Chapitre 4

Distribution des échantillons aléatoires

Université de Paris Ouest

2012–2013

Objectifs du chapitre

Rappel :

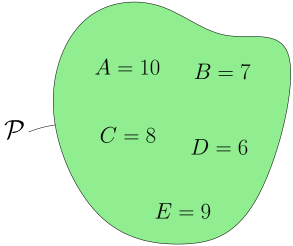
L'**inférence statistique** consiste à induire les caractéristiques inconnues d'une population à partir d'un échantillon.

- échantillon représentatif ?
- échantillon suffisamment grand ?

Sommaire

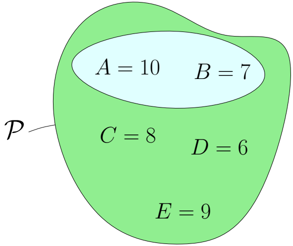
Un exemple avec une petite population

Le score d'*Agpar* mesure la santé d'un nouveau-né.
$\mathcal{P} = \{5 \text{ enfants}\}, \mu = 8$.



Un exemple avec une petite population

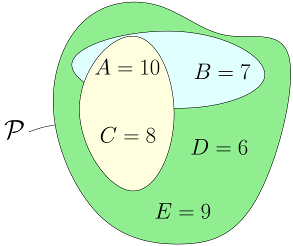
Le score d'*Agpar* mesure la santé d'un nouveau-né.
$\mathcal{P} = \{5 \text{ enfants}\}, \mu = 8.$
échantillon de taille 2



éch. AB			
$\bar{x} = 8,5$			

Un exemple avec une petite population

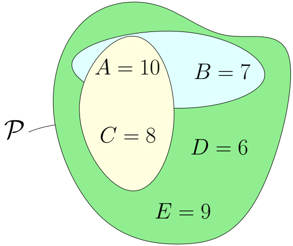
Le score d'*Agpar* mesure la santé d'un nouveau-né.
$\mathcal{P} = \{5 \text{ enfants}\}, \mu = 8$.
échantillon de taille 2



éch. AB	éch. AC		
$\bar{x} = 8,5$	$\bar{x} = 9$		

Un exemple avec une petite population

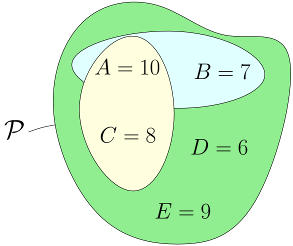
Le score d'*Agpar* mesure la santé d'un nouveau-né.
$\mathcal{P} = \{5 \text{ enfants}\}, \mu = 8.$
échantillon de taille 2



éch. *AB*	éch. *AC*	...	éch. *DE*
$\bar{x} = 8,5$	$\bar{x} = 9$	...	$\bar{x} = 7,5$

Un exemple avec une petite population

Le score d'*Agpar* mesure la santé d'un nouveau-né.
$\mathcal{P} = \{5 \text{ enfants}\}, \mu = 8.$
échantillon de taille 2

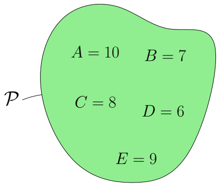


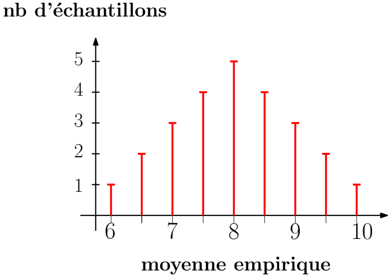
éch. AB	éch. AC	...	éch. DE
$\bar{x} = 8,5$	$\bar{x} = 9$	...	$\bar{x} = 7,5$

Si l'échantillon est **tiré au sort**, alors la moyenne observée $\bar{x}$ est **aléatoire** !

