① Sampling → Parameter-free

a) Random Sampling $\Longrightarrow$ $\boxed{\text{i.i.d}}$ $X_1, \dots, X_n \sim$ population $f(x)$

{ independent
identical distributed

Infinite Population { with replacement
without replacement

finite Population ← with replacement

b) Finite Population

$f(x)$

$\downarrow$ sample

$\{x_1, \dots, x_N\} \Rightarrow$ Population

$$\boxed{\hat{F}_N(t) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\{X_i \le t\}}$$

Empirical CDF (ECDF)

{ with replacement $\Rightarrow$ i.i.d
without replacement $\Rightarrow$ { independent ✗
identically distributed ✓

when N Big. 弱相关

简单随机抽样

$\Rightarrow X_1, \dots, X_n$ 近似独立.

only care about the 数字特征 of the population distribution

② Statistics & Sampling Distribution

$$\boxed{T = T(X_1, \dots, X_n)} \xrightarrow{\text{distribution}}$$

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

$$\boxed{f(x) \longrightarrow \begin{cases} \mu \\ \sigma^2 \end{cases}}$$

$$\begin{cases} E[\bar{X}] = \mu \\ E[S^2] = \sigma^2 \end{cases}$$

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

$= X^T \Lambda X$

$$\boxed{\begin{aligned} &\Lambda = (I_n - \tfrac{1}{n}\mathbb{1}\mathbb{1}^T) \\ &\operatorname{tr}\Lambda = n-1 \end{aligned}}$$

$$\boxed{\sum_{i=1}^{n} (X_i - \bar{X}) = 0}$$

$$S^2 = \frac{1}{n-1}\left[\sum_{i=2}^{n} (X_i - \bar{X})^2 + \boxed{(X_1 - \bar{X})^2}\right]$$

③ Sampling of Normality

a) $\left[\begin{array}{l} 1.\ \bar{X} \perp\!\!\!\perp S^2 \Leftarrow \text{by joint distribution of } \bar{X}, X_2 - \bar{X}, \dots, X_n - \bar{X} \\ 2.\ \bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \end{array}\right.$

$(X_i - \bar{X})^2$

3. $\boxed{\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)}$

$LHS = \sum_{i=1}^{n} \dfrac{(X_i-\bar{X})^2}{\sigma^2}$ U

$= \sum_{i=1}^{n} \dfrac{(X_i-\mu)^2}{\sigma^2} - \dfrac{(\bar{X}-\mu)^2}{\frac{\sigma^2}{n}}$ V

$\Rightarrow \begin{cases} ① \ U+V \sim \chi^2(n) \\ ② \ U \perp V, \ V \sim \chi^2(1) \end{cases} \Rightarrow \boxed{U \sim \chi^2(n-1)}$

⊗ another method

from $X^TAX \sim \chi^2(p)$

$\Leftrightarrow \begin{cases} r(A)=p \quad A^2=A \\ A\cdot\mu=0 \end{cases}$

b) t-dist & F-dist (also for Normal Population)

1. when σ^2 is unknown, consider

r.v. $\dfrac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} = \dfrac{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{S^2}{\sigma^2}}} \to \dfrac{Z(0,1)}{\sqrt{\frac{\chi^2(n-1)}{n-1}}} \Rightarrow t(n-1)$

(r.v.) t-distribution with n-1 degree of freedom

$\boxed{T} = \dfrac{U}{\sqrt{\frac{V}{n-1}}} \sim t_{n-1}$ where $\begin{cases} U \sim Z(0,1) \\ V \sim \chi^2(n-1) \end{cases}$

pdf $\to \boxed{f(t) \propto \left(1+\frac{1}{n-1}t^2\right)^{-\frac{n}{2}}}$ 自由度为n-1

→ achieved by joint distribution (U,V) & $\dfrac{U}{\sqrt{\frac{V}{n-1}}}$

Basis: F-statistics

2. F-dist ⇒ 方差变异性 (for Normal Population)

consider $\dfrac{\frac{S_x^2}{\sigma_x^2}}{\frac{S_y^2}{\sigma_y^2}}$, $\dfrac{S_x^2}{\sigma_x^2} \to \dfrac{\chi^2_{n-1}}{n-1}$, $\dfrac{S_y^2}{\sigma_y^2} \to \dfrac{\chi^2_{m-1}}{m-1}$ → $\boxed{F(n-1, m-1)}$

用 $\dfrac{S_x^2}{S_y^2}$ 来衡量 $\dfrac{\sigma_x^2}{\sigma_y^2}$

$\boxed{\text{Property}}$

① $X \sim F_{p,q} \Rightarrow X^{-1} \sim F_{q,p}$

② $X \sim t_q = \dfrac{u}{\sqrt{\frac{V}{q}}} \Rightarrow X^2 = \dfrac{u^2}{\frac{V}{q}} \sim F_{1,q}$

③ $X \sim F_{p,q} \Rightarrow \dfrac{\frac{p}{q}X}{1+\frac{p}{q}X} \sim \text{Beta}\left(\frac{p}{2}, \frac{q}{2}\right)$

③ Order Statistics

$X_{(i)} \longrightarrow \boxed{\text{i-th order}}$ in $\{X_1, \ldots, X_n\}$

$\downarrow$ cdf / pdf

$$f_{X_{(i)}}(x) = \binom{n}{i-1 \quad 1 \quad n-i} \cdot f_X(x) \cdot F_X(x)^{i-1} \cdot (1 - F_X(x))^{n-i}$$

④ LLN + CLT

$\Rightarrow$ give the convergence result

$$\boxed{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \text{ asymptotically}}$$

$\Rightarrow$ But we do not know in which situation it actually converges

$\downarrow$ slutsky theorem

$$\left.\begin{aligned} X_n &\xrightarrow{d} X \\ Y_n &\xrightarrow{d} a \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} X_n Y_n &\xrightarrow{d} aX \\ X_n + Y_n &\xrightarrow{d} a + X \end{aligned}\right.$$

⑤ Bias of S^2

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

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right]$$

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

$$\mathbb{E}[S^2] = \frac{1}{n-1}\left[n\,\mathbb{E}[X^2] - n\,\mathbb{E}[\bar{X}^2] \right]$$

$$= \frac{1}{n-1}\left[n \operatorname{var}[X_i] - n \operatorname{var}[\bar{X}] \right]$$

$$= \frac{n}{n-1}\left(\sigma^2 - \operatorname{var}[\bar{X}]\right) \in \left[0, \frac{n}{n-1}\sigma^2\right]$$