l’estimation de paramètres

Bensalem Hana

La statistique inductive

- **Etendre** (inférer) les propriétés constatées sur un échantillon à toute la population.

- De valider ou infirmer des hypothèses sur la population énoncées a priori ou formulées après une phase exploratoire.
-
- nécessite des méthodes d'échantillonnage:

- **L'Estimation** : approcher des paramètres de la population à partir de l'échantillon.
- **Les Tests** : valider ou infirmer des hypothèses émises sur ces paramètres.
- **La Modélisation et prévision** : recherche d'une relation entre une variable et plusieurs autres, valable pour l'ensemble de la population

Modèle paramétrique

- un échantillon $X = (X_1, \ldots, X_n)$ ou les variables aléatoires X_i sont indépendantes et identiquement distribuées.
- leur loi commune est dans une famille de probabilités $P = \{P_\theta, \theta \in \Theta\}$
- ensemble fini de paramètres : $\Theta \subset \Re^d$, d est la dimension du modèle.
- θ un vecteur réel de paramètres
- Exemple:
- -“Pile ou face” (Bernoulli) $\theta = p$; $\Theta = [0, 1]$, $d = 1$
- -Modèle gaussien : $\theta = (\mu, \sigma 2)$, $\Theta = \Re \times \Re^{+*}$, $d = 2$

Estimation

Problème

- Obtenir des estimations fiables des caractéristiques d'une population à partir d'un échantillon représentatif

Estimation de?

- Espérance mathématique
- Variance ou écart type
- La proportion p(pour un caractère dénombrable)
- …

Type d'estimation

- Estimation ponctuelle(valeur unique)
- Estimation par intervalle de confiance(intervalle de valeurs pour plus de chance de tomber sur la vraie valeur du paramètre)

Définir une statistique

- Définir une fonction statistique
- **Qualité d'un estimateur**

Échantillon réalisé

- Relevé les mesures et calculer la statistique

estimateur

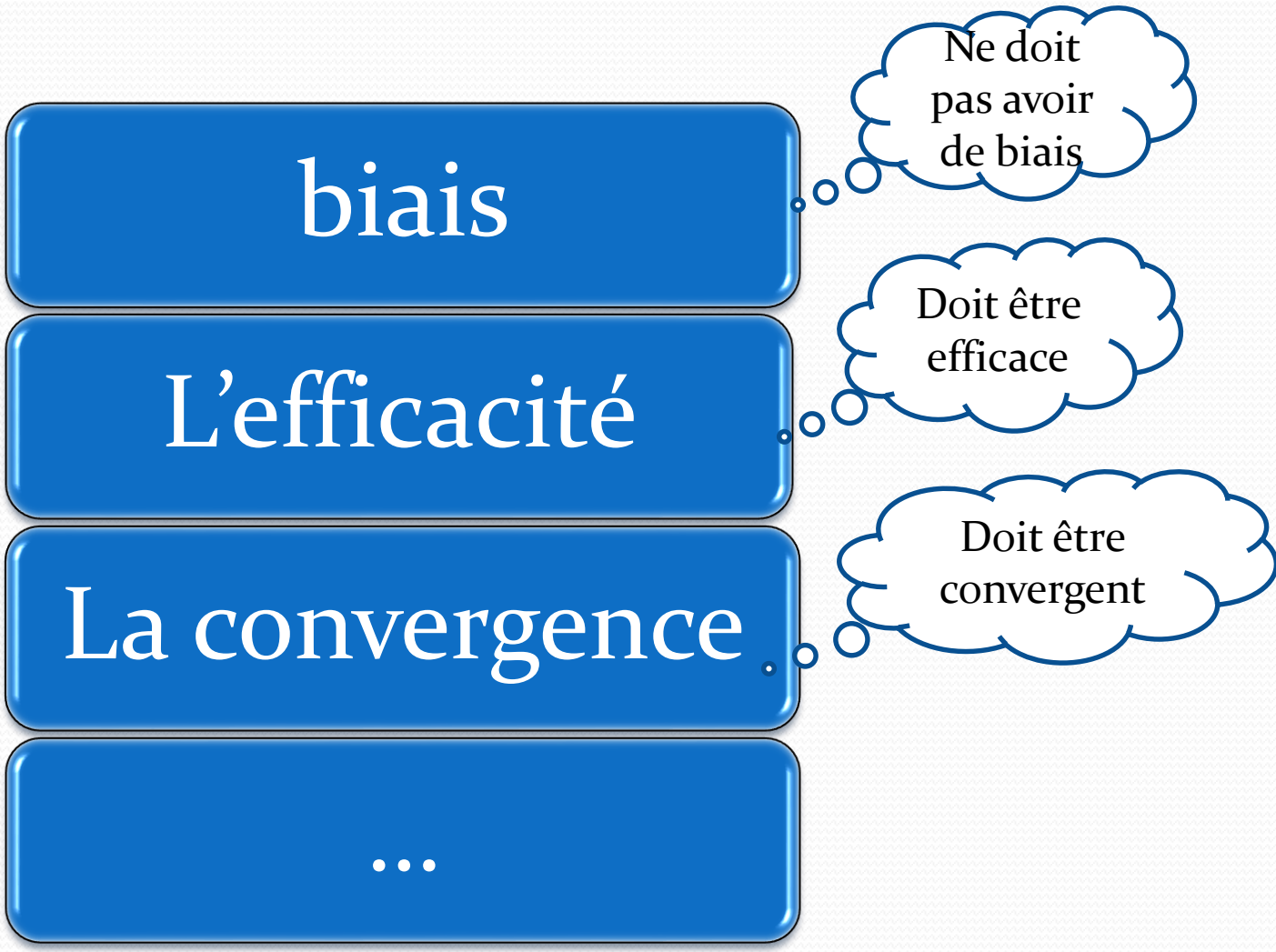
- L'estimateur prend une valeur ,noté $\hat{\theta}$ estimateur du paramètre θ

