

# Doing Bayesian Data Analysis

## §5.3~5.4

Chao-En Yu

National Tsing Hua University

# Bias in a Coin

- We now consider an set of examples that will build your understanding of how prior distributions and data interact to produce posterior distributions.
- The first step in Bayesian analysis is identifying the type of data being considered: Binary & categorical.
- The next step is creating a descriptive model with meaningful parameters,  $p(\gamma = 1|\theta) = \theta \in [0,1]$ :

$$p(\gamma|\theta) = \theta^\gamma (1 - \theta)^{(1-\gamma)}$$



Bernoulli distribution

# Likelihood

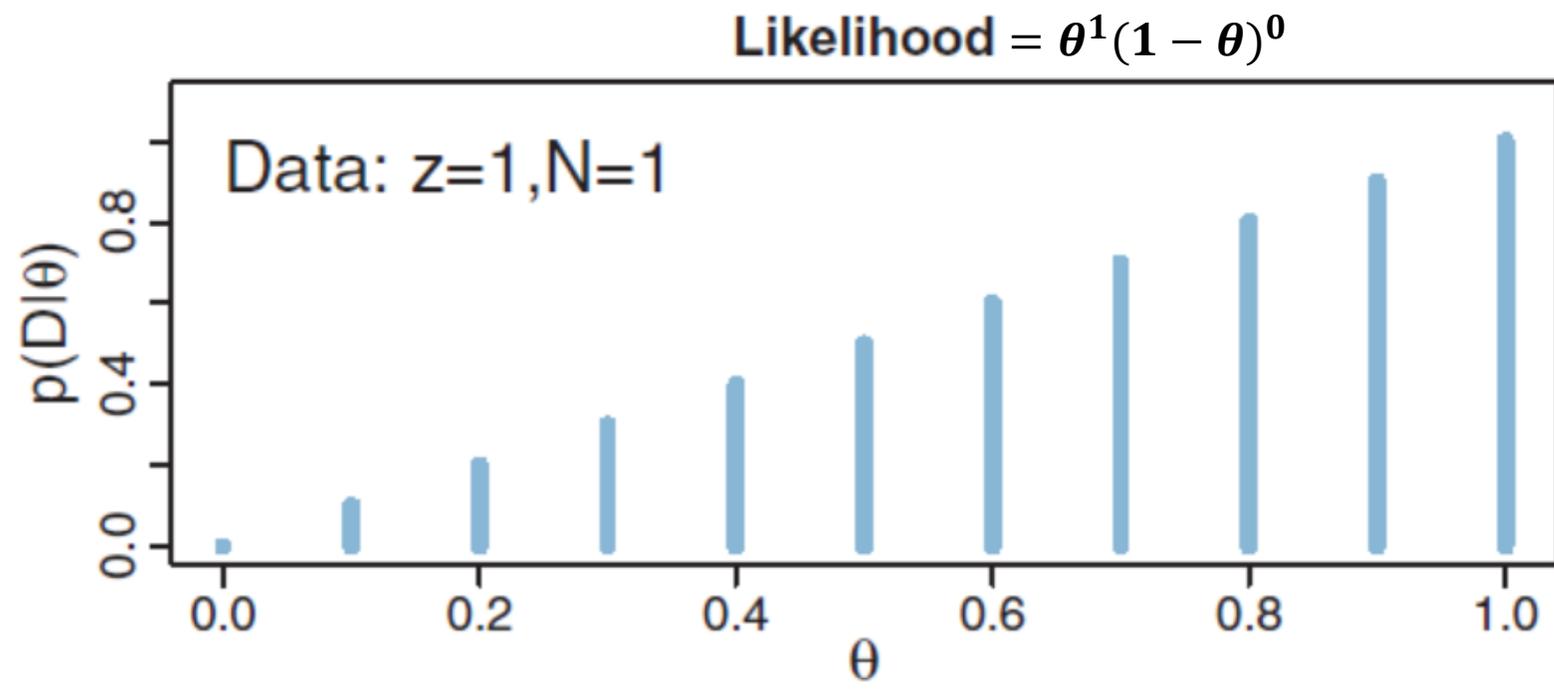
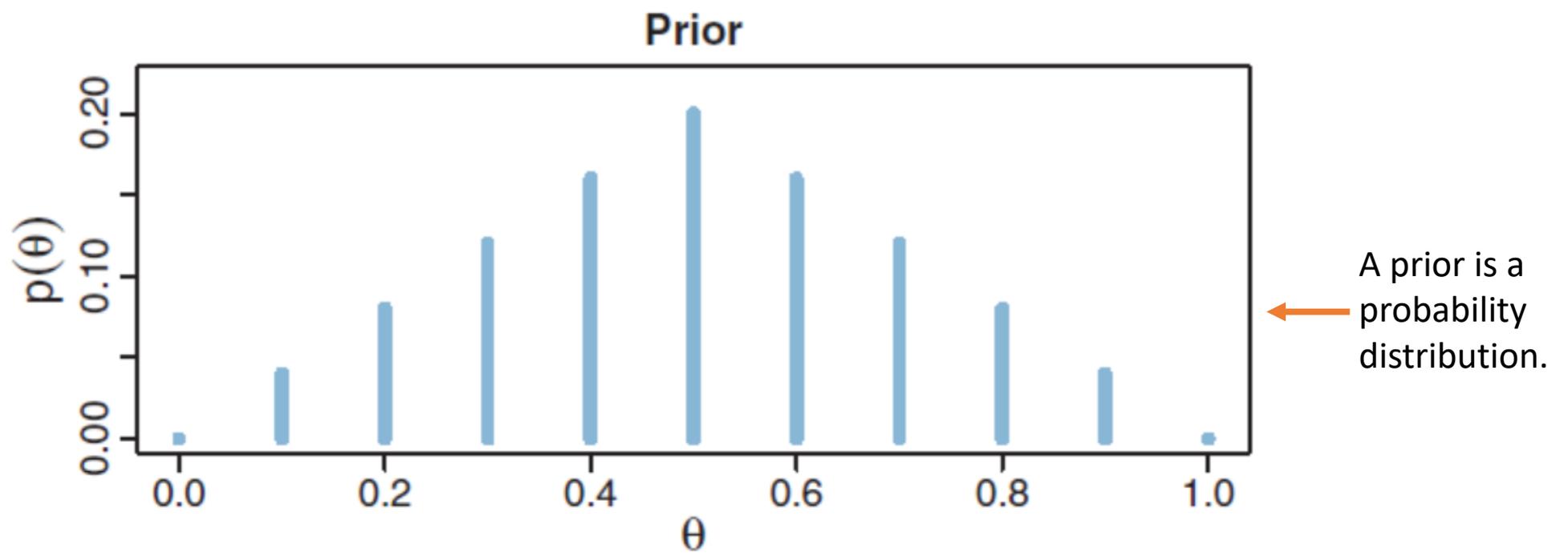
- Assume that the outcomes are independent of each other, we derive the following formula for the likelihood of a set of outcomes  $\{\gamma_i\}$ :

