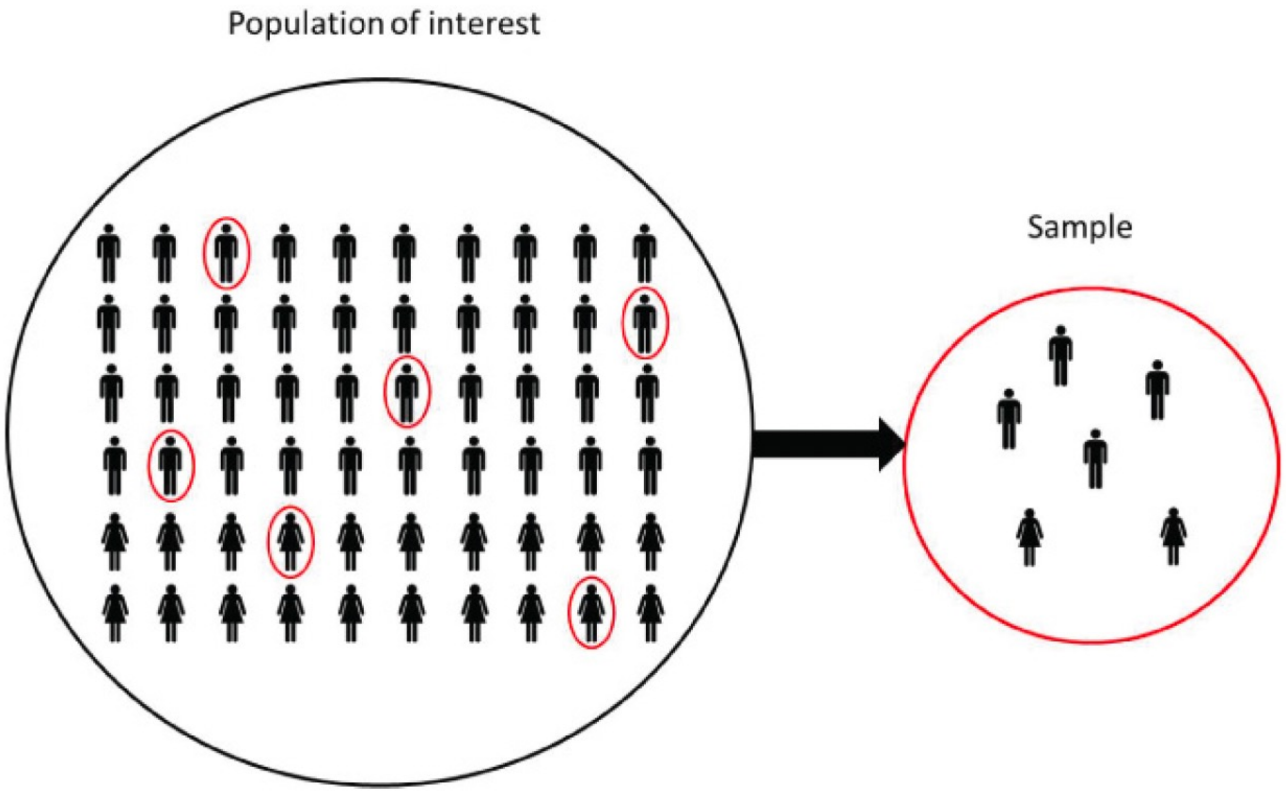
Population mean vector can be estimated from sample mean vector

01/03/2023

Jing Qin

Re-cap: univariate population and sample



$$Y \longrightarrow \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{pmatrix} \longrightarrow \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix}$$

Y_1, Y_2, Y_3, Y_4, Y_5 forms a *random sample* of $Y \sim$ Bernoulli (μ), where μ is unknown.

It takes a *vector* of (univariate) observations to estimate the distribution of a univariate random variable (r.v.).

Re-cap: univariate case: estimator and estimation

(Univariate) Y_1, Y_2, Y_3, Y_4, Y_5 forms a *random sample* of Y and

$$E(Y) = \mu$$

missing

Then we use sample mean $\bar{Y} = \frac{1}{5}\,(Y_1 + Y_2 + Y_3 + Y_4 + Y_5)$ as

Formula

an *estimator* to estimate μ

WHY?

