Bayesian Data Analysis

Sourish Das[1]

[1]Mathematics,
Chennai Mathematical Institute, INDIA

Beta-Binomial Models

Beta-Binomial Models

Agenda:

1. Bayesian Setup
2. Bayesian Analysis of Binomial Distribution with Beta Priors
3. Modeling Effectiveness of Marketing Campaign

Bayesian Setup

Modeling the Unknown Quantities

From the Bayesian perspective, there are known and unknown quantities.

- The known quantity is the data, denoted D.
- The unknown quantities are the parameters (e.g. mean, variance, missing data), denoted θ.

To make inferences about the unknown quantities, we stipulate a joint probability function that describes how we believe these quantities behave in conjunction, $p(\theta, D)$.

Bayesian Setup

- Using Bayes' Rule, this joint probability function can be rearranged to make inference about θ

$$
\begin{aligned}
p(\theta|D) &= \frac{p(D|\theta)p(\theta)}{p(D)} \\
&= \frac{L(\theta|D)p(\theta)}{\int L(\theta|D)p(\theta)d\theta}
\end{aligned}
$$

- $L(\theta|D)$ is the likelihood function of θ
- $p(D) = \int L(\theta|D)p(\theta)d\theta$ is the normalizing constant or prior predictive distribution
- It is the normalizing constant because it ensures that the posterior distribution of θ integrates to one.

Bayesian Setup

- It is the prior predictive distribution because it is not conditional on a previous observation of the data-generating process (prior) and it is the distribution of an observable quantity (predictive).

Popular Presentation

The posterior distribution often presented as

$$p(\theta|D) \propto L(\theta|D)p(\theta)$$

i.e., posterior $\propto$ likelihood $\times$ prior

- Why are we allowed to do this?
- Why might not be as useful?

Bayesian Analysis: Binomial Distribution

- Suppose X_1, X_2,...,X_n are independent random draws from same Bernoulli distribution with parameter π (unknown).
- Thus $X_i \sim Bernoulli(\pi)$ for $i = \{1, 2, ..., n\}$ or equivalently $Y = \sum_{i=1}^{n} X_i \sim Bin(n, \pi)$.
- The joint distribution of Y and π is the product of the conditional distribution of Y and the prior distribution π.
- What distribution might be a reasonable choice for the prior distribution of π? Why?

Bayesian Analysis: Binomial Distribution

- If Y $Bin(n, \pi)$, a reasonable prior distribution for p must be bounded between zero and one.

  One option is the uniform distribution $\pi \sim Unif(0, 1)$.

-
$$p(\pi|Y) \propto^{n} C_y \pi^y (1-\pi)^{n-y} \times 1$$

- As it happens, this is a proper posterior density function.

- How can you tell?

Bayesian Analysis: Binomial Distribution

Let $Y \sim Bin(n, \pi)$ and $\pi \sim unif(0, 1)$

$$\begin{aligned} p(\pi|Y) &\propto {}^{n}C_y \pi^y (1-\pi)^{n-y} \times 1 \\ &\propto \pi^y (1-\pi)^{n-y} \end{aligned}$$

- You cannot just call the posterior a binomial distribution because you are conditioning on Y and π is a random variable, not the other way around.

- The pdf of beta distribution which is known to be proper is:

$$Beta(x|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

  where $0 < x < 1$ and $\alpha > 0$, $\beta > 0$

  Note that $\Gamma(k)$ is the Gamma function.

Bayesian Analysis: Binomial Distribution

Let $Y \sim Bin(n, \pi)$ and $\pi \sim unif(0, 1)$

$$\begin{aligned} p(\pi|Y) &\propto {}^{n}C_{y}\pi^{y}(1-\pi)^{n-y} \times 1 \\ &\propto \pi^{y}(1-\pi)^{n-y} \end{aligned}$$

- The pdf of beta distribution which is known to be proper is:

  $$Beta(x|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

  where $0 < x < 1$ and $\alpha > 0$, $\beta > 0$
  Note that $\Gamma(k)$ is the Gamma function.

