

## Set 10 - Bayesian statistics

STAT 401 (Engineering) - Iowa State University

February 13, 2017

- Bayesian parameter estimation
  - Condition on what is known
  - Describe uncertainty using probability
  - Binomial example

# A Bayesian statistician

Let

- $y$  be the data we will collect from an experiment,
- $K$  be everything we know for certain about the world (aside from  $y$ ), and
- $\theta$  be anything we don't know for certain.

My definition of a Bayesian statistician is an individual who makes decisions based on the probability distribution of those things we don't know conditional on what we know, i.e.

$$p(\theta|y, K).$$

Typically, the  $K$  is dropped from the notation.

# Bayes' Rule

Bayes' Rule applied to a partition  $P = \{A_1, A_2, \dots\}$ ,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{\infty} P(B|A_i)P(A_i)}$$

Bayes' Rule also applies to probability density (or mass) functions, e.g.

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}$$

where the integral plays the role of the sum in the previous statement.

# Parameter estimation

Let  $y$  be data from some model with unknown parameter  $\theta$ . Then

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}$$

and we use the following terminology

Terminology	Notation
Posterior	$p(\theta y)$
Prior	$p(\theta)$
Model	$p(y \theta)$
Prior predictive distribution (marginal likelihood)	$p(y)$

If  $\theta$  is discrete (continuous),

then  $p(\theta)$  and  $p(\theta|y)$  are probability mass (density) functions.

If  $y$  is discrete (continuous),

then  $p(y|\theta)$  and  $p(y)$  are probability mass (density) functions.

# Binomial model

Suppose  $Y \sim \text{Bin}(n, \theta)$ , then

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

A reasonable default prior is the uniform distribution on the interval  $(0, 1)$

$$p(\theta) = \text{I}(0 < \theta < 1).$$

Using Bayes Rule, you can find

$$\theta|y \sim \text{Be}(1 + y, 1 + n - y).$$

# Beta distribution

## Definition

The **beta distribution** defines a distribution for a probability, i.e. a number on the interval  $(0,1)$ . The probability density function is

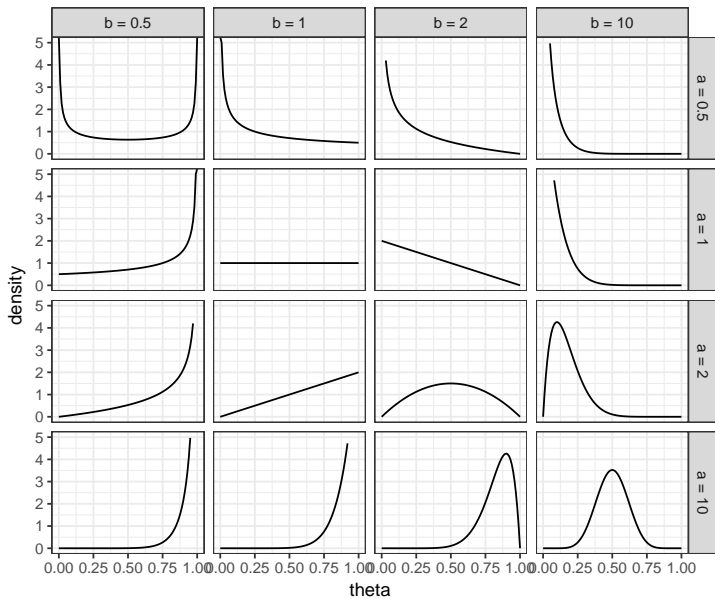
$$p(\theta) = \frac{\theta^{a-1}(1-\theta)^{b-1}}{Beta(a,b)} \mathbf{I}(0 < \theta < 1)$$

where  $a, b > 0$  and  $Beta$  is the beta function, i.e.

$$Beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \quad \text{and} \quad \Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx.$$

The beta distribution has the following properties:

