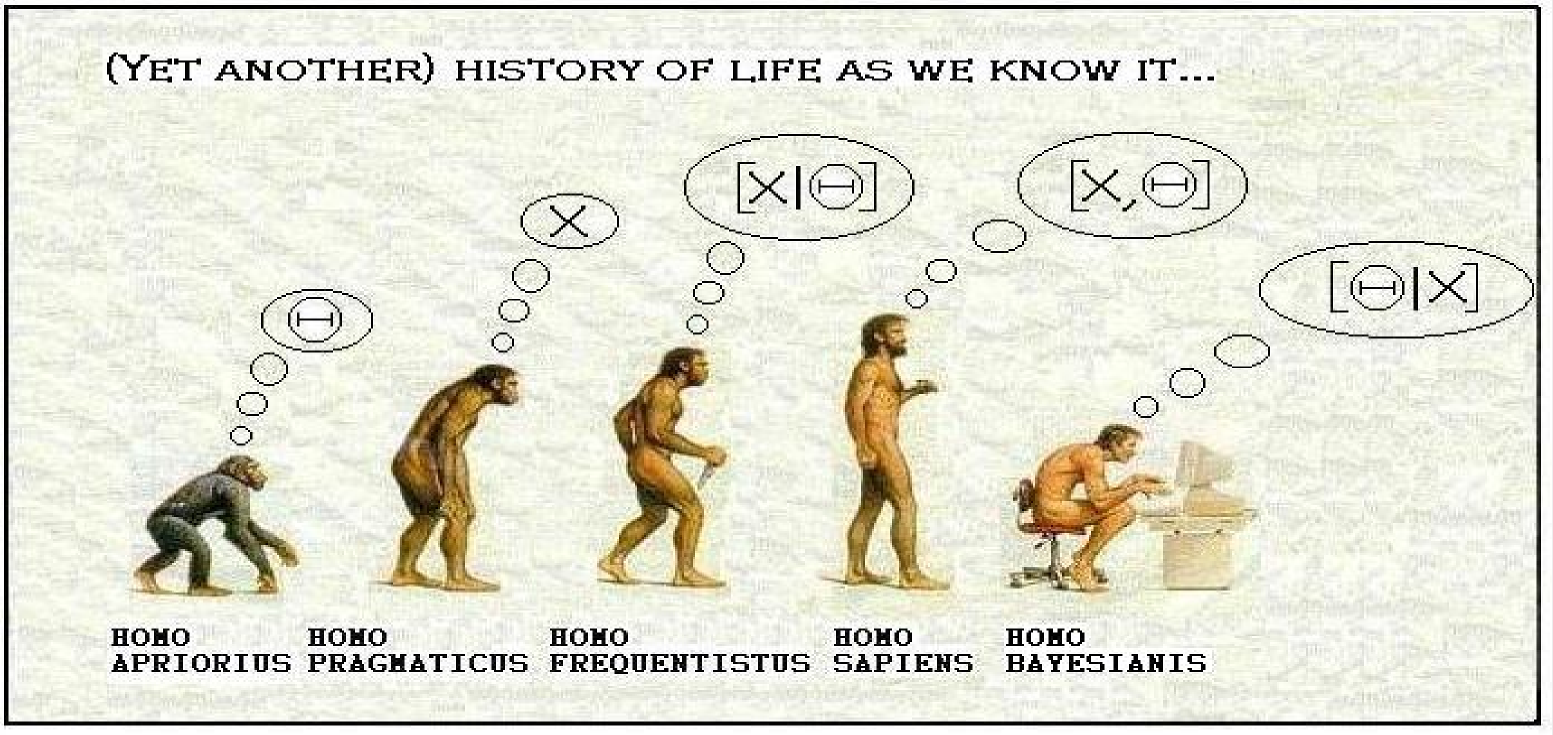
STATS331



Credit: Unknown to me at this time

Introduction to Bayesian Statistics
Semester 2, 2016

Today's Lecture

Point and Interval Estimation

or

Summarising Posterior Distributions

The Posterior is It

- In Bayesian stats, the posterior distribution is the complete answer

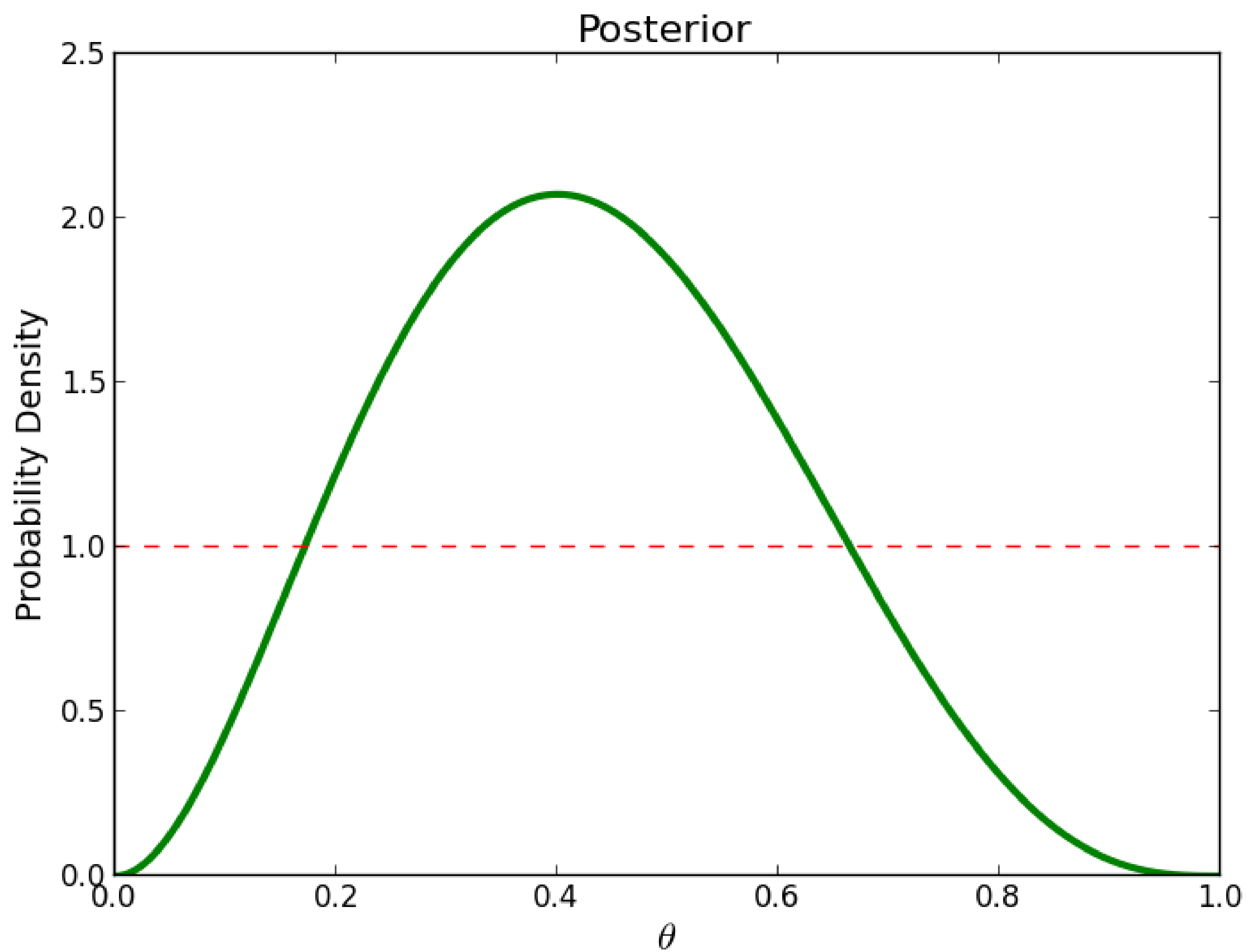
$$
\begin{aligned}
p(\theta|x) &= \frac{p(\theta)p(x|\theta)}{p(x)} \\
p(\theta|x) &\propto p(\theta)p(x|\theta) \\
\texttt{posterior} &\propto \texttt{prior} \times \texttt{likelihood}.
\end{aligned}
$$

