STA2101 Assignment 2*

Ahmed Nawaz Amanullah †

November 2020

1

1.1 Problem

Let $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ $\mathbf{B} = \begin{pmatrix} 0 & 2 \\ 2 & 1 \end{pmatrix}$ $\mathbf{C} = \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix}$

1.1.1 (a)

Calculate $\mathbf{AB}$ and $\mathbf{AC}$

1.1.2 (b)

Do we have $\mathbf{AB} = \mathbf{AC}$? Answer Yes or No.

1.1.3 (c)

Prove $\mathbf{B} = \mathbf{C}$. Show your work.

---

†with help from the Overleaf team

1.2 Solution

1.2.1 (a)

$$\mathbf{AB} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 0 & 2 \\ 2 & 1 \end{pmatrix} \tag{1}$$

$$\mathbf{AB} = \begin{pmatrix} 4 & 4 \\ 8 & 8 \end{pmatrix} \tag{2}$$

$$\mathbf{AC} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix} \tag{3}$$

$$\mathbf{AC} = \begin{pmatrix} 4 & 4 \\ 8 & 8 \end{pmatrix} \tag{4}$$

1.2.2 (b)

Yes.

1.2.3 (c)

From (b), we have $\mathbf{AB} = \mathbf{AC}$, from which we get,

$$\mathbf{AB} = \mathbf{AC} \tag{5}$$

$$\mathbf{A}^{-1}\mathbf{AB} = \mathbf{A}^{-1}\mathbf{AC} \tag{6}$$

$$\mathbf{B} = \mathbf{C} \tag{7}$$

N.B. The above proof assumes the inverse of $\mathbf{A}$ exists, which doesn't (i.e. determinant of the matrix is 0). Actually, we can just observe $\mathbf{B} \neq \mathbf{C}$ from the question, so my humble opinion is that the purpose of this question is to point out the inverse of $\mathbf{A}$ does not exist, so we can't pull off the proof we disclosed above.

2

2.1 Problem

Let $\mathbf{X}$ be an n by p matrix with $n \neq p$. Why is it incorrect to say that $(\mathbf{X}^T\mathbf{X})^{-1} = \mathbf{X}^{-1}\mathbf{X}^{T^{-1}}$?

2.2 Solution

It is incorrect as according to the question $\mathbf{X}$ is non-square and a non-square matrix does not have an inverse.

3

3.1 Problem

Let $\mathbf{a}$ be an $n \times 1$ matrix of real constants. How do you know $\mathbf{a}^T\mathbf{a} \geq 0$?

3.2 Solution

$$\mathbf{a}^T\mathbf{a} = \begin{pmatrix} a_1 & \dots & a_n \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \tag{8}$$

$$\mathbf{a}^T\mathbf{a} = a_1^2 + ... + a_n^2 \geq 0 \tag{9}$$

With the last line being a consequence of $a_i^2 \geq 0$ for $i = 1, ..., n$.

4

4.1 Problem

The $p \times p$ matrix $\mathbf{\Sigma}$ is said to be positive definite if $\mathbf{a}^T\mathbf{\Sigma}\mathbf{a} > 0$ for all $p \times 1$ vectors $\mathbf{a} \neq \mathbf{0}$. Show that the eigenvalues of a positive definite matrix are all strictly positive. A good approach is to start with the definition of an eigenvalue and the corresponding eigenvalue (presumably this is a typing error and question meant corresponding *eigenvector*): $\mathbf{\Sigma}\mathbf{v} = \lambda\mathbf{v}$. Eigenvectors are typically scaled to have length one, so you may assume $\mathbf{v}^T\mathbf{v} = 1$.

4.2 Solution

Take any eigenvalue λ and its corresponding eigenvector $\mathbf{v}$. We can then use the provided definition of the positive definite matrix to arrive at our desired result,

$$\boldsymbol{\Sigma}\mathbf{v} = \lambda\mathbf{v} \tag{10}$$
$$\mathbf{v}^T\boldsymbol{\Sigma}\mathbf{v} = \lambda\mathbf{v}^T\mathbf{v} \tag{11}$$
$$\lambda = \mathbf{v}^T\boldsymbol{\Sigma}\mathbf{v} > 0 \tag{12}$$

5

5.1 Problem

Recall the *spectral decomposition* of a symmetric matrix (for example, a variance-covariance matrix). Any such matrix $\boldsymbol{\Sigma}$ can be written as $\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^T$, where $\mathbf{P}$ is a matrix whose columns are the (orthonormal) eigenvectors of $\boldsymbol{\Sigma}$, $\boldsymbol{\Lambda}$ is a diagonal matrix of the corresponding eigenvalues, and $\mathbf{P}^T\mathbf{P} = \mathbf{P}\mathbf{P}^T = \mathbf{I}$. If $\boldsymbol{\Sigma}$ is real, the eigenvalues are real as well.

5.1.1 (a)

Let $\boldsymbol{\Sigma}$ be a square symmetric matrix with eigenvalues that are all strictly positive.

i. What is $\boldsymbol{\Lambda}^{-1}$?

ii. Show $\boldsymbol{\Sigma}^{-1} = \mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}^T$

5.1.2 (b)

Let $\boldsymbol{\Sigma}$ be a square symmetric matrix, and this time the eigenvalues are non-negative.

i. What do you think $\boldsymbol{\Lambda}^{1/2}$ might be?

ii. Define $\boldsymbol{\Sigma}^{1/2}$ as $\mathbf{P}\boldsymbol{\Lambda}^{1/2}\mathbf{P}^T$. Show $\boldsymbol{\Sigma}^{1/2}$ is symmetric.

iii. Show $\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma}$, justifying the notation.

5.1.3 (c)

Now return to the situation where the eigenvalues of the square symmetric matrix $\mathbf{\Sigma}$ are all strictly positive. Define $\Sigma^{-1/2}$ as $\mathbf{P\Lambda}^{-1/2}\mathbf{P}^T$, where the elements of the diagonal matrix $\mathbf{\Lambda}^{-1/2}$ are the reciprocals of the corresponding elements of $\mathbf{\Lambda}^{1/2}$.

i. Show that the inverse of $\mathbf{\Sigma}^{1/2}$ is $\mathbf{\Sigma}^{-1/2}$, justifying the notation.

ii. Show $\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{-1/2} = \mathbf{\Sigma}^{-1}$.

5.1.4 (d)

Let $\mathbf{\Sigma}$ be a symmetric, positive definite matrix. How do you know that $\mathbf{\Sigma}^{-1}$ exists?

5.2 Solution

5.2.1 (a)

i. $\mathbf{\Lambda}^{-1}$ is a diagonal matrix of the reciprocals of the corresponding eigenvalues.

ii.

$$\mathbf{\Sigma}^{-1} = (\mathbf{P\Lambda P}^T)^{-1} \tag{13}$$
$$\mathbf{\Sigma}^{-1} = (\mathbf{\Lambda P}^T)^{-1}\mathbf{P}^{-1} \tag{14}$$
$$\mathbf{\Sigma}^{-1} = \left(\mathbf{P}^T\right)^{-1}\mathbf{\Lambda}^{-1}\mathbf{P}^{-1} \tag{15}$$
$$\mathbf{\Sigma}^{-1} = \mathbf{P\Lambda}^{-1}\mathbf{P}^T \tag{16}$$

N.B.