- $E[\theta] = \frac{a}{a+b}$
- $Var[\theta] = \frac{ab}{(a+b)^2(a+b+1)}.$





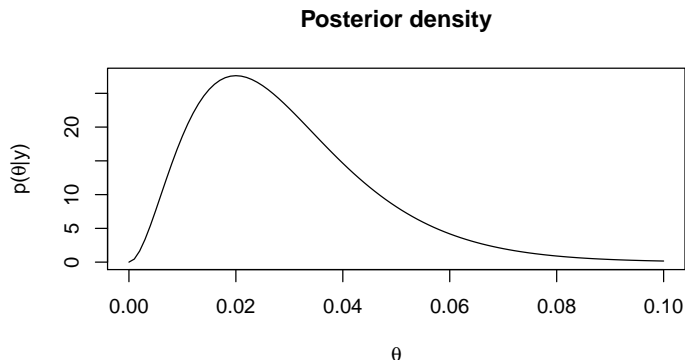
# Beta posterior

Suppose we have made 100 sensors according to a particular protocol and 2 have a sensitivity below a pre-determined threshold. Let  $Y$  be the number below the threshold. Assume  $Y \sim \text{Bin}(n, \theta)$  with  $n = 100$  and  $\theta \sim \text{Unif}(0, 1) \stackrel{d}{=} \text{Be}(1, 1)$ , then

$$\theta|y \sim \text{Be}(1 + y, 1 + n - y) \stackrel{d}{=} \text{Be}(3, 99).$$

# Posterior density

```
n <- 100
y <- 2
curve(dbeta(x, 1+y, 1+n-y), 0, 0.1,
      main = "Posterior density",
      xlab = expression(theta),
      ylab = expression(paste("p(", theta, "| y)")))
```



# Posterior expectation

Often times it is inconvenient to provide a full posterior and so we often summarize using a point estimate from the posterior. For a point estimate, we can use the posterior expectation:

$$\hat{\theta}_{Bayes} = E[\theta|y] = \frac{1+y}{(1+y) + (1+n-y)} = \frac{1+y}{1+n}$$

```
(1+y)/(2+n)
```

```
[1] 0.02941176
```

Not that this is close, but not exactly equal to  $\hat{\theta}_{MLE} = y/n$ . Since the MLE is unbiased, this posterior expectation will generally be biased but it is still consistent since  $\hat{\theta}_{Bayes} \rightarrow \hat{\theta}_{MLE}$ .

# Credible intervals

## Definition

A  $100(1 - \alpha)\%$  **credible interval** is any interval  $(L, U)$  such that

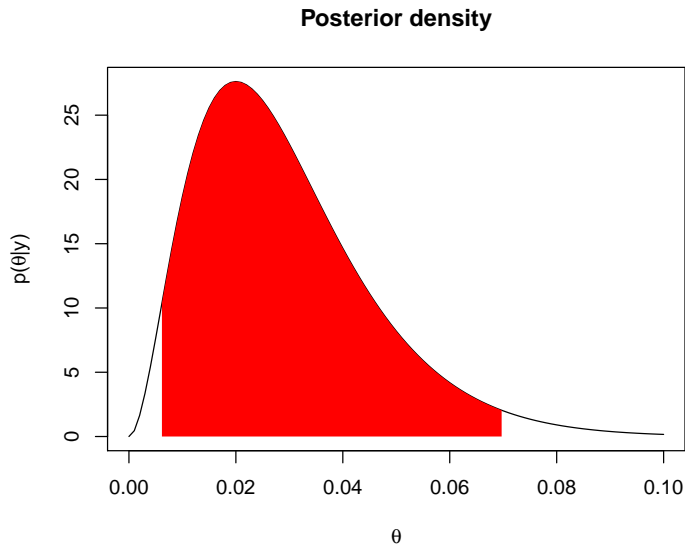
$$1 - \alpha = \int_L^U p(\theta|y)d\theta.$$

An **equal-tail**  $100(1 - \alpha)\%$  **credible interval** is the interval  $L, U)$  such that

$$\alpha/2 = \int_{-\infty}^L p(\theta|y)d\theta = \int_U^{\infty} p(\theta|y)d\theta.$$

[1] 0.006 0.070

# Equal-tail 95% credible interval



# Summary

Bayesian parameter estimation involves

1. Specifying a model  $p(y|\theta)$  for your data.
2. Specifying a prior  $p(\theta)$  for the parameter.
3. Deriving the posterior

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \propto p(y|\theta)p(\theta).$$

4. Calculating quantities of interest, e.g.
  - Posterior expectation,  $E[\theta|y]$
  - Credible interval