trouver une façon de construire des estimateurs de f(θ)?(définir la statistique ou sa formule)

Définir un estimateur

- la méthode des moments
- La méthode des moindres carrés ordinaire ou généralisées
- l'estimation bayésienne
- le maximum de vraisemblance

Qualité d'un estimateur



Biais d'un estimateur

Dit aussi sans biais

Non biaisé

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

En cas d'échantillonnage sans remise et avec le facteur d'exhaustivité cet estimateur restera toujours biaisé.

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

$E(\bar{x}) = E(\sum xi/n)$

$= 1/n(E(\sum xi)) = 1/n \sum E(xi)$

$= 1/n \sum m = 1/n*n*m = m$

$E(f) = p$; proportion pour une variable binomiale

$E(S^2) = (n-1)/n * \sigma 2 \Rightarrow$

S^2 est un estimateur biaisé de $\sigma 2$

(comportant des erreurs).

fréquence

$B(\hat{\theta}) = E(\hat{\theta}) - \theta$.....erreur ou biais de l'estimateur $\hat{\theta}$

Efficacité

- $Var(\hat{\theta}_1) < Var(\hat{\theta}_2) \Rightarrow \hat{\theta}_1$ est plus efficace que $\hat{\theta}_2$
- Si $\lim_{n \to +\infty} Var(\hat{\theta}) \to 0$ $\Rightarrow \hat{\theta}$ est efficace

Convergence

- Si $\lim_{n \to \infty} \hat{\theta} = \theta$ $\Rightarrow \hat{\theta}$ est un estimateur convergent de θ

paramètres	Estimateurs (tirage avec remise)
μ (L'espérance mathématique)	$\bar{x}$
p (la proportion d'une population)	f
σ^2 (la variance)	$S'^2 = n/(n-1) * S^2 = 1/(n-1) * \Sigma(x_i - \bar{x})$
σ(écart type)	$S' = S * \sqrt{\frac{n}{n-1}}$

Moyenne arithmétique

fréquence

Correction de la statistique S , pour qu' il soit sans biais

Aussi dit des estimateurs de maximum de vraisemblance (pour les loi B(p),N(μ, σ^2),P(λ)

Notes

tend vers zéro
lorsque n grand

- On peut avoir plusieurs estimateurs pour un même paramètre.
- L'estimateur Kn est dit **asymptotiquement sans biais** si $\lim_{n \to +\infty} E(Kn) = \theta$,
- Un estimateur **sans biais** et de **variance minimale** est appelé estimateur efficace.
- un estimateur peu biaisé, mais de variance très faible, pourrait même, en pratique, être préféré à un estimateur sans biais, mais de variance grande.