- Let $x = \pi$, $\alpha = Y + 1$ and $\beta = n - Y + 1$

- Thus $p(\pi|Y = y) \sim Beta(y + 1, n - y + 1) = \frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)}\pi^{(y+1)-1}(1-\pi)^{(n-y+1)-1}$

Bayesian Analysis: Binomial Distribution

Let $Y \sim Bin(n, \pi)$ and $\pi \sim unif(0, 1)$

$$\begin{aligned} p(\pi|Y) &\propto {}^{n}C_{y}\pi^{y}(1-\pi)^{n-y} \times 1 \\ &\propto \pi^{y}(1-\pi)^{n-y} \end{aligned}$$

Note that $\Gamma(k)$ is the Gamma function.

- Let $x = \pi$, $\alpha = Y + 1$ and $\beta = n - Y + 1$

- Thus $p(\pi|Y = y) \sim Beta(y + 1, n - y + 1) = \frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)}\pi^{(y+1)-1}(1-\pi)^{(n-y+1)-1}$

- Note that $\frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)}$ is the normalizing constant to transform $\pi^{y}(1-\pi)^{n-y}$

CHENNAI MATHEMATICAL INSTITUTE

Application-The Marketing Effectiveness Model

Description

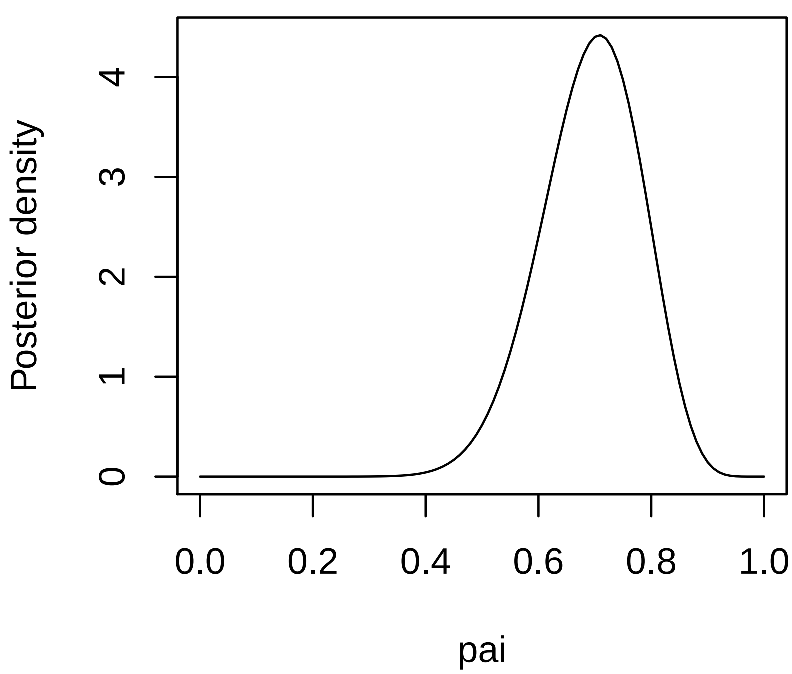
A data scientist examined the level of effectiveness of marketing denoted π among $n = 24$ store across India. In this case, 17 store met the target.

- Let $X_i = 1$ if store i met the target and $X_i = 0$ otherwise
- Let $\sum_{i=1}^{24} X_i = Y \sim Bin(24, \pi)$ and $\pi \sim Unif(0, 1) = Beta(1, 1)$
- $p(\pi|Y, n) \sim Beta(y + 1, n - y + 1)$
- Substitute $n = 24$ and $Y = 17$ we have the posterior distribution as

$$p(\pi|Y, n) = Beta(18, 8)$$

Application-The Campaign Effectiveness Model

Posterior density plot of π:



Posterior Summaries

The full posterior contains too much information, especially in multi-parameter models. So, we use summary statistics (e.g. mean, var, HDR).

- Methods for generating summary stats:
  - Analytical Solutions: use the well-known analytic solutions for the mean, variance, etc. of the various posterior distribution.
  - Numerical Solutions: use a random number generator to draw a large number of values from the posterior distribution, then compute summary stats from those random draws.

