Statistics 3858b : Bayesian Nonconjugate Prior Example

Bernoulli with Non-Conjugate Prior

For Binomial we usually use a conjugate prior : Θ having a Beta distribution

$X|\Theta = \theta \sim \text{Bernoulli}(\theta)$

pmf is

$$f(x|\theta) = (1-\theta)^{1-x}\theta^x, \ , x = 0, 1 \ .$$

Consider the prior proportional to

$$f_\Theta(\theta) \propto (1+\theta) I(0 \le \theta \le 1) \ .$$

This yields the prior as

$$f_\Theta(\theta) = \frac{2}{3}(1+\theta) I(0 \le \theta \le 1) \ .$$

Notice $\text{E}(\Theta) = \frac{5}{9}$ is the prior mean.

Below we consider what happens if we instead use a non-conjugate prior. We will see that while all our calculations go through we must carry out integrations to find the normalizing constants and again to find the expectation of the posterior in order to find the Bayes estimate.

Consider observing a single Bernoulli r.v. X.

There are two cases to consider, observing $X = x$, where $x = 0$ or 1.

Posterior calculations, for arguments $\theta \in [0, 1]$, and with posterior 0 otherwise :

$$f_{\text{Posterior}}(\theta|X=0) \quad \propto \quad (1-\theta)(1+\theta) = 1 - \theta^2 \ .$$

Therefore

$$f_{\text{Posterior}}(\theta|X=0) = \frac{3}{2}\left(1-\theta^2\right) \ .$$

$$f_{\text{Posterior}}(\theta|X=1) \quad \propto \quad \theta(1+\theta) \ .$$

Therefore

$$f_{\text{Posterior}}(\theta|X=1) = \frac{6}{5}\theta\,(1+\theta) \ .$$

Then

$$\mathrm{E}(\Theta|X=0) = \frac{3}{8}$$

and

$$\mathrm{E}(\Theta|X=1) = \frac{7}{10} \ .$$

Thus the Bayes estimator is

$$\hat{\Theta}_{\text{Bayes}} = h(X)$$

where

$$h(x) = \begin{cases} \frac{3}{8} & \text{if} \quad x = 0 \\ \frac{7}{10} & \text{if} \quad x = 1 \end{cases}$$

There are other statistical quantities that one can calculate from this posterior distribution. We can also obtain a central .95 prediction interval, which is also called the Bayes confidence interval. For this we need to find the .025 and .975 quantiles for the posterior. Of course this depends on x, the observed value of X. In this case the posterior is not of a nice parametric family so we can jut calculate these directly.

If we then take second observation, we then treat the posterior based on one observation as the prior for this second observation.