The *estimation* is given by $\bar{y} = \frac{1}{5}\,(y_1 + y_2 + y_3 + y_4 + y_5)$

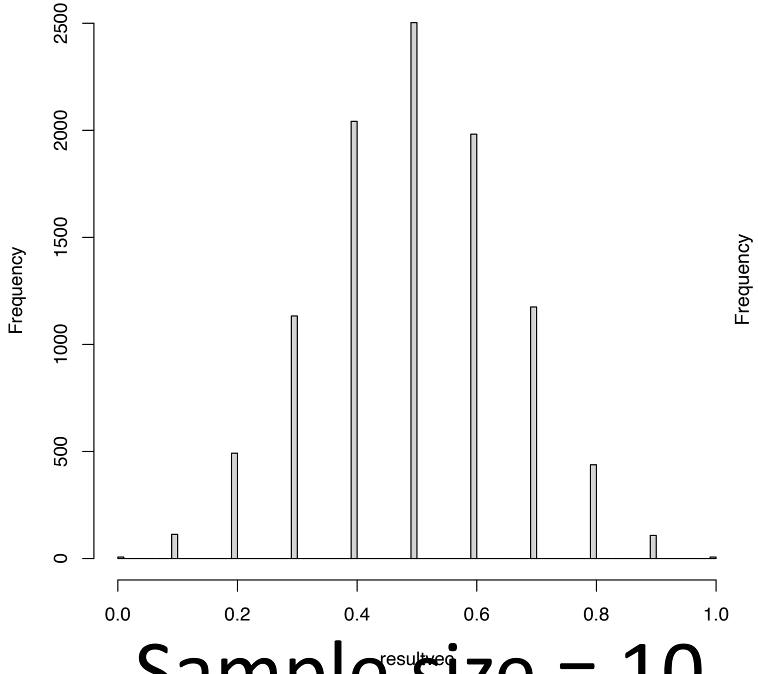
Number

Notice in general case: $\bar{Y} = \sum_{i=1}^{n} Y_i$ and $\bar{y} = \sum_{i=1}^{n} y_i$

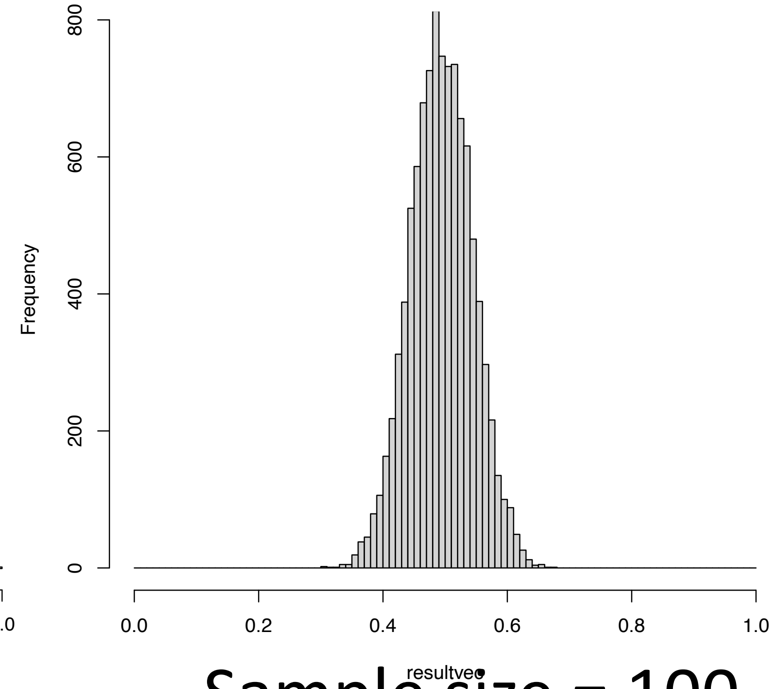
Re-cap: CLT Central limit theorem (uni-variate)

