Bayesian Data Analysis

Dr. Durba Bhattacharya

St. Xavier’s College (Autonomous), Kolkata.

Contents

- **Bayes Theorem Applied to Parameters and Data.**
- **Effect of prior and data on the posterior through Example**

Example: A series of m coin flips.

The coin is fair or not?

Assumption: The data were generated by a sequence of independent draws from a Bernoulli distribution, parameterized by θ

probability of flipping Heads.

Question: what's the value of θ?

That is, which Bernoulli distribution generated these data?

Let Y_i = 1 if ith flip was Heads,

0 otherwise.

Let y=$\sum_{i=1}^{m} Y_i$: number of heads in m tosses

likelihood model is: f (y|θ) =$\binom{m}{y}\theta^{y}(1-\theta)^{(m-y)}$

Y = $\sum_{i=1}^{m} Y_i$: number of heads in m tosses

likelihood model is: f (y|θ) = $\binom{m}{y}\theta^y(1-\theta)^{(m-y)}$----(i)

Frequentist's approach: θ is an unknown, but deterministic quantity.

To determine the value / range of values for θ that is supported by the data (Point Estimate/ Confidence interval).

The log-likelihood is: l(θ) = k + y log θ + (m − y)log(1 − θ).

$$\frac{dl(\theta)}{d\theta} = 0$$

$$\hat{\theta}_{ML} = \frac{y}{m}$$

Point estimate of θ: The proportion of the flips that are Heads.

Bayes Theorem Applied to Parameters and Data

Model of data specifies the following:

1. probability of particular data values given the model's structure and parameter values;

   p(data values | parameters values): $f(y|\theta)$

   probability that the data could be generated by the model with parameter value θ

2. The probability of the various parameter values, in other words, the prior,

   p(parameters values): $\pi(\theta)$ → credibility of the θ values without the data

Que: What do we really want to know?

Answer: How strongly we should believe in the various parameter values, given the data.
i.e, p(parameters values | data values): $\pi(\theta|y)$

credibility of θ values with the data y taken into account.

Posterior Distribution as Combination of Likelihood and Prior

likelihood function: How likely the data y are, given the model specified by any value of θ

Likelihood specification: $Y \mid \theta \sim f(y \mid \theta)$.

Prior specification: $\theta \sim \pi(\theta)$, where either Y or θ can be vectors.

represents any knowledge about how the data are generated prior to observing Y.

Posterior distribution: The posterior distribution of θ:

$$\pi(\theta \mid y) = \frac{\pi(\theta) f(y \mid \theta)}{m(y)},$$

Joint distn of y and θ

where $m(y) = \int \pi(\theta) f(y \mid \theta) d\theta$ is the marginal density of the data y: overall probability of the data according to the model, determined by averaging across all possible parameter values weighted by the strength of belief in those parameter values.

$\pi(\theta|y) = \dfrac{\pi(\theta)f(y|\theta)}{m(y)}$,

- $m(y) = \int \pi(\theta)f(y|\theta)d\theta$: <u>marginal density of the data y</u>: overall probability of the data according to the model, determined by averaging across all possible parameter values weighted by the strength of belief in those parameter values.

- $m(y)$ is the <u>normalizing constant</u> as it ensures that the posterior distribution $\pi(\theta|y)$ of θ integrates to one.

- $m(y)$ is the <u>prior predictive distribution.</u>

The posterior distribution often presented as $\pi(\theta|y) \propto \pi(\theta)\, f(y|\theta)$

Problem: What distn might be an appropriate choice for the prior $\pi(\theta)$? Why?

Bayesian inference ≡ learning from data using probability.

Start with a probability model (prior + likelihood).

Update beliefs after seeing data → posterior distribution.

Posterior summarizes uncertainty about the model parameters, unobserved or future observations.

Outcome - a full probabilistic description of what we know after seeing the data.

Example 1:

A factory generates coins with the following probability of heads:

$$\theta \rightarrow 0.000\ 0.125\ 0.250\ 0.375\ 0.500\ 0.625\ 0.750\ 0.875\ 1.000$$

