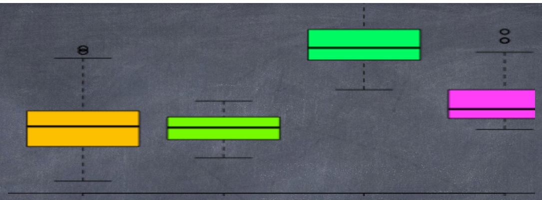
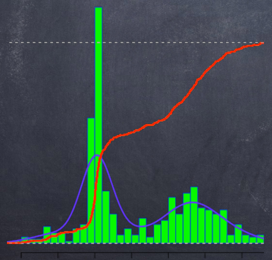


Statistique Inférentielle



Anne Philippe

Université de Nantes, LMJL 2021

Programme

1. Estimation ponctuelle dans un modèle paramétrique : Méthode des moments, Maximum de vraisemblance, delta-méthode, propriétés asymptotiques.
2. Région de confiance : Fonction pivotale Approche asymptotique
3. Efficacité : Borne de Cramer Rao, Théorème de Rao Blackwell.
4. Tests paramétriques : tests de Neymann Pearson, tests asymptotiques
5. Tests non paramétrique : test de Kolmogorov-Smirnov et test du Chi-Deux

Exemples et Définitions

Introduction
chapitre 1

Recensement / Sondage

- Soit P une population constituée de N individus.
- On suppose que P se décompose en K sous population $P = \cup_{i=1}^{K} A_i$ avec $A_i \cap A_j = \emptyset$ si $i \neq j$
- On note p_j la proportion de la population qui appartient à A_j
- On veut déterminer les valeurs de $(p_1, \ldots, p_K)$

Recensement :

- On interroge tous les individus de la population. On peut ainsi déterminer les valeurs des proportions $(p_1, \ldots, p_K)$
- Problème : non réalisable si N est grand

Sondage

- On interroge n individus ($n < N$)
- Pour n individus, on connait la sous population à laquelle ils appartiennent
- On peut alors calculer $\hat{p}_{1,n}, \ldots, \hat{p}_{K,n}$:
  $\hat{p}_{i,n}$ est la proportion des individus de l'échantillon qui appartiennent à A_i
- Quel est le lien entre $(\hat{p}_{1,n}, \ldots, \hat{p}_{K,n})$ et $(p_1, \ldots, p_K)$?

Problème de mesures

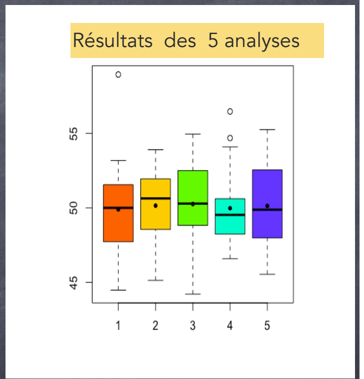
- Pour mesurer une quantité inconnue θ
  Exemple : une distance, un poids ou une température etc…
- On effectue n mesures : $x_1, \ldots, x_n$

Remarque : Si θ peut être mesuré sans erreur, alors $x_i = \theta \ \forall i$

- On peut calculer la mesure moyenne $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Quelle est la relation entre $\bar{x}$ et θ ?

Datation par luminescence



- Calcul de l'âge A d'un quartz : A = D/d
  - D est la dose de rayonnement absorbé par le quartz mesurée en laboratoire
  - d est le débit de dose (connu).
- On date les cristaux de quartz trouvés dans un sédiment
- On demande à 5 Laboratoires de mesurer la dose de rayonnement absorbé par n cristaux de quartz trouvés dans le sédiment
- Chaque laboratoire va fournir n mesures :
  $D_1^1, \ldots, D_n^1 \quad \ldots \quad D_1^5, \ldots, D_n^5$

Quel est le lien avec A ?

Problème de mesures : suite

- Les données $x_1, \ldots, x_n$ sont considérées comme le résultat d'une expérience aléatoire.
- On suppose qu'il existe $X = (X_1, \ldots, X_n)$ un vecteur aléatoire tel que $(x_1, \ldots, x_n)$ est une réalisation du vecteur aléatoire X.
- La moyenne empirique $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ est une variable aléatoire
- $\bar{x}$ est une réalisation de $\bar{X}_n$.
- Sous certaines hypothèses sur X et sa loi, on pourra établir un lien entre la variable aléatoire $\bar{X}_n$ et θ.

Définitions

1) $x_1, \ldots, x_n$ sont les données ou les observations

2)Le vecteur aléatoire $X = (X_1, \ldots, X_n)$ est appelé un échantillon et n est la taille de l'échantillon.