Statement:

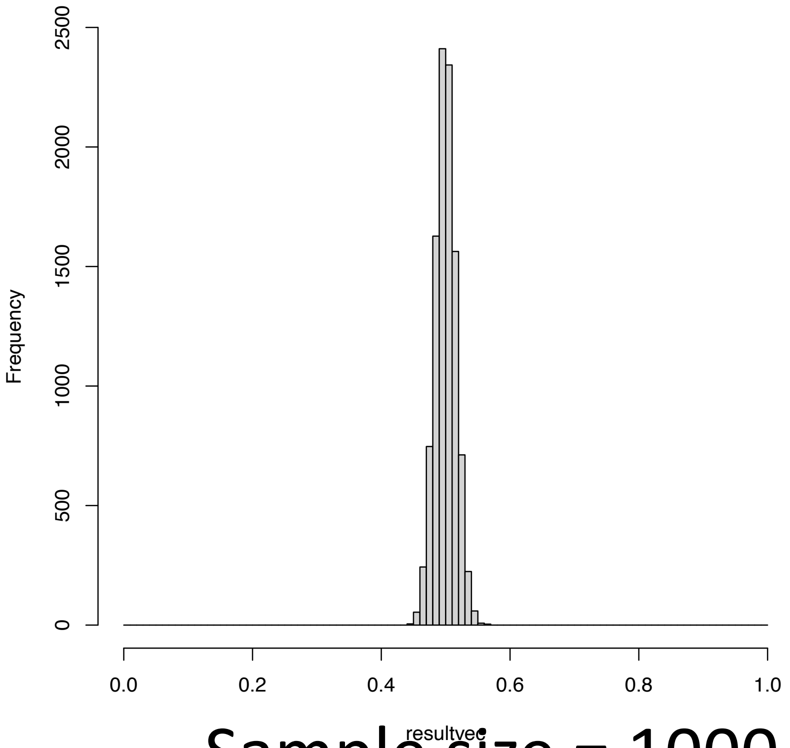
$$\bar{Y} \overset{\text{approx}}{\sim} N\left(\mu, \frac{Var(Y)}{n}\right)$$



Sample size = 10



Sample size = 100



Sample size = 1000

Result 4.13 (The central limit theorem). Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be independent observations from any population with mean $\boldsymbol{\mu}$ and finite covariance $\boldsymbol{\Sigma}$. Then

$$\sqrt{n}\,(\overline{\mathbf{X}} - \boldsymbol{\mu}) \text{ has an approximate } N_p(\mathbf{0}, \boldsymbol{\Sigma}) \text{ distribution}$$

for large sample sizes. Here n should also be large relative to p.

$$\overline{\boldsymbol{X}} \overset{\text{approx}}{\sim} \boldsymbol{N}_{\boldsymbol{p}}\left(\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{n}\right)$$

Terminology comes later

multivariate case

We need a $(n \times p)$ data matrix to estimate a p-dim random vector

$$\begin{array}{cccccc} X_1 & X_2 & \cdots & X_k & \cdots & X_p \\ \end{array}$$

$$\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1k} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2k} & \cdots & X_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ X_{j1} & X_{j2} & \cdots & X_{jk} & \cdots & X_{jp} \\ \vdots & \vdots & & \vdots & \cdots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nk} & \cdots & X_{np} \end{bmatrix}$$

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2k} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{j1} & x_{j2} & \cdots & x_{jk} & \cdots & x_{jp} \\ \vdots & \vdots & & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} & \cdots & x_{np} \end{pmatrix}$$

Compare to univariate

$$Y \longrightarrow \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{pmatrix} \longrightarrow \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix}$$