- The last line of the above proof follows from the provided information that $\mathbf{P}^T\mathbf{P} = \mathbf{PP}^T = \mathbf{I}$.
- In our proof we used the formula $(AB)^{-1} = B^{-1}A^{-1}$.

5.2.2 (b)

i. I think $\mathbf{\Lambda}^{1/2}$ might be a diagonal matrix of the square root of the corresponding eigenvalues.

ii. We need to show $(\mathbf{\Sigma}^{1/2})^T = \mathbf{\Sigma}^{1/2}$ to show that $\mathbf{\Sigma}^{1/2}$ is symmetric, which we can do as follows,

$$(\mathbf{\Sigma}^{1/2})^T = (\mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}^T)^T \tag{17}$$
$$(\mathbf{\Sigma}^{1/2})^T = (\mathbf{P}^T)^T(\mathbf{P}\mathbf{\Lambda}^{1/2})^T \tag{18}$$
$$(\mathbf{\Sigma}^{1/2})^T = \mathbf{P}(\mathbf{\Lambda}^{1/2})^T\mathbf{P}^T \tag{19}$$
$$(\mathbf{\Sigma}^{1/2})^T = \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}^T = \mathbf{\Sigma}^{1/2} \tag{20}$$

N.B.

- The last line of the above proof follows from the assumption that $\mathbf{\Lambda}^{1/2}$ is a diagonal matrix, which is trivially symmetric.
- In our proof we used the formula $(AB)^T = B^T A^T$.

iii.

$$\mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{1/2} = \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}^T\mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}^T \tag{21}$$
$$\mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{1/2} = \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{I}\mathbf{\Lambda}^{1/2}\mathbf{P}^T \tag{22}$$
$$\mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{1/2} = \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{P}^T \tag{23}$$
$$\mathbf{\Sigma}^{1/2}\mathbf{\Sigma}^{1/2} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^T = \mathbf{\Sigma} \tag{24}$$

N.B.

- The second line of the above proof follows from the provided information that $\mathbf{P}^T\mathbf{P} = \mathbf{P}\mathbf{P}^T = \mathbf{I}$.
- The last line of the above proof follows from our setup of $\mathbf{\Lambda}^{1/2}$ being a diagonal matrix of the square root of the corresponding eigenvalues, which results in the matrix multiplication of $\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}$ being a diagonal matrix of the corresponding eigenvalues, i.e. $\mathbf{\Lambda}$.

5.2.3 (c)

i. We want to show $\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{1/2} = \mathbf{I}$. Let's attempt to do that,

$$\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{1/2} = \mathbf{P}\mathbf{\Lambda}^{-1/2}\mathbf{P}^T\mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}^T \tag{25}$$

$$\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{1/2} = \mathbf{P}\mathbf{\Lambda}^{-1/2}\mathbf{I}\mathbf{\Lambda}^{1/2}\mathbf{P}^T \tag{26}$$

$$\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{1/2} = \mathbf{P}\mathbf{\Lambda}^{-1/2}\mathbf{\Lambda}^{1/2}\mathbf{P}^T \tag{27}$$

$$\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{1/2} = \mathbf{P}\mathbf{I}\mathbf{P}^T \tag{28}$$

$$\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{1/2} = \mathbf{P}\mathbf{P}^T \tag{29}$$

$$\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{1/2} = \mathbf{I} \tag{30}$$

- Throughout the proof we make use of the provided information that $\mathbf{P}^T\mathbf{P} = \mathbf{P}\mathbf{P}^T = \mathbf{I}$.
- We note that $\mathbf{\Lambda}^{-1/2}\mathbf{\Lambda}^{1/2} = \mathbf{I}$ (can verify this trivially by carrying out the matrix multiplication), which we make use of in the middle of the proof. Note that we are following our setup of $\mathbf{\Lambda}^{1/2}$ being a diagonal matrix of the square root of the corresponding eigenvalues.

ii.

$$\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{-1/2} = \mathbf{P}\mathbf{\Lambda}^{-1/2}\mathbf{P}^T\mathbf{P}\mathbf{\Lambda}^{-1/2}\mathbf{P}^T \tag{31}$$

$$\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{-1/2} = \mathbf{P}\mathbf{\Lambda}^{-1/2}\mathbf{I}\mathbf{\Lambda}^{-1/2}\mathbf{P}^T \tag{32}$$

$$\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{-1/2} = \mathbf{P}\mathbf{\Lambda}^{-1/2}\mathbf{\Lambda}^{-1/2}\mathbf{P}^T \tag{33}$$

$$\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{-1/2} = \mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{P}^T \tag{34}$$

$$\mathbf{\Sigma}^{-1/2}\mathbf{\Sigma}^{-1/2} = \mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{P}^T = \mathbf{\Sigma}^{-1} \tag{35}$$

- At the beginning of the proof we make use of the provided information that $\mathbf{P}^T\mathbf{P} = \mathbf{P}\mathbf{P}^T = \mathbf{I}$.
- We note that $\mathbf{\Lambda}^{-1/2}\mathbf{\Lambda}^{-1/2} = \mathbf{\Lambda}^{-1}$ (can verify this trivially by carrying out the matrix multiplication), which we make use of in the middle of the proof. Note that we are following our setup of $\mathbf{\Lambda}^{1/2}$ being a diagonal matrix of the square root of the corresponding eigenvalues.
- At the end of the proof we make use of what we showed earlier, $\mathbf{\Sigma}^{-1} = \mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{P}^T$.

5.2.4 (d)

Since we are provided that $\mathbf{\Sigma}$ is positive definite, all of its eigenvalues are strictly positive (we have shown this in an earlier question). We further note that the determinant of a matrix is a product of its eigenvalues (confirmed via Google). So the determinant of $\mathbf{\Sigma}$ is non-zero, and as a result $\mathbf{\Sigma}^{-1}$ exists (the fact that the inverse of a matrix exists if its determinant is non-zero was confirmed via Google).

6

6.1 Problem

Let $\mathbf{X}$ be an $n \times p$ matrix of constants. The idea is that $\mathbf{X}$ is the "design matrix" in the linear model $\mathbf{y} = \mathbf{X}\beta + \epsilon$, so this problem is really about linear regression.

6.1.1 (a)

Recall that $\mathbf{A}$ symmetric means $\mathbf{A} = \mathbf{A}^T$. Let $\mathbf{X}$ be an n by p matrix. Show that $\mathbf{X}^T\mathbf{X}$ is symmetric.

6.1.2 (b)

Recall the definition of linear independence. The columns of $\mathbf{A}$ are said to be *linearly dependent* if there exists a column vector $\mathbf{v} \neq \mathbf{0}$ with $\mathbf{Av} = \mathbf{0}$. If $\mathbf{Av} = \mathbf{0}$ implies $\mathbf{v} = \mathbf{0}$, the columns of $\mathbf{A}$ are said to be linearly *independent*. Show that if the columns of $\mathbf{X}$ are linearly independent, then $\mathbf{X}^T\mathbf{X}$ is positive definite.

6.1.3 (c)

Show that if $\mathbf{X}^T\mathbf{X}$ is positive definite then $(\mathbf{X}^T\mathbf{X})^{-1}$ exists.

6.1.4 (d)

Show that if $(\mathbf{X}^T\mathbf{X})^{-1}$ exists then the columns of $\mathbf{X}$ are linearly independent.

