Comparing groups

Stat 340, Fall 2021

Example 1. Comparing proportions across 3+ groups

SurveyUSA conducted a poll from October 29 to 31, 2020. A total of 1,500 registered voters were polled and asked to indicate their preference in the presidential election. $y_1 = 743$ supported Biden, $y_2 = 607$ supported Trump, and $y_3 = 150$ supported other candidates, expressed no opinion, or were unsure.

The counts an be modeled as arising from a multinomial distribution with sample size $n = 1500$ and respective probabilities θ_1, θ_2, and θ_3.

For this example, use a uniform (uninformative) prior on $(\theta_1, \theta_2, \theta_3)$; that is,

$$\pi(\theta_1, \theta_2, \theta_3) \propto 1.$$

(a) Show that the joint posterior distribution of $(\theta_1, \theta_2, \theta_3)$ is a Dirichlet distribution of the form:

$$f(\theta_1, \ldots, \theta_k | a_1, \ldots, a_k) = \frac{\Gamma(\alpha_1 + \ldots + \alpha_k)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_k)} \theta_1^{\alpha_1} \cdots \theta_k^{\alpha_k}, \ \theta_1 \in [0, 1], \ldots, \theta_k \in [0, 1], \ \sum_{j=1}^{k} \theta_j = 1.$$

(b) You can use the `rdirichlet()` function in the **LearnBayes** R package (which is loaded by **ProbBayes**) to simulate from your posterior from part (a). To do this, store the parameters from your posterior in the vector `alpha`, then draw 1000 simulations from the posterior.

```
# Fill in the blanks with the posterior parameter values
alpha <- c(___, ___, ___)

# Draw samples from the posterior
library(LearnBayes)
theta <- rdirichlet(1000, alpha)
```

(c) Using your posterior draws, find the posterior density of $\theta_1 - \theta_2$, the difference in proportions between Biden and Trump. Create a histogram of this posterior.

(d) What is the posterior probability that a larger proportion of registered voters support Biden?

Example 2. A Bayesian two-sample t-test

Suppose that we observe two independent normal samples, the first distributed according to an $\mathcal{N}(\mu_1, \sigma_1)$ distribution, the second according to an $\mathcal{N}(\mu_2, \sigma_2)$ distribution. Denote the first sample by $X_1, \ldots, X_m$ and the second sample by $Y_1, \ldots, Y_n$. Suppose also that the parameters $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$ are assigned the vague prior

$$\pi(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2) \propto \frac{1}{\sigma_1^2 \sigma_2^2}.$$

(a) Write down an expression for the posterior density.

(b) Show that the vectors (μ_1, σ_1^2) and (μ_2, σ_2^2) are independent. To do this, show that the posterior distribution derived in (a) can be factored into the product of two posteriors, where these posteriors are the same posterior we derived during class on Monday for the one group problem.

(c) Keeping in mind that (μ_1, σ_1^2) and (μ_2, σ_2^2) are independent, describe how to simulate from the joint posterior density of $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$. (Do this in words or using psuedo code. Hint - you can apply the grid approximation approach twice.)

(d) The following data give the mandible lengths in millimeters for 10 male and ten female golden jackals in the collection of the British Museum. Using simulation procedure you outlined in part (c), find the posterior density of the difference in mean mandible length between the sexes. Is there sufficient evidence to conclude that the males have a larger average?

```
males    <- c(120, 107, 110, 116, 114, 111, 113, 117, 114, 112)
females  <- c(110, 111, 107, 108, 110, 105, 107, 106, 111, 111)
```