Single-parameter models

θ - Scalar value

2.1 Estimating probability from binomial data

Estimate an unknown population proportion from n Bernoulli trails $Y_1, Y_2 \dots Y_n$ $\quad Y_i \in \{0, 1\}$

The sequence $Y_1 Y_2 \dots Y_n$ must consist of n exchangeable trails

Y - Number of successfull trails

θ - Proportion of success in the population.

$\cancel{Y = \begin{bmatrix} Y_1 \\ Y_1 \\ \vdots \\ Y_n \end{bmatrix}}$

Sampling model

$$P(Y|\theta) = Bin(Y|n, \theta) = \binom{n}{Y} \theta^Y (1-\theta)^{n-Y}$$

Example - Female births

θ - True proportion of female births
Y - Number of female births in a sample of n

Assume prior distribution for θ is uniform on $[0,1]$

Posterior density for θ

$$P(\theta|Y) \propto \theta^Y(1-\theta)^{n-Y}$$

For fixed n and Y the factor $\binom{n}{Y}$ can be regarded as a constant and removed.

Estimate $P_r\left(\theta \in (\theta_1, \theta_2) \mid Y\right)$

Assume θ has a prior uniform distribution on $[0,1]$

$$P_r\left(\theta \in (\theta_1,\theta_2) \mid Y\right) = \frac{P_r\left(\theta \in (\theta_1,\theta_2), Y\right)}{P(Y)}$$

$$= \frac{\int_{\theta_1}^{\theta_2} P(Y\mid\theta)\, \overset{1}{\overset{\shortparallel}{P(\theta)}}\, d\theta}{P(Y)}$$

$$= \frac{\int_{\theta_1}^{\theta_2} \binom{n}{Y} \theta^Y (1-\theta)^{n-Y} d\theta}{P(Y)}$$

$$P(Y) = \int_0^1 \binom{n}{Y} \theta^Y (1-\theta)^{n-Y} d\theta = \frac{1}{n+1} \qquad \text{for } Y = 0, 1, \dots n.$$

Example continued

Between 1745 - 1770 there were born 241 945 females and 251527 males in Paris.

θ - Probability that any birth is female.

Prior: uniform [0,1]

$$P_r\left(\theta \geq \frac{1}{2} \middle| Y = 241\,945,\ n = 241945 + 251527\right) \approx 1{,}15 \times 10^{-43}$$

<u>Prediction</u>

$\tilde{Y}$ ~ Result of a new trail, exchangeable with the n first.

$$P(\tilde{Y}=1 \mid Y) = \int_0^1 P_r(\tilde{Y}=1 \mid \theta, Y)\, P(\theta \mid Y)\, d\theta$$

$$= \int_0^1 P_r(\tilde{Y}=1/\theta)\, P(\theta/Y)\, d\theta$$

$$= \int_0^1 \theta \; P(\theta \mid Y)\, d\theta = E(\theta \mid Y) = \frac{Y+1}{n+2}$$

This result is known as "Laplace's law of succession."

2.2 Posterior as a compromise between data and prior information

The process of Bayesian inference involves passing from a prior distribution $p(\theta)$ to a posterior distribution $p(\theta|y)$.
We therefore expect there to be some general relations

Mean $$\boxed{E(\theta) = E(E(\theta|y))}$$

The prior mean is the average of all posterior means over the distributions of possible data.

$$E(\theta) = \iint \theta\, p(\theta, y)\, d\theta\, dy = \iint \theta\, p(\theta|y)\, d\theta\, p(y)\, dy$$
$$= \int E(\theta|y)\, p(y)\, dy = E(E(\theta|y)).$$

Variance

$$var(\theta) = E(var(\theta|y)) + var(E(\theta|y))$$

The posterior variance is on average smaller than the prior variance, by an amount that depends on the variation in posterior means over the distribution of all possible data.