Chapitre 4

DISTRIBUTIONS D'ECHANTILLONNAGE et INTERVALLES DE VARIATION

Chapitre 4 *(suite 2)*

1. Introduction

2. Estimation d'une moyenne
   - Distribution d'échantillonnage
   - Intervalles de variation

3. Estimation d'une proportion
   - Distribution d'échantillonnage
   - Intervalles de variation

4. Estimation d'une variance
   - Distribution d'échantillonnage
   - Estimation sans biais

Chapitre 4 *(suite 2)*

4. Estimation d'une variance

4.1 Variable quantitative

4.2 Variance empirique : statistique et estimateur

4.3 Fluctuations d'échantillonnage de la variance empirique

exemple théorique

4.4 Propriétés de la variance empirique

4.5 Variance empirique sans biais ou corrigée

exemple théorique

4.6 Propriétés de la variance empirique sans biais

4. Estimation d'une variance

X variable quantitative de $\mathcal{P}$ dans E

$\Rightarrow$deux paramètres

moyenne de X = μ ***connue*** dans $\mathcal{P}$
variance de X = σ^2 ***connue*** dans $\mathcal{P}$
(écart-type de X = σ ***connu*** dans $\mathcal{P}$)

échantillon de X issu de $\mathcal{P}$ de taille ***n***

- étudier les propriétés de l'estimation ponctuelle du paramètre variance σ^2 par
  la variance observée s^2 ou
  la variance observée sans biais s^{2*}
  sur un échantillon de taille ***n***

- les variances observées s^2 et s^{2*} ***varient*** d'un échantillon à l'autre

- étudier les variations (fluctuations) des variances observées s^2 et s^{2*} sur <u>tous</u> les échantillons de <u>même</u> taille ***n***

4.1 Variable quantitative

- **Exemple : âge des français**

$\mathcal{P}$ = *{ français recensés en 1999 }*

N *= 58 520 688*

X *= "âge" (en années)*

variable quantitative continue

$$\begin{cases} \mu = \text{âge moyen} = \textbf{39 connu} \text{ dans } \mathcal{P} \\ \sigma = \text{écart-type de l'âge} = \textbf{23 connu} \text{ dans } \mathcal{P} \end{cases}$$

échantillon de ***X*** *issu de* $\mathcal{P}$ *de taille* ***n*** *= 50*

➢ *la variance de l'âge est estimée par*
- la variance observée de l'âge s^2
- la variance observée sans biais (corrigée) de l'âge s^{2*}

➔ *quelles valeurs des variances* s^2 *et* s^{2*} *s'attend-on à observer sur l'échantillon de taille* ***n*** *= 50 ?*

4.2 Variance empirique : statistique et estimateur

La variable qui représente toutes les valeurs observées $\boldsymbol{s}^2$ sur tous les échantillons possibles de taille $\boldsymbol{n}$ est appelée **variance empirique** et notée $S_n^{\,2}$

$S_n^{\,2}$ est une **statistique** : une variable calculée à partir des observations (x_1, x_2, …, x_n) qui permet de résumer numériquement ces observations

- les propriétés de la statistique dépendent de la taille de l'échantillon ***n***

$\boldsymbol{s}^2$ est la valeur calculée sur l'échantillon observé de taille $\boldsymbol{n}$: c'est une **estimation** du paramètre $\boldsymbol{\sigma^2}$

- une estimation est propre à l'échantillon observé

$S_n^{\,2}$ est l'ensemble de toutes les estimations possibles $\boldsymbol{s}^2$ sur tous les échantillons de taille $\boldsymbol{n}$: c'est un **estimateur** de $\boldsymbol{\sigma^2}$

4.3 Fluctuations d'échantillonnage de la variance empirique

population de tous les échantillons de taille ***n***

- un individu est un échantillon de taille ***n***

S_n^2 = ***variance empirique***

⇨ **variable quantitative** définie sur la population des échantillons de taille ***n***

⇨ deux paramètres $\begin{cases} \textbf{moyenne} \text{ de } S_n^2 \\ \textbf{écart-type} \text{ de } S_n^2 \end{cases}$

