Topic 12

Overview of Estimation

Bayesian Statistics

Outline

Bayes Formula

Bayesian Estimation

Coin Tosses

Introduction

The Bayesian approach to statistics takes into account not only the density

$$\mathbf{f}_{X|\Theta}(\mathbf{x}|\psi)$$

for the data collected for any given experiment but also external information to determine a prior density π on the parameter space Ψ. Thus, in this approach, both the parameter and the data are modeled as random.

Estimation is based on Bayes formula.

Bayes Formula

Let $\tilde{\Theta}$ be a discrete random variable with state space a finite subset of the parameter space Ψ. Based on the values θ for $\tilde{\Theta}$, we form a partition $C_\theta = \{\tilde{\Theta} = \theta\}$ of the probability space. For any event A, Bayes formula is

$$P(C_\theta|A) = \frac{P(A|C_\theta)P(C_\theta)}{\sum_\psi P(A|C_\psi)P(C_\psi)}$$

Now choose $A = \{X = \mathbf{x}\}$ to be the event given by data $\mathbf{x}$.

$$P\{\tilde{\Theta} = \theta|X = \mathbf{x}\} = \frac{P\{X = \mathbf{x}|\tilde{\Theta} = \theta\}P\{\tilde{\Theta} = \theta\}}{\sum_\psi P\{X = \mathbf{x}|\tilde{\Theta} = \psi\}P\{\tilde{\Theta} = \psi\}}.$$

Bayesian Estimation

$$P\{\tilde{\Theta} = \theta | X = \mathbf{x}\} = \frac{P\{X = \mathbf{x} | \tilde{\Theta} = \theta\} P\{\tilde{\Theta} = \theta\}}{\sum_\psi P\{X = \mathbf{x} | \tilde{\Theta} = \psi\} P\{\tilde{\Theta} = \psi\}}.$$

We make the replacements

- $\pi\{\theta\} = P\{\tilde{\Theta} = \theta\}$, the prior density,
- $\mathbf{f}_{X|\Theta}(\mathbf{x}|\theta) = P\{X = \mathbf{x} | \tilde{\Theta} = \theta\}$, the density (or the likelihood) given by the data, and
- $\mathbf{f}_{\Theta|X}(\theta|\mathbf{x}) = P\{\tilde{\Theta} = \theta | X = \mathbf{x}\}$, the posterior density.

$$\mathbf{f}_{\Theta|X}(\theta|\mathbf{x}) = \frac{\mathbf{f}_{X|\Theta}(\mathbf{x}|\theta)\pi\{\theta\}}{\sum_\psi \mathbf{f}_{X|\Theta}(\mathbf{x}|\psi)\pi\{\psi\}}.$$

Estimation is based on the posterior density.

Bayes Formula

$$\mathbf{f}_{\Theta|X}(\theta|\mathbf{x}) = \frac{\mathbf{f}_{X|\Theta}(\mathbf{x}|\theta)\pi\{\theta\}}{\sum_{\psi}\mathbf{f}_{X|\Theta}(\mathbf{x}|\psi)\pi\{\psi\}}.$$

For a continuous distribution on the parameter space, π is now a density for a continuous random variable and the sum in Bayes formula becomes an integral.

$$f_{\Theta|X}(\theta|\mathbf{x}) = \frac{\mathbf{f}_{X|\Theta}(\mathbf{x}|\theta)\pi(\theta)}{\int_{\Psi}\mathbf{f}_{X|\Theta}(\mathbf{x}|\psi)\pi(\psi)\ d\psi} = c(\mathbf{x})\mathbf{f}_{X|\Theta}(\mathbf{x}|\theta)\pi(\theta),$$

where $c(\mathbf{x})$, the reciprocal of the integral in the denominator, is the value necessary so that the integral of the posterior density is 1. We might also write

$$f_{\Theta|X}(\theta|\mathbf{x}) \propto \mathbf{f}_{X|\Theta}(\mathbf{x}|\theta)\pi(\theta).$$

Coin Tosses

We consider independent flips of a biased coin and use a Bayesian approach to make some inference for the probability of heads. We first set a prior distribution for $\tilde{P}$. The beta family $Beta(\alpha, \beta)$ of distributions takes values in the interval $[0, 1]$ and provides a convenient prior density π. Thus,

$$\pi(p) = c_{\alpha,\beta} p^{(\alpha-1)}(1-p)^{(\beta-1)}, \quad 0 < p < 1.$$

Any density on the interval $[0, 1]$ that can be written a a power of p times a power of $1 - p$ times a constant chosen so that

$$1 = \int_0^1 \pi(p)\ dp$$

is a member of the beta family. This distribution has

$$\text{mean } \frac{\alpha}{\alpha+\beta} \quad \text{and} \quad \text{variance } \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Coin Tosses

If we perform n Bernoulli trials $\mathbf{x} = (x_1, \ldots, x_n)$, then the joint density

$$\mathbf{f}_{X|\tilde{P}}(\mathbf{x}|p) = p^{\sum_{k=1}^n x_k}(1-p)^{n-\sum_{k=1}^n x_k}.$$

