ISTA410/INFO510 Bayesian Modelling and Inference

Lecture 1 - Course Introduction

Dr. Kunal Arekar
College of Information
University of Arizona, Tucson

August 25, 2025

Bayesian statistic

is a theory in the field of statistics based on the Bayesian interpretation of probability, where probability expresses **a degree of belief in an event.** (*Wikipedia*)

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Bayes' theorem

Thomas Bayes
(1701 – 1761)

Pierre-Simon Laplace
(1749 – 1827)

But what does it mean?

Probability and statistics

Statistics is the study of uncertainty.

How do we measure it?

One of the ways to deal with uncertainties is **probability**

Examples,

$P(X=4)$ - rolling a fair six-sided die

P(fair) - probability that the die is fair

P(drop packet) – probability that your internet connection drops a packet

$P(C1 > C2)$ – what is the probability that router from company 1 is better than company 2

P (rain) – what's the probability that it will rain tomorrow.

P(universe expands) – what is the probability that the universe will expand forever

Probability and statistics

Three different frameworks under which probabilities can be defined

Classical: Outcomes that are equally likely have equal probabilities.
E.g., $P(X=4)$ - rolling a fair six-sided die. Since there are 6 equally likely outcomes possible, the probability of rolling a 4 is 1/6

Frequentist: Relative frequency of events in a hypothetical sequence of events.
E.g., Rolling fair six-sided die infinite number of times, $P(x=4)$ will be 1 in 6 or 1/6
P(drop packets) = 1/10000; if we lose 1 in 10000 packets

Bayesian: the probability represents your own perspective, it's your measure of uncertainty; it takes into account what you know about a particular problem. You may have a particular information about the events in question that help you change your perspective about it.
E.g., P(fair) – probability that the die is fair. If you have some information about fairness of the die, then your probability might differ from someone else's without that information

Conditional probability: the probability that an event B will occur given the knowledge that an event A has already occurred

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)}$$

Example:

30 students in a class of which 9 are female;
out of the 30, 12 students are Biology major, of which 4 are female

What is the prob. that someone is a female, P(F) = 9/30 = 3/10

P(Bio) = 12/30 = 2/5

P(F and Bio) = 4/30 = 2/15

Now let's ask conditional probability questions...

P(F | Bio) = P(F and Bio) / P (Bio) = 2/15 ÷ 2/5 = 1/3

P(F | Bio') = P(F and Bio') / P (Bio') = 5/30 ÷ 18/30 = 5/18

But when the events are **independent** of each other

$P(B|A) = P(B)$ $\qquad$ $P(B \text{ and } A) = P(B)\,P(A)$

it doesn't matter whether or not A occurred

In previous example

$P(F|Bio) \neq P(F)$, therefore these events are not independent

Bayes theorem

Bayes theorem is reverses the direction of conditioning

Likelihood

prior probability

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B|A)\,P(A) + P(B|A')P(A')}$$

sum of the product of all probabilities of mutually exclusive hypotheses and corresponding conditional probabilities

Posterior probability

$$P(Bio|F) = \frac{P(F|Bio)\,P(Bio)}{P(F|Bio)\,P(Bio) + P(F|Bio')P(Bio')} = \frac{(1/3)\,(2/5)}{(1/3)\,(2/5) + (5/18)(3/5)} = 4/9$$

When the above **posterior probability** compared to direct conditional probability

$$P(Bio|F) = P(Bio \text{ and } F) / P(F) = (4/30) \div (9/30) = (4/9)$$

Another example of early test for HIV antibodies, the ELISA test

$P(+|HIV) = 0.977$ $P(-|noHIV) = 0.926$ $P(HIV) = 0.0026$

Probability of HIV among North Americans

So now, we select one person randomly from north America, and they tested positive for HIV

$$P(HIV|+) = \frac{P(+|HIV)\,P(HIV)}{P(+|HIV)\,P(HIV) + P(+|HIV')P(HIV')}$$

$$P(HIV|+) = \frac{(0.977)\,(0.0026)}{(0.977)\,(0.0026) + (1 - 0.926)(1 - 0.0026)} = 0.033$$

The probability is less than 4 percent even though they tested +ve on a test that is normally quite accurate

HIV is a rare disease; false +ve >> true +ve

Have implications for policy making – make testing mandatory

*WEISS, S. H., JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION **253**: 221 (1985)*

