Haute école de gestion de Genève
Geneva School of Business Administration

Estimations ponctuelles
Distribution de la moyenne

Dr. Sacha Varone

Objectif

- Comprendre et maîtriser l'estimation ponctuelle
- Connaître la distribution d'une moyenne

Rappels

Rappel

La distribution de Student $\mathcal{T}_n$

- Ligne
  nombre de degrés de liberté n
- Colonne
  une erreur de première espèce α.
- Intersection ligne/colonne $t_{\alpha,n}$

$$P(\mathcal{T}_n \geq t_{\alpha,n}) = \alpha$$

La relation entre p et α est $p = 1 - \alpha$.

Rappel

Soit un échantillon aléatoire de taille n, de moyenne $\bar{x}$ et de variance s^2, issu d'une loi normale $\mathcal{N}(\mu, \sigma^2)$. Alors

$$\frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \sim \mathcal{T}_{n-1}$$

Utilité : inférence sur la moyenne d'une population suivant une loi normale de moyenne μ et de variance σ^2 inconnue.

Rappel

Loi du χ^2

Soit $Q_n \sim \chi^2_n$

$p = P(Q_n \leq q_{\alpha,n})$

$\alpha = P(Q_n > q_{\alpha,n})$ avec $p = 1 - \alpha$

Rappel

La statistique χ^2 à $n-1$ degrés de liberté vaut

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

où

$$\begin{aligned} \chi^2 &= \text{variable chi-2 standard} \\ s^2 &= \text{variance de l'échantillon} \\ \sigma^2 &= \text{variance de la population} \\ n &= \text{taille de l'échantillon} \end{aligned}$$

Utilité : inférence sur la variance d'une population

Estimation ponctuelle

Définition

Une *estimation ponctuelle*, ou point d'estimation, est une valeur calculée à partir d'un échantillon pour estimer un paramètre d'une population.

Objectif : attribuer une valeur à un paramètre
Comment : échantillon

Exemple

X ou Y = Taille des ménages (nombre de personnes composant un ménage)

échantillon X	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
	1	3	2	4	4	1	2	6

échantillon Y	y_1	y_2	y_3	y_4	y_5
	2	2	1	1	3

Échantillon X : $\bar{x} = \frac{23}{8}$
Échantillon Y : $\bar{y} = \frac{9}{5}$

$\bar{x} \approx \mu$ et $\bar{y} \approx \mu$ μ = paramètre

Estimateur

Une *distribution d'échantillonnage* d'une estimateur $\hat{\Theta}$ est la distribution des valeurs possible d'une statistique pour un échantillon de taille fixée, sélectionné à partir d'une population.

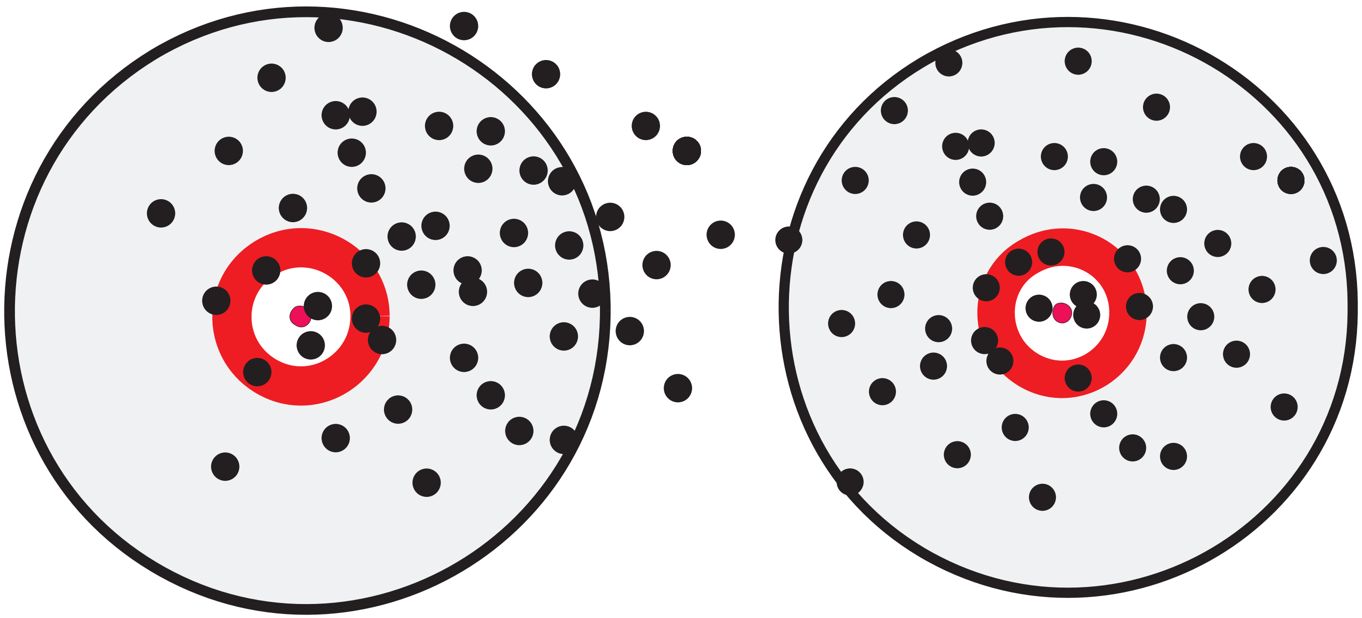
- comportement moyen :

