The importance of priors

Polychronis Kostoulas

HARMONY
Novel tools for test evaluation and disease prevalence estimation

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

UNIVERSITY OF THESSALY
creative years

Bayesian inference

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

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

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

The posterior is proportional to the likelihood and the prior

Bayesian inference

We estimate the posterior $L(\vartheta|data)$

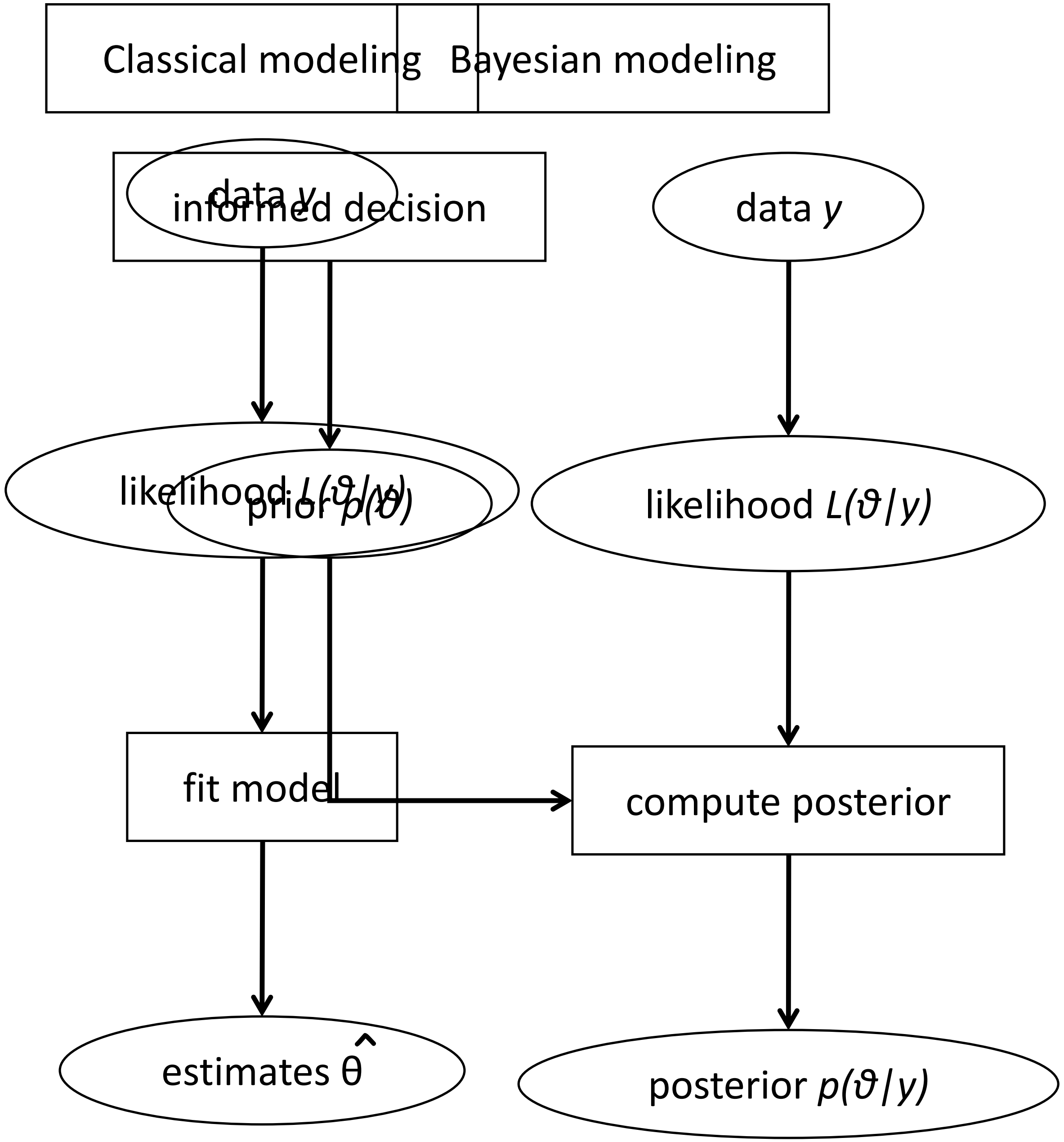
We need to

i. define the likelihood of the data $L(data|\vartheta)$

   assumption about the underlying distribution that the observed data follow

ii. specify prior information on the parameters ϑ

   choice of a distribution for the prior on ϑ


Classical modeling
Bayesian modeling
data y
informed decision
data y
likelihood L(ϑ|y)
prior p(ϑ)
likelihood L(ϑ|y)
fit model
compute posterior
estimates θ̂
posterior p(ϑ|y)