On dit aussi que X est un n-échantillon

3) Les variables aléatoires X_i sont à valeurs dans E muni de sa tribu $\mathscr{E}$ (tribu borélienne si E= $\mathbb{R}$)

4)Modèle stochastique : On considère une famille de lois $\mathscr{F}$ et on suppose que la loi de X appartient à $\mathscr{F}$: $P_X \in \mathscr{F}$

Modèle paramétrique

- On dit que le modèle est paramétrique si la famille de lois $\mathscr{F}$ est indexée par un paramètre $\theta \in \Theta$

$$\mathscr{F}_n = \{P_\theta^{(n)}, \theta \in \Theta\}$$

- Exemples :
  - famille des vecteurs gaussiens $\mathscr{F}_n = \{\mathcal{N}_n(\mu, \Sigma), \theta = (\mu, \Sigma)\}$
  - $P_\theta^{(n)}$ est la loi de n variables aléatoires indépendantes et identiquement distribuées (iid) suivant P_θ avec $\theta \in \Theta$

  $\mathscr{F} = \{P_\theta, \theta \in \Theta\}$

    - $\mathscr{F} = \{\mathcal{N}(\mu, \sigma^2), \theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+\}$ modèle gaussien
    - $\mathscr{F} = \{\mathscr{P}(\theta), \theta \in \mathbb{R}^+\}$ modèle de Poisson

Problème de mesures : Modèle 1

- Toutes les mesures sont réalisées de façon indépendante dans les mêmes conditions

$\Rightarrow X = (X_1, \ldots, X_n)$ sont des variables aléatoires iid.

Construction de la famille de lois $\mathscr{F}$:

- Pour tout i=1,...,n $X_i = \theta + \epsilon_i$
  - ϵ_i représente l'erreur de mesure $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$
- Modèle : $X = (X_1, \ldots, X_n)$ iid suivant la loi gaussienne $\mathcal{N}(\theta, \sigma^2)$
- Lien entre $\bar{X}_n$ et θ : $\bar{X}_n \xrightarrow[n\to\infty]{ps} \theta$ et $\mathbb{E}(\bar{X}_n) = \theta$

Modèle 2

- Si toutes les mesures sont effectuées avec la même appareil alors l'hypothèse d'indépendance des $X_1, \ldots, X_n$ est fausse
- On doit ajouter une erreur commune liée à l'appareil

  $X_i = \theta + \epsilon_i + \epsilon_0$ pour tout i=1,..,n

- Si $(\epsilon_1, \ldots, \epsilon_n) \perp \epsilon_0$ et $\epsilon_0 \sim \mathcal{N}(0, \tau^2)$ alors

$$X \sim \mathcal{N}_n((\theta, \ldots, \theta)^T, \Sigma) \text{ avec } \Sigma_{i,j} = \begin{cases} \tau^2 + \sigma^2 & \text{si } i = j \\ \tau^2 & \text{si } i \neq j \end{cases}$$

- Lien entre $\bar{X}_n$ et θ : $\mathbb{E}(\bar{X}_n) = \theta$ attention $\bar{X}_n \xrightarrow[n\to\infty]{ps} \theta + \epsilon_0$

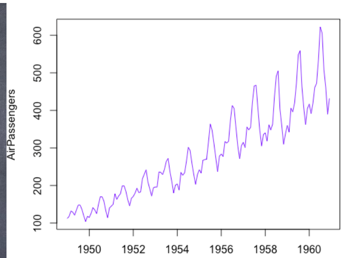
Durée de vie

- Soit $x_1, \ldots, x_n$ la durée de vie de certains composants électroniques.
- Hypothèses :
  1. Les composants fonctionnent de façon indépendante,
  2. Tous les composants sont du même type.
- Modèle : $X_1, \ldots, X_n$ iid suivant la loi qui admet pour fonction de répartition F.
- La quantité d'intérêt est 1 - F (t) avec un t > 0.
  C'est la probabilité que la durée de vie du composant dépasse t.
- C'est un problème d'estimation fonctionnelle.

Essai clinique

- On veut tester l'effet d'un médicament
- On constitue 2 groupes
  $$\begin{cases} G_1 : & n_1 \text{ individus recoivent le traitement} \\ G_2 : & n_2 \text{ individus recoivent un placebo} \end{cases}$$
- On observe x_i le nombre d'individus du groupe G_i qui ressentent un effet positif.
- x_i est une réalisation de X_i
- Si les individus réagissent de façon indépendante : $X_i \sim \mathcal{B}(n_i, p_i)$
- Tester l'effet d'un médicament $\Rightarrow$ Tester statistiquement si $p_1 \geq p_2$

Prévision



- On observe le nombre de passagers par mois dans un aéroport
- Données : $x_1, \ldots, x_n$
- On veut prévoir x_{n+1}
- Modèle on considère $X_1, \ldots, X_n, X_{n+1}$ n+1 variables aléatoires
- $x_1, \ldots, x_n$ est une réalisation de $X_1, \ldots, X_n$,
- X_{n+1} n'est pas observée, on veut prévoir x_{n+1}.
- Si les variables $X_1, \ldots, X_n, X_{n+1}$ appartiennent à L^2 alors $\mathbb{E}(X_{n+1} | X_1, \ldots, X_n)$ est la meilleure approximation de X_{n+1} au sens L^2
- Problème statistique : estimation de l'espérance conditionnelle.

