Estimation

Intervalle de Confiance

Pr Roch Giorgi

roch.giorgi@univ-amu.fr

SESSTIM, Faculté de Médecine, Aix-Marseille Université, Marseille, France
http://sesstim.univ-amu.fr

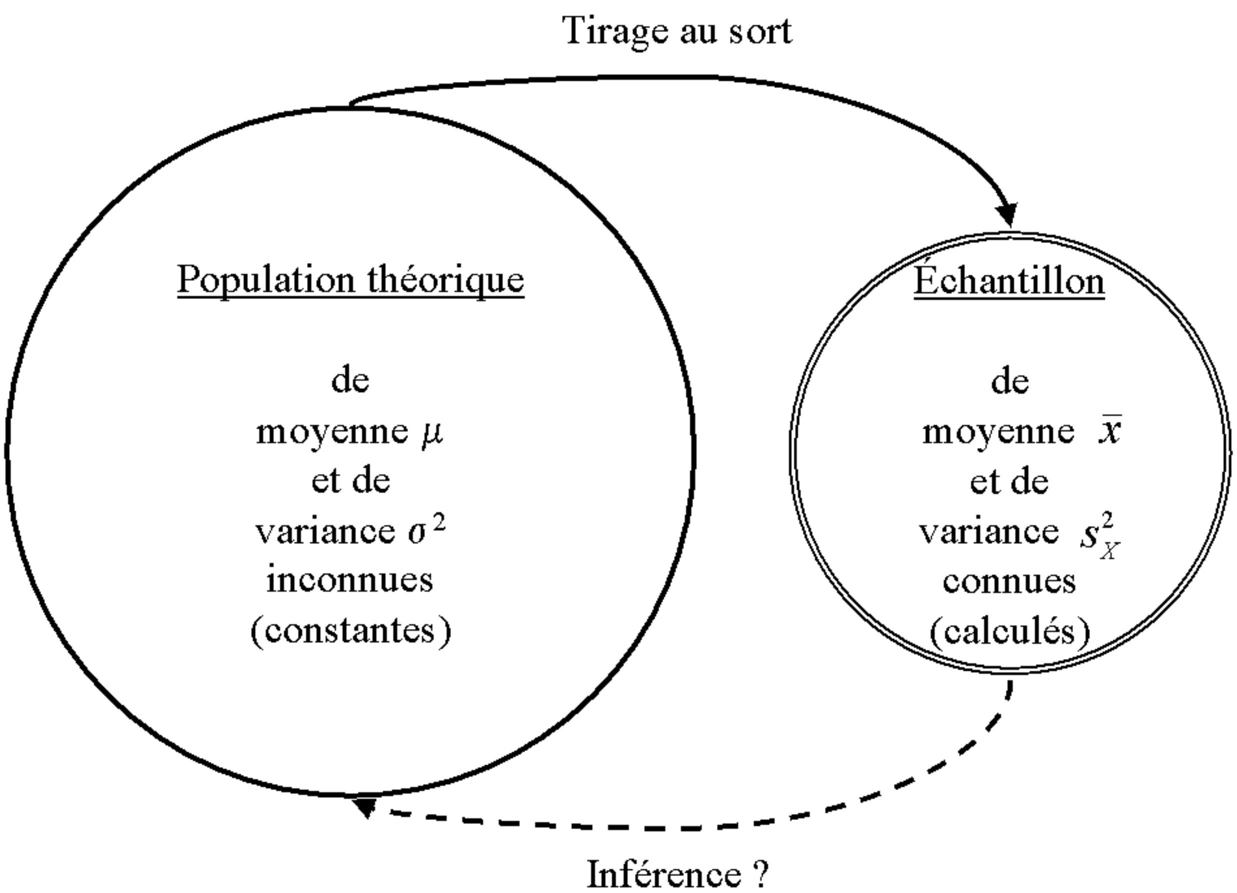
Introduction

- Connaître des valeurs de certaines grandeurs grâce à des observations réalisées sur un échantillon
  - ✓ Fréquence de la survenue du mélanome malin ?
  - ✓ Fréquence des infections nosocomiales ?
  - ✓ Valeur de la glycémie d'un patient ?
  - ✓ Variance de la glycémie mesurée chez ce patient ?