This is a good problem because it establishes that the least squares estimator $\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ exists if and only if the columns of $\mathbf{X}$ are linearly independent.

6.2 Solution

6.2.1 (a)

$$(\mathbf{X}^T\mathbf{X})^T = \mathbf{X}^T(\mathbf{X}^T)^T \tag{36}$$
$$(\mathbf{X}^T\mathbf{X})^T = \mathbf{X}^T\mathbf{X} \tag{37}$$

N.B. In our proof we used the formula $(AB)^T = B^TA^T$.

6.2.2 (b)

Take any $p \times 1$ vector $\mathbf{v} \neq \mathbf{0}$. We want to show $\mathbf{v}^T\mathbf{X}^T\mathbf{X}\mathbf{v} > 0$ (see info provided in Problem 4). Observe that $\mathbf{v}^T\mathbf{X}^T\mathbf{X}\mathbf{v} = (\mathbf{X}\mathbf{v})^T\mathbf{X}\mathbf{v}$. Now let $\mathbf{y} = \mathbf{X}\mathbf{v}$. Realize that $(\mathbf{X}\mathbf{v})^T\mathbf{X}\mathbf{v} = \mathbf{y}^T\mathbf{y} \geq 0$. Now, we are provided that the columns of $\mathbf{X}$ are linearly independent, i.e. $\mathbf{X}\mathbf{v} = \mathbf{0}$ implies $\mathbf{v} = \mathbf{0}$. Since we have taken $\mathbf{v} \neq \mathbf{0}$, then $\mathbf{y} = \mathbf{X}\mathbf{v} \neq \mathbf{0}$ (using a contrapositive argument), and thus we must have $\mathbf{v}^T\mathbf{X}^T\mathbf{X}\mathbf{v} = \mathbf{y}^T\mathbf{y} > 0$. So $\mathbf{X}^T\mathbf{X}$ is positive definite as required.

6.2.3 (c)

Use the same argument that was used for 5 (d) earlier.

6.2.4 (d)

We want to show that the columns of $\mathbf{X}$ are linearly independent, i.e. $\mathbf{X}\mathbf{v} = \mathbf{0}$ implies $\mathbf{v} = \mathbf{0}$. We are provided that $(\mathbf{X}^T\mathbf{X})^{-1}$ exists. Let's proceed,

$$\mathbf{X}\mathbf{v} = \mathbf{0} \tag{38}$$
$$\mathbf{X}^T\mathbf{X}\mathbf{v} = \mathbf{X}^T\mathbf{0} \tag{39}$$
$$\mathbf{X}^T\mathbf{X}\mathbf{v} = \mathbf{0} \tag{40}$$
$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\mathbf{v} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{0} \tag{41}$$
$$\mathbf{I}\mathbf{v} = \mathbf{0} \tag{42}$$

$$\mathbf{v} = \mathbf{0} \tag{43}$$

With this we have shown that if $(\mathbf{X}^T\mathbf{X})^{-1}$ exists, $\mathbf{X}\mathbf{v} = \mathbf{0}$ implies $\mathbf{v} = \mathbf{0}$, i.e. the columns of $\mathbf{X}$ are linearly independent, as required.

7

7.1 Problem

Women and men are coming into a store according to independent Poisson processes with rates λ_1 for women and λ_2 for men. You don't have to know anything about Poisson processes to do this question. We have that the number of women and the number of men entering the store in a given time period are independent Poisson random variables, with expected values λ_1 for women and λ_2 for men. Because the Poisson process is an independent increments process, we can treat the numbers from n time periods as a random sample.

Management wants to know the expected number of male customers and the expected number of female customers. Unfortunately, the total number of customers were recorded, but not their sex. Let $y_1, ..., y_n$ denote the total number of customers who enter the store in n time periods. That's all the data we have.

7.1.1 (a)

What is the distribution of y_i? If you know the answer, just write it down without proof.

7.1.2 (b)

What is the parameter space?

7.1.3 (c)

Find the MLE of the parameter vector (λ_1, λ_2). Show your work.

7.1.4 (d)

How is the question related to the Zipper example?

7.2 Solution

7.2.1 (a)

Poisson distribution with expected value $\lambda_1 + \lambda_2$.

7.2.2 (b)

$$\{(\lambda_1, \lambda_2) : \lambda_1 \geq 0, \lambda_2 \geq 0\} \tag{44}$$

7.2.3 (c)

Firstly, since we established $y_i \sim Po(\lambda_1 + \lambda_2)$, we note that the probability mass function for y_i is

$$P(Y_i = y_i) = \frac{(\lambda_1 + \lambda_2)^{y_i} e^{-(\lambda_1+\lambda_2)}}{y_i!} \tag{45}$$

Using this, let's set up the likelihood function

$$L(\lambda_1, \lambda_2) = \prod_{i=1}^{n} P(Y_i = y_i) \tag{46}$$

$$L(\lambda_1, \lambda_2) = \prod_{i=1}^{n} \frac{(\lambda_1 + \lambda_2)^{y_i} e^{-(\lambda_1+\lambda_2)}}{y_i!} \tag{47}$$

$$L(\lambda_1, \lambda_2) = \frac{(\lambda_1 + \lambda_2)^{\sum_{i=1}^{n} y_i} e^{-n(\lambda_1+\lambda_2)}}{\prod_{i=1}^{n} y_i!} \tag{48}$$

Then, obtain the log likelihood function

$$l(\lambda_1, \lambda_2) = \log(L(\lambda_1, \lambda_2)) \tag{49}$$

$$l(\lambda_1, \lambda_2) = \log\left(\frac{(\lambda_1 + \lambda_2)^{\sum_{i=1}^{n} y_i} e^{-n(\lambda_1+\lambda_2)}}{\prod_{i=1}^{n} y_i!}\right) \tag{50}$$

$$l(\lambda_1, \lambda_2) = \left(\sum_{i=1}^{n} y_i\right) \log(\lambda_1 + \lambda_2) - n(\lambda_1 + \lambda_2) - \sum_{i=1}^{n} \log(y_i!) \tag{51}$$

Next, we derive the partial derivatives with respect to the parameters,

$$\frac{\partial l}{\partial \lambda_1} = \frac{\sum_{i=1}^{n} y_i}{\lambda_1 + \lambda_2} - n \tag{52}$$

$$\frac{\partial l}{\partial \lambda_2} = \frac{\sum_{i=1}^{n} y_i}{\lambda_1 + \lambda_2} - n \tag{53}$$

We note that both the partial derivatives we obtained are the same, which if we set to zero,

$$\frac{\partial l}{\partial \lambda_1} = \frac{\partial l}{\partial \lambda_2} = 0 \tag{54}$$

$$\frac{\sum_{i=1}^{n} y_i}{\lambda_1 + \lambda_2} - n = 0 \tag{55}$$

$$\lambda_1 + \lambda_2 = \frac{\sum_{i=1}^{n} y_i}{n} = \bar{y} \tag{56}$$

We realize that any pair (λ_1, λ_2) with $\lambda_1 + \lambda_2 = \frac{\sum_{i=1}^{n} y_i}{n} = \bar{y}$ will maximize the likelihood, i.e. that the MLE is not unique.

7.2.4 (d)

As with the Zipper model, the parameters in our model are not identifiable.