⇨ **forme** de la distribution (loi) de S_n^2

➢ **étudier les variations (fluctuations) de la variance empirique** S_n^2

➢ étudier la ***"loi des variances"***

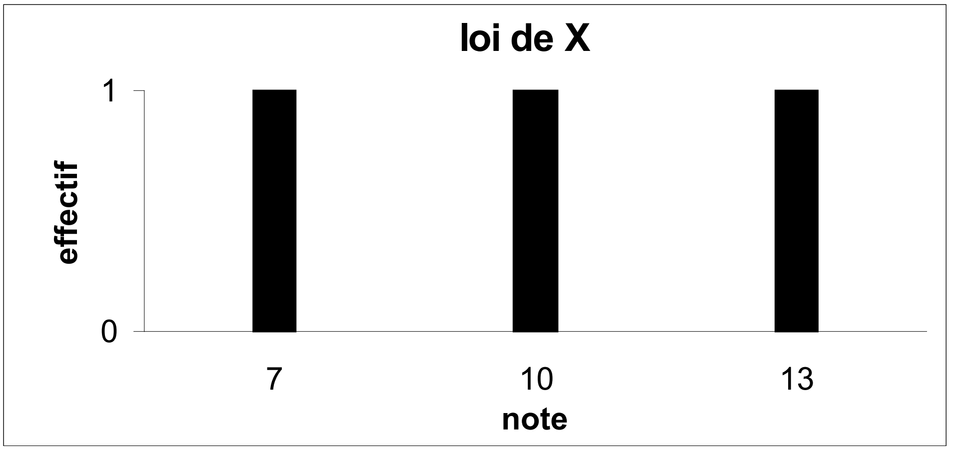
Exemple théorique (1)

- **Exemple : note à un examen**

  $\mathcal{P}$ = *{ étudiants }* ***N*** *= 3*

  X *= "note à un examen" (en points)*
  variable quantitative discrète sur
  E = { 0, 1, … 20 }

variable X	n_i	$n_i\,x_i$	$n_i\,x_i^2$
7	1	7	49
10	1	10	100
13	1	13	169
total	3	30	318



$\Sigma\, n_i\, x_i = 30 \quad \Sigma\, n_i\, x_i^2 = 318$

➔ *deux paramètres* **connus** *dans* $\mathcal{P}$

$\begin{cases} \mu \ = \textit{note moyenne} = \mathbf{10} \\ \sigma^2 = \textit{variance de la note} = \mathbf{6} \end{cases}$

Exemple théorique (2)

- **Exemple : note à un examen**

  $\mathcal{P}$ = { étudiants } ***N*** *= 3*

  X *= "note à un examen" (en points)*
  variable quantitative discrète sur
  E = { 0, 1, … 20 }

  ➔ *deux paramètres* **connus** *dans $\mathcal{P}$*

  μ *= note moyenne =* ***10***
  σ^2 *= variance de la note =* ***6***

  échantillons de ***X*** *issu de $\mathcal{P}$ de taille*
  n *= 2 : il y en a* **9**

  pour chacun des 9 échantillons la variance observée s^2 ***varie***

échantillon	(X_1, X_2)	$\overline{X}_2$	S_2	S_2^2
1	(7, 7)	7,0	0,0	0,00
2	(7, 10)	8,5	1,5	2,25
3	(7, 13)	10,0	3,0	9,00
4	(10, 7)	8,5	1,5	2,25
5	(10, 10)	10,0	0,0	0,00
6	(10, 13)	11,5	1,5	2,25
7	(13, 7)	10,0	3,0	9,00
8	(13, 10)	11,5	1,5	2,25
9	(13, 13)	13,0	0,0	0,00

s^2

<u>Exemple théorique</u> (3)

- **Exemple : note à un examen**

  population des échantillons de taille ***n*** *= 2* ***de taille 9***

  un individu est un échantillon de taille ***n*** *= 2*

  S_2^2 = *"variance de la note à l'examen"*

  variable quantitative

  distribution *(loi) de* S_2^2 :