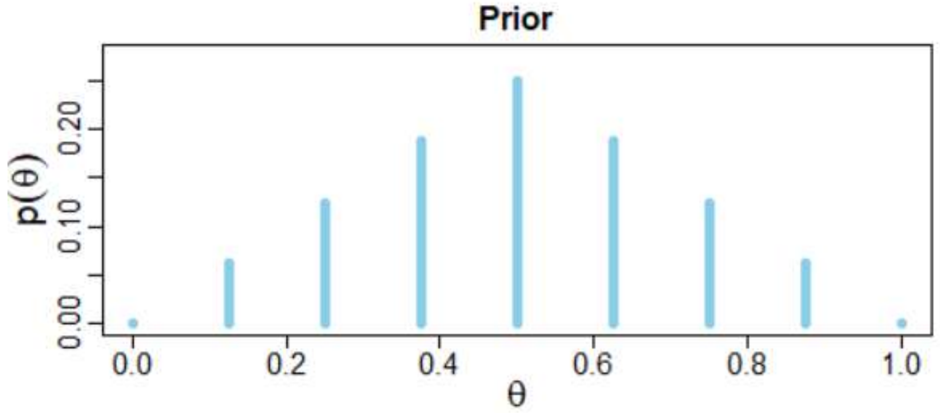
Data: Head/ Tail in a single flip of a coin.

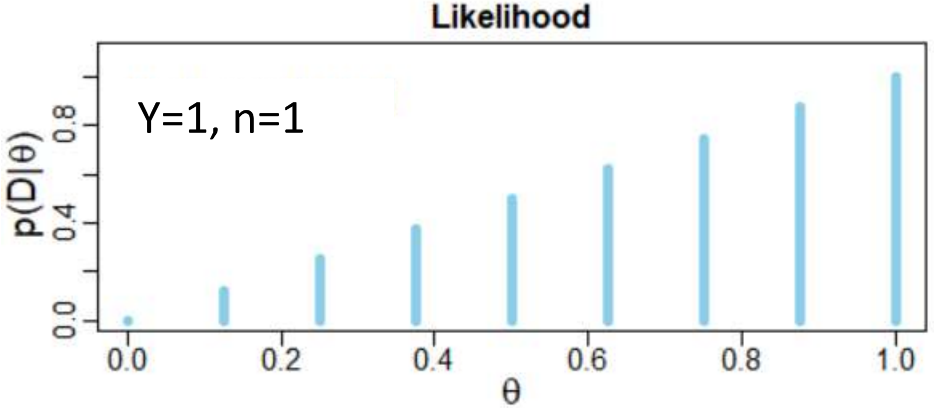
Likelihood: $f(y|\theta) = \theta^{y}(1-\theta)^{1-y}$

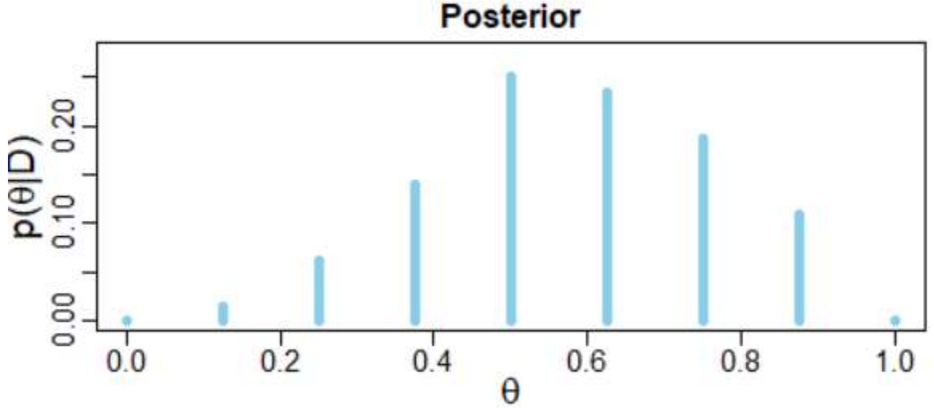
The prior distribution indicates our belief about the factory's production of these types of coins.

Suppose we believe that the factory tends to produce fair coins, i.e, θ near 0.5, and we assign lower prior credibility to biases far above or below θ = 0.5.

The next step is: collecting the data and applying Bayes' rule to re-allocate credibility across the possible parameter values.







We toss the coin once and observe H.