CHENNAI MATHEMATICAL INSTITUTE

Analytic Summaries of the Posterior

- Analytic summaries are based on standard results from probability theory
- Continuing our example, $p(\pi|Y,n) = Beta(18,8)$
- If $\theta \sim Beta(\alpha,\beta)$

$$E(\theta) = \frac{\alpha}{\alpha+\beta} \qquad E(\pi|Y,n) = \frac{18}{18+8} = 0.69$$

$$Var(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \qquad Var(\pi|Y,n) = \frac{18\ 8}{(18+8)^2(18+8+1)} = 0.01$$

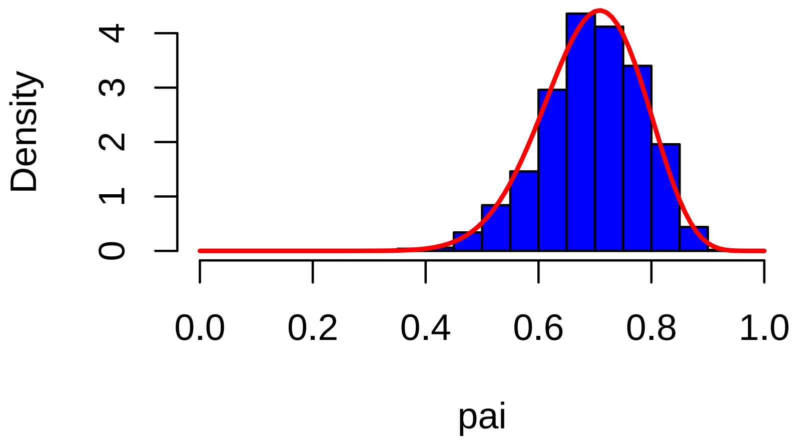
$$Mode(\theta) = \frac{\alpha-1}{\alpha+\beta-2} \qquad Mode(\pi|Y,n) = \frac{18-1}{18+8-2} = 0.71$$

