# STAT 4880/5088 Exam 1

Name: ________________________________________

Tuesday, October 1, 2024

Write directly on this exam. You may "show work" by handing in and R script, .Rmd file, or knit Markdown document.

You may use R, the internet, and any reference material. You are not allowed to communicate with anyone - no email, messaging, internet forums, AI, etc. If you happen to find exact copies of the exam questions on a "homework help" website, please bring that to the instructor's attention.

### Honor Pledge (10 points)

The work I have submitted represents my own effort. While working on this exam, I did not use AI or communicate in any form with individuals other than the instructor.

**Signed:**

________________________________________

### Problem 1 (10 points)

Approximately 1 in 10 people in the US are left handed, and 32/1000 are a twin (or triplet, etc.). Interestingly, 21% of twins are left handed.

What is the probability that a left handed person is a twin?

**Solution** P(Twin | Lefty) = P(Twin) * P(Lefty | Twin) / P(Lefty) = $\frac{(32/1000)\cdot 0.21}{1/10} = 0.0672$.

That's a little more than double the prior probability of 0.032.

### Problem 2 (10 points)

The racial breakdown of the U.S. is 59% White, 13% Black, 19% Hispanic, 6% Asian, and 3% multi-racial or other.

Pet ownership varies greatly by race, with 68% of Whites, 34% of Blacks, 66% of Hispanics, 37% of Asians, and 62% of multi/other saying they own a pet.

Given that an adult owns a pet, what is the probability they are White? What is the probability they are Black?

**Solution**

```r
prior <- c(white = .59, black = .13, hispanic = .19, asian = .06, other = .03)
likelihood <- c(.68, .34, .66, .37, .62)
post <- prior*likelihood / sum(prior*likelihood)
post
```

```
##      white      black   hispanic      asian      other
## 0.65598430 0.07226946 0.20503597 0.03629823 0.03041203
```

P(white | pet) = 0.656, P(black | pet) = 0.072.

## Problem 3 (10 points)

a. The Beta($\alpha$,$\beta$) distribution has no mode when $\alpha < 1$. Explain why not.

b. What about when $\beta < 1$?

**Solution** When $\alpha < 1$, there is a vertical asymptote at $x = 0$, so there is no $x$ where the function has a maximum. When $\beta < 1$ there is an asymptote at $x = 1$, causing the same problem.

## Problem 4 (10 points)

The Houston Astros baseball team is about to play the Detroit Tigers in the first game of the 2024 playoffs. Let $\pi$ be the probability that Houston wins the game.

Choose a Beta prior that represents your knowledge of this probability and explain your reasoning.

**Solution** I don't know much about these two teams, but I do know that baseball games tend to be pretty even. So I'll choose a Beta(2,2) prior which has a mean of 0.5 and falls to zero at the endpoints 0 and 1. If you know even less about baseball, you might consider a Beta(1,1) prior. If you want to get fancy, you could search and find that these teams played each other six times this year and the Astros won four of them. Then it would be reasonable to use a Beta(5,3) prior.

## Problem 5 (10 points)

Recall the Beta-Binomial model is $Y|\pi \sim \text{Binom}(n, \pi)$ and $\pi \sim \text{Beta}(\alpha, \beta)$.

Suppose we observe 10 trials with 9 successes and 1 failure.

a. Write the likelihood $L(\pi) = P(Y = 9|\pi)$ as a function of $\pi$.

b. Show that the maximum likelihood occurs at $\pi = 0.9$.

**Solution** $L(\pi) = \binom{10}{9}\pi^9(1-\pi)^1 \propto \pi^9 - \pi^{10}$

Then $L'(\pi) \propto 9\pi^8 - 10\pi^9$

Setting $L' = 0$ gives $\pi = 0$ where $L = 0$ and the maximum, $\pi = 9/10$.

## Problem 6 (10 points)

Let $\pi$ be the proportion of Missouri voters that intend to vote to re-elect Senator Josh Hawley in the November election. We will use a Beta-Binomial model for the election. Since we expect elections to be reasonably close, let's use a Beta(100,100) prior for $\pi$.

In August, the SLU/YouGov poll asked 900 likely voters and found that 477 of them planned to vote for Sen. Hawley.

a. What is the posterior distribution for $\pi$ given the poll results?

b. What is the posterior probability $P(\pi > 0.5 \mid \text{poll results})$? That is, what is the probability that Sen. Hawley wins re-election?

