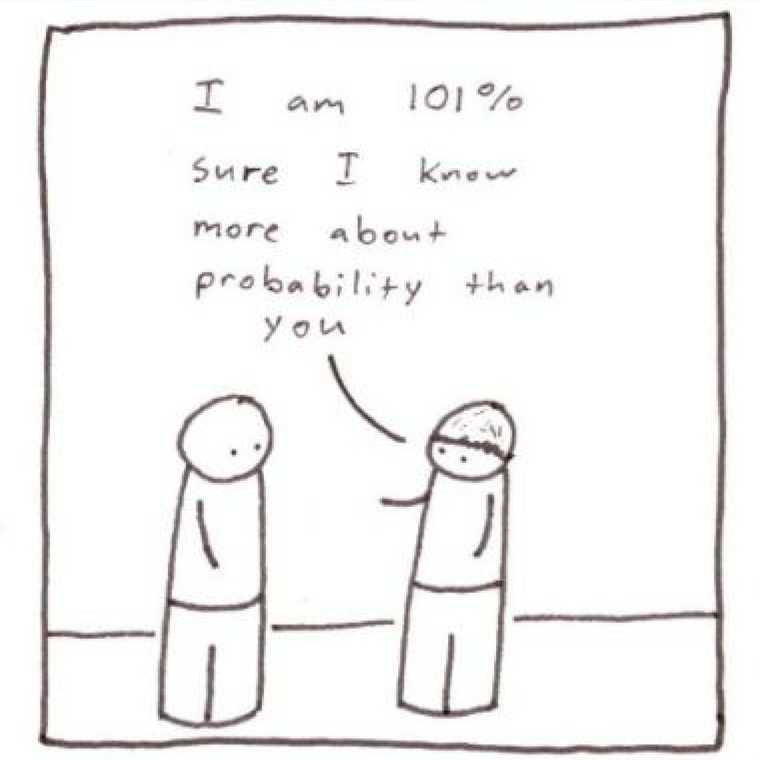
STATS331



Credit: www.afewpanels.com

Introduction to Bayesian Statistics
Semester 2, 2016

Today

- Review!
- What I expect you to know

History

- 1600s-1800s: Probability was developed and applied, including the “Bayesian” approach

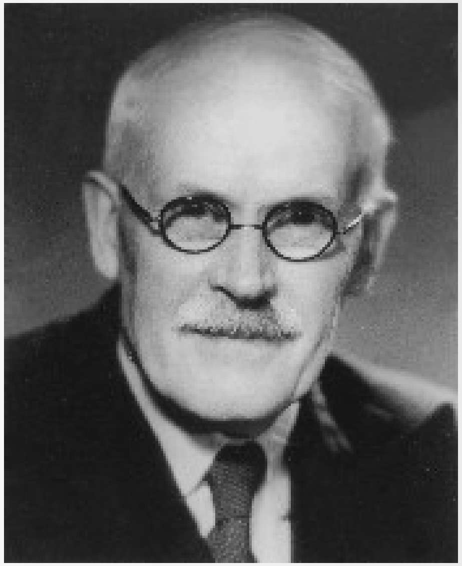
- 1900-1990: Frequentists questioned, developed alternatives

- 1990-forever: The Bayesian Revival

Key Players

Bayes

Laplace



Jeffreys

- History **will not be** in the exam!

You could always look it up if you felt like it...

:-)

Probability and Uncertainty

- In Bayesian Statistics, a probability is a measure of **how confident you are that a hypothesis is true**
- These probabilities change over time as you get more information.

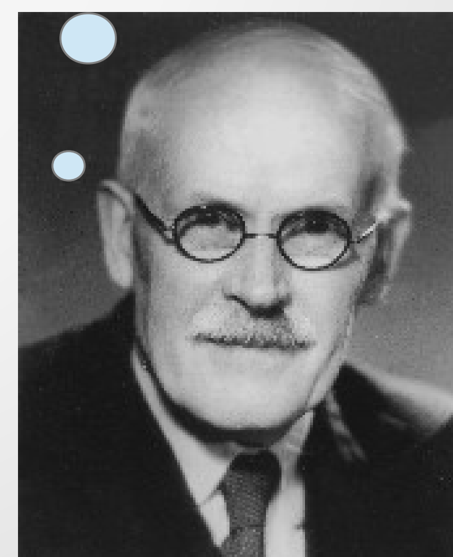
Probability can describe either...

95% of the time

I'm 95% sure

Bayesian statistics is
based on using this →
on every problem

Classical statistics is
based on using this only
←



Probability Rules!