CHENNAI MATHEMATICAL INSTITUTE

Numerical Summaries of the Posterior

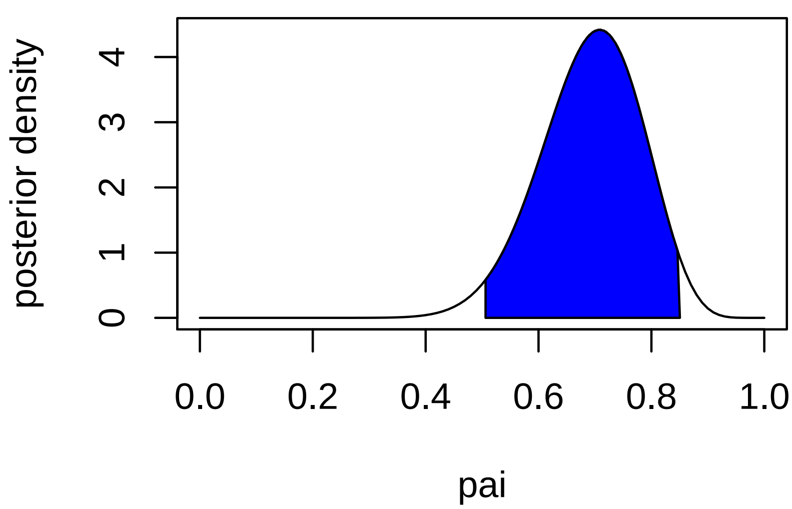
Simulate 1000 samples from `Beta(18,8)`

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.3890  0.6404  0.6995  0.6964  0.7608  0.9094
```



Credible Intervals or Highest Posterior Density Region

Highest Density Regions (HDR's) are intervals containing a specified posterior probability. The figure below plots the 95% highest posterior density region.



Confidence Intervals vs. Bayesian Credible Intervals

Bayesian credible interval

The Bayesian credible interval is the probability that a true value of θ lies in the interval. Technically,

$$P(\theta \in Interval \mid Data) = \alpha$$

Note that here probability statement is direct.

Frequentist Confidence interval

The Frequentist Confidence interval is the region of sampling distribution for θ such that given the observed data one would expect (1-α) percent of the future estimates of θ to be outside that interval. Note that here understanding of probability is implicit. It is not a direct probability statement.

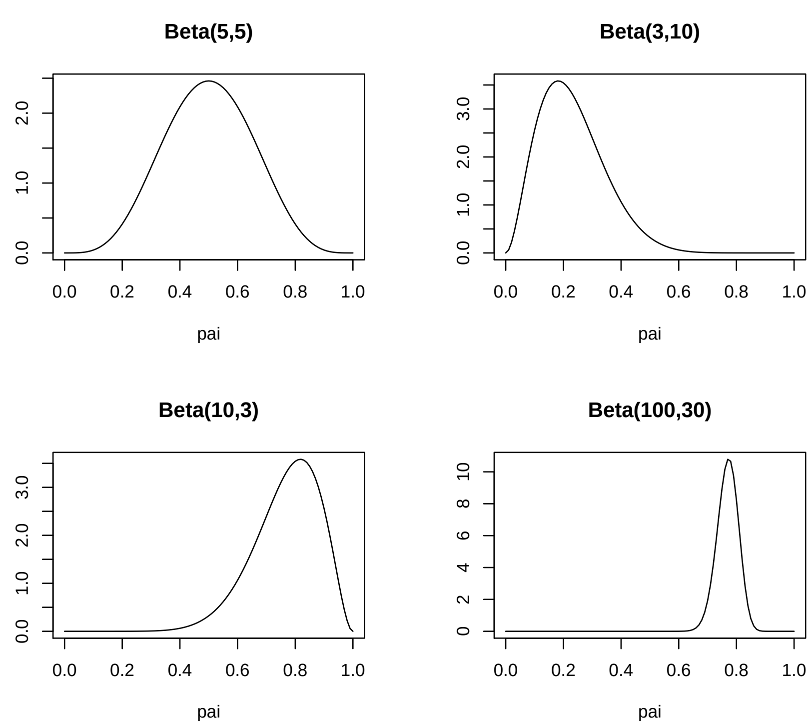
Confidence Intervals vs. Bayesian Credible Intervals

- But often the results appear similar.
- If Bayesians use "non-informative priors" and there is a large number of observations, often several dozen will do, HDRs and frequentist confidence intervals will coincide numerically.