In this course…

- Introduction to Bayesian Modeling
- Bayesian Methodology and Graphical Models
- Exact Inference Methods
- Markov Random Fields and Applications
- Gaussian Mixture Models (GMM)
- Expectation Maximization (EM)
- Hidden Markov Models (HMM) and Linear Dynamical Systems (LDS)
- Introduction to MCMC Sampling
- Practical Applications of Bayesian Methods
- Bayesian Networks
- Non-Parametric Bayesian Models (If time permits)

*****The order of the topics can change over the course of this Semester.*****

Bayes theorem

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{\int_{\theta} P(X|\theta)P(\theta)d\theta}$$

$$P(\theta|X) \propto P(X|\theta)P(\theta)$$

Frequentist vs Bayesian

Frequentist	Bayesian
Probability is "long-run frequency"	Probability is "degree of certainty"
$Pr(X \mid \theta)$ is a sampling distribution (function of X with θ fixed)	$Pr(X \mid \theta)$ is a likelihood (function of θ with X fixed)
No prior	Prior
P-values (NHST)	Full probability model available for summary/decisions
Confidence intervals	Credible intervals

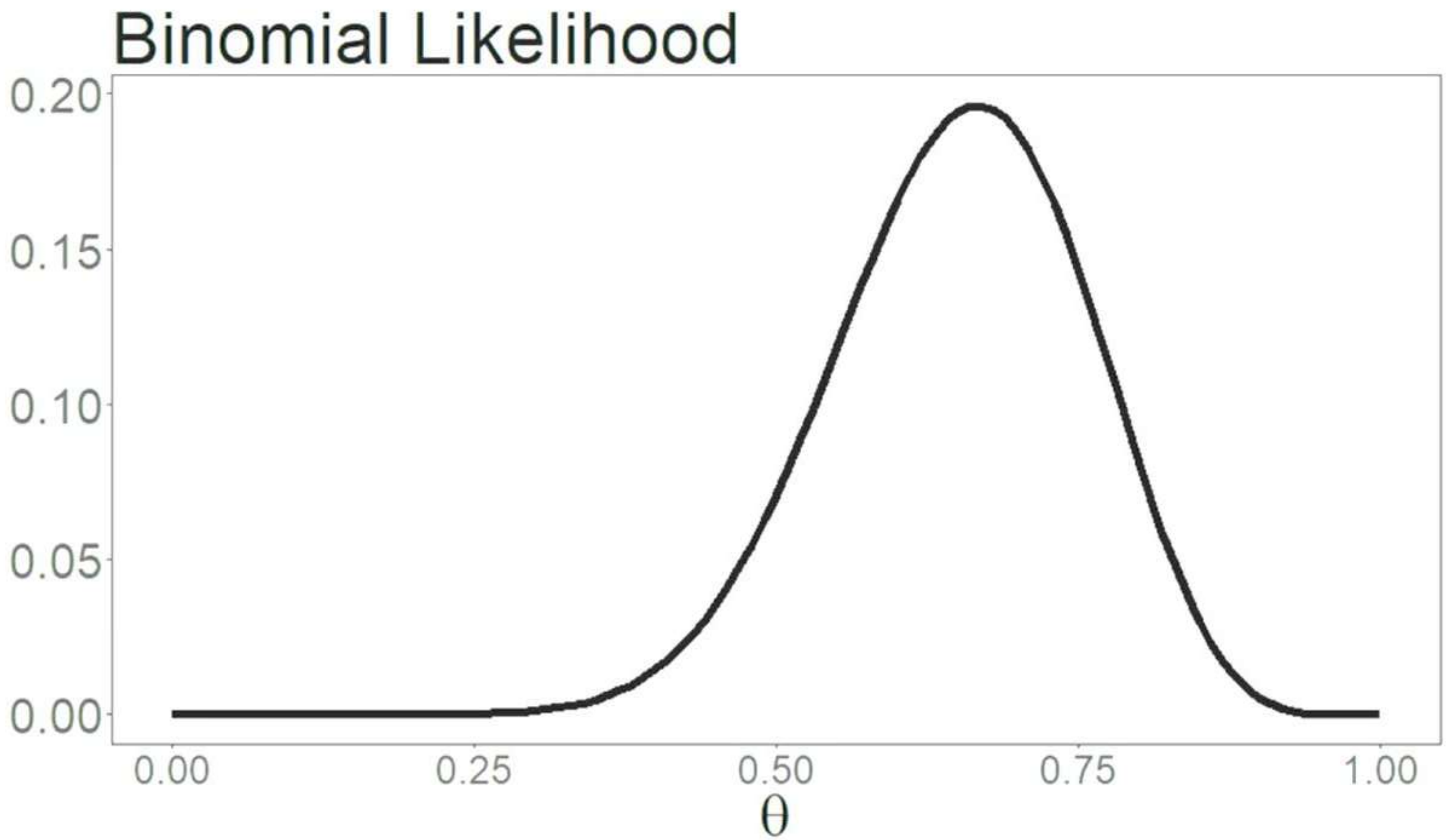
Objective? Subjective?

