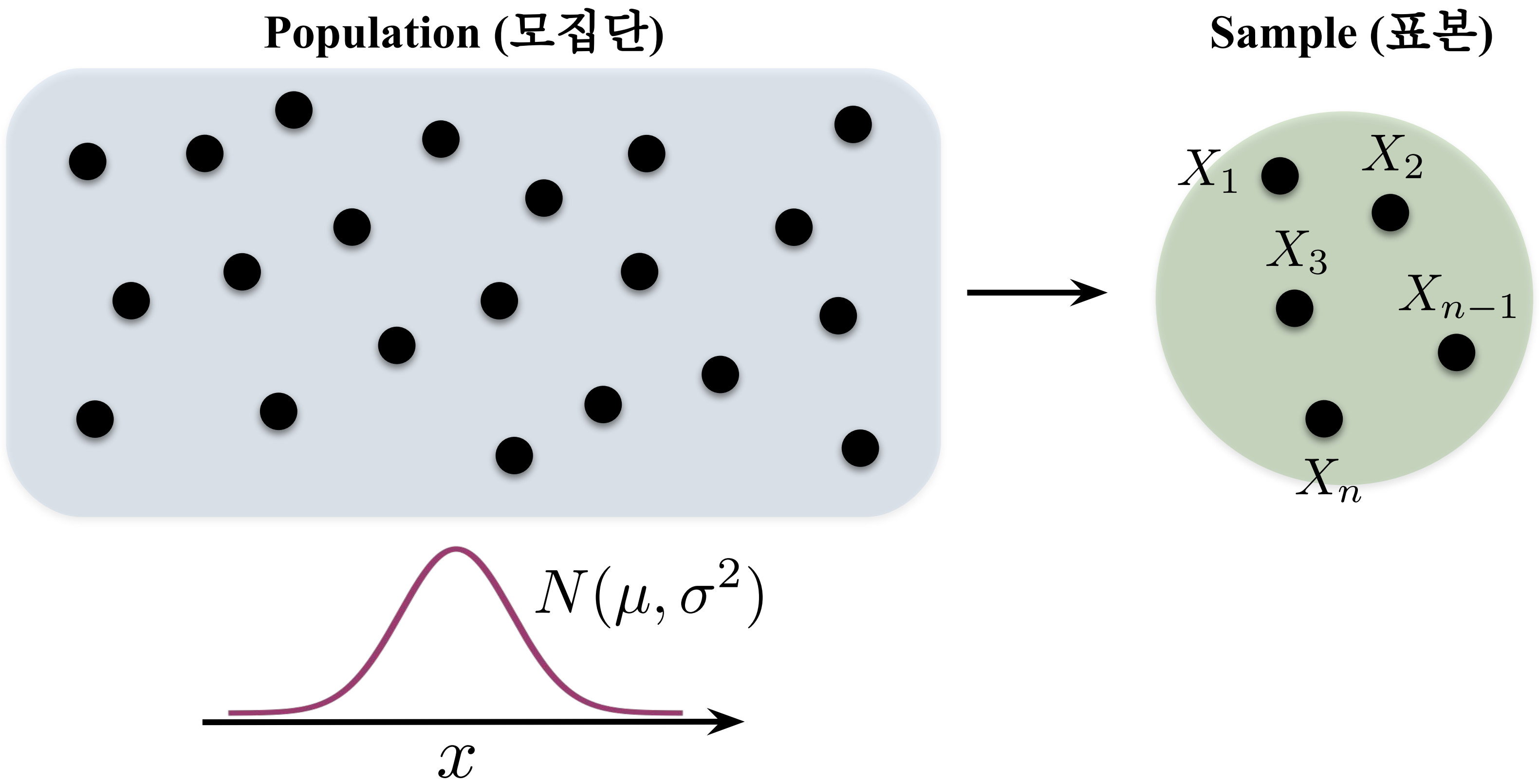
통계분석
Statistical Analysis

Random Sample from Normal Distributions

$X_1, X_2, \cdots, X_n$: Random sample from $N(\mu, \sigma^2)$

- Every X_i follows a normal distribution $N(\mu, \sigma^2)$.
- $X_1, X_2, \cdots, X_n$ is independent of one another.



Random Sample from Normal Distributions

$X_1, X_2, \cdots, X_n$: Random sample from $N(\mu, \sigma^2)$

- $\boxed{\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i}$ is normally distributed $\sim N(\mu, \sigma^2/n)$

  sample mean

- As n increases, the normal distribution of $\bar{X}$ becomes sharper.

