STATS 2107
Statistical Modelling and Inference II
Welcome!

Matt Ryan

School of Mathematical Sciences, University of Adelaide

Semester 2 2021

Distribution of the sample variance - Part 1

The single normal sample

So far in Module 1, we have considered only the case of $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} N(\mu, \sigma^2)$, where σ^2 is known.

The key distributional result that underpins the hypothesis test and confidence intervals is that of Lemma 3:

$$\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)\,.$$

Case when σ^2 is not known

We use the sample variance

$$\boxed{S^2} = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2$$

in its place.

In this case, the key distribution result used in inference is

$$\boxed{\frac{\bar{Y}-\mu}{s/\sqrt{n}} \sim t_{n-1}\,.}$$

Case when σ^2 is not known

An outline of the proof of this result is as follows:

1. We first prove that S^2 is an unbiased estimator for σ^2. (Theorem 4)
2. We then derive the distribution of S^2 under the assumption $Y_i \sim N(\mu, \sigma^2)$. (Lemma 4, Definition 2.5, Theorem 5)
3. We find the distribution of $\frac{\bar{Y}-\mu}{s/\sqrt{n}}$. (Definition 2.6, Theorem 6)

Theorem 4

Suppose $Y_1, Y_2, \ldots, Y_n$ are independent random variables with

$$\mathrm{E}[Y_i] = \mu \quad \text{and} \quad \mathrm{var}(Y_i) = \sigma^2 \, ,$$

and let S^2 be the sample variance. Then

$$\mathrm{E}[S^2] = \sigma^2$$

Proof: See Assignment 1 QED

Remarks on Theorem 4

- We made no assumptions about the distribution of Y_i other than independence, $\mathrm{E}[Y_i] = \mu$ and $\mathrm{var}(Y_i) = \sigma^2$.
- Although S^2 is unbiased for σ^2, it is not true that S is unbiased for σ.

$$Var(S) = E[S^2] - E[S]^2$$

$$\Rightarrow E[S]^2 = E[S^2] - Var(S)$$

$$= \boxed{\sigma^2} - \boxed{Var(S)} > 0$$

$$\leq \sigma^2 \Rightarrow E[S] \leq \sigma$$

Lemma 4

Suppose $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} N(\mu, \sigma^2)$ random variables. Then $\bar{Y}$ and S^2 are independent.

Proof:

1. S^2 is a f^n of $(Y_1 - \bar{Y}, Y_2 - \bar{Y}, \ldots, Y_{n-1} - \bar{Y})$
2. $cov(\bar{Y}, Y_i - \bar{Y}) = 0$
3. Use (a) if X, Y are normal and $cov(X, Y) = 0$ X, Y are independent.

Proof of Lemma 4

3 b if X, Y are independent, then $f(X), g(Y)$ are independent.

1 $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$

$= \frac{1}{n-1} \left[\boxed{(Y_n - \bar{Y})^2} + \sum_{i=1}^{n-1} (Y_i - \bar{Y})^2 \right]$

$Y_n - \bar{Y}$? $\sum_{i=1}^{n} (Y_i - \bar{Y}) = \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} \bar{Y}$

$= n\bar{Y} - n\bar{Y} = 0$

$\therefore (Y_n - \bar{Y})^2 = \left[-\sum_{i=1}^{n-1} (Y_i - \bar{Y}) \right]^2$

$\therefore S^2 = \frac{1}{n-1} \left[\left[\sum_{i=1}^{n-1} (Y_i - \bar{Y}) \right]^2 + \sum_{i=1}^{n-1} (Y_i - \bar{Y})^2 \right]$

Proof of Lemma 4

(2) $\operatorname{cov}(\bar{Y}, Y_i - \bar{Y}) = 0$

$$\operatorname{cov}(\bar{Y}, Y_i - \bar{Y}) = \operatorname{cov}(\bar{Y}, Y_i) - \operatorname{cov}(\bar{Y}, \bar{Y})$$

$$= \operatorname{cov}\left(\frac{1}{n}\sum_{j=1}^{n} Y_j, Y_i\right) - \operatorname{var}(\bar{Y})$$

$$= \frac{1}{n}\sum_{j=1}^{n} \operatorname{cov}(Y_j, Y_i) - \frac{\sigma^2}{n}$$

$$= \frac{1}{n}(\sigma^2) - \frac{\sigma^2}{n} = 0.$$

$\operatorname{cov}(Y_j, Y_i) = \begin{cases} 0 & \text{if } i \neq j \\ \sigma^2 & \text{if } i = j \end{cases}$

by independence

(3) Since $\bar{Y}$, $Y_i - \bar{Y}$ are normal, they are independent.
Since S^2 is a f^n of the $(Y_i - \bar{Y})$, it is also independent of $\bar{Y}$! $\square$.