Binomial example

In 18 trials, we observe 12 successes

The likelihood function is

$$p(x = 12 \mid \theta) \propto \theta^{12}(1 - \theta)^{6}$$



$$p(x = 12 \mid \theta) \propto \theta^{12}(1 - \theta)^{6}$$

Binomial example

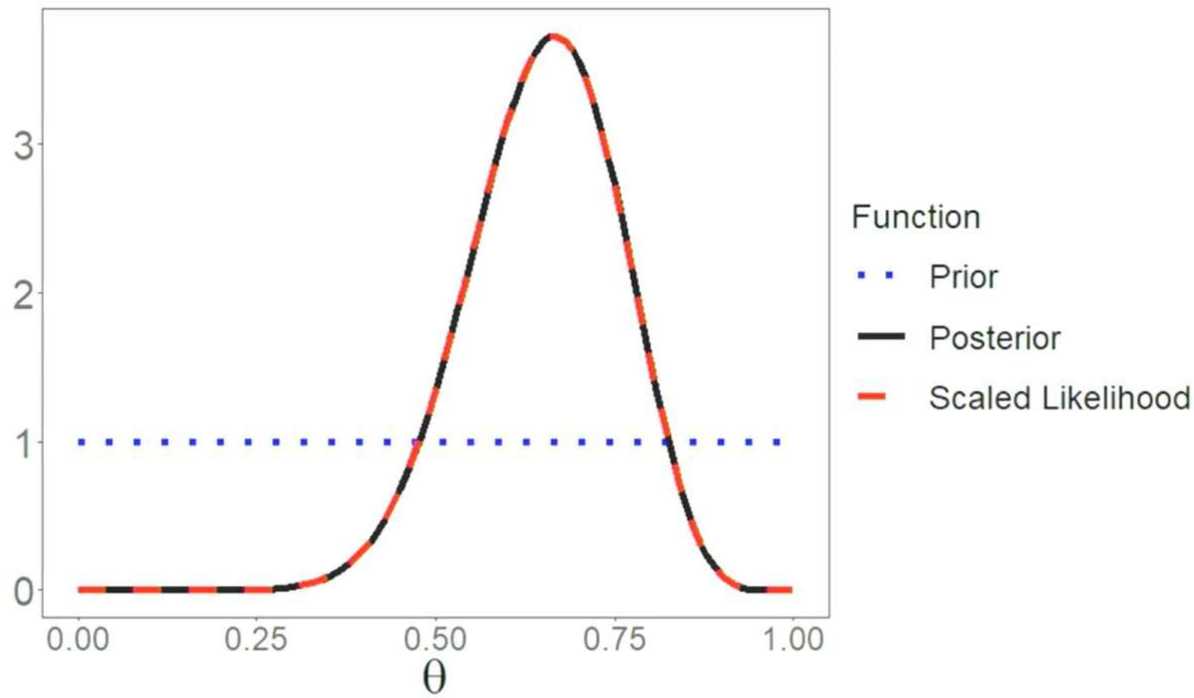