- The interpretation of the two quantities are entirely different.

Returning to the Binomial Distribution

- If $Y \sim Bin(n, \pi)$, the uniform prior is just one of an infinite number of possible prior distributions.
- What other distributions could we use?
- A reasonable alternative to the unif(0,1) distribution is the beta distribution.
- Can you show that Beta(1,1) is a uniform(0,1) distribution?

Prior Consequences
Plots of 4 Different Beta Distributions



Modeling Expert's Opinion with Conjugate Beta Prior

- Suppose a subject matter expert believes the chance that the value of π is less than 0.5 is less than 0.05, i.e.,

$$P(\pi < 0.5) \leq 0.05$$

In addition the expert believes the chance that the value of π is more than 0.9 is less than 0.05, i.e.,

$$P(\pi > 0.9) \leq 0.05$$

So effectively the expert believes

$$P(0.5 < \pi < 0.9) \geq 0.9$$

We can model it with Beta distribution as `Beta(9.2,4.3)`

Modeling Expert's Opinion with Conjugate Beta Prior

- The expert believes

$$P(0.5 < \pi < 0.9) = \int_{0.5}^{0.9} f(\pi) d\pi = 0.9$$

- Choose $f(\pi)$ a conjugate prior - so that it satisfies the above equation, i.e.,

$$\int_0^1 \frac{1}{Beta(\alpha, \beta)} \pi^{\alpha-1}(1-\pi)^{\beta-1} d\pi = 0.9$$