End video 1

Distribution of the sample variance - Part 2

Outline of the proof

1. We first prove that S^2 is an unbiased estimator for σ^2. (Theorem 4 ✓)
2. We then derive the distribution of S^2 under the assumption $Y_i \sim N(\mu, \sigma^2)$. (Lemma 4 ✓, Definition 2.5, Theorem 5)
3. We find the distribution of $\frac{\bar{Y}-\mu}{s/\sqrt{n}}$. (Definition 2.6, Theorem 6)

Lemma 4: If the data are normal

$\bar{Y}$, S^2 are independent.

Definition 2.5

Suppose $Z_1, Z_2, \ldots, Z_n \overset{iid}{\sim} N(0,1)$ random variables. Then the random variable

$$X = \sum_{i=1}^{n} Z_i^2$$

is said to have a χ^2-distribution with n degrees of freedom, written

$$\boxed{X \sim \chi_n^2 .}$$

Remarks on Definition 2.5

- The χ^2_n distribution can be proved to be the same as the gamma distribution with parameters $\frac{n}{2}$ and $\frac{1}{2}$. MGF.
- You should be able to show that the moment generating function of the $X \sim \chi^2_n$ distribution is

$$\boxed{M_X(t) = \frac{1}{(1-2t)^{n/2}}.}$$

Tutorial 1 Q2

$(1-2t)^{-\frac{n}{2}}$

$E[X] = n$

$Var(X) = 2n$

Theorem 5

Suppose $Y_1, Y_2, \ldots, Y_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$ random variables.

Then

$$\text{MGF} \longrightarrow \boxed{\frac{(n-1)S^2}{\sigma^2}} \sim \boxed{\chi^2_{n-1}.}$$

We are going to show that the MGF is χ^2_{n-1} $\left(\text{i.e. } (1-2t)^{-\frac{(n-1)}{2}}\right)$

Proof of Theorem 5

$$\sum_{i=1}^{n}(Y_i-\mu) = \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n}\mu$$

$$= n\bar{Y} - n\mu$$

$$= n(\bar{Y}-\mu)$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i-\mu-(\bar{Y}-\mu)\right)^2$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(Y_i-\mu)^2 - 2(\bar{Y}-\mu)\boxed{\sum_{i=1}^{n}(Y_i-\mu)} + \sum_{i=1}^{n}(\bar{Y}-\mu)^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(Y_i-\mu)^2 - \cancel{2}n(\bar{Y}-\mu)^2 + \cancel{n(\bar{Y}-\mu)^2}\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(Y_i-\mu)^2 - n(\bar{Y}-\mu)^2\right]$$

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\cancel{(n-1)}}{\boxed{\sigma^2}} \cdot \frac{1}{\cancel{(n-1)}}\left[\sum_{i=1}^{n}(Y_i-\mu)^2 - \boxed{n}(\bar{Y}-\mu)^2\right]$$

$$n = \frac{1}{\left[\frac{1}{\sqrt{n}}\right]^2}$$

Proof of Theorem 5

$$\frac{(n-1)S^2}{\sigma^2} = \underbrace{\sum_{i=1}^{n}\left(\frac{Y_i-\mu}{\sigma}\right)^2}_{\chi^2_n} - \underbrace{\left(\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}\right)^2}_{\chi^2_1}$$

So
$$\underbrace{\overbrace{\frac{(n-1)S^2}{\sigma^2}}^{M_1(t)}}_{\text{independent!}} + \underbrace{\left(\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}\right)^2}_{M_2(t)} = \sum_{i=1}^{n}\left(\frac{Y_i-\mu}{\sigma}\right)^2 \quad M_3(t)$$

If X, Y are independent
$M_{X+Y}(t) = M_X(t)M_Y(t)$

MGF

$\Rightarrow M_1(t)M_2(t) = M_3(t)$ by independence.

$M_1(t) \times (1-2t)^{-\frac{1}{2}} = (1-2t)^{-\frac{n}{2}}$

$\Rightarrow M_1(t) = (1-2t)^{-\frac{n}{2}+\frac{1}{2}} = (1-2t)^{-\frac{(n-1)}{2}}$

which the MGF of a χ^2_{n-1} $\square$