Chapitre 2 : Estimation

Said, El Melhaoui

Faculté des Sciences Juridiques, Économiques et Sociales Oujda

http://said-el-melhaoui.e-monsite.com

Outline

Exemple introductif

1. **Population** U
   $X \sim \mathcal{N}(\mu, 1)$, où μ est un paramètre inconnu appartenant à un espace de paramètres Θ (par exemple, $\Theta = \mathbb{R}$)
2. **Objectif**
   On veut estimer μ
3. **Échantillon aléatoire** $\{X_1, X_2, \ldots, X_n\}$
   Les X_i sont supposées i.i.d (issu par un ESAR)
   Rappel : X_i désigne la v. a. et x_i la valeur prise par celle-ci
4. **Loi de** X_i

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}} \quad x \in \mathbb{R}$$

$\Rightarrow f(x) = f(x, \mu)$: la loi dépend de μ

Exemple introductif (suite)

5 **Loi du vecteur** $\{X_1, X_2, \ldots, X_n\}$
Les X_i sont des v.a. i.i.d. $\Rightarrow$ la loi jointe du vecteur est

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i)$$

Notation: $\prod_{i=1}^{n} a_i = \prod_{i=1} a_i = a_1 \times a_2 \ldots \times a_n$

$$\begin{aligned} f(x_1, \ldots, x_n) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2}} \\ &= \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2} \\ \Rightarrow f(x_1, \ldots, x_n) &= f(x_1, \ldots, x_n\,;\,\mu) \end{aligned}$$

Exemple introductif (suite)

6 **Comment estimer μ?**
Estimateur: $\hat{\mu} = \hat{\mu}(X_1, X_2, \ldots, X_n)$
Exemples:

- $\hat{\mu}_1 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i = \overline{X}$
- $\hat{\mu}_2 = \dfrac{1}{k}\displaystyle\sum_{i=1}^{k} X_i \quad k \in \{1, \ldots, n\}$
- $\hat{\mu}_3 = X_i \quad i \in \{1, \ldots, n\}$
- $\hat{\mu}_4 = X_{1/2}$ (la médiane)

Comment choisir un estimateur? Quelles sont ses propriétés?

N. B. La v. a. $\hat{\mu}(X_1, \ldots, X_n)$ est un estimateur de μ
la valeur observée $\hat{\mu}(x_1, \ldots, x_n)$ est une estimation de μ

Définitions et notations

- Population U : v. a. X où la loi-type est donnée:
  a. Cas discret: $\{(x, p_x), \quad x \in \mathbb{N}\}$
  b. Cas continu : $f(x), \quad x \in \mathbb{R}$
- Paramètres à estimer : $\theta_1, \ldots, \theta_K$ $(\theta_k \in \Theta_k)$
  N. B. L'ensemble des paramètres est noté θ $(\theta \in \Theta)$
- Échantillon aléatoire $\{X_1, X_2, \ldots, X_n\}$ où la loi de X_i dépend de θ:
  a. Cas discret: $p_{x_i} = p_{x_i}(\theta)$
  b. Cas continu : $f(x_i) = f(x_i; \theta)$

  N. B. Les X_i sont supposées i.i.d.

Méthode des moments: Rappel

- Moments simples de la loi de X:

$$\mu_k = E(X^k) = \mu_k(\theta_1, \ldots, \theta_K) \text{ où } k = 1, \ldots, K$$

- Moments simples (empiriques) de l'échantillon:

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \text{ où } k = 1, \ldots, K$$

Méthode des moments : Définition des estimateurs

Définition

Les estimateurs $\hat{\theta}_1, \dots, \hat{\theta}_K$ des paramètres $\theta_1, \dots, \theta_K$ sont déterminés en résolvant le système de K équations à K inconnus:

$$\begin{cases} \mu_1(\theta_1, \dots, \theta_K) & = m_1 \\ \mu_2(\theta_1, \dots, \theta_K) & = m_2 \\ \vdots & \vdots \\ \mu_K(\theta_1, \dots, \theta_K) & = m_K \end{cases}$$

Méthode des moments: Exemple 1

- Population : $X \sim \mathcal{P}(\lambda), \quad \lambda \in \mathbb{R}^{+*}$
  On désire estimer λ (un seul paramètre inconnu)
- Moment population : $\mu_1(\lambda) = E(X) = \lambda$
  Moment échantillon : $m_1 = \overline{X}$
- Équation:

$$\mu_1(\lambda) = m_1 \Leftrightarrow \hat{\lambda} = \overline{X}$$

- Conclusion : L'estimateur du paramètre inconnu λ obtenu par la méthode des moments est donné par $\hat{\lambda} = \overline{X}$

