Quiz 6

Dustin Leatherman

October 25, 2020

1 1

A survey of commuters in 10 counties likely to be effected by a proposed addition of a high occupancy vehicle (HOV) lane was conducted. Let n_i be the number of respondents in county i and Y_i be the number in favor of the HOV lane. Consider two potential models for these data (responses are independent over county for both models).

$$\text{Model 1} \quad Y_i \sim Binomial(n_i, p_i) \text{ with priors } p_i \sim Beta(0.5, 0.5)$$

$$\text{Model 2} \quad Y_i \sim Binomial(n_i, p) \text{ with priors } p \sim Beta(0.5, 0.5)$$

1.1 a

Give the posterior distribution of p under Model #2.

Likelihood: $Y_i|p \sim Bin(n_i, p)$
Prior: $p \sim Beta(0.5, 0.5)$
Posterior: $p|Y_i \sim Beta(Y_i + 0.5, n_i - Y_i + 0.5)$
Given that a Beta Prior is conjugate when the Likelihood is Binomial.

1.2 b

Give a scenario where inspecting the posterior distributions under the two models would suggest that Model #2 is invalid.

Model 2 suggests that all i counties have a similar distribution in opinions regarding HOV lanes. It is an unrealistic assumption. For example, opinions on an HOV lane would likely be different between Cook and Kendall counties in Illinois.

2 2

Each of n study participants slept in a cold room for a month and a warm room for a month. Let Y_i be the difference between the average number of hours of sleep per night in these two months. Assum $Y_1, ..., Y_n$ are independent and

$$Y_1, ..., Y_n|\mu, \sigma^2 \sim N(\mu, \sigma^2)$$

with priors $\mu \sim N(0, 100)$ and $\sigma^2 \sim InvGamma(0.01, 0.01)$. The goal is to test whether the mean of Y_i is positive or negative. The hypothesis are

$$H_1 : \mu < 0$$

$$H_2 : \mu \geq 0$$

2.1 a

Write an integral expression, denoting $p(\mu, \sigma^2|Y)$ as the posterior distribution for the posterior probability of H_1. Note that you do not need to give an equation for p.

Likelihood: $Y|\mu, \sigma^2 \sim N(\mu, \sigma^2)$
Priors:

$$\begin{aligned} \mu \sim& N(0, 100) \\ \sigma^2 \sim& InvGamma(0.01, 0.01) \end{aligned} \tag{1}$$

$$
\begin{aligned}
P(\mu, \sigma^2|Y) =& \frac{P(Y|\mu, \sigma^2)P(\mu, \sigma^2)}{\int P(Y|\mu, \sigma^2)P(\mu, \sigma^2)d\mu d\sigma^2} \\
=& \frac{P(Y|\mu, \sigma^2)P(\mu)P(\sigma^2)}{\int P(Y|\mu, \sigma^2)P(\mu)P(\sigma^2)d\mu d\sigma^2} \text{ (Assumes } \mu \text{ and } \sigma^2 \text{ are independent)} \\
=& \frac{\int_{-\infty}^{\infty} P(Y|\mu, \sigma^2) \cdot \int_{-\infty}^{0} P(\mu) \cdot \int_{0}^{\infty} P(\sigma^2)}{\int P(Y|\mu, \sigma^2)P(\mu)P(\sigma^2)d\mu d\sigma^2}
\end{aligned} \quad (2)
$$

2.2 b

Give three computational approaches to computing/approximating the posterior probability of H_1. Give a Pro and Con of each.

1. MCMC via Gibbs Sampling. This problem decomposes nicely into known Full Conditional distributions making it straightforward to implement but involves computation over a presumably largish number of iterations.

2. Compute MAP Estimators of priors, then MAP Estimator of posterior. Very fast but only provides a point estimate with no description of uncertainty.

3. Invoke Bayesian CLT by using MAP Estimators of the priors as parameters of the Likelihood function. This is faster than MCMC with Gibbs Sampling but ignores the uncertainty of μ and σ which makes the credible interval narrower than they should be.