- Any linear combination of $X_1, \cdots, X_n$ is normally distributed.

$$a_1X_1 + a_2X_2 + \cdots + a_nX_n \sim N(\mu', \sigma'^2)$$

NOT APPROXIMATION, BUT **EXACT**
Central Limit Theorem NOT APPLIED HERE
NOT PROVED; RESULTS JUST GIVEN

Q. What is the distribution of sample variance? We haven't yet discussed so far.

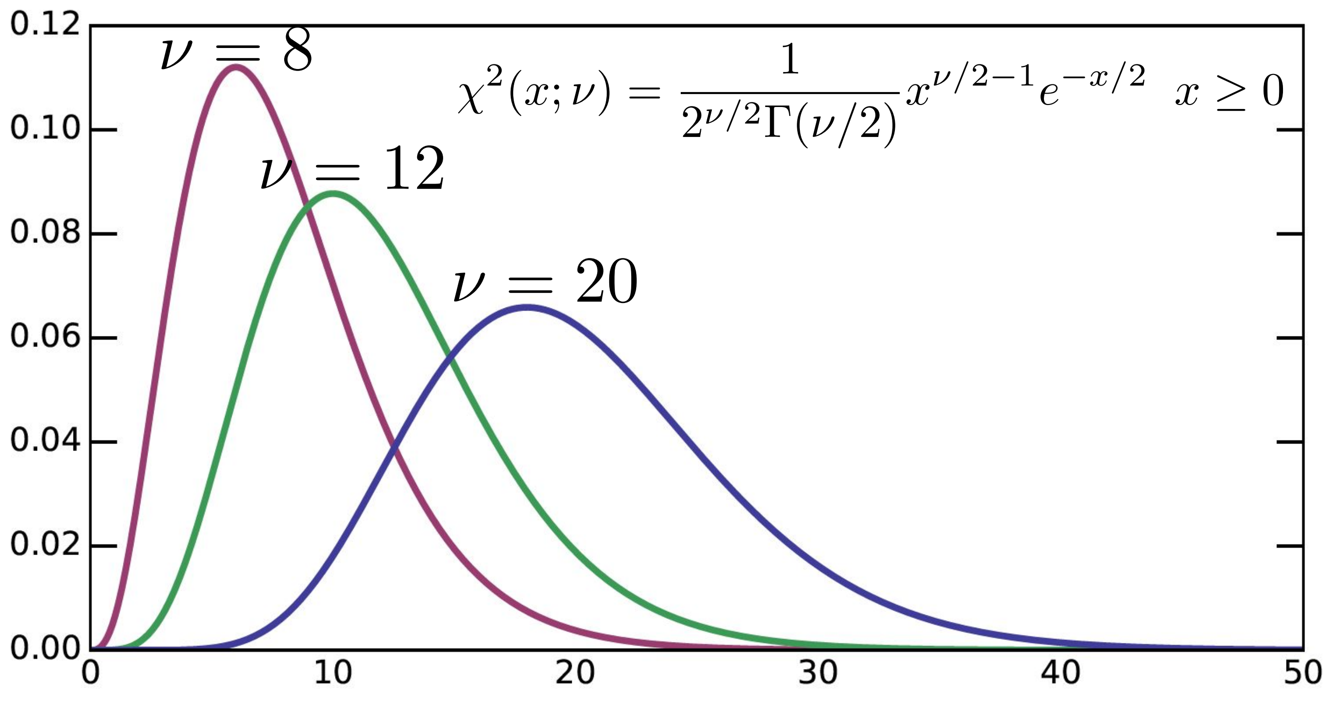
Random Sample from Normal Distributions:
Distribution of Sample Variances

$X_1, X_2, \cdots, X_n$: Random sample from $N(\mu, \sigma^2)$

- What is the distribution of the sample variance?

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

- It is related to the ***chi-squared distribution***!



Sample Mean and Sample Variance of Normal Random Samples**

$$X_1, X_2, \cdots, X_n : \text{Random sample from } N(\mu, \sigma^2)$$

- $\bar{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n)$

- $Y = \displaystyle\sum_{i=1}^{n} \dfrac{\left(X_i - \bar{X}\right)^2}{\sigma^2} \sim \chi^2(\nu = n-1)$

Note that the degree of freedom ν is NOT n, but n-1.

$\rightarrow$ What does the degree of freedom mean?

[DEFINITION] **A ~ B** means that the random variable A follows the probability distribution B.

NOTE 01. The random variable Y has different prefactor, compared with the sample variance.
<u>NOTE 02. Y follows chi-squared distribution of df(degree of freedom) n-1.</u>

Q. What is the degree of freedom?

Degree of Freedom

- **the degree of freedom = the number of independent variables**

Example I : $X_1,\ X_2$ without any relation between them

$\rightarrow X_1,\ X_2$ are independent.

