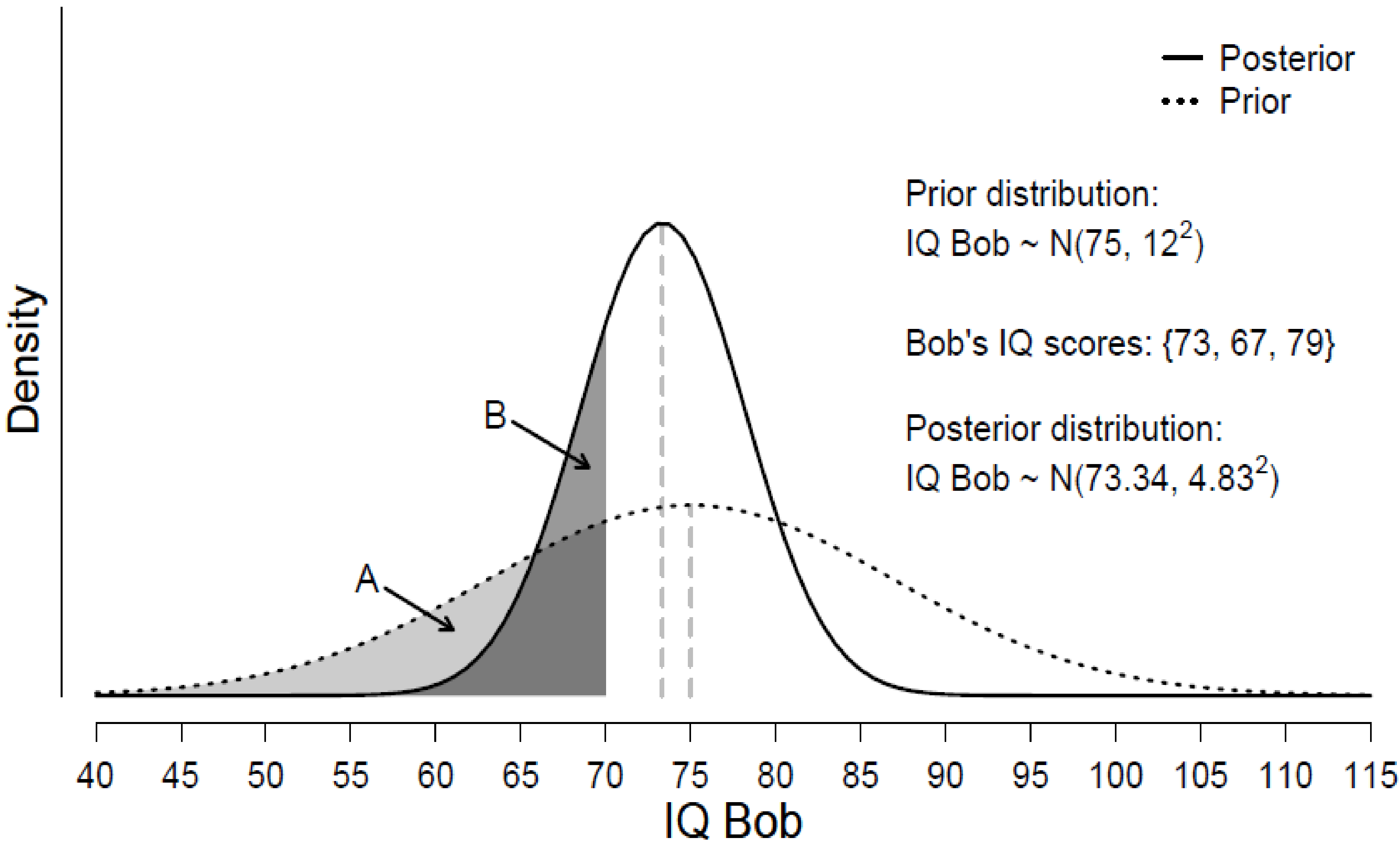
Posterior distribution



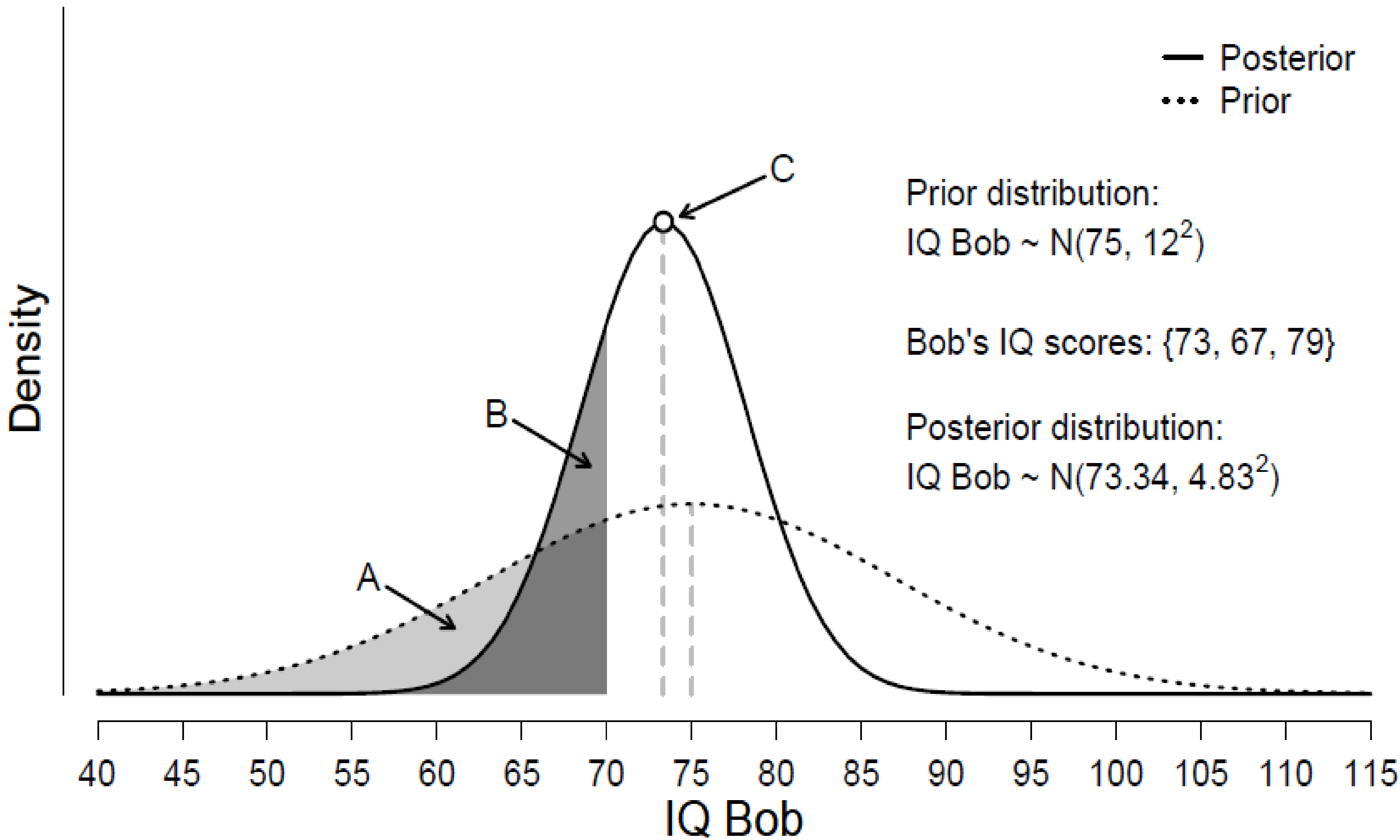
A = Prior probability Bob’s IQ < 70



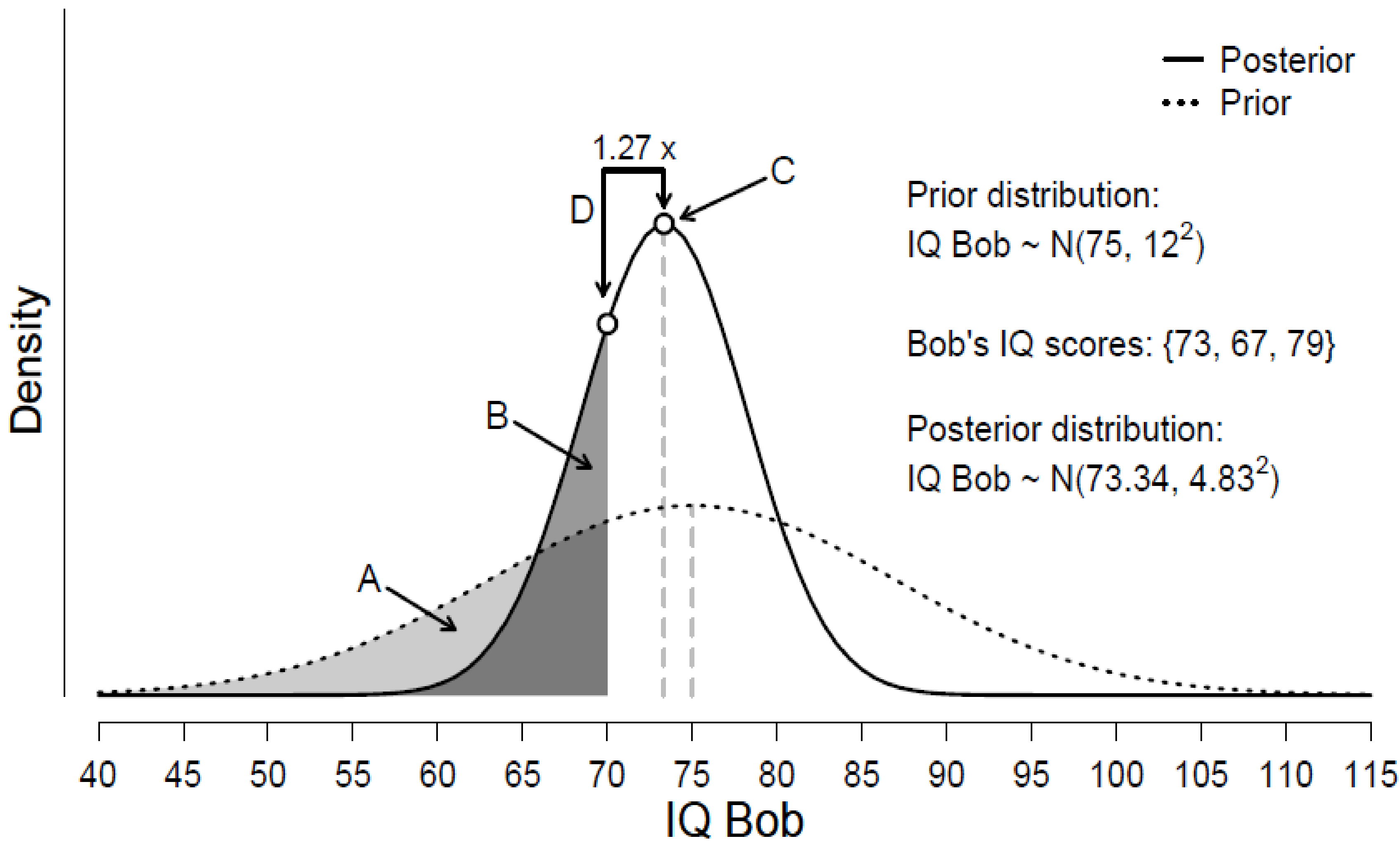
