

S&DS 238 Problem Set #4

Shaun Radgowski

October 2, 2020

Problem 1

(1a)

If we consider the laser pole fulcrum height to be S and the position on the X-axis to be θ , then we can write an equation with the laser angle to the vertical α and the X-axis intercept point for the laser, x : $\tan(\alpha) = \frac{x-\theta}{S}$. From there, we can say:

$$\begin{aligned} F_X(x) &= P\{X \leq x\} = P\{S \tan(\alpha) + \theta \leq x\} \\ &= P\{A \leq \tan^{-1}(\frac{x-\theta}{S})\} = \frac{\tan^{-1}(\frac{x-\theta}{S}) - (-\frac{\pi}{2})}{\frac{\pi}{2} - (-\frac{\pi}{2})} = \frac{1}{\pi}(\tan^{-1}(\frac{x-\theta}{S}) + \frac{\pi}{2}) \\ f_X(x) &= F'_X(x) = \frac{1}{\pi S(1+(\frac{x-\theta}{S})^2)} = \frac{1}{S} \kappa(\frac{x-\theta}{S}) \end{aligned}$$

Where $\kappa(x) = \frac{1}{\pi(1+x^2)}$, the Cauchy distribution. The final line above is true due to the Chain Rule of differentiation, where the derivative of the expression inside of the arctan is inverse S . This is then multiplied by the derivative of the whole function with respect to that interior expression, which is $\frac{1}{\pi}$ times the derivative of \tan^{-1} .

(1b)

```
prob <- function(x, s, theta) {
  inside <- (x - theta) / s
  cauchy <- 1/(pi*(1 + inside^2))
  return((1/s)*cauchy)
}

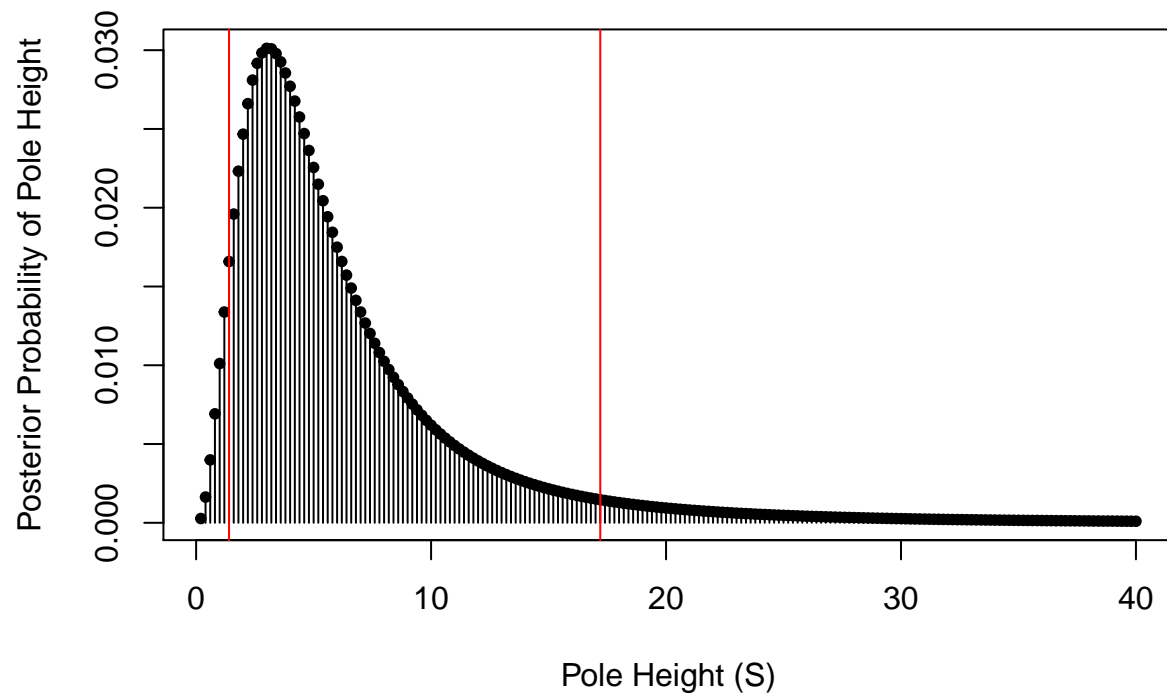
mat <- matrix(nrow=200, ncol=201)

x <- c(1, 2, 7.8, 9.2)
ss <- seq(.2, 40, 0.2)
thetas <- seq(-10, 30, 0.2)

for (i in 1:200) {
  for (j in 1:201) {
    mat[i, j] <- prod(prob(x, ss[i], thetas[j]))
  }
}
```