We looked at (and used) the sum rule

$$P(A \lor B) = P(A) + P(B) - P(A, B)$$

and the product rule

$$P(A, B) = P(A)P(B|A)$$

Bayes' Rule

- How exactly do the probabilities get updated?
- Bayes' Rule (comes from the product rule):

Prob. that we would have got this data if *H* were true – "**likelihood**"

Original probability "**prior**"

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

Updated probability "**posterior**"

Prob. that we would have got this data whether *H* were true or not

Interpretation of Bayes' Rule

- The posterior probability depends on:
- **The prior probability**. If the hypothesis was plausible before the data, then that helps it to be plausible after the data!
- **The likelihood**. If the hypothesis assigned a high probability to the data (i.e. it *predicted the data*) then it is rewarded!
- **The P(D) in the denominator (marginal likelihood)***.* If the data would have happened anyway *even if H were false*, then a high likelihood doesn't mean much!

Checklist

- Know Bayes' Rule and the sum rule from memory! **They will not be given in the exam**
- Know how to use them (or a Bayes' Box, which accomplishes the same thing).
- Sequential use: “today's posterior is tomorrow's prior”
- Sum rule to calculate prior/posterior probabilities of “or” statements

Parameter Estimation

- Bayes' rule can be applied to a set of hypotheses about the value of an unknown parameter

- A hypothesis might be "$\theta = 1$"
- Another might be "$\theta = 2$"
- Suppose we observed "$x = 3$"

Bayes' Rule Lots of Times

$$P(\theta = 1 | x = 3) = \frac{P(\theta = 1) P(x = 3 | \theta = 1)}{P(x = 3)}$$

$$P(\theta = 2 | x = 3) = \frac{P(\theta = 2) P(x = 3 | \theta = 2)}{P(x = 3)}$$

$$\cdots$$

$$P(\theta = 10 | x = 3) = \frac{P(\theta = 10) P(x = 3 | \theta = 10)}{P(x = 3)}$$

Posterior Distribution

Prior Distribution

Bayes' Rule Lots of Times

$$P(\theta = 1 | x = 3) = \frac{P(\theta = 1)\boxed{P(x = 3 | \theta = 1)}}{\boxed{P(x = 3)}}$$

$$P(\theta = 2 | x = 3) = \frac{P(\theta = 2)\boxed{P(x = 3 | \theta = 2)}}{\boxed{P(x = 3)}}$$

$$\ldots$$

$$P(\theta = 10 | x = 3) = \frac{P(\theta = 10)\boxed{P(x = 3 | \theta = 10)}}{\boxed{P(x = 3)}}$$

Green are likelihoods. Orange is a **common** normalisation constant, the marginal likelihood

$$p(\theta|x) \propto p(\theta)p(x|\theta)$$

$$\texttt{posterior} \propto \texttt{prior} \times \texttt{likelihood}$$

This works for discrete and continuous distributions

Canonical Example: A Proportion

- Remember when I moved to Auckland I took the correct bus 2 times out of 5

- Wanted to estimate θ, the true proportion of correct buses

Likelihood

We usually get our likelihood by having a model for the *probability distribution for the data (sometimes called the sampling distribution).*

It always *depends on the unknown parameter(s)*

$$x|\theta \sim \mathrm{Binomial}(N, \theta)$$

$$p(x|\theta) = \left(\begin{array}{c} N \\ x \end{array} \right) \theta^x (1-\theta)^{N-x}$$

The Bayes' Box

- The Bayes' Box is a way of showing *discrete* parameter estimation in a table

- In the proportion example, θ is continuous and between 0 and 1, but can make a discrete approximation

A Bayes' Box

Possible Answers for θ	Prior $p(\theta)$	Likelihood $p(x\|\theta)$	Prior x Likelihood h	Posterior $p(\theta\|x)$
0	0.0909	0	0	0
0.1	0.0909	0.0729	0.0066	0.0437
0.2	0.0909	0.2048	0.0186	0.1229
0.3	0.0909	0.3087	0.0281	0.1852
0.4	0.0909	0.3456	0.0314	0.2074
0.5	0.0909	0.3125	0.0284	0.1875
0.6	0.0909	0.2304	0.0209	0.1383
0.7	0.0909	0.1323	0.0120	0.0794
0.8	0.0909	0.0512	0.0047	0.0307
0.9	0.0909	0.0081	0.0007	0.0049
1	0.0909	0	0	0
Totals	*1*		*0.1515*	*1*

Probabilities

Possible Answers for θ	Prior p(θ)	Likelihood p(x|θ)	Prior x Likelihood h	Posterior p(θ|x)
0	0.0909	0	0	0
0.1	0.0909	0.0729	0.0066	0.0437
0.2	0.0909	0.2048	0.0186	0.1229
0.3	0.0909	0.3087	0.0281	0.1852
0.4	0.0909	0.3456	0.0314	0.2074
0.5	0.0909	0.3125	0.0284	0.1875
0.6	0.0909	0.2304	0.0209	0.1383
0.7	0.0909	0.1323	0.0120	0.0794
0.8	*0.3636* 0.0909	0.0512	0.0047	*0.1150* 0.0307
0.9	0.0909	0.0081	0.0007	0.0049
1	0.0909	0	0	0
Totals	*1*		*0.1515*	*1*