Estimateur ponctuel

- Soit $X = (X_1, \ldots, X_n)$ un n-échantillon à valeurs dans $(E^n, \mathcal{E}_n)$
- On suppose que la loi de X appartient à $\mathcal{F}_n = \{P_\theta^{(n)}, \theta \in \Theta\}$
  c'est à dire il existe $\theta_0 \in \Theta$ tel que $P_X = P_{\theta_0}^{(n)}$
- Définition : un estimateur est une variable aléatoire de la forme $h(X_1, \ldots, X_n)$ avec
  $h : E^n \mapsto \Theta$
  h est connue elle ne dépend pas du paramètre θ
- Notation : $h(X_1, \ldots, X_n) := \hat{\theta}_n$ ou $\hat{\theta}(X_1, \ldots, X_n)$

Espace probabilisé

- $(\Omega, \mathscr{A}, \mathbb{P})$ espace probabilisé, la loi de X est définie par $P_X(A) = \mathbb{P}(X^{-1}(A))$ et $\mathbb{E}(g(X)) = \int g(X) d\mathbb{P}$ si $g(X) \in L^1$
- Dans un modèle paramétrique on introduit une famille d'espaces probabilisés $(\Omega, \mathscr{A}, \mathbb{P}_\theta)_{\theta \in \Theta}$
- À θ fixé, la loi de X est $P_\theta^{(n)}(A) = \mathbb{P}_\theta(X^{-1}(A)) = \mathbb{P}_\theta(X \in A)$ et $\int g(X) d\mathbb{P}_\theta = \mathbb{E}_\theta(g(X))$

Cas continu

- Si la loi de X est continue, on définit le modèle par une famille de densités $\mathscr{F}_n = \{f_\theta^{(n)}, \theta \in \Theta\}$
- On a $\begin{cases} \mathbb{P}_\theta(X \in A) = \int_A f_\theta^{(n)}(x)\, dx \\ \mathbb{E}_\theta(g(X)) = \int_{E^n} g(x) f_\theta^{(n)}(x)\, dx \end{cases}$

Cas discret

- Si la loi de X est discrète, on définit le modèle par une famille $\mathscr{F}_n = \{p_\theta^{(n)}, \theta \in \Theta\}$
- On a

$$\begin{cases} \mathbb{P}_\theta(X = x) = \mathbb{P}_\theta(X_1 = x_1, \ldots, X_n = x_n) = p_\theta^{(n)}(x_1, \ldots, x_n) = p_\theta^{(n)}(x) \\ \mathbb{P}_\theta(X \in A) = \sum_{x \in A} p_\theta^{(n)}(x) \\ \mathbb{E}_\theta(g(X)) = \sum_{x \in E^n} g(x)\, p_\theta^{(n)}(x) \end{cases}$$

Les qualités d'un estimateur : le biais

- $\mathscr{F}_n = \{P_\theta^{(n)}, \theta \in \Theta\}$ le modèle paramétrique
- Soit $X = (X_1, \ldots, X_n)$ un n-échantillon et $\hat{\theta}_n$ un estimateur de θ
- le biais d'un estimateur est défini par $b_n(\theta) = \mathbb{E}_\theta(\hat{\theta}_n) - \theta$
- On dit que $\hat{\theta}_n$ est un estimateur sans biais de θ si pour tout $\theta \in \Theta$ on a $\mathbb{E}_\theta(\hat{\theta}_n) = \theta$
- On dit que $\hat{\theta}_n$ est un estimateur asymptotiquement sans biais de θ si pour tout $\theta \in \Theta$ on a $\mathbb{E}_\theta(\hat{\theta}_n) \xrightarrow{n \to \infty} \theta$

Les qualités d'un estimateur : la consistance

- On dit que $\hat{\theta}_n$ est un estimateur fortement consistant de θ si pour tout $\theta \in \Theta$ on a $\hat{\theta}_n \xrightarrow[P.S.]{n\to\infty} \theta$

- On dit que $\hat{\theta}_n$ est un estimateur (faiblement) consistant de θ si pour tout $\theta \in \Theta$ on a $\hat{\theta}_n \xrightarrow[proba]{n\to\infty} \theta$

- Soit $\hat{\theta}_n$ un estimateur de θ tel que pour tout $\theta \in \Theta$ on a $\hat{\theta}_n \in L^2_\theta$.
On dit que $\hat{\theta}_n$ est un estimateur consistant au sens L^2 de θ si pour tout $\theta \in \Theta$ on a $\mathbb{E}_\theta\left((\hat{\theta}_n - \theta)^2\right) \xrightarrow{n\to\infty} 0.$

Erreur quadratique

- Soit $\hat{\theta}_n$ un estimateur de θ tel que pour tout $\theta \in \Theta$ on a $\hat{\theta}_n \in L^2_\theta$.