Classical estimation for binomial data

- Estimate an unknown population proportion (θ) from the results of a sequence of Bernoulli trials; data $y_1, \ldots, y_n$, each of which is either 0 or 1.

$$p(y|\theta) = Bin(y \mid n, \theta) = \binom{n}{y} \theta^{y} (1-\theta)^{n-y}$$

$$\theta = \frac{y}{n}$$

$$\theta \pm z_{\left(1-\frac{\alpha}{2}\right)} * \sqrt{\frac{\theta(1-\theta)}{n}}$$

TB prevalence – Classical approach

Suppose that we are interested in estimating the prevalence of tuberculosis (TB) among hospitalized men in Greece. Data from a retrospective study indicated that 5 men out of a random sample of 1000 were TB+.

$$p(y|\theta)=Bin(y\,|\,n,\theta)=\binom{n}{y}\theta^{y}(1-\theta)^{n-y}$$

$$\theta=\frac{y}{n}=\frac{5}{1000}=0.005$$

$$\theta \pm z_{\left(1-\frac{\alpha}{2}\right)} * \sqrt{\frac{\theta(1-\theta)}{n}}$$

$$\theta=0.005\pm 1.96*\sqrt{0.005*0.995/1000}$$

$$\theta=0.005\pm 0.004$$

$$\theta=0.005(0.001;0.009)$$

Bayesian inference for binomial data

We estimate the posterior $L(\vartheta|data)$

We need to

i. define the likelihood of the data $L(data|\vartheta)$

   assumption about the underlying distribution that the observed data follow

ii. specify prior information on the parameters ϑ

   choice of a distribution for the prior on ϑ

Bayesian inference for binomial data

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

$$p(y|\theta) = Bin(y\mid n,\theta) = \binom{n}{y}\theta^{y}(1-\theta)^{n-y}$$

$$p(\theta) = Beta(a,b) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$p(y|\theta) \propto \theta^{y}(1-\theta)^{n-y}$$