Posterior Mode. cf “Maximum Likelihood”

Possible Answers for θ	Prior $p(\theta)$	Likelihood $p(x\|\theta)$	Prior x Likelihood h	Posterior $p(\theta\|x)$
0	0.0909	0	0	0
0.1	0.0909	0.0729	0.0066	0.0437
0.2	0.0909	0.2048	0.0186	0.1229
0.3	0.0909	0.3087	0.0281	0.1852
0.4	0.0909	0.3456	0.0314	0.2074
0.5	0.0909	0.3125	0.0284	0.1875
0.6	0.0909	0.2304	0.0209	0.1383
0.7	0.0909	0.1323	0.0120	0.0794
0.8	0.0909	0.0512	0.0047	0.0307
0.9	0.0909	0.0081	0.0007	0.0049
1	0.0909	0	0	0
Totals	*1*		*0.1515*	*1*

Posterior Mean

Possible Answers for θ	Prior p(θ)	Likelihood p(x|θ)	Prior x Likelihood h	Posterior p(θ|x)
0	0.0909	0	0	0
0.1	0.0909	0.0729	0.0066	0.0437
0.2	0.0909	0.2048	0.0186	0.1229
0.3	0.0909	0.3087	0.0281	0.1852
0.4	0.0909	0.3456	0.0314	0.2074
0.5	0.0909	0.3125	0.0284	0.1875
0.6	0.0909	0.2304	0.0209	0.1383
0.7	0.0909	0.1323	0.0120	0.0794
0.8	0.0909	0.0512	0.0047	0.0307
0.9	0.0909	0.0081	0.0007	0.0049
1	0.0909	0	0	0
Totals	*1*		*0.1515*	*1*

Need to Know

- Know what all parts of a Bayes' Box mean
- How to calculate some numbers from other numbers (e.g. summaries)
- PREDICTION!

Hypothesis Testing

- In classical statistics, parameter estimation and hypothesis testing are considered different topics

- Not for us!

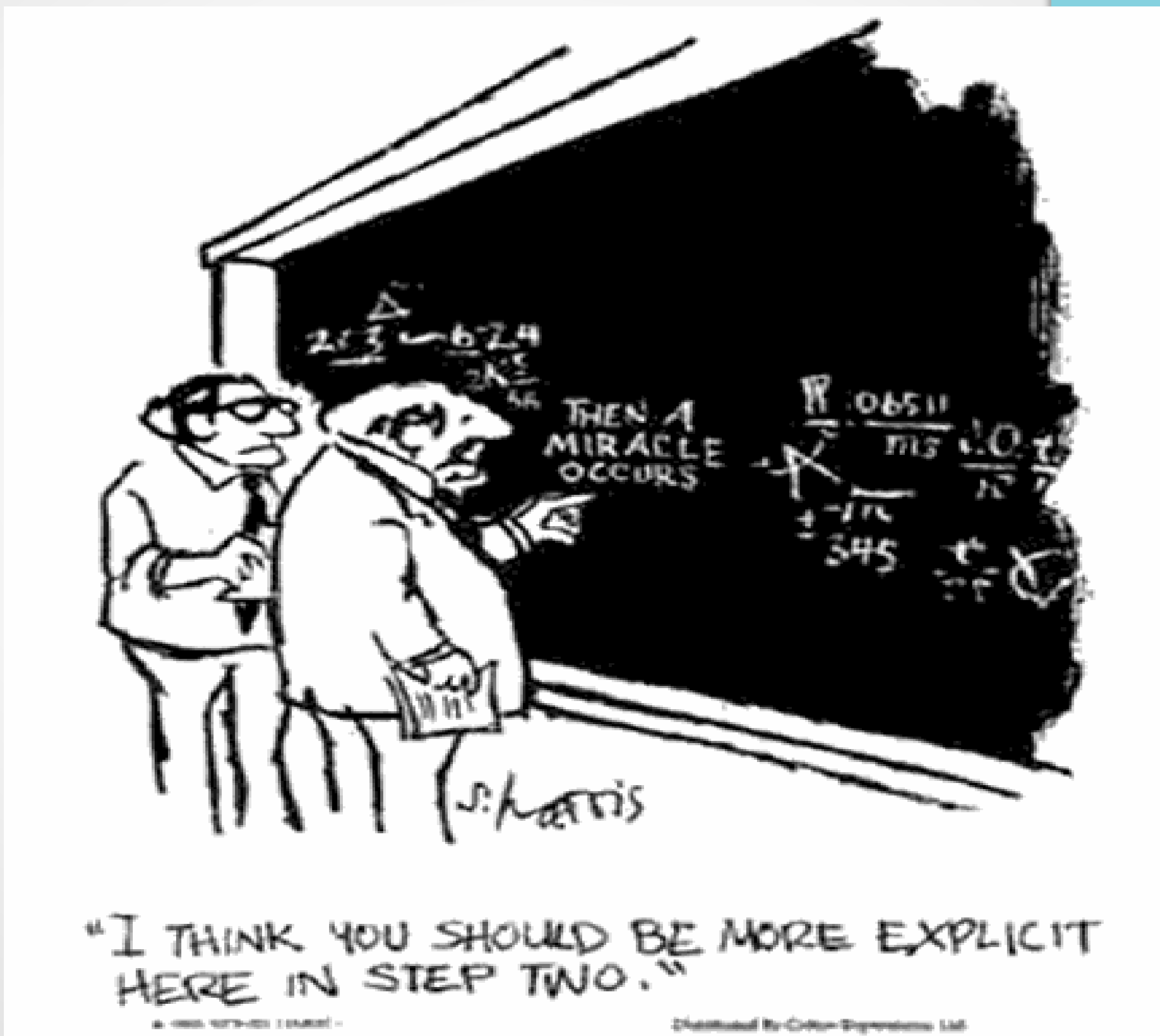
  e.g. “θ=0.5” is just a hypothesis.

