Ch. 8 INTERVALLE DE CONFIANCE

Méthodes d'estimation

Estimation du paramètre $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_m]^\top$

1. **L'estimation ponctuelle :**
   Pour estimer la vraie valeur $\boldsymbol{\theta_0}$ du paramètre inconnu $\boldsymbol{\theta}$ on utilise une fonction (statistique) qui dépend de l'échantillon $(X_1, \ldots, X_n)$.

   (a) On utilise les estimateurs dits *naturels* :

      i. $\overline{X}$ pour estimer $\mu = E(X)$

      ii. S^2 pour estimer $V(X) = \sigma^2$

      iii. f_n pour estimer p

   (b) On construit l'E.M.V.

2. **L'estimation par Intervalle De Confiance :**

   (a) On associe au paramètre $\boldsymbol{\theta}$ un intervalle, dit de *confiance*, dans lequel se trouve la vraie valeur $\boldsymbol{\theta_0}$ du paramètre $\boldsymbol{\theta}$, avec une probabilité fixée d'avance, $1-\alpha$, α étant *le risque ou le niveau de confiance*.

   (b) Cet intervalle dépend **seulement de l'échantillon** et il sera calculé à l'aide d'une **fonction (statistique) dont on connaˆla loi**.

Intervalle De Confiance

Modélisation

Definition 1. *Soient :*

- *Un espace probabilisé* $(\Omega, \mathcal{A}, P)$
- *Une v.a.* X *dont la loi dépend d'un paramètre* Θ *avec* $\theta \in \Theta \subset \mathbb{R}$
- *La famille* $\mathcal{I}(\Theta) = \{I_x(\theta) \mid x \in \mathcal{B}_{\mathbb{R}}\}$ *d'intervalles de* Θ
- *On se fixe* $\alpha \in [0, 1]$ *;* α *est appelé niveau de confiance.*
- *On souhaite trouver deux statistiques*
  $T_1 = T_1(X_1, ..., X_n)$ *et* $T_2 = T_2(X_1, ..., X_n)$.
  t.q. $P_\theta(\theta \in I_\alpha) = 1 - \alpha \quad$ *avec* $I_\alpha = [T_1, T_2]$ *;*

  I_α *ainsi défini, est appelé I.D.C. pour* θ *au niveau ou au risque* α.

Remarques

– On peut avoir $P_\theta(\theta \in I_\alpha) \geq 1 - \alpha$
  i.e. un intervalle de confiance avec risque minimal α .
– Les bornes étant construites d'une façon aléatoire à partir d'une v.a. sont aussi des v.a.
– Il n'y a pas d'unicité des I.D.C. et il faut donner des critères associés aux problèmes posés.

– **Choix optimal** : C'est le choix de l'intervalle de longueur minimale.

Type de l'I.D.C.

La détermination des bornes de l'I.D.C. dépend du partage du risque α en α_1 et α_2 .

Deux cas se présentent :

1. **Intervalle Bilatéral** de la forme $[a, b]$
   avec $\alpha = \alpha_1 + \alpha_2$; $\alpha_1 \neq 0$, $\alpha_2 \neq 0$.

   (a) *Symétrique* $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$.

   (b) *Asymétrique* ou *Dissymétrique* $\alpha = \alpha_1 + \alpha_2$; $\quad \alpha_1 \neq \alpha_2$.

I.D.C. bilatéral

Symétrique *Dissymétrique*

2. **Intervalle unilatéral**

(a) *Bloqué à gauche :* $[a, +\infty[$: $\alpha_1 = \alpha$ et $\alpha_2 = 0$.

Intervalle utilisé dans le cas où $\theta > a$
(exemples : durée de vie, résistance à la rupture, etc.).

(b) *Bloqué à droite :* $]-\infty, b]$: $\alpha_1 = 0$ et $\alpha_2 = \alpha$

Intervalle utilisé dans le cas où $\theta < b$
(exemples : nombre de pièces défectueuses, temps d'attente, etc.).

Intervalle unilatéral

Bloqué à gauche :

Bloqué à droite :

Construction d'un I.D.C.

Soit $\theta \in \Theta$ le paramètre inconnu.

1. On considère une fonction monotone $u(x,\theta)$ qui dépend d'une statistique.
2. On définit α , $\alpha \in [0,1]$
   On pose :
   $P_\theta \left[u_1 < u(x,\theta) < u_2\right] = 1 - \alpha$;
3. On peut écrire :

   $$P_\theta \left[\beta_1(u_1) \leq \theta \leq \beta_2(u_2)\right] = 1 - \alpha \quad ;$$

4. L'**I.D.C.** pour θ est : $I_\alpha = [\beta_1(x); \beta_2(x)]$

On utilise les fonctions dites *pivotales*

Supposons que $(X_1, \ldots, X_n) \sim \mathcal{N}\left(\mu, \sigma^2\right)$.

