Bayesian Statistical Inference

Introductory Toy Example

Suppose that we're interested in

$$\wp = \text{probability that randomly chosen person is left-handed.}$$

The classical ("frequentist") approach involves drawing a random sample

$$X_1, \ldots, X_n$$

where

$$X_i = \begin{cases} 1 & \text{if } i\text{th person is left-handed} \\ 0 & \text{otherwise} \end{cases}$$

and estimate $\wp$ via the sample proportion

$$\widehat{\wp} = \frac{\sum_{i=1}^{n} X_i}{n}.$$

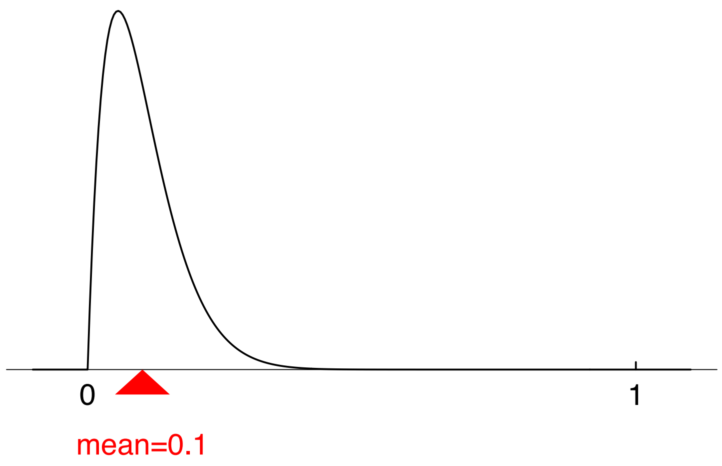
For large n we can get an approximate 95% confidence interval for $\wp$ via e.g.

$$\widehat{\wp} \pm 1.96 \sqrt{\frac{\widehat{\wp}(1-\widehat{\wp})}{n}}.$$

A Bayesian approach treats $0 \leq \wp \leq 1$ as a random variable.

The Bayesian analyst then puts a prior distribution on $\wp$ expressing his/her belief before seeing data.

Based on my belief about $\wp$ I will specify a prior that looks like:



A density function that looks like this (and has mean 0.1) is

$$p(\wp) = 342\wp(1-\wp)^{17}, \quad 0 < \wp < 1,$$

the Beta(2,18) density.

My model for the data and its dependence on $\wp$ is

$$p(X_i|\wp) = \begin{cases} \wp, & X_i = 1 \\ 1-\wp, & X_i = 0 \end{cases}$$

independently for $1 \leq i \leq n$.

We can write this neatly as

$$p(X_i|\wp) = \wp^{X_i}(1-\wp)^{1-X_i}, \quad 1 \leq i \leq n.$$

By independence

$$p(X_1, \ldots, X_n | \wp) = \prod_{i=1}^{n} \wp^{X_i} (1 - \wp)^{1 - X_i}.$$

I now update my belief via the posterior density:

$$p(\wp | X_1, \ldots, X_n).$$

$$\begin{aligned} p(\wp | X_1, \ldots, X_n) &= \frac{p(\wp, X_1, \ldots, X_n)}{p(X_1, \ldots, X_n)} \\ &= \frac{p(X_1, \ldots, X_n | \wp)\, p(\wp)}{p(X_1, \ldots, X_n)} \\ &\propto p(X_1, \ldots, X_n | \wp) p(\wp) \\ &= \prod_{i=1}^{n} \wp^{X_i} (1 - \wp)^{1 - X_i} \\ &\quad \times 342 \wp (1 - \wp)^{17} \\ &\propto \wp^{1 + \sum_{i=1}^{n} X_i} (1 - \wp)^{17 + n - \sum_{i=1}^{n} X_i} \end{aligned}$$

which is the Beta $\left(2 + \sum_{i=1}^{n} X_i, 18 + n - \sum_{i=1}^{n} X_i\right)$ density function.

Now collect data on class participants:

$$n =$$

$$\sum_{i=1}^{n} X_i =$$

Posterior is

Beta(,).

The posterior density looks like:

The most common single number summary is:

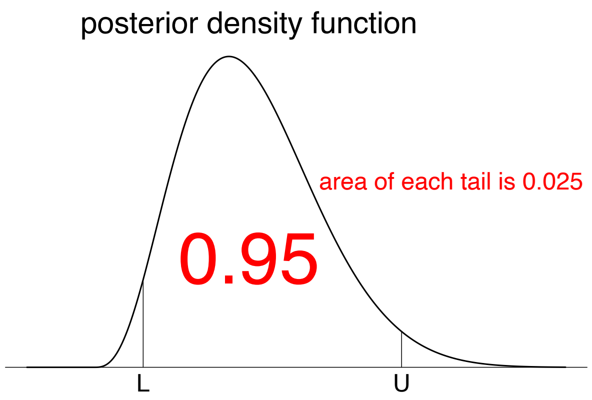
$$\widehat{\wp}_{\text{Bayes}} = \boldsymbol{E}(\wp|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = \quad .$$

Compare this with the frequentist answer:

$$\widehat{\wp} = \frac{}{} = \quad .$$

A common interval answer in Bayesian statistics is a

95% credible interval



(L, U) is the 95% credible interval for $\wp$.

These days we can do the required calculations quickly in R ...

PROBLEM!

Most practical Bayesian inference problems have integrals that cannot be solved analytically.

Bayes estimates and 95% credible intervals require quadrature (e.g. trapezoidal rule).

But quadrature is hard to impossible in higher dimensions.

Until 1990 problems like this made practical Bayesian analysis very difficult.

The last 34 years has seen a revolution in Bayesian analysis, fuelled mainly by

Markov Chain Monte Carlo (MCMC)

methodology.

A Short History of Bayesian Inference

- mid 1700s: Bayes Theorem established by Reverend Thomas Bayes.
- next 250 years: Lots of philosophical discussion and debate on Bayesian versus frequentist inference. But most practical statistics was frequentist.
- 1990: MCMC introduced to Statistics in *Journal of the American Statistical Association* paper by A.E. Gelfand & A.F.M. Smith.

- mid 1990s: First ‘professional’ MCMC software package started. Named BUGS (Bayesian inference Using Gibbs Sampling). However, clumsy to use.
- 2005: R package BRugs released. It allows script-based Bayesian analyses, run from inside R. Soon after rjags released.
- 2024: Even better R packages being developed such as rstan. We will use rstan in Laboratories 2–4 (so your laptops soon need to be set up for rstan).

♦ ♦ ♦