Méthode des moments : Exemple 2

- Population : $X \sim \mathcal{N}(\mu, \sigma^2), \;\; \mu \in \mathbb{R}, \; \sigma^2 \in \mathbb{R}^{+*}$
  On désire estimer μ et σ^2 (2 paramètres inconnus)
- Moments population :

$$\begin{aligned} \mu_1(\mu, \sigma^2) &= E(X) = \mu \\ \mu_2(\mu, \sigma^2) &= E(X^2) = \sigma^2 + \mu^2 \end{aligned}$$

Moments échantillon :

$$\begin{aligned} m_1 &= \overline{X} \\ m_2 &= s^2 + \overline{X}^2 \end{aligned}$$

Rappel: $s^2 = \frac{1}{n} \sum_i X_i^2 - \overline{X}^2 = m_2 - \overline{X}^2$

Méthode des moments : Exemple 2 (suite)

- Équations:

$$\begin{cases} \mu_1(\mu,\sigma^2) & = m_1 \\ \mu_2(\mu,\sigma^2) & = m_2 \end{cases} \Leftrightarrow \begin{cases} \mu & = \overline{X} \\ \sigma^2+\mu^2 & = s^2+\overline{X}^2 \end{cases}$$

- Conclusion : Les estimateur des paramètres inconnus μ et σ^2 obtenus par la méthode des moments sont donnés par

$$\begin{cases} \hat{\mu} = \overline{X} \\ \hat{\sigma}^2 = s^2 \end{cases}$$

Idée de la méthode du maximum de vraisemblance

- Supposons devoir estimer $\theta \in \Theta$
- Pour estimer θ on choisit dans Θ la valeur qui assure à l'échantillon la plus grande '***vraisemblance***': apparence de vérité ; crédibilité ; probabilité d' d'apparence (de réalisation)
- Loi de l'échantillon $\{X_1, X_2, \dots, X_n\}$
  a. Cas discret:
  $$p(x_1, \dots, x_n; \theta) = \Pi_i p_{x_i}(\theta)$$
  Fonction de vraisemblance :
  $$L(\theta) = p(x_1, \dots, x_n; \theta)$$
  b. Cas continu :
  $$f(x_1, \dots, x_n; \theta) = \Pi_i f(x_i)$$
  Fonction de vraisemblance :
  $$L(\theta) = f(x_1, \dots, x_n; \theta)$$

Méthode du maximum de vraisemblance

Définition des estimateurs MV

Valeur $\hat{\theta} \in \Theta$ telle que $L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta)$

Équation du maximum de vraisemblance : Si $L(\theta)$ est différentiable et possède au point $\hat{\theta}$ un maximum sur Θ, $\hat{\theta}$ est solution de l'équation du M. V. :

$$\frac{dL(\theta)}{d\theta} \Big|_{\theta=\hat{\theta}} = 0$$

Remarque

Si $\hat{\theta}$ correspond au maximum de $L(\theta)$, il correspond aussi au maximum de $L^*(\theta) = \log(L(\theta))$ obtenu en résolvant l'équation:

$$\frac{d \log(L(\theta))}{d\theta} \Big|_{\theta=\hat{\theta}} = 0$$

Exemple 1

- **Population :**régie par une loi normale $X \sim \mathcal{N}(\mu, 1)$, où $\mu \in \mathbb{R}$
- **Échantillon aléatoire:** $\{X_1, X_2, \ldots, X_n\}$
- **Loi de X_i**

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}} \quad x \in \mathbb{R}$$

- Fonction de vraisemblance:

$$\begin{aligned} L(\mu) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2}} \\ &= \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\displaystyle\sum_{i=1}^{n}(x_i-\mu)^2} \end{aligned}$$

Exemple 1 (suite)

- Fonction de log-vraisemblance:

$$L^*(\mu) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_i (x_i - \mu)^2$$

- Équation du M. V. : $\frac{d\log(L(\mu))}{d\mu}\,|_{\mu=\hat{\mu}} = 0$

$$\sum_i (x_i - \hat{\mu}) = 0 \Leftrightarrow \sum_i x_i - n\hat{\mu} = 0$$

$\Rightarrow$ Estimateur du M. V. : $\hat{\mu} = \frac{1}{n}\sum_i X_i = \overline{X}$

N. B. Maximum car $\frac{d^2L^*(\mu)}{d\mu^2} = -n < 0$

Exemple 2

- **Population:** régie par une loi uniforme $X \sim \mathcal{U}_{[0,\theta]}$, où $0 < \theta < +\infty$
- **Loi de X_i :**

$$f(x) = \begin{cases} \frac{1}{\theta} & 0 \leq x < \theta \\ 0, & \text{aileurs} \end{cases}$$