- Valeur la plus vraisemblable : estimation ponctuelle

- Intervalle de valeurs possibles, compatibles avec les observations : intervalle de confiance

Estimateur - Estimation

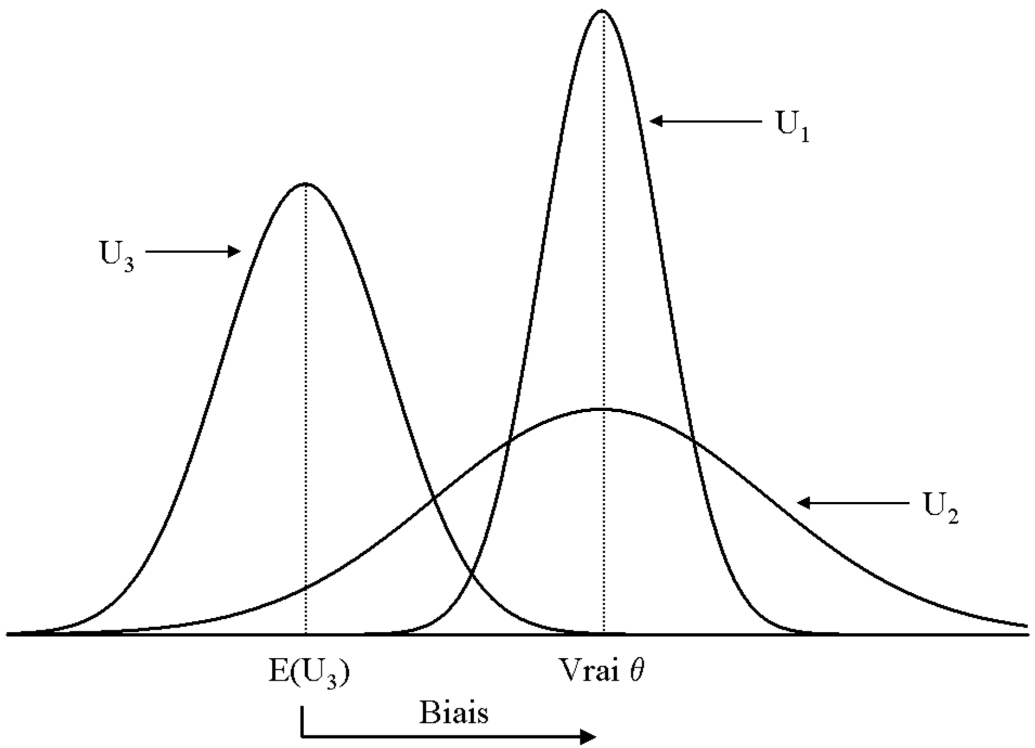


Estimateur

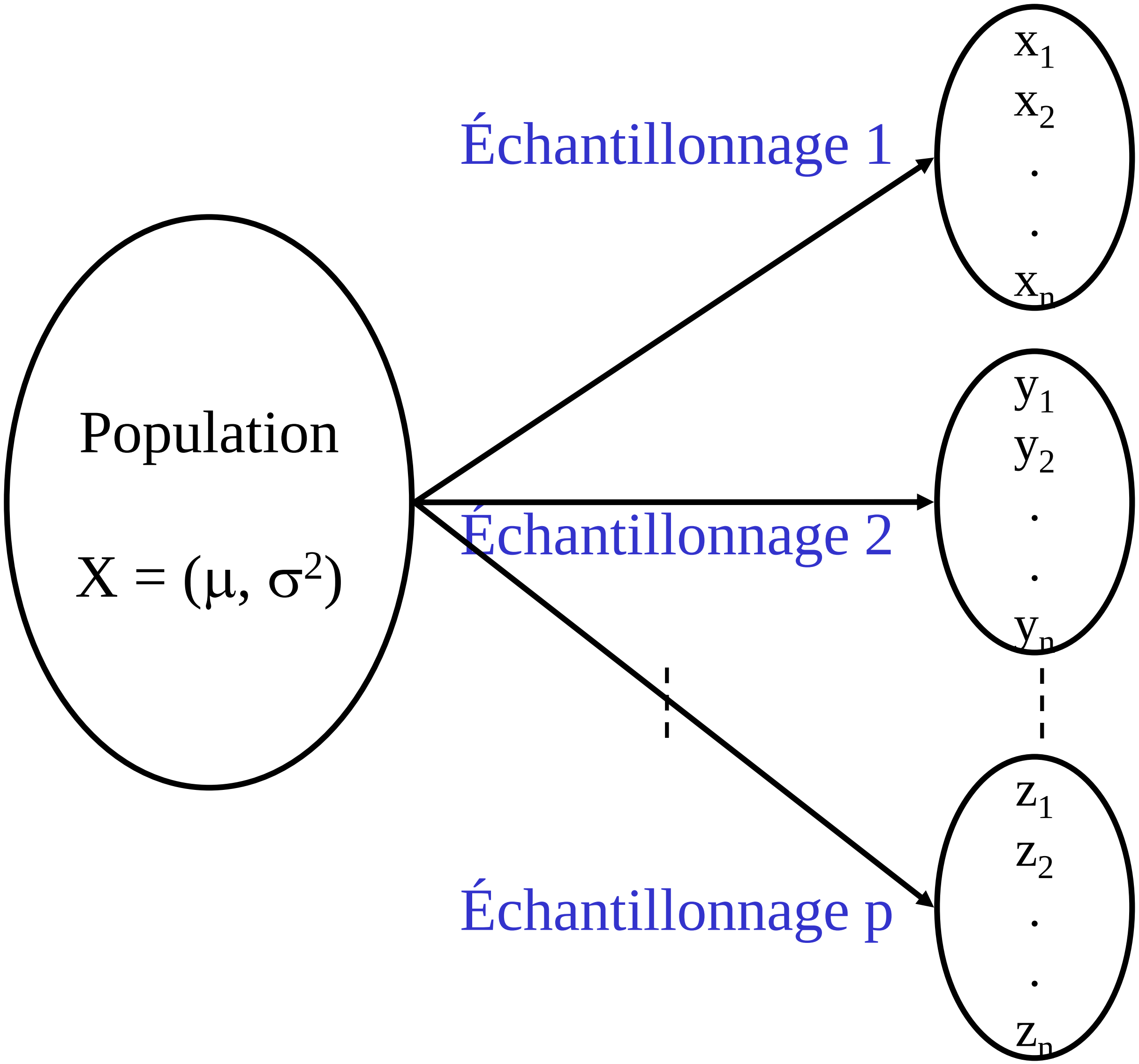
Estimation

Qualité d'un Estimateur

- U : estimateur sans biais de θ si $E(U) = \theta$
- U : estimateur biaisé de θ si $E(U) \neq \theta$ et le biais = $E(U) - \theta$



Estimation de la Moyenne d'une Population (1)



$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\bar{y} = \frac{y_1 + y_2 + \ldots + y_n}{n} = \frac{\sum_{i=1}^{n} y_i}{n}$$

$$\bar{z} = \frac{z_1 + z_2 + \ldots + z_n}{n} = \frac{\sum_{i=1}^{n} z_i}{n}$$

Estimation de la Moyenne d'une Population (2)

$$\boxed{\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}}$$

$\bar{X}$ est l'estimateur de μ

C'est un estimateur sans biais car $E\left(\bar{X}\right) = \mu$

$$E\left(\bar{X}\right) = \frac{1}{n} E\left(X_1 + X_2 + \ldots + X_n\right) = \frac{1}{n}\left[E\left(X_1\right) + E\left(X_2\right) + \ldots + E\left(X_n\right)\right]$$

Comme l'espérance de la moyenne arithmétique des n variables est la moyenne arithmétique des espérances

$E\left(\bar{X}\right) = \mu$ et est **sans biais**

Estimation de la Variance d'une Population

- Soit x1, x2, …, xn un échantillon tiré au hasard, d'effectif n et de moyenne $\bar{x} = \sum x_i / n$

- L'estimation de la variance de la population est

$$\boxed{s_x^2 = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x} \right)^2}{n-1}}$$

- S_x^2 est un estimateur sans biais convergent de σ^2
- s_x est une bonne estimation de l'écart-type de la population

Estimation d'une Proportion d'une Population

- Soit k le nombre de fois où un caractère donné est présent dans un échantillon tiré au hasard d'effectif n
- Soit p la proportion inconnue du caractère étudié dans la population
- Fréquence f du caractère étudié dans l'échantillon