- Now problem is for what choice of α and β the above equation satisfies.
- Solving the above equation, the target prior is `Beta(9.2,4.3)`

Binomial Distribution with Conjugate Beta Prior

- If $Y \sim Bin(n, \pi)$ and $\pi \sim Beta(\alpha, \beta)$

- The posterior distribution:

$$p(\pi|Y,n) = \frac{{}^{n}C_{y}\pi^{y}(1-\pi)^{n-y} \times \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\pi^{\alpha-1}(1-\pi)^{\beta-1}}{p(Y)},$$

where

$$p(Y) = \int_0^1 .{}^{n}C_{y}\pi^{y}(1-\pi)^{n-y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\pi^{\alpha-1}(1-\pi)^{\beta-1}d\pi$$

This is a very complicated integral in the denominator. Though this particular integral can be solved; but we will pretend that it is difficult integral and we shall use a standard trick in the Bayesian toolbox to solve this problem.

CHENNAI MATHEMATICAL INSTITUTE

The Posterior of Binomial Model with Beta Prior

- The posterior distribution is :

$$f(\pi|y) = \frac{f(y|\pi)f(\pi)}{f(y)}$$

$$f(\pi|y) = \frac{\frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)}\pi^{y}(1-\pi)^{n-y}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\pi^{\alpha-1}(1-\pi)^{\beta-1}}{f(y)}$$

- Simplifying the above expression:

$$f(\pi|y) = \frac{\Gamma(\alpha+n+\beta)}{\Gamma(y+\alpha)\Gamma(n+\beta-y)}\pi^{y+\alpha-1}(1-\pi)^{n+\beta-y-1}$$