- **Fonction de vraisemblance:**

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^{n} f(x_i) = \begin{cases} \frac{1}{\theta^n} & \text{si tous les } x_i \in [0, \theta] \\ 0, & \text{aileurs} \end{cases}$$

- **Équation du M. V.** : $\dfrac{dL(\theta)}{d\theta} = \dfrac{-n}{\theta^{n+1}} < 0$

  $\Rightarrow$ La plus grande valeur de $L(\theta)$ est donnée par la plus petite valeur possible de θ; comme $\theta \geq x_i, \quad \forall i = 1, \dots, n$, l'estimateur du M. V.est

$$\hat{\theta} = X_{(n)} = \max_i X_i$$

Existence et unicité de l'estimateur du M. V.

Propriété

Un estimateur MV n'existe pas toujours, n'as pas nécessairement une forme explicite et n'est pas nécessairement unique

Estimateur sans biais

Définition

$\hat{\theta}$ est un estimateur sans biais de θ si

$$E[\hat{\theta}] = \theta \quad \forall \theta \in \Theta$$

- **Exemple:** $\hat{\mu} = \overline{X}$ comme estimateur de μ la moyenne de X: $E(\hat{\mu}) = E(\overline{X}) = \mu$
- **Biais d'un estimateur:**

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta$$

- **Estimateur asymptotiquement sans biais:** Estimateur $\hat{\theta}_n$ de θ tel que

$$\lim_{n \mapsto \infty} E(\hat{\theta}_n) = \theta$$

Mesure de précision d'un estimateur: MSE

Définition

Erreur quadratique moyenne [Mean Squared Error (MSE)] d'un estimateur $\hat{\theta}$ d'un paramètre θ est:

$$\text{MSE}(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right]$$

- **MSE=Biais et variance:**

$$\begin{aligned}
\text{MSE}(\hat{\theta}) &= E\left[(\hat{\theta} - \theta)^2\right] = E\left[(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2\right] \\
&= E\left[(\hat{\theta} - E(\hat{\theta}))^2\right] + \left[E(\hat{\theta}) - \theta\right]^2 \\
&\quad -2\left[E(\hat{\theta}) - \theta\right]\underbrace{E\left[(\hat{\theta} - E(\hat{\theta}))\right]}_{0} \\
&= V(\hat{\theta}) + B^2(\hat{\theta})
\end{aligned}$$

MSE (suite)

- Cas d'un estimateur sans biais:

$$\text{MSE}(\hat{\theta}) = V(\hat{\theta})$$

- Mesure de précision d'un estimateur:

$$\frac{1}{\text{MSE}(\hat{\theta})}$$

Efficacité relative

- **Efficacité relative:** Si $\hat{\theta}_1$ et $\hat{\theta}_2$ sont des estimateurs quelconque de θ, l'efficacité relative de $\hat{\theta}_1$ par rapport à $\hat{\theta}_2$ est définie par

$$E_r(\hat{\theta}_1, \hat{\theta}_2) = \frac{\text{MSE}(\hat{\theta}_2)}{\text{MSE}(\hat{\theta}_1)}$$

- $\hat{\theta}_1$ est plus efficace que $\hat{\theta}_2$ si

$$E_r(\hat{\theta}_1, \hat{\theta}_2) \geq 1$$

- **Exemple:** (estimators d'une moyenne μ)

$$\left.\begin{array}{lcl} \hat{\mu}_1 = \overline{X} & : & MSE(\overline{X}) = \frac{\sigma^2}{n} \\ \hat{\mu}_2 = X_1 & : & MSE(X_1) = \sigma^2 \end{array}\right\} \Rightarrow E_r(\hat{\mu}_1, \hat{\mu}_2) = n$$

$\overline{X}$ est plus efficace que X_1

Efficacité absolue

- **Objectif poursuivi:** Rechercher un estimateur sans biais $\hat{\theta}^*$ d'un paramètre θ tel que
$$V(\hat{\theta}^*) \leq V(\hat{\theta})$$
pour tout $\hat{\theta}$ estimateur sans biais de θ
$\Rightarrow$ Recherche d'un estimateur sans biais de variance minimum (SBVM)
- **Borne de Cramer-Rao:** On suppose remplies certaines conditions de régularités de la fonction $L(\theta)$ alors :
$$V(\hat{\theta}) \geq \frac{1}{E\left[\left(\frac{d \log L(\theta)}{d\theta}\right)^2\right]}$$

Efficacité absolue (suite)