$$\boxed{f = \frac{k}{n}}$$

- On montre que E(F) = p

Estimation de la Variance d'une Proportion

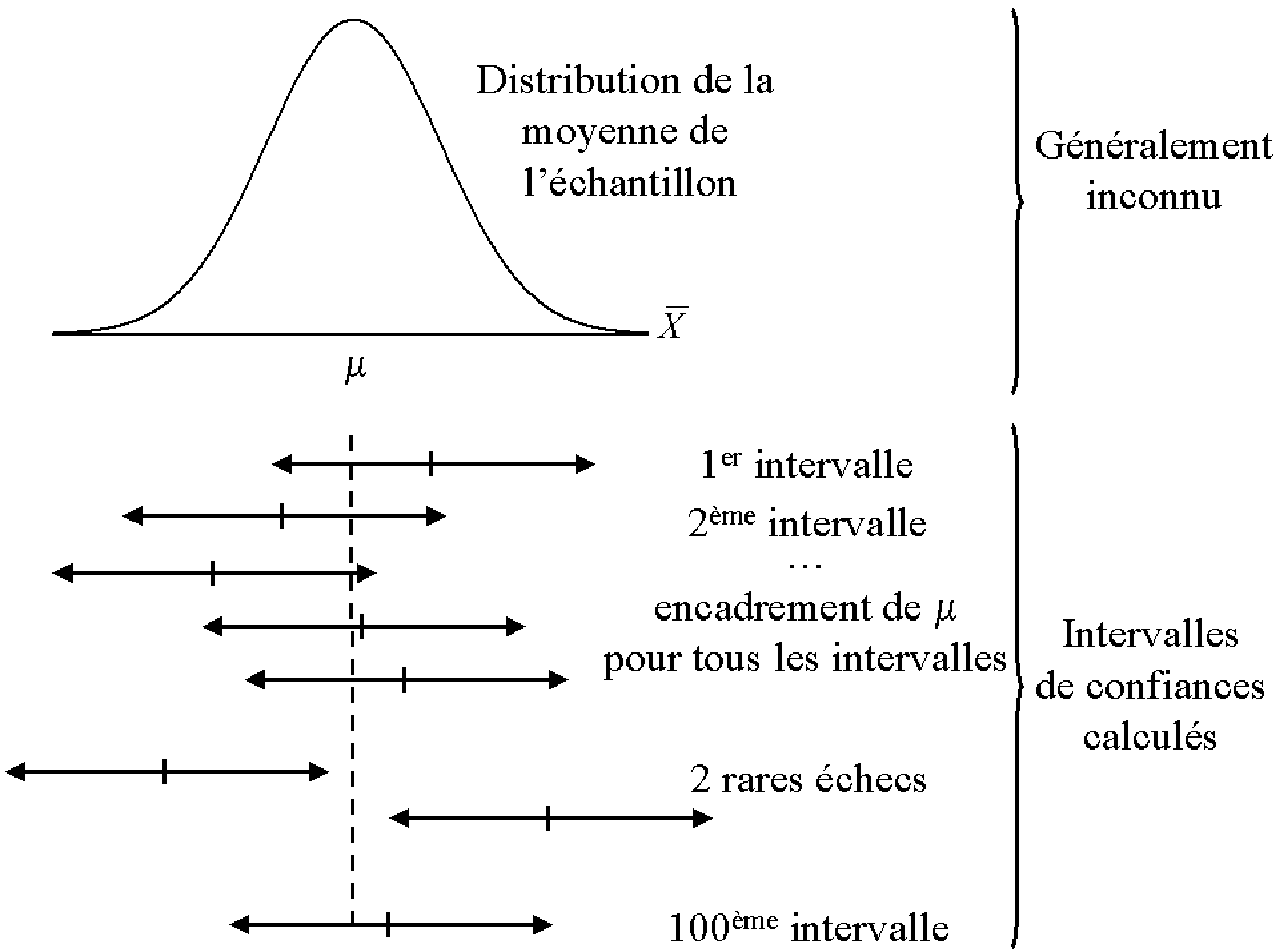
$$\boxed{Var(F) = \frac{p \cdot (1 - p)}{n}}$$

- F est un estimateur convergent de p
- On estime la variance p.(1 – p)/n par f.(1-f)/n

Estimation par Intervalle (1)

- Soit Σ un paramètre inconnu (une moyenne ou une proportion) estimé à partir d'un échantillon au hasard par θ
- On souhaite avoir un degré de confiance acceptable comme quoi θ approche bien Σ
- Intervalle de confiance : déterminé à partir des données d'un échantillon dans lequel on peut parier, avec un risque de se tromper qui soit acceptable, que se situe réellement Σ dans la population

Estimation par Intervalle (2)



Estimation par Intervalle (3)

- Risque (α) généralement = 0,05 ; il correspond aux erreurs d'échantillonnages jugées acceptables
- Intervalle de confiance de Σ est de la forme