**Solution** a. The posterior is Beta(577,523). b. About 94.8%.

```r
1 - pbeta(0.5,577,523)
```

```
## [1] 0.9483547
```

---

The final three questions all use the `space-objects.csv` data, available on our course web page at https://turtlegraphics.org/bayes/data/space-objects.csv

## Problem 7 (10 points)

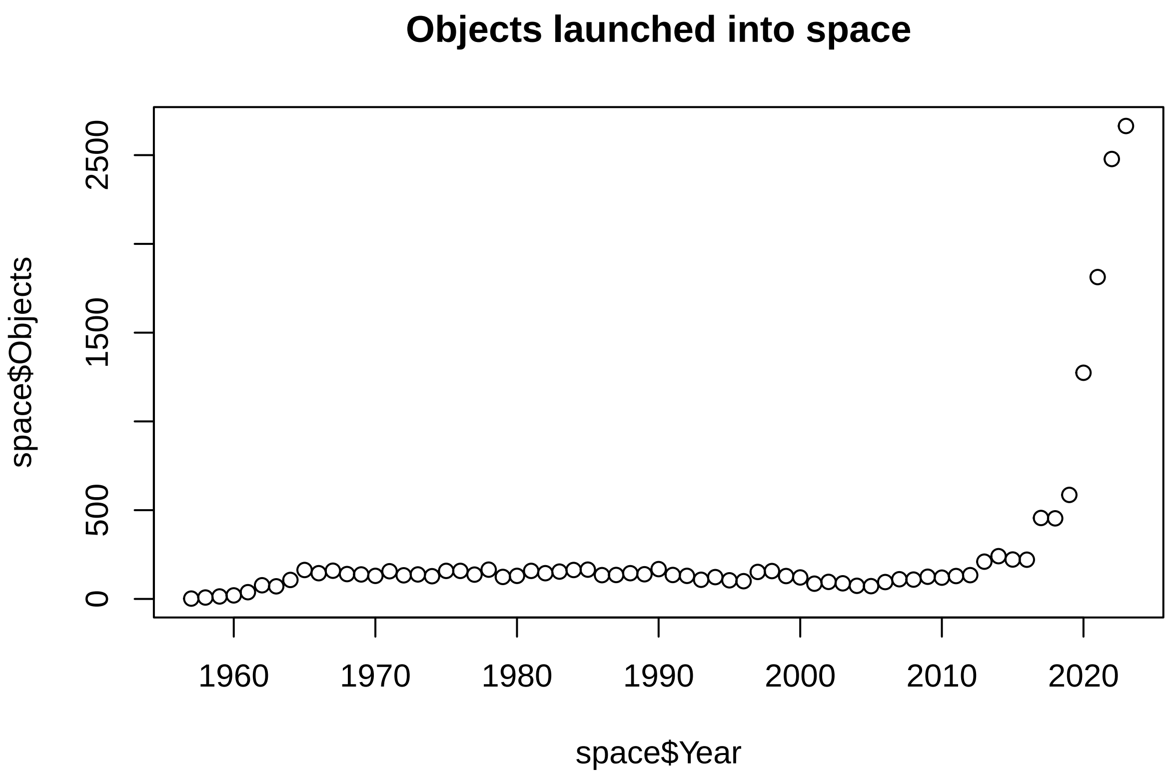
Since 1957, the world has been using rockets to send objects into space. The number for each year is stored in the `Objects` variable. Make a plot showing the number of Objects launched.

Looking at the plot, it seems reasonable to model these launches as a Poisson process from 1965 until 2017.

a. Explain why those years look like good cutoffs.
b. What were the total number of objects launched in the 53 years 1965-2017?

**Solution**

```r
space <- read.csv("https://turtlegraphics.org/bayes/data/space-objects.csv")
plot(space$Year, space$Objects, main="Objects launched into space")
```




```r
suppressMessages(library(dplyr))
space |> filter(Year >= 1965, Year <= 2017) |> summarize(sum(Objects))
```

```
##   sum(Objects)
## 1         7657
```

Before 1965, the rate was clearly lower, and after 2017 the rate got dramatically higher. But in that range it was fairly stable for a long time.

Total objects in those 53 years was 7657.

## Problem 8 (10 points)

(Continuing with the space objects data)

Let's use the Gamma-Poisson model to estimate the annual launch rate $\lambda$ for 1965-2017.

a. With a vague prior of $\lambda \sim \text{Gamma}(1,1)$ and data for 1965-2017, find the posterior distribution for $\lambda$ given the data,

b. What is the posterior mean of $\lambda$?

c. What is the posterior SD of $\lambda$?

**Solution** The posterior is Gamma(7658, 54) which has mean $7658/54 \approx 141.8$ and sd given by $\sqrt{7658}/54 = 1.62$.

## Problem 9 (10 points)

In the six full years 2018-2023, there have been 9269 objects launched into space.

a. With the posterior from the previous problem as your prior, incorporate this new information and give the new posterior distribution and mean for $\lambda$.

b. Does this accurately model the current state of space launches?

**Solution** The new posterior is Gamma(16927, 60) and has mean $16927/60 \approx 282.1$.

Clearly the model is not accurate anymore, because we're launching over 2000 objects per year now.