Theoretical Questions

Mohammad Mahdi Abdollahpour - 9531057

1.

$$\hat{F}_n(x) = \sum_{i=1}^{n} \frac{I(X_i \le x)}{n} = \sum_{i=1}^{n} \frac{Y_i}{n}$$

$$Y_i \sim Bernouli\big(F(x)\big)$$
$$E\left(\hat{F}_n(x)\right) = \frac{1}{n}\big(n * E(Y_i)\big) = F(x)$$
$$V\left(\hat{F}_n(x)\right) = \frac{1}{n^2}\big(n * V(Y_i)\big) = \frac{\Big(F(x) * \big(1 - F(x)\big)\Big)}{n}$$
$$MSE = bias^2(\hat{F}_n(x)) + V(\hat{F}_n(x))$$
$$bias = E\left(\hat{F}_n(x)\right) - F(x) = 0, So\ MSE$$
$$= V\left(\hat{F}_n(x)\right), V\left(\hat{F}_n(x)\right)\ goes\ to\ 0\ as\ n\ goes\ to\ \infty, so\ MSE\ goes\ to\ zero\ as\ n\ goes\ to\ \infty$$
$$MSE = E\left(\hat{F}_n(x) - F(x)\right)^2, so\ if\ MSE \to 0, then\ \hat{F}_n(x) \to (qm)\ F(x), so\ \hat{F}_n(x) \to (p)\ F(x)$$

2.

$$\hat{P}_n = \frac{1}{n} * \sum_{i=1}^{n} X_i\ , se = \sqrt{V(\hat{P}_n)} = \sqrt{\left(\hat{P}_n * \frac{(1 - \hat{P}_n)}{n}\right)}$$
$$\hat{P}_n \pm z_{0.05} * se = \hat{P}_n \pm 2 * se$$
$$\hat{q}_n = \frac{1}{n} * \sum_{i=1}^{n} Y_i$$
$$estimator\ for\ p - q :\ \hat{p}_n - \hat{q}_n$$
$$se = \sqrt{V(\hat{p}_n - \hat{q}_n)} = \sqrt{\frac{(\hat{p}_n - \hat{q}_n)\big(1 - (\hat{p}_n + \hat{q}_n)\big)}{n}}$$
$$90\ confidence\ interval :\ (\hat{p}_n - \hat{q}_n) \pm 2 * \sqrt{\frac{(\hat{p}_n - \hat{q}_n)\big(1 - (\hat{p}_n + \hat{q}_n)\big)}{n}}$$

3.

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} X_i$$
$$by\ CLT, \frac{\sqrt{n}(\bar{F}_n(x) - \mu)}{V(\bar{F}_n(x))} \to Z$$
$$\lim_{n\to\infty}\big(\bar{F}_n(x) - F(x)\big) = N(0, \frac{F(x)\big(1 - F(x)\big)}{2})$$

4.

$$se(\hat{\theta}) = \sqrt{V(\hat{\theta})} = \sqrt{V\left(\widehat{F_n}(b)\right) + V\left(\widehat{F_n}(a)\right) - 2 * COV\left(\widehat{F_n}(b), \widehat{F_n}(a)\right)}$$

$$computing\ COV\left(\widehat{F_n}(b), \widehat{F_n}(a)\right)\ as\ done\ in\ excercise\ 5\ of\ chapter\ 8$$

$$V\left(\widehat{F_n}(b)\right) = \widehat{F_n}(b) * \frac{\left(1 - \widehat{F_n}(b)\right)}{n}$$

$$V\left(\widehat{F_n}(a)\right) = \widehat{F_n}(a) * \frac{\left(1 - \widehat{F_n}(a)\right)}{n}$$

Confidence interval will be $(\widehat{F_n}(b) - \widehat{F_n}(a)) \pm z_{\frac{\alpha}{2}} * se$

5.

$$CDF(\theta) = \prod_i P(X_i \leq c\ ) = P(X_i \leq c)^n = \left(\frac{c}{\theta}\right)^n$$

$$PDF(\theta) = n\ c^{n-1}\ , \theta = 1$$

$$\hat{\theta} \sim Beta(n, 1)$$

$$in\ parametric\ bootstrap\ \hat{\theta} \sim Unif\left(0, \hat{\theta}\right)\ so\ P\left(\hat{\theta} = c\right) = 0$$

$$in\ nonparametric\ we\ have\ \hat{\theta} = \widehat{\theta^*}\ any\ time\ X_i s\ include\ the\ max,$$