(1c)

```
myplot <- function(x, y, x_lab, y_lab) {  
  plot(x, y, type="h", xlab=x_lab, ylab=y_lab)  
  points(x, y, pch=20)  
}  
  
like_s <- rowSums(mat)  
  
prior <- rep(1/length(ss), length(ss))  
post <- prior*like_s / sum(prior*like_s)  
myplot(ss, post, y_lab="Posterior Probability of Pole Height", x_lab="Pole Height (S)")  
  
cs <- cumsum(post)  
L <- min(ss[cs > 0.05])  
R <- min(ss[cs >= 0.95])  
abline(v=L, col="red")  
abline(v=R, col="red")
```



The 5th percentile is at $S = 1.4$, and the 95th percentile is at $S = 17.2$ (as shown in the red lines above).

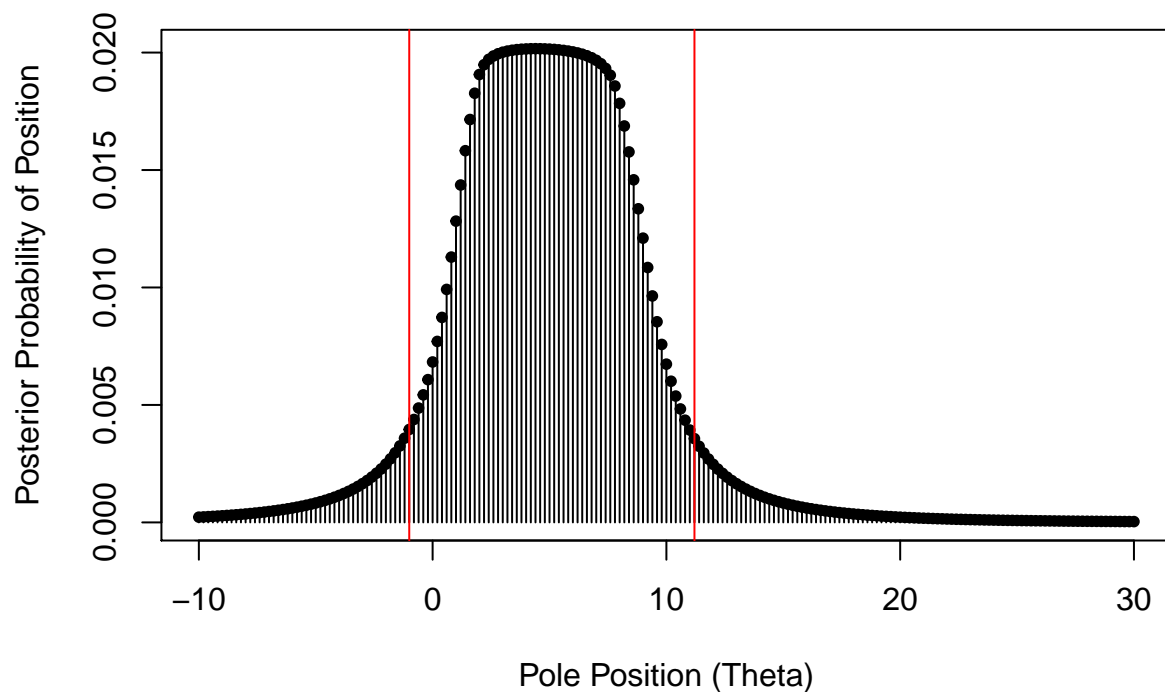
(1d)

```
like_theta <- colSums(mat)

prior <- rep(1/length(thetas), length(thetas))
post <- prior*like_theta / sum(prior*like_theta)

myplot(thetas, post, y_lab="Posterior Probability of Position",
        x_lab="Pole Position (Theta)")

cs <- cumsum(post)
L <- min(thetas[cs > 0.05])
R <- min(thetas[cs >= 0.95])
abline(v=L, col="red")
abline(v=R, col="red")
```