Assume a uniform prior

$$p(\theta) = 1.$$

Then the posterior is

$$p(\theta \mid X) \propto p(\theta)p(X \mid \theta) = p(X \mid \theta).$$

Binomial example



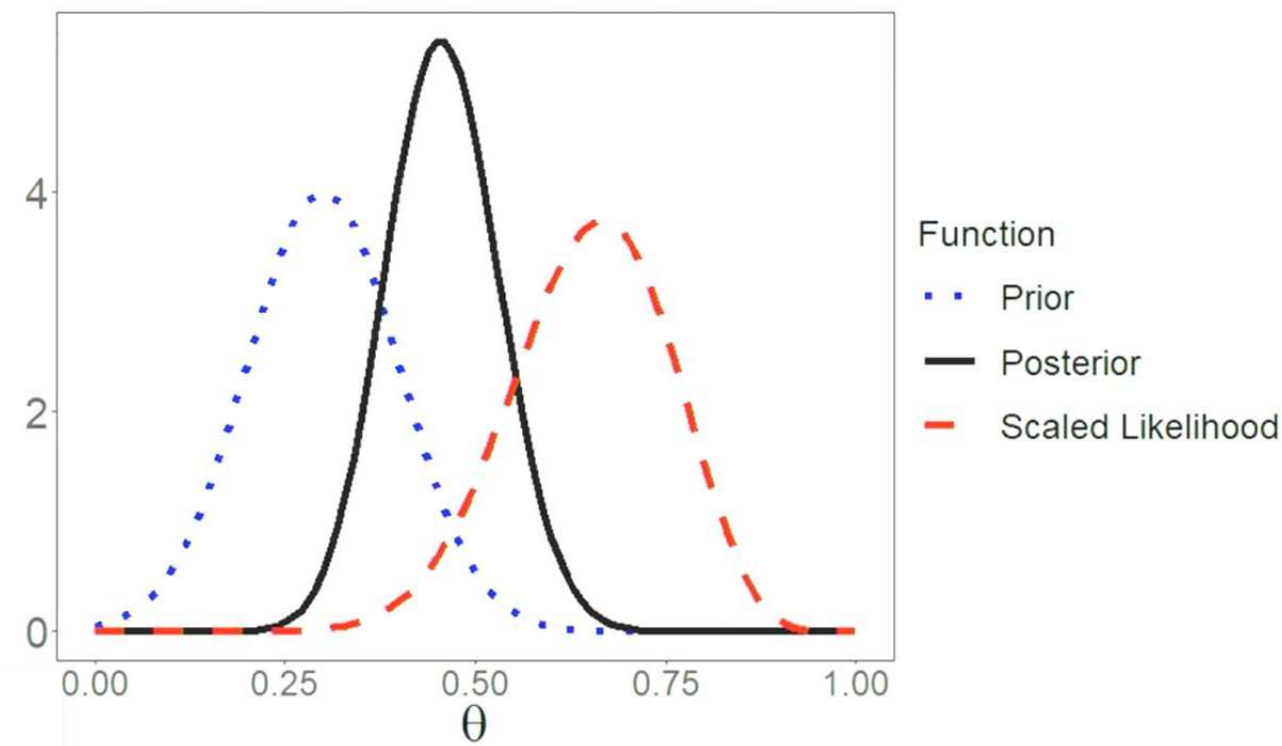
Binomial example

Now lets take a prior which is relatively far away from the data.