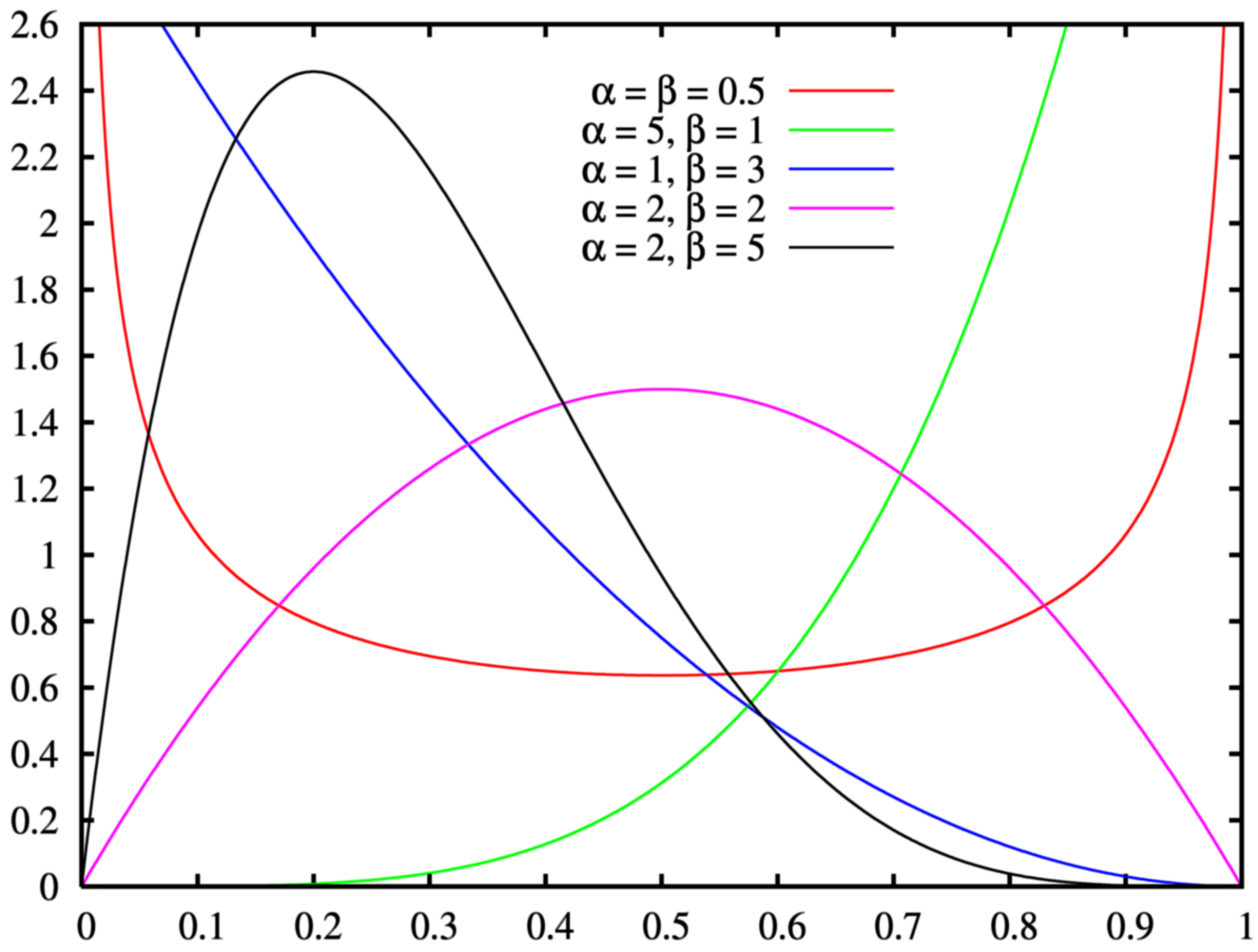
$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

$$p(\theta|y) \propto \theta^{y}(1-\theta)^{n-y}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

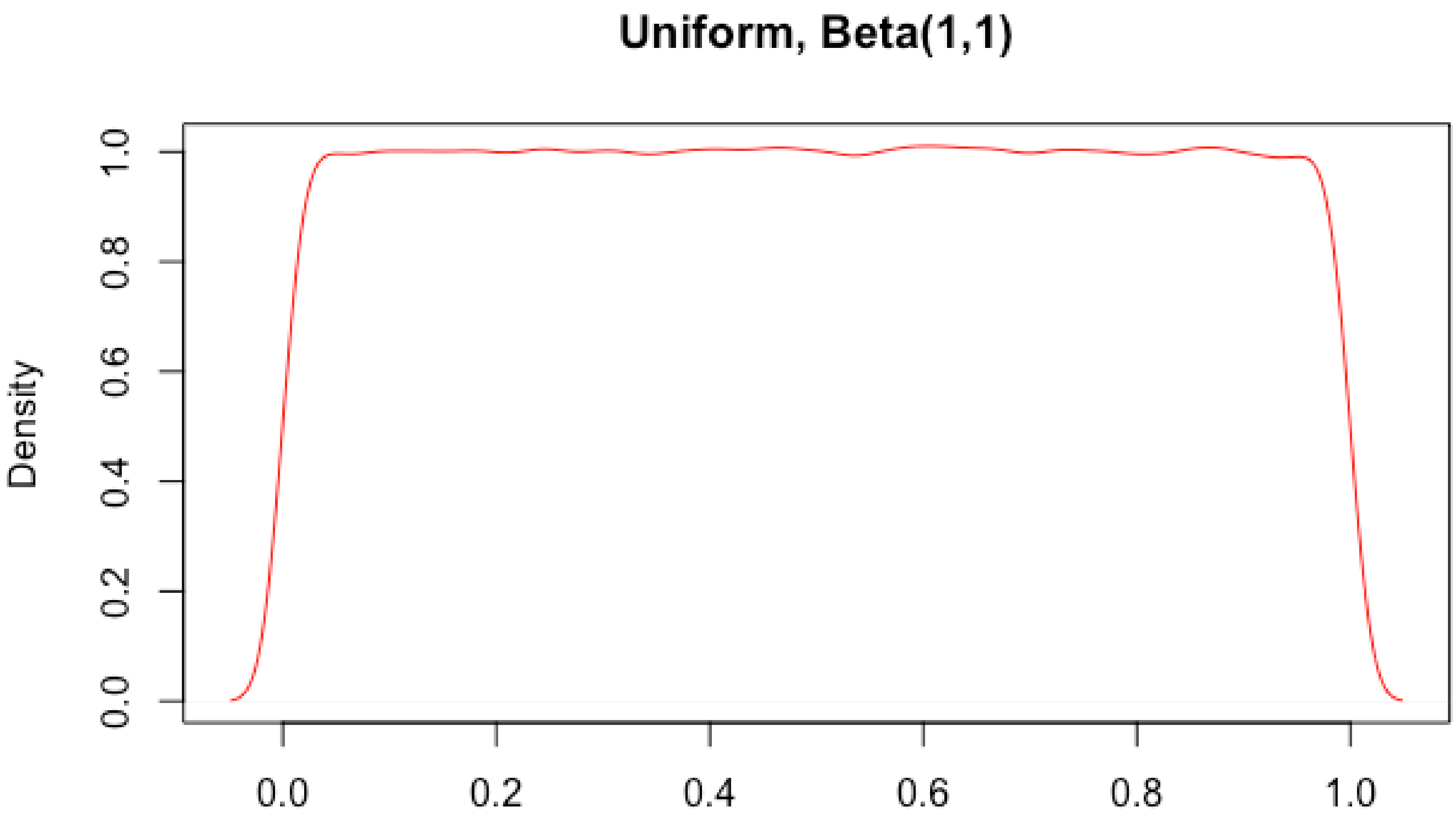
$$p(\theta|y) \propto \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}$$

