Assignment 8

Adarsh Sai - AI20BTECH11001

Download all python codes from

```
https://github.com/Adarsh541/AI1103−prob−and−ranvar/blob/main/Assignment8.1/codes/Assignment8.1.py
```

and latex-tikz codes from

```
https://github.com/Adarsh541/AI1103−prob−and−ranvar/blob/main/Assignment8.1/Assignment8.1.tex
```

1 Problem

Let X_1, X_2, X_3, X_4, X_5 be a random sample of size 5 from a population having standard normal distribution.If $\overline{X} = \frac{1}{5}\sum_{i=1}^{5} X_i$ and $T = \sum_{i=1}^{5}\left(X_i - \overline{X}\right)^2$ then $E[T^2\overline{X}^2]$ is equal to

1) 3
2) 3.6
3) 4.8
4) 5.2

2 Solution

2.1 Terminology

Let $\mathbf{X}$ be a standard normal random vector

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{pmatrix} \tag{2.1.1}$$

Then $\overline{X}$ can be written as

$$\overline{X} = \frac{1}{5}\mathbf{X}^\top \mathbf{u} \tag{2.1.2}$$

where

$$\mathbf{u} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \tag{2.1.3}$$

Let $\mathbf{Y}$ be another random vector defined as

$$\mathbf{Y} = \mathbf{X} - \frac{1}{5}\mathbf{X}^\top \mathbf{u}^2 \tag{2.1.4}$$

$$\mathbf{Y} = \begin{pmatrix} X_1 - \overline{X} \\ X_2 - \overline{X} \\ X_3 - \overline{X} \\ X_4 - \overline{X} \\ X_5 - \overline{X} \end{pmatrix} \tag{2.1.5}$$

then

$$T = \mathbf{Y}^\top \mathbf{Y} \tag{2.1.6}$$

2.2 Theorem

Let $\overline{X_n}$ be the sample mean of size n from a normal distribution with mean μ and variance σ^2.Then

1) $\overline{X_n} \sim N(\mu, \frac{\sigma^2}{n})$
2) $\overline{X}$ and S^2 are independent.
3) $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

where χ^2_{n-1} is chi-square distribution with $(n-1)$ degrees of freedom and S^2 is defined as

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{2.2.1}$$

2.3 Useful concepts

If X,Y are independent random variables.then

$$E[XY] = E[X]E[Y] \tag{2.3.1}$$

$$E[X^2] = Var[X] + (E[X])^2 \tag{2.3.2}$$

If X is chi-square distributed with parameter k,then

$$E[X] = k \tag{2.3.3}$$

$$Var[X] = 2k \tag{2.3.4}$$

2.4 solution

For standard normal distribution $\mu = 0, \sigma^2 = 1$.So from the above theorem for $n = 5$ we have

1) T/4 and $\overline{X}$ are independent.
2) $\overline{X} \sim N(0, 1/5)$

3) $T \sim \chi_4^2$

So from (2.3.3) and (2.3.4)

$$E[T] = 4 \tag{2.4.1}$$
$$Var[T] = 8 \tag{2.4.2}$$

Since $\frac{T}{4}$ and $\overline{X}$ are independent,T and $\overline{X}$ are also independent

$$E[T^2\overline{X}^2] = E[T^2]E[\overline{X}^2] \tag{2.4.3}$$

from (2.3.2)

$$E[\overline{X}^2] = \frac{1}{5} \tag{2.4.4}$$
$$E[T^2] = 24 \tag{2.4.5}$$

So from (2.4.3)

$$E[T^2\overline{X}^2] = 4.8 \tag{2.4.6}$$

3 Independency

In this section we will prove that for a normally distributed population the sample mean and sample variance are independent.The sample mean is defined as

$$\overline{X} = \frac{1}{5}\sum_{i=1}^{5} X_i = \frac{1}{5}\mathbf{X}^\top\mathbf{X} \tag{3.0.1}$$

the sample variance is defined as

$$S^2 = \frac{1}{4}\sum_{i=1}^{5}(X_i - \overline{X})^2 = \frac{1}{4}\mathbf{Y}^\top\mathbf{Y} \tag{3.0.2}$$

where X_i are i.i.d and normally distributed and n is the sample size.

