INTRODUCTION

➢ EMV valeur la plus probable du paramètre θ mais aucune information sur la véracité de cette valeur

➢ Intervalle de confiance est un intervalle dans lequel on peut supposer que le paramètre θ se trouve avec une probabilité de se tromper fixée par l'utilisateur :

$$P[a \leq \theta \leq b] = 1 - \alpha$$

où α est la probabilité de se tromper

CONSTRUCTION DE L'INTERVALLE

• Soit $T=t(X_1,\ldots,X_n;\theta)$ un estimateur de θ dont on connaît la loi de probabilité

Tables $\Rightarrow$ a', b' tels que $P[a'\leq T\leq b']=1-\alpha$

• Supposons que t est monotone /r à θ, on en déduit

$\Rightarrow$ $a=t^{-1}(a')$, $b=t^{-1}(b')$ tels que $P[a\leq\theta\leq b]=1-\alpha$

Exemple : Supposons $X \sim \mathcal{N}(\mu,1)$. On cherche un intervalle de confiance pour μ à 95%.

$\overline{X}$ est un estimateur de μ et on sait que $\sqrt{n}(\overline{X}-\mu)\sim\mathcal{N}(0,1)$ donc

$$\begin{aligned} P[a\leq\mu\leq b] &= P[\sqrt{n}(\overline{x}-b)\leq\sqrt{n}(\overline{X}-\mu)\leq\sqrt{n}(\overline{x}-a)] \\ &= F\left(\sqrt{n}(\overline{x}-a)\right)-F\left(\sqrt{n}(\overline{x}-b)\right)=F(a')-F(b') \end{aligned}$$

Risque symétrique : a'=-b'

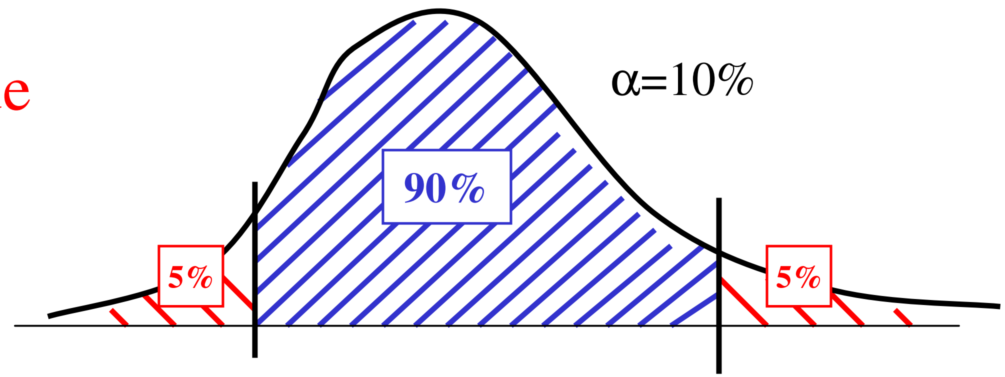
$$P[a\leq\mu\leq b]=0.95 \Leftrightarrow 2F(a')-1=0.95 \Leftrightarrow F(a')=0.975 \Leftrightarrow a'=1.96$$

$$\Rightarrow a=\overline{x}-\frac{a'}{\sqrt{n}}=\overline{x}-\frac{1.96}{\sqrt{n}} \quad \text{et} \quad b=\overline{x}+\frac{a'}{\sqrt{n}}=\overline{x}+\frac{1.96}{\sqrt{n}}$$

INTERVALLE BILATERAL OU UNILATERAL

Les bornes de l'intervalle de confiance dépendent du partage de α en $\alpha=\alpha_1+\alpha_2$

- Intervalle bilatéral : [a,b].
On utilise en général le risque symétrique, *i.e* $\alpha_1=\alpha_2=\alpha/2$