  "loi des variances"

s^2

variable S_2^2	n_i	$n_i\ s_i^2$
0,00	3	0
2,25	4	9
9,00	2	18
total	9	27

$$\sum n_i s_i^2 = 27$$

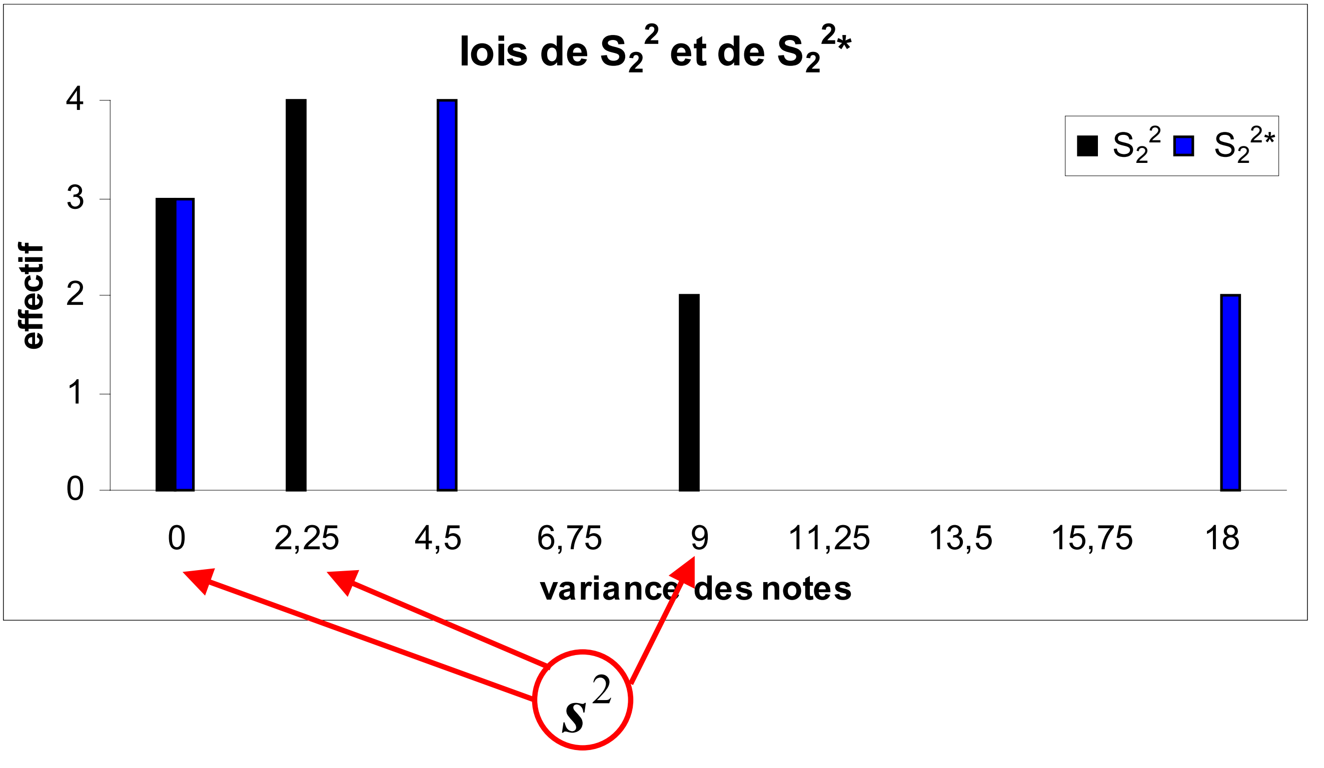
<u>Exemple théorique</u> (4)