Say a normal distribution centred around 0.3 with standard deviation of 0.1

$$\theta \sim N(0.3, 0.1)$$

Binomial example

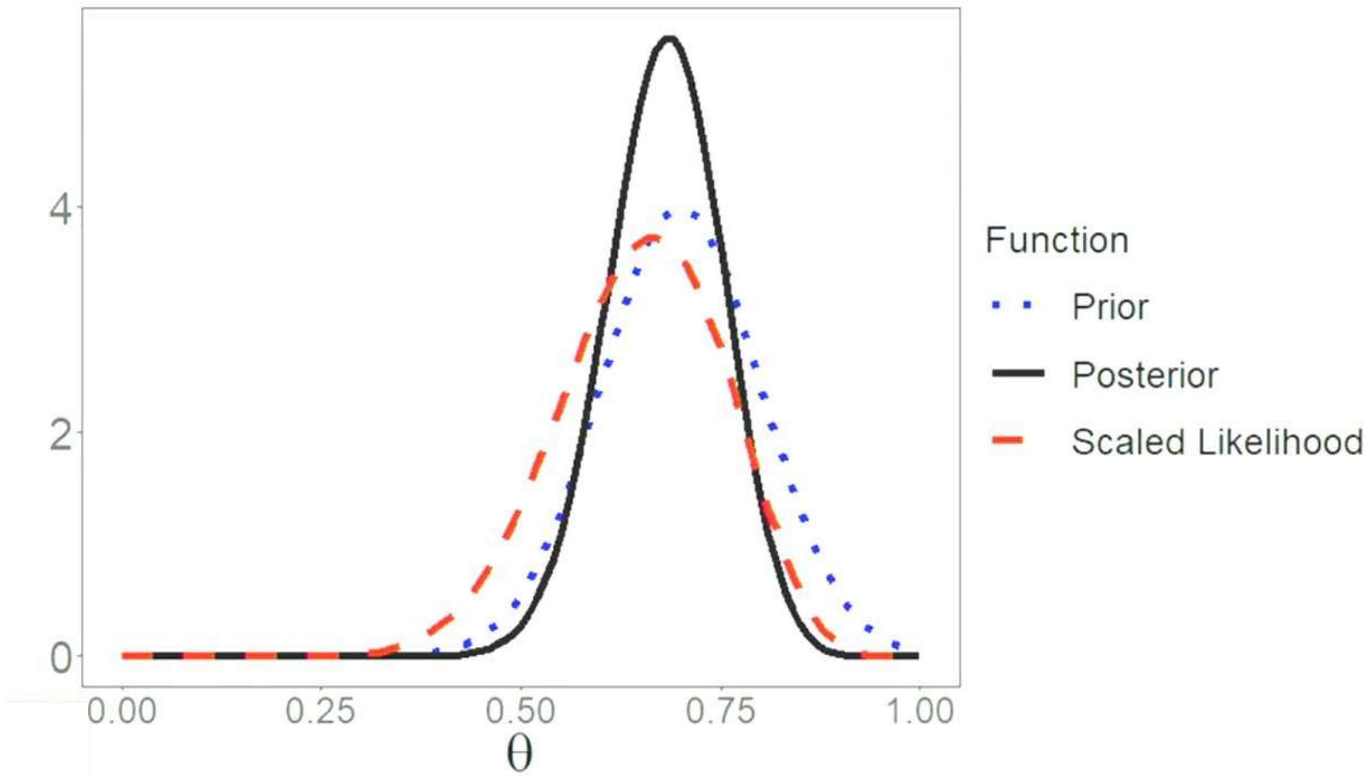


Binomial example