- From this, you can work out the probability of anything

Why We Study This

- Summaries: The full posterior can be “too much information”. Good for communication
- Point of contact with frequentist stats, which has “estimators” and “confidence intervals”

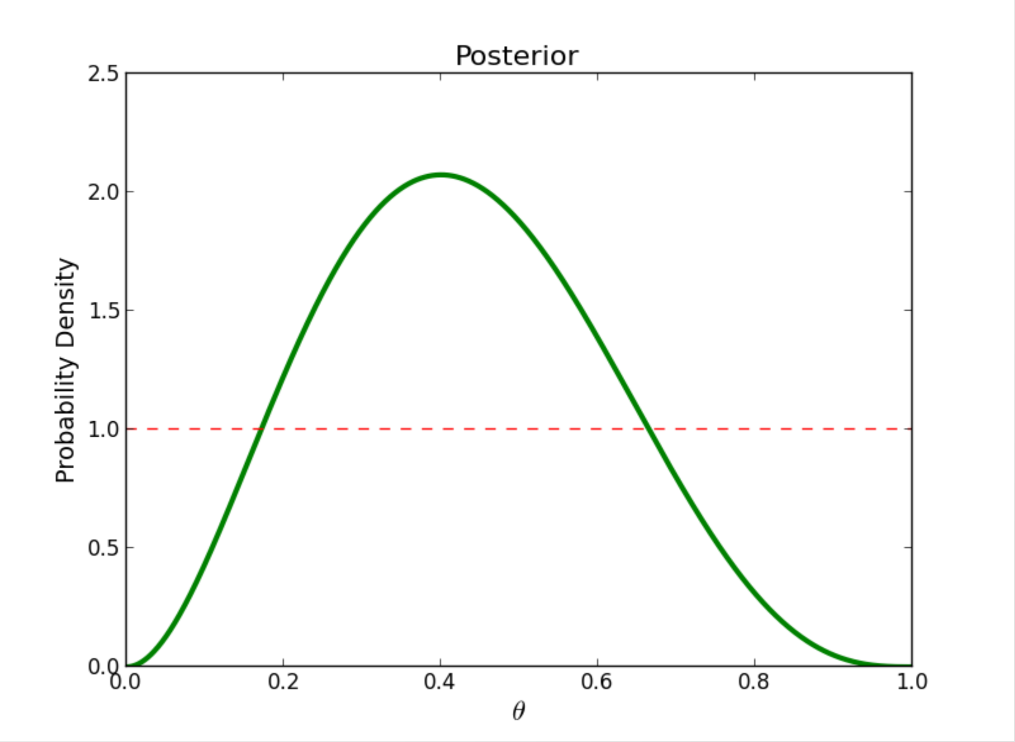
Bus Example


Posterior
Probability Density
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0
θ

Reminder of Bus Results

- Uniform prior
- Binomial sampling distribution to get the likelihoods
- 5 trials
- 2 successes

$$\theta | x \sim \mathrm{Beta}(3, 4)$$



Decision Theory

When you are forced to make a choice, how do you make the *best* choice?

"You acted unwisely," I cried, "as you see

by the outcome." He calmly eyed me:

"When choosing the course of my action," said he, “I had not the outcome to guide me."

Ambrose Bierce (via Ed Jaynes)

But We Have *Some* Information

So some choices are “probably” better than others!

Point Estimation

If we are estimating a parameter θ from data, we could give a guess $\hat{\theta}$