8

8.1 Problem

Suppose $\sqrt{n}(T_n - \theta) \xrightarrow{d} T$. Show $T_n \xrightarrow{P} \theta$. Please use Slutsky lemmas rather than definitions. Hint: Think of the sequence of constants $\frac{1}{\sqrt{n}}$ as a sequence of degenerate random variables (variance zero) that converge almost surely and hence in probability to zero. Now you can use a Slutsky lemma.

8.2 Solution

Firstly, we take note of the hint provided, $\frac{1}{\sqrt{n}} \xrightarrow{a.s.} 0 \Rightarrow \frac{1}{\sqrt{n}} \xrightarrow{P} 0$. Then making use of the third Slutsky lemma for Convergence in Distribution from the lecture slides we have,

$$\begin{pmatrix} \sqrt{n}(T_n - \theta) \\ \frac{1}{\sqrt{n}} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} T \\ 0 \end{pmatrix} \tag{57}$$

Then, consider the function $f\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = xy$, which is continuous. Then, making use of the first Slutsky lemma for Convergence in Distribution from the lecture slides on what we established above we have,

$$f\left(\begin{pmatrix} \sqrt{n}(T_n - \theta) \\ \frac{1}{\sqrt{n}} \end{pmatrix}\right) \xrightarrow{d} f\left(\begin{pmatrix} T \\ 0 \end{pmatrix}\right) \tag{58}$$

$$\sqrt{n}(T_n - \theta)\frac{1}{\sqrt{n}} \xrightarrow{d} (T)(0) \tag{59}$$

$$T_n - \theta \xrightarrow{d} 0 \tag{60}$$

Since 0 is a constant, we have $T_n - \theta \xrightarrow{d} 0 \Rightarrow T_n - \theta \xrightarrow{P} 0$. Thus, we have $T_n - \theta \xrightarrow{P} 0 \Rightarrow T_n \xrightarrow{P} \theta$, as required.

9

9.1 Problem

Let $X_1, ..., X_n$ be a random sample from a Binomial distribution with parameters 3 and θ. That is,

$$P(X_i = x_i) = \binom{3}{x_i} \theta^{x_i}(1-\theta)^{3-x_i}, \tag{61}$$

for $x_i = 0, 1, 2, 3$. Find the maximum likelihood estimator θ, and show that it is strongly consistent.

9.2 Solution

First let's setup the likelihood function,

$$L(\theta) = \prod_{i=1}^{n} P(X_i = x_i) \tag{62}$$

$$L(\theta) = \prod_{i=1}^{n} \binom{3}{x_i} \theta^{x_i} (1-\theta)^{3-x_i} \tag{63}$$

$$L(\theta) = \left(\prod_{i=1}^{n} \binom{3}{x_i} \right) \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{3n-\sum_{i=1}^{n} x_i} \tag{64}$$

Next, let's obtain the log likelihood function,

$$l(\theta) = \log L(\theta) \tag{65}$$

$$l(\theta) = \sum_{i=1}^{n} \log \left(\binom{3}{x_i} \right) + \left(\sum_{i=1}^{n} x_i \right) \log \theta + \left(3n - \sum_{i=1}^{n} x_i \right) \log(1-\theta) \tag{66}$$

Then, we differentiate the log likelihood function with respect to θ,

$$\frac{dl}{d\theta} = \frac{\sum_{i=1}^{n} x_i}{\theta} + \left(\frac{3n - \sum_{i=1}^{n} x_i}{1-\theta} \right) (-1) \tag{67}$$

$$\frac{dl}{d\theta} = \frac{\sum_{i=1}^{n} x_i}{\theta} - \frac{3n - \sum_{i=1}^{n} x_i}{1-\theta} \tag{68}$$

Now, we set the obtained derivative to 0 yo obtain the MLE,

$$\frac{dl}{d\theta} = 0 \tag{69}$$

$$\frac{\sum_{i=1}^{n} x_i}{\theta} - \frac{3n - \sum_{i=1}^{n} x_i}{1-\theta} = 0 \tag{70}$$

$$\sum_{i=1}^{n} x_i - \theta \sum_{i=1}^{n} x_i = 3n\theta - \theta \sum_{i=1}^{n} x_i \tag{71}$$

$$\sum_{i=1}^{n} x_i = 3n\theta \tag{72}$$

$$\theta = \frac{\sum_{i=1}^{n} x_i}{3n} = \frac{\bar{x}}{3} \tag{73}$$

Thus, as required, we derived the MLE $\hat{\theta}_{MLE} = \frac{\bar{x}}{3}$. Next, we want to show strong consistency, i.e. that $\hat{\theta}_{MLE} = \frac{\bar{x}}{3} \xrightarrow{a.s.} \theta$. By SLLN, $\bar{x} \xrightarrow{a.s.} E(X) = 3\theta$. Since the function $f(x) = \frac{x}{3}$ is continuous,

$$\hat{\theta}_{MLE} = f(\bar{x}) \xrightarrow{a.s.} f(3\theta) = \frac{3\theta}{3} = \theta \tag{74}$$

With this we have shown our derived MLE is strongly consistent as well, as required.

10

10.1 Problem

Let $X_1, ..., X_n$ be a random sample from a continuous distribution with density

$$f(x;\tau) = \frac{\tau^{1/2}}{\sqrt{2\pi}} e^{-\frac{\tau x^2}{2}}, \tag{75}$$

where the parameter $\tau > 0$. Let

$$\hat{\tau} = \frac{n}{\sum_{i=1}^{n} X_i^2}. \tag{76}$$

Is $\hat{\tau}$ a consistent estimator of τ? Answer Yes or No and prove your answer. Hint: You can just write down $E(X^2)$ by inspection. This is a very familiar distribution.

10.2 Solution

First, let's consider the provided hint. By inspection we note the similarity of the provided density with that for the normal distribution. Take a random variable, Y, with a normal distribution, i.e. $Y \sim N(\mu, \sigma^2)$. The density for this random variable is

$$f(y;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{77}$$

Additionally we know that $E(Y)=\mu$ and $\text{Var}(Y)=\sigma^2$. Now getting back to our provided density, and after doing some algebraic manipulation,

$$f(x;\tau)=\frac{\tau^{1/2}}{\sqrt{2\pi}}e^{-\frac{\tau x^2}{2}} \tag{78}$$

$$f(x;\tau)=\frac{1}{\tau^{-1/2}\sqrt{2\pi}}e^{-\frac{x^2}{2\tau^{-1}}} \tag{79}$$

we realize that X_i is from a normal distribution with parameters $\mu=0$ and $\sigma^2=\tau^{-1}$. Thus, as indicated by the hint, we can write by inspection, $E(X^2)=\sigma_X^2+E(X)^2=\tau^{-1}+0=\tau^{-1}$. So, by SLLN, we can write $\frac{\sum_{i=1}^n X_i^2}{n}\xrightarrow{a.s.}E(X^2)=\tau^{-1}$. Since the function $f(x)=\frac{1}{x}$ is continuous, except possibly on a set with probability zero (?),

$$\hat{\tau}=\frac{n}{\sum_{i=1}^n X_i^2}=f\left(\frac{\sum_{i=1}^n X_i^2}{n}\right)\xrightarrow{a.s.}f(\tau^{-1})=\frac{1}{\tau^{-1}}=\tau \tag{80}$$

Since we observe that $\hat{\tau}\xrightarrow{a.s.}\tau$, we say that **yes**, $\hat{\tau}$ is a (strongly) consistent estimator of τ.