1. **I.D.C pour** $\theta = \mu = E(X)$. On utilise la loi de $\overline{X}$

   (a) **Si la variance σ^2 est connue**.

   $$\sqrt{n}\,\frac{\overline{X} - \theta}{\sigma} \sim \mathcal{N}\,(0, 1)$$

   (b) **Si la variance σ^2 est inconnue.**
   On utilise l'estimateur de σ^2,
   $s^2 = \frac{1}{n-1} \sum_{j=1}^{n} \left(X_j - \overline{X}\right)^2$ ou $s_1^2 = \frac{1}{n} \sum_{j=1}^{n} \left(X_j - \overline{X}\right)^2$.

   $$\sqrt{n}\,\frac{\overline{X} - \theta}{s} \sim t(\nu), \quad \nu = n - 1 \text{ d.d.l.}$$

   (loi de Student ou t-distribution)

2. **I.D.C pour** $\sigma^2 = V(X)$, On pose $\theta = \sigma^2$.

(a) **Si la moyenne** $\mu = E(X)$ **est connue**.

On sait que $\frac{X-\mu}{\sigma} \sim N(0,1)$; $\left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi^2(\nu)$ $\nu = 1$ d.d.l.

et donc : $\frac{1}{\theta}\sum_{j=1}^{n}(X_j - \mu)^2 \sim \chi^2(\nu), \nu = n$ d.d.l.

(b) **Si la moyenne est inconnue** on utilise la loi de S^2

$\frac{1}{\theta}(n-1)\cdot s^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(\nu),\quad \nu = n-1$ d.d.l.

Avec $s^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2$

Pour le cas où $X \sim \mathcal{L}$, $\quad \mathcal{L}$ loi quelconque et un **grand** échantillon on utilise le **T.C.L.**

1. Supposons que $X \sim \mathcal{B}\left(n,p\right)$. On pose $\theta = p$.

$$t(X;\theta) = \lim_{n\to\infty} \frac{X - n\theta}{\sqrt{n\theta\left(1-\theta\right)}} \sim \mathcal{N}\left(0,1\right)$$

2. Si $X \sim \mathcal{P}\left(\lambda\right)$. On pose $\theta = \lambda$.

$$\lim_{n\to\infty} \sqrt{n}\frac{\overline{X} - \theta}{\sqrt{\theta}} \sim \mathcal{N}\left(0,1\right)$$

Intervalles de Confiance pour la Moyenne $\mu = E(X)$

Il y a deux cas :

1. **Loi normale $\mathcal{N}\left(\mu, \sigma^2\right)$ avec variance connue**
   (a) On construit la fonction $Z_n(\mu) = \frac{\overline{X}-\mu}{\sigma}\sqrt{n} \sim N(0,1)$
   (b) On cherche a, b tels que : $P_\mu\left[a \leq \frac{\overline{X}-\mu}{\sigma}\sqrt{n} \leq b\right] = 1-\alpha$
   (c) Pour un I.D.C. symétrique, on prend $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$
   (d) L'intervalle pour μ sera : $I_\alpha = \left[\overline{X} - \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}} \quad ; \quad \overline{X} + \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}\right]$

(e) La longueur de l'intervalle est : $2\, z_{\alpha/2}\, \frac{\sigma}{\sqrt{n}} = 2c$

(f) Á partir de cette formule on peut calculer la taille minimale de l'échantillon nécessaire pour obtenir une longueur $2c$

L'intervalle symétrique, pour un α fixé, est l'intervalle optimal, c.a.d. le plus précis.

2. **Loi normale $\mathcal{N}\left(\mu,\sigma^2\right)$ avec variance inconnue**

(a) On utilise l'estimateur $s^2 = \frac{1}{n-1}\sum\limits_{j=1}^{n}\left(X_j - \overline{X}\right)^2$.

(b) La v.a. $T_n\left(\mu\right) = \sqrt{n}\,\frac{\overline{X}-\mu}{s} \sim t(\nu)$, avec $\nu = n-1$ suit une *Loi de Student ou t-distribution.*

(c) On cherche a, b t.q : $P_\mu\left[a \leq \sqrt{n}\,\frac{\overline{X}-\mu}{s} \leq b\right] = 1-\alpha$

(d) Pour un intervalle symétrique on a :

$$I_\alpha = \left[\overline{X} - t_{n-1;\frac{\alpha}{2}}\frac{s}{\sqrt{n}} \quad ; \quad \overline{X} + t_{n-1;\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right]$$

Rq :1. La longueur de l'intervalle est minimale si l'intervalle est symétrique.