The value that we choose is a *decision*

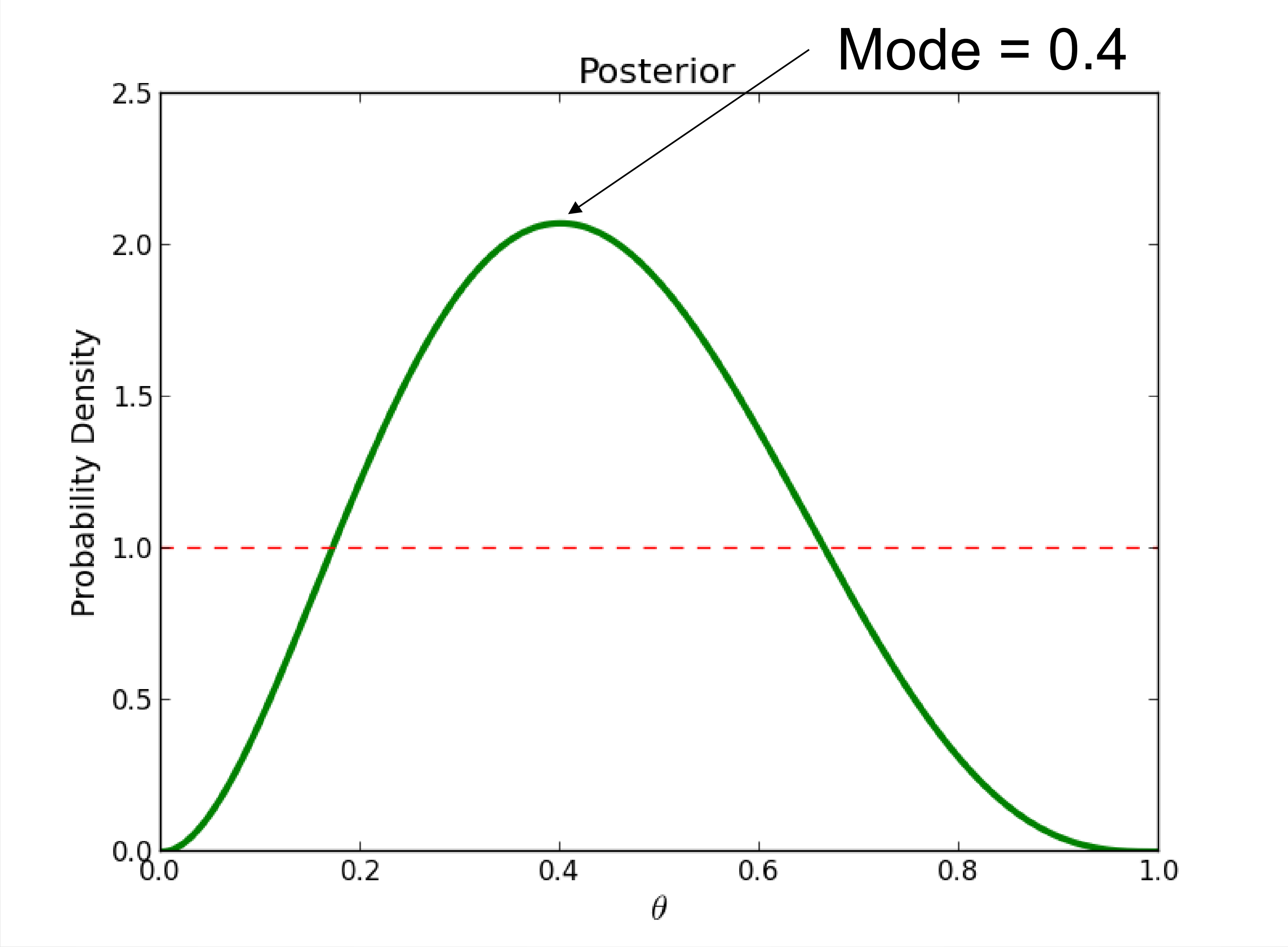
Best Point Estimate

The best estimate would be the true value

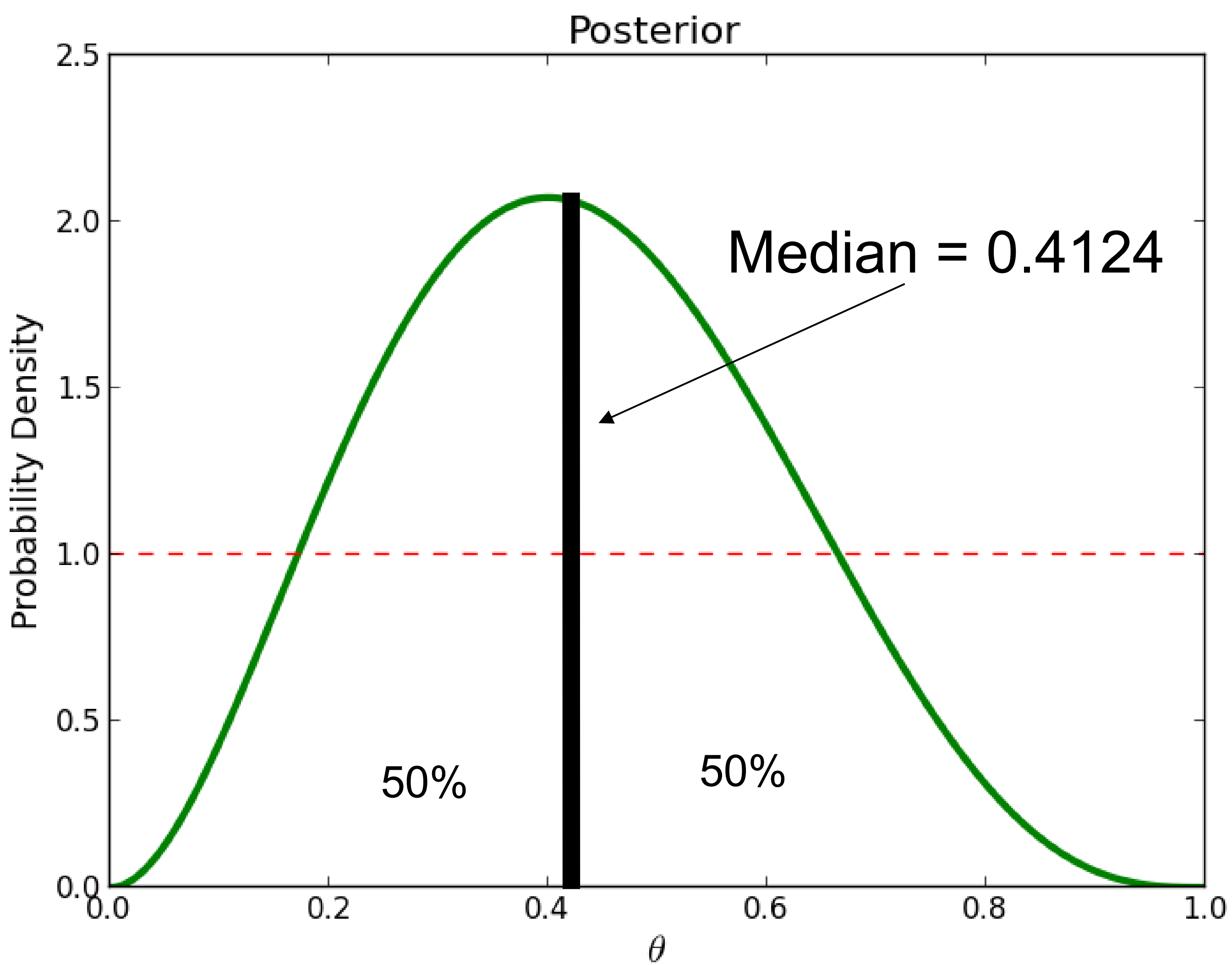
$$\hat{\theta} = \theta$$

Take home message: be wary of any stats method that is "optimal", such statements should always have qualifiers

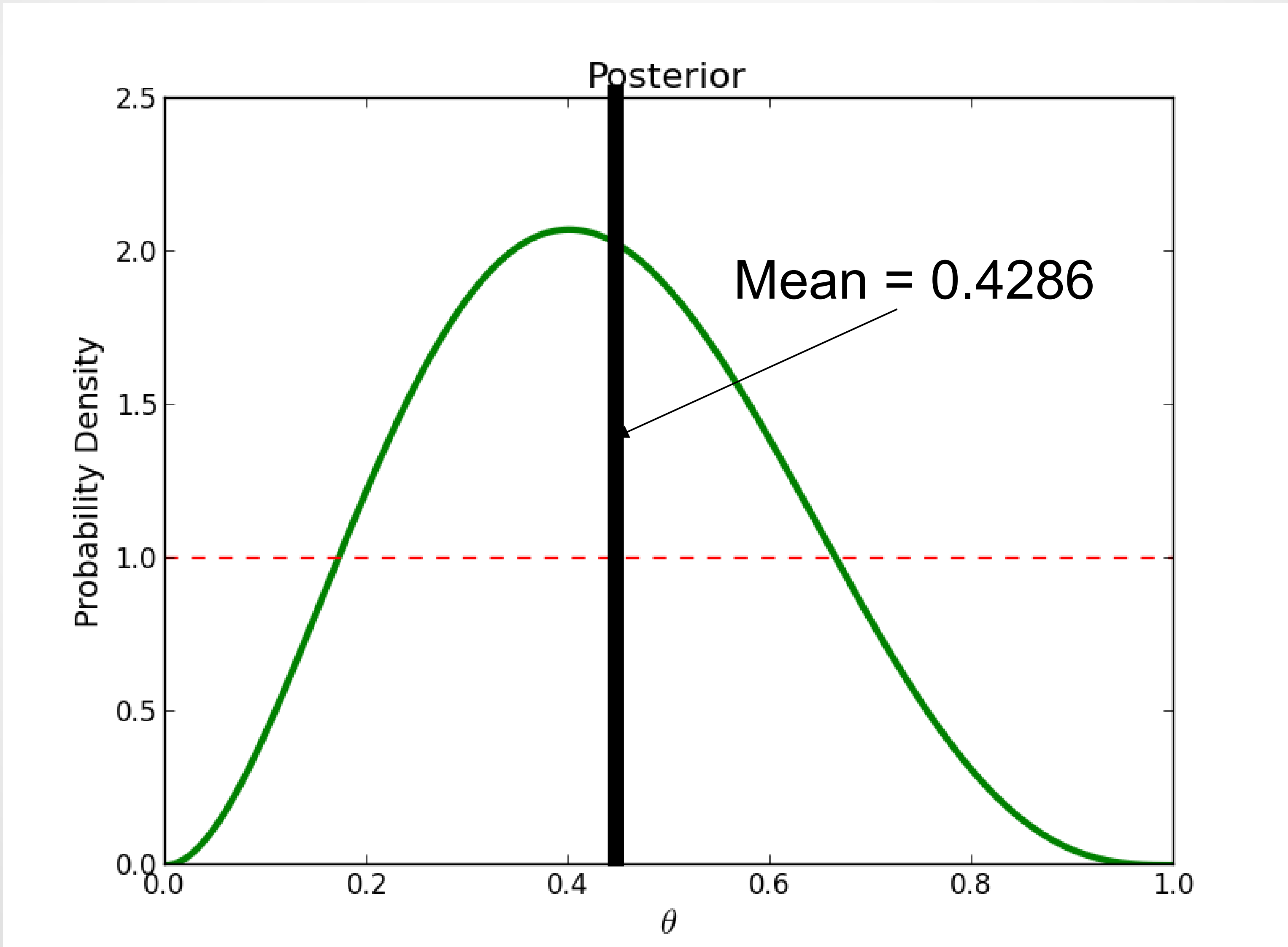
Point Estimate: Mode



Point Estimate: Median



Point Estimate: Mean/Expectation



Calculating Summaries

Calculating posterior mode, mean, and median in R

NOTE: This will be different when we use MCMC/JAGS.

```
# Mean
posterior_mean = sum(theta*posterior)

# Mode
highest_probability = max(posterior)
posterior_mode = theta[posterior == highest_probability]

# Median
F = cumsum(posterior)
dist = abs(F - 0.5)
posterior_median = theta[dist == min(dist)]
```

Choosing Between Them

All of these estimates are well within the range of uncertainty, in this example. But this isn't always true!

Utility

If the true value is θ and I guess $\hat{\theta}$ how "good" is my guess?

