HBS-3: Bayesian Statistics - Pre-reads and Pre-workshop Exercise

Pre-reads:
1. Spiegelhalter DJ, Abrams KR, Miles JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Wiley, Chichester, 2004. Read chapters 1 to 2, and sections 3.5 to 3.6 from this book.

2. *Optional reading for those already familiar with Bayesian methods:* Neuenschwander B, Branson M, Gsponer T (2008). Critical aspects of the Bayesian approach to phase I cancer trials. Statistics in Medicine 27: 2420 - 2439.

Pre-workshop Exercise
Objective: Bayesian statistics brings probability calculations to the forefront. This exercise aims to provide familiarity with simulating from probability distributions in the context of Bayesian analysis for binary and count data variables.

Note: Some hints for the exercises are provided for simulations using `R` *software. In case using SAS, you may use the rand() function to carry out the simulations for the exercises.*

1. Binary endpoints are analyzed using binomial distributions which is characterized by the proportion parameter π. Here π indicates the probability of the outcome of interest. Beta distribution is a common choice for specifying prior distributions for the proportion parameter π.

Simulate the proportion π from Beta(α, β) prior distributions under five different scenarios. Use `rbeta()` function from the `R` software for simulations.

Draw 10,000 samples from each of the following five Beta distributions for proportion (π): (i) Beta(8, 12), (ii) Beta(4, 6), (iii) Beta(2, 2), (iv) Beta(1, 1) and (v) Beta(0.5, 0.5).

Plot the density for these five beta distributions (Hint: use `p = rbeta(...)` function for drawing random samples from the beta distribution and use `plot(density(p))` to plot the density).

What do these different prior distributions tell about our prior belief / knowledge on the response rate (proportion of subjects who would experience the endpoint) before the conduct of the trial?

2. *Comparing the summary statistics for the beta distribution* Beta$(\alpha = 4, \beta = 6)$:
(a) by directly calculating from the known expressions for mean and variance, and
(b) by drawing samples from the density to calculate them.

For (a), the mean $= \frac{\alpha}{(\alpha+\beta)}$ and the variance $= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$. Calculate the mean and variance.
For (b), draw 10,000 random samples from Beta$(\alpha = 4, \beta = 6)$ and calculate the mean and variance from this sample.

Provide your findings on how the summary statistics compare between the two methods. Does a large sample of random draws from the given density approximate the summary features of the density one expects to observe?

3. *Simulating from the predictive distribution for a binary endpoint to check the typical data that are expected to be observed*:
First draw M = 10,000 samples for $\pi \sim \text{Beta}(\alpha = 4, \beta = 6)$ using the `rbeta()` function. The M samples are denoted by $\{\pi^{(m)}, m = 1, \ldots, M\}$, where $\pi^{(m)}$ indicates the m^{th} sample drawn for π. Then, for each value of $\pi^{(m)}$, draw a binomial variable (y) where $y \sim \text{Bin}(n = 12, \pi^{(m)})$ using the `rbinom()` function.

Plot the density for π and the histogram for y. Repeat these steps for options (iii) and (iv) from question 1. What type of change in the distribution of y is observed by using different prior distributions for π?

4. *Analyzing a phase-2 single-arm trial with binary endpoint*:
A single arm trial is designed to evaluate the response rate for a binary endpoint (i.e., proportion of subjects who would experience the outcome of interest) for an experimental drug.

The input obtained from the scientific team were utilized in specifying the prior distribution for π, which is approximated by $\text{Beta}(\alpha = 8, \beta = 12)$, i.e., the mean response rate for the experimental treatment would be centered around 0.4, and that the mean response rate would lie between 0.2 and 0.6 with about 95% certainty.

Based on the data observed in the trial, this distribution is updated to $\text{Beta}(\alpha + y, \beta + n - y) = \text{Beta}(8 + 21, 12 + 45 - 21)$.

Draw 10,000 samples from the updated Beta distribution to calculate the following summaries:
(i) the mean, the median and the 2.5th and 97.5th percentiles
(ii) The response rate in the standard of care (SoC) is known to be 0.265 from previous trials. Calculate the probability that the response rate in the current trial with the experimental drug is larger than 0.265. To compute this quantity, calculate the proportion of samples in which the sampled values are larger than 0.265.

5. *Simulating from the predictive distribution for endpoints that are counts such as number of infections, number of hospitalizations, etc. per subject and to check the typical data that are expected to be observed*:
Count data are usually modelled using the Poisson distribution with mean λ which characterizes the average number of events experienced per subject. The gamma distribution is a common choice for specifying a prior distribution for λ.

First draw $M = 10,000$ samples for $\lambda \sim \text{Gamma}(\alpha = 8, \beta = 2)$ using the `rgamma()` function. Here, α and β correspond to shape and rate (i.e., 1/scale) parameters, respectively. Then, for each value of $\lambda^{(m)}$ in $\{\lambda^{(m)}, m = 1, \ldots, M\}$, draw a Poisson variable (y) where $y \sim \text{Pois}(n = 20, \lambda^{(m)})$ using the `rpois()` function. Here, $\lambda^{(m)}$ denotes the m^{th} sample drawn for λ.

Plot the density for λ and the histogram for the sample mean $\overline{y}$. Repeat these steps for $\lambda \sim \text{Gamma}(\alpha = 4, \beta = 4)$. What type of change in the distribution of $\overline{y}$ is observed by using different prior distributions for λ?

Refresher to some commonly (and not-so-commonly) used distributions in Bayesian inference:
Passing familiarity with the following distributions will be assumed for the workshop: beta, gamma, inverse-gamma, normal, half-normal, bivariate normal, binomial and Poisson. Chapter 2, in particular, Section 2.6 of the book by Spiegelhalter et al., 2004 (included as pre-read) covers these distributions.