11

11.1 Problem

Let $X_1,...,X_n$ be a random sample from a distribution with mean μ. Show that $T_n=\frac{1}{n+400}\sum_{i=1}^n X_i$ is a strongly consistent estimator of μ.

11.2 Solution

First, consider we can re-express our estimator as follows,

$$T_n=\frac{1}{n+400}\sum_{i=1}^n X_i=\frac{1}{n+400}\frac{n}{n}\sum_{i=1}^n X_i=\frac{n}{n+400}\frac{1}{n}\sum_{i=1}^n X_i \tag{81}$$

Now, as was done in an earlier question, we can think of the sequence of constants $\frac{n}{n+400}$ as a sequence of degenerate random variables that converge almost surely to one, i.e. $\frac{n}{n+400} = \frac{1}{1+\frac{400}{n}} \xrightarrow{a.s.} \frac{1}{1+0} = 1$. Also, by SLLN, $\frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n \xrightarrow{a.s.} \mu$. Since the function $f(x, y) = xy$ is continuous,

$$T_n = f\left(\frac{n}{n+400}, \frac{1}{n}\sum_{i=1}^{n} X_i\right) \xrightarrow{a.s.} f(1, \mu) = (1)(\mu) = \mu \tag{82}$$

Since $T_n \xrightarrow{a.s.} \mu$, T_n is a strongly consistent estimator of μ, as required.

12

12.1 Problem

Let $X_1, ..., X_n$ be a random sample from a distribution with mean μ and variance σ^2. Prove that the sample variance $S^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}$ is a strongly consistent estimator of σ^2.

12.2 Solution

We want to first perform some algebraic manipulations on S^2 to re-express it in another form,

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} \tag{83}$$

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}\frac{n}{n} \tag{84}$$

$$S^2 = \frac{n}{n-1}\frac{1}{n}\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right) \tag{85}$$

$$S^2 = \frac{n}{n-1}\frac{1}{n}\left(\sum_{i=1}^{n}(X_i^2 - 2X_i\bar{X} + \bar{X}^2)\right) \tag{86}$$

$$S^2 = \frac{n}{n-1}\frac{1}{n}\left(\sum_{i=1}^{n}X_i^2 - 2\bar{X}\sum_{i=1}^{n}X_i + n\bar{X}^2\right) \tag{87}$$

$$S^2 = \frac{n}{n-1}\frac{1}{n}\left(\sum_{i=1}^{n} X_i^2 - 2n\bar{X}^2 + n\bar{X}^2\right) \tag{88}$$

$$S^2 = \frac{n}{n-1}\frac{1}{n}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) \tag{89}$$

$$S^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2\right) \tag{90}$$

Now, as was done in an earlier question, we can think of the sequence of constants $\frac{n}{n-1}$ as a sequence of degenerate random variables that converge almost surely to one, i.e. $\frac{n}{n-1} = \frac{1}{1-\frac{1}{n}} \xrightarrow{a.s.} \frac{1}{1-0} = 1$. Also, by SLLN, $\bar{X} \xrightarrow{a.s.} \mu$ and $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{a.s.} E(X^2) = \sigma^2 + \mu^2$. Since the function $f(x, y, z) = x(y - z^2)$ is continuous,

$$S^2 = f\left(\frac{n}{n-1}, \frac{1}{n}\sum_{i=1}^{n} X_i^2, \bar{X}\right) \xrightarrow{a.s.} f(1, \sigma^2 + \mu^2, \mu) = (1)(\sigma^2 + \mu^2 - \mu^2) = \sigma^2 \tag{91}$$

Since $S^2 \xrightarrow{a.s.} \sigma^2$, sample variance S^2 is a strongly consistent estimator of σ^2, as required.

13

13.1 Problem

Independently for $i = 1, ..., n$, let

$$Y_i = \beta X_i + \epsilon_i, \tag{92}$$

where $E(X_i) = E(\epsilon_i) = 0$, $Var(X_i) = \sigma_X^2$, $Var(\epsilon_i) = \sigma_\epsilon^2$, and ϵ_i is independent of X_i. Let

$$\hat{\beta}_n = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}. \tag{93}$$

Is $\hat{\beta}_n$ a consistent estimator of β? Answer Yes or No and prove your answer.

13.2 Solution

We want to first perform some algebraic manipulations on $\hat{\beta}_n$ to re-express it in another form,

$$\hat{\beta}_n = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2} \tag{94}$$

$$\hat{\beta}_n = \frac{\frac{1}{n}\sum_{i=1}^n X_i Y_i}{\frac{1}{n}\sum_{i=1}^n X_i^2} \tag{95}$$

Now, by SLLN, $\frac{1}{n}\sum_{i=1}^n X_i Y_i \xrightarrow{a.s.} E(XY)$ and $\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{a.s.} E(X^2)$, where

$$E(XY) = E[X(\beta X + \epsilon)] = E(\beta X^2 + X\epsilon) \tag{96}$$

$$E(XY) = \beta E(X^2) + E(X\epsilon) \tag{97}$$

$$E(XY) = \beta(Var(X) + E(X)^2) + E(X)E(\epsilon) \tag{98}$$

$$E(XY) = \beta(\sigma_X^2 + 0^2) + (0)(0) \tag{99}$$

$$E(XY) = \beta\sigma_X^2 \tag{100}$$

$$E(X^2) = Var(X) + E(X)^2 \tag{101}$$

$$E(X^2) = \sigma_X^2 + 0^2 \tag{102}$$

$$E(X^2) = \sigma_X^2 \tag{103}$$

Since the function $f(x,y) = \frac{x}{y}$ is continuous, except possibly on a set with probability zero (?),

$$\hat{\beta}_n = f\left(\frac{1}{n}\sum_{i=1}^n X_i Y_i, \frac{1}{n}\sum_{i=1}^n X_i^2\right) \xrightarrow{a.s.} f(\beta\sigma_X^2, \sigma_X^2) = \frac{\beta\sigma_X^2}{\sigma_X^2} = \beta \tag{104}$$

Since $\hat{\beta}_n \xrightarrow{a.s.} \beta$, we say, **yes**, $\hat{\beta}_n$ is a (strongly) consistent estimator of β.

14

14.1 Problem

In this problem, you'll use (without proof) the *variance rule*, which says that if θ is a real constant and $T_1, T_2, ...$ is a sequence of random variables with

$$\lim_{n\to\infty} E(T_n) = \theta \text{ and } \lim_{n\to\infty} Var(T_n) = 0, \tag{105}$$

then $T_n \xrightarrow{P} \theta$.
In Problem 13, the independent variables are random. Here they are fixed constants, which is more standard (though a little strange if you think about it). Accordingly, let

$$Y_i = \beta x_i + \epsilon_i \tag{106}$$

for $i = 1, ..., n$, where $\epsilon_1, ..., \epsilon_n$ are a random sample from a distribution with expected value zero and variance σ^2, and β and σ^2 are unknown constants.

14.1.1 (a)

What is $E(Y_i)$?

14.1.2 (b)

What is $Var(Y_i)$?

14.1.3 (c)

Use the same estimator as in Problem 13. Is $\hat{\beta}_n$ unbiased? Answer Yes or No and show your work.

14.1.4 (d)

Suppose that the sequence of constants $\sum_{i=1}^n x_i^2 \to \infty$ as $n \to \infty$. Does this guarantee $\hat{\beta}_n$ will be consistent? Answer Yes or No. Show your work.

