Stat 400, Lecture 25

Sampling from $N(\mu, \sigma^2)$ and the CLT

$\boxed{X \sim N(\mu, \sigma^2)}$ - - - → $X_1, X_2, \ldots, X_n$

Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a normal population. We have seen that we should use the <u>sample</u> mean $\overline{X}$ to estimate the population mean μ and the <u>sample</u> variance S^2 to estimate the population variance σ^2

$\overline{X}$ and S^2 are random variables.

<u>\$64,000 question</u>

How are $\overline{X}$ and S^2 distributed?

The answer is given by the following considerations

Any linear combination of independent normal random variables is again normal so $\overline{X}$ is normal. Since $E(\overline{X}) = \mu$ and

$$V(\overline{X}) = \frac{\sigma^2}{n}$$

we have

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Suppose $Z_1, Z_2, \ldots, Z_n$ are independent standard normal random variables. Then

$$Z_1^2 + \cdots + Z_n^2 \sim \chi^2(n) \quad (*)$$

chi-squared with n degrees of freedom

Now $Z_i = \dfrac{X_i - \mu}{\sigma} \sim N(0,1)$

So

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2 \sim \chi^2(n)$$

Now replace μ by its estimator $\overline{X}$

Rule of thumb - <u>every time you replace a quantity by its estimator you lose one degree of freedom in the chi-squared distribution</u>

So by the "rule of thumb"

$$Y = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \overline{X})^2 \sim \chi^2(n-1)$$

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

So $Y = \frac{n-1}{\sigma^2} S^2$ and we

obtain the critical

$$\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1) \qquad (**)$$

<u>Remark</u>

This isn't a proof because we used "the rule of thumb" but the result is true

Bottom Line

Theorem

Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal population with mean μ and variance σ^2. Then

(i) $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$

(ii) $\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$

(iii) $\overline{X}$ and S^2 are independent.

The above statement is an exact statement but if we take a large sample $(n > 30)$ from <u>any</u> population with mean μ and variance σ^2 we may assume to a good approximation that the population has $N(\mu, \sigma^2)$ distribution and we have by CLT

Theorem

If $X_1, X_2, \ldots, X_n$ is a large ($n > 30$) random sample from any population with mean μ and variance σ^2 then

(i) $\bar{X} \approx N(\mu, \frac{\sigma^2}{n})$

(ii) $S^2 \approx \chi^2(n-1)$

(iii) $\bar{X}$ and S^2 are approximately independent.
(they are not independent unless the population is normal)