- This is $Beta(y+\alpha, n-y+\beta)$ distribution.

CHENNAI MATHEMATICAL INSTITUTE

The Posterior of Binomial Model with Beta Prior

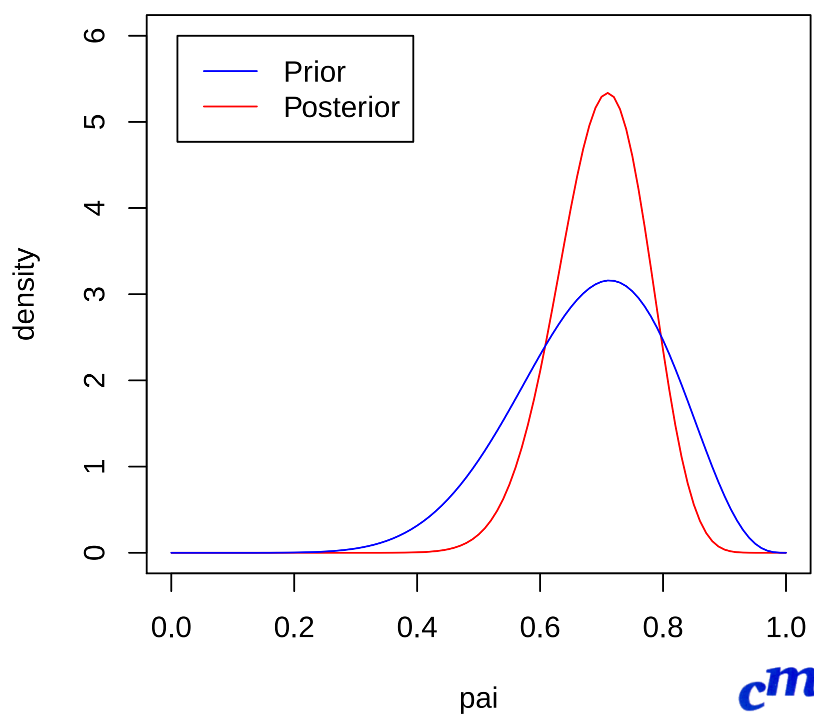
Note

You can see posterior distribution has the same distribution as prior distribution updated by new data. In general, when this happens we say the prior is conjugate prior.

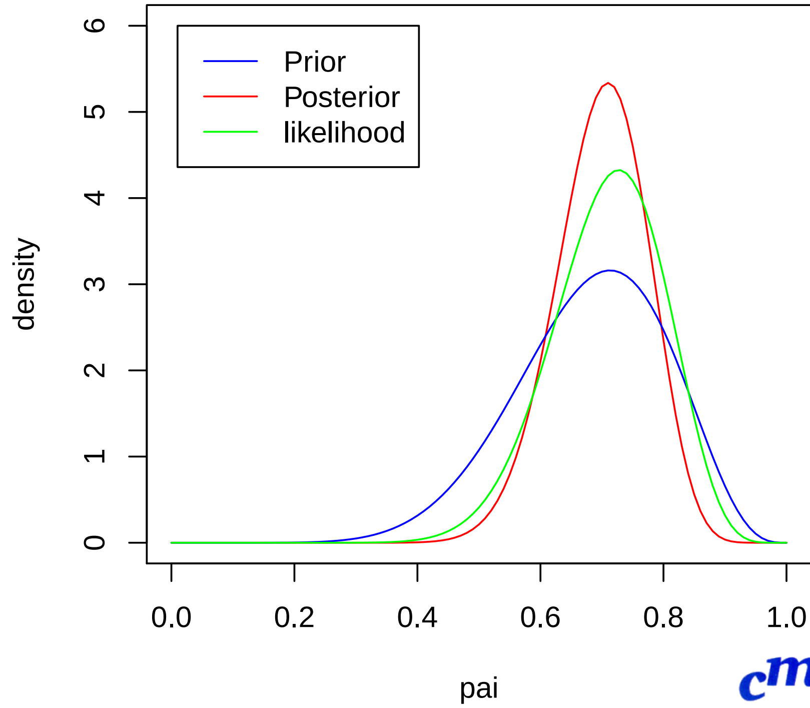
Application

Lets continue to the previous example. You remeber 17 of 24 store met the target (so $y = 17$ and $n = 24$ where y is a realization from binomial) and you use $Beta(9.2, 4.3)$ prior; the posterior distribution is $Beta(17 + 9.2, 24 - 17 + 4.3) = Beta(26.2, 11.3)$

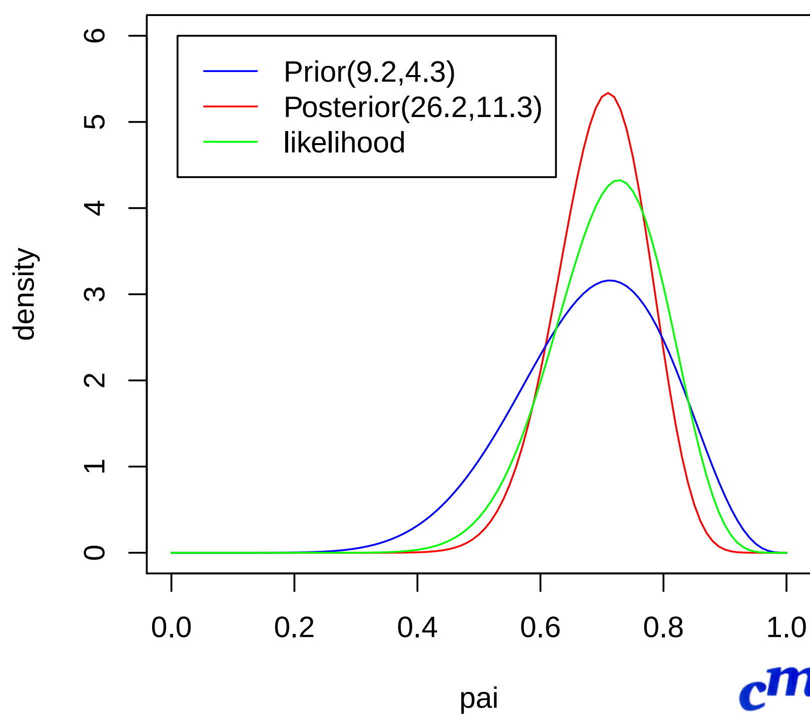
Application



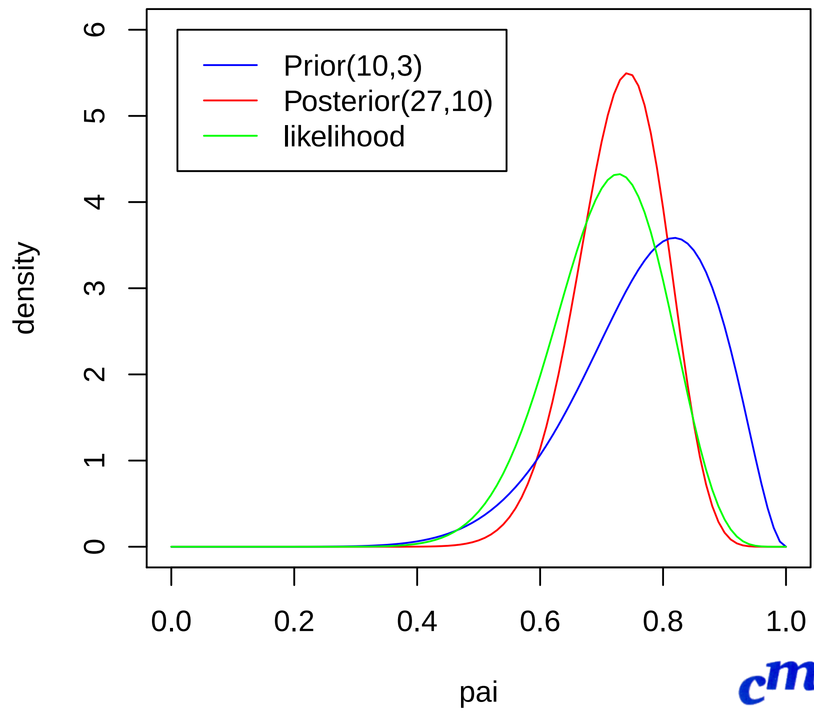
Application



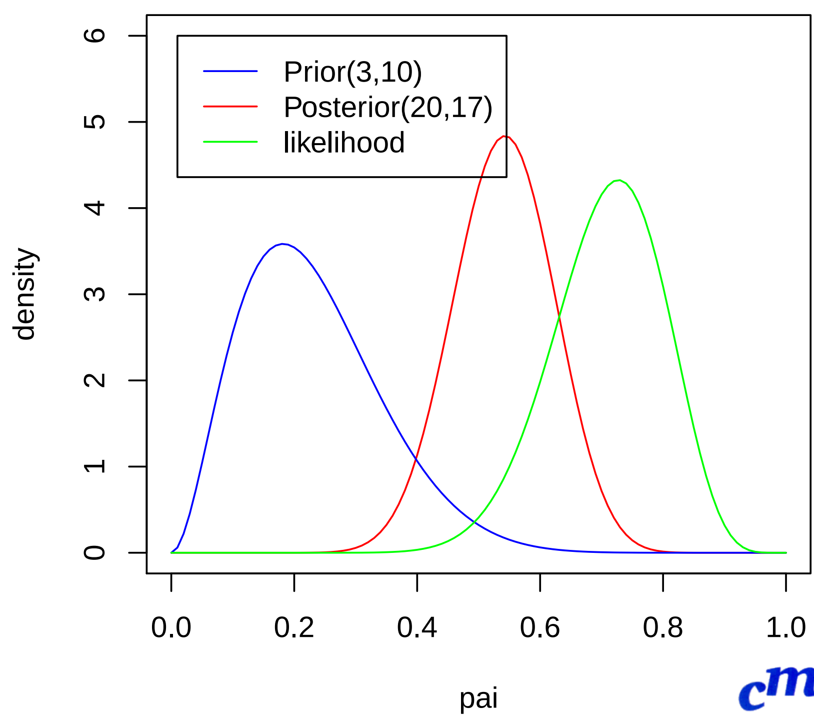
Consequence of Different Priors



Consequence of Different Priors

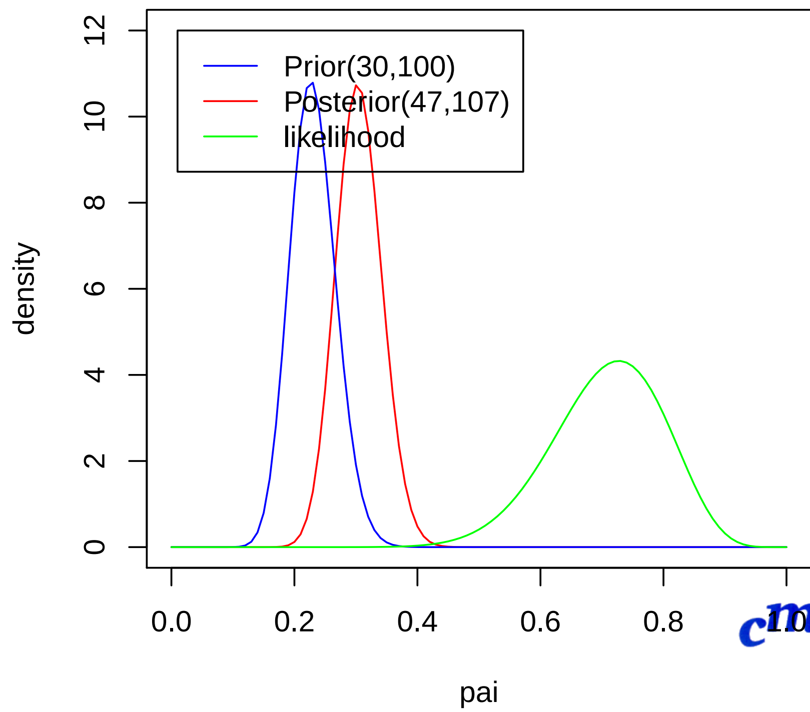


Consequence of Different Priors



Consequence of Different Priors

Bad prior or Bad Data??



Conclusion

- Conjugate prior can be used to model expert's opinion.
- For Conjugate prior - you don't have to solve the complicated integration.
- Solution for Conjugate prior is known
- Time for Hands-on.

Thank You

sourish@cmi.ac.in

www.cmi.ac.in/~sourish

CHENNAI MATHEMATICAL INSTITUTE