Moyenne arithmétique $\bar{x}$

Estimateurs d'une loi normale $X \sim N(\mu, \sigma)$

Paramètre estimé	estimateur	qualificatif
μ	$\hat{\mu}=1/n\sum_{i=1}^{n} Xi$/ $E(\hat{\mu})=\mu$, $var(\hat{\mu})=1/n*\sigma^2$	Estimateur efficace
σ^2/μ connue	$T_n^2=1/n\sum_{i=1}^{n}(Xi-\mu)^2$	Estimateur efficace
σ^2/μ inconnue ° °	$S_n^2=1/n\sum_{i=1}^{n}(Xi-\hat{\mu})^2$/ $E(S_n^2)=(n-1)/n*\sigma^2$	est un estimateur asymptotiquement sans biais de σ^2

Noté $\hat{\sigma}$ aussi variance empirique

Estimateurs d'une loi binomiale

X ~B(p) & Y ~P(λ)

- La moyenne empirique =k/n , est un estimateur efficace de p /k est le nombre de réalisation de p.
- 1/n*Σxi(moyenne empirique) est un estimateur de λ.

Précision d'un estimateur

Erreur quadratique moyenne

- $EQM(T) = E[(T - \theta)^2]$ ….(mesure la précision de l'estimateur)
- $EQM(T) = var(T) + [E(T) - \theta]^2 = var(T) + B(T)^2$
-
- En particulier, l'erreur quadratique moyenne des estimateurs sans biais est égale à leur variance.
- le meilleur estimateur est celui qui présente l'erreur quadratique moyenne la plus faible.

Intervalle de confiance

- Lorsque l'on cherche à estimer un paramètre, il est souvent plus utile de donner un renseignement du type a $\leq \theta \leq$ b, avec une estimation de la confiance que l'on peut avoir en cette affirmation, plutôt qu'une valeur précise. On dit alors qu'on fournit une estimation par intervalle de θ.

confiance

- Quelle confiance peut on accorder à une valeur d'un estimateur d'un paramètre, d'un échantillon de taille n?

Fixer α

- Une probabilité α (un risque) est fixer/ $\exists IC=[I_1^n,I_2^n]$ et $P(\theta \notin [I_1^n,I_2^n])=\alpha \Rightarrow P(\theta \in [I_1^n,I_2^n])=1-\alpha$

Appelé niveau de confiance

Trouver IC

- Pour α_1,α_2 tels que $\alpha_1+\alpha_2=\alpha$:
- Risque bilatéral classique: $\alpha_1=\alpha_2=\alpha/2$
- Risque unilatéral droit: $\alpha_1=\alpha,\alpha_2=0$
- Risque unilatéral gauche: $\alpha_1=0,\alpha_2=\alpha$

Écart type, voir la variance de cet estimateur page 12

Intervalle de confiance

- Rechercher un intervalle de confiance d'une ***moyenne*** μ d'une population, $X \sim N(\mu, \sigma^2)$
- **<u>Choix de la statistique:</u>** On utilise la statistique $\bar{x}$ qui est un bon estimateur de la ***moyenne*** μ.
- $\bar{x} \sim N(\mu , \sigma /\sqrt{n}) \Rightarrow \sqrt{n}\,(\bar{x} - \mu)/\sigma \sim N(0,1)$
  D'après la table de la loi N(0 ; 1) et α étant fixé, il est possible de trouver $U\alpha 1$, $U\alpha 2$ tel que

$$P(-u_{\alpha 2} < \sqrt{n}\,\frac{\bar{x}-\mu}{\sigma} < u_{\alpha 1}) = 1 - \alpha$$

- cette formule est la base de l'intervalle de confiance .