$$\mathrm{E}(\hat{\Theta})$$

- dispersion :

$$\mathrm{Var}(\hat{\Theta})$$

Propriétés



Propriétés

Le *biais* : $\mathrm{E}(\hat{\Theta} - \theta) = \mathrm{E}(\hat{\Theta}) - \theta$

Estimateur *non-biaisé* lorsque son espérance est égale à la vraie valeur du paramètre estimé, i.e. $\mathrm{E}(\hat{\Theta}) = \theta$

Estimateur *convergent* si, lorsque la taille n de l'échantillon devient grande

1. le biais disparaît : $\lim_{n\to\infty} \mathrm{Biais}(\hat{\Theta}) = 0$
2. la variance devient nulle : $\lim_{n\to\infty} \mathsf{Var}(\hat{\Theta}) = 0$

Un estimateur sans biais et convergent est dit *absolument correct*.

Estimateurs

Estimation de la moyenne

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

est un estimateur absolument correct de la moyenne μ

Remarque :
La moyenne tronquée, le mode et la médiane calculés à partir d'un échantillon sont aussi des estimateurs absolument corrects pour respectivement la moyenne tronquée, le mode et la médiane de la population.

Estimation de la variance

Première idée :

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

S^2 estimateur biaisé, mais convergent.

Meilleur estimateur

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

s^2 estimateur absolument correct.

Exemple

Population $1, 2, 6$

- Moyenne et variance ?

Exemple

Population $1, 2, 6$

- Moyenne et variance ?
  $\mu = \frac{1+2+6}{3} = 3$ et $\sigma^2 = \frac{(1-3)^2+(2-3)^2+(6-3)^2}{3} = \frac{14}{3}$
- Échantillons possibles de taille 2, avec remise ?

Exemple

Population $1, 2, 6$

- Moyenne et variance ?
  $\mu = \frac{1+2+6}{3} = 3$ et $\sigma^2 = \frac{(1-3)^2+(2-3)^2+(6-3)^2}{3} = \frac{14}{3}$
- Échantillons possibles de taille 2, avec remise ?
  (1,1) (1,2) (1,6) (2,1) (2,2) (2,6) (6,1) (6,2) (6,6)
- Moyenne et variance de ces couples (en tant que populations) ?

Exemple

Population $1, 2, 6$

- Moyenne et variance ?
  $\mu = \frac{1+2+6}{3} = 3$ et $\sigma^2 = \frac{(1-3)^2+(2-3)^2+(6-3)^2}{3} = \frac{14}{3}$
- Échantillons possibles de taille 2, avec remise ?
  (1,1) (1,2) (1,6) (2,1) (2,2) (2,6) (6,1) (6,2) (6,6)
- Moyenne et variance de ces couples (en tant que populations) ?
  moyennes : $1, \frac{3}{2}, \frac{7}{2}, \frac{3}{2}, 2, 4, \frac{7}{2}, 4, 6$
  variances : $0, \frac{1}{4}, \frac{25}{4}, \frac{1}{4}, 0, 4, \frac{25}{4}, 4, 0$
- Moyenne de ces variances ?