In this case, the data consist of a single head in a single toss, so that y = 1

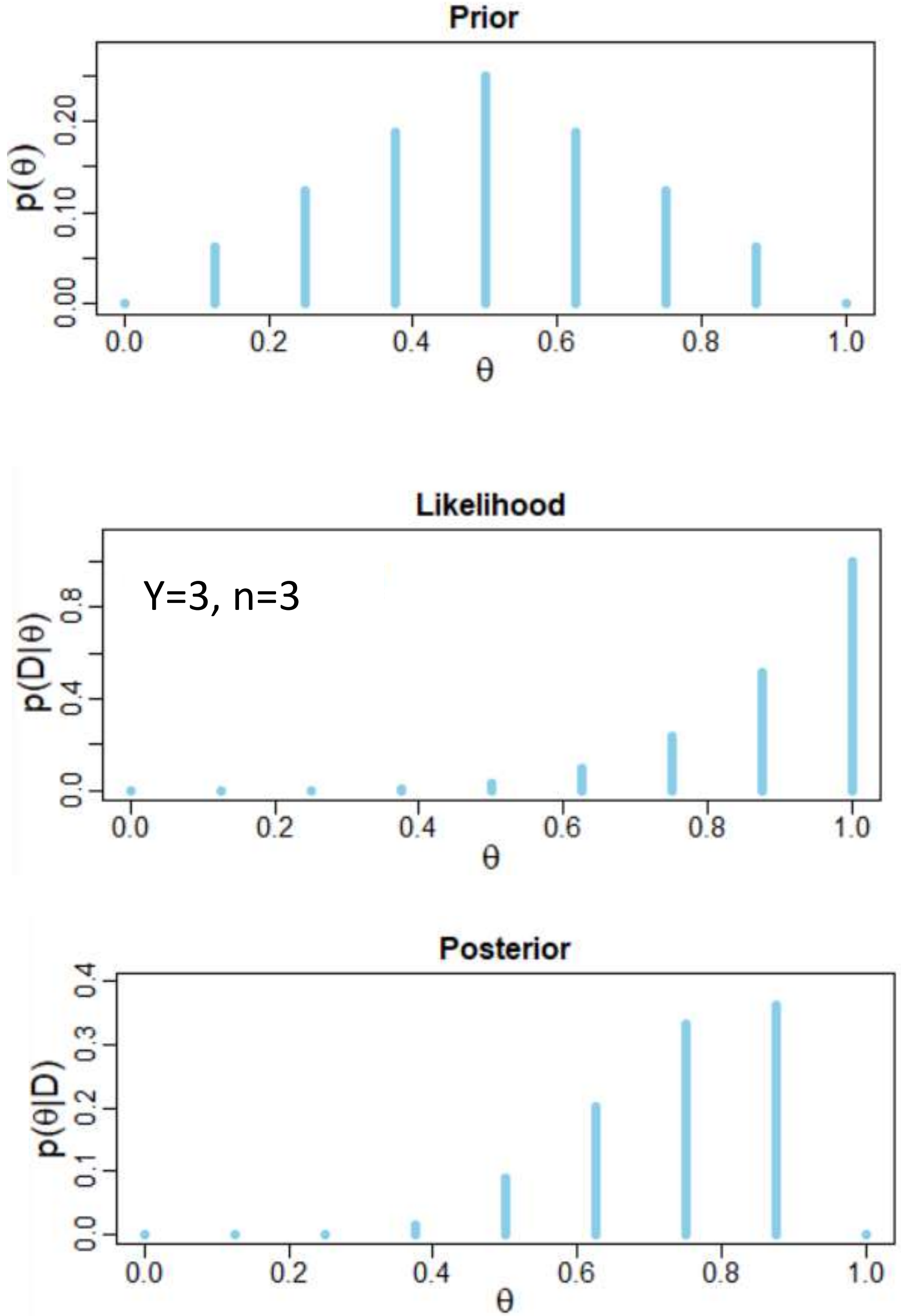
Likelihood: $f(y=1|\theta) = \theta^{y}(1-\theta)^{1-y}$