B = Posterior probability Bob's IQ < 70



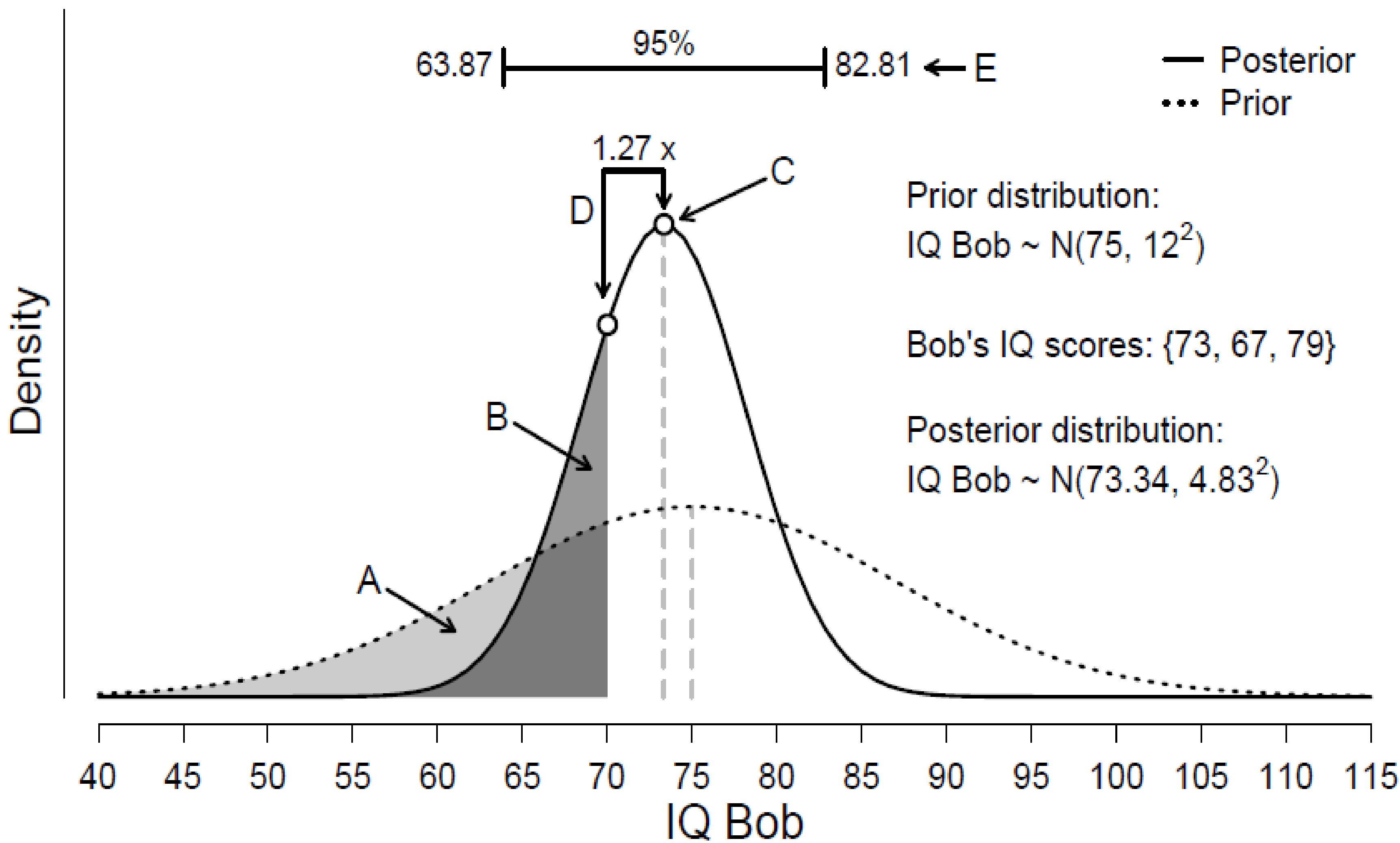
C = Most likely value of Bob's IQ



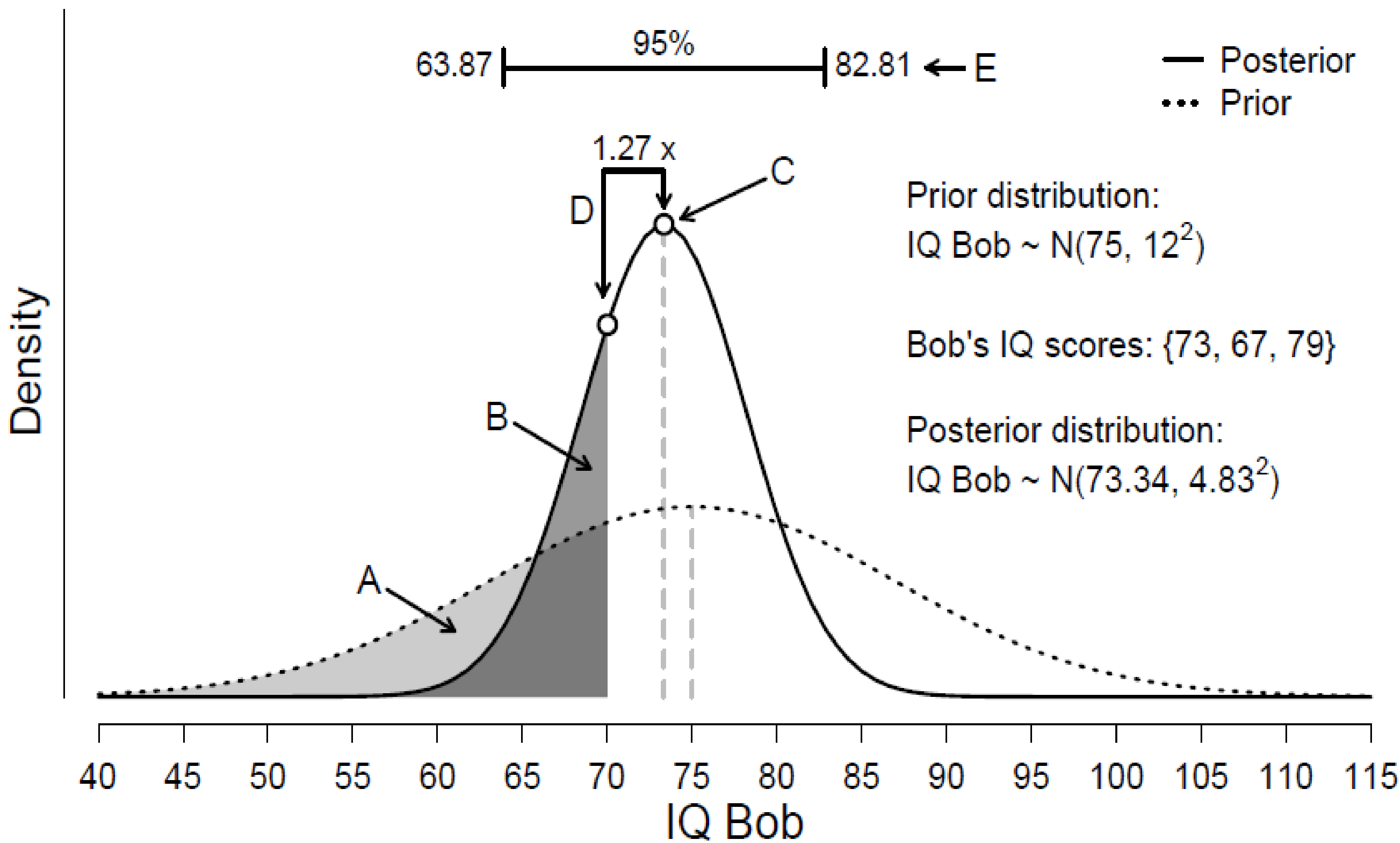
D = How much more likely Bob's IQ is 73 than 70



E = 95% Credible interval for Bob's IQ



NONE of statements A-E can be made in classical framework!



Outline

- Bayesian inference
- Bayesian parameter estimation
- Example: Bob's IQ
- <u>Bayesian hypothesis testing</u>
- Example: Adam Sandler

Bayesian hypothesis test

- Suppose we have two models, H_0 and H_1.
- Which model is better supported by the data?
- The model that *predicted* the data best!
- The ratio of predictive performance is known as the **Bayes factor** (Jeffreys, 1961).

