# Problem Section 8

## The Beta Binomial Model

**Learning Outcomes**

The problems are designed to build conceptual understanding and problem-solving skills. The emphasis is on learning to find, evaluate and build confidence. The specific tasks include:

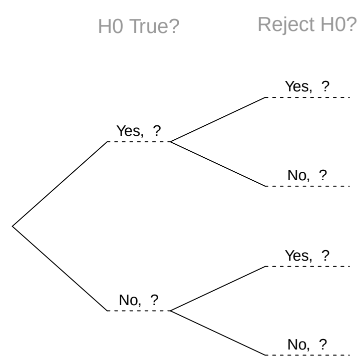
- Calculate the posterior distribution for discrete $\theta$
- Elicit the hyperparameters of a subjective beta prior
- Calculate the posterior distribution in a beta binomial model
- Back up and support work with relevant explanations

---

## Exercises

1. Are many medical discoveries actually Type 1 errors? In medical research, suppose that 10% of null hypotheses are actually false, and that when a null hypothesis is false, the chance of making a Type II error and failing to reject it (for example, due to insufficient sample size) is 0.55.

a. Given that we reject a null hypothesis at level $\alpha = 0.05$, use Bayes' Theorem to calculate the False Discovery rate: $P(H_0 \text{ is true}|\text{Reject } H_0)$. Use the tree diagram to structure your thinking.




---

Note: The Benjamini Hochberg procedure tries to control the FDR by changing the significance threshold when testing multiple hypotheses. The smaller P-values face stricter thresholds, while the larger ones have more lenient thresholds. So it adaptively controls the Type I error, rather than for every hypothesis being tested.

The Bonferroni approach tries to minimize the Type 1 error across all the tests by conducting each test at the same lower $\alpha$ level. This also reduces the FDR, however, the Type II error is increased since this approach makes it much harder to reject any hypothesis.

---

2. Suppose $X = 3$ is a realization of geometric random variable indexed by a probability $\pi$. In other words, the PMF is

$$f_{\pi_0}(x) = (1 - \pi_0)^x \, \pi_0, \qquad x = 0, 1, 2, \ldots$$

Suppose $\pi$ which can only take one of three values $\frac{1}{4}, \frac{1}{2}, \frac{3}{4}$.

a. (Frequentist) Find the MLE $\widehat{\pi}_0$.

b. (Bayesian) Derive the posterior distribution $P(\pi = t|X = 3), \qquad t = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$ assuming the discrete uniform prior for $\pi$. Then find the posterior mean.

3. Suppose a random variable $X$ has PDF given by

$$f(x) = x^4 \cdot (1 - x)^5, \qquad 0 \le x < 1$$

a. What is the scaling constant of this PDF.

b. Can you name the distribution? What are the parameters?

c. What is the mean of this distribution? Find $E[X]$.

d. Use `qbeta` to find the median of this distribution.

4. The R function `leaflet_map` from the **leaflet** package can be used to generate random locations on the globe and to use Google maps to show you where these locations are:

```r
set.seed(3535)

leaflet_map(position=rgeo(n=20),mark=TRUE)
```

Use this to gather a sample of size 20 and calculate the proportion of the sample that represents a location which is covered with water. (You may need to zoom in for locations that are near where land and water meet)

Let's assume that $X$, the number out of the 20 randomly selected locations which are covered with water, is a binomial random variable with success probability $\pi$.

a. What is the MLE of $\pi_0$, the true value?

b. Calculate the posterior distribution for the proportion of the earth that is covered with water. Use a uniform prior for $\pi$. Make a plot of the uniform prior and the resulting posterior distribution on the same graph. What is the posterior mean?

c. Now suppose a Bayesian wants to use a beta distribution which reflects the prior knowledge that a majority of the earth is covered with water. Specifically, they assume that the mean of the prior is 0.70 and the variance is 0.05.

   i. Keeping in mind that for a $Beta(a, b)$ distribution,

$$\mu = E[X] = \frac{a}{a+b}$$

   and the variance is

$$\sigma^2 = Var[X] = \frac{ab}{(a+b)^2 \cdot (a+b+1)} = \mu \cdot (1 - \mu) \cdot \frac{1}{a+b+1}$$

   elicit the values for $a$ and $b$.

Hint: You have two equations in two unknowns $a$ and $b$. You are given $\mu = 0.7$ and $\sigma^2 = 0.05$ and you want to find $a$ and $b$.

ii. Now calculate the posterior distribution using this informative prior from part i. What is the posterior mean?

iii. Make a plot of your prior and the resulting posterior distribution in one graph.

d. Now suppose you consider a prior with a median of 0.7 and 95% of the distribution lies above 0.68. Use the function `beta.select` function from the **LearnBayes** package to find the values of $a$ and $b$ corresponding to these prior beliefs. Then repeat parts ii and iii from part b with this prior.