Lab 5

Name here

Suppose you have been hired as a statistical consultant by a NBA expansion team that will be moving to Big Sky, MT. Your goal is to help identify basketball players to bring to Montana.

Specifically, the team has one more spot to fill and is looking to add a free throw shooting specialist. Here are two options:

- Bugs Bunny. Bugs Bunny is a shooting guard that was 2 for 7 on free throws last year.
- Tasmanian Devil. Tasmanian Devil is a center that was 25 for 40 on free throws last year.

1.

Use a binomial model, along with uniform prior of the probability parameter, to model the free throw shooting for Bugs Bunny and Tasmanian Devil.

a (4 points)

Plot and summarize the posterior distribution for each player.

b (4 points)

Using the posteriors above, compute the probability that Bugs Bunny has a higher shooting percentage (θ parameter in the binary setting).

c (2 points) If both players had 25 free throws, who do you think will make more free throws? Why?

2.

Now assume that we know that shooting guards, like Bugs Bunny, tend to make about 80 percent of free throws and centers, like Tasmanian Devil, tend to make about 60 percent of free throws. This information can be imparted by more informative priors: Use Beta(40,10) for Bugs Bunny and Beta(30,20) for Tasmanian Devil as the prior distributions.

a (4 points)

Plot and summarize the posterior distribution for each player.

b (4 points)

Using the posteriors above, compute the probability that Bugs Bunny has a higher shooting percentage (θ parameter in the binary setting).

c (2 points)

If both players had 25 free throws, who do you think will make more free throws? Why?

3. (4 points)

Reflect on the results from question a and question b. Which analysis, and associated prior, do you prefer, why?

4. Bayesian Analysis on a continuous outcome

Let's model housing prices using our King County, WA dataset.

A normal distribution seems like a reasonable stating point, although we will see some potential issues.

$$price = \beta_0 + \epsilon \;\; \epsilon \sim N(0, \sigma^2)$$

So we have two parameters in this model β_0 and σ^2

a. Specify and plot a prior distributions for β_0 (4 points)

b. (4 points)

Using your prior from part a, along with a prior such that $\sigma \sim Gamma(.00001, .00001)$, model the mean housing price in the Seattle dataset.

```r
seattle <- read_csv('http://www.math.montana.edu/ahoegh/teaching/stat408/datasets/SeattleHousing.csv')
```

```
## Rows: 869 Columns: 14
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (14): price, bedrooms, bathrooms, sqft_living, sqft_lot, floors, waterfr...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
y <- seattle$price
n <- nrow(seattle)
```

For reference, you should expect to get similar values to `summary(lm(price ~ 1, data=seattle))`

```r
model_normal<- "model{
  # Likelihood
  for (i in 1:n){
    y[i] ~ dnorm(beta0, 1/sigma^2)
  }

  # Prior
  beta0 ~ dnorm(750000, 1/200000^2)
  sigma ~ dgamma(.00001, .00001)
}"
```