Utility function

$$U(\hat{\theta}, \theta)$$

Loss

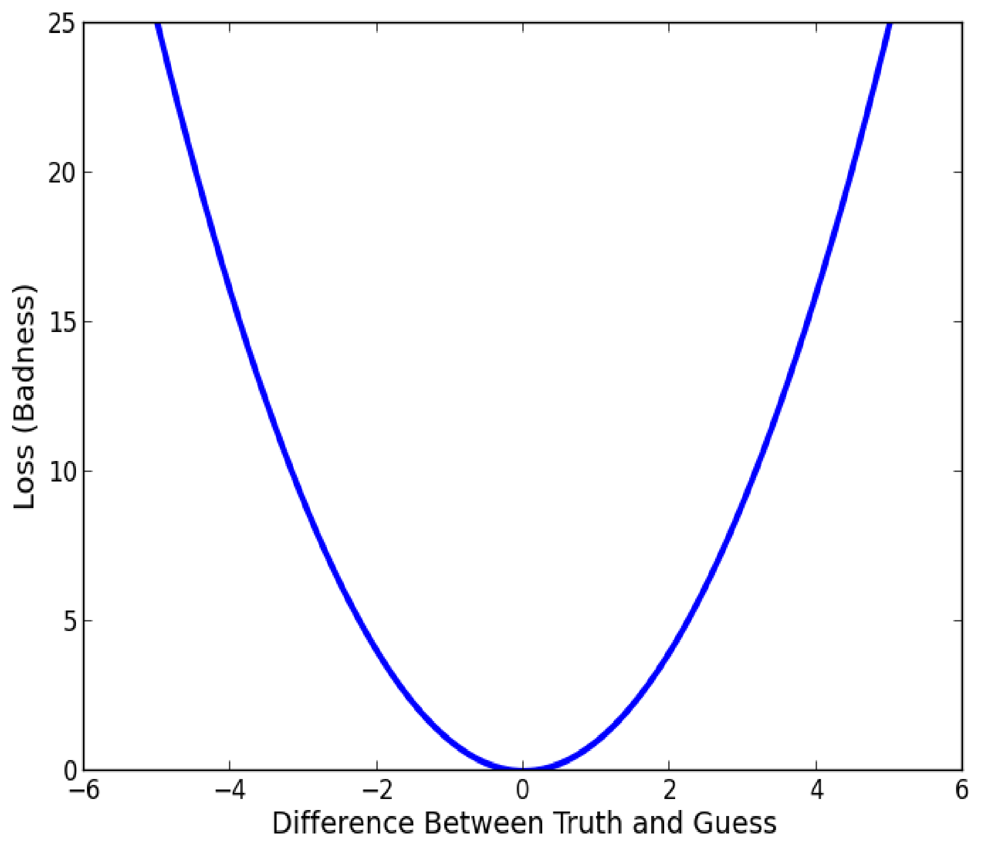
- Loss is just the negative of utility
- For pessimists...

$$L = -U$$

Loss = “badness”

Quadratic Loss

$$L(\hat{\theta}, \theta) = \left(\hat{\theta} - \theta\right)^2$$



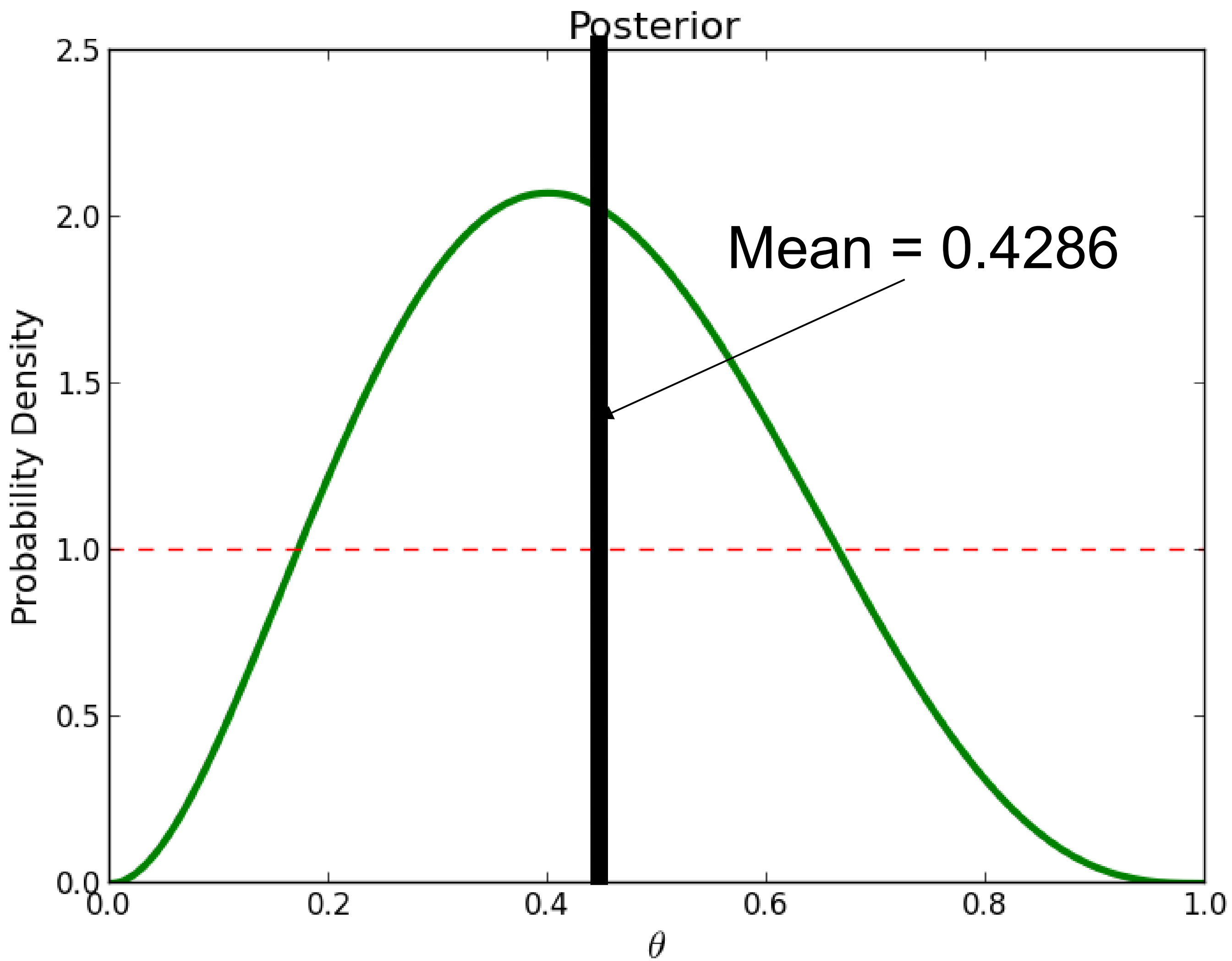
Minimise Loss?

- Can't – need to know true θ value
- Minimise POSTERIOR EXPECTED LOSS

$$E\left[L(\hat{\theta},\theta)\right] = \int p(\theta|x)\left(\hat{\theta}-\theta\right)^2 \, d\theta$$

- Minimise by differentiation wrt estimate and then setting to zero, solve for estimate. See lecture notes for proof

Posterior Mean is the Best!