  And we have tested it.

Testing Prior

Possible Answers for θ	Prior p(θ)	Likelihood p(x|θ)	Prior x Likelihood h	Posterior p(θ|x)
0	**0.05**	0	0	0
0.1	**0.05**	0.0729	0.0036	0.0162
0.2	**0.05**	0.2048	0.0102	0.0457
0.3	**0.05**	0.3087	0.0154	0.0689
0.4	**0.05**	0.3456	0.0172	0.0772
0.5	**0.5**	0.3125	0.1562	**0.6977**
0.6	**0.05**	0.2304	0.0115	0.0514
0.7	**0.05**	0.1323	0.0066	0.0295
0.8	**0.05**	0.0512	0.0026	0.0114
0.9	**0.05**	0.0081	0.0004	0.0018
1	**0.05**	0	0	0
Totals	*1*		*0.2240*	*1*

Analytical Methods



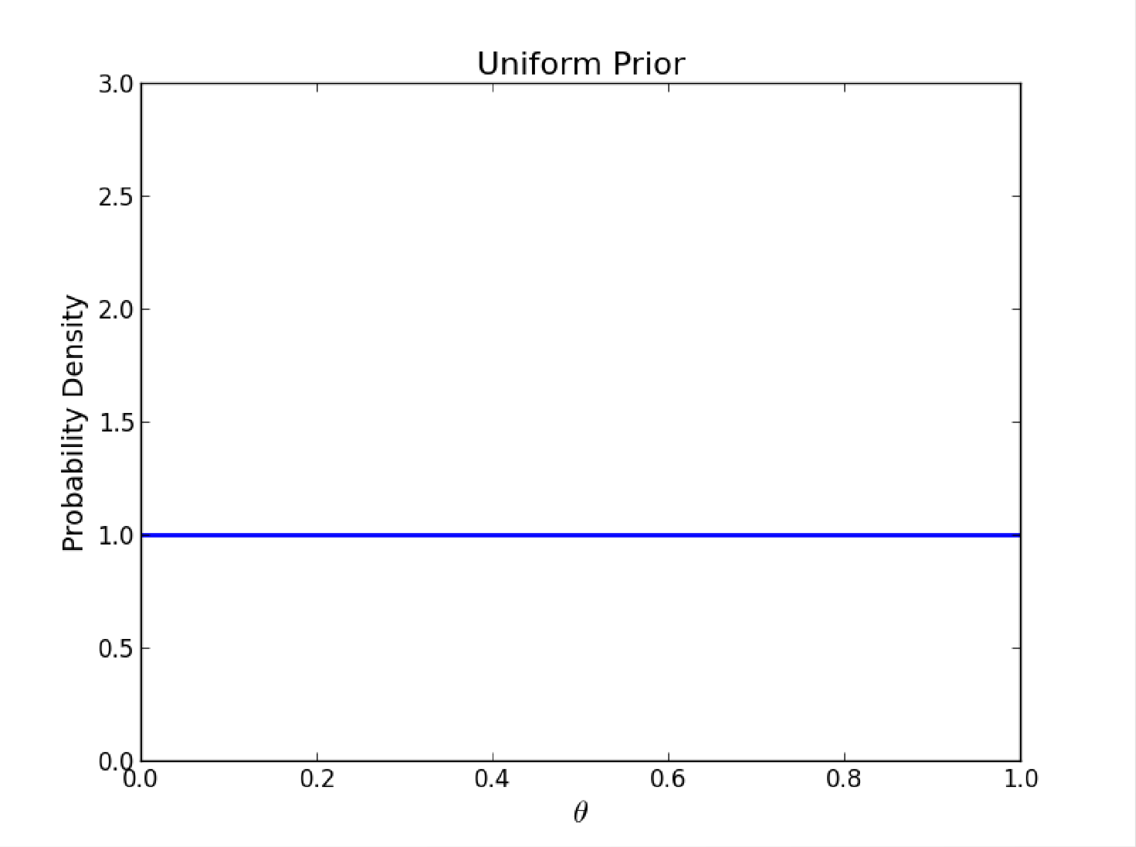
Analytical Methods

- Solve parameter estimation problems
- Write down the equation for the prior distribution
- Write down the equation for the likelihood (this depends on the data and the parameter(s))

Bus Problem

- Uniform Prior

$$p(\theta) = \begin{cases} 1, & 0 \leq \theta \leq 1 \\ 0, & \text{otherwise} \end{cases}$$