The 5th percentile is at $\Theta = -1$, and the 95th percentile is at $\Theta = 11.2$ (as shown in the red lines above). Qualitatively, this graph is similar to the Cauchy distribution that we observed in class, but with a flatter middle plateau instead of a thin peak.

Problem 2

(2a)

The marginal pmfs are as follows:

$$f_X(x) = \begin{cases} 0.25 & x = 1 \\ 0.5 & x = 2 \\ 0.25 & x = 3 \\ 0 & \text{otherwise} \end{cases}$$

$$f_Y(y) = \begin{cases} 0.5 & y = 0 \\ 0.5 & y = 2 \\ 0 & \text{otherwise} \end{cases}$$

(2b)

The pmf of the random variable XY is:

$$f_{XY}(xy) = \begin{cases} 0.5 & xy = 0 \\ 0.25 & xy = 2 \\ 0.25 & xy = 6 \\ 0 & \text{otherwise} \end{cases}$$

(2c)

$$E(XY) = (0.5) * 0 + (0.25) * 2 + (0.25) * 6 = 2$$

$$E(X) = (0.25) * 1 + (0.5) * 2 + (0.25) * 3 = 2$$

$$E(Y) = (0.5) * 0 + (0.5) * 2 = 1$$

$$\text{Thus, } E(XY) = E(X)E(Y)$$

(2d)

For X and Y to be independent, it must hold that:

$$P(X = x)P(Y = y) = P(X = x, Y = y)$$

For all values of x and y . This can be proven to be false with the counterexample:

$$P(X = 1) = 0.25, \quad P(Y = 2) = 0.5, \quad P(X = 1, Y = 2) = 0.25 \neq P(X = 1)P(Y = 2)$$

Problem 3

(3a)

$$E(X) = \int x f_X(x) dx = \int_0^1 2x^2 = \frac{2}{3} x^3 \Big|_0^1 = \frac{2}{3}$$

(3b)

For $y < 1$, $f_Y(y) = F_Y(y) = 0$. For $y \geq 1$:

$$F_Y(y) = P\{Y \leq y\} = P\{\frac{1}{X} \leq y\} = P\{X \geq \frac{1}{y}\} = 1 - F_X(\frac{1}{y})$$

$$F_X(x) = \int f_X(x)dx = \int 2x \, dx = x^2, \quad F_Y(y) = 1 - \frac{1}{y^2}$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} (1 - F_X(\frac{1}{y})) = -F'_X(\frac{1}{y}) \frac{d}{dy} (\frac{1}{y}) = \frac{1}{y^2} f_X(\frac{1}{y}) = \frac{2}{y^3}$$

(3c)

While x spans from 0 to 1, y spans from 1 to infinity.

$$E(Y) = \int y f_Y(y) dy = \int_1^\infty \frac{2}{y^2} = -\frac{2}{y} \Big|_1^\infty = 2$$

(3d)

According to the LOTUS, if $Y = g(X)$, then:

$$E(\frac{1}{X}) = \int g(x) f_X(x) \, dx = \int \frac{1}{x} 2x \, dx = 2x \Big|_0^1 = 2$$

Problem 4

(4a)

For 2 colors and 4 total socks, then we can model the random variable N (the number of pairs formed) by randomly drawing the first 2 socks as a pair and then leaving the remaining 2 socks as the other pair. Thus, there is a $\frac{1}{3}$ chance that $N = 2$ and a $\frac{2}{3}$ chance that $N = 0$. This is because there are 4 ways to draw a matching pair of socks and leave another pair of socks and 8 ways to draw a mismatched pair and leave another mismatched pair. If we consider the two colors to be white and black, then the 4 ways to draw a matching pair are: $\{w_1 w_2, w_2 w_1, b_1 b_2, b_2 b_1\}$. The 8 ways to draw a mismatched pair are: $\{w_1 b_1, w_1 b_2, w_2 b_1, w_2 b_2, b_1 w_1, b_1 w_2, b_2 w_1, b_2 w_2\}$.