The degree of freedom is two.

Example II : $X_1,\ X_2$ satisfying $X_2 = f(X_1)$

$\rightarrow$ When X_1 is known, $f(X_1)$ determines X_2.

$\rightarrow X_1,\ X_2$ are NOT independent.

The degree of freedom is one.

Example III : $X_1,\ X_2,\ X_3$ satisfying

$$f_1(X_1, X_2, X_3) = 2X_1 + 3X_2 + X_3 - 1 = 0$$

$$f_2(X_1, X_2, X_3) = 3X_1 + X_2 + X_3 - 3 = 0$$

Degree of freedom = (the number of variables) – (the number of relations) = 3 - 2 = 1

Sample Variance of Normal Random Samples

$X_1, X_2, \cdots, X_n$: Random sample from $N(\mu, \sigma^2)$

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

$(X_1 - \bar{X}), (X_2 - \bar{X}), \cdots, (X_n - \bar{X})$ are also n random variables.

But, they are NOT independent of one another.

One relation among them: $\displaystyle\sum_{i=1}^{n} (X_i - \bar{X}) = 0$

from the expression of sample mean

$\longrightarrow$ Among $(X_1 - \bar{X}), (X_2 - \bar{X}), \cdots, (X_n - \bar{X})$,

there is $(n-1)$ independent variables.

$\nu =$ degree of freedom $=$ the number of independent variables

Sample Variance of Normal Random Samples

$X_1, X_2, \cdots, X_n$: Random sample from $N(\mu, \sigma^2)$

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

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

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

$$S^2 \sim \chi^2(n-1) \text{ [WRONG!]}$$

DO NOT FORGET THE PREFACTOR!!

Sample Mean and Sample Variance of Normal Random Samples**

$X_1, X_2, \cdots, X_n$: Random sample from $N(\mu, \sigma^2)$

- $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n)$

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

  Note that the degree of freedom ν is NOT n, but n-1.

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

Student's *t* distribution from Normal Random Sample

$$X_1, X_2, \cdots, X_n : \text{Random sample from } N(\mu, \sigma^2)$$

$$\text{What is the distribution of } T = \frac{\bar{X} - \mu}{S/\sqrt{n}} ?$$

Here we define a new random variable T, which consists of sample mean, variance, population mean and sample size.

Q1. What is the distribution of T? : **Student's t distribution [DISCUSSED HERE]**

Q2. Why is this distribution and this variable useful? : Can use for population mean estimation (inference) **[DISCUSSED NEXT TIME]**

Student's *t* distribution from Normal Random Sample

$$X_1, X_2, \cdots, X_n : \text{Random sample from } N(\mu, \sigma^2)$$

$$\text{What is the distribution of } \boxed{T = \frac{\bar{X} - \mu}{S/\sqrt{n}}?}$$

(1)

- $\bar{X} \sim N(\mu, \sigma^2/n) \xrightarrow{\textit{standardization}} Z \sim N(0,1)$

$$\boxed{Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}$$

(2) $Y = \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2 = \dfrac{n-1}{\sigma^2}\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2 = \dfrac{n-1}{\sigma^2}S^2$

- $\dfrac{\bar{X} - \mu}{S/\sqrt{n}} = \dfrac{Z}{\sqrt{Y/(n-1)}}$

1. Using (1) and (2), we can show that T variable can be written in term of variables Z and Y.
2. Z ~ N(0,1), Y ~ chi2(n-1)

Student's *t* distribution from Normal Random Sample

$X_1, X_2, \cdots, X_n$: Random sample from $N(\mu, \sigma^2)$

- $\bar{X} \sim N(\mu, \sigma^2) \longrightarrow Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$
- $Y = \dfrac{n-1}{\sigma^2} S^2 \sim \chi^2(\nu = n-1)$
- Further, it is known that Z and Y are independent.
- When two independent random variables X_1 and X_2 satisfy