$$\begin{aligned} p(\{\gamma_i\}|\theta) &= \prod_i p(\gamma_i|\theta) = \prod_i \theta^{\gamma_i} (1 - \theta)^{(1-\gamma_i)} \\ &= \theta^{\sum_i \gamma_i} (1 - \theta)^{\sum_i (1-\gamma_i)} = \theta^{\text{\#heads}} (1 - \theta)^{\text{\#tails}}. \end{aligned}$$

- Later, we will refer to #heads as  $z$ , the number of flips as  $N$ , and hence the #tails as  $N - z$ .

# Prior

- Suppose that there are only 11 discrete values of the parameter  $\theta$ , namely  $\theta = 0.0, \theta = 0.1, \dots, \theta = 1.0$ .
- The prior distribution indicates what we believe about the factory's production of those types.
- Ex. A prior for fair coins has the peak at  $\theta = 0.5$ , and the heights decreases toward either extreme where  $p(\theta = 0) = p(\theta = 1) = 0$ .



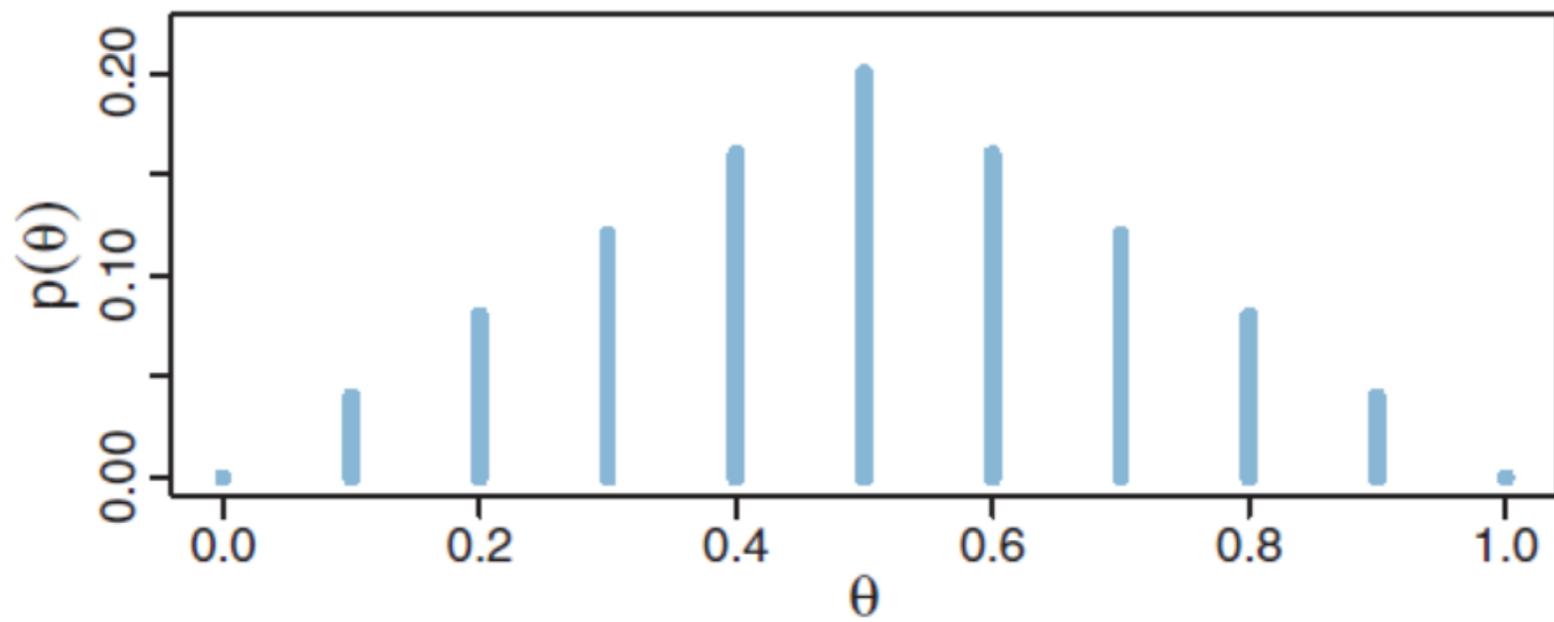
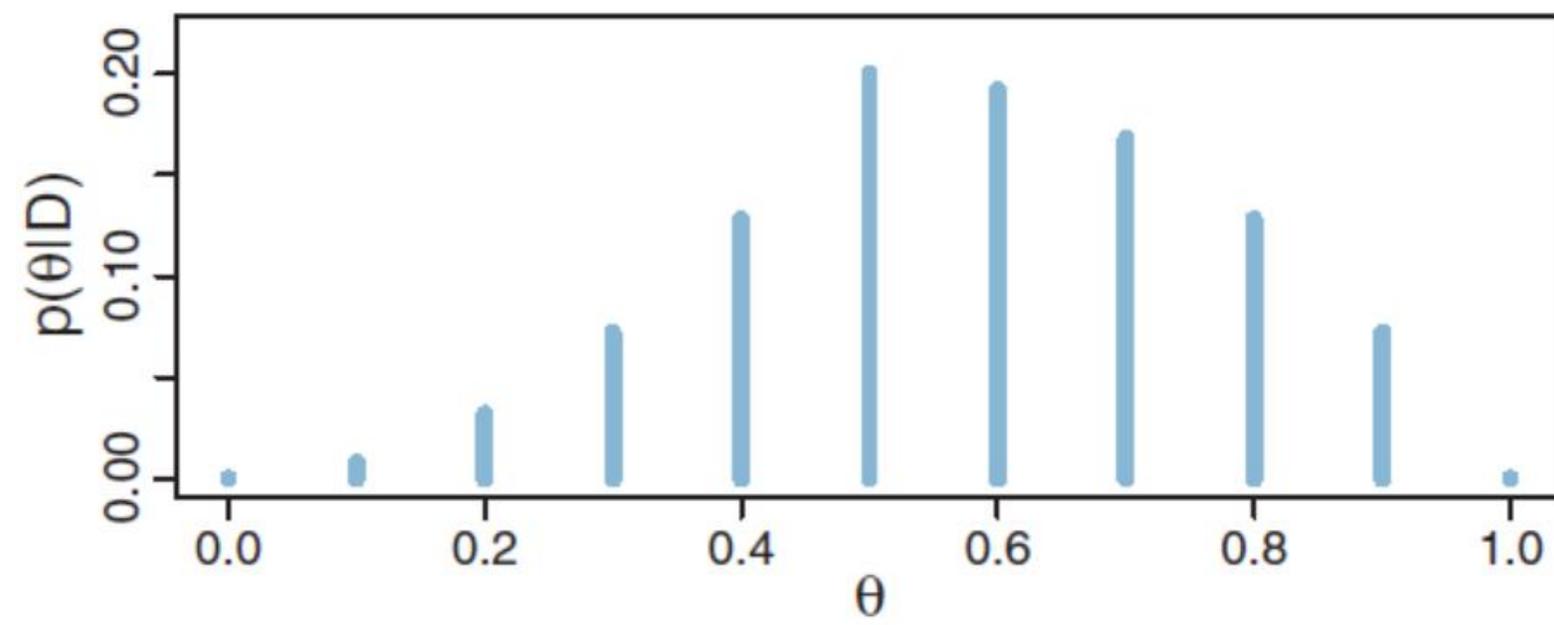
# Posterior

- Given  $z = N = 1$ , the posterior probability is computed from Bayes' rule as the likelihood at  $\theta$  times the prior at  $\theta$  divided by  $p(D) = \sum_{\theta^*} p(D|\theta^*)p(\theta^*)$ .
- Since the data showed a head, the credibility of higher  $\theta$  values has increased:  
 $p(\theta = 0.4|D)$  is less than  $p(\theta = 0.6|D)$ .
- In fact, the posterior becomes a left-skewed distribution.

