# 5.1 I.i.d random samples

**Def** Suppose $X_1, \dots, X_n \overset{iid}{\sim} f$ (pmf or pdf)

then we say $(X_1, \dots, X_n)$ is a Random sample from the population (with pmf / pdf) $f(x)$.

**Remark:** other variations

- $n$-sample of $f$.
- $n$-iid sample of $f$ (vs. non-iid sample)
- a sample of size $n$ (vs. a sample of size 1)
- any of the above, without mentioning of $f$.

**Example**

$X_1, \dots, X_n \overset{iid}{\sim} \mathrm{Exp}(\beta)$, $\beta > 0$

Then we may compute, say

$P(X_1 > a_1, X_2 > a_2, \dots, X_n > a_n)$, and so on.

**Def.** 
- Let $(X_1, \dots, X_n)$ be a $n$-sample from a population
- Let $T(x_1, \dots, x_n)$ be a real-valued function

Then $Y = T(X_1, \dots, X_n)$ is called a **statistic**.

i.e. a function of a random sample is a statistic.

**Remarks**

- A statistic is a Random variable too
- A stat is telling us something about an (underlying) population with pdf / pmf $f$.
- it does so through <u>only</u> Random samples (data)
  (it is not a function of parameters, but it may inform us about the parameters of $f$)
- Modern viewpoint

$$X_1, \dots, X_n \longrightarrow \boxed{\quad T \quad} \longrightarrow Y = T(X_1, \dots, X_n).$$

an algorithm / an app.

# Examples of Statistics

① $\overline{X} := \frac{1}{n}(X_1 + \dots + X_n)$ sample mean

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

③ $X_{(1)} := \min \{X_1, \dots, X_n\}$

$X_{(2)} := \min \{X_1, \dots, X_n\} \setminus \{X_{(1)}\}$

...

$X_{(k)} := \min \{X_1, \dots, X_n\} \setminus \{X_{(1)}, \dots, X_{(k-1)}\}$

this is the k-th smallest member of the sample

$(X_{(1)}, \dots, X_{(n)})$ is called the order statistics of of the n-sample

**Theorem** Let $(X_1, \dots, X_n)$ be an $n$-iid sample from a population with $\begin{cases} \text{mean } \mu \\ \text{variance } \sigma^2 < \infty \end{cases}$

Then
$$\mathbb{E}\bar{X} = \mu \qquad (*)$$
$$\operatorname{var}\bar{X} = \sigma^2/n$$
$$\mathbb{E}S^2 = \sigma^2. \qquad (**)$$

Remark. ◦ This theorem provides a statistical justification for using $\bar{X}$ and $S^2$ as estimates of $\mu$ and $\sigma^2$ respectively.

$(*)$ and $(**)$ say these are unbiased estimators

**Proof.**

◦ $\mathbb{E}\bar{X} = \mathbb{E}\frac{1}{n}(X_1 + \dots + X_n)$

$$\overset{\text{L.O.E}}{=} \frac{1}{n}(\mathbb{E}X_1 + \dots + \mathbb{E}X_n)$$
$$= \frac{1}{n} n\mu = \mu$$

◦ $\operatorname{var}\bar{X} = \operatorname{var}\frac{1}{n}(X_1 + \dots + X_n)$

$$= \frac{1}{n^2} \operatorname{var}(X_1 + \dots + X_n)$$

$$= \frac{1}{n^2} \left( \operatorname{var} X_1 + \dots + \operatorname{var} X_n \right)$$ , since $X_i$ are indep.

$$= \frac{1}{n^2} n \cdot \operatorname{var} X_1$$

$$= \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n} .$$

(**) is left as an exercise. $\square$

$\boxed{\text{What is the Distribution of } \bar{X} \text{ ?}}$

Two main methods ① method of moments
② using change of variable formula (convolution formula)

## ① Method of moments

$$M_{\bar{X}}(t) = E\, e^{it\bar{X}}$$

$$= E\, e^{i\frac{t}{n}(X_1 + \dots + X_n)}$$

$$= \left(M_X(t/n)\right)^n$$

if $M_{\bar{X}}$ can be recognized as MGF of a known family then we can find the distribution of $\bar{X}$ easily

Example
if $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

then $\bar{X} \sim N(\mu, \sigma^2/n)$

## ② Convolution formula

**Theorem.** if $X \sim f_X$, $Y \sim f_Y$ both pdf

then $Z := X+Y$ has the pdf which is the convolution of $f_X$ and $f_Y$:

$$\boxed{\begin{aligned} f_Z(z) &= f_X * f_Y(z) \\ &:= \int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)\, dx . \end{aligned}}$$

**Proof.** Use the change-of-variable formula for the mapping $(X,Y) \mapsto (X+Y, X)$.

Remark.

- Useful for deriving pdf / pmf when the transformation $T(X_1, X_2, \dots, X_n)$ does not belong to a known / well-recognized family (such as exponential family, location-scale family…)