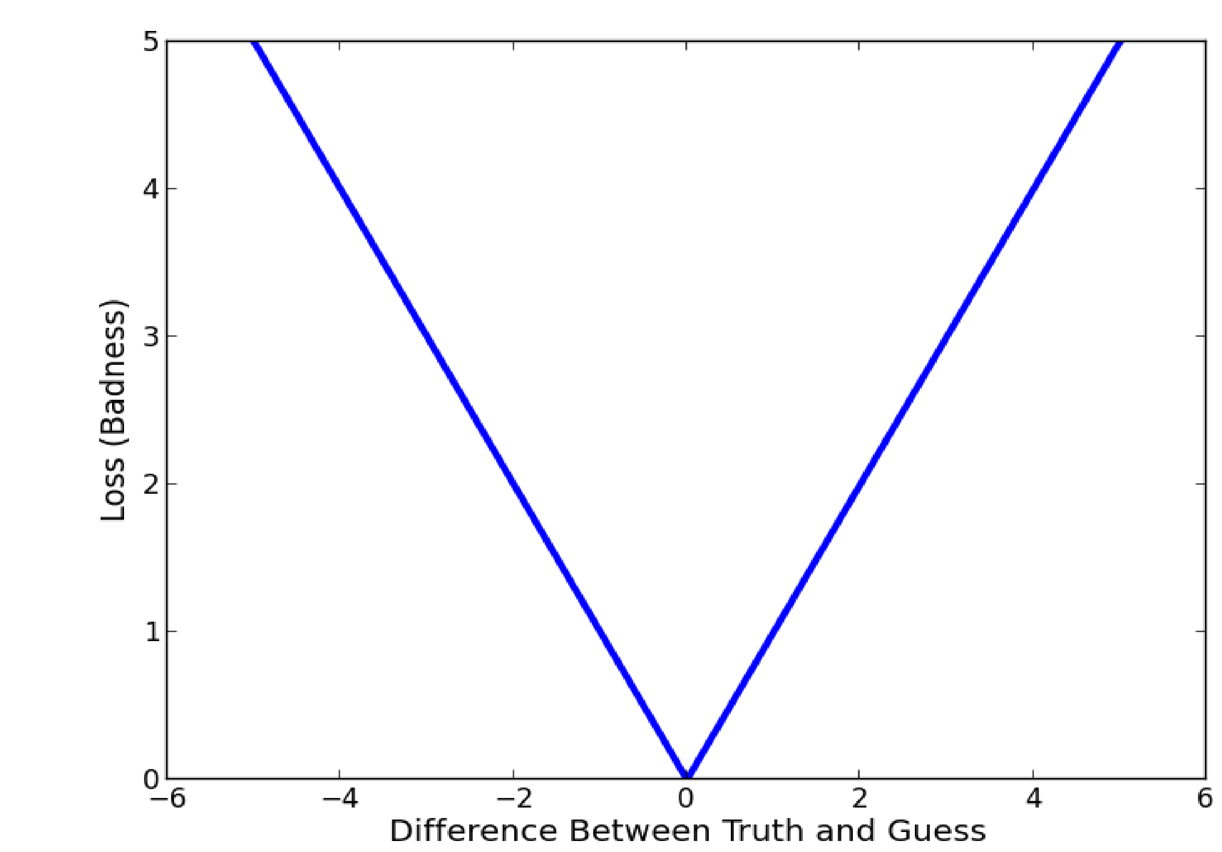
Take Home Message

- You're best off using the posterior mean if you think a quadratic loss function is reasonable

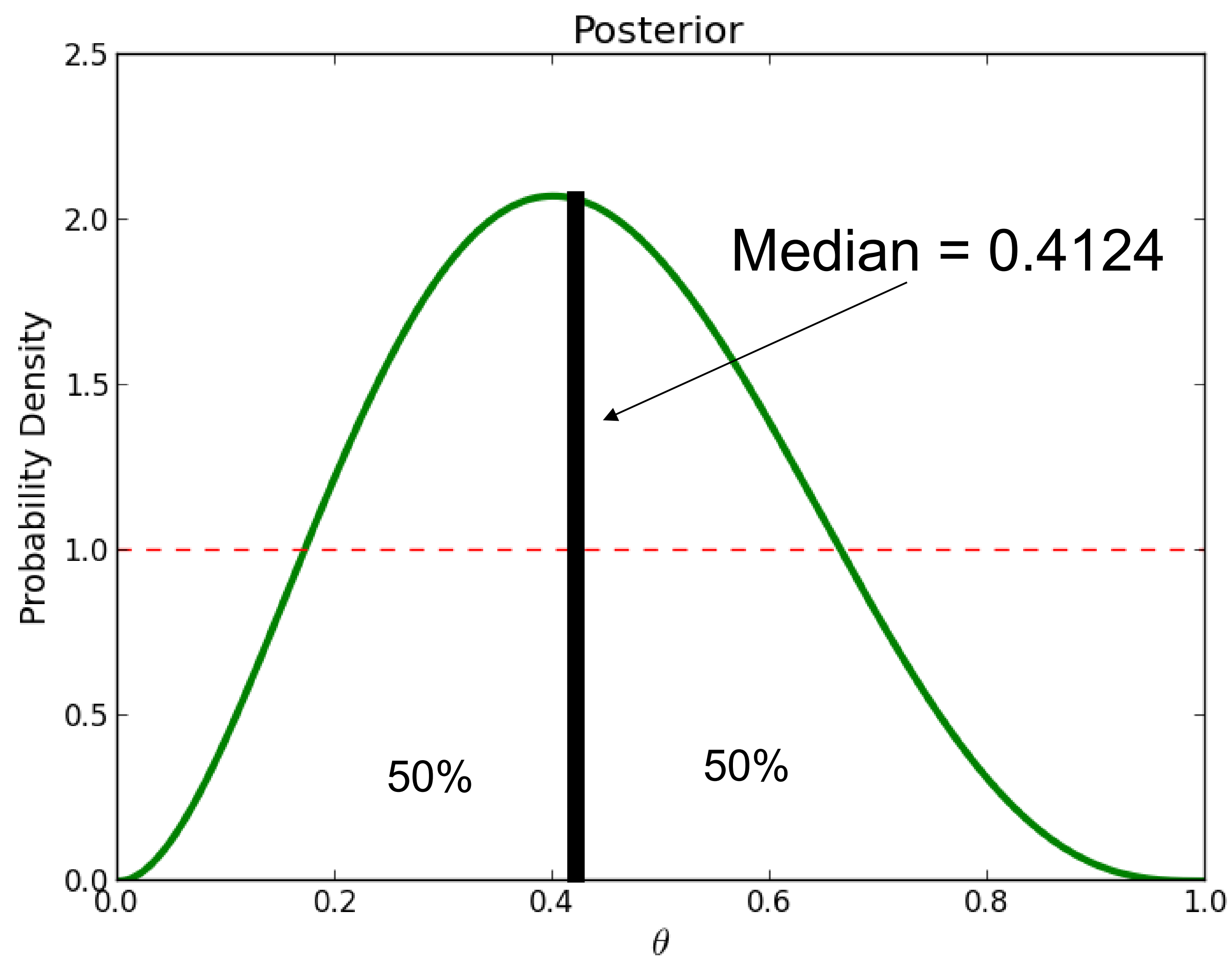
- This is sometimes called the **Bayes Estimate**

Linear/Absolute Loss

$$L(\hat{\theta}, \theta) = |\hat{\theta} - \theta|$$

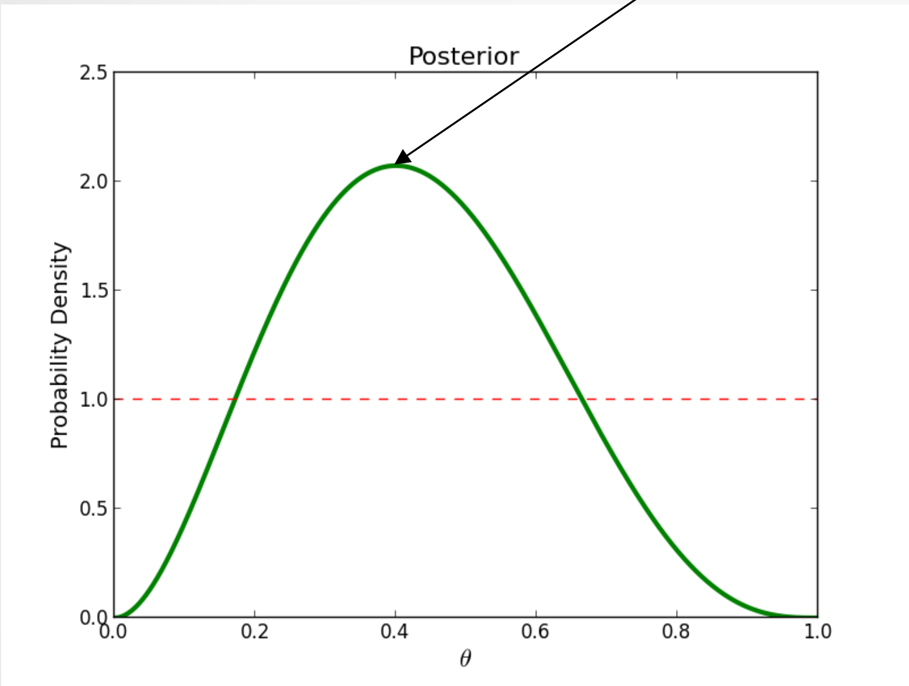


Posterior Median Wins!