- **Exemple : note à un examen**

population des échantillons de taille ***n*** *= 2* ***de taille 9***

S_2^2 = *"variance de la note à l'examen" variable quantitative*

➔ forme *de la loi de* S_2^2



➔ *deux* paramètres *pour la loi de* S_2^2

moyenne de S_2^2 = 27 / 9 = 3 = σ^2/ **2**

variance de S_2^2

4.4 Propriétés de la variance empirique (1)

- La moyenne de la variance empirique S_n^2 est égale à $\dfrac{n-1}{n}\sigma^2$

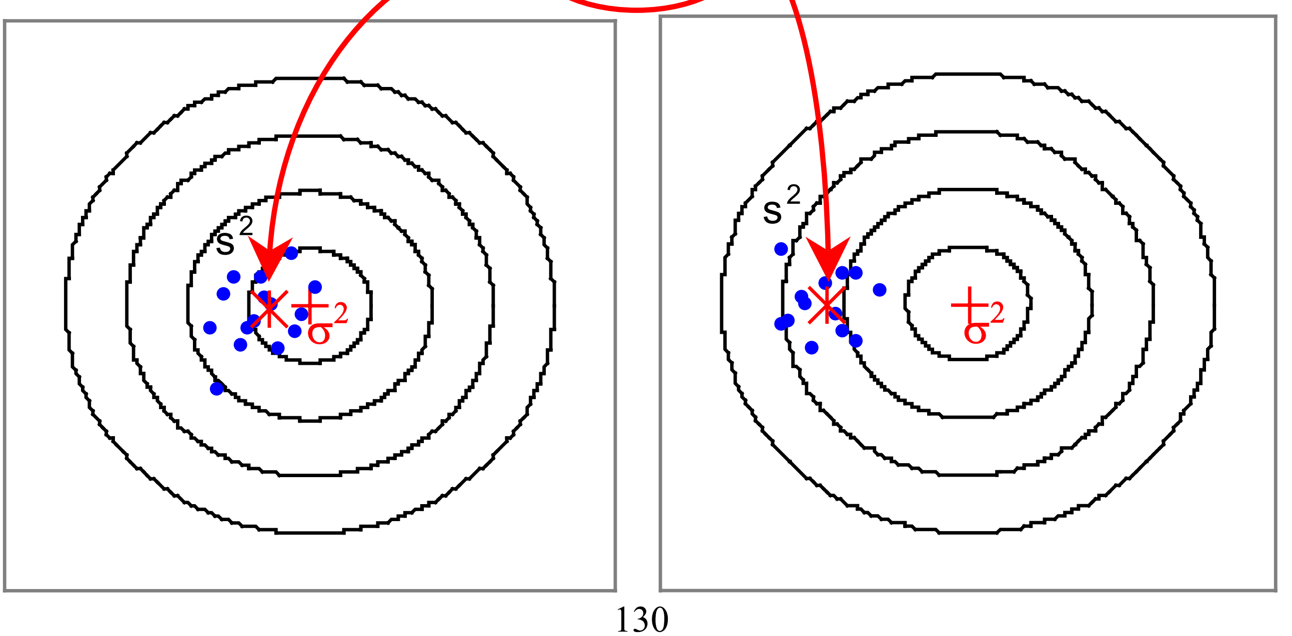
$\Rightarrow$ la variance empirique S_n^2 est un estimateur biaisé de $\boldsymbol{\sigma^2}$
elle *sous-estime* systématiquement $\boldsymbol{\sigma^2}$

$\Rightarrow$ la variance empirique S_n^2 *n*'est *pas* un *bon* estimateur de $\boldsymbol{\sigma^2}$

- ***n*** "grand"
biais faible

- ***n*** "petit"
biais élevé

$\dfrac{n-1}{n}\boldsymbol{\sigma}^2$



4.4 Propriétés de la variance empirique (2)

- **Exemple : âge des français**

$\mathcal{P}$ = *{ français recensés en 1999 }*

N *= 58 520 688*

X *= "âge" (en années)*

variable quantitative continue

deux paramètres ***connus*** *dans* $\mathcal{P}$

$$\begin{cases} \mu = \text{âge moyen} = \mathbf{39} \text{ ans} \\ \sigma^2 = \text{variance de l'âge} = \mathbf{23}^2 = \mathbf{529} \end{cases}$$

échantillon de ***X*** *issu de* $\mathcal{P}$ *de taille*

n *= 50*

➢ *la moyenne de* $S_n^2 = (49/50)\,\sigma^2$

$= 0{,}98 \times 529 = \mathbf{518{,}42}$

① *on s'attend <u>en moyenne</u> à observer une variance de l'âge égale à 518,42 sur l'échantillon de taille* ***n*** *= 50 ; cette quantité change avec la taille de l'échantillon*

échantillon de ***X*** *issu de* $\mathcal{P}$ *de taille*

n *= 100*

➢ *la moyenne de* $S_n^2 = (99/100)\,\sigma^2$

$= 0{,}99 \times 529 = \mathbf{523{,}71}$

① *on s'attend <u>en moyenne</u> à observer une variance de l'âge égale à 523,71 sur l'échantillon de taille* ***n*** *= 100*