$$\boxed{p(\theta|y) = Beta(\theta\mid y+\alpha, n-y+\beta)}$$

Beta distribution



Uniform beta distribution



TB prevalence – Bayesian Inference

Suppose that we are interested in estimating the prevalence of tuberculosis (TB) among hospitalized men in Greece. Data from a retrospective study indicated that 5 men out of a random sample of 1000 were TB+.

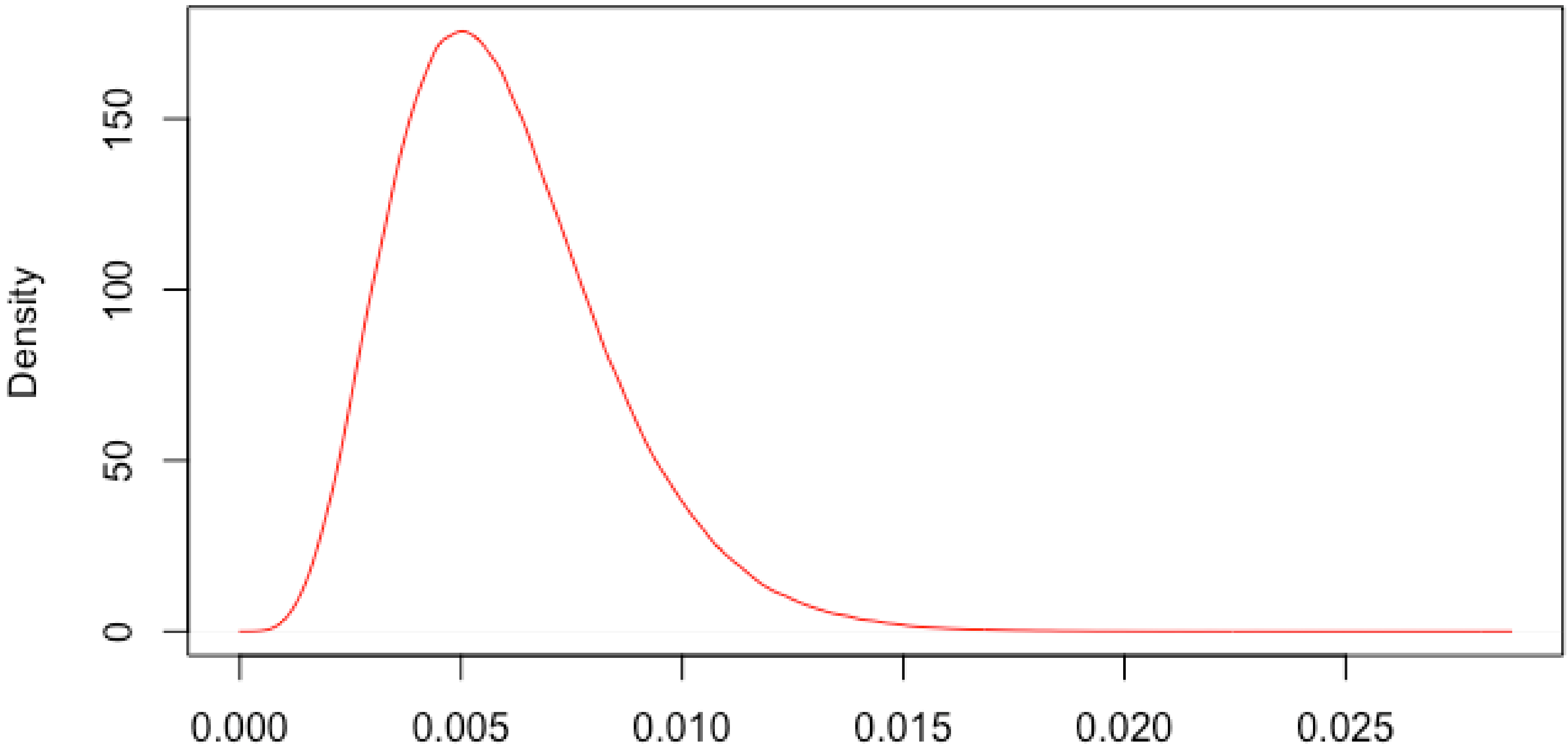
Using a non-informative Beta(1,1) prior we estimate the prevalence of TB as:

$$p(\theta|y) = Beta(\theta \,|\, \alpha + y, \beta + n - y)$$

$$p(\theta|y) = Beta(1+5, 1+1000-5)$$

$$p(\theta|y) = Beta(6, 996)$$

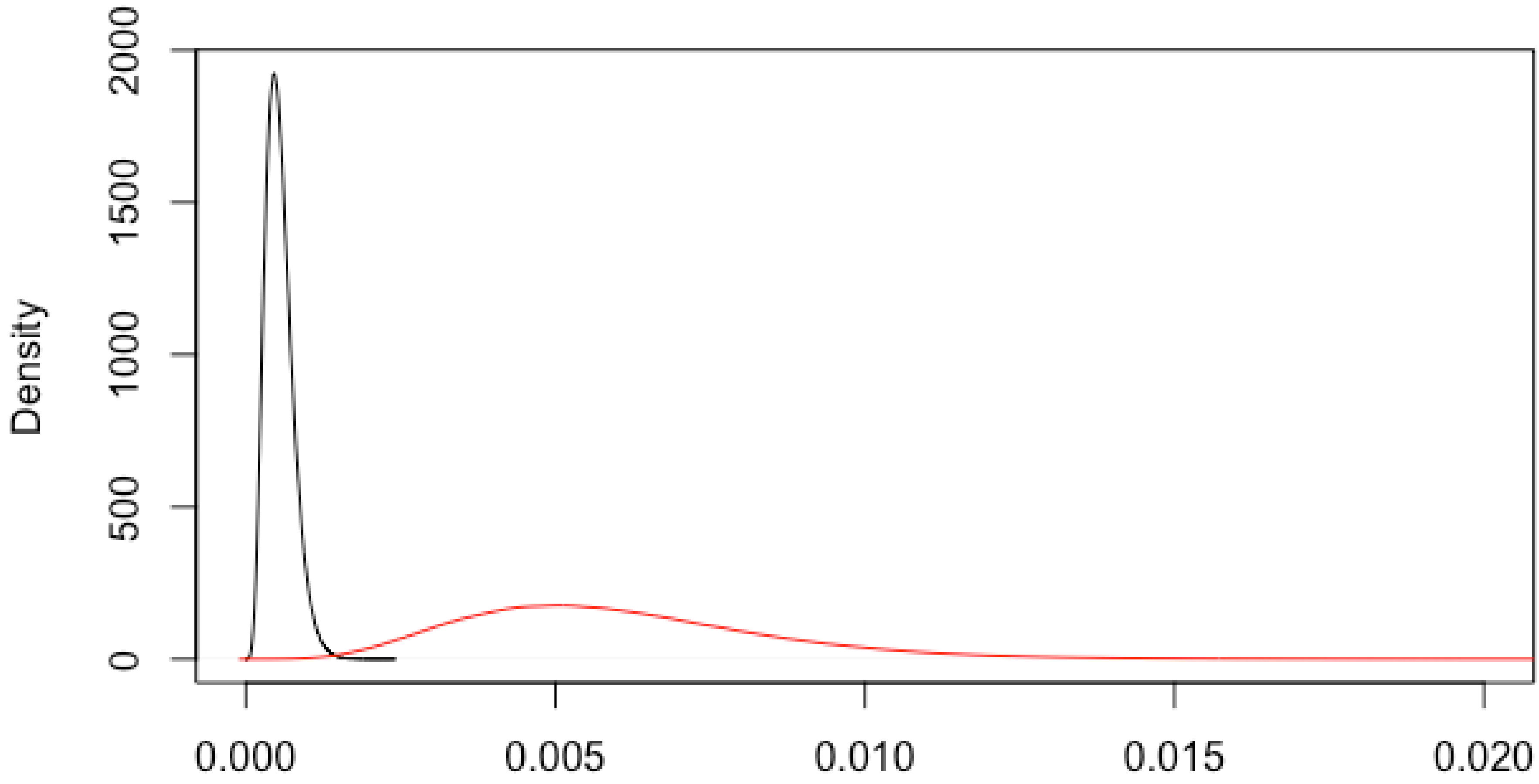
Beta(6,996)