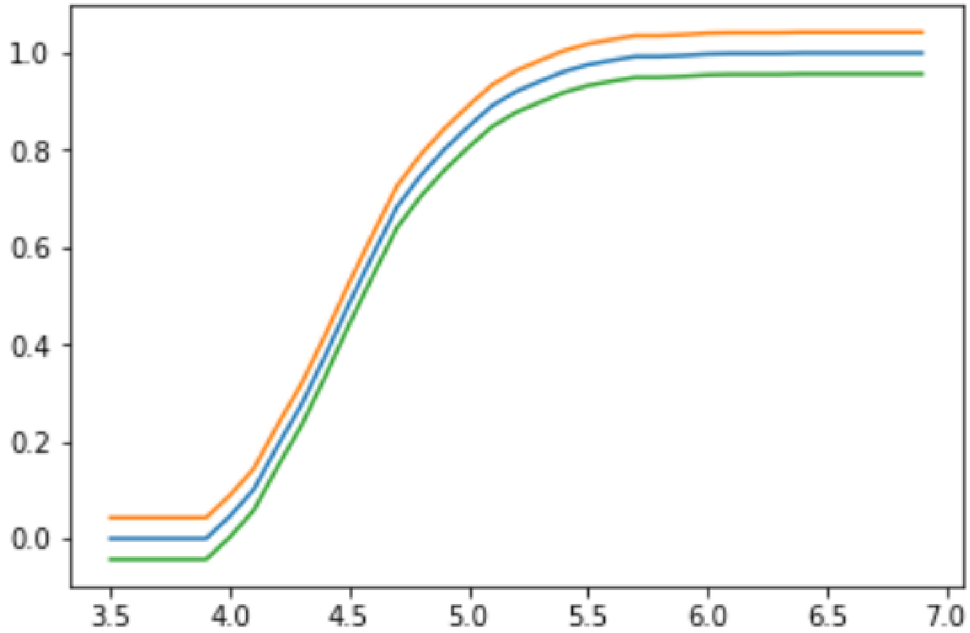
$$so\ probability\ of\ omitting\ the\ \max is\ \left(1 - \left(\frac{1}{n}\right)\right)^n$$

$$= e^{-1}\ , chance\ of\ having\ the\ \max is\ 1 - e^{-1} = 0.63$$

Codes Report

Question 1

A and b



Blue line is estimated CDF and orange and green lines are 95 percent confidence envelope for F.

C.

Point estimated would be $\hat{F}(4.9) - \hat{F}(4.3)$

And standard error is $se = \sqrt{V\big(F(4.9) - F(4.3)\big)} = \sqrt{V\left(\hat{F}(4.9)\right) + V\left(\hat{F}(4.9)\right)}$

The interval is $\hat{F}(4.9) - \hat{F}(4.3) \pm 2 * se$

95%

-0.584178140300015 , 1.636178140300015
93%
-0.5044730430477062 , 1.5564730430477063

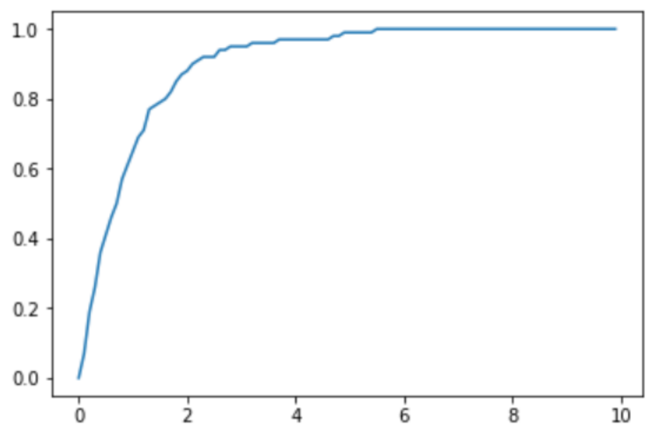
Question 2

Estimated Lambda 1.006595871904141

97% confidence interval (0.9349665290299697 , 1.0782252147783125)
the interval contains the param 0.975 % of times

Question 3

Estimated lambda 1.0212193328795964



Empirical Distribution Function

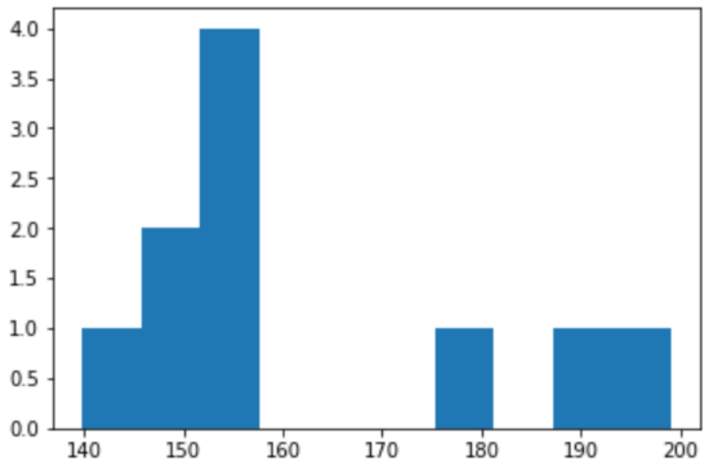
Plugin Estimators
Mean: 0.989556249796454
Sigma: 1.0241083502783939
the other way (defined in course book): 1.029267618952513
skewness: 2.1584672388251156

