Sampling distribution of variance

Saturday, 2 December 2023 10:21 AM

- Let a random sample of size n is drawn from a normal population with mean $= \mu$ and variance $= \sigma^2$ and the sample variance is computed, we obtain a value of the statistic s^2.

- Now we are interested in knowing the distribution of s^2.

- We proceed as follows:-

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{(n-1)}$$

$$(n-1)s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$(n-1)s^2 = \sum_{i=1}^{n} (x_i - \bar{x} + \mu - \mu)^2 \quad \text{(adding \& subtracting } \mu)$$

$$= \sum_{i=1}^{n} \left((x_i - \mu) - (\bar{x} - \mu)\right)^2$$

$$= \sum_{i=1}^{n} (x_i - \mu)^2 + (\bar{x} - \mu)^2 - 2(x_i - \mu)(\bar{x} - \mu)$$

$$= \sum_{i=1}^{n} (x_i - \mu)^2 + \sum_{i=1}^{n} \underbrace{(\bar{x} - \mu)^2}_{\text{independent of } i} - 2 \sum_{i=1}^{n} (x_i - \mu) \underbrace{(\bar{x} - \mu)}_{\text{independent of } i}$$

$$= \sum_{i=1}^{n} (x_i - \mu)^2 + (\bar{x} - \mu)^2 \underbrace{\sum_{i=1}^{n} 1}_{= n} - 2(\bar{x} - \mu) \underbrace{\sum_{i=1}^{n} (x_i - \mu)}_{n\bar{x} - n\mu}$$

$$= \sum_{i=1}^{n} (x_i - \mu)^2 + n(\bar{x} - \mu)^2 - 2n(\bar{x} - \mu)(\bar{x} - \mu)$$

$$(n-1)s^2 = \sum_{i=1}^{n} (x_i - \mu)^2 + n(\bar{x} - \mu)^2 - 2n(\bar{x} - \mu)^2$$

$$(n-1)s^2 = \sum_{i=1}^{n} (x_i - \mu)^2 - n(\bar{x} - \mu)^2 \quad — \star$$

$$\frac{(n-1)s^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma}\right)^2 - \left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right)^2$$

Now x_i is sample from a normal population with mean $= \mu$ & S.D. $= \sigma^2$

$\therefore \frac{x_i - \mu}{\sigma}$ is a standard normal distribution

This is sampling distribution of mean

$\therefore$ a standard normal distribution

Note this subtraction sign

So we get a total of $(n-1)$ squares of standard normal random variables on the right hand side. → That is the definition of chi square distribution.

$\therefore \frac{(n-1)s^2}{\sigma^2}$ is a chi square distribution with $(n-1)$ degrees of freedom.

↓

This is our result.

Que A manufacturer of car batteries guarntees that the batteries will last, on average, 3 years with a standard deviation of 1 year. If five of these batteries have the lifetimes of 1.9, 2.4, 3.0, 3.5 & 4.2 years, should the manufacturer still be convinced that the batteries have a S.D. of 1 year? Assume that the battery lifetime follows a normal distribution.

Soln

$$s^2 = \frac{1}{(n-1)} \sum_{i=1}^{5} (x_i - \bar{x})^2 = 0.815$$

(do the calculations)

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(4)(0.815)}{1} = 3.26 \quad \star$$

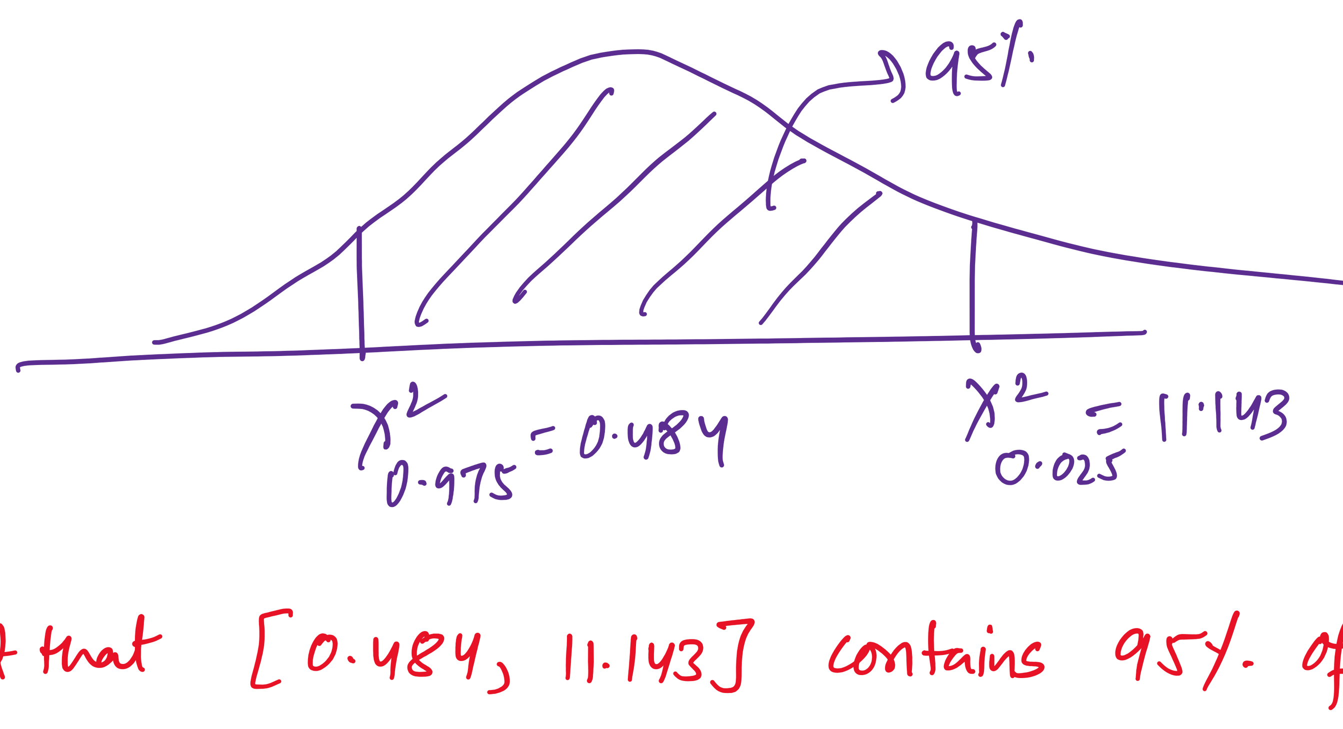
with degree of freedom $= 4$. $(5-1)$

Now we know that 95% χ^2 data lies between $\chi^2_{0.975}$ and $\chi^2_{0.025}$

Look at the table for these values in the row of degree of freedom $= 4$

we get the numbers $\chi^2_{0.975} = 0.484$

$\chi^2_{0.025} = 11.143$



So we get that $[0.484, 11.143]$ contains 95% of χ^2 data with dof $= 4$.

& the value of $\frac{(n-1)s^2}{\sigma^2}$ with assumed value of $\sigma = 1$, is 3.26 (from $\star$), which $\in [0.484, 11.143]$

$\therefore$ The assumed value of σ is correct.

Had the value of statistics $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$ lied outside the interval, we would have said that assumed value of σ^2 is wrong.