

## Bayesian Statistical Inference

### Introductory Toy Example

Suppose that we're interested in

$p$  = probability that randomly chosen person is left-handed.

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

$$X_1, \dots, X_n$$

where

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

and estimate  $p$  via the sample proportion

$$\hat{p} = \frac{\sum_{i=1}^n X_i}{n}.$$

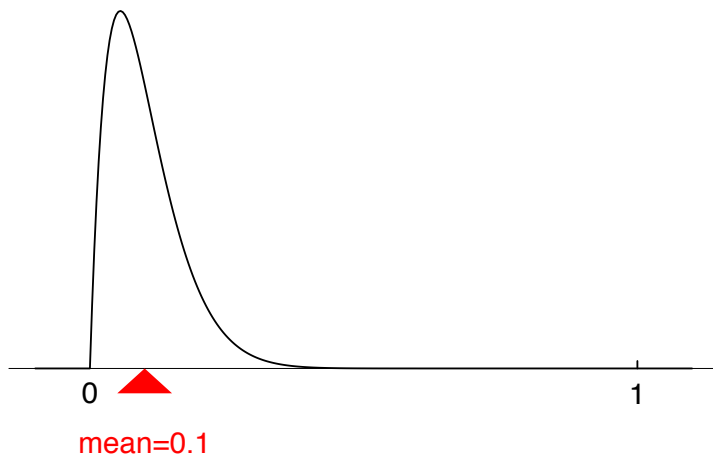
For large  $n$  we can get an approximate 95% confidence interval for  $p$  via e.g.

$$\hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}.$$

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

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

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



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

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

the Beta(2,18) density.

My model for the data and its dependence on  $p$  is

$$[X_i|p] = \begin{cases} p, & X_i = 1 \\ 1 - p, & X_i = 0 \end{cases}$$

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

We can write this neatly as

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

By independence

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

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

I now update my belief via the **posterior** density:

$$[p | X_1, \dots, X_n].$$

$$\begin{aligned} [p | X_1, \dots, X_n] &= \frac{[p, X_1, \dots, X_n]}{[X_1, \dots, X_n]} \\ &= \frac{[X_1, \dots, X_n | p][p]}{[X_1, \dots, X_n]} \\ &\quad \text{(Bayes Theorem)} \\ &\propto [X_1, \dots, X_n | p][p] \\ &= \prod_{i=1}^n p^{X_i} (1-p)^{1-X_i} \\ &\quad \times 342 p (1-p)^{17} \\ &\propto p^{1+\sum_{i=1}^n X_i} (1-p)^{17+n-\sum_{i=1}^n X_i} \end{aligned}$$

Now collect data on class participants:

$$\begin{aligned} n &= \\ \sum_{i=1}^n X_i &= \end{aligned}$$

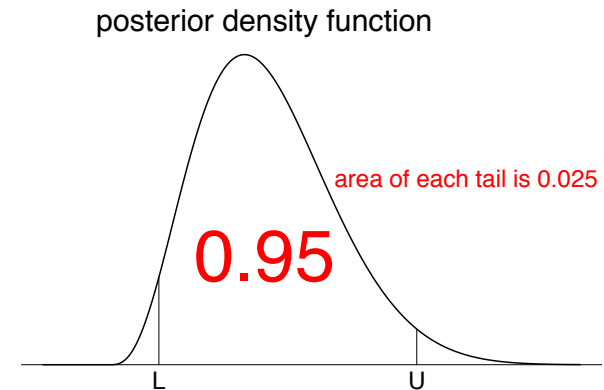
Posterior is

Beta( , ).

The posterior density looks like:

A common **interval** answer in Bayesian statistics is a

**95% credible interval**



The most common single number summary is:

$$\hat{p}_{\text{Bayes}} = E(p|X_1, \dots, X_n) = \quad .$$

Compare this with the frequentist answer:

$$\hat{p} = \text{---} = \quad .$$

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

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

## More Realistic Example

How would a Bayesian treat the logistic regression situation:

$$[y_i | \beta_0, \beta_1] \stackrel{\text{ind.}}{\sim} \frac{e^{y_i(\beta_0 + \beta_1 x_i)}}{1 + e^{\beta_0 + \beta_1 x_i}}$$

$$\text{logit}\{P(\text{BPD}_i = 1)\} = \beta_0 + \beta_1 \text{birthweight}_i \quad ?$$

More generally:

$$\text{logit}\{P(y_i = 1)\} = \beta_0 + \beta_1 x_i, \quad 1 \leq i \leq n.$$

Parameters of interest are  $\beta_0$  and  $\beta_1$ .

$\beta_0$  and  $\beta_1$  are now treated as **random**.

Usually (although not always) the  $x_i$ 's are treated as fixed.

So we have

A common prior for the  $\beta_0, \beta_1$  is

$$N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\beta_0}^2 & 0 \\ 0 & \sigma_{\beta_1}^2 \end{bmatrix} \right)$$

for fixed numbers  $\sigma_{\beta_0}^2$  and  $\sigma_{\beta_1}^2$ .

The posterior for  $\beta_1$  is

$$\begin{aligned}
[\beta_1 | \mathbf{y}] &= \frac{[\mathbf{y}, \beta_1]}{[\mathbf{y}]} \\
&\propto [\mathbf{y}, \beta_1] \\
&= \int_{-\infty}^{\infty} [\mathbf{y}, \beta_0, \beta_1] d\beta_0 \\
&= \int_{-\infty}^{\infty} [\mathbf{y} | \beta_0, \beta_1] [\beta_0, \beta_1] d\beta_0 \\
&\propto \int_{-\infty}^{\infty} \prod_{i=1}^n \frac{e^{y_i(\beta_0 + \beta_1 x_i)}}{1 + e^{\beta_0 + \beta_1 x_i}} e^{-\beta_0^2 / (2\sigma_{\beta_0}^2)} e^{-\beta_1^2 / (2\sigma_{\beta_1}^2)} d\beta_0 \\
&= e^{-\frac{\beta_1^2}{2\sigma_{\beta_1}^2}} \int_{-\infty}^{\infty} e^{\sum_{i=1}^n \{y_i(\beta_0 + \beta_1 x_i) - \ln(1 + e^{\beta_0 + \beta_1 x_i})\} - \frac{\beta_0^2}{2\sigma_{\beta_0}^2}} d\beta_0
\end{aligned}$$

## PROBLEM!

We can't solve this integral analytically.

$[\beta_1 | \mathbf{y}]$  has no closed form.

Bayes estimates for  $\beta_1$  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 21 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.

