Stat 111: Homework 7

Daniel de Castro

April 1, 2022

1. (a) $\hat{\mu} = 124.327$

   (b) Chapter 3 of the Stat 111 textbook gives the following result:

$$\sqrt{n}\left(Y_{(\lceil np \rceil} - Q_{Y_1}(p)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{p(1-p)}{(f_{Y_1}(Q_{Y_1}(p))^2}\right)$$

Thus,

$$\sqrt{n}\left(Y_{(\lceil n/2 \rceil)} - Q_{Y_1}(0.5)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{0.25}{f_{Y_1}(Q_{Y_1}(0.5))^2}\right)$$

$$\sqrt{n}\frac{f_{Y_1}(Q_{Y_1}(0.5))}{0.5}\left(Y_{(\lceil n/2 \rceil)} - Q_{Y_1}(0.5)\right) \xrightarrow{d} \mathcal{N}(0,1)$$

In this way,

$$P\left(-1.96 \leq \sqrt{n}\frac{f_{Y_1}(Q_{Y_1}(0.5))}{0.5}\left(Y_{(\lceil n/2 \rceil)} - Q_{Y_1}(0.5)\right) \leq 1.96\right) = 0.95$$

$$P\left(-\frac{0.98}{\sqrt{n}f_{Y_1}(Q_{Y_1}(0.5))} \leq Y_{(\lceil n/2 \rceil)} - Q_{Y_1}(0.5) \leq \frac{0.98}{\sqrt{n}f_{Y_1}(Q_{Y_1}(0.5))}\right) = 0.95$$

$$P\left(Y_{(\lceil n/2 \rceil)} - \frac{0.98}{\sqrt{n}f_{Y_1}(Q_{Y_1}(0.5))} \leq Q_{Y_1}(0.5) \leq Y_{(\lceil n/2 \rceil)} + \frac{0.98}{\sqrt{n}f_{Y_1}(Q_{Y_1}(0.5))}\right) = 0.95$$

This yields

$$C(\boldsymbol{Y}) = \left[Y_{(\lceil n/2 \rceil)} - \frac{0.98}{\sqrt{n}f_{Y_1}(Q_{Y_1}(0.5))}, Y_{(\lceil n/2 \rceil)} + \frac{0.98}{\sqrt{n}f_{Y_1}(Q_{Y_1}(0.5))}\right]$$

Given that $\hat{\mu} = Y_{(\lceil n/2 \rceil)}$ and $n = 83$,

$$C(\boldsymbol{Y}) = \left[\hat{\mu} - \frac{0.98}{\sqrt{83}f_{Y_1}(\hat{\mu})}, \hat{\mu} + \frac{0.98}{\sqrt{83}f_{Y_1}(\hat{\mu})}\right]$$

We estimate $f_{Y_1}(Q_{Y_1}(0.5))$ and calculate $C(\boldsymbol{Y})$ using the following code:

```python
import numpy as np
import pandas as pd
import scipy.stats as stats

jill = pd.read_csv("jill.csv")
mu_hat = np.median(jill.Y)
f = stats.gaussian_kde(jill.Y)
(mu_hat - 1.96 * 0.5 / (np.sqrt(83) * f(mu_hat)[0]), mu_hat + 1.96 * 0.5 /
↪   (np.sqrt(83) * f(mu_hat)[0]))
```

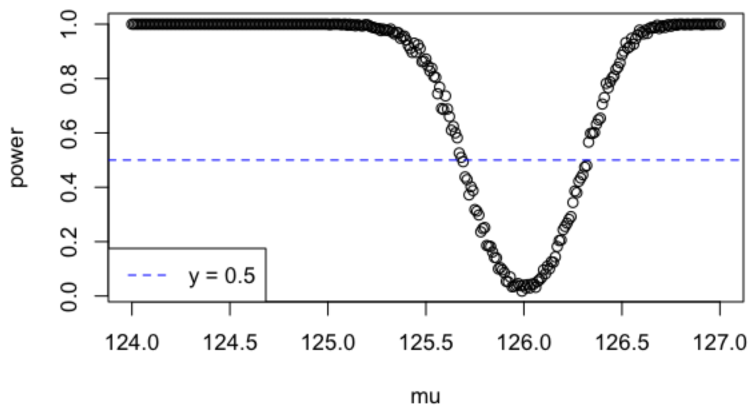
This returned

$$C(\boldsymbol{Y}) = [123.58136288231937, 125.07253905316664]$$

(c) Let $H_0 : \mu = \mu_0 = 126$ and $H_1 : \mu \neq \mu_0$. Let $C(\boldsymbol{Y})$ be a 95% confidence interval for μ based on $\hat{\mu}$. Then by Stat 111 Theorem 8.5.2, we can construct an $\alpha = 0.05$ sized test by retaining the null hypothesis if $\mu_0 \in C(\boldsymbol{Y})$.

(d) $\mu_0 = 126$ is not within the 95% confidence interval given by part (b). Thus, we would reject the null hypothesis.