$\pm U_{1-\alpha/2} = \pm U_{\alpha/2}$

Pour la loi normale

Intervalle de confiance

Paramètre estimé	intervalle de confiance
μ avec X~ N(μ,σ^2) et σ connue	$I=[\hat{\mu}-u_{1-\alpha/2}\,\sigma/\sqrt{n}\ ,\ \hat{\mu}+u_{1-\alpha/2}\,\sigma/\sqrt{n}]$ $u_{1-\alpha/2}$ est le fractile d'ordre 1- α/2 de N(0,1)
μ avec X~N (μ,σ) et σ inconnue	$I=[\hat{\mu}-u_{1-\alpha/2}\,*S_n/\sqrt{n-1}\ ,\ \hat{\mu}+u_{1-\alpha/2}\,*S_n/\sqrt{n-1}]$ $I=[\hat{\mu}-u_{1-\alpha/2}\,\hat{\sigma}/\sqrt{n}\ ,\ \hat{\mu}+u_{1-\alpha/2}\,\hat{\sigma}/\sqrt{n}]$ $u_{1-\alpha/2}$ est le fractile d'ordre 1- α/2 de loi t(n-1)
Si n≥30 ,$\forall$ la loi de X=> $((\bar{X}_n-\mu)/(\sigma/\sqrt{n}))\sim N(0,1)$ $((\bar{X}_n-\mu)/\sqrt{(S_n^2/n)})\sim N(0,1)$	Même intervalle précédant. $u_{1-\alpha/2}$ des deux intervalles selon une loi normale centrée réduite
σ^2 avec X~N(μ, σ^2) et μ connue	Soit l'estimateur T_n^2, $n\,T_n^2/\sigma^2 \sim \chi^2(n)$ $I=[n\,T_n^2/u_{1-\alpha/2}\ ,\ n\,T_n^2/u_{\alpha/2}]$, $u_{1-\alpha/2}$ et $u_{\alpha/2}$ fractile de loi $\chi^2(n)$
σ^2 avec X~N(μ,σ^2) et μ inconnue	Soit l'estimateur S_n, $nS_n^2/\sigma^2 \sim \chi^2(n-1)$ $I=[n\,S_n^2/u_{1-\alpha/2}\ ,\ nS_n^2/u_{\alpha/2}]$, $u_{1-\alpha/2}$ et $u_{\alpha/2}$ fractile de loi $\chi^2(n-1)$
Proportion **X=Σxi~ B (n, p)**	$\hat{p}=\Sigma xi/n$, $I=[\hat{p}-u_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ ,\ \hat{p}+u_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]=[p_{min},p_{max}]$ Si n* p≥5 et n*(1-p)≥5(en remplaçant p par : p_{min} et p_{max}), $u_{1-\alpha/2}$ est le fractile d'ordre 1- α/2 de N(0,1)

Exercice-1-

- Une entreprise vend des bouchons de liège pour bouteilles de l'huile d'olive. Dans un souci de productivité, elle décide de traiter ses chênes-lièges avec des produits chimiques pour qu'ils développent leur écorce (aspect extérieur) plus vite. Ces traités chimiques peuvent altérer le liège et donner par la suite un goût bouchonné aux bouteilles.
- Dans la suite, on notera **p** la proportion de bouchons présentant un tel défaut.
- Un groupe d'experts goûte **215** de ces bouteilles et en compte **50** bouchonnées.
- 1. Proposer une estimation ponctuelle de **p**.
- 2. Construire un intervalle de confiance pour **p** au niveau **99**%.

solution

- Soit l'estimation d'une proportion p, ayant un caractère (les bouteilles de l'huile d'olive ayant un gout bouchonné)
- $\hat{p}$ = 50/215≈ 23.25%,
- Calcule de $I=[\hat{p}- t_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p}}\ /\sqrt{n}, \hat{p}+ t_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p}}\ /\sqrt{n}\ ]$
- avec α = 1% (le niveau de risque de l'intervalle)et n = 215
- $P(U<t_{1-\alpha/2})=1-\alpha/2=1-0.005=0.995$(de la table de la loi normale centrée réduite) $=>t_{1-\alpha/2}=2.5758$

$I= [0.2325- 2.5758\sqrt{0.2325(1-0.2325)}\ /\sqrt{215}, 0.2325+ 2.5758\sqrt{0.2325(1-0.2325}\ /\sqrt{215}\ ]$

$I=[0.2325-2.5758* 0.4224/ 14.6628, 0.2325+2.5758* 0.4224/ 14.6628]$

$I=[0.1583, 0.3067]=[15.83\%, 30.67\%]$

215*0.1583=34.0345 ,2015*(1-0.3067)=149.0595

N> 30, $N*P_{min}>5$, $N*(1-P_{max})>5$=> condition d'approximation de la loi binomiale à la loi normale respectée