Bayesian vs. Classical Statistics **Math 422 - Workshop**

Last year, I surveyed my 47 introductory statistics students and found out that 29 were born in Virginia. Suppose I want to estimate the percent of all HSC students who were born in Virginia. We could model the number of students in the survey who were born in Virginia with a Binom$(47, p)$ distribution.

1. What's the likelihood function for p and what is the posterior distribution for p if we start with a Unif$(0, 1)$ prior?

2. According to the posterior distribution, what is the probability that more than half of the students at Hampden-Sydney were born in Virginia? (Hint: Use something like WolframAlpha to do the integration for you.)

3. In classical statistics, you can do a hypothesis test to find out if more than half of HSC students were born in Virginia. You use the formula below to test the null hypothesis $H_0 : p_{\text{HSC}} = 0.5$ vs. the alternative hypothesis $H_A : p_{\text{HSC}} \neq 0.5$.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{29}{47} - 0.5}{\sqrt{0.25/47}} = 1.60.$$

   Then you would find the corresponding p-value using the normal distribution. Here the 2-sided p-value is: 11%. What is the right conclusion to draw from this p-value about Hampden-Sydney students?

4. The p-value in the last problem, and the answer to problem 2 are both conditional probabilities that could be expressed using the notation $P(A \mid B)$. What are the events A and B for the answer to problem 2? What about for the p-value in problem 3?

Not only does Bayesian statistics give us a much more direct approach to hypothesis testing, it also has something that works like a confidence interval called a **credible interval**. To make a 95% credible interval, find the locations where the CDF of the posterior distribution is 2.5% and 97.5%. Then you can be 95% sure that the parameter of interest is between those two values.

5. The posterior distribution you found in problem 1 should have been a beta distribution. Using the R command `qbeta(percentile,alpha,beta)` you can find the quantiles for the beta distribution with parameters α and β. Use that command to find the 95% credible interval for the proportion of all HSC students who were born in Virginia.

6. Use the formula $\hat{p} \pm 1.96\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$ to make a classical 95% confidence interval for the proportion of all HSC students that were born in Virginia. Is your answer similar to the credible interval in the last problem?

Remark. Both classical and Bayesian techniques are effective and widely used. Bayesian statistics is a little harder to learn, and the answers depend on what you choose for your prior distribution. On the other hand, the answers you get from Bayesian statistics are often easier to interpret.