We do end up picking a good prior,

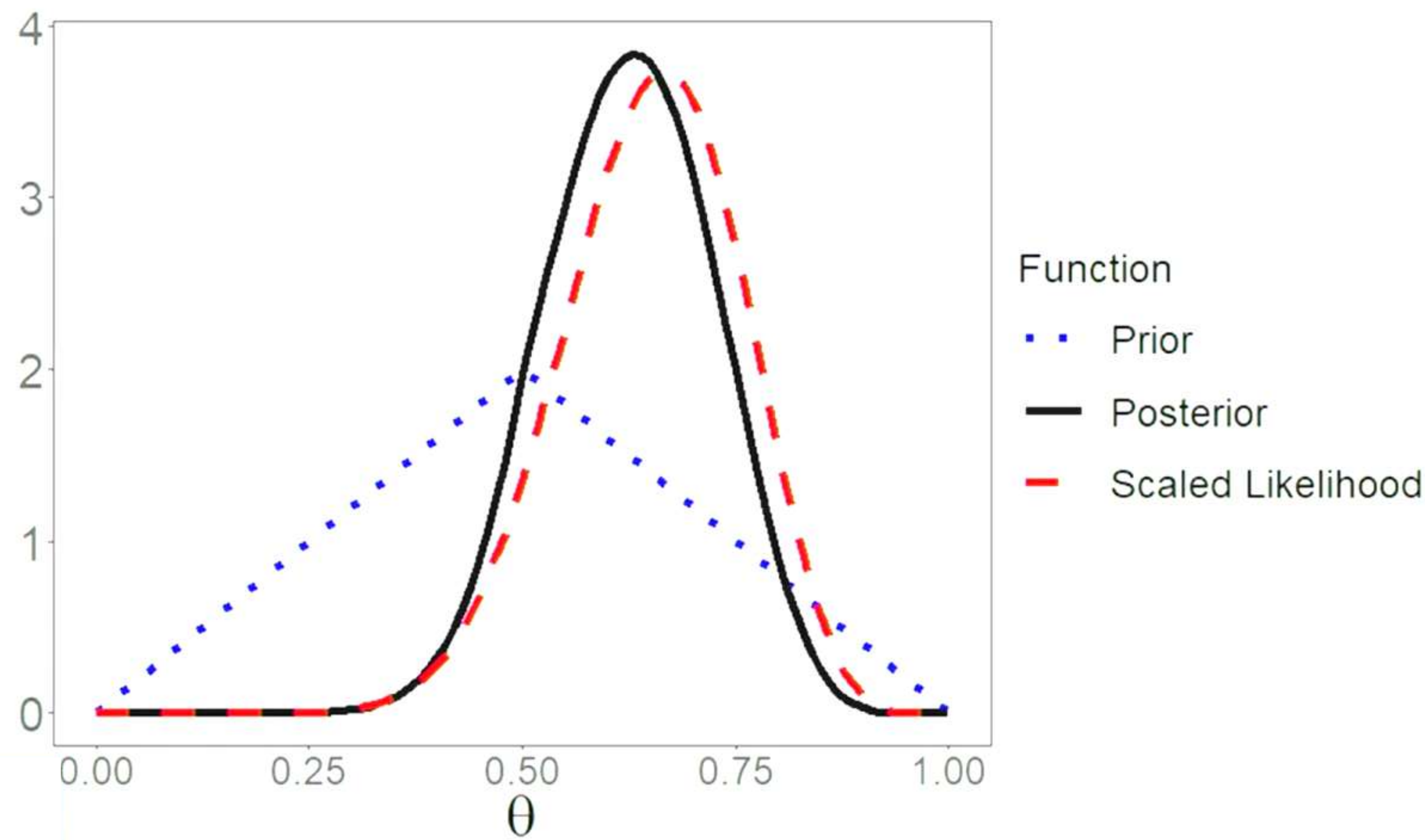
$$\theta \sim N(0.7, 0.1)$$

Binomial example



Binomial example

Weakly informative prior