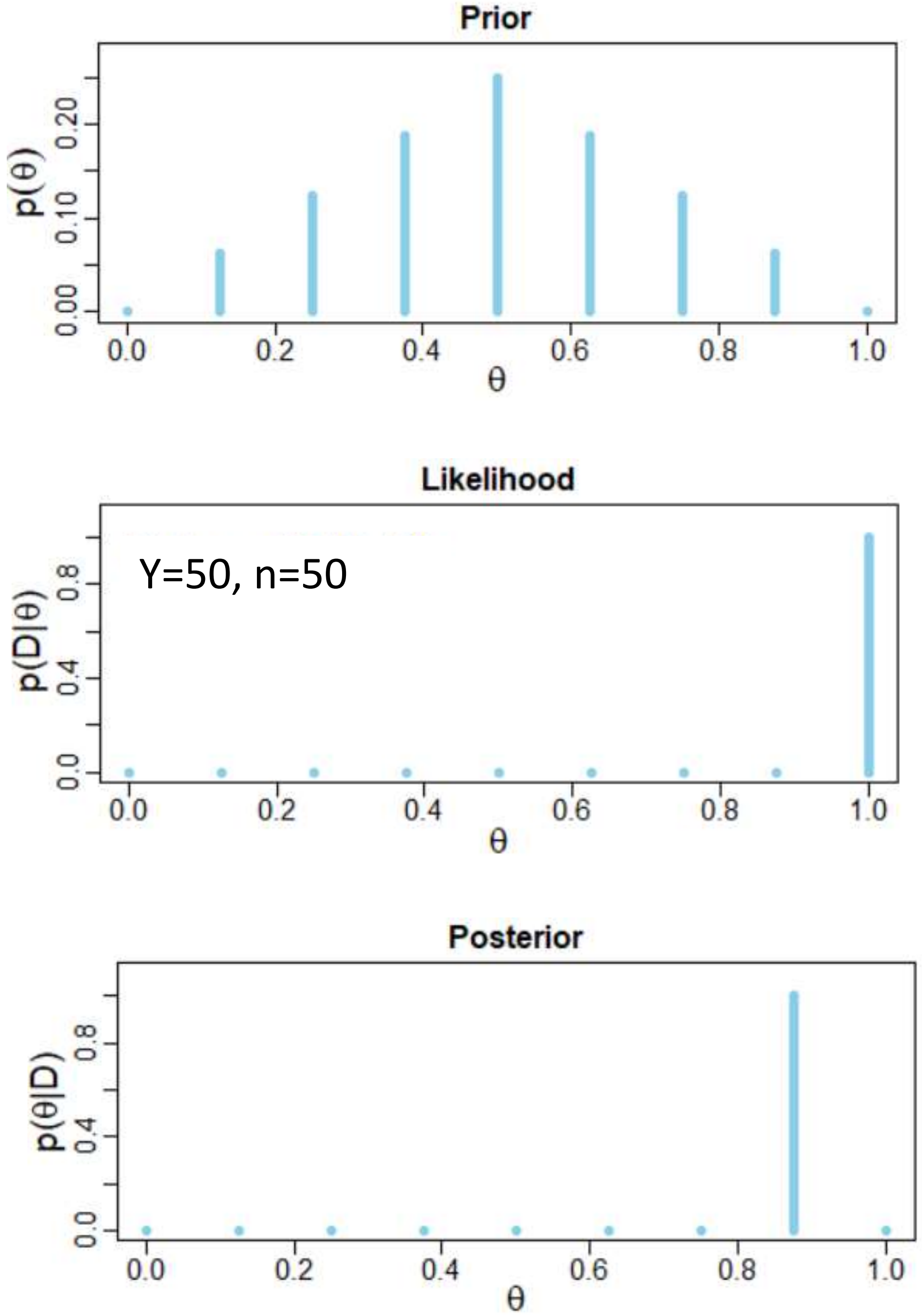
As the data showed a head, the credibility of higher θ values has increased.

- In the prior distribution, $p(\theta = 0.4) = p(\theta = 0.6)$, but in the posterior distribution, $p(\theta = 0.4|D) < p(\theta = 0.6|D)$.
- Despite the data showing 100% heads (single toss), the posterior probability of large θ values such as 0.9 is low.

⇒ The prior distribution has a notable effect on the posterior distribution because of only a single observation.

- This illustrates a general phenomenon in Bayesian inference: The posterior is a compromise between the prior distribution and the likelihood function


Prior
p(θ)
0.00 0.10 0.20
0.0 0.2 0.4 0.6 0.8 1.0
θ
Likelihood
Y=3, n=3
p(D|θ)
0.0 0.4 0.8
0.0 0.2 0.4 0.6 0.8 1.0
θ
Posterior
p(θ|D)
0.0 0.1 0.2 0.3 0.4
0.0 0.2 0.4 0.6 0.8 1.0
θ


Prior
p(θ)
0.00 0.10 0.20
0.0 0.2 0.4 0.6 0.8 1.0
θ
Likelihood
Y=50, n=50
p(D|θ)
0.0 0.4 0.8
0.0 0.2 0.4 0.6 0.8 1.0
θ
Posterior
p(θ|D)
0.0 0.4 0.8
0.0 0.2 0.4 0.6 0.8 1.0
θ

References

1. **Bayesian Data Analysis; Andrew Gelman, John B. Carlin, Hel S. Stern, David B. Dunson, Aki Vehtari and Donald B. Rubin.**

2. **An Introduction to Bayesian analysis: theory and Methods; Jayanta Kumar Ghosh, Mohan Delampady, Tapas Samanta.**