What we already knew about models

$$\underbrace{\frac{p(\mathcal{H}_1)}{p(\mathcal{H}_0)}}_{\substack{\text{Prior beliefs} \\ \text{about hypotheses}}}$$

What we knew about models after seeing data

$$\underbrace{\frac{p(\mathcal{H}_1 \mid \text{data})}{p(\mathcal{H}_0 \mid \text{data})}}_{\substack{\text{Posterior beliefs} \\ \text{about hypotheses}}}$$

How we used the data to update

$$\underbrace{\frac{p(\text{data} \mid \mathcal{H}_1)}{p(\text{data} \mid \mathcal{H}_0)}}_{\substack{\text{Predictive} \\ \text{updating factor}}}$$

Bayesian hypothesis test

$$\underbrace{\frac{p(\mathcal{H}_1 \mid \text{data})}{p(\mathcal{H}_0 \mid \text{data})}}_{\substack{\text{Posterior beliefs} \\ \text{about hypotheses}}} = \underbrace{\frac{p(\mathcal{H}_1)}{p(\mathcal{H}_0)}}_{\substack{\text{Prior beliefs} \\ \text{about hypotheses}}} \times \underbrace{\frac{p(\text{data} \mid \mathcal{H}_1)}{p(\text{data} \mid \mathcal{H}_0)}}_{\substack{\text{Predictive} \\ \text{updating factor}}}$$

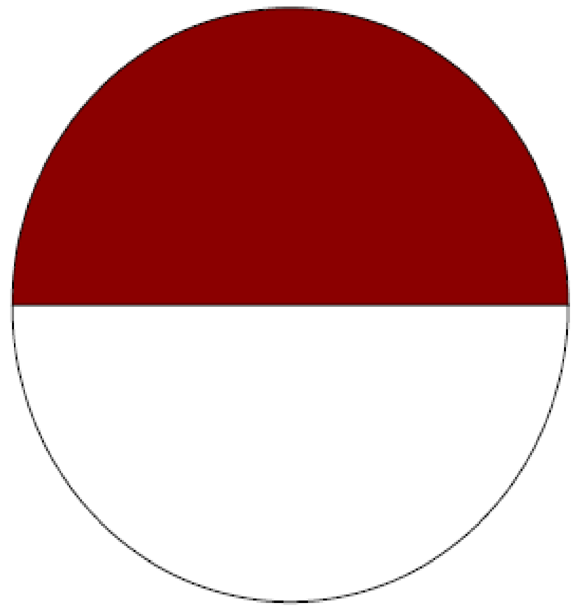
Guidelines for interpreting the Bayes factor

BF	Evidence
1 – 3	Anecdotal
3 – 10	Moderate
10 – 30	Strong
30 – 100	Very strong
>100	Extreme