- **Information de Fisher:**

$$I_n(\theta) = E\left[\left(\frac{d \log L(\theta)}{d\theta}\right)^2\right]$$

- **Estimateur efficace (SBVM):** Si $\hat{\theta}^*$ est un estimateur sans biais de θ, il est efficace si

$$V(\hat{\theta}^*) = \frac{1}{I_n(\theta)}$$

- **N. B.**

$$I_n(\theta) = E\left(\frac{-d^2 \log L(\theta)}{d\theta^2}\right)$$

Exemple 1

- **Population :**régie par une loi normale
  $X \sim \mathcal{N}(\mu, 1)$,   où $\mu \in \mathbb{R}$ et $\hat{\mu} = \overline{X}$
- **Échantillon aléatoire:** $\{X_1, X_2, \ldots, X_n\}$
- Fonction de log-vraisemblance:

$$L^*(\mu) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_i (x_i - \mu)^2$$

- $\dfrac{d \log(L(\mu))}{d\mu} = \sum_i (x_i - \mu)$

Exemple 1 (suite)

- L'information de Fisher :

$$\begin{aligned} I_n(\mu) &= E\left[\left(\frac{d \log L(\mu)}{d\mu}\right)^2\right] = E\left[\left(\sum_i (x_i - \mu)\right)^2\right] \\ &= \sum_i E\,(x_i - \mu)^2 \text{ car } X_i \text{ indépendants} \\ &= \sum_i \underbrace{E\left[(x_i - \mu)^2\right]}_{\sigma^2 = 1} = n \end{aligned}$$

- Or $V(\hat{\mu}) = \sigma^2/n = 1/n = 1/I_n(\mu) \Rightarrow$ cet estimateur est SBVM (efficace)

Estimateur convergent

Définition

Soit $\hat{\theta}_n$ un estimateur de θ, il est convergent si

$$\lim_{n \mapsto +\infty} P\left(|\hat{\theta}_n - \theta| < \varepsilon\right) = 1$$

Notation: $\hat{\theta}_n \xrightarrow{P} \theta$

Propriété : CS de convergence

Si $\hat{\theta}_n$ est un estimateur de θ tel que

$$\lim_{n \mapsto +\infty} E(\hat{\theta}) = \theta \text{ et } \lim_{n \mapsto +\infty} V(\hat{\theta}) = 0$$

alors : $\hat{\theta}_n \xrightarrow{P} \theta$

Exemple

- $X \sim \mathcal{N}(\mu, \sigma^2)$
- Échantillon : $\{X_1, X_2, \ldots, X_n\}$
- $\hat{\mu} = \overline{X}$
- $E(\hat{\mu}) = \mu$
- $V(\hat{\mu}) = \sigma^2/n \xrightarrow{n} 0$

$$\Rightarrow \overline{X} \xrightarrow{P} \mu$$

Exercice

Soit l'estimateur s^2 de la variance σ^2 défini par:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

1. Montrer que l'estimateur s^2 est biaisé, mais asymptotiquement sans biais
2. Montrer qu' estimateur sans biais et convergent de σ^2 est:

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

Intervalle de confiance

- **Objectif:** Rechercher un intervalle $\hat{\Theta} \subset \Theta$ tel que la probabilité $P(\theta \in \hat{\Theta})$ soit fixé a priori à un niveau 'élevé'
- **Niveau de confiance:**
  - Soit un seuil d'incertitude $\alpha \in]0,1[$, alors

$$P(\theta \notin \hat{\Theta}) = \alpha \Leftrightarrow P(\theta \in \hat{\Theta}) = 1 - \alpha$$

  $\Rightarrow \quad 1 - \alpha =$ niveau de confiance
  - Niveaux de confiance usuels: 0.90; 0.95 ; 0.99
- **Intervalle de confiance:** Soit un paramètre $\theta \in \Theta \subset \mathbb{R}$. On désire construire un intervalle (ℓ_1, ℓ_2) tel que

$$P(\ell_1 \leq \theta \leq \ell_2) = 1 - \alpha$$

  où $1 - \alpha$ est un niveau de confiance choisi a priori

Intervalle de confiance

Remarques

- ℓ_1 et ℓ_2 sont des variables aléatoires
- L'intervalle peut être bilatéral ou unilatéral
  - Cas bilatéral : on impose que

$$P(\theta < \ell_1) = \alpha/2 \text{ et } P(\ell_2 < \theta) = \alpha/2$$

  - Cas unilatéral :

$$\begin{aligned} \ell_2 &= +\infty \text{ et } P(\ell_1 > \theta) = \alpha \\ \ell_1 &= -\infty \text{ et } P(\ell_2 < \theta) = \alpha \end{aligned}$$