Suite de l'exemple : distribution des échantillons

Le score d'*Agpar* mesure la santé d'un nouveau-né.
$\mathcal{P} = \{5 \text{ enfants}\}, \mu = 8$.
échantillon de taille 2

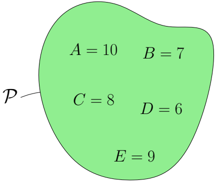


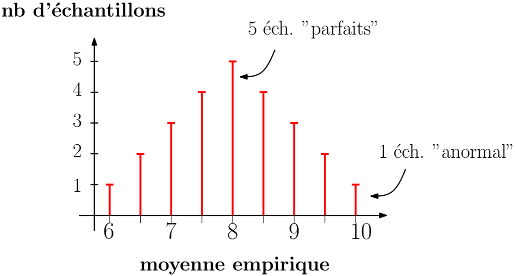


Si l'échantillon est **tiré au sort**, alors la moyenne empirique $\bar{x}$ est **aléatoire** !

Suite de l'exemple : distribution des échantillons

Le score d'*Agpar* mesure la santé d'un nouveau-né.
$\mathcal{P} = \{5 \text{ enfants}\}, \mu = 8$.
échantillon de taille 2





Si l'échantillon est **tiré au sort**, alors la moyenne empirique $\bar{x}$ est **aléatoire** !

Exemple : à retenir

Si l'échantillon est **tiré au sort**, alors la moyenne empirique $\bar{x}$ est **aléatoire**.

- On aimerait connaître la distribution de $\bar{x}$
- Si l'échantillon est grand, peu de chances d'avoir un échantillon "anormal"?

Notations : $\bar{x}$ vs $\overline{X}_n$

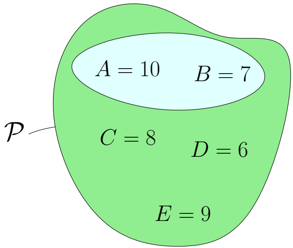
En fonction du contexte, deux notations

- $\bar{x}$: moyenne observée sur un **un échantillon fixé**
- $\overline{X}_n$: moyenne empirique d'un **échantillon aléatoire**

Notations : $\bar{x}$ vs $\overline{X}_n$

En fonction du contexte, deux notations

- $\bar{x}$: moyenne observée sur un **un échantillon fixé**
- $\overline{X}_n$: moyenne empirique d'un **échantillon aléatoire**



Pour l'échantillon AB, $\bar{x} = 8,5$ mais $\overline{X}_n$ peut valoir 8.5/9/7.5/ ...

Sommaire

Protocole : qu'est-ce qu'un échantillon aléatoire ?

On a besoin d'un modèle mathématique précis pour le tirage d'un échantillon de n personnes.

Hypothèses

- n tirages **uniformes** : chaque individu a la même chance d'être tiré au sort.
- Les tirages sont **indépendants**.
- Les tirages se font **avec remise**.

Échantillons aléatoires pour une variable quantitative

X variable continue, de moyenne μ et d'écart-type σ.
$\overline{X}_n$ désigne la moyenne empirique d'un échantillon **tiré au sort**.

Formule

Si $n \geq 30$, alors

$$\overline{X}_n \overset{\text{approx.}}{\sim} \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Échantillons aléatoires pour une variable quantitative

X variable continue, de moyenne μ et d'écart-type σ.
$\overline{X}_n$ désigne la moyenne empirique d'un échantillon **tiré au sort**.

Formule

Si $n \geq 30$, alors

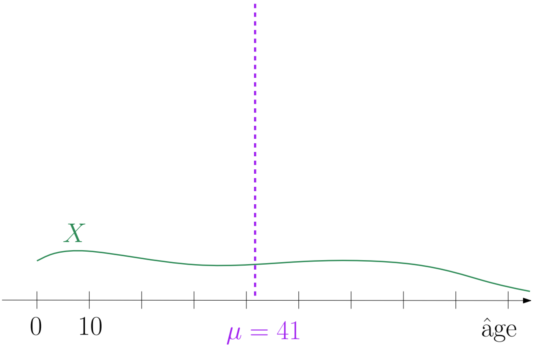
$$\overline{X}_n \overset{\text{approx.}}{\sim} \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

Exemple : $X=$ "Âge", $\mu = 41$, $\sigma = 23$.
Si $n = 30$, $\overline{X}_{30}$ est l'âge moyen de 30 Français tirés au hasard.

$$\overline{X}_{30} \overset{\text{approx.}}{\sim} \mathcal{N}\left(41, \frac{23}{\sqrt{30}}\right) = \mathcal{N}\left(41; 4,20\right).$$