14.1.5 (e)

Let $\hat{\beta}_{2,n} = \frac{\bar{Y}_n}{\bar{x}_n}$. Is $\hat{\beta}_{2,n}$ unbiased? Consistent? Answer Yes or No to each question and show your work. Do you need a condition on the x_i values?

14.1.6 (f)

Prove that $\hat{\beta}_n$ is a more accurate estimator than $\hat{\beta}_{2,n}$ in the sense that it has smaller variance. Hint: The sample variance of the explanatory variable values cannot be negative.

14.2 Solution

14.2.1 (a)

$$E(Y_i) = E(\beta x_i + \epsilon) \tag{107}$$
$$E(Y_i) = \beta x_i + E(\epsilon) \tag{108}$$
$$E(Y_i) = \beta x_i + 0 \tag{109}$$
$$E(Y_i) = \beta x_i \tag{110}$$

14.2.2 (b)

$$Var(Y_i) = Var(\beta x_i + \epsilon) \tag{111}$$
$$Var(Y_i) = Var(\epsilon) \tag{112}$$
$$Var(Y_i) = \sigma^2 \tag{113}$$

14.2.3 (c)

$$E(\hat{\beta}_n) = E\left(\frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}\right) \tag{114}$$
$$E(\hat{\beta}_n) = \frac{\sum_{i=1}^n x_i E(Y_i)}{\sum_{i=1}^n x_i^2} \tag{115}$$
$$E(\hat{\beta}_n) = \frac{\sum_{i=1}^n x_i (\beta x_i)}{\sum_{i=1}^n x_i^2} \tag{116}$$
$$E(\hat{\beta}_n) = \frac{\beta \sum_{i=1}^n x_i^2}{\sum_{i=1}^n x_i^2} \tag{117}$$
$$E(\hat{\beta}_n) = \beta \tag{118}$$

Since $E(\hat{\beta}_n) = \beta$, we say, **yes**, $\hat{\beta}_n$ is unbiased.

14.2.4 (d)

As indicated by the problem, let's attempt to use the provided variance rule for this. As such let's proceed to check the two conditions of the rule,

$$\lim_{n\to\infty} E(\hat{\beta}_n) = \lim_{n\to\infty} \beta = \beta \tag{119}$$

Before wanting to test out the second condition, we want to derive,

$$Var(\hat{\beta}_n) = Var\left(\frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}\right) \tag{120}$$

$$Var(\hat{\beta}_n) = Var\left(\frac{\sum_{i=1}^n x_i(\beta x_i + \epsilon_i)}{\sum_{i=1}^n x_i^2}\right) \tag{121}$$

$$Var(\hat{\beta}_n) = Var\left(\frac{\beta\left(\sum_{i=1}^n x_i^2\right) + \sum_{i=1}^n x_i\epsilon_i}{\sum_{i=1}^n x_i^2}\right) \tag{122}$$

$$Var(\hat{\beta}_n) = Var\left(\beta + \frac{\sum_{i=1}^n x_i\epsilon_i}{\sum_{i=1}^n x_i^2}\right) \tag{123}$$

$$Var(\hat{\beta}_n) = Var\left(\frac{\sum_{i=1}^n x_i\epsilon_i}{\sum_{i=1}^n x_i^2}\right) \tag{124}$$

$$Var(\hat{\beta}_n) = \frac{\sum_{i=1}^n x_i^2 Var(\epsilon_i)}{\left(\sum_{i=1}^n x_i^2\right)^2} + \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \frac{x_i x_j Cov(\epsilon_i, \epsilon_j)}{\left(\sum_{i=1}^n x_i^2\right)\left(\sum_{j=1}^n x_j^2\right)} \tag{125}$$

$$Var(\hat{\beta}_n) = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{\left(\sum_{i=1}^n x_i^2\right)^2} + 0 \tag{126}$$

$$Var(\hat{\beta}_n) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2} \tag{127}$$

Now, we can test out the second condition,

$$\lim_{n\to\infty} Var(\hat{\beta}_n) = \lim_{n\to\infty} \frac{\sigma^2}{\sum_{i=1}^n x_i^2} = 0 \tag{128}$$

which follows from the provided condition that the sequence of constants $\sum_{i=1}^n x_i^2 \to \infty$ as $n \to \infty$. Having shown the two conditions of the variance rule we can say that $\hat{\beta}_n \xrightarrow{P} \beta$, that is, **yes**, the provided condition on the x_i values guarantee $\hat{\beta}_n$ will be consistent.

14.2.5 (e)

First, let's check if it is unbiased,

$$E(\hat{\beta}_{2,n}) = E\left(\frac{\bar{Y}_n}{\bar{x}_n}\right) \tag{129}$$

$$E(\hat{\beta}_{2,n}) = E\left(\frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} x_i}\right) \tag{130}$$

$$E(\hat{\beta}_{2,n}) = \frac{\sum_{i=1}^{n} E(Y_i)}{\sum_{i=1}^{n} x_i} \tag{131}$$

$$E(\hat{\beta}_{2,n}) = \frac{\sum_{i=1}^{n} \beta x_i}{\sum_{i=1}^{n} x_i} \tag{132}$$

$$E(\hat{\beta}_{2,n}) = \frac{\beta \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i} \tag{133}$$

$$E(\hat{\beta}_{2,n}) = \beta \tag{134}$$

Since we have found $E(\hat{\beta}_{2,n}) = \beta$, we say, **yes**, $E(\hat{\beta}_{2,n})$ is unbiased. Next, we inspect the consistency, and as with previous question, we will make use of the variance rule. Let's check the first condition,

$$\lim_{n\to\infty} E(\hat{\beta}_{2,n}) = \lim_{n\to\infty} \beta = \beta \tag{135}$$

As done in the previous question, before wanting to test out the second condition, we want to derive,

$$Var(\hat{\beta}_{2,n}) = Var\left(\frac{\bar{Y}_n}{\bar{x}_n}\right) \tag{136}$$

$$Var(\hat{\beta}_{2,n}) = Var\left(\frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} x_i}\right) \tag{137}$$

$$Var(\hat{\beta}_{2,n}) = Var\left(\frac{\sum_{i=1}^{n}(\beta x_i + \epsilon_i)}{\sum_{i=1}^{n} x_i}\right) \tag{138}$$

$$Var(\hat{\beta}_{2,n}) = Var\left(\frac{\beta \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \epsilon_i}{\sum_{i=1}^{n} x_i}\right) \tag{139}$$

$$Var(\hat{\beta}_{2,n}) = Var\left(\beta + \frac{\sum_{i=1}^{n} \epsilon_i}{\sum_{i=1}^{n} x_i}\right) \tag{140}$$

$$Var(\hat{\beta}_{2,n}) = Var\left(\frac{\sum_{i=1}^n \epsilon_i}{\sum_{i=1}^n x_i}\right) \tag{141}$$

$$Var(\hat{\beta}_{2,n}) = \frac{\sum_{i=1}^n Var(\epsilon_i)}{\left(\sum_{i=1}^n x_i\right)^2} + \sum_{i=1}^n \sum_{j=1,j\neq 1} \frac{Cov(\epsilon_i, \epsilon_j)}{\left(\sum_{i=1}^n x_i\right)\left(\sum_{j=1}^n x_j\right)} \tag{142}$$

$$Var(\hat{\beta}_{2,n}) = \frac{\sum_{i=1}^n \sigma^2}{\left(\sum_{i=1}^n x_i\right)^2} + 0 \tag{143}$$

$$Var(\hat{\beta}_{2,n}) = \frac{n\sigma^2}{\left(\sum_{i=1}^n x_i\right)^2} \tag{144}$$

$$Var(\hat{\beta}_{2,n}) = \frac{\sigma^2}{\frac{1}{n}\left(\sum_{i=1}^n x_i\right)^2} = \frac{\sigma^2}{n\bar{x}_n^2} \tag{145}$$