- L'erreur quadratique est définie par $EQ_n(\theta) = \mathbb{E}_\theta\left((\hat{\theta}_n - \theta)^2\right)$

- Pour tout $\theta \in \Theta$ on a $EQ_n(\theta) = b_n(\theta)^2 + Var_\theta(\hat{\theta}_n)$ (var+biais²)

Proposition : Si $\hat{\theta}_n$ est un estimateur asymptotiquement sans biais de θ et si $Var_\theta(\hat{\theta}_n) \xrightarrow[n\to\infty]{} 0$ pour tout $\theta \in \Theta$ alors $\hat{\theta}_n$ est un estimateur consistant au sens L^2de θ

Vitesse de Convergence

- On suppose que $\hat{\theta}_n$ un estimateur consistant de θ

- Soit $(\alpha_n)_{n\in N}$ une suite de réels positifs telle que $\alpha_n \xrightarrow{n\to\infty} +\infty$

  On dit que $\hat{\theta}_n$ est un estimateur α_n−consistant de θ si
  $\forall \theta \in \Theta \quad \forall \epsilon > 0 \quad \exists M > 0 \ : \ \sup_{n\in N} P_\theta\left(\alpha_n|\hat{\theta}_n - \theta| > M\right) \leq \epsilon$

  Proposition : Si $\forall \theta \in \Theta \quad \alpha_n(\hat{\theta}_n - \theta) \xrightarrow[loi]{n\to\infty} Z$ et Z est non dégénérée alors $\hat{\theta}_n$ est un estimateur α_n−consistant

Modèle exponentiel

- Soit $X = (X_1, \ldots, X_n)$ un n-échantillon iid suivant la loi exponentielle

- $\mathcal{F} = \{\mathcal{E}xp(\theta)\ , \theta \in \Theta\}$: la densité est $f_\theta(x) = \theta e^{-x\theta}\, \mathbb{I}_{R^+}(x)$

  et $f^{(n)}_\theta(x_1, \ldots, x_n) = \theta^n e^{-\theta \sum_{i=1}^n x_i}\, \mathbb{I}_{R^n_+}(x_1, \ldots, x_n)$

- $\hat{\theta}_n = \dfrac{n}{\sum_{i=1}^n X_i} = \dfrac{1}{\bar{X}_n}$ est un estimateur de θ qui est asymptotiquement sans biais, fortement consistant et $\sqrt{n}$−consistant. De plus il est consistant au sens L^2.

Reparamétrisation

- Soit $X = (X_1, \ldots, X_n)$ un n-échantillon suivant $P_X \in \{P_\theta^{(n)}, \theta \in \Theta\}$
- Soit $h : \Theta \to \Lambda$. On veut estimer le paramètre $\lambda = h(\theta) \in \Lambda$
- Soit $\hat{\theta}_n$ est un estimateur de θ.
- On pose $\hat{\lambda}_n = h(\hat{\theta}_n)$. C'est un estimateur de λ
- Quelles sont les propriétés de $\hat{\lambda}_n$ héritées de $\hat{\theta}_n$?

Remarque : si h est une bijection, on peut définir le modèle $\{\tilde{P}_\lambda^{(n)}, \lambda \in \Lambda\}$ avec $\tilde{P}_\lambda^{(n)} = P_{h^{-1}(\lambda)}^{(n)}$

Transfert des propriétés

- Si $\hat{\theta}_n$ est fortement consistant et h est continue sur Θ alors $\hat{\lambda}_n$ est fortement consistant
- Si $\hat{\theta}_n$ est consistant et h est continue sur Θ alors $\hat{\lambda}_n$ est consistant
- Si $\hat{\theta}_n$ est α_n-consistant et h est C^1 alors $\hat{\lambda}_n$ est α_n-consistant

Rappel sur les convergences

Soit $(X_n)_n, X, (Y_n)_n$ des vecteurs aléatoires et y une constante.
Soit h une application, on note $C(h)$ les points de continuité de la fonction h

(1) Si $X_n \overset{ps}{\to} X$ et h continue alors $h(X_n) \overset{ps}{\to} h(X)$

(2) Si $X_n \xrightarrow{proba} X$ et h continue alors $h(X_n) \xrightarrow{proba} h(X)$

(3) Si $X_n \overset{loi}{\to} X$ et $P(X \in C(h)) = 1$ alors $h(X_n) \overset{loi}{\to} h(X)$

- Cas particulier : Si $Y_n \xrightarrow{proba} y$ et $y \in C(h)$ alors $h(Y_n) \xrightarrow{proba} h(y)$

(4) Théorème de Slutky

Si $X_n \overset{loi}{\to} X$ et $Y_n \xrightarrow[loi]{proba} y$ alors $(X_n, Y_n) \overset{loi}{\to} (X, y)$

Rappel sur les convergences

(5) Corollaire du Théorème de Slutsky

Si $\begin{cases} X_n \overset{loi}{\to} X \\ Y_n \xrightarrow{proba} y \end{cases}$ alors $\begin{cases} X_n + Y_n \overset{loi}{\to} X + y \\ < X_n, Y_n > \overset{loi}{\to} < X, y > \end{cases}$ et en $\dim(E) = 1$ $\begin{cases} X_n / Y_n \overset{loi}{\to} X/y \text{ si } y \neq 0 \\ X_n Y_n \overset{loi}{\to} Xy \end{cases}$

(6) Δ-méthode (ou théorème de Cramer)

Soit $(X_n)_n$ des vecteurs aléatoires $(X_1 \in R^p)$ et $x \in R^p$ une constante.
$(\alpha_n)_n$ une suite déterministe tel que $\alpha_n \to \infty$