Multi-varaite: Expected value of random matrix

- (2-23) Expected value of the random matrix $\boldsymbol{X}$ is a matrix as well:

$$E(\boldsymbol{X}) = \begin{pmatrix} E(X_{11}) & E(X_{12}) & \cdots & E(X_{1k}) & \cdots & E(X_{1p}) \\ E(X_{21}) & E(X_{22}) & \cdots & E(X_{2k}) & \cdots & E(X_{2p}) \\ \vdots & \vdots & & \vdots & & \vdots \\ E(X_{j1}) & E(X_{j2}) & \cdots & E(X_{jk}) & \cdots & E(X_{jp}) \\ \vdots & \vdots & & \vdots & \cdots & \vdots \\ E(X_{n1}) & E(X_{n2}) & \cdots & E(X_{nk}) & \cdots & E(X_{np}) \end{pmatrix}$$

$$E(\boldsymbol{X}) = (E(X_{j,k}))_{(n \times p)}$$

Population mean vector $\boldsymbol{\mu}$

When the rows are independent and identially distributed random vectors

$$\begin{array}{c} \begin{array}{cccccc} X_1 & X_2 & \cdots & X_k & \cdots & X_p \end{array} \\ \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1k} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2k} & \cdots & X_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ X_{j1} & X_{j2} & \cdots & X_{jk} & \cdots & X_{jp} \\ \vdots & \vdots & & \vdots & \cdots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nk} & \cdots & X_{np} \end{bmatrix} \end{array}$$

$$\begin{pmatrix} E(X_{11}) = \mu_1 & E(X_{12}) = \mu_2 & \cdots & E(X_{1p}) = \mu_p \\ E(X_{21}) = \mu_1 & E(X_{22}) = \mu_2 & \cdots & E(X_{2p}) = \mu_p \\ \vdots & \vdots & \vdots & \vdots \\ E(X_{j1}) = \mu_1 & E(X_{j2}) = \mu_2 & \cdots & E(X_{jp}) = \mu_p \\ \vdots & \vdots & \vdots & \cdots \\ E(X_{n1}) = \mu_1 & E(X_{n2}) = \mu_2 & \cdots & E(X_{np}) = \mu_p \end{pmatrix}$$

$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}$ is referred to as (population) mean vector

Sample mean (as an estimator) $\bar{\boldsymbol{X}}$ (3–8)

$\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_n$ form a *random sample of size n* on $\boldsymbol{X}$ and $E(\boldsymbol{X}) = \boldsymbol{\mu}$

$$
\begin{array}{c}
\begin{array}{cccccc}
X_1 & X_2 & \cdots & X_k & \cdots & X_p
\end{array} \\
\begin{bmatrix}
X_{11} & X_{12} & \cdots & X_{1k} & \cdots & X_{1p} \\
X_{21} & X_{22} & \cdots & X_{2k} & \cdots & X_{2p} \\
\vdots & \vdots & & \vdots & & \vdots \\
X_{j1} & X_{j2} & \cdots & X_{jk} & \cdots & X_{jp} \\
\vdots & \vdots & & \vdots & \cdots & \vdots \\
X_{n1} & X_{n2} & \cdots & X_{nk} & \cdots & X_{np}
\end{bmatrix}
\end{array}
=
\begin{pmatrix}
\boldsymbol{X}_1' \\ \boldsymbol{X}_2' \\ \vdots \\ \boldsymbol{X}_j' \\ \vdots \\ \boldsymbol{X}_n'
\end{pmatrix}
$$

$\bar{\boldsymbol{X}}$ in its matrix form

$$\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1k} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2k} & \cdots & X_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ X_{j1} & X_{j2} & \cdots & X_{jk} & \cdots & X_{jp} \\ \vdots & \vdots & & \vdots & \cdots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nk} & \cdots & X_{np} \end{bmatrix}$$

$$\frac{1}{n}\sum_{i=1}^{n} X_{i1} \quad \frac{1}{n}\sum_{i=1}^{n} X_{i2} \quad \frac{1}{n}\sum_{i=1}^{n} X_{ik} \quad \frac{1}{n}\sum_{i=1}^{n} X_{ip}$$

$$\bar{\boldsymbol{X}} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} X_{i1} \\ \frac{1}{n}\sum_{i=1}^{n} X_{i2} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} X_{ik} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} X_{ip} \end{pmatrix} = \frac{1}{n} \cdot \begin{pmatrix} \sum_{i=1}^{n} X_{i1} \\ \sum_{i=1}^{n} X_{i2} \\ \vdots \\ \sum_{i=1}^{n} X_{ik} \\ \vdots \\ \sum_{i=1}^{n} X_{ip} \end{pmatrix}$$

$$= \frac{1}{n} \cdot \sum_{i=1}^{n} \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{ik} \\ \vdots \\ X_{ip} \end{pmatrix} = \boxed{\frac{1}{n} \cdot \sum_{i=1}^{n} \boldsymbol{X}_i}$$