Intervalle de confiance pour la moyenne: Cas des petits échantillons

- Soit $X \sim \mathcal{N}(\mu, \sigma^2)$ où $\mu \in \mathbb{R}$ et $\sigma^2 \in \mathbb{R}^{*+}$
- Échantillon : $\{X_1, X_2, \ldots, X_n\}$, où la taille $n < 30$
- Estimateur de μ : $\hat{\mu} = \overline{X}$
- Loi de $\hat{\mu}$: $\overline{X} \sim \mathcal{N}(\mu, \sigma^2/n)$
- Loi centrée réduite :
$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$
- La distribution dépend de σ, on distingue alors deux cas selon σ connu ou inconnue

I. C. pour moyenne: Cas des petits échantillons avec σ connu

Construction un intervalle de confiance de μ au degré de confiance $(1-\alpha)$: $IC_{1-\alpha}(\mu)$

- **Rappel:** si $Z \sim \mathcal{N}(0,1)$ alors

$$P(Z \leq z_{1-\alpha/2}) = 1 - \frac{\alpha}{2}$$

Il en résulte que

$$P(-z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 1 - \alpha$$

N. B. $-z_{1-\alpha/2} = z_{\alpha/2}$

I. C. pour moyenne: Cas des petits échantillons avec σ connu

Comme $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$

$$\begin{aligned}
&\Rightarrow \quad P\left(-z_{1-\alpha/2} \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq z_{1-\alpha/2}\right) = 1-\alpha \\
&\Leftrightarrow \quad P\left(-z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \overline{X}-\mu \leq z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha \\
&\Leftrightarrow \quad P\left(-\overline{X}-z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\overline{X}+z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha \\
&\Leftrightarrow \quad P(\underbrace{\overline{X}-z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\ell_1} \leq \mu \leq \underbrace{\overline{X}+z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\ell_2}) = 1-\alpha
\end{aligned}$$

$$\Rightarrow IC_{1-\alpha}(\mu) = \left[\overline{X}-z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X}+z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

I. C. pour moyenne: Cas des petits échantillons avec σ inconnu

- Puisque σ^2 est inconnu, il faut l'estimer par $\hat{\sigma}^2$
- Soit la variance empirique :

$$s^2 = \frac{1}{n}\sum_i (X_i - \overline{X})^2$$

On montre que $E(s^2) = \dfrac{n-1}{n}\sigma^2 \neq \sigma^2$, le s^2 est biaisé

- Ajustons le afin d'éliminer le biais, soit

$$\begin{aligned} S^2 &= \frac{n}{n-1}s^2 \\ &= \frac{1}{n-1}\sum_i (X_i - \overline{X})^2 \end{aligned}$$

Il est clair que $E(S^2) = \dfrac{n}{n-1}E(s^2) = \sigma^2$, le S^2 est sans biais

- Remplaçons donc $\sigma/\sqrt{n}$ par $S/\sqrt{n}$ ou par $s/\sqrt{n-1}$

I. C. pour moyenne: Cas des petits échantillons avec σ inconnu

La substitution de σ change la loi de la statistique: $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$

$$\Rightarrow \quad P(-t_{1-\alpha/2} \leq \frac{\overline{X}-\mu}{S/\sqrt{n}} \leq t_{1-\alpha/2}) = 1-\alpha$$

$$\Leftrightarrow \quad P(-t_{1-\alpha/2}\frac{S}{\sqrt{n}} \leq \overline{X}-\mu \leq t_{1-\alpha/2}\frac{S}{\sqrt{n}}) = 1-\alpha$$

$$\Leftrightarrow \quad P(-\overline{X} - t_{1-\alpha/2}\frac{S}{\sqrt{n}} \leq -\mu \leq -\overline{X} + t_{1-\alpha/2}\frac{S}{\sqrt{n}}) = 1-\alpha$$

$$\Leftrightarrow \quad P(\underbrace{\overline{X} - t_{1-\alpha/2}\frac{S}{\sqrt{n}}}_{\ell_1} \leq \mu \leq \underbrace{\overline{X} + t_{1-\alpha/2}\frac{S}{\sqrt{n}}}_{\ell_2}) = 1-\alpha$$

$$\Rightarrow IC_{1-\alpha}(\mu) = \left[\overline{X} - t_{1-\alpha/2;n-1}\frac{S}{\sqrt{n}}, \overline{X} + t_{1-\alpha/2;n-1}\frac{S}{\sqrt{n}}\right]$$