Visual interpretation of the Bayes factor

$p(\text{data} \mid \mathcal{H}_1)$

$\text{BF}_{10} = 1$



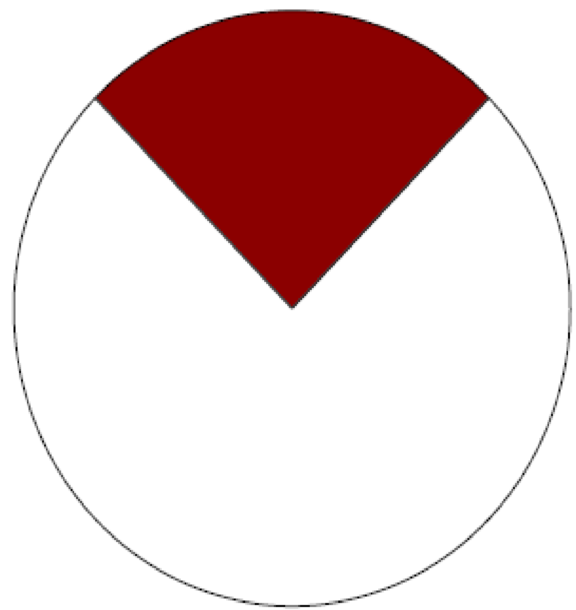
$p(\text{data} \mid \mathcal{H}_0)$

Visual interpretation of the Bayes factor

$$p(\text{data} \mid \mathcal{H}_1)$$

$$\mathrm{BF}_{10} = \frac{1}{3}$$

$$\mathrm{BF}_{01} = 3$$

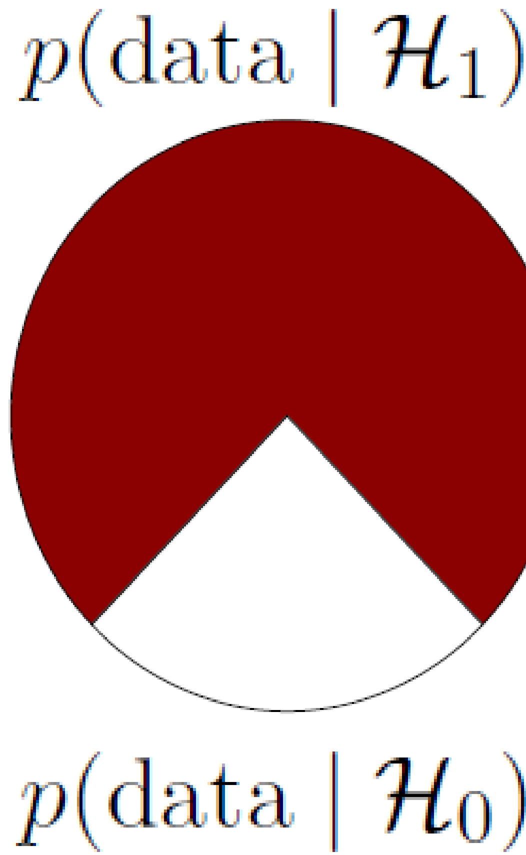


$$p(\text{data} \mid \mathcal{H}_0)$$

Visual interpretation of the Bayes factor

$$\mathrm{BF}_{10} = 3$$

$$\mathrm{BF}_{01} = \frac{1}{3}$$



Advantages of the Bayes factor

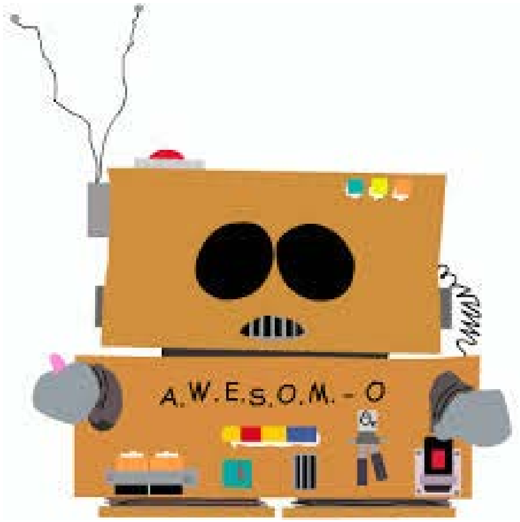
- Quantifies evidence instead of forcing an all-or-none decision
- Discriminates “evidence of absence” from “absence of evidence”
  - In particular, allows evidence to be found **for** the null hypothesis
- Allows evidence to be monitored as data accumulate
- Applies to data from the real world, for which no sampling plan can be articulated

Outline

- Bayesian inference
- Bayesian parameter estimation
- Example: Bob's IQ
- Bayesian hypothesis testing
- <u>Example: Adam Sandler</u>

AWESOME-O

- Southpark episode 166.
- Eric Cartman pretends to be a robot, the A.W.E.S.O.M.-O 4000.