- Intervalle unilatéral
  - [a,+∞[(*i.e.* $\alpha_1=\alpha$, $\alpha_2=0$) utiliser dans le cas où le paramètre θ doit vérifier $\theta \geq a$ (durée de vie, résistance à la rupture, etc…)
  -]-∞,b[(*i.e.* $\alpha_1=0$, $\alpha_2=\alpha$) utiliser dans le cas où le paramètre θ doit vérifier $\theta \leq b$ (nombre de pièces défectueuses, temps d'attente, etc…)

CAS USUELS (1/2)

➢ Loi normale $\mathcal{N}(\mu,\sigma^2)$

Estimation de μ

- σ^2 connu, $\theta=\mu$ $\Rightarrow t(X_1,\ldots,X_n;\theta) = \boxed{\sqrt{n}\,\dfrac{\overline{X}-\theta}{\sigma} \sim \mathcal{N}(0,1)}$
- σ^2 inconnu, $\theta=\mu$ $\Rightarrow t(X_1,\ldots,X_n;\theta) = \boxed{\sqrt{n}\,\dfrac{\overline{X}-\theta}{S^*} \sim t(n-1)}$

où $S^{*2} = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$

Estimation de σ²

- μ connu, $\theta=\sigma^2$ $\Rightarrow t(X_1,\ldots,X_n;\theta) = \boxed{\dfrac{1}{\theta}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2_n}$
- μ inconnu, $\theta=\sigma^2$ $\Rightarrow t(X_1,\ldots,X_n;\theta) = \boxed{\dfrac{1}{\theta}\sum_{i=1}^{n}(X_i-\overline{X})^2 \sim \chi^2_{n-1} \Leftrightarrow \dfrac{(n-1)}{\theta}S^{*2} \sim \chi^2_{n-1}}$

CAS USUELS (2/2)

Dans le cas où n est grand, on peut utiliser les résultats de convergence de la loi binomiale ou de la loi de Poisson vers la loi normale pour connaître la loi de l'estimateur.

- Loi Binomiale B(n,p)

$$\theta = p \quad \Rightarrow \quad t(X_1,\dots,X_n;\theta) = \boxed{\frac{\overline{X} - n\theta}{\sqrt{n\theta(1-\theta)}} \sim \mathcal{N}(0,1)}$$

- Loi de Poisson P(λ)

$$\theta = \lambda \quad \Rightarrow \quad t(X_1,\dots,X_n;\theta) = \boxed{\sqrt{n}\,\frac{\overline{X} - \theta}{\sqrt{\theta}} \sim \mathcal{N}(0,1)}$$

Attention, ces lois sont des approximations, il faut donc s'assurer que n est suffisamment grand

TAILLE DE L'ECHANTILLON

Les lois précédentes permettent de calculer un I.D.C. [a,b] tel que $P[a \leq \theta \leq b]=1-\alpha$. Cet intervalle dépend

- L'estimateur de θ ($\overline{X}$, S^2,…)
- Les quantiles de la loi (obtenus dans les tables)
- La taille de l'échantillon n

Il est alors possible de déterminer la taille n telle que la longueur de l'intervalle $\ell=b-a$ vérifie une propriété donnée.

<u>Exemple</u> : Supposons $X \sim \mathcal{N}(\mu,1)$. Quelle doit être la taille de l'échantillon pour que la longueur de l'intervalle de confiance de μ à 95% soit inférieure à 0,5. D'après ce qui précède, nous avons

$$P[a \leq \mu \leq b] = P[\overline{x} - \frac{1{,}96}{\sqrt{n}} \leq \mu \leq \overline{x} + \frac{1{,}96}{\sqrt{n}}]$$

D'où

$$\ell = 2 \times \frac{1{,}96}{\sqrt{n}} \qquad \text{et alors} \qquad \ell < 0{,}5 \quad \Leftrightarrow \quad n > \left(\frac{2 \times 1{,}96}{0{,}5}\right)^2 \approx 61{,}47$$

Donc il faut avoir un échantillon d'au moins 62 individus.