Question 4

A and B

vboot 179.73283867041596
se 13.406447652917455
95% confidence interval(103.36720117089159 , 155.92047597032803)

C



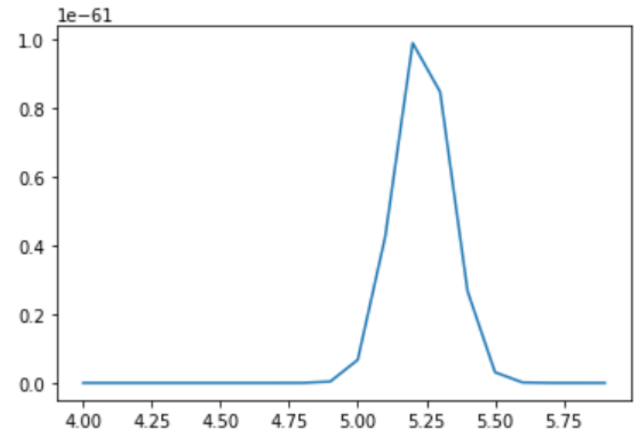
Histogram of Bootstrap Replications

And the True sampling distribution of $\hat{\theta}$ is N(~148,se) which quite looks like the histogram

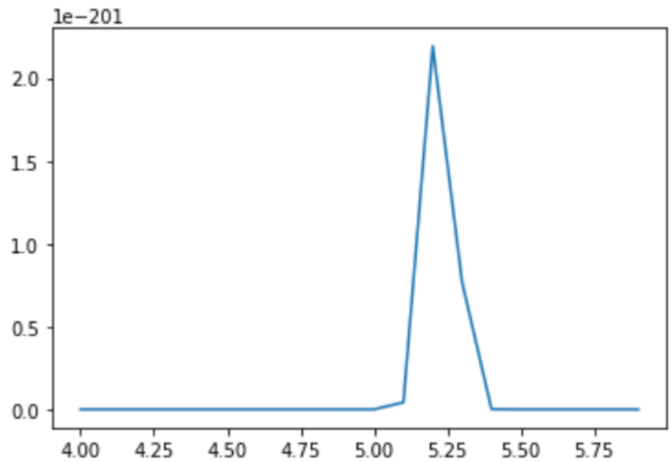
Question 5

A and B

Plot of Posterior



Plot of simualted posterior



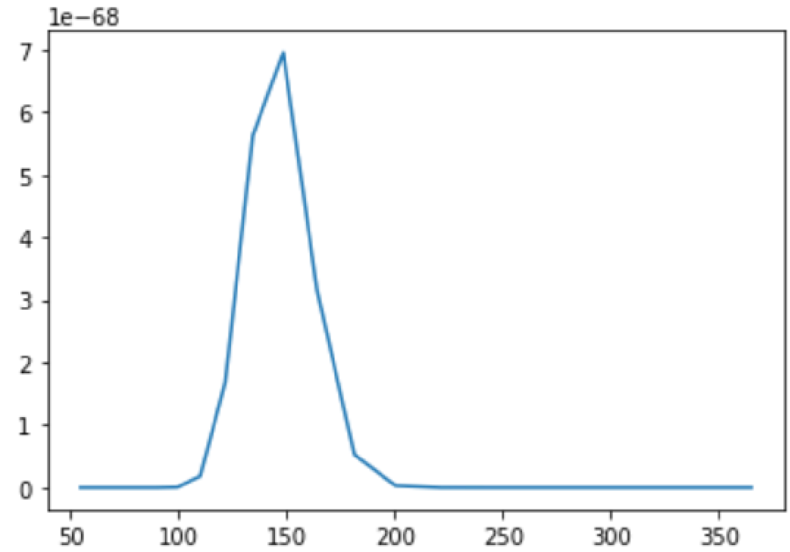
D

Analytically

$CDF(\theta) = \int_A f(\mu|x^n)\, d\mu, A = \{\mu: e^\mu \leq \theta\}$

By simulation

Using the guide on page 212 of course book we have:



E

```
97% confidence interval ( -61.88483753233056 , 76.84934247091029 )
93% confidence interval ( -50.37453361005844 , 65.33903854863817 )
```

Question 6

In 0.0637 of times we reject the null, which is the error rate for type-1 error and is close to 0.05

Question 7
A
Delta 0.022175
97% confidence interval (-0.013458074389523171 , 0.05780807438952317)
P-Value 0.17686214169208325

Since both negative and positive values are in confidence interval, we can not judg e to which author it belongs to.

B

Based on permutation test, P-value is 0.00062
So, the two distributions are the same.

Question 8

test accuracy 0.72
test confusion matrix
 [[31. 14. 0.]
 [9. 30. 3.]
 [10. 6. 47.]]

train accuracy 0.78
train confusion matrix
 [[33. 6. 1.]
 [10. 38. 3.]
 [7. 6. 46.]]

Question 9

theta_hat -0.07130609590256017
se: 0.10018298605515501
95% confidence interval (-0.2716720680128702 , 0.12905987620774984)