Bus Problem

- Binomial

$$
\begin{aligned}
p(x|\theta) &= \binom{N}{x}\theta^x(1-\theta)^{N-x} \\
&= \binom{5}{2}\theta^2(1-\theta)^3
\end{aligned}
$$

Bayes' Rule

$$p(\theta|x) \quad \propto \quad p(\theta)p(x|\theta)$$

$$\texttt{posterior} \quad \propto \quad \texttt{prior} \times \texttt{likelihood}$$

$$p(\theta|x) \quad \propto \quad \theta^2(1-\theta)^3$$

Recognise the Posterior

$$X \sim \mathrm{Beta}(\alpha, \beta)$$

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}$$

$$p(\theta | x) \propto \theta^2 (1-\theta)^3$$

Need to Know

- Know how to identify the posterior as being from a particular family (e.g. beta, gamma)

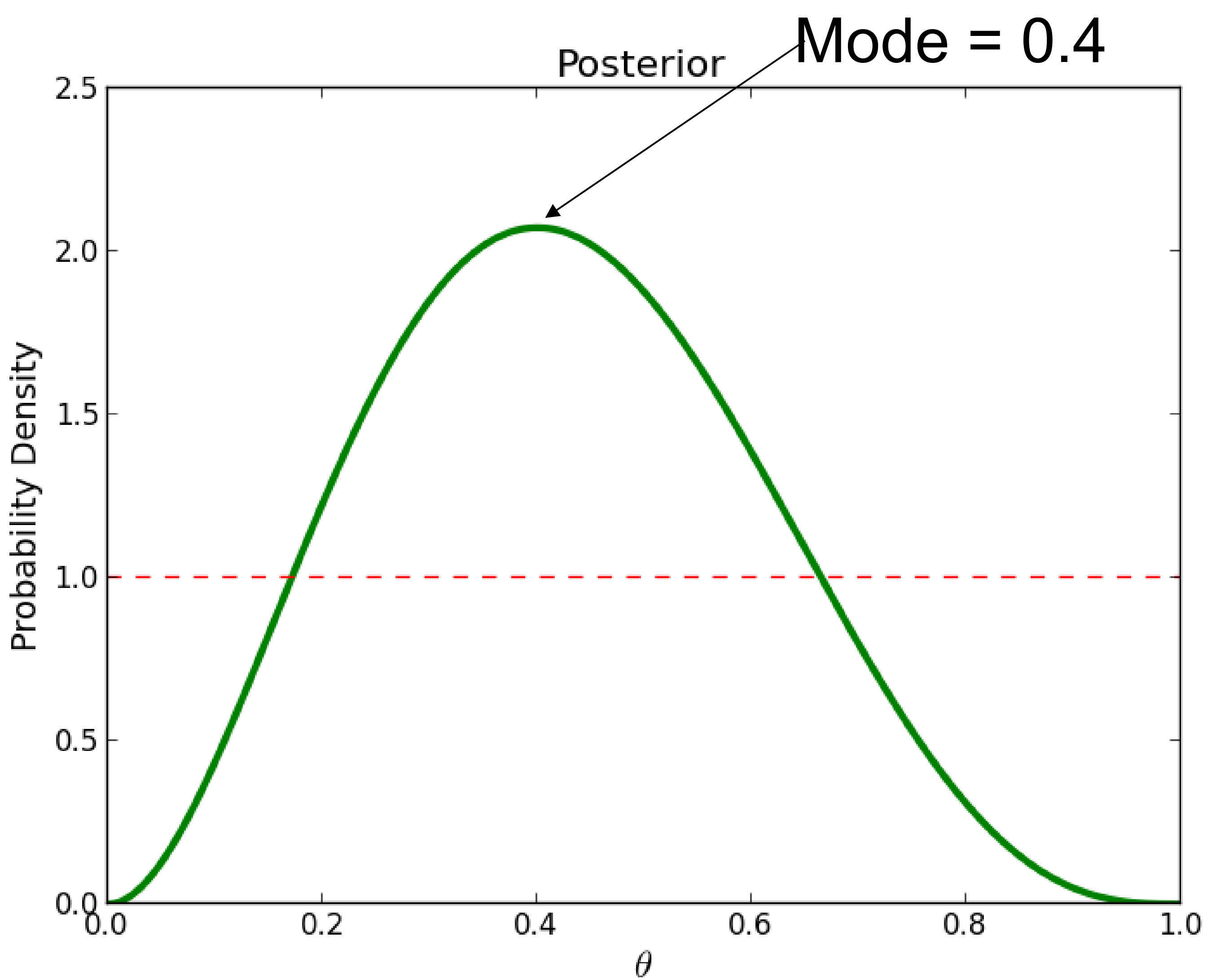
- **You will not need to memorise the equation for a beta or a gamma or any other distribution** but you need to *know how to use them*!