  θ - erreur d'échantillonnage ; θ + erreur d'échantillonnage
- Interprétation
  - ✓ On accepte qu'il y ait α.100 chances sur cent de se tromper en disant que Σ appartient à l'intervalle
  - ✓ On accepte qu'il y ait (1-α).100 chances sur cent de ne pas se tromper en disant que Σ appartient à l'intervalle

Intervalle de Confiance d'une Moyenne

$$\boxed{\bar{x} - L_\alpha \cdot \frac{s_x}{\sqrt{n}} \ ; \ \bar{x} + L_\alpha \cdot \frac{s_x}{\sqrt{n}}}$$

- Si $n \geq 30$ alors, $L_\alpha = N_\alpha$ ($N_{\alpha=5\%} = 1{,}96$)
- Si $n < 30$ et la loi de distribution de la variable dans la population est Normale alors, $L_\alpha = t_{\alpha,\nu}$

Intervalle de Confiance d'une Proportion

- Si f = k/n n'est pas voisin de 1 ou de 0
- Si f*n ≥ 5 et (1 - f)*n ≥ 5

$$\boxed{f - N_{\alpha} \cdot \sqrt{\frac{f \cdot (1 - f)}{n}} \ ; \ f + N_{\alpha} \cdot \sqrt{\frac{f \cdot (1 - f)}{n}}}$$

Estimation de la Prévalence

- Mesure du risque de maladie dans la population
- Proportion de malades présents (M^+) dans la population (N) à un instant donné

$$\boxed{P = \frac{M^+}{N}}$$

✓ C'est une probabilité

✓ Intègre la notion de durée de la maladie (↗ durée maladie $\Rightarrow$ ↗ nombre de M^+ $\Rightarrow$ ↗ prévalence)

✓ Intègre la notion de vitesse d'apparition des nouveaux M^+ (↗ vitesse d'apparition $\Rightarrow$ ↗ prévalence)

Estimation de l'Incidence

- Quantifie la production de nouveaux cas de maladie dans la population dans un certain intervalle de temps

$$\boxed{I = \frac{\text{nb nouveaux cas pendant } \Delta t}{N \cdot \Delta t}}$$

Estimation de l'Effet d'un Facteur Pronostique

- Risque = probabilité d'apparition d'un événement, d'une maladie
  - ✓ Risque d'infarctus du myocarde
  - ✓ Risque de récidive d'un cancer après rémission

- La probabilité d'apparition est-elle modifiée par la présence ou l'absence d'un facteur (pronostique) ?

Risque Relatif

- Considérons le cas où la maladie (M) est présente (M+) ou absente (M-) avec un seul facteur (F) qui peut être présent (F+) ou absent (F-)
  - ✓ Risque d'apparition de la maladie chez les exposés au facteur F : P(M+/F+)
  - ✓ Risque d'apparition de la maladie chez les non exposés au facteur F : P(M+/F-)
- Risque relatif : indicateur de l'influence du facteur

$$\boxed{RR = \frac{P(M+/F+)}{P(M+/F-)}}$$

Le RR varie de 0 à l'infini

Interprétation du Risque Relatif

- Si RR > 1
  - ✓ $\Leftrightarrow$ P(M+/F+) > P(M+/F-)
  - ✓ La présence du facteur F « favorise la maladie » = facteur de risque
- Si RR < 1
  - ✓ $\Leftrightarrow$ P(M+/F+) < P(M+/F-)
  - ✓ La présence du facteur F « favorise la non maladie » = facteur protecteur
- Si RR = 1
  - ✓ $\Leftrightarrow$ P(M+/F+) = P(M+/F-)
  - ✓ Le facteur F n'a pas d'effet sur la maladie

Estimation du Risque Relatif (1)

	M+	M-	Total
F+	a	b	m_1
F-	c	d	m_2
Total	n_1	n_2	N

$$P(M+/F+) = a/m_1$$

$$P(M+/F-) = c/m_2$$

$$\boxed{RR = \frac{a/m_1}{c/m_2}}$$

Estimation du Risque Relatif (2)

- Le RR peut être estimé
  - ✓ Dans les enquêtes sur un seul échantillon au hasard
  - ✓ Dans les enquêtes exposés / non exposés
- Le RR ne peut pas être estimé
  - ✓ Dans les enquêtes cas-témoins. Dans ce cas on peut estimer un Odds Ratio, (a.d)/(b.c), qui s'interprète qualitativement comme le RR par rapport à 1
- En plus de l'estimation ponctuelle du RR il est nécessaire de calculer son intervalle de confiance