2. Si on veut que l'intervalle ait une longueur égale à $2c$, alors il faut prendre

$$n = \left[\frac{t_{n-1;\alpha/2} \cdot s}{c} \right]^2$$

Intervalle De Confiance pour les Variances

Il y a eux cas :

1. **Loi normale $\mathcal{N}\left(\mu, \sigma^2\right)$ avec moyenne connue**
   Intervalle pour $\theta = \sigma^2$

   (a) On pose $\theta = \sigma^2$

   (b) $U\left(\sigma^2\right) = \frac{n s_\mu^2}{\sigma^2} \sim \chi^2(\nu); \quad \nu = n \quad d.d.l.$
   avec $s_\mu^2 = \frac{1}{n} \sum\limits_{j=1}^{n} \left(X_j - \mu\right)^2$.

(c) On cherche a, b t.q :

$$P_{\sigma^2}\left[a \leq \frac{ns_\mu^2}{\sigma^2} \leq b\right] = 1 - \alpha$$

(d) $I_\alpha = \left[\frac{ns_\mu^2}{b},\ \frac{ns_\mu^2}{a}\right]$

(e) La longueur de l'intervalle est : $\left(\frac{1}{a} - \frac{1}{b}\right) ns_\mu^2$

(f) Dans le cas d'un intervalle symétrique :

$$a = \chi^2_{n;1-\alpha/2} \quad ; \quad b = \chi^2_{n;\alpha/2}$$

2. **Loi normale $\mathcal{N}\left(\mu, \sigma^2\right)$ avec moyenne inconnue**
   Intervalle pour $\theta = \sigma^2$

   (a) On utilise pour μ son estimateur $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

   (b) On utilise la fonction $S^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2$.

   (c) On a vu que : $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.

   (d) On cherche a, b t.q.

$$P_{\sigma^2}\left[a \leq \frac{ns^2}{\sigma^2} \leq b\right] = 1 - \alpha$$

(e) $I_\alpha = \left[\frac{ns^2}{b}, \;\; \frac{ns^2}{a} \right]$

(f) Longueur de l'intervalle est : $\left(\frac{1}{a} - \frac{1}{b}\right)(n-1)\; s_X^2$

(g) Pour un intervalle symétrique :

$$a = \chi^2_{n-1;1-\alpha/2} \quad ; \quad b = \chi^2_{n-1;\alpha/2}$$

I.D.C pour $\theta = p$

Pour n=40 essais on a calculé $f_n = \frac{24}{40} = 0.6$

Étapes

1. $u(x;\theta) = \frac{f_n - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$
2. Soit $\alpha = 0.05$
3. $P\left[t_1 \leq \frac{f_n - p}{\sqrt{\frac{p(1-p)}{n}}} \leq t_2\right] = 1 - \alpha$
4. $I_\alpha = \left[f_n - t_1\sqrt{\frac{p(1-p)}{n}} \quad ; \quad f_n + t_2\sqrt{\frac{p(1-p)}{n}}\right]$

Problème : $\theta = p$ se trouve dans les bornes ! !

5. **Calcul de l'I.D.C.**

   (a) Méth.-1 : **Par excès :**

      i. On maximise $f(p) = p(1-p)$

      ii. $f'(p) = 1 - 2p$ et $f'(p) = 0 \;\Rightarrow\; p = \frac{1}{2}$

      iii. $I_\alpha = \left[f_n - t_{\frac{\alpha}{2}} \sqrt{\frac{\frac{1}{2}(1-\frac{1}{2})}{n}} \quad ; \quad f_n + t_{\frac{\alpha}{2}} \sqrt{\frac{\frac{1}{2}(1-\frac{1}{2}}{n}} \right]$

      iv. **A.N :** $I_\alpha = [0.445\,;\,0.7549]$

(b) Méth.-2 : On utilise f_n qui est l'estimateur de p

i. On pose , $\widehat{p} = f_n$

ii. $I_\alpha = \left[f_n - t_1 \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \quad ; \quad f_n + t_2 \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \right]$

iii. **A.N** : $I_\alpha = [0.448\,; \, 0.752]$

(c) Méth.-3 :

On utilise les *Abaques* ou méthode des ellipses, pour un I.D.C. symétrique.

i. On a : $P\left[t_1 \le \frac{f_n - p}{\sqrt{\frac{p(1-p)}{n}}} \le t_2\right] = 1 - \alpha$

ii. La résolution en p se fait de façon direct :
$(\frac{f_n - p}{\sqrt{\frac{p(1-p)}{n}}})^2 \le t^2 \Leftrightarrow p^2(1 + \frac{t^2}{n}) - 2p(f_n + \frac{t^2}{2n}) + f_n^2 \le 0 \ (I)$

iii. On vérifie qu'il existe deux racines p_i, $i = 1, 2$ avec $0 \le p_i \le 1$
et l'inégalité (I) est vraie pour $p_1 \le p \le p_2$

iv. $I_\alpha = [p_1 \, , \, p_2]$

R.q :

- La méthode par excès donne l'intervalle le moins précis
- L'intervalle le plus précis est est donné par les *Abaques*.