Exercice

- **Énonce:** Un échantillon de taille 17, provenant d'une population normale, a les caractéristiques suivantes : $\bar{x} = 4.7$ et $S^2 = 5.76$. Déterminons un I.C de μ au niveau de confiance 95%
- **Solution:** Il s'agit du cas des petit échantillon, avec σ inconnu, ainsi

$$
\begin{aligned}
IC_{1-\alpha}(\mu) &= \left[\overline{x} \pm t_{1-\alpha/2} \frac{S}{\sqrt{n}}\right] \\
&= \left[4.7 \pm t_{0.975;16} \frac{\sqrt{5.76}}{\sqrt{17}}\right] \\
&= \mathbf{[3.47, 5.93]} \quad (t_{0.975;16} = 2.12)
\end{aligned}
$$

Intervalle de confiance pour la moyenne: Cas des grands échantillons

- Soit $X \sim \mathcal{L}(\mu, \sigma^2)$ où $\mu \in \mathbb{R}$ et $\sigma^2 \in \mathbb{R}^{*+}$
- Échantillon : $\{X_1, X_2, \ldots, X_n\}$, où la taille $n \geq 30$
- Estimateur de μ : $\hat{\mu} = \overline{X}$
- Loi de $\hat{\mu}$ (TCL): $\overline{X} \approx \mathcal{N}(\mu, \sigma^2/n)$
- Loi centrée réduite :

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \approx \mathcal{N}(0, 1)$$

I. C. pour moyenne: Cas des grands échantillons (suite)

- **Cas σ connu:** Comme $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \approx \mathcal{N}(0,1)$ alors

$$IC_{1-\alpha}(\mu) = \left[\overline{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

- **Cas σ inconnu:** puisque S^2 est sans biais et convergeant vers σ^2 alors

$$IC_{1-\alpha}(\mu) = \left[\overline{X} - z_{1-\alpha/2}\frac{S}{\sqrt{n}}, \overline{X} + z_{1-\alpha/2}\frac{S}{\sqrt{n}}\right]$$

Intervalle de confiance pour la variance: Cas des petits échantillons

- Soit $X \sim \mathcal{N}(\mu, \sigma^2)$ où $\mu \in \mathbb{R}$ et $\sigma^2 \in \mathbb{R}^{*+}$
- Échantillon : $\{X_1, X_2, \ldots, X_n\}$, où la taille $n < 30$
- Le choix de l'estimateur de σ^2 diffère selon que μ soit connu ou inconnu

I. C. pour la variance : Cas des petits échantillons avec μ connu

- Estimateur de σ^2 :

$$\hat{\sigma}^2 = \frac{1}{n}\sum_i (X_i - \mu)^2$$

- Loi de $\hat{\sigma}^2$:

$$\begin{aligned} \frac{n\hat{\sigma}^2}{\sigma^2} &= \sum_i \left(\frac{X_i - \mu}{\sigma}\right)^2 \\ &= \sum_{i=1}^{n} Z_i^2 \sim \chi_n^2 \ \left(\sum \text{ v.a normales standards indépendantes}\right) \end{aligned}$$

I. C. pour la variance : Cas des petits échantillons avec μ connu

- **Rappel:** si $Y \sim \chi^2_n$ alors

$$P(Y \leq \chi^2_{n;\alpha/2}) = \alpha/2$$

$$P(Y \geq \chi^2_{n;1-\alpha/2}) = \alpha/2$$

Il en résulte que

$$P\left(\chi^2_{n;\alpha/2} \leq Y \leq \chi^2_{n;1-\alpha/2}\right) = 1 - \alpha$$

I. C. pour moyenne: Cas des petits échantillons avec μ connu

Construction un intervalle de confiance de σ^2 au degré de confiance $(1-\alpha)$:

$$
\begin{aligned}
P\left(\chi^2_{n;\alpha/2} \leq \frac{n\hat{\sigma}^2}{\sigma^2} \leq \chi^2_{n;1-\alpha/2}\right) &= 1-\alpha \\
\Leftrightarrow P\left(\underbrace{\frac{n\hat{\sigma}^2}{\chi^2_{n;1-\alpha/2}}}_{\ell_1} \leq \sigma^2 \leq \underbrace{\frac{n\hat{\sigma}^2}{\chi^2_{n;\alpha/2}}}_{\ell_2}\right) &= 1-\alpha
\end{aligned}
$$

$$
\Rightarrow IC_{1-\alpha}(\sigma^2) = \left[\frac{n\hat{\sigma}^2}{\chi^2_{n;1-\alpha/2}}, \frac{n\hat{\sigma}^2}{\chi^2_{n;\alpha/2}}\right]
$$