# Quiz 5a

- What is the ratio of blue binder clips before you see the data?
- What is the ratio of blue binder clips after you see the data?

<https://forms.gle/mj7jwQvxhaaTSt1r6>

**Prior****Posterior**

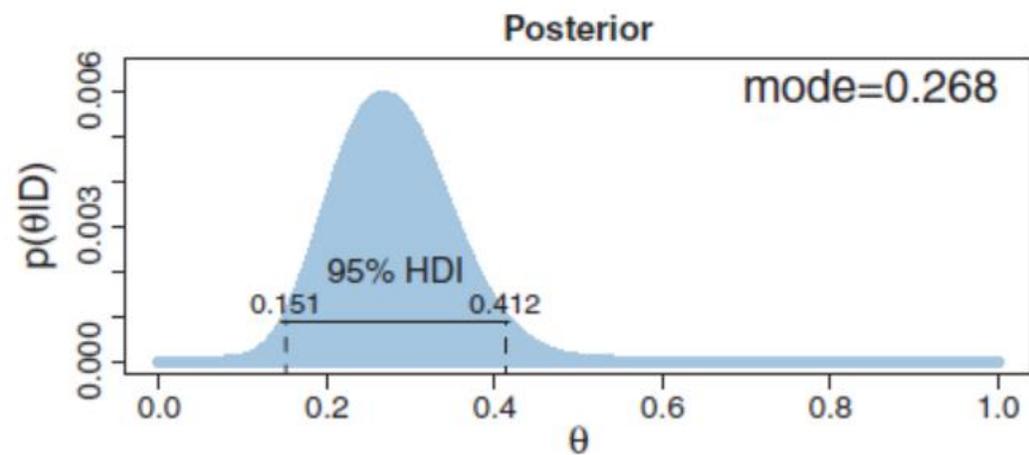
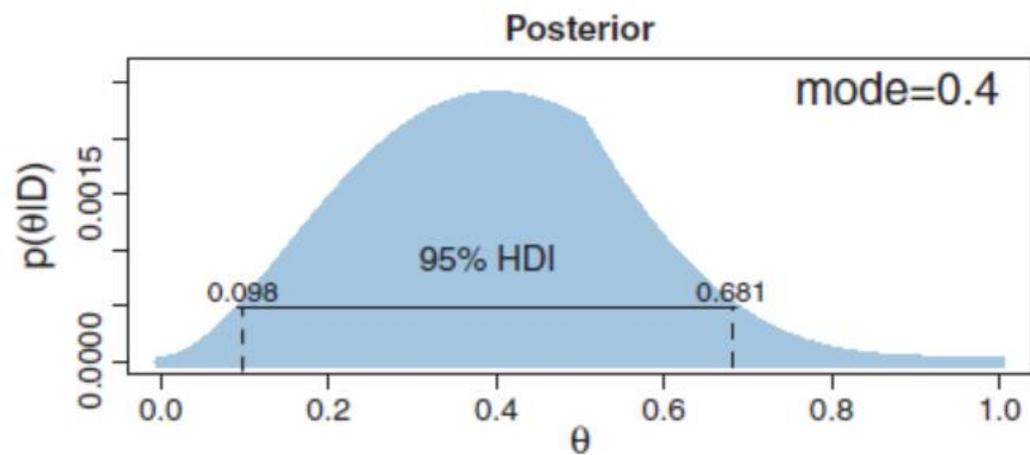
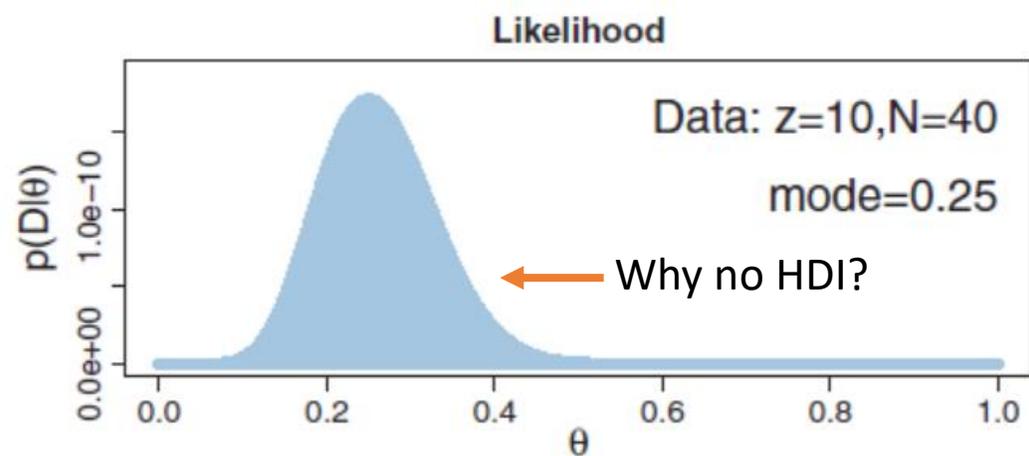
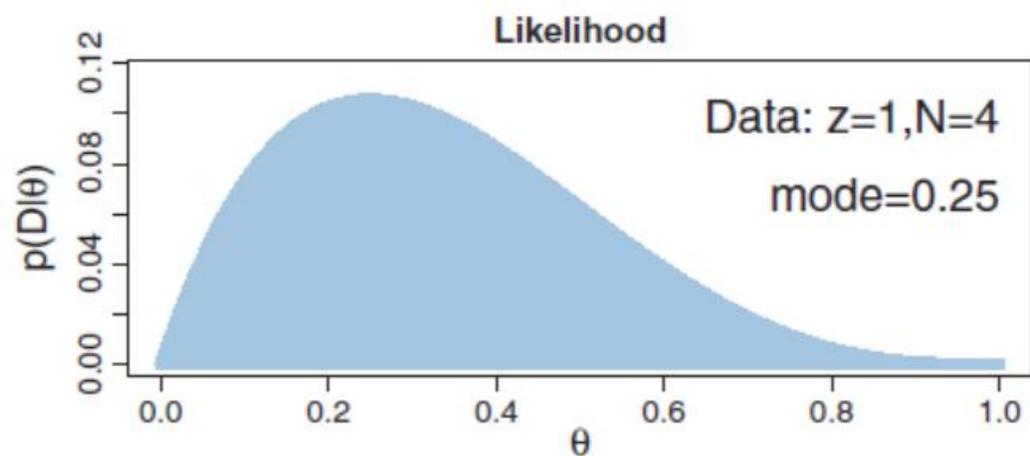
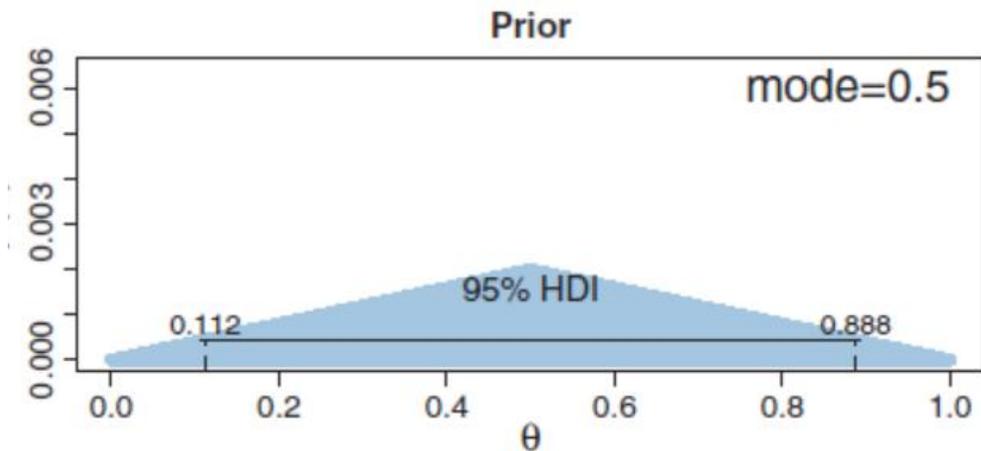