Si $\alpha_n(X_n - x) \overset{loi}{\to} X$ et $h : R^p \to R^k$ est C^1 au voisinage de x
alors $\alpha_n(h(X_n) - h(x)) \overset{loi}{\to} Dh(x)\, X$

$$Dh(x) = \begin{pmatrix} \nabla h_1(x) \\ \vdots \\ \nabla h_k(x) \end{pmatrix}$$

$$\nabla = \left(\frac{\partial}{\partial x_1}, \cdots, \frac{\partial}{\partial x_p}\right)$$

Vraisemblance

Chapitre 2

1- Définition

Contexte

(A) $X = (X_1, \ldots, X_n)$ est un n-échantillon de variables aléatoires de loi P_X

(B) On suppose que la loi P_X appartient à une famille paramétrique $\mathscr{F} = \{P_\theta^{(n)} : \theta \in \Theta\}$

1. La loi de X est continue on note $f_\theta^{(n)}$ la densité de la loi $P_\theta^{(n)}$
2. La loi de X est discrète on note $f_\theta^{(n)}(x_1, \ldots, x_n) = P_\theta^{(n)}(X_1 = x_1, \ldots, X_n = x_n)$

On peut ainsi unifier la présentation en décrivant $\mathscr{F}$ par $\{f_\theta^{(n)} : \theta \in \Theta\}$

Exemples

A. Modèle continu :

- Exemple du modèle exponentiel iid

$$f_\theta^{(n)}(x_1, \ldots, x_n) = \theta^n e^{-\theta \sum_{i=1}^n x_i} 1_{\mathbb{R}_+^n}(x_1, \ldots, x_n)$$

B. Modèle discret :

- Exemple du modèle de Poisson iid

$$f_\theta^{(n)}(x_1, \ldots, x_n) = \frac{e^{-n\theta} \theta^{\sum_{i=1}^n x_i}}{x_1! \ldots x_n!}$$

Vraisemblance

Définition : La vraisemblance V d'un n-échantillon X pour le modèle $\mathscr{F}$ est une fonction définie sur Θ à valeurs aléatoires positives.

Elle est définie par

$$V(\theta) = f_\theta^{(n)}(X_1, \ldots, X_n) \qquad \forall \theta \in \Theta$$

- Si V est strictement positive on définit aussi la log-vraisemblance : $L(\theta) = \log(\ V(\theta)\ )$ pour tout $\theta \in \Theta$.

Calcul de vraisemblance

- Modèle exponentiel iid : pour tout $\theta > 0$

$$V(\theta) = \theta^n e^{-\theta \sum_{i=1}^n X_i}$$

$$L(\theta) = n\log(\theta) - \theta \sum_{i=1}^n X_i$$

- Modèle de Poisson iid : pour tout $\theta > 0$

$$V(\theta) = \frac{e^{-n\theta}\theta^{\sum_{i=1}^n X_i}}{X_1! \ldots X_n!}$$

$$L(\theta) = -n\theta + \log(\theta)\sum_{i=1}^n X_i - \sum_{i=1}^n \log(X_i!)$$

Théorème de séparabilité

Hypothèses de régularité

- HR-0 [identifiabilité] Si $f_\theta^{(n)} = f_\eta^{(n)}$ alors $\theta = \eta$
- HR-1 Toutes les lois $f_\theta^{(n)}, \theta \in \Theta$ ont le même support
- HR-2 $(X_1, \ldots X_n) \sim f_{\theta_0}^{(n)}$ avec $\theta_0 \in \Theta$ et Θ ouvert. [θ_0 est la vraie valeur du paramètre]

Théorème : On suppose que HR-0-1-2 sont vérifiées.
Si $(X_1, \ldots X_n)$ sont iid et les variables aléatoires $\dfrac{f_\theta(X_i)}{f_{\theta_0}(X_i)}$ et $\log\left(\dfrac{f_\theta(X_i)}{f_{\theta_0}(X_i)}\right)$ sont $L^1_{\theta_0}$

alors $\lim P_{\theta_0}(L(\theta_0) > L(\theta)) = 1$ pour tout $\theta \neq \theta_0$.

Autrement dit, asymptotiquement la vraisemblance atteint sont maximum au point θ_0

Estimateur du maximum de vraisemblance (EMV)

Définition :
Soit X un n-échantillon et V sa vraisemblance.

Si la vraisemblance V admet un maximum global atteint en un unique point c'est l'estimateur du maximum de vraisemblance. On le note $\hat{\theta}_n^{MV}$

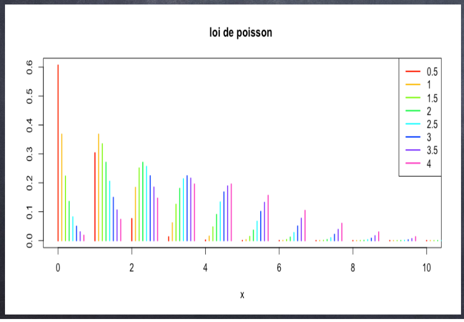
Si $\hat{\theta}_n^{MV}$ existe alors pour tout $\theta \in \Theta$ on a $V(\hat{\theta}_n^{MV}) \geq V(\theta)$

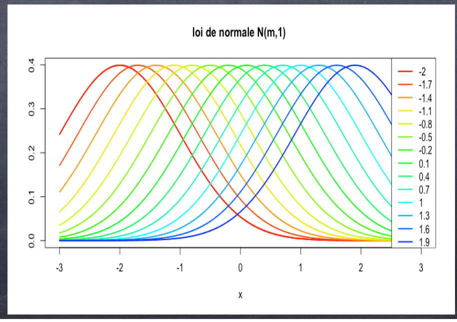
autrement dit $\hat{\theta}_n^{MV} = \text{argmax}_{\theta \in \Theta}\ V(\theta)$.