4.4 Propriétés de la variance empirique (3)

- Le biais de la variance empirique S_n^2 dans l'estimation de σ^2 est égal à $-\dfrac{\sigma^2}{n}$

$\Rightarrow$si la taille de l'échantillon n augmente le biais diminue en se rapprochant de 0

taille de l'échantillon n	$\dfrac{n-1}{n}$	biais $-\dfrac{1}{n}$	$\dfrac{n}{n-1}$
2	0,5	– 0,5	2
10	0,9	– 0,1	1,1111
20	0,95	– 0,05	1,0526
30	0,9667	– 0,0333	1,0345
50	0,98	– 0,02	1,0204
100	0,99	– 0,01	1,0101
1000	0,999	– 0,001	1,0010

4.5 Variance empirique sans biais ou corrigée (1)

➢ **X variable quantitative définie de $\mathcal{P}$ dans E**
$$\begin{cases} \text{moyenne de X} = \mu \\ \text{écart-type de X} = \sigma \end{cases}$$
échantillon de X issu de $\mathcal{P}$ de taille *n*

➢ **S_n^2 variance empirique**

moyenne de $S_n^2 = \sigma^2(n-1)/n$

➢ **on corrige l'estimateur S_n^2 en définissant**

S_n^{2*} variance empirique sans biais

variance empirique corrigée

variable quantitative définie sur E

$$S_n^2 * = \frac{n}{n-1} S_n^2$$

⇒ **moyenne de $S_n^{2*} = \sigma^2$**

Exemple théorique (1)

- **Exemple : note à un examen**

$\mathcal{P}$ = { étudiants } ***N*** *= 3*

X *= "note à un examen" (en points)*
variable quantitative discrète sur
E = { 0, 1, … 20 }

➔ *deux paramètres* **connus** *dans $\mathcal{P}$*

$\begin{cases} \mu = \text{note moyenne} = \mathbf{10} \\ \sigma^2 = \text{variance de la note} = \mathbf{6} \end{cases}$

échantillons de ***X*** *issu de $\mathcal{P}$ de taille* ***n*** *= 2 : il y en a* **9**

pour chacun des 9 échantillons la variance observée sans biais** s^{2*}* ***varie

échantillon	(X_1, X_2)	$\overline{X}_2$	$S_2^{\,2}$	$S_2^{\,2*}$
1	(7, 7)	7,0	0,00	0,0
2	(7, 10)	8,5	2,25	4,5
3	(7, 13)	10,0	9,00	18,0
4	(10, 7)	8,5	2,25	4,5
5	(10, 10)	10,0	0,00	0,0
6	(10, 13)	11,5	2,25	4,5
7	(13, 7)	10,0	9,00	18,0
8	(13, 10)	11,5	2,25	4,5
9	(13, 13)	13,0	0,00	0,0

s^{2*}

<u>Exemple théorique</u> (2)

- **Exemple : note à un examen**

  $\mathcal{P}$ = { étudiants } ***N*** *= 3*

  X *= "note à un examen" (en points)*
  variable quantitative discrète sur E = { 0, 1, … 20 }

  ➔ *deux paramètres* **connus** *dans $\mathcal{P}$*

  $$\begin{cases} \mu = \text{note moyenne} = \mathbf{10} \\ \sigma^2 = \text{variance de la note} = \mathbf{6} \end{cases}$$

  population des échantillons de taille ***n*** *= 2* ***de taille 9***
  un individu est un échantillon de taille ***n*** *= 2*

  S_2^{2*} = *"variance de la note à l'examen"*
  variable quantitative
  distribution *(loi) de* S_2^{2*} :
  "loi des variances corrigées"