AWESOME-O

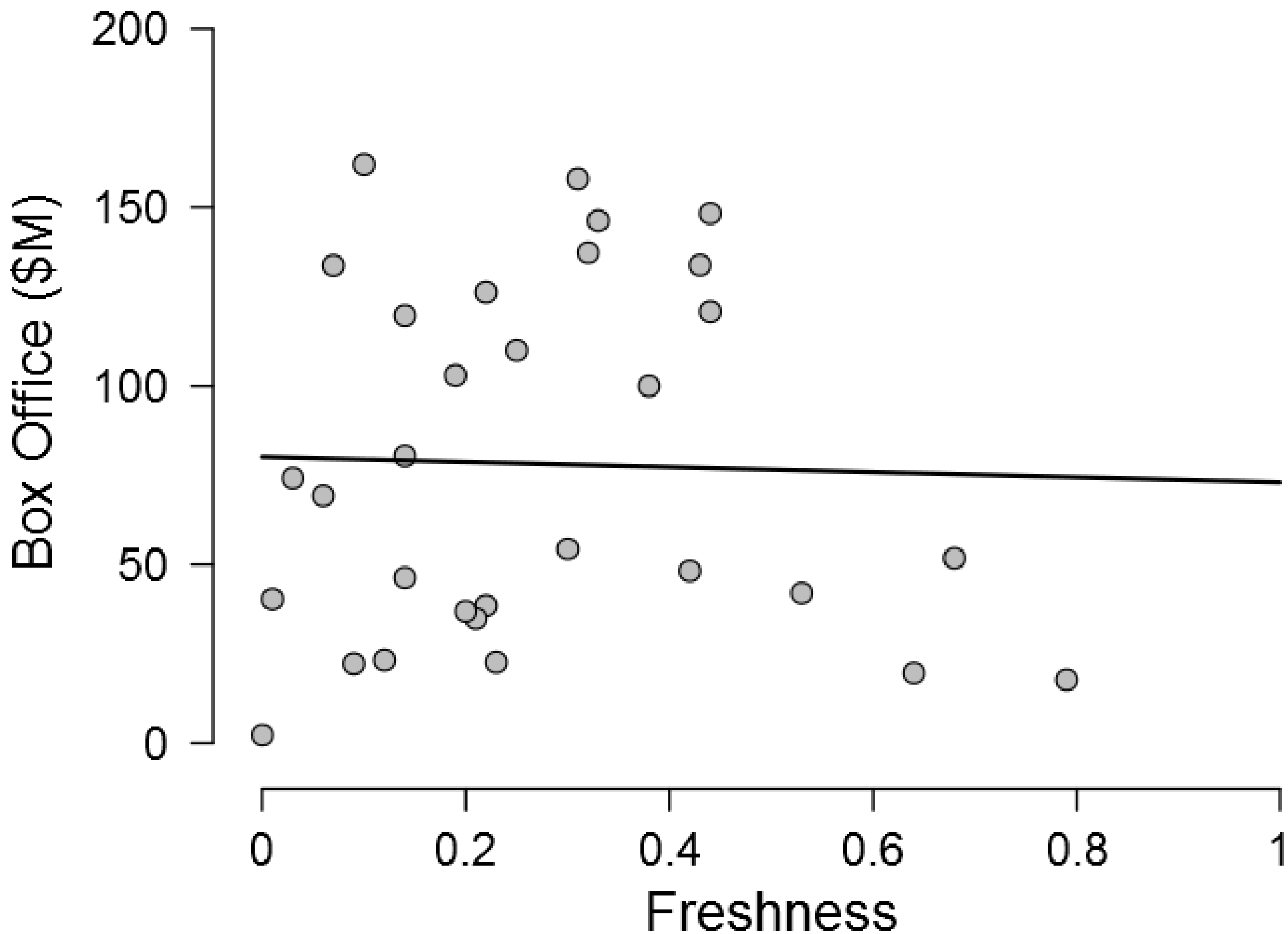
- Hollywood movie-producers kidnap the robot and force it to generate profitable movie ideas.
- The A.W.E.S.O.M.-O 4000 generates more than 2,000 silly movie ideas, 800 of which star Adam Sandler

South Park hypothesis (implied)

- General: “Adam Sandler movies are profitable regardless of their quality”
- Specific: “For Adam Sandler movies, box office success does not correlate with freshness ratings on Rotten Tomatoes”

Statistical inference question

- Is there a correlation or not?



Demo in JASP
(jasp-stats.org)

Outline

- Bayesian inference
- Bayesian parameter estimation
- Example: Bob's IQ
- Bayesian hypothesis testing
- Example: Adam Sandler