EXEMPLE INTRODUCTIF

Estimateur de la moyenne d'une population quantitative univariée (EAS)

Considérons la population des âges des enfants d'une famille

$U = \{2, 4, 12\}$

- $N = 3$
- $\mu = \frac{2+4+12}{3} = 6$
- $\sigma^2 = \frac{1}{3}(4+16+144) - 6^2 = \frac{56}{3}$

1) On prélève, dans la population, un échantillon EAS de taille $n = 2$

. on prélève un premier enfant (en attribuant à chacun la même probabilité d'être sélectionné, soit $p = \frac{1}{3}$)
L'âge de l'enfant n'est pas connu à priori : il constitue donc une variable aléatoire $\underline{X}_1$ dont la distr. de Pr. est :

$$\underline{X}_1 \sim \begin{pmatrix} 2 & 4 & 12 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$$

avec $\begin{cases} E(\underline{X}_1) = \mu = 6 \\ var(\underline{X}_1) = \sigma^2 = \frac{56}{3} \end{cases}$

. on prélève un second enfant ("tirages" indépendants)

$$\rightarrow \quad \underline{X}_2 \sim \begin{pmatrix} 2 & 4 & 12 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$$

avec $\begin{cases} E(\underline{X}_2) = \mu = 6 \\ var(\underline{X}_2) = \sigma^2 = \frac{56}{3} \end{cases}$

On a donc $(\underline{X}_1, \underline{X}_2)$, un ensemble de 2 v.a. (correspondant respectivement à la 1ère et à la 2e observation) INDÉPENDANTES (schéma de Bernoulli) et IDENTIQUEMENT DISTRIBUEES (elles admettent toutes deux une distribution uniforme sur l'ensemble des 3 valeurs 2, 4, 12) $\Rightarrow \begin{cases} E(\underline{X}_i) = \mu = 6 \\ Var(\underline{X}_i) = \sigma^2 = \frac{56}{3} \end{cases}$ $(i = 1, 2)$

Soit $\{2, 12\}$ l'échantillon prélevé : il s'agit de valeurs réalisées de 2 v.a. $(\underline{X}_1, \underline{X}_2)$ iid (*)

② Estimation de la moyenne de la population

Dans cet exemple, on sait que $\mu = 6$. Mais, de façon générale, lorsqu'on étudie une population $\mathcal{U}$, on ne dispose pas d'une telle information :

> LES PARAMETRES de $\mathcal{U}$ SONT INCONNUS ET ON DESIRE LES ESTIMER
>
> → PROBLEME DE L'ESTIMATION

Une façon d'atteindre cet objectif est d'utiliser l'information contenue dans l'échantillon prélevé dans $\mathcal{U}$.

Comme $\underline{X}_1$ et $\underline{X}_2$ sont des v.a., il en est de même de leur moyenne :

$$\underline{\overline{X}} = \frac{\underline{X}_1 + \underline{X}_2}{2} = \text{ESTIMATEUR DE } \mu = \underline{\hat{\mu}} = \text{variable aléatoire}$$

FONCTION DE $\underline{X}_1, \underline{X}_2$

La moyenne arithmétique d'un échantillon peut jouer le rôle d'UNE estimation de μ

$$\text{une estimation de } \mu = \hat{\mu} = \frac{x_1 + x_2}{2} = \text{une valeur réalisée de } \underline{\overline{X}}$$

$\hat{\mu}$ = un réel ; Fct de x_1, x_2 ; une valeur réalisée de $\underline{\overline{X}}$ = un $\overline{x}$

Avec l'échantillon (*), on a $\hat{\mu} = \frac{2+12}{2} = 7$

fct de 2 et 12

↳ une valeur réalisée de $\underline{\overline{X}}$

(NB) $\underline{\hat{\mu}} = f(\underline{X}_1, \underline{X}_2)$

↳ la Fct doit être judicieusement choisie pour que l'estimateur remplisse son rôle !!

③ Enumérons TOUS les échantillons possibles EAS de taille $n=2$ ainsi que leur moyenne, extraits de la population $\mathcal{U}$

Echantillons ω_i	Moyenne d'échantillon $\bar{x}$	Probabilité de tirer l'échantillon i
$\{2,2\}$	$\frac{2+2}{2} = 2$	$\frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$
$\{2,4\}$	3	1/9
$\{2,12\}$	7	1/9
$\{4,2\}$	$\frac{4+2}{2} = 3$	1/9
$\{4,4\}$	4	1/9
$\{4,12\}$	8	1/9
$\{12,2\}$	$\frac{12+2}{2} = 7$	1/9
$\{12,4\}$	8	1/9
$\{12,12\}$	12	1/9

Distribution de probabilité de la v.a. $\bar{X}$:

$$\bar{X} \sim \begin{pmatrix} 2 & 3 & 4 & 7 & 8 & 12 \\ \frac{1}{9} & \frac{2}{9} & \frac{1}{9} & \frac{2}{9} & \frac{2}{9} & \frac{1}{9} \end{pmatrix}$$

$$E(\bar{X}) = 2 \cdot \frac{1}{9} + 3 \cdot \frac{2}{9} + 4 \cdot \frac{1}{9} + 7 \cdot \frac{2}{9} + 8 \cdot \frac{2}{9} + 12 \cdot \frac{1}{9} = 6$$

Conclusion : dans notre exemple, la moyenne d'un échantillon de taille $n=2$ ne sera "jamais" égale à la moyenne de la population (cf *)

MAIS

$$\boxed{\begin{array}{l} E(\hat{\mu}) = E(\bar{X}) = \mu \\ \text{cad la v.a. } \bar{X} \text{ est en moyenne } = \mu \\ \text{cad la moyenne d'échantillonnage de la moyenne d'échantillon est égale à la moyenne de la population} \end{array}}$$

$$\mathrm{var}(\underline{\bar{X}}) = E\left[\left(\underline{\bar{X}} - \text{la moyenne de } \underline{\bar{X}}\right)^2\right]$$

$$= (2-6)^2 \cdot \frac{1}{9} + (3-6)^2 \cdot \frac{2}{9} + (4-6)^2 \cdot \frac{1}{9} + (7-6)^2 \cdot \frac{2}{9}$$

$$+ (8-6)^2 \cdot \frac{2}{9} + (12-6)^2 \cdot \frac{1}{9}$$

$$= \frac{84}{9}$$

$$= \frac{28}{3}$$

$$= \frac{56/3}{2}$$

$$= \frac{\sigma^2}{n(=2)}$$

Conclusion : la moyenne $\underline{\bar{X}}$ prend des valeurs qui se répartissent autour de μ avec une certaine dispersion mesurée par $\mathrm{var}(\underline{\bar{X}}) = \frac{28}{3}$

$$\boxed{\mathrm{var}(\underline{\hat{\mu}}) = \mathrm{var}(\underline{\bar{X}}) = \frac{\sigma^2}{n}}$$

cad la variance d'échantillonnage de la moyenne d'échantillon est égale à la variance de la population divisée par la taille de l'échantillon

REMARQUES : ECHANTILLON EASR

Si EASR, on a toujours $E(\underline{\hat{\mu}}) = \mu$

mais $\mathrm{var}(\underline{\hat{\mu}}) = \frac{\sigma^2}{n} \frac{N-n}{N-1}$

Si $N \gg n \Rightarrow \frac{N-n}{N-1} \to 1$ car n négligeable devant N

$$\Rightarrow \left[\mathrm{var}(\underline{\hat{\mu}})\right]_{EAS} \simeq \left[\mathrm{var}(\underline{\hat{\mu}})\right]_{EASR}$$

si $N \gg n$