I03 - Bayesian parameter estimation

STAT 5870 (Engineering)
Iowa State University

October 4, 2024

Outline

- Bayesian parameter estimation
    - Condition on what is known
    - Describe **belief** using probability
    - Terminology
        - Prior $\rightarrow$ posterior
        - Posterior expectation
        - Credible intervals
    - Binomial example
        - Beta distribution

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
- θ 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 $P = \{A_1, A_2, \ldots\}$,

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

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 (vector) θ. 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 (likelihood)	$p(y\|\theta)$
Prior predictive (marginal likelihood)	$p(y)$

Bayesian parameter estimation involves updating your prior belief about θ, $p(\theta)$, into a posterior belief about θ, $p(\theta|y)$, based on the data observed.

Bayesian notation

We now have two distributions for our parameter θ: prior and posterior. To distinguish these two, we will have no conditioning in the prior and we will condition on y in the posterior. For example,

	Prior	Posterior
Density	$p(\theta)$	$p(\theta\|y)$
Expectation	$E[\theta]$	$E[\theta\|y]$
Variance	$Var[\theta]$	$Var[\theta\|y]$
Probabilities	$P(\theta < c)$	$P(\theta < c\|y)$

Binomial model

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

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

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

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

Using Bayes Rule, you can find

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

Beta distribution

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

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

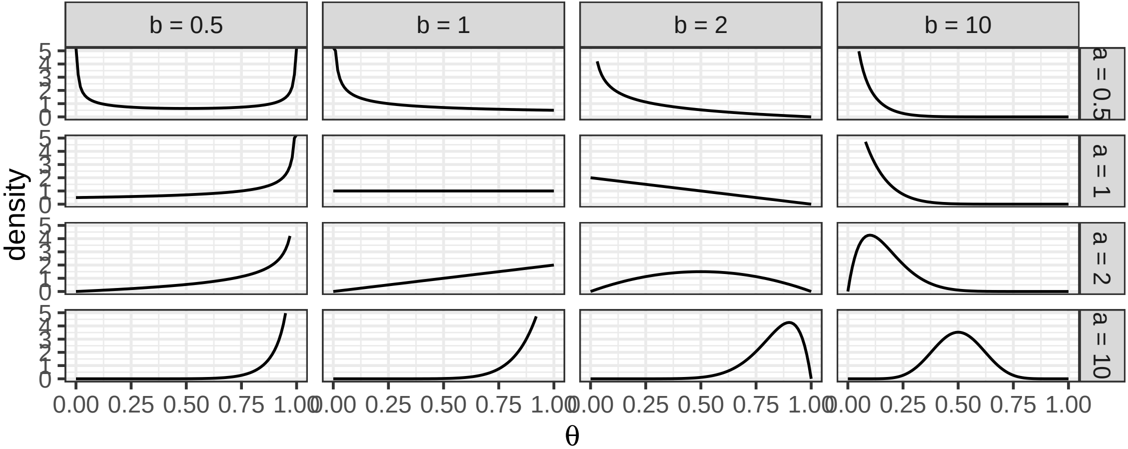
where $a, b > 0$ and $Beta$ is the beta function, i.e.

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

The beta distribution has the following properties:

- $E[\theta] = \frac{a}{a+b}$,
- $Var[\theta] = \frac{ab}{(a+b)^2(a+b+1)}$, and
- $Be(1,1) \stackrel{d}{=} Unif(0,1)$.