When is the Mode Best?

Mode = 0.4



- All-or-nothing loss
- Reward for getting it right
- Anything wrong is equally bad

All-or-nothing loss

- You are hungry and in the dark
- Based on some rustling noises, you have a posterior distribution for the location of an animal
- Where should you shoot?
- **Hint:** a hit is good, but any miss is equally as bad as any other

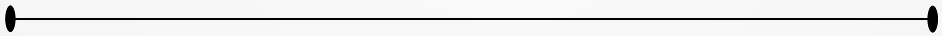
Where should you shoot?

a) Posterior mean

b) Posterior median

c) Posterior mode

Interval Estimation



Credible Intervals: the idea

Find an interval that encloses some specified amount of probability, e.g. 90%, 95%

Credible Intervals in R

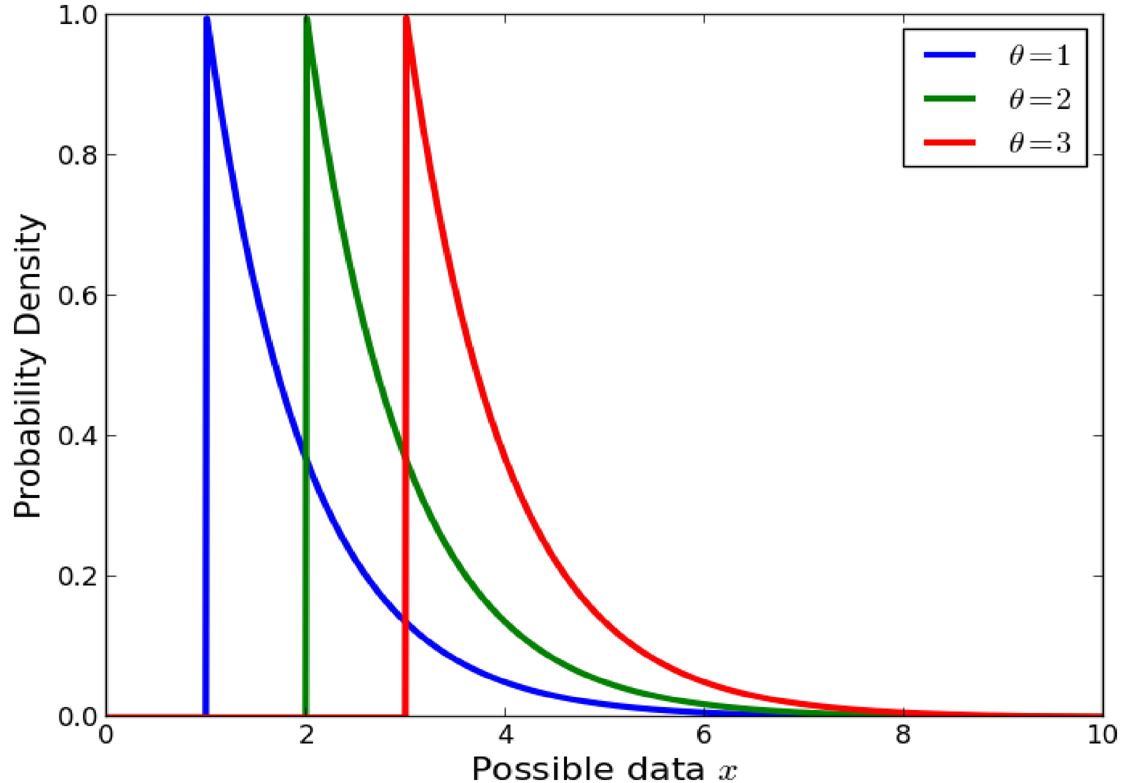
Once again, the method for this will be different when we use MCMC/JAGS

Credible Intervals vs. Confidence Intervals

- New Widgets are injected with a chemical that helps them to function
- They are guaranteed to work for a certain duration
- After that, they fail with a mean lifetime of 1 year.

Sampling Distribution

$$p(x|\theta) = \begin{cases} \exp\left(\theta - x\right), & x \geq \theta \\ 0, & \text{otherwise} \end{cases}$$



Multiple Data

$$p(x_1, x_2, x_3|\theta) = \begin{cases} \prod_{i=1}^{3} \exp\left(\theta - x_i\right), & \text{all } x\text{s} \geq \theta \\ 0, & \text{otherwise} \end{cases}$$

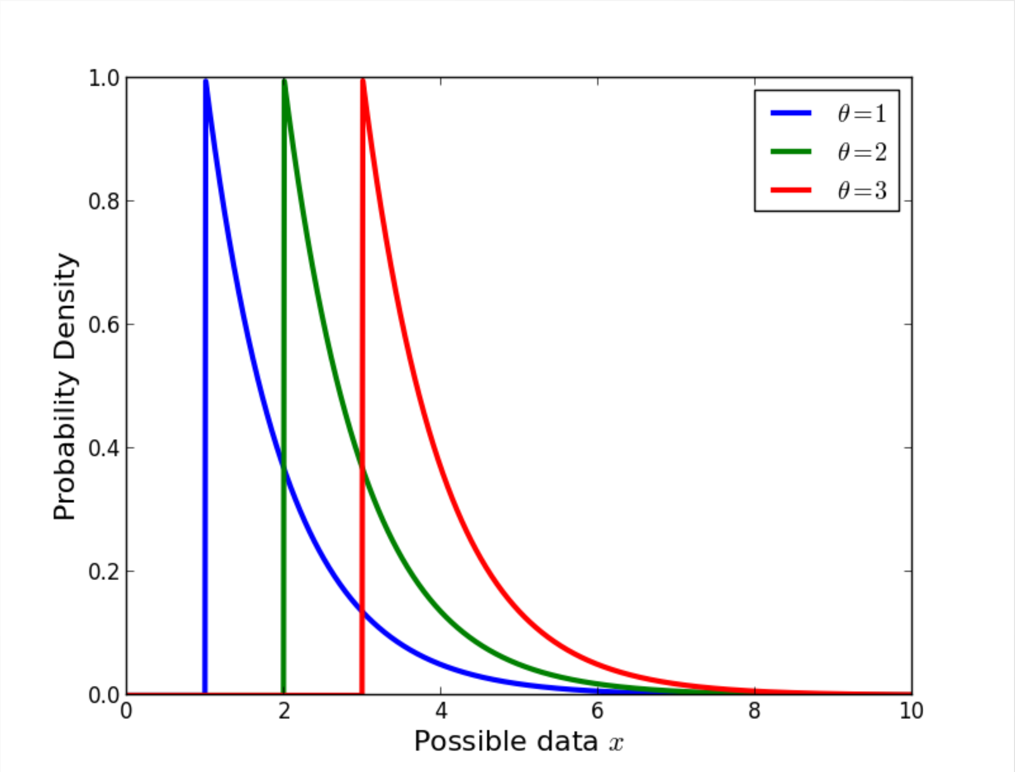
We observe x = {12, 14, 16}

We observe

x = {12, 14, 16}

What is the value of the *likelihood* for θ**=20?**

Likelihood is the probability of observing this data if the hypothesis were true



Bayesian Ingredients

Likelihood

$$p(x_1, x_2, x_3|\theta) = \begin{cases} \prod_{i=1}^{3} \exp\left(\theta - x_i\right), & \text{all } xs \geq \theta \\ 0, & \text{otherwise} \end{cases}$$

$$p(x_1, x_2, x_3|\theta) = \begin{cases} \exp\left(3\theta - 42\right), & \theta < 12 \\ 0, & \text{otherwise} \end{cases}$$