(e)
```
library(stats)

jill <- read.csv("jill.csv")
n <- length(jill$Y)

mu <- c(12400:12700/100)
mu_0 <- 126
B <- 500

power <- c()
for (m in mu) {
  reject <- c()
  for (i in 1:B) {
    U <- rnorm(n)
    Y <- m * (1 + exp(-4 + 0.5 * U)) / (1 + exp(-4))
    mu_hat <- median(Y)

    dx <- density(Y)
    fden <- approx(dx$x, dx$y, xout=m)$y

    SE <- sqrt(1 / (4 * n * fden ** 2))

    lower <- mu_hat - 1.96 * SE
    upper <- mu_hat + 1.96 * SE

    reject <- append(reject, as.integer((mu_0 < lower)|(mu_0 > upper)))
  }
  power <- append(power, mean(reject))
}

png(file="1e.png", width=6, height=4, units="in", res=100)
plot(mu, power)
abline(h = 0.5, col="blue", lty=2)
legend("bottomleft", legend=c("y = 0.5"), col=c("blue"), lty=c(2))
dev.off()
```



(f)
```
for (i in 1:length(power)) {
  if (power[i] < 0.5) {
    print(i)
    print(mu[i])
    print(power[i])
    break;
  }
}

for (i in 170:length(power)) {
  if (power[i] > 0.5) {
    print(i)
    print(mu[i])
    print(power[i])
    break;
  }
}
```

The *approximate* range of μ for which this test has reasonable power (above 0.5) is $\mu \notin [125.69, 126.33]$

2. (a) Under the null hypothesis, $\theta = \theta_0$, so

$$\sqrt{n}\left(\hat{\theta}_{\text{MoM}} - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, \text{Var}(h(Y_1)))$$

$$\hat{\theta}_{\text{MoM}} \xrightarrow{d} \mathcal{N}\left(\theta_0, \frac{\text{Var}(h(Y_1)))}{n}\right)$$

(b) Given the result from part (a),

$$T(\boldsymbol{Y}) = \frac{\sqrt{n}\left(\hat{\theta}_{\text{MoM}} - \theta_0\right)}{\text{Var}(h(Y_1)))} \xrightarrow{d} \mathcal{N}(0, 1)$$

We could thus construct an $\alpha = 0.01$ sized test by rejecting the null hypothesis if $|T(\boldsymbol{Y})| > Q_{\mathcal{N}(0,1)}(0.995)$. This is to say, we reject the null hypothesis if $T(\boldsymbol{Y}) \notin \left[Q_{\mathcal{N}(0,1)}(0.005), Q_{\mathcal{N}(0,1)}(0.995)\right] \approx [-2.576, 2.576]$, and accept the null hypothesis otherwise.

(c) It holds that

$$\frac{\sqrt{n}\left(\hat{\theta}_{\text{MoM}} - \theta_0\right)}{\text{Var}(h(Y_1)))} - \frac{\sqrt{n}\,(\theta - \theta_0)}{\text{Var}(h(Y_1)))} \xrightarrow{d} \mathcal{N}(0, 1)$$

because

$$\frac{\sqrt{n}\left(\hat{\theta}_{\text{MoM}} - \theta_0\right)}{\text{Var}(h(Y_1)))} - \frac{\sqrt{n}\,(\theta - \theta_0)}{\text{Var}(h(Y_1)))} = \frac{\sqrt{n}\left(\hat{\theta}_{\text{MoM}} - \theta\right)}{\text{Var}(h(Y_1)))}$$

which is asymptotically distributed $\mathcal{N}(0,1)$ by the central limit theorem. Let

$$A = \left[Q_{\mathcal{N}(0,1)}(0.005), Q_{\mathcal{N}(0,1)}(0.995)\right] \approx [-2.576, 2.576]$$

Then the power function $\beta(\theta)$ is such that

$$\beta(\theta) = P(T(\boldsymbol{Y}) \notin A)$$

$$\beta(\theta) = P(T(\boldsymbol{Y}) < Q_{\mathcal{N}(0,1)}(0.005)) + 1 - P(T(\boldsymbol{Y}) < Q_{\mathcal{N}(0,1)}(0.995))$$

$$\beta(\theta) = \Phi\left(Q_{\mathcal{N}(0,1)}(0.005)) - \frac{\sqrt{n}\,(\theta - \theta_0)}{\text{Var}(h(Y_1)))}\right) + 1 - \Phi\left(Q_{\mathcal{N}(0,1)}(0.995)) - \frac{\sqrt{n}\,(\theta - \theta_0)}{\text{Var}(h(Y_1)))}\right)$$

where Φ is the CDF of the $\mathcal{N}(0,1)$ distribution. This limit is useful in that it allows us to calculate $\beta(\theta)$ using the CDF of the normal distribution.