TB prevalence – Bayesian Inference

Let’s use a very narrow informative Beta(1,10000) prior:

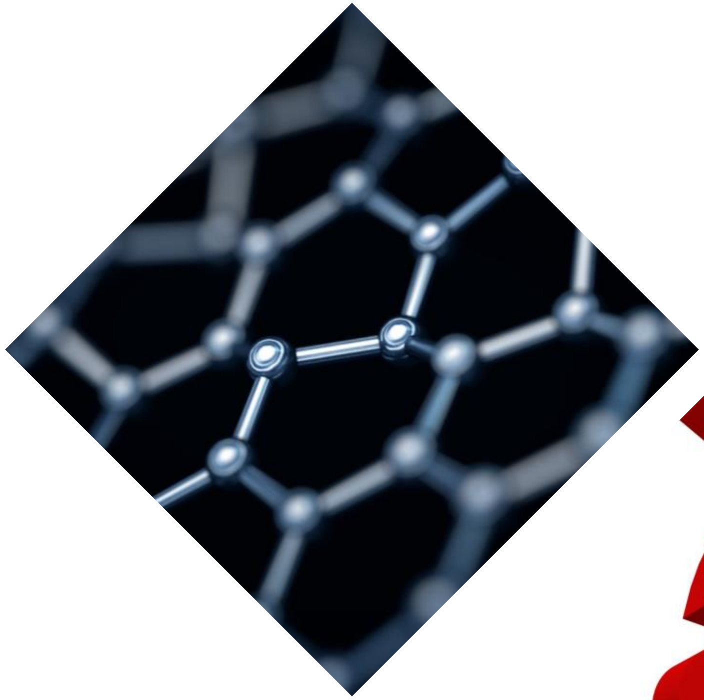
$$P(\theta|y) = Beta(\theta|\alpha + y, \beta + n - y$$
$$P(\theta|y) = Beta(1 + 5, 10000 + 1000 - 5)$$
$$P(\theta|y) = Beta(6, 10995)$$


Density
0
500
1000
1500
2000
0.000
0.005
0.010
0.015
0.020

Prior + Data

IDENTIFIABLE MODELS

NON-IDENTIFIABLE MODELS



HARMONY

Novel tools for test evaluation and disease prevalence estimation

Thank you!