We note that $n\bar{x}_n^2 \to \infty$ as $n \to \infty$. Now let's check the second condition of the variance rule,

$$\lim_{n\to\infty} Var(\hat{\beta}_{2,n}) = \lim_{n\to\infty} \frac{\sigma^2}{n\bar{x}_n^2} = 0 \tag{146}$$

Having shown the two conditions of the variance rule we can say that $\hat{\beta}_{2,n} \xrightarrow{P} \beta$, that is, **yes**, $\hat{\beta}_{2,n}$ is consistent. We **did not need** a condition on the x_i values (note that we have $n\bar{x}_n^2 \to \infty$ as $n \to \infty$, regardless of the x_i values).

14.2.6 (f)

$$Var(\hat{\beta}_n) - Var(\hat{\beta}_{2,n}) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2} - \frac{\sigma^2}{n\bar{x}_n^2} \tag{147}$$

$$Var(\hat{\beta}_n) - Var(\hat{\beta}_{2,n}) = \sigma^2\left(\frac{1}{\sum_{i=1}^n x_i^2} - \frac{1}{n\bar{x}_n^2}\right) \tag{148}$$

$$Var(\hat{\beta}_n) - Var(\hat{\beta}_{2,n}) = \sigma^2\left(\frac{n\bar{x}_n^2 - \sum_{i=1}^n x_i^2}{n\bar{x}_n^2 \sum_{i=1}^n x_i^2}\right) \tag{149}$$

$$Var(\hat{\beta}_n) - Var(\hat{\beta}_{2,n}) = \frac{\sigma^2}{n\bar{x}_n^2 \sum_{i=1}^n x_i^2}\left(n\bar{x}_n^2 - \sum_{i=1}^n x_i^2\right) \tag{150}$$

$$Var(\hat{\beta}_n) - Var(\hat{\beta}_{2,n}) = \frac{\sigma^2}{n\bar{x}_n^2 \sum_{i=1}^n x_i^2}\left(-\sum_{i=1}^n (x_i^2 - \bar{x}_n^2)\right) \tag{151}$$

$$Var(\hat{\beta}_n) - Var(\hat{\beta}_{2,n}) = \frac{\sigma^2}{n\bar{x}_n^2 \sum_{i=1}^n x_i^2} \left(-\sum_{i=1}^n (x_i^2 - 2\bar{x}_n^2 + \bar{x}_n^2) \right) \tag{152}$$

$$Var(\hat{\beta}_n) - Var(\hat{\beta}_{2,n}) = \frac{\sigma^2}{n\bar{x}_n^2 \sum_{i=1}^n x_i^2} \left(-\sum_{i=1}^n (x_i^2 + \bar{x}_n^2) + \sum_{i=1}^n 2\bar{x}_n^2 \right) \tag{153}$$

$$Var(\hat{\beta}_n) - Var(\hat{\beta}_{2,n}) = \frac{\sigma^2}{n\bar{x}_n^2 \sum_{i=1}^n x_i^2} \left(-\sum_{i=1}^n (x_i^2 + \bar{x}_n^2) + 2n\bar{x}_n^2 \right) \tag{154}$$

$$Var(\hat{\beta}_n) - Var(\hat{\beta}_{2,n}) = \frac{\sigma^2}{n\bar{x}_n^2 \sum_{i=1}^n x_i^2} \left(-\sum_{i=1}^n (x_i^2 + \bar{x}_n^2) + 2\bar{x}_n \sum_{i=1}^n x_i \right) \tag{155}$$

$$Var(\hat{\beta}_n) - Var(\hat{\beta}_{2,n}) = \frac{\sigma^2}{n\bar{x}_n^2 \sum_{i=1}^n x_i^2} \left(-\sum_{i=1}^n (x_i^2 + \bar{x}_n^2) + \sum_{i=1}^n 2\bar{x}_n x_i \right) \tag{156}$$

$$Var(\hat{\beta}_n) - Var(\hat{\beta}_{2,n}) = \frac{\sigma^2}{n\bar{x}_n^2 \sum_{i=1}^n x_i^2} \left(-\sum_{i=1}^n (x_i^2 - 2\bar{x}_n x_i + \bar{x}_n^2) \right) \tag{157}$$

$$Var(\hat{\beta}_n) - Var(\hat{\beta}_{2,n}) = \frac{\sigma^2}{n\bar{x}_n^2 \sum_{i=1}^n x_i^2} \left(-\sum_{i=1}^n (x_i - \bar{x}_n)^2 \right) \tag{158}$$

We note that $\frac{\sigma^2}{n\bar{x}_n^2 \sum_{i=1}^n x_i^2} \geq 0$ and $-\sum_{i=1}^n (x_i - \bar{x}_n)^2 \leq 0$ (the latter particularly we know from the provided hint that the sample variance of the explanatory variable cannot be negative). So, we have established that $Var(\hat{\beta}_n) - Var(\hat{\beta}_{2,n}) \leq 0$, i.e. that $\hat{\beta}_n$ is a more accurate estimator than $\hat{\beta}_{2,n}$ in the sense that it has smaller variance, as required.

15

15.1 Problem

Let X be a random variable with expected value μ and variance σ^2. Show $\frac{X}{n} \xrightarrow{P} 0$.

15.2 Solution

My humble idea is to make use of Slutsky Theorems for Convergence in Probability provided in the lecture slides. We can perhaps say that trivially

$X \xrightarrow{P} X$. Additionally, as we did in previous questions, we can think of the sequence of constants $\frac{1}{n}$ as a sequence of degenerate random variables that converge almost surely to one and hence in converge in probability to one, i.e. $\frac{1}{n} \xrightarrow{a.s.} 0 \Rightarrow \frac{1}{n} \xrightarrow{P} 0$. By the third of the Slutsky Theorems for Convergence in Probability provided in the lecture slides,

$$\begin{pmatrix} X \\ \frac{1}{n} \end{pmatrix} \xrightarrow{P} \begin{pmatrix} X \\ 0 \end{pmatrix} \tag{159}$$

Then, consider the function $f\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = xy$, which is continuous. Then, making use of the first Slutsky lemma for Convergence in Probability from the lecture slides on what we established above we have,

$$f\left(\begin{pmatrix} X \\ \frac{1}{n} \end{pmatrix}\right) \xrightarrow{P} f\left(\begin{pmatrix} X \\ 0 \end{pmatrix}\right) \tag{160}$$

$$\frac{X}{n} \xrightarrow{P} (X)(0) \tag{161}$$

$$\frac{X}{n} \xrightarrow{P} 0 \tag{162}$$

as required.

N.B. Previously we proved this using variance rule. That is, let $T_n = \frac{X}{n}$. Then,

$$\lim_{n\to\infty} E(T_n) = \lim_{n\to\infty} \frac{\mu}{n} = 0 \tag{163}$$

$$\lim_{n\to\infty} Var(T_n) = \lim_{n\to\infty} \frac{\sigma^2}{n^2} = 0 \tag{164}$$

So, by Variance Rule, $\frac{X}{n} \xrightarrow{P} 0$, as required.