In the case of 2 colors, there is no possibility to only make one pair because that one pair of matched socks by necessity leaves the other two matched socks for the second pair. To differentiate the symbol for the number of colors (n) from the symbol for the individual values of the random variable N , I will call the latter n' . Thus, the functions can be expressed:

$$f_N(n') = \begin{cases} \frac{1}{3} & n' = 2 \\ \frac{2}{3} & n' = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$E(N) = \sum_{n'} n' f_N(n') = \frac{2}{3}$$

(4b)

N does **not** have a binomial distribution, because this graph will have more than one peak while a binomial distribution will only have one. Qualitatively, forming one pair is not independent from forming a second pair, and binomial trials are independent. This situation features selection without replacement, so each trial of drawing two socks affects more than just that pair.

(4c)

If we define indicator variables such that $N = I_1 + I_2 + \dots + I_n$, then:

$$I_j = \begin{cases} 1 & \text{if pair } j \text{ is matched} \\ 0 & \text{otherwise} \end{cases}$$

Each I_j has distribution:

$$I_j = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

The indicator variables here are not independent of each other, because drawing one pair theoretically changes the probability of drawing the next pair, but this doesn't necessarily matter because of symmetries in this problem. Here, the probability of a given pair being matched is just the probability of drawing the second sock of the pair from the full pool of socks minus the first sock. Because the expectation value is just the individual probability summed over across all indicators, this can be used to find the final expectation value:

$$p = \frac{1}{2n-1}, \quad E(N) = \frac{n}{2n-1}$$

Problem 5

(5a)

We can begin with the equation $F_X(x) = P\{X \leq x\}$. From here, we want the probability that the median of T_1 , T_2 , and T_3 is $\leq x$. Let N be the number of random variables among $\{T_1, T_2, T_3\}$ that are $\leq x$. X definitely $\leq x$ when $N = 3$, 2, and definitely not $\leq x$ when $N = 1$, 0. This is to say that the event $X \leq x$ is equivalent to the event $N \geq 2$. N has a binomial distribution, because each random variable in $\{T_1, T_2, T_3\}$ can either be greater than or equal to 2 with probability p or less than 2 with probability $1 - p$. In this way, p can be considered the success probability for each of the random variables. In this way, $P\{X \leq x\} = P\{N \geq 2\}$. To then find probability that $N = 2$ or 3, we can use the Binomial distribution. Given that these random variables are drawn from an exponential distribution with $\lambda = 1$, we know:

For exponential distributions, $P\{X \geq x\} = 1 - e^{-\lambda x}$. Here, $p = 1 - e^{-x}$

$$P\{N = 2\} = \binom{3}{2}p^2(1-p)^1 = \binom{3}{2}(1 - e^{-x})^2(1 - (1 - e^{-x}))^1 = 3e^{-x}(1 - e^{-x})^2$$

$$P\{N = 3\} = \binom{3}{3}p^3(1-p)^0 = \binom{3}{3}(1 - e^{-x})^3(1) = (1 - e^{-x})^3$$

$$F_X(x) = P\{N = 2\} + P\{N = 3\} = 3e^{-x}(1 - e^{-x})^2 + (1 - e^{-x})^3$$

Thus, the full CDF is:

$$F_X(x) = \begin{cases} 1 - 3e^{-2x} + 2e^{-3x} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

(5b)

To get the pdf of X , we can differentiate with respect to x .

$$\begin{aligned} f_X(x) &= \frac{d}{dx} F_X(x) = -3e^{-3x}(e^x - 3)(e^x - 1) + 3e^{-3x}(e^x - 1)^2 \\ &= 3e^{-3x}((e^x - 1)^2 - (e^x - 3)(e^x - 1)) \\ &= 3e^{-3x}(2e^x - 2) = 6e^{-2x} - 6e^{-3x} \end{aligned}$$