Exemple

Population $1, 2, 6$

- Moyenne et variance ?
  $\mu = \frac{1+2+6}{3} = 3$ et $\sigma^2 = \frac{(1-3)^2+(2-3)^2+(6-3)^2}{3} = \frac{14}{3}$
- Échantillons possibles de taille 2, avec remise ?
  (1,1) (1,2) (1,6) (2,1) (2,2) (2,6) (6,1) (6,2) (6,6)
- Moyenne et variance de ces couples (en tant que populations) ?
  moyennes : $1, \frac{3}{2}, \frac{7}{2}, \frac{3}{2}, 2, 4, \frac{7}{2}, 4, 6$
  variances : $0, \frac{1}{4}, \frac{25}{4}, \frac{1}{4}, 0, 4, \frac{25}{4}, 4, 0$
- Moyenne de ces variances ? $\frac{7}{3} \neq \frac{14}{3}$

mais

$$\frac{n}{n-1} \cdot \frac{7}{3} = \frac{2}{2-1} \cdot \frac{7}{3} = \frac{14}{3}$$

Remarque

Calculatrices : variance par défaut s^2.

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

Distribution $\bar{x}$

Moyenne dans une population normale

Si une population est normalement distribuée, de moyenne μ et d'écart type σ, alors la distribution d'échantillonnage de la moyenne $\bar{x}$ est

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

Remarque
La distribution d'échantillonnage de la moyenne est composée de toutes les moyennes possibles sur tous les échantillons de même taille.

Erreur standard de la moyenne

L'écart type de la distribution d'échantillonnage de la moyenne, aussi appelée *erreur standard de la moyenne*, est le terme

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Remarque :
$\sigma_{\bar{x}} \leq \sigma$

Variable centrée réduite

La variable centrée réduite associée à la moyenne d'échantillonnage est la variable

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$\begin{array}{lcl} \bar{x} & = & \text{moyenne de l'échantillon} \\ \mu & = & \text{moyenne de la population} \\ \sigma & = & \text{écart type de la population} \\ n & = & \text{taille de l'échantillon} \end{array}$$

Exemple

Étiquette : poids en poisson d'un plat surgelé de 1 kg : $\mathcal{N}(500, 100^2)$ en grammes.
Échantillon : 25 plats achetés pour vérifier la quantité de poisson.
Résultats : moyenne de 490 g par plat.
Question : Y a-t-il tromperie du consommateur ?

1. La moyenne pour cet échantillon :

Exemple

Étiquette : poids en poisson d'un plat surgelé de 1 kg : $\mathcal{N}(500, 100^2)$ en grammes.
Échantillon : 25 plats achetés pour vérifier la quantité de poisson.
Résultats : moyenne de 490 g par plat.
Question : Y a-t-il tromperie du consommateur ?

1. La moyenne pour cet échantillon : $\bar{x} = 490$
2. La distribution d'échantillonnage de la moyenne :

Exemple

Étiquette : poids en poisson d'un plat surgelé de 1 kg : $\mathcal{N}(500, 100^2)$ en grammes.
Échantillon : 25 plats achetés pour vérifier la quantité de poisson.
Résultats : moyenne de 490 g par plat.
Question : Y a-t-il tromperie du consommateur ?

1. La moyenne pour cet échantillon : $\bar{x} = 490$
2. La distribution d'échantillonnage de la moyenne : $\mathcal{N}(500, \frac{100^2}{25} = 400)$
3. L'événement d'intérêt :

Exemple

Étiquette : poids en poisson d'un plat surgelé de 1 kg : $\mathcal{N}(500, 100^2)$ en grammes.
Échantillon : 25 plats achetés pour vérifier la quantité de poisson.
Résultats : moyenne de 490 g par plat.
Question : Y a-t-il tromperie du consommateur ?

1. La moyenne pour cet échantillon : $\bar{x} = 490$
2. La distribution d'échantillonnage de la moyenne : $\mathcal{N}(500, \frac{100^2}{25} = 400)$
3. L'événement d'intérêt : $P(\bar{x} \leq 490) =?$
4. Conversion en une valeur centrée réduite z

Exemple

Étiquette : poids en poisson d'un plat surgelé de 1 kg : $\mathcal{N}(500, 100^2)$ en grammes.
Échantillon : 25 plats achetés pour vérifier la quantité de poisson.
Résultats : moyenne de 490 g par plat.
Question : Y a-t-il tromperie du consommateur ?

1. La moyenne pour cet échantillon : $\bar{x} = 490$
2. La distribution d'échantillonnage de la moyenne : $\mathcal{N}(500, \frac{100^2}{25} = 400)$
3. L'événement d'intérêt : $P(\bar{x} \leq 490) =?$
4. Conversion en une valeur centrée réduite z
   $z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{490-500}{\frac{100}{\sqrt{25}}} = -0.5$
5. Probabilité désirée

Exemple

Étiquette : poids en poisson d'un plat surgelé de 1 kg : $\mathcal{N}(500, 100^2)$ en grammes.
Échantillon : 25 plats achetés pour vérifier la quantité de poisson.
Résultats : moyenne de 490 g par plat.
Question : Y a-t-il tromperie du consommateur ?