3.1 properties of mean and variance

If X and Y are independent random variables

$$E[aX + bY] = aE[X] + bE[Y] \tag{3.1.1}$$
$$Var[aX + b] = a^2Var[X] \tag{3.1.2}$$
$$Var[aX + bY] = a^2Var[X] + b^2Var[Y] \tag{3.1.3}$$

3.2 properties of covariance

If X,Y,Z are random variables and a,b are constants

$$Cov[aX, bY] = ab \times Cov[X, Y] \tag{3.2.1}$$
$$Cov[X + Z, Y] = Cov[X, Y] + Cov[Z, Y] \tag{3.2.2}$$
$$Cov[X, X] = Var[X] \tag{3.2.3}$$

3.3 Multivariate normal distribution

One definition is that a random vector is said to be a k-variate normally distributed if every linear combination of its k components has univariate normal distribution.**If a random vector has multivariate normal distribution then any two or more of its components that are uncorrelated are independent.**

3.4 proof

X_i are i.i.d and standard normally distributed

$$\overline{X} - X_1 = \frac{1}{5}(X_2 + X_3 + X_4 + X_5) - \left(\frac{4}{5}\right)X_1 \tag{3.4.1}$$

from properties of mean and variance

$$\overline{X} - X_1 = N\left(0, \frac{4}{5}\right) \tag{3.4.2}$$

So $\mathbf{Y}$ is also a normal random vector. We also note that

1)

$$Cov\left[X_j, \overline{X}\right] = Cov\left[X_j, \frac{1}{5}\sum_{i=1}^{5} X_i\right] \tag{3.4.3}$$
$$= \frac{1}{5}\sum_{i=1}^{5} Cov\left[X_j, X_i\right] \tag{3.4.4}$$
$$= \frac{1}{5} \tag{3.4.5}$$

2) and so for $j \neq k$ it follows that

$$Cov\left[X_j - \overline{X}, X_k - \overline{X}\right] = Cov\left[X_j - X_k\right] - Cov\left[X_j, \overline{X}\right] - Cov\left[\overline{X}, X_k\right] + Cov\left[\overline{X}, \overline{X}\right] \tag{3.4.6}$$

$$= 0 - \frac{1}{5} - \frac{1}{5} + \frac{1}{5} \tag{3.4.7}$$
$$= -\frac{1}{5} \tag{3.4.8}$$

3) similarly

$$Cov\left[\overline{X}, X_j - \overline{X}\right] = Cov\left[X_j, \overline{X}\right] - Cov\left[\overline{X}, \overline{X}\right] \tag{3.4.9}$$

$$= 0 \tag{3.4.10}$$

Thus,we see that $\left(\overline{X}, X_1 - \overline{X}, X_2 - \overline{X}, .., X_5 - \overline{X}\right)^\top$ is $N\left(0, \Lambda\right)$ where

$$\Lambda = \begin{pmatrix} \frac{1}{5} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{4}{5} & -1/5 & -1/5 & -1/5 & -1/5 \\ 0 & -1/5 & \frac{4}{5} & -1/5 & -1/5 & -1/5 \\ 0 & -1/5 & -1/5 & \frac{4}{5} & -1/5 & -1/5 \\ 0 & -1/5 & -1/5 & -1/5 & \frac{4}{5} & -1/5 \\ 0 & -1/5 & -1/5 & -1/5 & -1/5 & \frac{4}{5} \end{pmatrix} \tag{3.4.11}$$

So from the property of multivariate normal distribution we can conclude from the form of Λ that $\overline{X}$ and $\mathbf{Y} = \left(X_1 - \overline{X}, ..., X_5 - \overline{X}\right)^\top$ are independent normal vectors.Since $\overline{X}$ and $\mathbf{Y}$ are independent so too are $\mathbf{Y}^\top\mathbf{Y}$ and $\overline{X}$