- The exam sheet about probability distributions will be the same as in the midterm test.

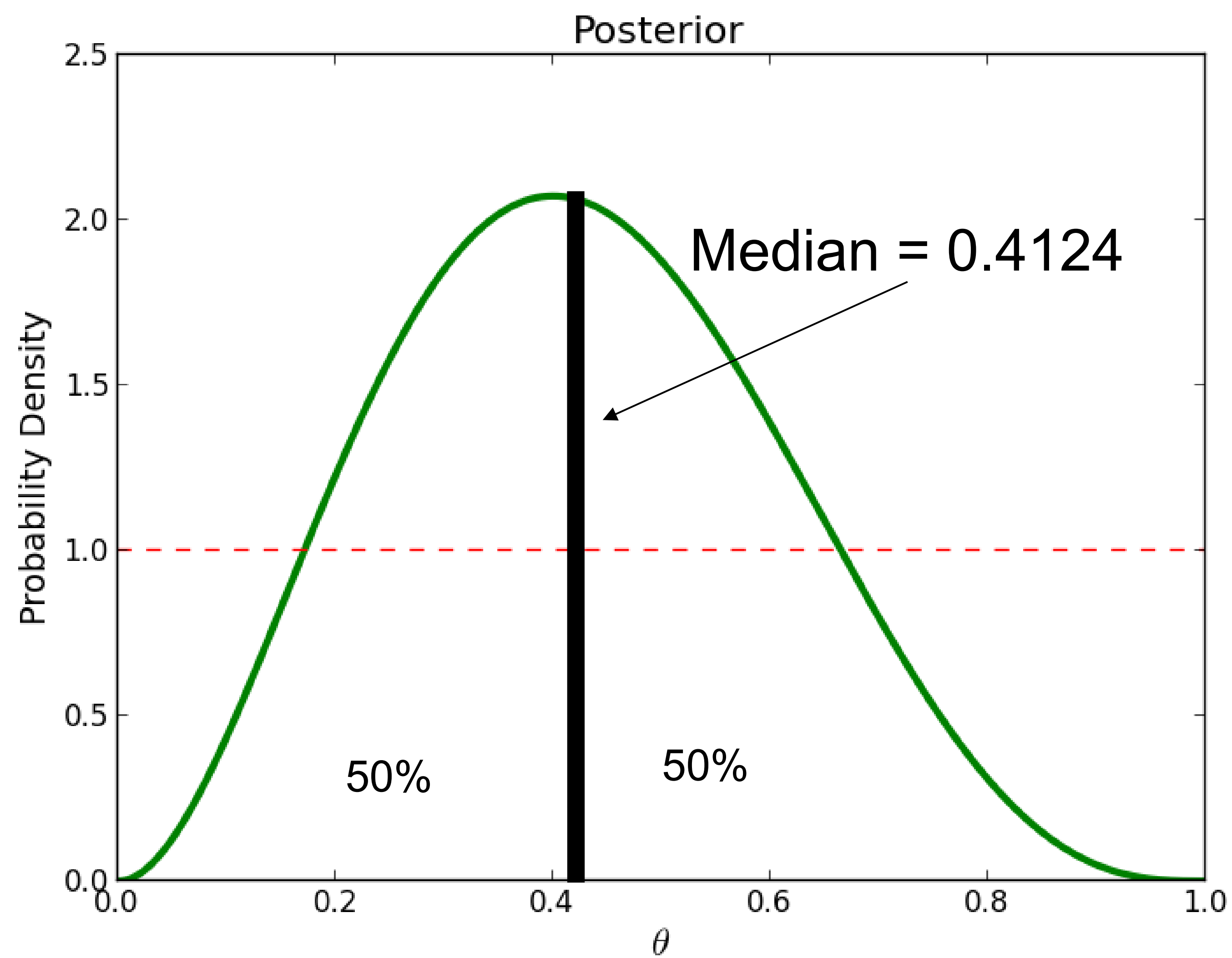
Point and Interval Estimation

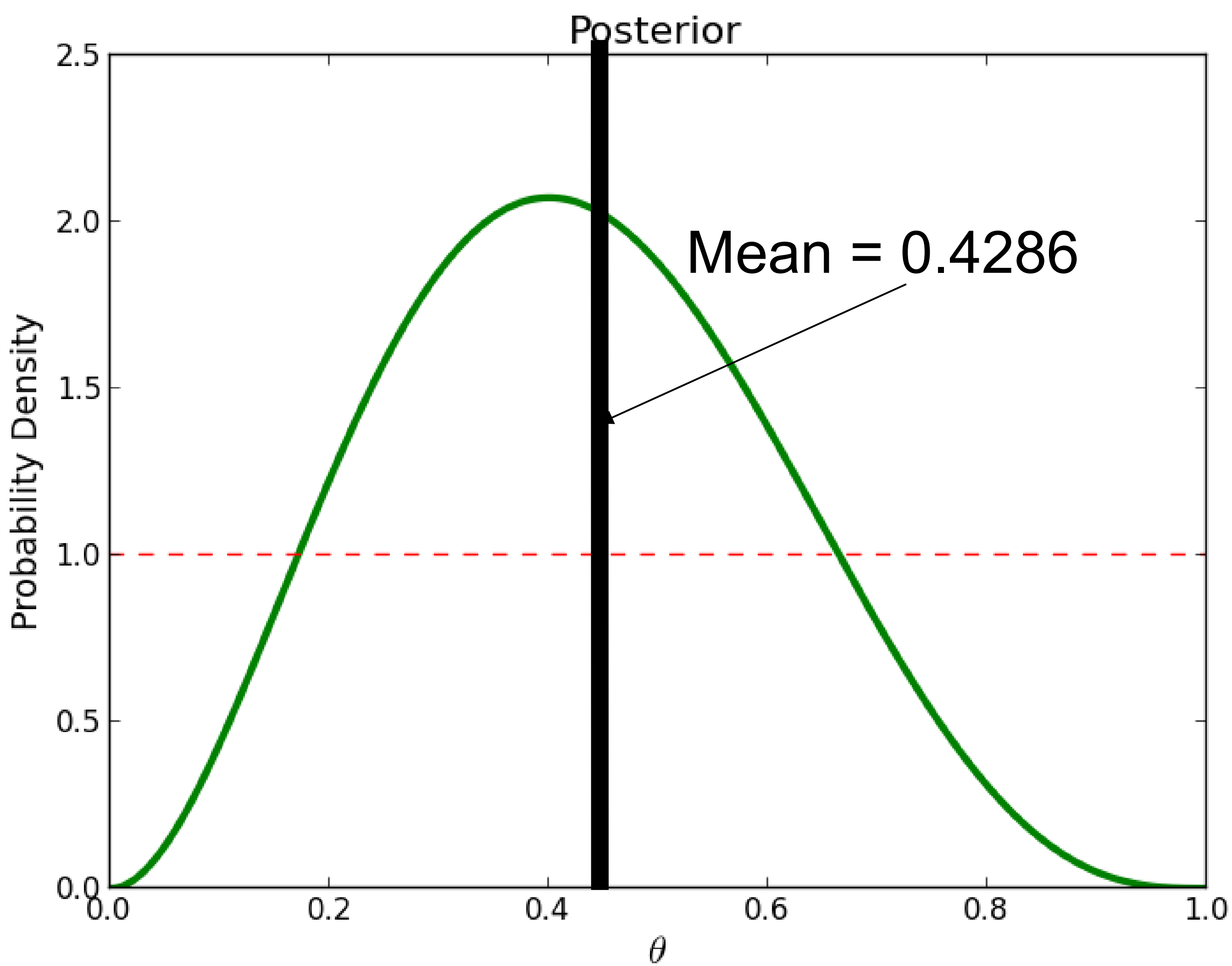
- These are useful as summaries of the posterior distribution

- Point estimate = “make a single guess for the value of the parameter(s)”



Mode = 0.4
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


Posterior
Probability Density
Median = 0.4124
50%
50%
θ
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0



Posterior
Mean = 0.4286
Probability Density
θ
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0

“Best” Estimate

- Best estimate is the true value → we wish!
- Depends on **loss/utility function** (how bad is it to be wrong in certain ways?)
- Quadratic loss → posterior mean
- Linear (really absolute value) loss → posterior median
- All-or-nothing loss → posterior mode

Interval Estimation

- We have seen a lot of credible intervals
- Interval that contains a specified amount of probability (e.g. 95%).