Remarque: si L est bien définie on a aussi $\hat{\theta}_n^{MV} = \text{argmax}_{\theta \in \Theta}\ L(\theta)$ par monotonie de la fonction log

Interprétation

Remarque : Si la loi des X est discrète alors $\hat{\theta}_n^{MV}(x_1, \ldots, x_n)$ est la valeur de θ pour laquelle l'observation $x_1, \ldots x_n$ est la plus probable.





Estimateur du MV

- Modèle exponentiel iid

$$V(\theta) = \theta^n e^{-\theta \sum_{i=1}^n X_i} \quad \text{et} \quad \theta_n^{MV} = \frac{1}{\bar{X}_n}$$

- Modèle de Poisson iid

$$V(\theta) = \frac{e^{-n\theta}\theta^{\sum_{i=1}^n X_i}}{X_1! \ldots X_n!} \quad \text{et} \quad \theta_n^{MV} = \bar{X}_n$$

- Modèle Gamma iid : on n'a pas de forme explicite de l'estimateur.
- Modèle de Gaussien iid $N(\theta, 1)$ avec $\theta \geq 0$

$$\theta_n^{MV} = max(0, \bar{X}_n)$$

Reparamétrisation

- Soit g est une application de $\Theta \mapsto \Lambda$ bijective. On pose $\lambda = g(\theta)$.
- λ est le paramètre d'intérêt.
- Modèle reparamétrisé s'écrit $\{g_\lambda^{(n)} : \lambda \in \Lambda\}$ avec $g_\lambda^{(n)} = f_{g^{-1}(\lambda)}^{(n)}$

Proposition : Si $\hat{\theta}_n^{mv}$ est l'estimateur du Maximum de vraisemblance de θ alors $g(\hat{\theta}_n^{mv})$ est l'estimateur du maximum de vraisemblance de $\lambda = g(\theta)$

Consistance

Théorème (Admis): On suppose que les hypothèse HR-0-1-2 sont vérifiées. Si $(X_1, \ldots X_n)$ sont iid et si la vraisemblance est dérivable par rapport à θ alors l'équation de vraisemblance définie par

$$\nabla V(\theta) = 0 \text{ (ou } \nabla L(\theta) = 0) \quad \text{avec} \quad \nabla = \left(\frac{\partial}{\partial \theta_1} \ldots \frac{\partial}{\partial \theta_p}\right)^T$$

admet une solution $\hat{\theta}_n$ telle que $\hat{\theta}_n \xrightarrow[proba]{} \theta_0$.

Corollaire Sous les hypothèses du Th précédent. Si l'équation vraisemblance admet une unique solution alors $\hat{\theta}_n^{MV} \xrightarrow[proba]{} \theta_0$

2-Information de Fisher

Conditions de Fisher

- On complète les conditions de régularité HR
  - HR-3 $\theta \mapsto f_\theta^{(n)}(x)$ est une fonction C^2 sur Θ pour tout x
  - HR-4 $\theta \mapsto \int f_\theta^{(n)}(x)\,dx$ est deux fois dérivable sous le signe $\int$ et les dérivées secondes sont continues.
    On peut échanger l'intégration et la différenciation par rapport à θ

Définition on dit que le modèle est régulier s'il vérifie les hypothèses HR-0 à HR-4 (aussi appelées conditions de Fisher).

Remarque
si $X_1, \ldots X_n$ sont iid alors il suffit de vérifier les hypothèses HR pour n=1

Propriétés d'un modèle régulier
Cas particulier $dim(\Theta) = 1$

- Si le modèle est régulier, la variable aléatoire $\frac{\partial}{\partial\theta}\log(f_\theta^{(n)}(X))$ vérifie les propriétés suivantes
  - $$E_\theta\left(\frac{\partial}{\partial\theta}\log(f_\theta^{(n)}(X))\right) = 0$$
  - $$Var_\theta\left(\frac{\partial}{\partial\theta}\log(f_\theta^{(n)}(X))\right) = E_\theta\left(\left(\frac{\partial}{\partial\theta}\log(f_\theta^{(n)}(X))\right)^2\right)$$
    $$= -E_\theta\left(\frac{\partial^2}{\partial\theta^2}\log(f_\theta^{(n)}(X))\right)$$

Information de Fisher

- On définit l'information de Fisher pour des modèles réguliers
- En dimension 1 l'information de Fisher apportée par un n-échantillon est définie par

$$I_n(\theta) = E_\theta\left(\left(\frac{\partial}{\partial\theta}\log(f_\theta^{(n)}(X))\right)^2\right)$$

On peut aussi exprimer l'information de Fisher de la forme :

$$I_n(\theta) = -E_\theta\left(\frac{\partial^2}{\partial\theta^2}\log(f_\theta^{(n)}(X))\right) = Var_\theta\left(\frac{\partial}{\partial\theta}\log(f_\theta^{(n)}(X))\right)$$

Proposition Si $X_1, .., X_n$ est un n-échantillon de variables aléatoires iid et si le modèle est régulier alors

$$I_n(\theta) = nI_1(\theta)$$

Toutes les observations apportent la même information.