$$X_1 \sim \chi^2(m) \text{ and } X_2 \sim N(0,1),$$

then, the random variable $X = \dfrac{X_2}{\sqrt{X_1/m}} \sim t(m)$

t-distribution with d.f. m

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{Z}{\sqrt{Y/(n-1)}}$$

NOT PROVED HERE, KNOWN FACT

Student's *t* distribution from Normal Random Sample

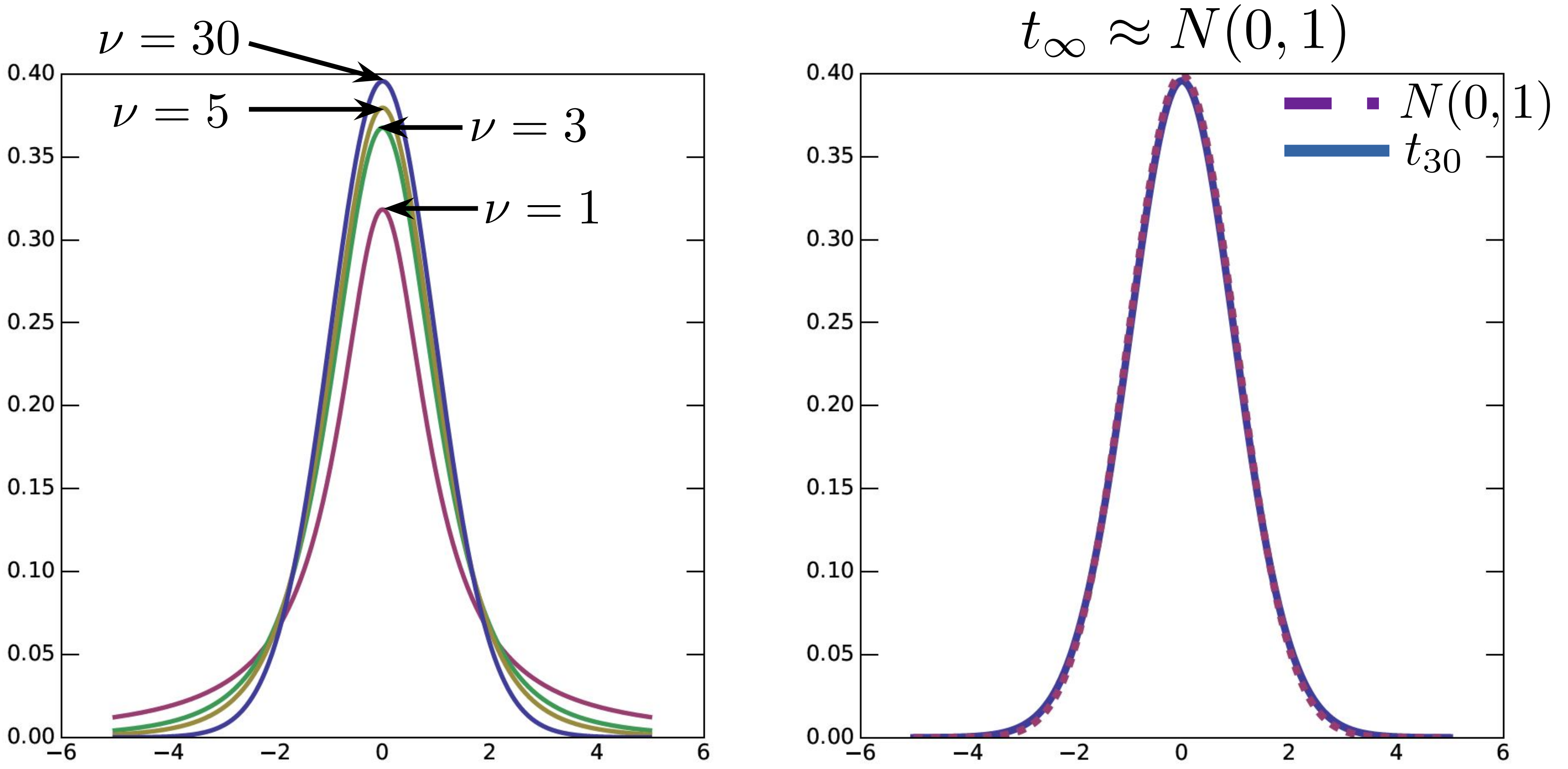
$X_1, X_2, \cdots, X_n$: Random sample from $N(\mu, \sigma^2)$

- $\bar{X} \sim N(\mu, \sigma^2) \longrightarrow Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$
- $Y = \dfrac{n-1}{\sigma^2} S^2 \sim \chi^2(\nu = n-1)$
- Further, it is known that Z and Y are independent.

$$\rightarrow T = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{Z}{\sqrt{Y/(n-1)}} \sim t(n-1)$$

Recall: Student's *t*-distribution

$$f(t;\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} = t_\nu$$



Distribution of *T* for very large *n*

$$X_1, X_2, \cdots, X_n : \text{Random sample from } N(\mu, \sigma^2)$$

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1) \approx N(0,1) \text{ when n is very large}$$

Distribution of T

$$\text{Why is } T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \text{ useful?}$$

Sample mean

Population mean (unknown)

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

Sample variance ***The number of elements in the sample***

- Once you take a sample from the population, you know ***the number of elements in the sample (sample size n)***, and you can calculate ***the sample mean***, and ***the sample variance.***

- Using ***three known values*** and ***the T variable***, you can ***estimate the population mean***.
= Answer for Q2.