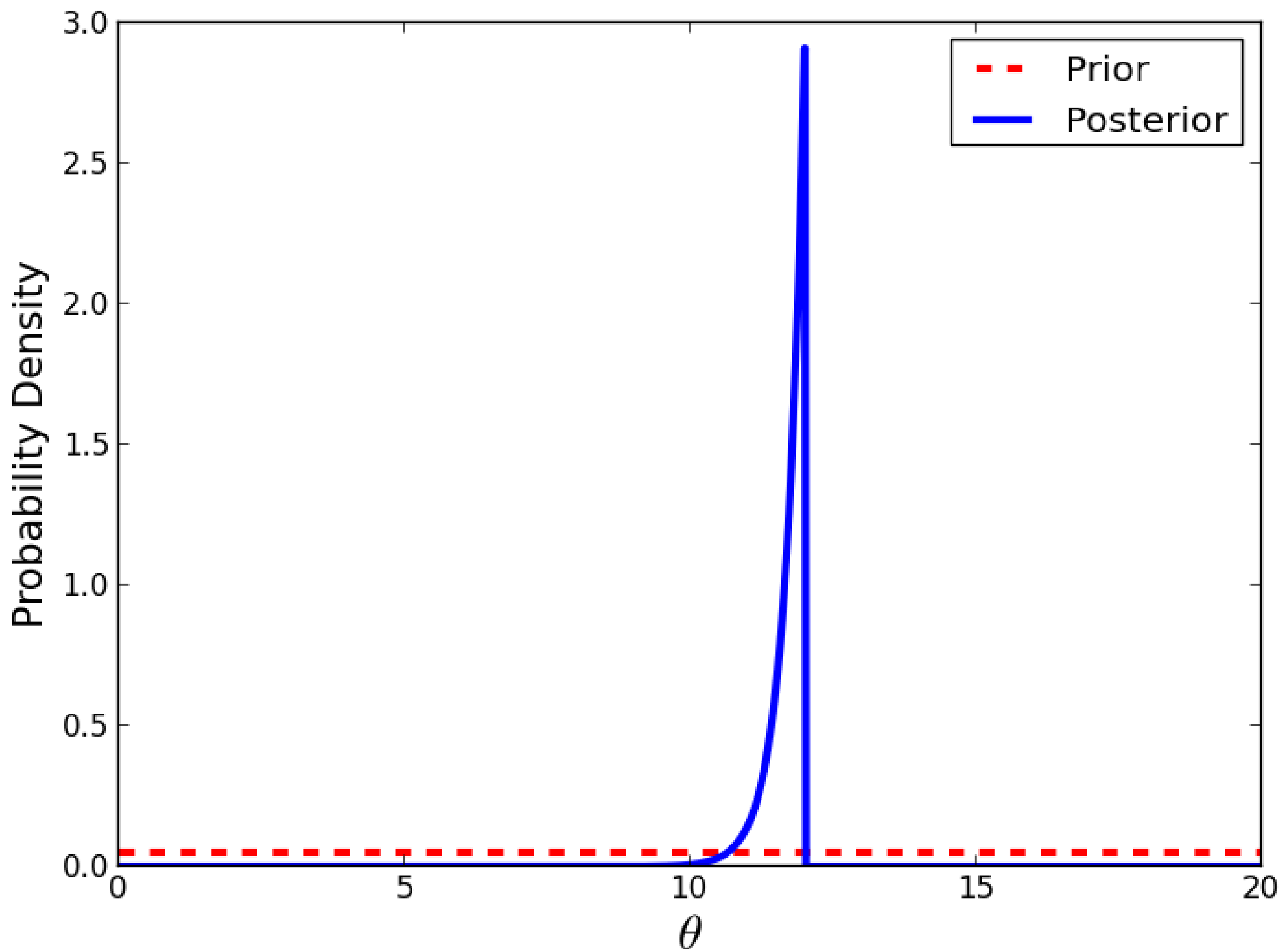
Prior

θ ~ Uniform(0, large upper limit)

Stirring the Pot

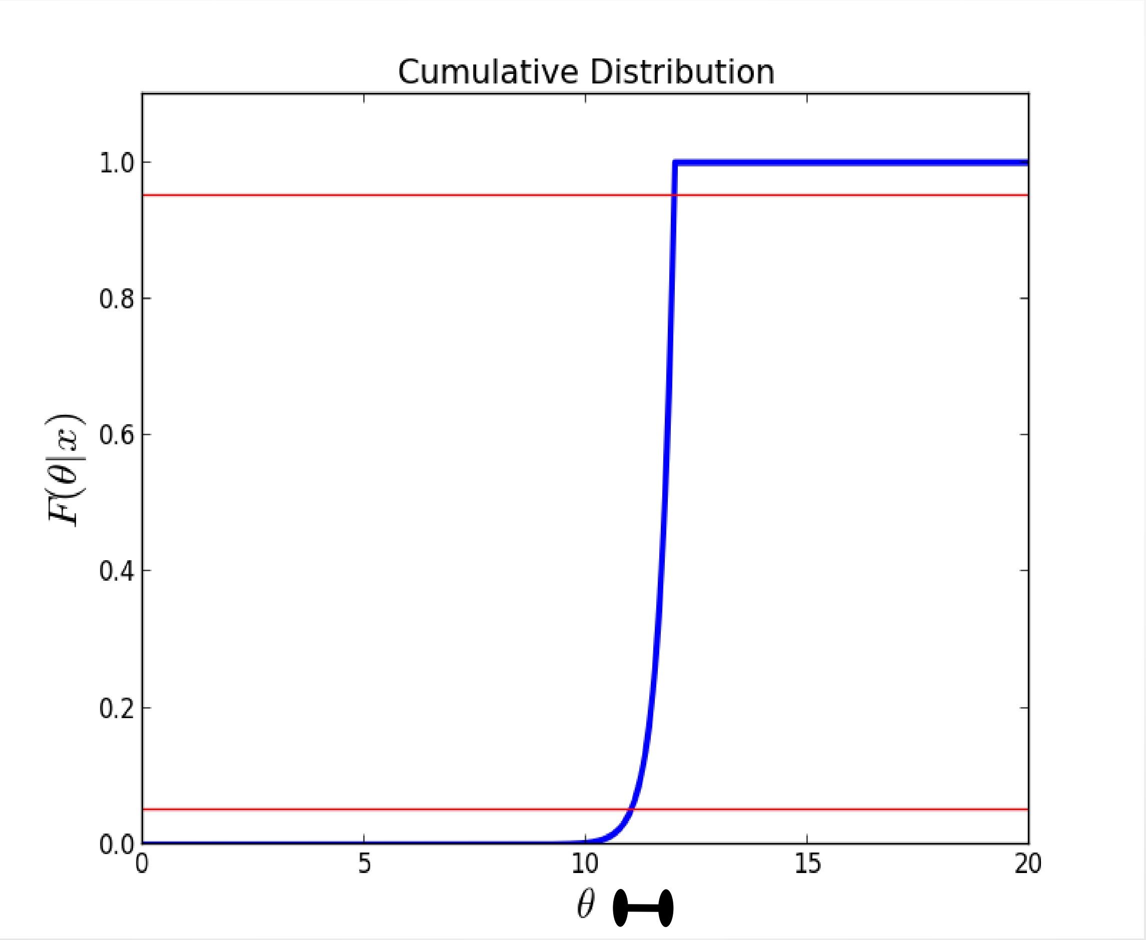
```
theta = seq(0, 20, by=0.01)
prior = rep(1, length(theta))
prior = prior/sum(prior)
lik = exp(3*theta - 42)
lik[theta > 12] = 0

h = prior*lik
post = h/sum(h)
plot(theta, post, type='l', xlab='Theta',
  ylab='Posterior Probability')
```

Posterior PDF



Cumulative Distribution

- 90% credible interval = {11.00, 11.98}



Frequentist Confidence Interval

- Prob(θ in interval) = 0.9

- Bayesian Credible interval

Prob(θ in interval | x)

- Frequentist Confidence interval

Prob(θ in interval | θ)

Widget Confidence Interval

- An estimator and a confidence interval

$$\theta^* = \frac{1}{N}\sum_{i=1}^{N}(x_i - 1)$$

$$(\theta^* - 0.8529, \theta^* + 0.8264)$$

Fake Data Sets

```
theta = 10

x = theta = log(runif(3)) # Generate from
  exponential

theta_star = mean(x - 1)

left = theta_star – 0.8529

right = theta_star + 0.8264

inside = (theta > left && theta < right)

TRUE
```

Repeat 1,000,000 times

```
for(i in 1:1000000)
{
      # do stuff
}

> mean(inside)
[1] 0.899118
```