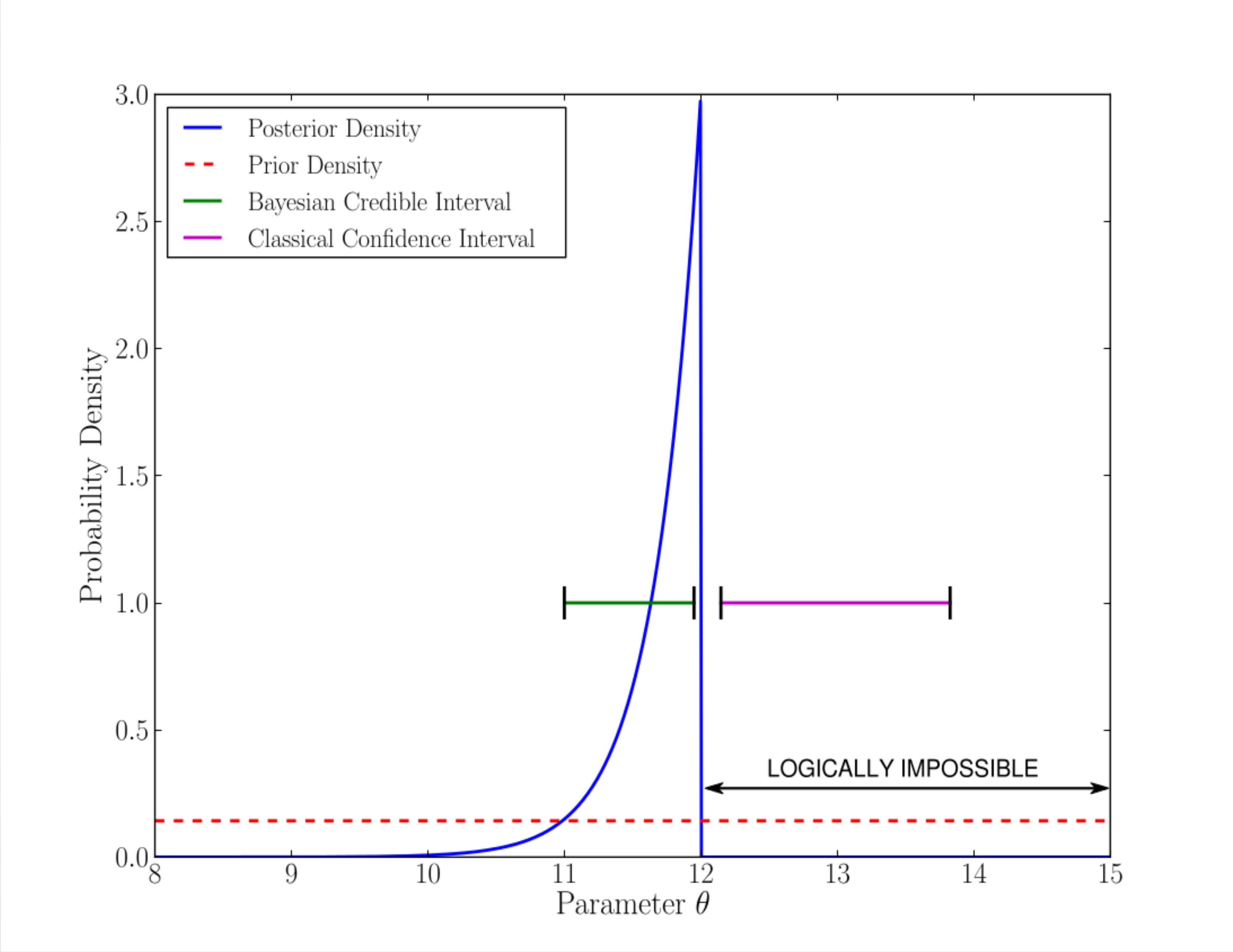
Question

What's the difference between a Bayesian credible interval and a frequentist confidence interval?

Bayesian: the probability the parameter is within the interval, given the data, is (for example) 95%

Frequentist: for 95% of possible data sets, the interval will contain the parameter

Remember this?



Credible Intervals

- Credible intervals are constructed from QUANTILES of the posterior distribution

Need to Know

- Which point estimate corresponds to which kind of loss function

- What credible intervals are and how they are calculated

  (e.g. see 2013 exam: some R output was given, and students had to know which part was calculating the credible interval)

On Wednesday

- Will quickly run through what we covered in the second half of the course, and comment on the exam