16

16.1 Problem

Let $X_1, ..., X_n$ be a random sample from a Gamma distribution with $\alpha = \beta = \theta > 0$. That is, the density is

$$f(x;\theta) = \frac{1}{\theta^\theta \Gamma(\theta)} e^{-x/\theta} x^{\theta-1}, \tag{165}$$

for $x > 0$. Let $\hat{\theta} = \bar{X}_n$. Is $\hat{\theta}$ a consistent estimator of θ? Answer Yes or No and prove your answer.

16.2 Solution

By SLLN we know $\hat{\theta} = \bar{X}_n \xrightarrow{a.s.} E(X)$. So we are interested in deriving $E(X)$. So, let's proceed,

$$E(X) = \int_{-\infty}^{\infty} x f(x;\theta)\, dx \tag{166}$$

$$E(X) = \int_0^{\infty} x \frac{1}{\theta^\theta \Gamma(\theta)} e^{-x/\theta} x^{\theta-1}\, dx \tag{167}$$

$$E(X) = \frac{1}{\theta^\theta \Gamma(\theta)} \int_0^{\infty} x e^{-x/\theta} x^{\theta-1}\, dx \tag{168}$$

now, we will be doing a change of variables, specifically $x = \theta z$,

$$E(X) = \frac{1}{\theta^\theta \Gamma(\theta)} \int_0^{\infty} \theta z e^{-(\theta z)/\theta} (\theta z)^{\theta-1}\, \theta dz \tag{169}$$

$$E(X) = \frac{1}{\theta^\theta \Gamma(\theta)} \theta^{\theta+1} \int_0^{\infty} e^{-z} z^{\theta}\, dz \tag{170}$$

$$\frac{\Gamma(\theta)}{\theta} E(X) = \int_0^{\infty} e^{-z} z^{\theta}\, dz \tag{171}$$

$$\frac{\Gamma(\theta)}{\theta} E(X) = z^\theta \int_0^{\infty} e^{-z}\, dz - \theta \int_0^{\infty} z^{\theta-1} \left(\int e^{-z}\, dz \right) dz \tag{172}$$

$$\frac{\Gamma(\theta)}{\theta} E(X) = -z^\theta e^{-z} \Big|_0^{\infty} + \theta \int_0^{\infty} z^{\theta-1} e^{-z}\, dz \tag{173}$$

$$\frac{\Gamma(\theta)}{\theta}E(X) = \theta \int_0^\infty z^{\theta-1}e^{-z}\, dz \tag{174}$$

we make the observation that integrating $\int_0^\infty e^{-z}z^\theta \, dz$ is a recursive process using integration by parts and we can deduce the end result by inspection,

$$\frac{\Gamma(\theta)}{\theta}E(X) = \theta \int_0^\infty z^{\theta-1}e^{-z}\, dz \tag{175}$$

$$\frac{\Gamma(\theta)}{\theta}E(X) = \theta(\theta-1)(\theta-2)...(2)(1) \int_0^\infty e^{-z}\, dz \tag{176}$$

$$\frac{\Gamma(\theta)}{\theta}E(X) = \theta! \left(-e^{-z}\big|_0^\infty\right) \tag{177}$$

$$\frac{\Gamma(\theta)}{\theta}E(X) = \theta! \tag{178}$$

$$\frac{(\theta-1)!}{\theta}E(X) = \theta! \tag{179}$$

$$E(X) = \theta^2 \tag{180}$$

As a result, since $\hat{\theta} = \bar{X}_n \xrightarrow{a.s.} E(X) = \theta^2 \neq \theta$, we say **no**, $\hat{\theta}$ is **not** a consistent estimator of θ.

17

17.1 Problem

Here is an integral you cannot do in closed form, and numerical integration is challenging. For example, R's **integrate** function fails.

$$\int_0^{1/2} e^{\cos(1/x)}\, dx \tag{181}$$

Using R, approximate the integral with Monte Carlo integration, and give a 99% confidence interval for your answer. You need to produce 3 numbers: the estimate, a lower confidence limit and an upper confidence limit. See lecture slides. **Bring your printout to the quiz.**

17.2 Solution

First we want a density, $f(x)$ with $f(x) > 0$ wherever the integrand $e^{\cos(1/x)} \neq 0$. A uniform distribution over the integral limits seems to be simple enough and satisfactory to our purposes. So, let $X_1, ..., X_n$ be a random sample from $unif(0, 1/2)$ and X be a general random variable from the distribution. Then, we note,

$$\int_0^{1/2} e^{\cos(1/x)}\, dx = \int_0^{1/2} \frac{1}{2} e^{\cos(1/x)} 2\, dx = E\left[\frac{1}{2} e^{\cos(1/X)}\right] = E[g(X)] \quad (182)$$

Now, let's lay out our process. First we want to sample $X_1, ..., X_n$ from $unif(0, 1/2)$ (say for our example let's use a sample size of $n = 10000$). Then we compute $Y_i = g(X_i) = \frac{1}{2} e^{\cos(1/X_i)}$ for $i = 1, ..., n$. Next we calculate our estimate of the integral $\bar{Y}_n$, i.e. $\bar{Y}_n \xrightarrow{a.s.} E[Y] = E[g(X)] = \int_0^{1/2} e^{\cos(1/x)}\, dx$. Finally, we compute the confidence interval using the CLT, $\bar{Y}_n \pm z_{\alpha/2}\sqrt{\frac{S_Y^2}{m}}$, where $\bar{Y}_n$ is our Monte Carlo estimate, $z_{\alpha/2}$ is our critical value for our chosen confidence level, S_Y^2 is the sample variance of the Y_i values, and m is our Monte Carlo sample size. The code written in R for carrying these computations out is disclosed below

```
# Question 17 computation
# set seed
set.seed(9999)
# Monte Carlo sample size
m <- 10000
# generate uniform density variates
# density function, f, being used is that of the uniform distribution
# with endpoints 0 and 0.5 (the integral bounds)
X <- runif(m, 0, 0.5)
# next compute Y from X
# note that density function, f(x) = 2 for 0<=x<=0.5 and 0 elsewhere
# so Y = (h(X)/f(X)) = (1/2)*h(X), where h(X) is integrand provided
Y <- 0.5*exp(cos(1/X))
# compute mean of Y - this is the approximate integral by SLLN
I <- mean(Y)
# need variance of Y for computing confidence interval
s <- var(Y)
```

```r
# next need the standard normal test stat for the significance level
alpha <- 0.01
z_c <- -qnorm(alpha/2)
# now compute lower and upper bounds
# interval below derived from dist'n sqrt(m)*(Y_bar-mu) ~ N(0,sigma^
# which is provided by the CLT
# not using t-stat as random variates not from a normal dist'n
lb <- I - z_c*(sqrt(s)/sqrt(m))
ub <- I + z_c*(sqrt(s)/sqrt(m))
# print the estimate
sprintf('Integral: %.4f',I)
# print the cf
sprintf('Lower confidence limit: %.4f', lb)
sprintf('Upper confidence limit: %.4f', ub)
```

Executing this code in R, we get an integral estimate of 0.4489 and and a confidence interval of (0.4397, 0.4581).