

Single-parameter models

θ - Scalar value

2.1 Estimating probability from binomial data

Estimate an unknown population proportion from n Bernoulli trials y_1, y_2, \dots, y_n $y_i \in \{0, 1\}$

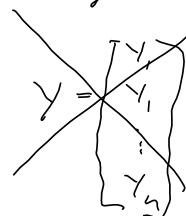
The sequence y_1, y_2, \dots, y_n must consist of n exchangeable trials

Y - Number of successful trials

θ - Proportion of success in the population.

Sampling model

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



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 $\Pr(\theta \in (\theta_1, \theta_2) | Y)$

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

$$\begin{aligned} \Pr(\theta \in (\theta_1, \theta_2) | Y) &= \frac{\Pr(\theta \in (\theta_1, \theta_2), Y)}{P(Y)} \\ &= \frac{\int_{\theta_1}^{\theta_2} P(Y|\theta) P(\theta) d\theta}{P(Y)} \\ &= \frac{\int_{\theta_1}^{\theta_2} \binom{n}{Y} \theta^Y (1-\theta)^{n-Y} d\theta}{P(Y)} \end{aligned}$$

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

Example continued

Between 1745 - 1770 there were born 241 945 females and 251 527 males in Paris.

θ - Probability that my birth is female.

Prior: uniform $[0, 1]$

$$\Pr(\theta \geq \frac{1}{2} | Y = 241945, n = 241945 + 251527) \approx 1.15 \times 10^{-43}$$

Prediction

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

$$\begin{aligned} P(\tilde{Y}=1 | Y) &= \int_0^1 P_r(\tilde{Y}=1 | \theta, Y) P(\theta | Y) d\theta \\ &= \int_0^1 P_r(\tilde{Y}=1 | \theta) P(\theta | Y) d\theta \\ &= \int_0^1 \theta P(\theta | Y) d\theta = E(\theta | Y) = \frac{Y+1}{n+2} \end{aligned}$$

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.

$$\begin{aligned} E(\theta) &= \iint \theta P(\theta, y) d\theta dy = \iint \theta P(\theta|y) d\theta P(y) dy \\ &= \int E(\theta|y) P(y) dy = E(E(\theta|y)). \end{aligned}$$

Variance

$$\text{var}(\theta) = E(\text{var}(\theta|y)) + \text{var}(E(\theta|y))$$

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.