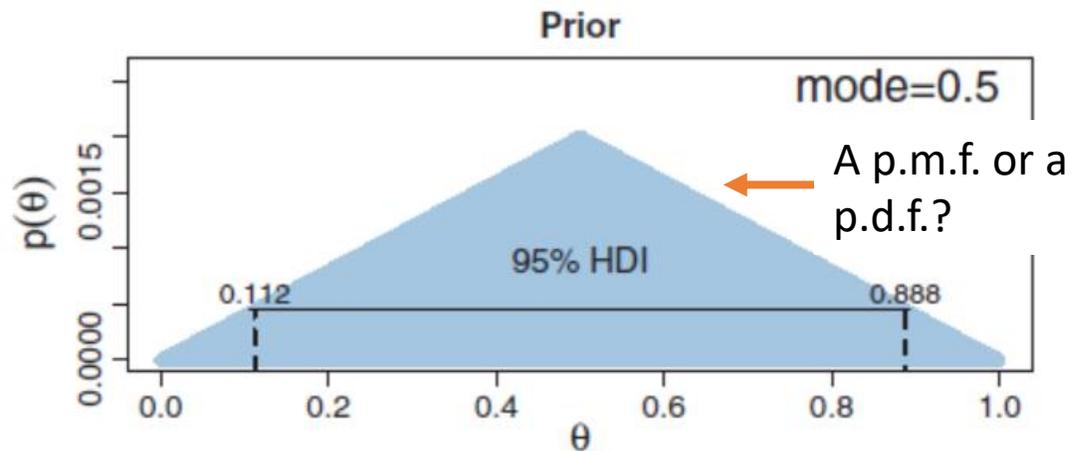
# Variance of Posterior

- Also notice that, with only a single flip of the coin, the posterior probability of large  $\theta$  values such as 0.9 is low.
- The prior distribution has a notable residual effect on the posterior distribution.
- A general phenomenon in Bayesian inference: The posterior is a *compromise* between the prior distribution and the likelihood function (data).