I. C. pour moyenne: Cas des petits échantillons avec μ inconnu

- Estimateur de σ^2 :

$$\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_i (X_i - \overline{X})^2$$

- Loi de S^2:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

- L'intervalle de confiance

$$IC_{1-\alpha}(\sigma^2) = \left[\frac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1;\alpha/2}}\right]$$

IC pour la variance : Cas des grands échantillons $n \geq 50$

- Soit $X \sim \mathcal{L}(\mu, \sigma^2)$ où $\mu \in \mathbb{R}$ et $\sigma^2 \in \mathbb{R}^{*+}$; $n \geq 50$
- Estimateur de μ : $\hat{\sigma}^2 = S^2$
- Loi centrée réduite (TCL):

$$\frac{S^2 - \sigma^2}{\sqrt{(\hat{\mu}_4 - S^4)/n}} \approx \mathcal{N}(0, 1)$$

avec

$$\hat{\mu}_4 = \frac{1}{n}\sum_i (X_i - \overline{X})^4$$

- L'intervalle de confiance :

$$IC_{1-\alpha}(\sigma^2) = \left[S^2 - z_{1-\alpha/2}\sqrt{(\hat{\mu}_4 - S^4)/n}, S^2 + z_{1-\alpha/2}\sqrt{(\hat{\mu}_4 - S^4)/n}\right]$$

N. B Puisque n est assez grand on a $S^2 \simeq s^2 \simeq \sigma^2$ et $\hat{\mu}_4 \simeq \mu_4$

Exercice

Énoncé: Une assurance s'intéresse à la distance moyenne parcourue par les voitures qu'elle assure, ainsi qu'a la variance des distances parcourues. Elle a observée 900 de ces véhicules les résultats sont :
$\bar{x} = 15u$, $S = 5u$ et $\hat{\mu}_4 = 800u^4$ où $u = 10^3$ km.
Déterminez les un I.C. de μ et σ^2 au niveau de confiance 95%

Intervalle de confiance pour la proportion

- Soit X une v.a. qui suit une loi de Bernoulli $B(p)$ où p est le paramètre d'intérêt
- p présente la proportion d'individus possédant une certaine modalité d'intérêt 'A'
- Soit un échantillon ESAR $\{X_1, X_2, \ldots, X_n\}$ de X
  - $X_i = \begin{cases} 1, & \text{si i possède la modalité A} \\ 0, & \text{sinon.} \end{cases}$
  - $n_A =$ nombre des élément de l'échantillon possédant la modalité A

$$n_A = \sum_{i=1}^{n} X_i$$

  - $n_A \sim \mathcal{B}(n, p)$

$$\Longrightarrow E(n_A) = np \;\; ; \;\; V(n_A) = np(1-p)$$

Intervalle de confiance pour la proportion (suite)

- Estimation de p:

$$\hat{p} = \frac{n_A}{n} = \overline{X} := \overline{p}$$

$$\Longrightarrow E(\overline{p}) = p \;\; ; \;\; V(\overline{p}) = \frac{p(1-p)}{n}$$

N. B. $\overline{p}$ est la solution unique de l'équation M.V. et il est SBMV (efficace)

- Loi de $\overline{p}$:
  1. Loi exacte : $n\overline{p} \sim \mathcal{B}(n, p)$
  2. Loi asymptotique ($n \geq 30$)

$$\frac{\overline{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \approx \mathcal{N}(0, 1)$$

N. B. Loi valable si $np(1-p) \geq 10$ ou

$$n \geq 20, \;\; np \geq 10, \;\; n(1-p) \geq 10$$

I. C. pour proportion: Cas des grands échantillons

- Construction de $I.C_{1-\alpha}$ pour p, Comme

$$\frac{\overline{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \approx \mathcal{N}(0,1)$$

$$\Rightarrow \quad P\left(-z_{1-\alpha/2} \leq \frac{\overline{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_{1-\alpha/2}\right) = 1-\alpha$$

$$\Leftrightarrow \quad P\left(\underbrace{\overline{p}-z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}}_{\ell_1} \leq p \leq \underbrace{\overline{p}+z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}}_{\ell_2}\right) = 1-\alpha$$

I. C. pour proportion: Cas des grands échantillons (suite)

- Valeurs approchées de ℓ_1 et ℓ_2 :

$$\begin{aligned} \ell_1 &\simeq \ell_1^* = \overline{p} - z_{1-\alpha/2}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} \\ \ell_2 &\simeq \ell_1^* = \overline{p} + z_{1-\alpha/2}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} \end{aligned}$$