Now, finally add them together

(2-24) $E(\boldsymbol{X} + \boldsymbol{Y}) = E(\boldsymbol{X}) + E(\boldsymbol{Y})$

$$\overline{\boldsymbol{X}} = \frac{1}{n} \cdot \sum_{i=1}^{n} \boldsymbol{X}_i = \frac{1}{n} \cdot \boldsymbol{X}_1 + \frac{1}{n} \cdot \boldsymbol{X}_2 + \cdots + \frac{1}{n} \cdot \boldsymbol{X}_n$$

We have

$$E(\overline{\boldsymbol{X}}) = \frac{1}{n} \cdot \sum_{i=1}^{n} E(\boldsymbol{X}_i) = \frac{1}{n} \cdot \sum_{i=1}^{n} \boldsymbol{\mu} = \boldsymbol{\mu}$$

Population covariance matrix $\mathbf{\Sigma}$

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)'$.

- Its covariance matrix is defined in (2-31) and denoted by Σ or Cov($\boldsymbol{X}$).

$$\Sigma = \begin{pmatrix} \text{Cov}(X_1, X_1) = \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_p) \\ \text{Cov}(X_2, X_1) & \text{Cov}(X_2, X_2) & \cdots & \text{Cov}(X_2, X_p) \\ \vdots & \vdots & \vdots & \vdots \\ \text{Cov}(X_p, X_1) & \text{Cov}(X_p, X_2) & \cdots & \text{Cov}(X_p, X_p) \end{pmatrix}$$

- Recall that $\text{Cov}(X_1, X_2) = E[(X_1 - \mu_1)(X_2 - \mu_2)]$, therefore

$$\text{Cov}(X_1, X_2) = \text{Cov}(X_2, X_1)$$

and thus Σ is symmetric.

Properties which we will need later

- Useful results

uncorrelated

1. $\text{Cov}(X_j, X_j) = \text{Var}(X_j)$
2. $\text{Cov}(X_i, X_j) = 0$ does not indicate two random variables are independent.
3. $\text{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$

- (2-32) $\text{Cov}(\boldsymbol{X}) = E[(\boldsymbol{X} - \boldsymbol{\mu}) \cdot (\boldsymbol{X} - \boldsymbol{\mu})']$

Try it out

- (2-45) Let $\boldsymbol{C}$ be a numeric $(q \times p)$ matrix, then we have

Exercise 2.41 needs this

$$\boxed{E(\boldsymbol{C} \cdot \boldsymbol{X}) = \boldsymbol{C} \cdot \boldsymbol{\mu}}$$

and

$$\boxed{\text{Cov}(\boldsymbol{C} \cdot \boldsymbol{X}) = \boldsymbol{C} \cdot \Sigma \cdot \boldsymbol{C}'.}$$

$$(Result\ \ 3.1) \quad Cov(\overline{\boldsymbol{X}}) = \frac{1}{n}\Sigma$$

Result 4.13 (The central limit theorem). Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be independent observations from any population with mean $\boldsymbol{\mu}$ and finite covariance $\boldsymbol{\Sigma}$. Then

$$\sqrt{n}\,(\overline{\mathbf{X}} - \boldsymbol{\mu}) \text{ has an approximate } N_p(\mathbf{0}, \boldsymbol{\Sigma}) \text{ distribution}$$

for large sample sizes. Here n should also be large relative to p.

$$\overline{\boldsymbol{X}} \overset{\text{approx}}{\sim} \boldsymbol{N}_{\boldsymbol{p}}\left(\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{n}\right)$$