Exemple : $X_1, .., X_n$ iid suivant la loi exponentielle

On a a $I_n(\theta) = \frac{n}{\theta^2}$

Cas multivarié : $dim(\Theta) = d > 1$

- Pour un modèle régulier multivarié L'information de Fisher est une matrice

$$I_n(i,j) = E_\theta\left(\left(\frac{\partial}{\partial\theta_i}\log(f_\theta^{(n)}(X))\right)\left(\frac{\partial}{\partial\theta_j}\log(f_\theta^{(n)}(X))\right)\right)$$

On peut aussi l'exprimer sous la forme

$$I_n(i,j) = -E_\theta\left(\frac{\partial^2}{\partial\theta_i\theta_j}\log(f_\theta^{(n)}(X))\right)$$

propriétés

- La matrice d'information de Fisher I_n peut s'exprimer comme la variance du vecteur aléatoire $\nabla log(f_\theta^{(n)}(X))$
- Si $X_1, .., X_n$ est un n-échantillon de variables aléatoires iid et si le modèle est régulier alors la matrice d'information de Fisher vérifie

$$I_n(\theta) = nI_1(\theta)$$

Famille exponentielle

- Définition : Une famille de loi $\{f_\theta^{(n)}, \theta \in \theta\}$ est une famille exponentielle s'il existe quatre fonctions η, q, K, H telles que pour tout $\theta \in \Theta \subset R^d$ et $x \in E^n$

$$f_\theta^{(n)}(x) = \begin{cases} e^{\eta(\theta)\bullet K(x) + H(x) + q(\theta)} & \text{pour tout } x \in S \\ 0 & \text{sinon} \end{cases}$$

Le support S de la loi ne dépend pas de θ

Notation : $x \bullet y$ est le produit scalaire

- Propriété si $X_1, ... X_n$ sont iid et et si la loi commune appartient à une famille exponentielle alors la loi du n-échantillon appartient à une famille exponentielle

Exemples

- La famille des lois gaussiennes $\{N(0,\sigma^2)\ :\ \sigma^2 > 0\}$ est une famille exponentielle
- La famille des lois uniformes $\{Unif(0,\theta)\ :\ \theta > 0\}$ n'est pas une famille exponentielle

Forme canonique

- On peut reparamétriser la densité de la famille exponentielle en fonction de $\lambda = \eta(\theta)$
- On définit une nouvelle famille de lois :
- $g_\lambda^{(n)}(x) = \begin{cases} e^{\lambda \bullet K(x) + H(x) + \tilde{q}(\lambda)} & \text{pour tout } x \in S \\ 0 & \text{sinon} \end{cases}$
- $\{g_\lambda^{(n)}, \lambda \in \Lambda\}$ constitue une famille exponentielle dite canonique.

Régularité dans une famille exponentielle

- Si $\mathcal{F}$ est une famille exponentielle canonique $\{e^{\theta \bullet K(x) + H(x) + q(\theta)}\ :\ \theta \in \Theta\}$ alors le modèle est régulier
- Si $\mathcal{F}^*$ est une famille exponentielle de la forme $\{e^{\eta(\theta) \bullet K(x) + H(x) + q(\theta)}\ :\ \theta \in \Theta\}$ et η est une fonction C^2 alors le modèle est régulier
- Application $\{N(0,\sigma^2)\ :\ \sigma^2 > 0\}$ définit un modèle régulier

2- Optimalité asymptotique

Convergence du MV

Rappel : si $X_1, \ldots X_n$ sont iid suivant un modèle régulier alors l'équation de vraisemblance définie par $\nabla V(\theta) = 0$ admet une solution $\hat{\theta}_n$ telle que $\hat{\theta}_n \xrightarrow[proba]{} \theta_0$.

Théorème TCL-MV : si $X_1, \ldots X_n$ sont iid suivant un modèle régulier tel que $\frac{\partial^2}{\partial \theta^2} \log f_\theta^{(n)}(x)$ est continue en θ uniformément en x et si $0 < I(\theta) < \infty$ alors la suite $\hat{\theta}_n$ est $\sqrt{n}$−consistante et

$$\sqrt{n}(\theta_n - \theta) \underset{loi}{\rightarrow} N(0, I^{-1}(\theta))$$

Efficacité asymptotique

Théorème (Admis) Soit $X_1, \ldots X_n$ sont iid suivant un modèle régulier tel que

$$\frac{\partial^2}{\partial \theta^2} \log f_\theta(X_1) \leq M(\theta_0, X_1) \in L^1 \quad \forall \theta \in \mathcal{V}(\theta_0)$$

Si $\sqrt{n}(\theta_n - \theta_0) \underset{loi}{\rightarrow} N(0, \sigma^2(\theta_0))$ alors $\sigma^2(\theta_0) \geq I^{-1}(\theta_0)$