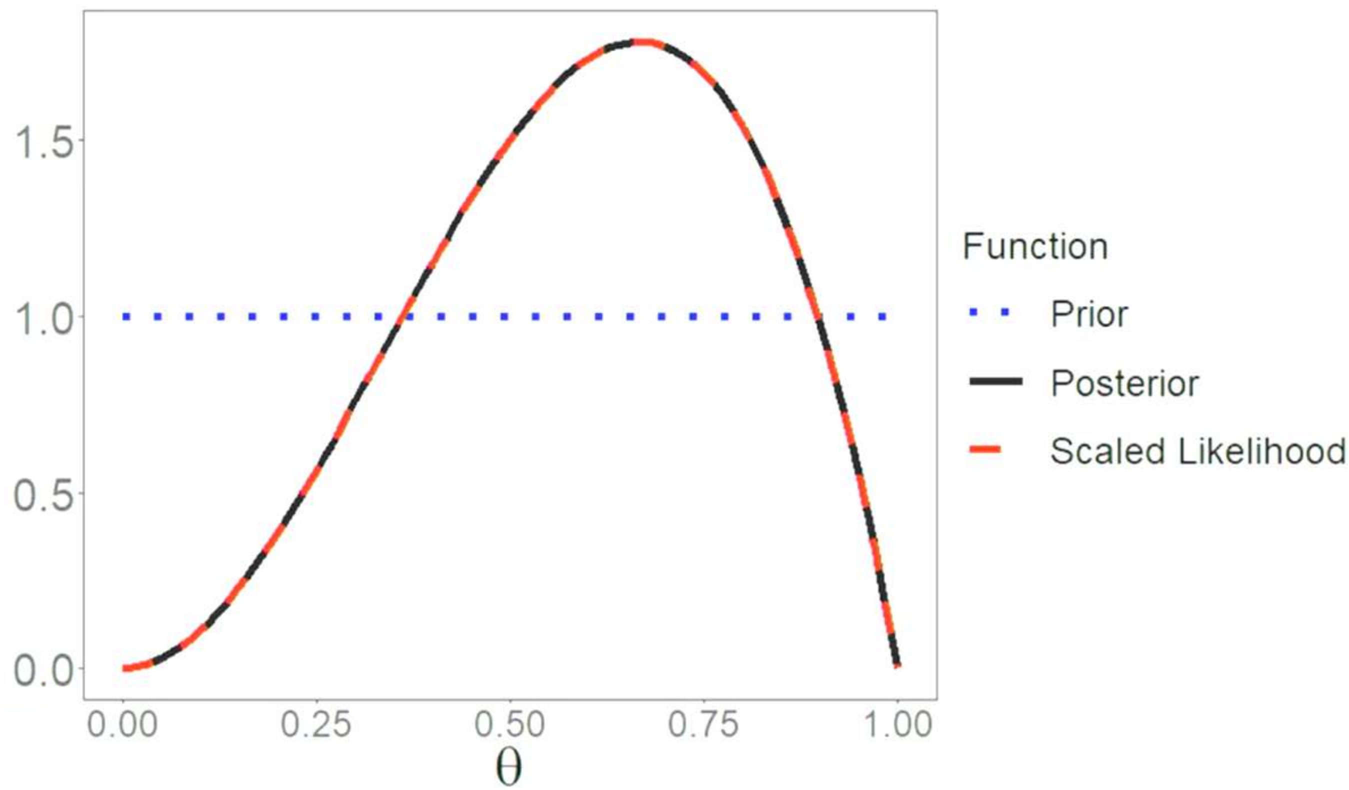


Binomial example

What happens when the data is bad

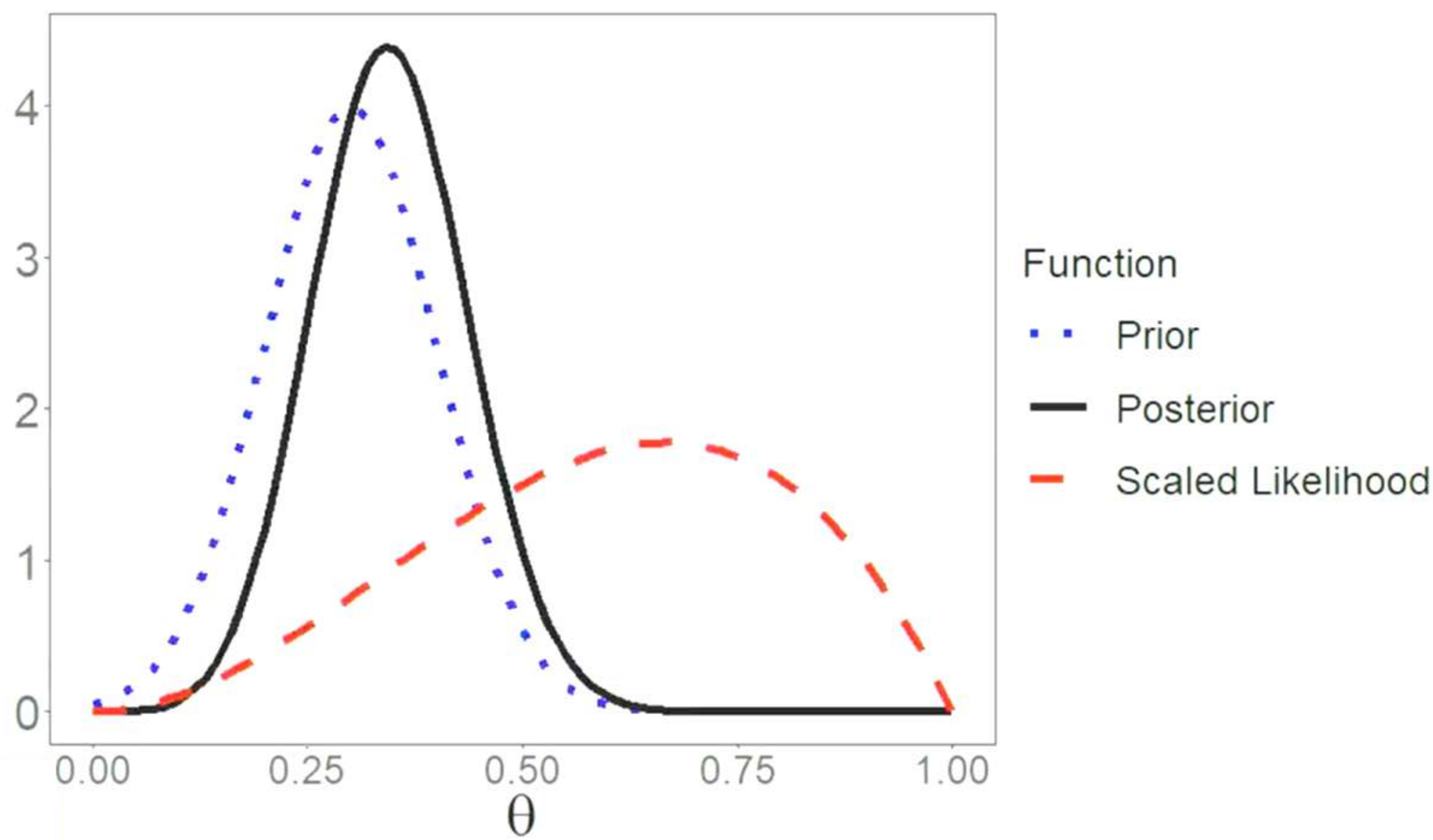
We have 2 successes out of 3 trials – 2/3