s^{2*}

variable S_2^{2*}	n_i	$n_i\, s_i^{2*}$
0,0	3	0
4,5	4	18
18,0	2	36
total	9	54

$$\sum n_i s_i^{2*} = 54$$

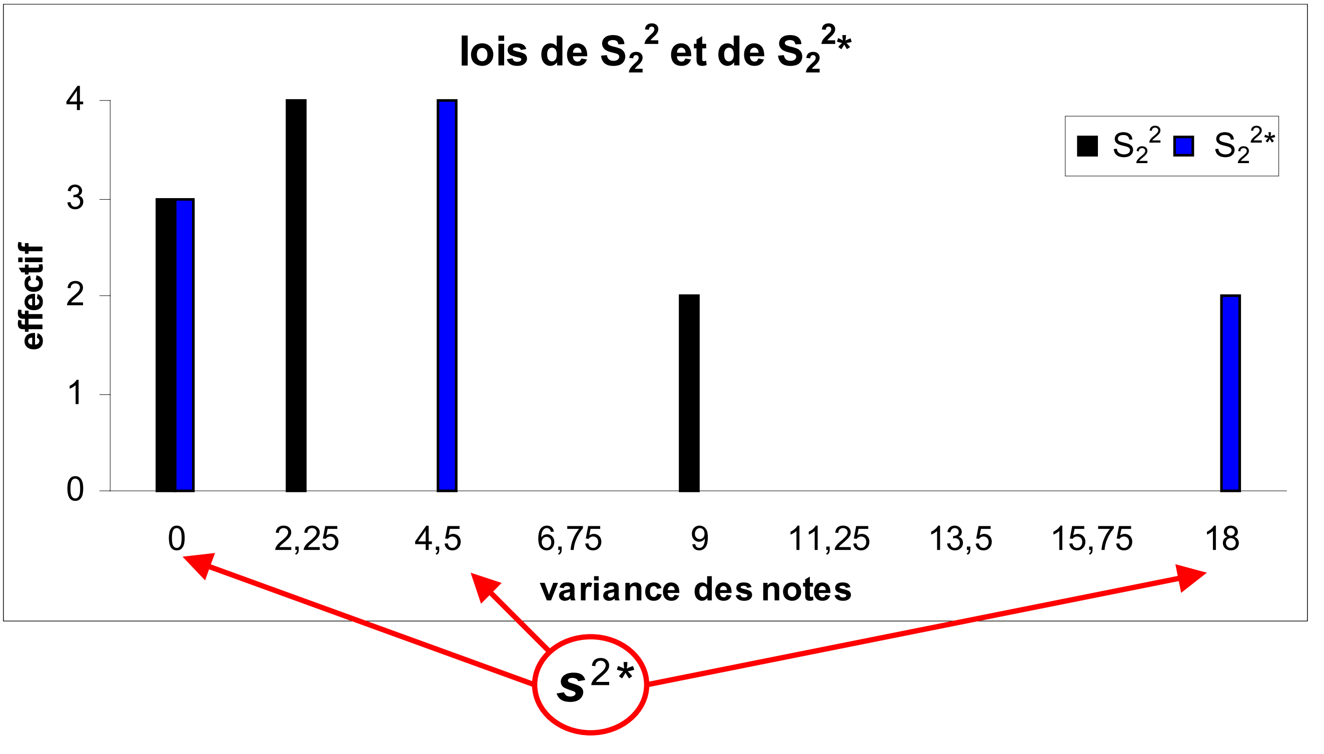
Exemple théorique (3)

- **Exemple : note à un examen**

population des échantillons de taille ***n*** = *2* ***de taille 9***

S_2^{2*} = *"variance de la note à l'examen" variable quantitative*

➔ forme *de la loi de* S_2^{2*}



➔ *deux* paramètres *pour la loi de* S_2^{2*}

moyenne de $S_2^{2*} = 54 / 9 = 6 = \sigma^2$

variance de S_2^{2*}

4.6 Propriétés de la variance empirique sans biais (1)

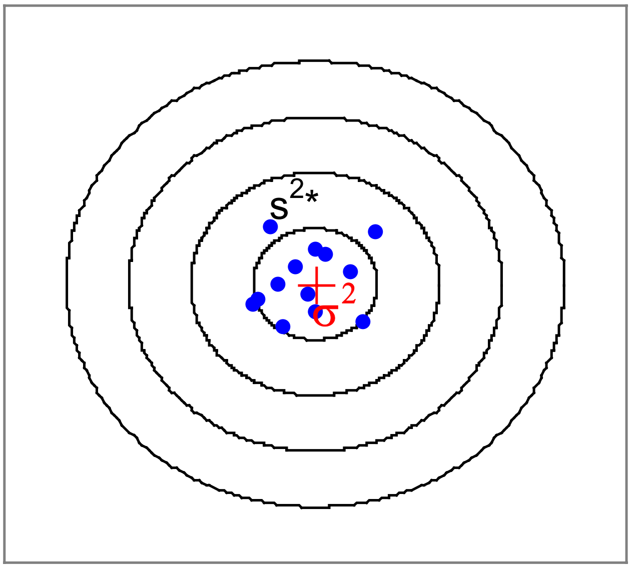
➢ La moyenne de la variance empirique sans biais (corrigée) S_n^{2*} est égale à σ^2