Exemples de $\bar{X}_n$ quand n varie

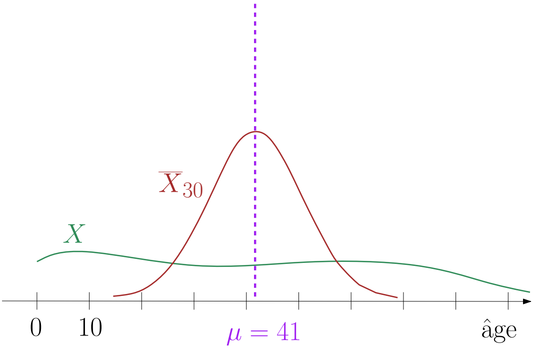
X= "Âge", $\mu = 41$, $\sigma = 23$.



Exemples de $\bar{X}_n$ quand n varie

X= "Âge", $\mu = 41$, $\sigma = 23$.
Si $n = 30$, $\overline{X}_{30}$ est l'âge moyen de 30 Français tirés au hasard.

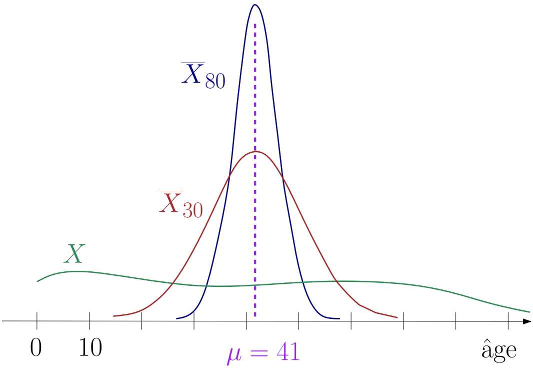
$$\overline{X}_{30} \sim \mathcal{N}(41; 4,20)$$



Exemples de $\bar{X}_n$ quand n varie

X= "Âge", $\mu = 41$, $\sigma = 23$.
Si $n = 30$, $\overline{X}_{30}$ est l'âge moyen de 30 Français tirés au hasard.
Si $n = 80$, $\overline{X}_{80}$ est l'âge moyen de 80 Français tirés au hasard.

$$\overline{X}_{30} \sim \mathcal{N}(41; 4,20) \qquad \overline{X}_{80} \sim \mathcal{N}(41; 2,57)$$



Échantillons aléatoires pour une variable qualitative

X variable qualitative, de proportion p.
On note F_n la moyenne empirique d'un échantillon **tiré au sort**.

Formule

Si $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$, alors

$$F_n \overset{\text{approx.}}{\sim} \mathcal{N}\left(p, \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right).$$

Échantillons aléatoires pour une variable qualitative

X variable qualitative, de proportion p.
On note F_n la moyenne empirique d'un échantillon **tiré au sort**.

Formule

Si $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$, alors

$$F_n \overset{\text{approx.}}{\sim} \mathcal{N}\left(p, \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right).$$

Exemple : $\mathcal{P} =$ "utilisateurs de Facebook en France",$X=$ "Sexe",
$p=$ "proportion de femmes" $= 0.54$.
Soit F_n la proportion de femmes parmi 30 utilisateurs tirés au hasard.

Échantillons aléatoires pour une variable qualitative

X variable qualitative, de proportion p.
On note F_n la moyenne empirique d'un échantillon **tiré au sort**.

Formule

Si $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$, alors

$$F_n \overset{\text{approx.}}{\sim} \mathcal{N}\left(p, \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right).$$