# Posterior as a Compromise

- The compromise favors the prior to the extent that the prior distribution is sharply peaked (stubborn) and the data are few.
- The compromise favors the likelihood function (i.e., the data) to the extent that the prior distribution is flat and the data are many.
- We start with the triangular prior used previously but filled in with 1001 candidate values for  $\theta$ .

← In frequentism, sample size has no influence on the mean of the sampling distribution.



# Influence of Sample Size

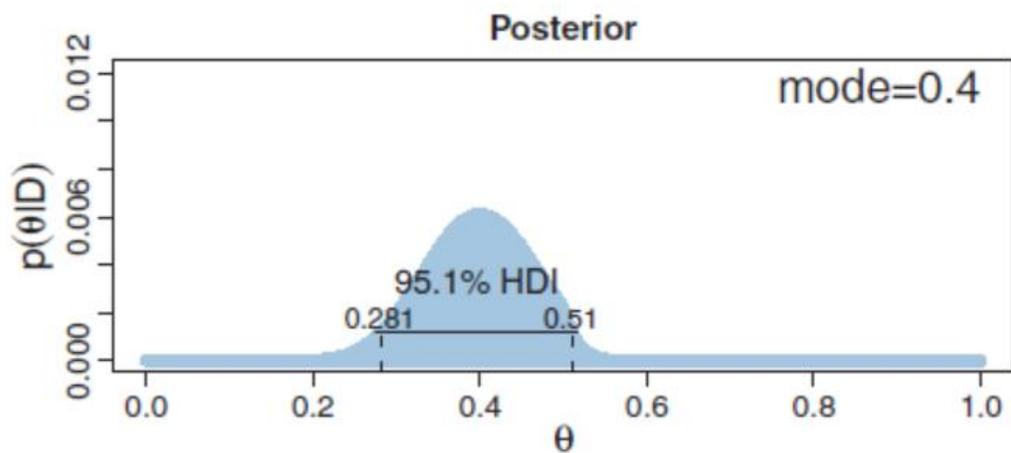