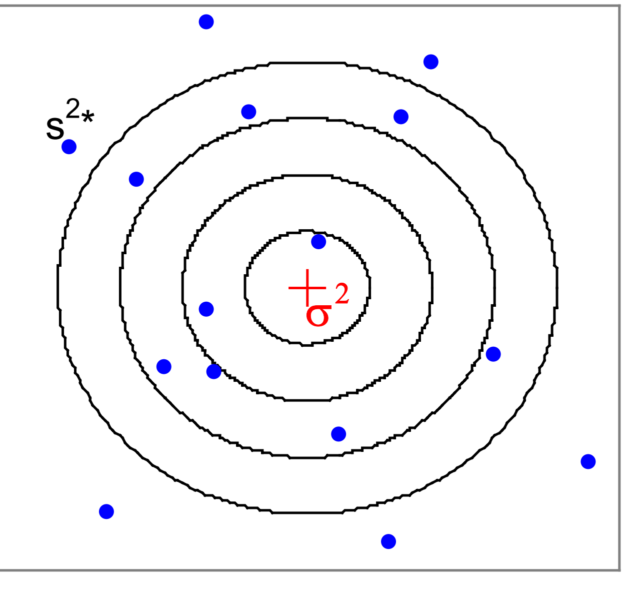
$\Rightarrow$la variance empirique S_n^{2*} est un estimateur sans biais de σ^2

$\Rightarrow$la variance empirique **sans biais** $\mathbf{S_n^{2*}}$ est un ***bon*** estimateur de la variance σ^2 de X dans $\mathcal{P}$: il est sans biais et convergent

- ***n*** "grand"
précision élevée

- ***n*** "petit"
précision faible





4.6 Propriétés de la variance empirique sans biais (2)

- **X variable quantitative définie de $\mathcal{P}$ dans E** $\begin{cases} \textbf{moyenne de X} = \mu \\ \textbf{écart-type de X} = \sigma \end{cases}$

  échantillon de X issu de $\mathcal{P}$ de taille *n*

- en pratique, on utilisera :

– la ***variance observée sans biais*** s^2* sur l'échantillon comme ***estimation ponctuelle*** de σ^2

$$s^2* = \frac{n}{n-1} s^2$$

– l'***écart-type observé sans biais*** $s*$ sur l'échantillon comme ***estimation ponctuelle*** de σ

4.6 Propriétés de la variance empirique sans biais (3)

- **Exemple : âge des français**

  $\mathcal{P}$ = *{ français recensés en 1999 }*

  N = *58 520 688*

  X = *"âge" (en années)*

  variable quantitative continue

  deux paramètres ***connus*** *dans* $\mathcal{P}$

  $$\begin{cases} \mu = \text{âge moyen} = \mathbf{39} \text{ ans} \\ \sigma^2 = \text{variance de l'âge} = \mathbf{23}^2 = \mathbf{529} \end{cases}$$

  échantillon de ***X*** *issu de* $\mathcal{P}$ *de taille*

  n *= 50*

  ➢ *la moyenne de* $S_n^2 = 518{,}42$

  ➢ *la moyenne de* $S_n^{2*} = \sigma^2 = \mathbf{529}$

  ① *on s'attend <u>en moyenne</u> à observer une variance sans biais (corrigée) de l'âge égale à* ***529*** *sur l'échantillon de taille* ***n*** *= 50, comme sur n'importe quel autre échantillon de taille* ***n*** *quelconque*

  échantillon de ***X*** *issu de* $\mathcal{P}$ *de taille*

  n *= 100*

  ➢ *la moyenne de* $S_n^2 = 523{,}71$

  ➢ *la moyenne de* $S_n^{2*} = \sigma^2 = \mathbf{529}$