Thus the posterior distribution of the parameter $\tilde{P}$ given the data $\mathbf{x}$,

$$\begin{aligned} f_{\tilde{P}|X}(p|\mathbf{x}) \propto \mathbf{f}_{X|\tilde{P}}(\mathbf{x}|p)\pi(p) &= p^{\sum_{k=1}^n x_k}(1-p)^{n-\sum_{k=1}^n x_k} \cdot c_{\alpha,\beta}\, p^{(\alpha-1)}(1-p)^{(\beta-1)}. \\ &= c_{\alpha,\beta}\, p^{\alpha+\sum_{k=1}^n x_k - 1}(1-p)^{\beta+n-\sum_{k=1}^n x_k - 1}. \end{aligned}$$

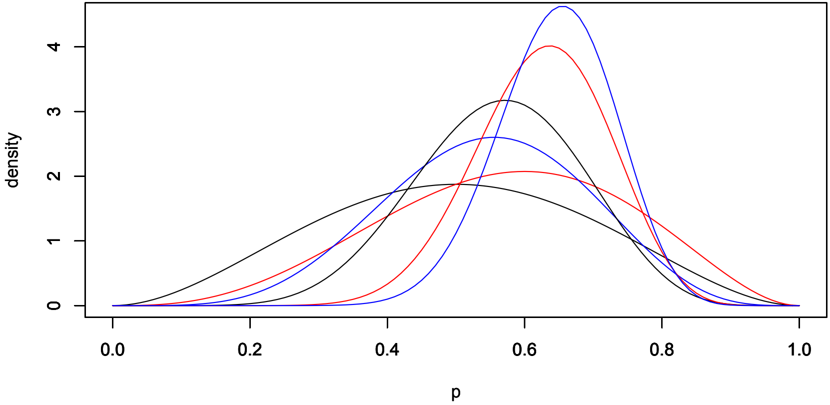
Consequently, the posterior distribution is also from the beta family with parameters

$$\alpha + \sum_{k=1}^n x_k \quad \text{and} \quad \beta + n - \sum_{k=1}^n x_k = \beta + \sum_{k=1}^n (1 - x_k).$$

$$\alpha + \#\ \text{successes} \quad \text{and} \quad \beta + \#\ \text{failures}.$$

Coin Tosses

Posterior densities based on $Beta(3,3)$ prior.



Data : *H T H T H T H H T H H T H T H H H H T T H T H H H*

Coin Tosses

Notice that the posterior mean can be written as

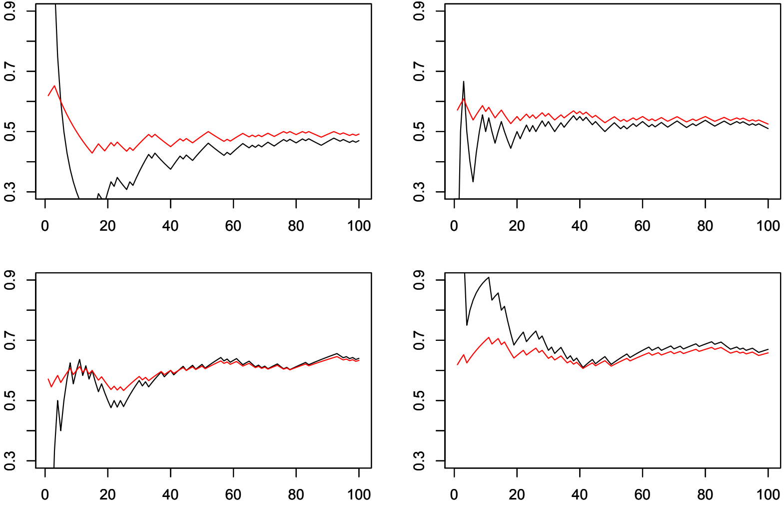
$$\begin{aligned}
\frac{\alpha + \sum_{k=1}^n x_k}{\alpha+\beta+n} &= \frac{\alpha}{\alpha+\beta+n} + \frac{\sum_{k=1}^n x_k}{\alpha+\beta+n} \\
&= \frac{\alpha}{\alpha+\beta}\cdot\frac{\alpha+\beta}{\alpha+\beta+n} + \frac{1}{n}\sum_{k=1}^n x_k \cdot \frac{n}{\alpha+\beta+n} \\
&= \frac{\alpha}{\alpha+\beta}\cdot\frac{\alpha+\beta}{\alpha+\beta+n} + \bar{x}\cdot\frac{n}{\alpha+\beta+n}.
\end{aligned}$$

This expression allow us to see that the posterior mean can be expresses as a weighted average. The relative weights are

$\alpha+\beta$ from the prior and n, the number of observations.

Coin Tosses

Posterior means based on $Beta(12, 8)$ prior.



Bayesian Estimation

This brings forward two central issues in the use of the Bayesian approach to estimation.

- If the number of observations is small, then the estimate relies heavily on the quality of the choice of the prior distribution π. Thus, an unreliable choice for π leads to an unreliable estimate.
- As the number of observations increases, the estimate relies less and less on the prior distribution. In this circumstance, the prior may simply be playing the roll of a catalyst that allows the machinery of the Bayesian methodology to proceed.

Exercise. Produce the two graphs like those on slide 11 for the classical estimate and the Bayesian estimate based on the posterior mean for the data set and prior on slide 9.