- L'intervalle de confiance pour p de confiance $I.C_{1-\alpha}$ pour p

$$\Rightarrow IC_{1-\alpha}(p) = \left[\overline{p} - z_{1-\alpha/2}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}, \overline{p} + z_{1-\alpha/2}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\right]$$

Précision d'une estimation par intervalle de confiance

1. ***Incertitude absolue I_a***
   - C'est le rayon (demi longueur) de l'intervalle de confiance bilatérale (ℓ_1, ℓ_2) :
   $$I_a = \frac{\ell_2 - \ell_1}{2} = \text{ Rayon de I.C.}$$
   - **Exemple:** pour l'estimation de la moyenne μ de $X \sim \mathcal{N}(\mu, \sigma^2)$
   $$I_a = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Précision d'une estimation par intervalle de confiance (suite)

2. *Incertitude relative I_r*

- Elle est obtenue en divisant l'incertitude relative par l'estimation du paramètre

$$I_r = \frac{I_a}{\hat{\theta}} = \frac{\ell_2 - \ell_1}{\ell_2 + \ell_1}$$

- **Exemple:** pour l'estimation de la moyenne μ de $X \sim \mathcal{N}(\mu, \sigma^2)$

$$I_r = \frac{z_{1-\alpha/2}\,\sigma/\sqrt{n}}{\overline{X}}$$

Remarque

L'incertitude est proportionnelle à l'écart-type-population σ et à au degré de confiance $1-\alpha$ et inversement proportionnelle à la racine de la taille échantillon n

Taille d'un échantillon aléatoire

- **Problème:** Comment déterminer la taille n en fonction d'une incertitude maximale souhaitée?
- **Exemples:**
  - Estimation d'une moyenne: Si on veut que $I_a \leq \Delta\mu$ il faut que

$$z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \Delta\mu \Leftrightarrow n \geq (z_{1-\alpha/2}\sigma/\Delta\mu)^2$$

  - Estimation d'une proportion: Si on veut que $I_a \leq \Delta p$ il faut que

$$z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leq \Delta p$$

$$\Leftrightarrow \quad n \geq \frac{z^2_{1-\alpha/2}p(1-p)}{(\Delta p)^2}$$

Comme $p(1-p) \leq \frac{1}{4}$, il suffit de choisir

$$n \geq (z_{1-\alpha/2}/2\Delta p)^2$$

Exercice

Un comptable pense que les problèmes de liquidité d'une entreprise sont une conséquence directe de l'encaissement lent des comptes fournisseurs. Le comptable prétend qu'au moins 70% des comptes fournisseurs datent de plus de deux mois. Un échantillon de 120 comptes fournisseurs a révélé que 88 dataient de plus de deux mois

1. Quelle est l'estimation ponctuelle de la proportion p de comptes fournisseurs datant de plus de deux mois
2. Donnez un intervalle de confiance à 99% pour la vraie proportion p de comptes fournisseurs datant de plus de deux mois
3. Déterminez la taille minimum n de l'échantillon à prélevé pour que l'incertitude absolue de l'estimation (au niveau de confiance 99%) soit inférieur à 0.03.

Solution

1. L'estimation ponctuelle de la proportion p de comptes fournisseurs ... est $\hat{p}$ la proportion de comptes fournisseurs ... dans l'échantillon :

$$\bar{p} = 88/120 = 0.73.$$

2. Puisque $n = 120 \geq > 30$, on est dans le cas d'un grand échantillon :

$$\frac{\bar{p} - p}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}} \approx N(0,1)$$

$$\Rightarrow \quad P[-z_{1-\alpha/2} \leq \frac{\bar{p} - p}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}} \leq z_{1-\alpha/2}] = 1 - \alpha$$

$$\Rightarrow \quad P[\bar{p} - \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} 2.58 \leq p \leq \bar{p} + \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} 2.58] = 0.99$$

Solution (suite)

Donc

$$\begin{aligned} I_{0.99}(p) &= [\bar{p} - \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} 2.58; \bar{p} + \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} 2.58] \\ &= [0.73 - 2.58 * 0.04; 0.73 + 2.58 * 0.04] = [\mathbf{0.63}; \mathbf{0.84}] \end{aligned}$$

3. Soit I_a l'incertitude absolue (le rayon de l'intervalle de confiance).

$$\begin{aligned} I_a &\leq 0.03 \\ \sqrt{\frac{p(1-p)}{n}} z_{1-\alpha/2} &\leq 0.03 \\ n &\geq \frac{z^2_{1-\alpha/2}}{4 * (0.03)^2} \quad \text{car } p(1-p) \leq 1/4, \forall p \\ n &\geq 1849. \end{aligned}$$