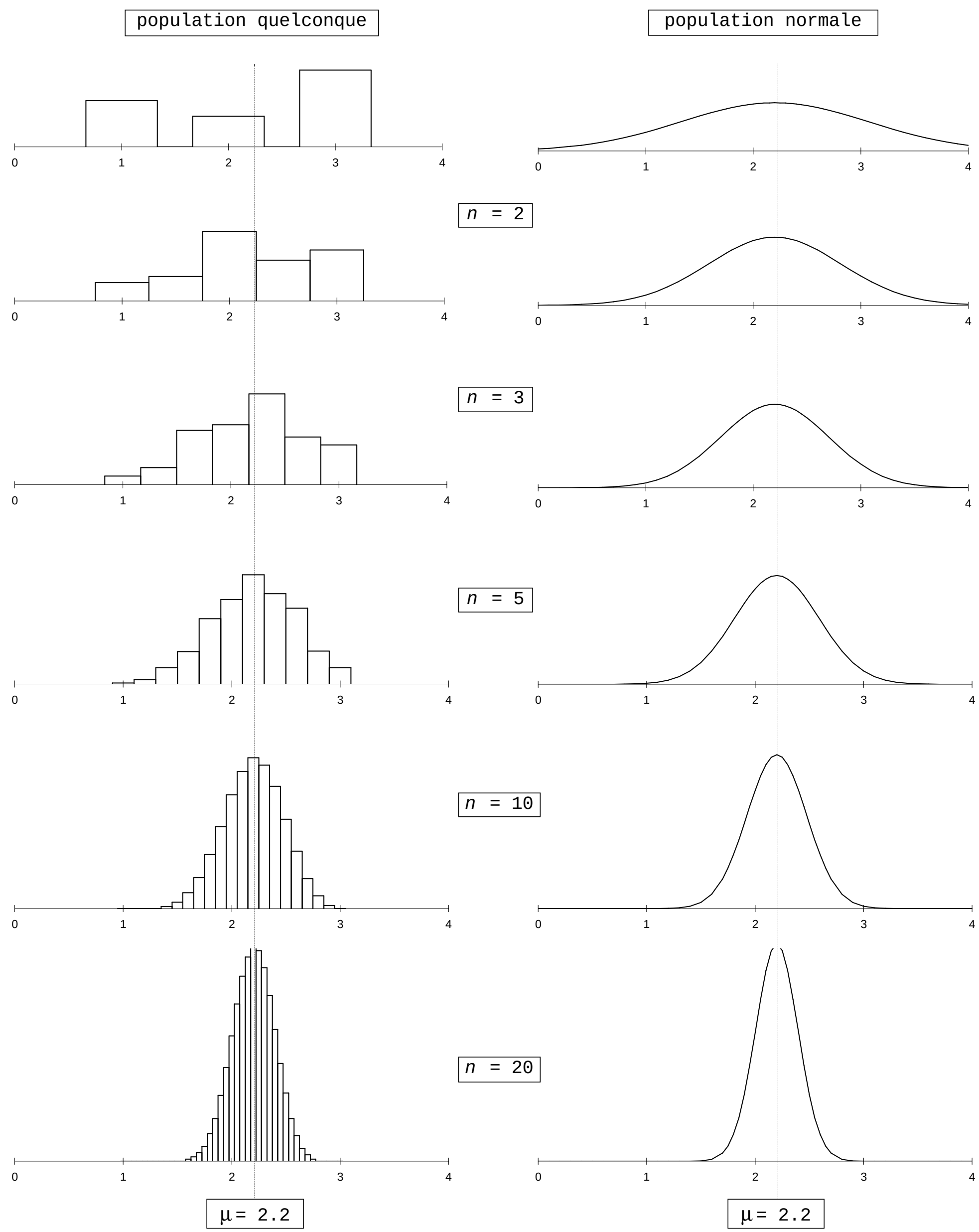
1. La moyenne pour cet échantillon : $\bar{x} = 490$
2. La distribution d'échantillonnage de la moyenne :
   $\mathcal{N}(500, \frac{100^2}{25} = 400)$
3. L'événement d'intérêt :
   $P(\bar{x} \leq 490) =?$
4. Conversion en une valeur centrée réduite z
   $z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{490-500}{\frac{100}{\sqrt{25}}} = -0.5$
5. Probabilité désirée
   $P(z \leq -0.5) = 0.3085$