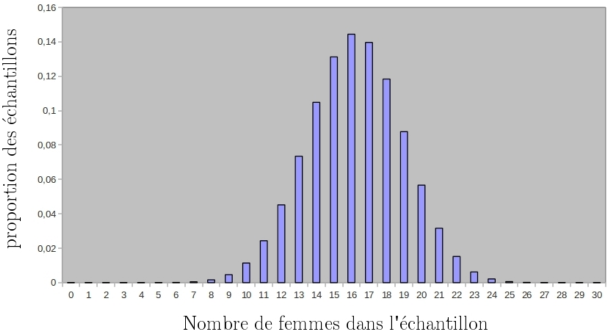
Exemple : $\mathcal{P}$ = "utilisateurs de Facebook en France",X= "Sexe",
p= "proportion de femmes" = 0.54.
Soit F_n la proportion de femmes parmi 30 utilisateurs tirés au hasard.
Comme $30 \geq 30$, $30 \times 0.54 \geq 5$, $30 \times 0.46 \geq 5$,

$$F_n \overset{\text{approx.}}{\sim} \mathcal{N}\left(0.54, \frac{\sqrt{0.54 \times 0.46}}{\sqrt{30}}\right) = \mathcal{N}\left(0.54, 0.09\right).$$

Échantillons aléatoires pour une variable qualitative

$\mathcal{P}$ = "utilisateurs de Facebook en France",X= "Sexe", $p = 0.54$
F_n = proportion de femmes parmi 30 utilisateurs tirés au hasard.

$$F_n \overset{\text{approx.}}{\sim} \mathcal{N}(0.54, 0.09) .$$



Sommaire

Un exemple de calcul

X= "Sexe", p= "proportion de femmes" $= 0.54$
F_n= proportion de femmes dans un échantillon de 30 individus.

Question :

On tire 30 utilisateurs au hasard, quelle est la probabilité d'avoir plus de 60% d'hommes ?

Un exemple de calcul

$X=$ "Sexe", $p=$ "proportion de femmes" $= 0.54$
$F_n=$ proportion de femmes dans un échantillon de 30 individus.

Question :

On tire 30 utilisateurs au hasard, quelle est la probabilité d'avoir plus de 60% d'hommes ?

Cela revient à calculer $P(F_n \leq 0.40)$.
Au transparent précédent, nous avons vu que $F_n \sim \mathcal{N}(0.54; 0.09)$.

Un exemple de calcul

$X=$ "Sexe", $p=$ "proportion de femmes" $= 0.54$
$F_n=$ proportion de femmes dans un échantillon de 30 individus.

Question :

On tire 30 utilisateurs au hasard, quelle est la probabilité d'avoir plus de 60% d'hommes ?

Cela revient à calculer $P(F_n \leq 0.40)$.
Au transparent précédent, nous avons vu que $F_n \sim \mathcal{N}(0.54; 0.09)$.
On **centre et on réduit** F_n :

$$P(F_n \leq 0.40) = P\left(\frac{F_n - 0.54}{0.09} \leq \frac{0.40 - 0.54}{0.09}\right)$$
$$= P\left(Z \leq -1.56\right) = F(-1.56) = 1 - F(1.56) = 0.0594.$$

Un exemple de calcul

X= "Sexe", p= "proportion de femmes" $= 0.54$
F_n= proportion de femmes dans un échantillon de 30 individus.

Question :

On tire 30 utilisateurs au hasard, quelle est la probabilité d'avoir plus de 60% d'hommes ?

Cela revient à calculer $P(F_n \leq 0.40)$.
Au transparent précédent, nous avons vu que $F_n \sim \mathcal{N}(0.54; 0.09)$.
On **centre et on réduit** F_n :

$$P(F_n \leq 0.40) = P\left(\frac{F_n - 0.54}{0.09} \leq \frac{0.40 - 0.54}{0.09}\right)$$
$$= P\left(Z \leq -1.56\right) = F(-1.56) = 1 - F(1.56) = 0.0594.$$

Réponse :

On a environ 6% de chance de tomber sur un échantillon comptant plus de 60% d'hommes.

Sommaire

Conclusion

- Échantillon **tiré au sort** $\Rightarrow$ moyenne empirique/fréquence empirique **aléatoire**.
- Si $n \geq 30$, on connaît la distribution de $\overline{X}_n, F_n$.

Conclusion

- Échantillon **tiré au sort** $\Rightarrow$ moyenne empirique/fréquence empirique **aléatoire**.
- Si $n \geq 30$, on connaît la distribution de $\overline{X}_n, F_n$.

Rappel

Les formules ne sont valables que si l'échantillon est **aléatoire** et **uniforme** : chaque individu de $\mathcal{P}$ a la même chance de faire partie de l'échantillon.