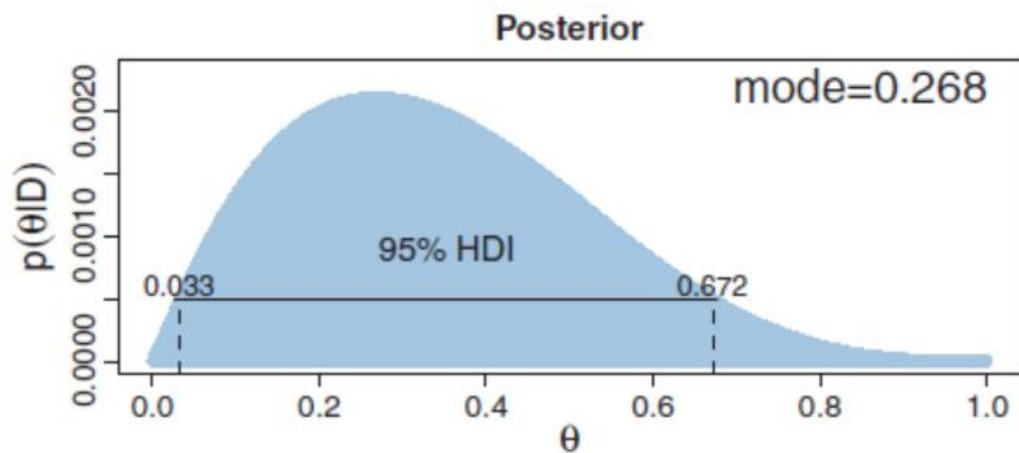
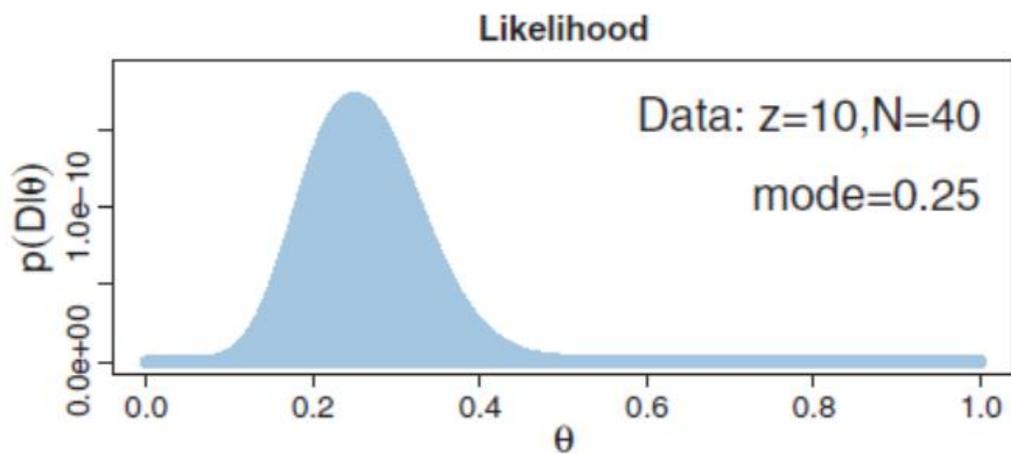
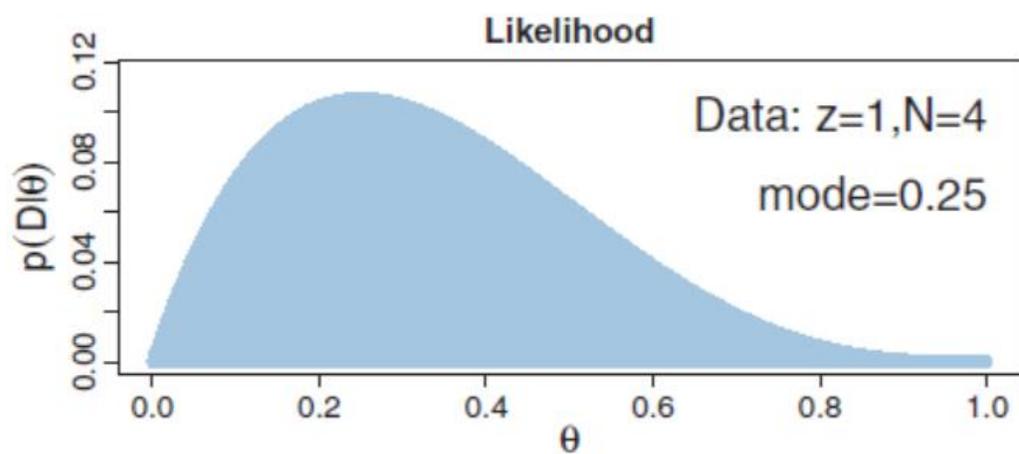
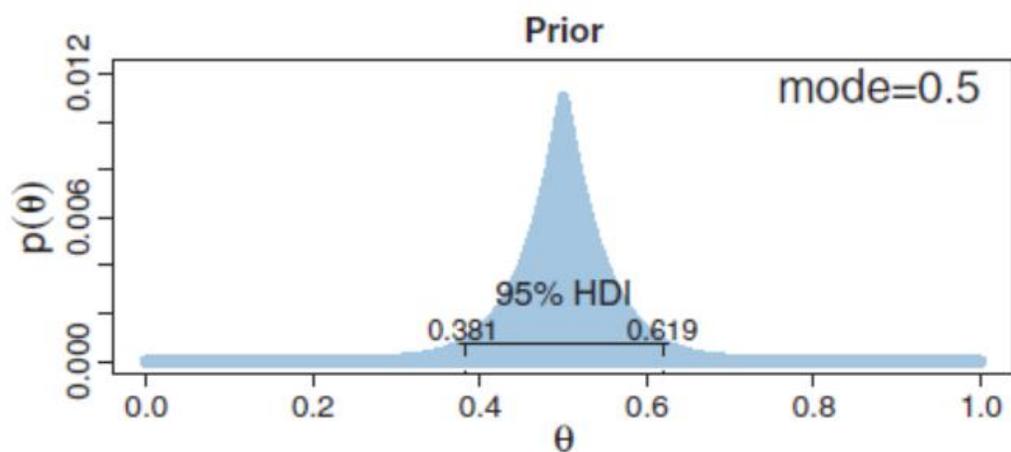
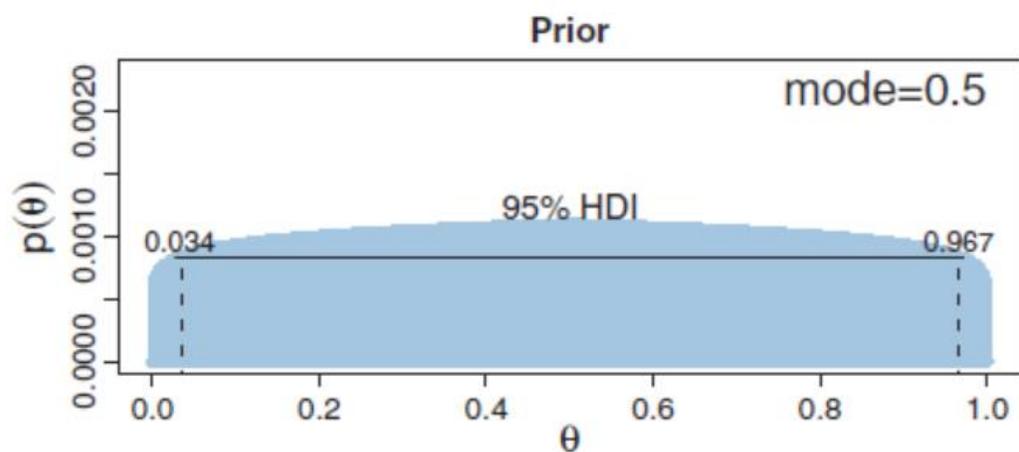
- The data consist of 25% heads for  $N = 4$  (left) and  $N = 40$  flips (right), respectively. So, the likelihood functions both have their modal value at  $\theta = 0.25$ .
- The mode of the posterior is between the mode of the prior and the mode of the likelihood function.
- The posterior mode is closer to the likelihood mode for larger sample size.