Facteur de correction

Si la taille de l'échantillon est *plus du 5%* de la taille de la population, et que l'échantillon tiré est fait *sans remise*
$\Rightarrow$ facteur de correction sur l'écart type

$$\sqrt{\frac{N-n}{N-1}}$$

où
N = taille de la population
n = taille de l'échantillon


population quelconque
population normale
n = 2
n = 3
n = 5
n = 10
n = 20
μ = 2.2
μ = 2.2
0
1
2
3
4

Théorème central limite

Soit une suite $(X_1, X_2, \ldots, X_n)$ de n variables aléatoires identiquement et indépendamment distribuées (μ, σ^2). Lorsque $n \to \infty$, la distribution de

$$\overline{X} = \frac{1}{n} \sum_i X_i$$

tend vers la loi $N\big(\mu, \frac{\sigma^2}{n}\big)$

Remarque
Plus la taille de l'échantillon augmente, meilleure est l'approximation par la loi normale.

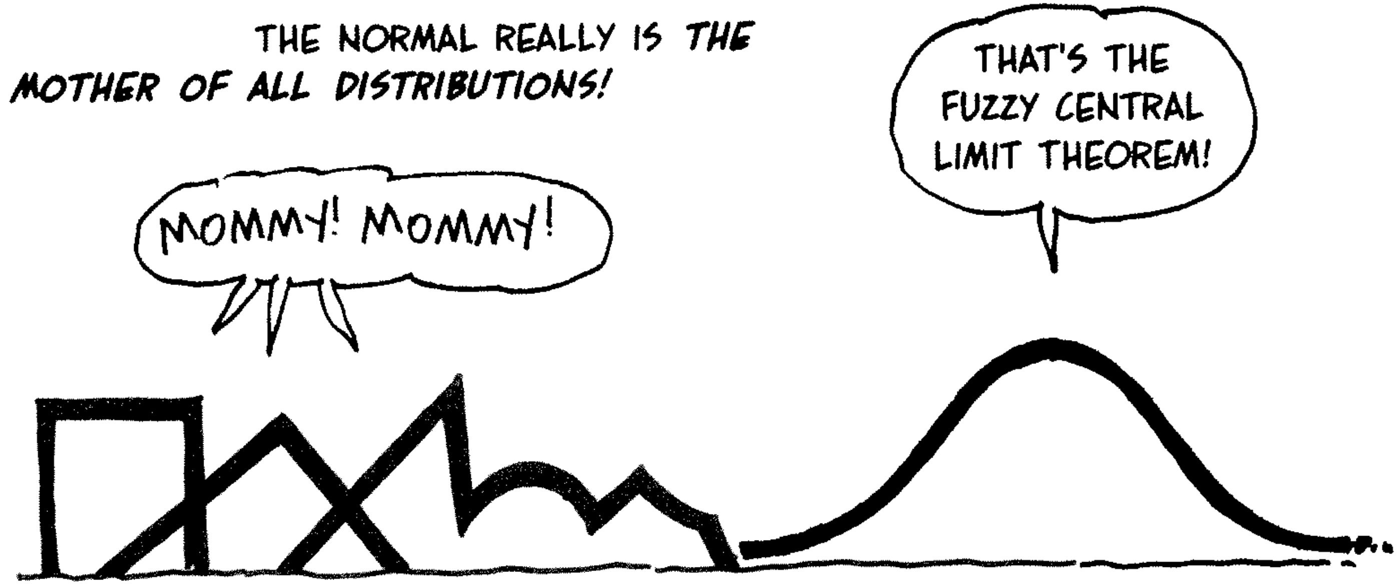
Importance du TCL

Hypothèse de nombreuses méthodes statistiques : loi normale.
En pratique : pas toujours vérifié.

Mais, grâce au **théorème central limite**, même si la population ne satisfait pas à la normalité, la moyenne d'un échantillon de grande taille issu de celle-ci est distribuée de façon normale

Donc : ok pour employer la plupart des outils statistiques.

Illustration



source : "The Cartoon Guide to Statistics", L. Gonick & W. Smith

Jeu 1

Chacun indique un budget pour ses soirées.
Budget moyen ?
Les budgets sont triés par ordre croissant. Quel échantillon semble le plus représentatif ?

1. Les 6 premiers budgets
2. Les 6 derniers budgets
3. les 6 valeurs centrales
4. 6 nombres pris au hasard