We know $f_X(0) = 0$ because the limit is 0 from both sides. Otherwise:

$$f_X(x) = \begin{cases} 6e^{-2x}(1 - e^{-x}) & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

(5c)

Given the integral $\int_0^\infty u^k e^{-u} du = k!$, we can find the expectation value:

$$\begin{aligned} E(X) &= \int_0^\infty x f_X(x) dx = \int_0^\infty 6e^{-2x}x - 6e^{-3x}x dx \\ &= \int_0^\infty 6e^{-2x}x dx - \int_0^\infty 6e^{-3x}x dx = \frac{3}{2} - \frac{2}{3} = \frac{5}{6} \end{aligned}$$

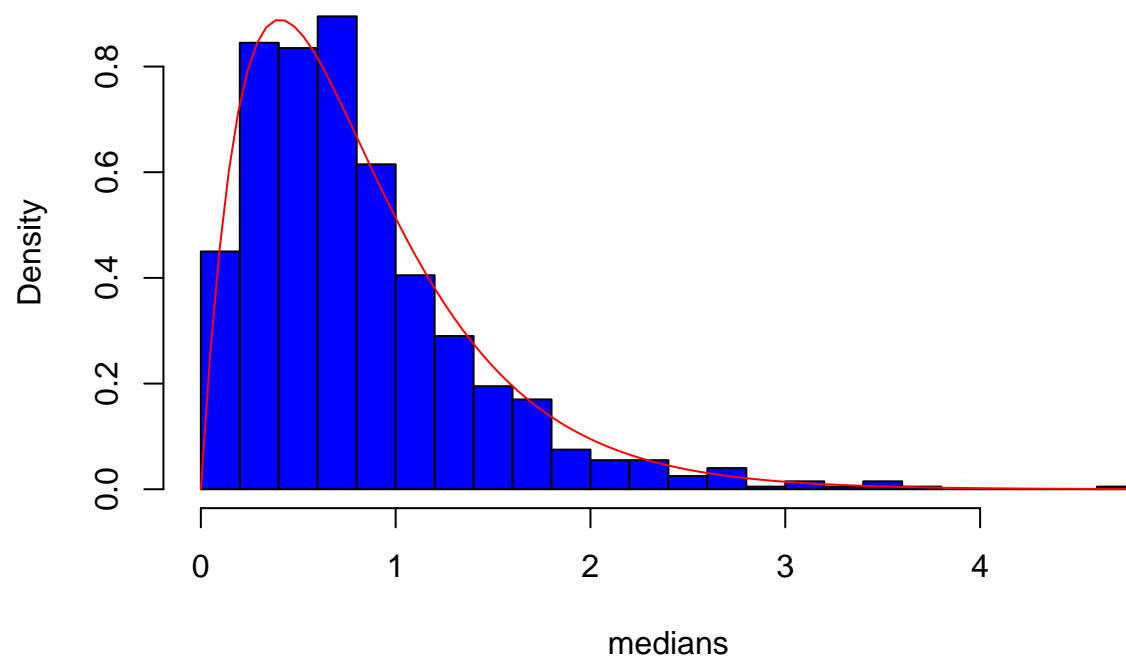
(5d)

```
t1 <- rexp(1000)
t2 <- rexp(1000)
t3 <- rexp(1000)
medians <- rep(NA, 1000)
for (i in 1:1000) {
  selection <- c(t1[i], t2[i], t3[i])
  medians[i] <- median(selection)
}

# Proving B
fx <- function(x) {
  return(6*exp(-2*x) - 6*exp(-3*x))
}

hist(medians, breaks=20, col="blue", freq=FALSE)
curve(fx, add=TRUE, col="red")
```

Histogram of medians



```
# Proving C  
5/6
```

```
## [1] 0.8333333
```

```
mean(medians)
```

```
## [1] 0.8067508
```