Math 170E

Lecture Notes Section 5.5 [*,†]

Sample mean and variance

Instructor: Swee Hong Chan

NOTE: Materials that appear in the textbook but do not appear in the lecture notes might still be tested. Please send me an email if you find typos.

[*] Version date: Friday 5th March, 2021, 13:19.

[†] This notes is based on March Boedihardjo's and Jamie Haddock's notes from the past quarters, and I would like to thank them for their generosity. "*Nanos gigantum humeris insidentes* (I am but a dwarf standing on the shoulders of giants)".

1 Sample mean and variance: Motivation

Let X be independent random variables with **unknown** mean μ and **unknown** variance σ^2.

We would like to determine the value of μ and σ^2.

Input: X_1, X_2, $\ldots X_n$ independent samples for X.

Output:

The sample mean $\overline{\mathrm{X}}$ is the "best guess" on what should be the mean μ is.

The sample variance S^2 is "the best guess" on what should be the variance σ^2.

2 Sample mean and variance: Definition

Let $X_1, \ldots, X_n$ be arbitrary random variables.

The **sample mean** $\overline{\mathrm{X}}$ is

$$\overline{\mathrm{X}} = \frac{X_1 + \ldots + X_n}{n}.$$

The **sample variance** S^2 is

$$S^2 = \frac{1}{n-1}\left[(X_1 - \overline{\mathrm{X}})^2 + (X_2 - \overline{\mathrm{X}})^2 + \ldots + (X_n - \overline{\mathrm{X}})^2\right].$$

These two concepts will be important in Math 170S for hypothesis testing.

3 Sample mean and variance: Before vs After

Samples $X_1, \ldots, X_n$ are just numbers **after** you perform the experiment.

However, samples $X_1, \ldots, X_n$ are random variables **before** you perform the experiment.

Thus sample mean $\overline{\mathrm{X}}$ and sample variance S^2 are deterministic numbers **after** you perform the experiments, and are random variables **before** you perform the experiments.

4 Sample mean, variance for normals: Theorem

Theorem 1. *Let* $X_1, X_2 \ldots, X_n$ *be independent normal random variables with mean* μ *and variance* σ^2. *Then*

- $\overline{\mathrm{X}}$ *is the normal random variable with mean* μ *and variance* σ^2/n*;*
- S^2 *is the chi-square random variable* $\chi^2(\sigma^2)$ *with mean* σ^2*;*
- $\overline{\mathrm{X}}$ *and* S^2 *are independent.*

5 Sample mean and variance: Example

Let $X_1, \ldots, X_5$ be independent standard normal RVs. Compute the probability

$$P(X_1 + \ldots + X_5 \le 5,\ X_1^2 + \ldots + X_5^2 \le 10.824).$$

Answer: We have

$$\begin{aligned}
& P(X_1 + \ldots + X_5 \le 5,\ X_1^2 + \ldots + X_5^2 \le 10.824) \\
= \ & P\left(\frac{X_1 + \ldots + X_5}{5} \le 1,\ \frac{X_1^2 + \ldots + X_5^2}{4} \le 2.706\right) \\
= \ & P(\overline{\mathrm{X}} \le 1, S^2 \le 2.706) \\
= \ & P(\overline{\mathrm{X}} \le 1)\ P(S^2 \le 2.706) \qquad \text{(by Theorem 1)} \\
= \ & (0.8413)(0.9) \qquad \text{(by Table Va and IV)} \\
= \ & 0.75717.
\end{aligned}$$

6 Sampling error: real mean

Let $X_1, X_2, \ldots, X_n$ be independent normals with mean μ and variance σ^2.

Let U be the random variables

$$U = \frac{(X_1-\mu)^2}{\sigma^2} + \frac{(X_2-\mu)^2}{\sigma^2} + \ldots + \frac{(X_n-\mu)^2}{\sigma^2}.$$

U measures the error of sampling $X_1, \ldots X_n$, with respect to the **real mean** μ.

We know from the moment generating function section (Section 5.4) that U is the chi-square random variable $\chi^2(n)$.

7 Sampling error: sample mean

Let W be the random variables

$$W = \frac{(X_1 - \overline{\mathrm{X}})^2}{\sigma^2} + \frac{(X_2 - \overline{\mathrm{X}})^2}{\sigma^2} + \ldots + \frac{(X_n - \overline{\mathrm{X}})^2}{\sigma^2},$$

where every instance of **real mean** μ in U is replaced with **sample mean** $\overline{\mathrm{X}}$.

W measures the error of sampling $X_1, \ldots X_n$, with respect to the **sample mean** $\overline{\mathrm{X}}$.

We know from the previous theorem that W is the chi-square random variable $\chi^2(n-1)$.

The loss of one degree of freedom for W is because of the **bias** created when using **sample mean**.

In statistics, we are usually more interested in W, since the **real mean** μ is usually unknown, while **sample mean** $\overline{\mathrm{X}}$ can be computed from the samples.

8 Student's t distribution: Example

Let $X_1, \ldots, X_{12}$ be independent standard normal RVs. Let T be given by

$$T = \frac{\overline{\mathrm{X}}}{\sqrt{S^2}},$$

where $\overline{\mathrm{X}}$ is the sample mean, and S^2 is the sample variance.

Find c such that

$$P(T \leq c) = 0.975.$$

9 Student's t distribution

Theorem 2. *Let Z be the standard normal random variable, and let U be the $\chi^2(r)$ random variable such that Z and U are independent. Let*

$$T = \frac{Z}{\sqrt{U/r}}.$$

Then T has support and pdf given by

$$\begin{aligned} \textit{Support} &= (-\infty, \infty); \\ f(t) &= \frac{\Gamma((r+1)/2)}{\sqrt{r\,\pi}\,\Gamma(r/2)} \frac{1}{(1+t^2/r)^{(r+1)/2}}. \end{aligned}$$

T is called the **student's t distribution** with r degrees of freedom.

We usually use Table VI in Appendix B to calculate the cdf of T.

10 Student's t distribution: Answer

By Theorem 1, $\overline{\mathrm{X}}$ is standard normal RV, S^2 is $\chi^2(1)$ RV, and $\overline{\mathrm{X}}$ and S^2 are independent.

Let $Z = \overline{\mathrm{X}}$, let $U = (n-1)S^2$, and let $r = n-1$.

Then Z is standard normal, U is $\chi^2(r)$ RV, Z and U are independent, and

$$T = \frac{\overline{\mathrm{X}}}{\sqrt{S^2}} = \frac{Z}{\sqrt{U/r}}.$$

By Theorem 2, T is student's t distribution with $r = 11$ degrees of freedom.

By Table VI, it then follows that $c = 2.201$.