Binomial example



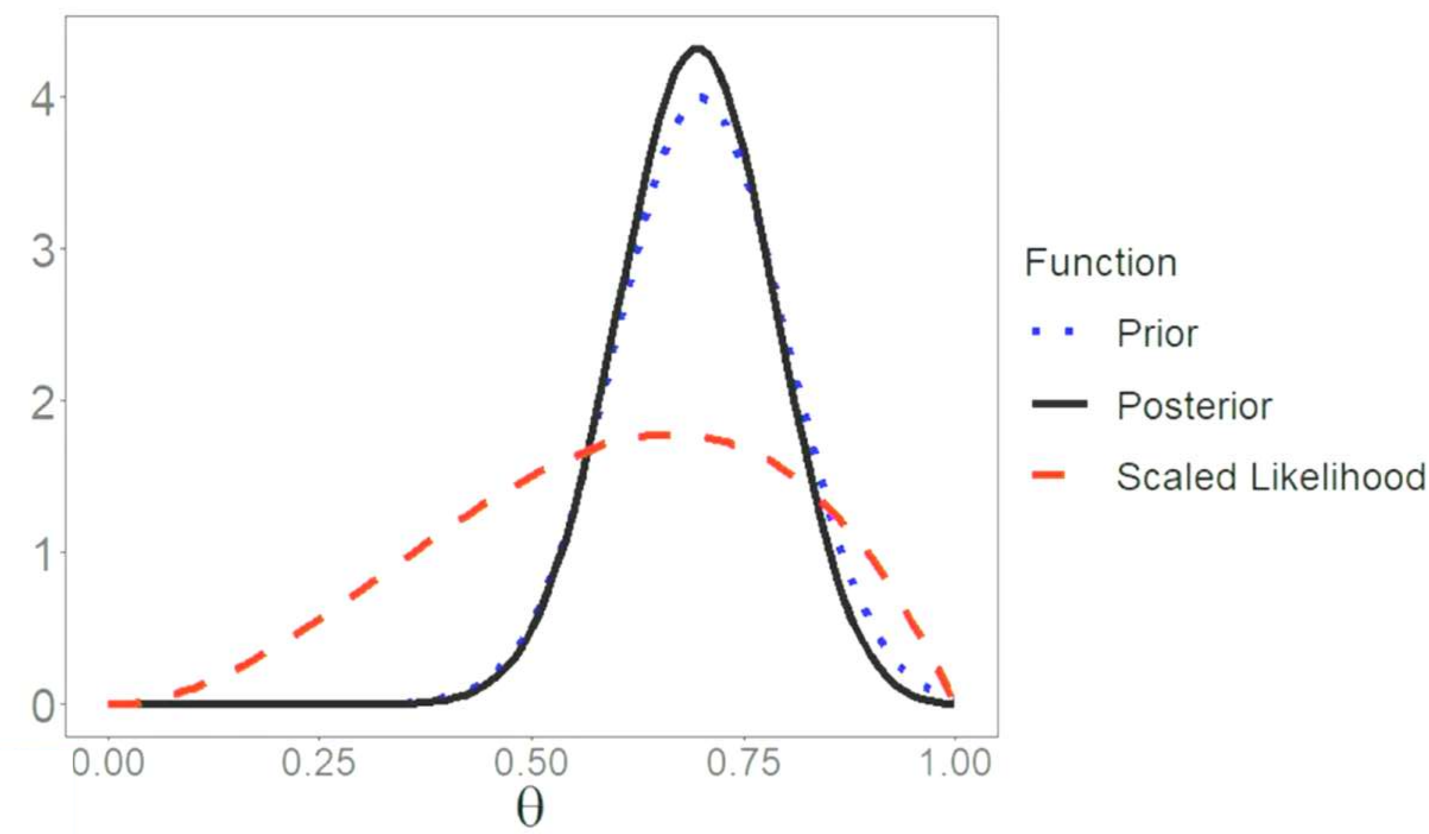
Binomial example

$$\theta \sim N(0.3, 0.1)$$



Binomial example

$$\theta \sim N(0.7, 0.1)$$



Binomial example

Weakly informative prior and bad data