3. (a) $\gamma = [\text{Corr}(X,Y)]^2$

   (b) i. $\gamma \le 1$, because $\text{Corr}(X,Y) \in [-1,1]$ for any random variables X, Y. In fact, we can go a step further and say that $\gamma \in [0,1]$.

   ii.

$$-1 \le \text{Corr}(X,Y) \le 1$$
$$0 \le \text{Corr}(X,Y)^2 \le 1$$
$$0 \le \beta_{Y\sim X} \times \beta_{X\sim Y} \le 1$$
$$0 \le |\beta_{Y\sim X} \times \beta_{X\sim Y}| \le 1$$
$$|\beta_{Y\sim X}| \times |\beta_{X\sim Y}| \le 1$$
$$|\beta_{Y\sim X}| \le \frac{1}{|\beta_{X\sim Y}|}$$

   iii. Consider that

$$\beta_{Y\sim X} = \frac{\text{Cov}(X,Y)}{\text{Var}(X)}$$

and

$$\beta_{X\sim Y} = \frac{\text{Cov}(X,Y)}{\text{Var}(Y)}$$

Since $\text{Var}(Y) \ge 0$ and $\text{Var}(X) \ge 0$, we know that $\text{sign}\left(\beta_{Y\sim X}\right) = \text{sign}\left(\text{Cov}(X,Y)\right)$ and $\text{sign}\left(\beta_{X\sim Y}\right) = \text{sign}\left(\text{Cov}(X,Y)\right)$. Thus,

$$\text{sign}\left(\beta_{Y\sim X}\right) = \text{sign}\left(\beta_{X\sim Y}\right)$$

4. (a) i. According to Adam's Law,

$$E\left[(Y - bX)^2\right] = E\left[E\left[(Y - bX)^2|X = x\right]\right]$$

$$E\left[(Y - bX)^2\right] = E\left[E\left[Y^2 - 2bXY - b^2X^2|X = x\right]\right]$$

$$E\left[(Y - bX)^2\right] = E\left[E\left[Y^2|X = x\right] - 2bXE\left[Y|X = x\right] - b^2X^2\right]$$

$$E\left[(Y - bX)^2\right] = E\left[E\left[Y^2|X = x\right] - 2bX\mu(X) - b^2X^2\right]$$

Similarly,

$$E[(Y - \mu(X))^2] + E[(\mu(X) - bX)^2] = E\left[E[(Y - \mu(X))^2 + (\mu(X) - bX)^2|X = x]\right]$$

$$E[(Y-\mu(X))^2]+E[(\mu(X)-bX)^2] = E\left[E[Y^2 - 2Y\mu(X) + 2\mu(X)^2 - 2bX\mu(X) + b^2X^2|X=x]\right]$$
$$E[(Y-\mu(X))^2]+E[(\mu(X)-bX)^2] = E\left[E[Y^2|X=x] - 2\mu(X)^2 + 2\mu(X)^2 - 2bX\mu(X) + b^2X^2\right]$$
$$E[(Y-\mu(X))^2]+E[(\mu(X)-bX)^2] = E\left[E[Y^2|X=x] - 2bX\mu(X) + b^2X^2\right]$$

Thus,

$$E\left[(Y-bX)^2\right] = E[(Y-\mu(X))^2] + E[(\mu(X)-bX)^2]$$

ii. This inequality is indeed true. We can contextualize this inequality in an attempt to predict Y given the information that $X = x$. If we let t be our prediction for Y, we can measure the success of this prediction with the value of $E[(Y-t)^2|X=x]$. Intuitively, we know that $t = \mu(X)$ would be the prediction that would yield the lowest $E[(Y-t)^2|X=x]$. This implies that if we were to use any other value for t, then $E[(Y-t)^2|X=x] \geq E[(Y-\mu(X))^2|X=x]$. This includes the case where $t = bX$. Thus, $E[(Y-bX)^2|X=x] \geq E[(Y-\mu(X))^2|X=x]$.

(b)

$$E[(\mu(X)-bX)^2] = E[\mu(X)^2] - 2bE[\mu(X)X] + b^2E[X^2]$$
$$\frac{\partial}{\partial b}E[(\mu(X)-bX)^2] = -2E[\mu(X)X] + 2bE[X^2] = 0$$
$$\theta = \frac{E[\mu(X)X]}{E[X^2]}$$

To verify that this is where $E[(\mu(X)-bX)^2]$ reaches a minimum, we conduct the second derivative test:

$$\frac{\partial^2}{\partial b^2}E[(\mu(X)-bX)^2] = 2E[X^2] \geq 0$$

$E[(\mu(X)-bX)^2]$ is thus convex, making this a minimum.

(c) i.

$$\theta = \frac{E[\mu(X)X]}{E[X^2]}$$
$$\theta = \frac{E[X(dX + cX^2 - cE[X^2])]}{E[X^2]} = \frac{E[dX^2 + cX^3 - cXE[X^2]]}{E[X^2]}$$
$$\theta = \frac{dE[X^2] + cE[X^3] - cE[X]E[X^2]}{E[X^2]}$$
$$\theta = \frac{2d}{2} = d$$

ii. Below is a plot of $\mu(x)$ and θx versus x over the interval $[-3, 3]$.