Confidence Interval

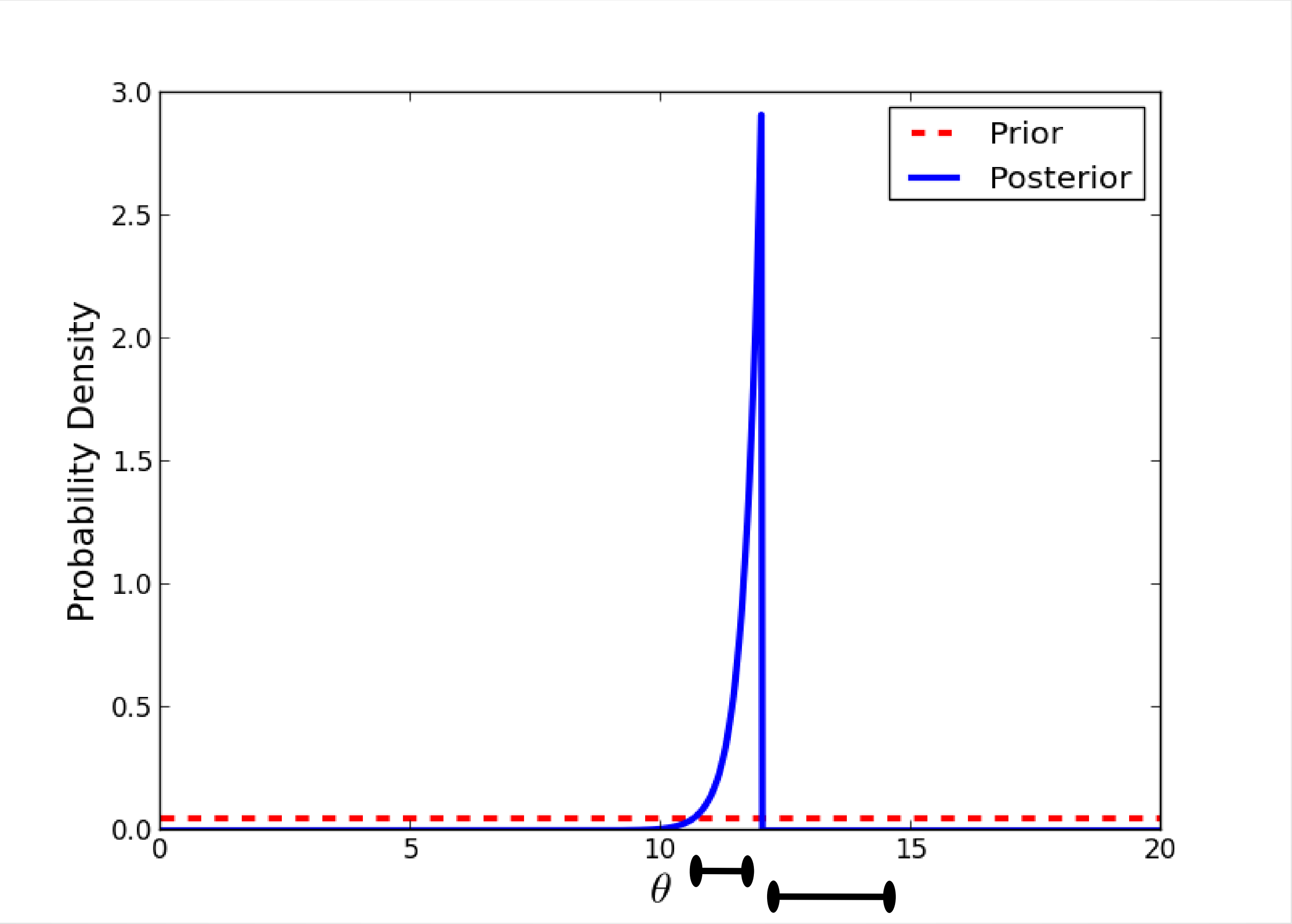
- Our data was x = {12, 14, 16}
- Frequentist confidence interval

  {12.1471, 13.8264}

- Bayesian credible interval

  {11.00, 11.98}

OOPS



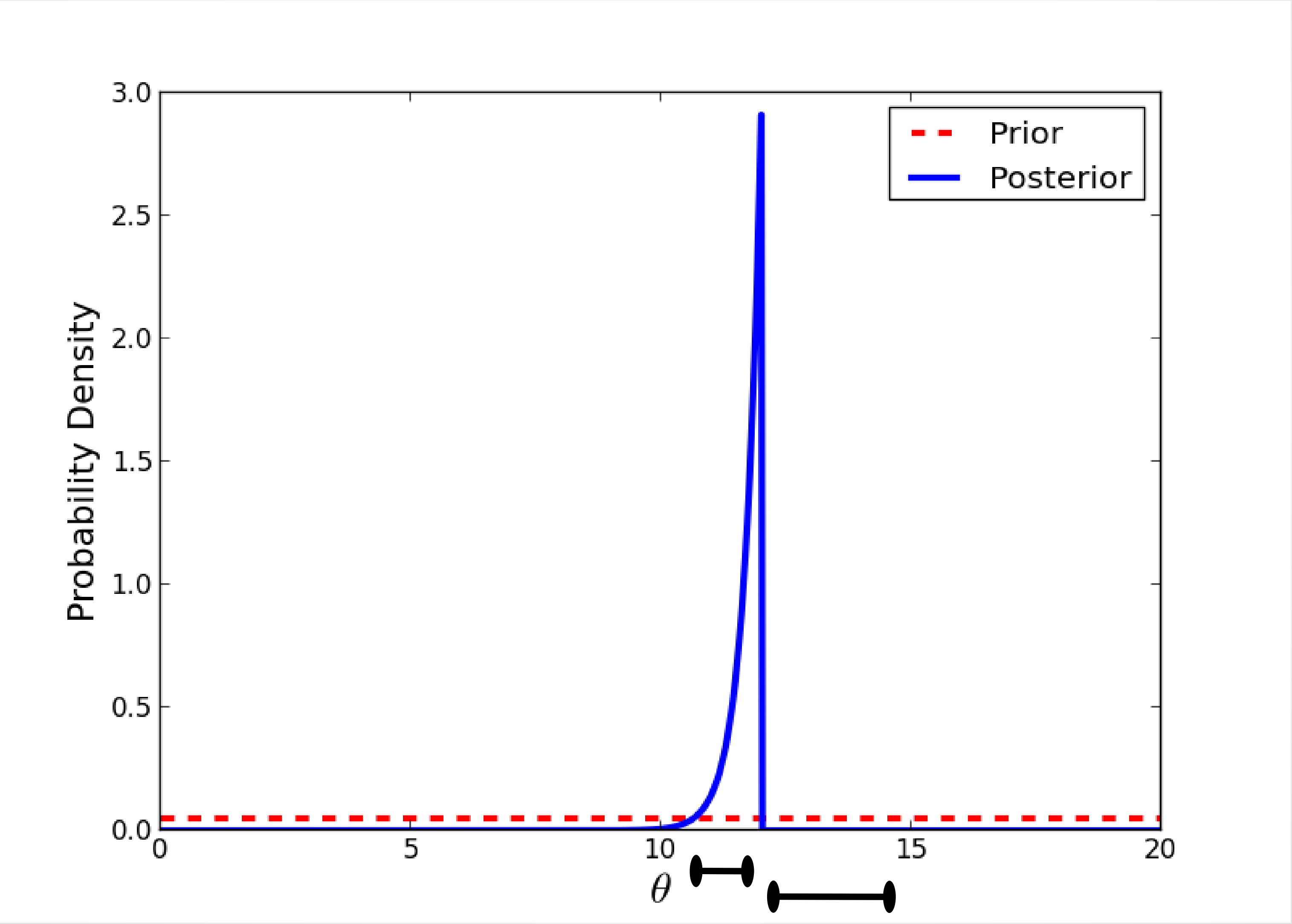
Prior
Posterior
Probability Density
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
5
10
15
20
θ

Summary

- Today we saw how the “best” estimate depends on the utility function

- Credible intervals and confidence intervals are different concepts!

OOPS



3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
5
10
15
20
Probability Density
θ
Prior
Posterior

Next Week

- We will learn about MCMC. Why we need it and how it works
- See you in the labs!