Beta densities

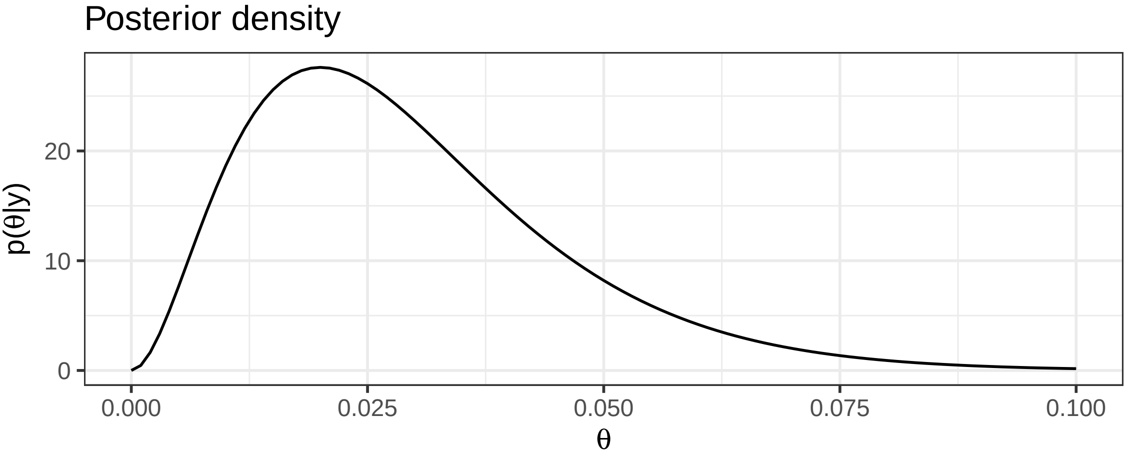


Beta posterior

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 $Y \sim Bin(n, \theta)$ with $n = 100$ and $\theta \sim Be(1, 1)$, then

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

Posterior density



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:

$$\hat{\theta}_{Bayes} = E[\theta|y] = \frac{1+y}{(1+y)+(1+n-y)} = \frac{1+y}{2+n}$$

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

[1] 0.02941176
```

Note 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 $\hat{\theta}_{Bayes} \to \hat{\theta}_{MLE}$.

Credible intervals

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

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

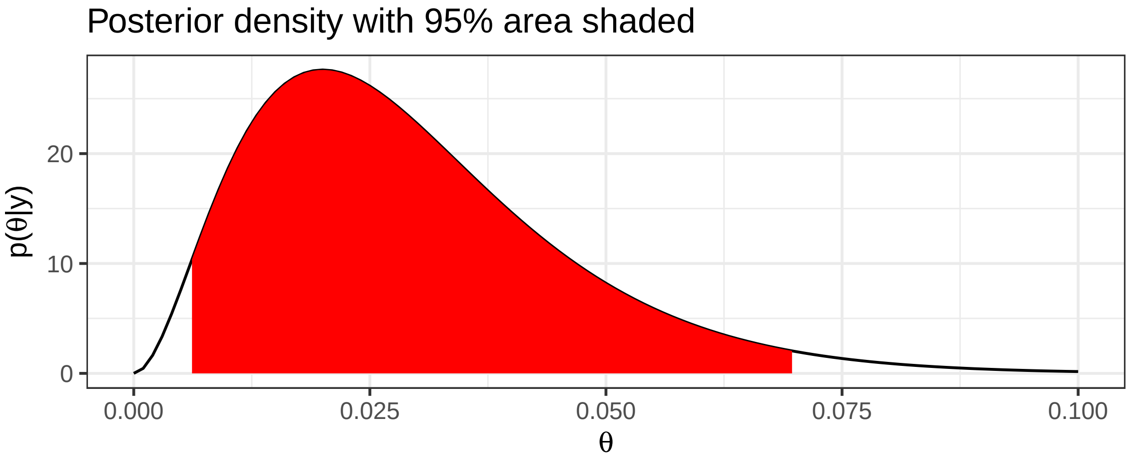
An equal-tail $100(1-a)\%$ credible interval is the interval $L, U)$ such that

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

```
# 95% credible interval is
ci = qbeta(c(.025,.975), 1+y, 1+n-y)
round(ci, 3)

[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

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

   This equation updates your prior belief, $p(\theta)$, about the unknown parameter θ into your posterior belief, $p(\theta|y)$, about θ.

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

Bayesian analysis for binomial model summary

Let $Y \sim Bin(n, \theta)$ and assume $\theta \sim Be(a, b)$. Then

$$\theta|y \sim Be(a + y, b + n - y).$$

A default prior is $\theta \sim Be(1, 1) \stackrel{d}{=} Unif(0, 1)$.

R code for binomial analysis:

```
a <- 1; b <- 1                  # default uniform prior
y <- 3; n <- 10                 # data

curve(dbeta(x,ay,b+n-y))        # posterior (pdf)
(a+y)/(a+b+n)                   # posterior mean
qbeta(.5, a+y, b+n-y)           # posterior median
qbeta(c(.025,.975), a+y, b+n-y) # 95% equal tail credible interval

# Probabilities
pbeta(0.5, a+y, b+n-y)          # P(theta<0.5|y)

# Special cases
qbeta(c(0,.95), a+y, b+n-y)     # if y=0, use a lower one-sided CI
qbeta(c(.05,1), a+y, b+n-y)     # if y=n, use a upper one-sided CI
```