# HDI of Posterior

- Notice also that the width of the posterior highest density intervals (HDI) is smaller for the larger sample size.
- This is because of a smaller range of credible underlying biases (likelihood) in the coin implied by the larger sample size.
- In general, the more data we have, the more precise is the estimate of the parameter(s) in the model. ← ≠ accurate

# Influence of Prior

- Whenever the prior distribution is relatively broad compared with the likelihood function, the prior has fairly little influence on the posterior..
- Despite the fact that the sample has a larger size of  $N = 40$ , the prior is so sharp that the posterior distribution is noticeably influenced by the prior.
- This is reasonable: In real applications a sharp prior has been informed by genuine prior knowledge or experience.



# Rational Inference

- In other words, Bayesian inference is intuitively rational.
- With a strongly informed prior that uses a lot of previous data to put high credibility over a narrow range of parameter values, it takes a lot of novel contrary data to budge beliefs away from the prior.
- With a weakly informed prior, it takes relatively little data to shift the peak of the posterior distribution toward the data.

# Tracking the Limits of Knowledge Propagation: How LLMs Fail at Multi-Step Reasoning with Conflicting Knowledge

Yiyang Feng<sup>♣♦</sup>

yiyang.feng@stonybrook.edu

Zeming Chen<sup>♣</sup>

zeming.chen@epfl.ch

Haotian Wu<sup>♣</sup>

haotian.wu@epfl.ch

Jiawei Zhou<sup>♦</sup>

jiawei.zhou.1@stonybrook.edu

Antoine Bosselut<sup>♣</sup>

antoine.bosselut@epfl.ch

♣ EPFL   ♦ Stony Brook University

## Abstract

A common solution for mitigating outdated or incorrect information in Large Language Models (LLMs) is to provide updated facts in-context or through knowledge editing. However, these methods introduce knowledge conflicts when the knowledge update fails to overwrite the model's parametric knowledge, which propagate to faulty reasoning. Current benchmarks for this problem, however, largely focus only on single knowledge updates and fact recall without evaluating how these updates affect

### Knowledge Probing

Who is the current UK Prime Minister?  
The current UK PM is Keir Starmer. Rishi Sunak. ❌

Who is the spouse of Keir Starmer?  
KS's spouse is Victoria Starmer. Victoria Starmer. ✅

Where was Victoria Starmer born?  
VS was born in London. London. ✅

### Knowledge Injection

Where was the spouse of the current UK Prime Minister born?  
Please update your knowledge with following facts:  
The current UK Prime Minister is Keir Starmer. Knowledge Propagation

The current UK Prime Minister is ~~Rishi Sunak~~ Keir Starmer. Rishi Sunak Keir Starmer's spouse is Victoria Starmer. Victoria Starmer was born in London. So the answer is London. ✅

# How to Choose a Prior

- Prior beliefs used in data analysis must be admissible by a skeptical scientific audience.
- When scientists disagree about prior beliefs, the analysis can be conducted with *multiple priors*, to assess the robustness of the posterior against changes in the prior.
- Or, the multiple priors can be *mixed together* into a joint prior, with the posterior thereby incorporating the uncertainty in the prior.

# Difficulty

- Determining the posterior distribution directly from Bayes' rule involves computing the evidence (a.k.a. marginal likelihood).
- In the usual case of continuous parameters, the integral in the denominator of Bayes' rule can be impossible to solve analytically.
- Historically, Bayesian models are thus restricted to relatively simple likelihoods with conjugate priors.

# Approximation Helps

- Instead of analytical mathematical approaches, another class of methods involves numerical approximation of the integral.
- When the parameter space is small, the integral can be approximated by a sum over the many discrete representative values.
- If we represent each parameter with a comb of 1000 values, then for  $P$  parameters there are  $1000^P$  combinations of parameter values.  curse of dimensionality

# MCMC Methods

- In recent decades, Markov chain Monte Carlo (MCMC) methods can generate representative parameter-value combinations from the posterior.
- The MCMC methods help in bypassing the integral in Bayes' rule even in complex models.
- It is the development of these MCMC methods and available fast computers that has allowed Bayesian statistical methods to gain practical use.

# ML26 Exam 1

- Given  $N=10$  and the  $n$  assigned to you is presented as follows:

[https://mx.nthu.edu.tw/~chaoenyu/ml\\_grade.html](https://mx.nthu.edu.tw/~chaoenyu/ml_grade.html)

- What are the 95% HDI limits\*1000?

Answer sheet:

<https://forms.gle/bCM9HEVQDDinijYV8>