Définition on dit qu'un estimateur $\hat{\theta}_n$ est asymptotiquement efficace si

$$\sqrt{n}(\hat{\theta}_n - \theta) \underset{loi}{\rightarrow} N(0, I^{-1}(\theta))$$

Amélioration d'un estimateur $\sqrt{n}$ consistant

On définit pour tout $\theta \in \Theta$: $h_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log(f_\theta(X_i))$

Théorème :

Sous les hypothèses du Th TCL-MV si $\sqrt{n}(Y_n - \theta) \overset{loi}{\rightarrow} Z$ avec Z non dégénérée alors $\delta_n = Y_n - \frac{h_n(Y_n)}{h_n'(Y_n)}$ est un estimateur asymptotiquement efficace.

Reparamétrisation

Soit $g : \theta \mapsto \Lambda$ une bijection.

Théorème d'invariance :

Soit $X_1, \ldots X_n$ sont iid suivant un modèle régulier.

Si l'estimateur du MV $\hat{\theta}_n^{ML}$ est asymptotiquement efficace et g est un difféomorphisme alors $g(\hat{\theta}_n^{ML})$ est l'estimateur du MV de $g(\theta)$, et il est asymptotiquement efficace.

Optimalité

Chapitre 3

-1-
Amélioré de Rao Blackwell

Exhaustivité

Soit $X_1, ..X_n \sim P_\theta^{(n)}$ avec $\theta \in \Theta$

Définition On dit que la statistique $S_n = S_n(X_1, ..X_n)$ est exhaustive si la loi conditionnelle de $(X_1, ..X_n)$ sachant S_n ne dépend pas de θ.

Autrement dit

Pour tout A : $P_\theta((X_1, \ldots, X_n) \in A \,|\, S_n)$ ne dépend pas de θ

Ou

Pour toute fonction intégrable h : $E_\theta(h(X_1, \ldots, X_n) \,|\, S_n)$ ne dépend pas de θ

Exemple Si $X_1, ..X_n$ sont iid suivant une loi de Poisson alors $\sum_{i=1}^{n} X_i$ est exhaustive

Théorème de Factorisation

la statistique $S_n = S_n(X_1, ..X_n)$ est exhaustive

si et seulement si

La loi de $(X_1, ..X_n)$ s'écrit de la forme

$$f_\theta^{(n)}(x_1, \ldots, x_n) = g(x_1, \ldots, x_n) h_\theta(S_n(x_1, \ldots, x_n))$$

où g, h_θ sont des fonctions positives

Estimateur du MV et statistique exhaustive

Soit S_n une statistique exhaustive

D'après le théorème de factorisation

$V(\theta) = g(X_1, \ldots, X_n)h_\theta(S_n) \Rightarrow L(\theta) = \log(g(X_1, \ldots, X_n)) + \log(h_\theta(S_n))$

Donc

$\nabla L(\theta) = O \Leftrightarrow \nabla \log(h_\theta(S_n)) = 0$

Conclusion si l'estimateur du maximum de vraisemblance existe il est de la forme

$$\hat{\theta}_n^{MV} = \psi(S_n)$$

On dit que l'estimateur factorise à travers la statistique S_n

Rappel sur l'espérance conditionnelle

Soit (X,Y) des variables aléatoires appartenant à L^1

1. $E(X\,|\,Y)$ est une variable aléatoire de la forme $\phi(Y)$
2. $E(X\,|\,Y) \in L^1$ et $E(E(X\,|\,Y)) = E(X)$
3. Si X et Y sont indépendantes $E(X\,|\,Y) = E(X)$
4. Si $E(g(X)f(Y)\,|\,Y) = f(Y)E(g(X)\,|\,Y)$
5. Si $(X, Y) \in L^2$ alors $var(E(X\,|\,Y)) = var(X) - E(var(X\,|\,Y))$

Théorème de Rao Blackwell

Soit S_n une statistique exhaustive

Si $\hat{\theta}_n$ est un estimateur sans biais du paramètre θ alors

1. L'espérance conditionnelle $E_\theta(\hat{\theta}_n\,|\,S_n)$ est aussi un estimateur du paramètre θ . On le note $\hat{\theta}_n^{RB}$
2. $\hat{\theta}_n^{RB}$ est un estimateur sans biais du paramètre θ
3. Si $\hat{\theta}_n \in L^2$ pour tout $\theta \in \Theta$ alors $\hat{\theta}_n^{RB}$ est meilleur que $\hat{\theta}_n$ au sens L^2 c'est à dire $var_\theta(\hat{\theta}_n^{RB}) \leq var_\theta(\hat{\theta}_n)$ pour tout $\theta \in \Theta$

Définition : $\hat{\theta}_n^{RB}$ est l'estimateur de Rao Blackwell ou l'amélioré de $\hat{\theta}_n$ par Rao Blackwell

Remarque : $E_\theta(\hat{\theta}_n^{RB}\,|\,S_n) = \hat{\theta}_n^{RB}$ car $\hat{\theta}_n^{RB}$ est par construction une fonction de S_n
On n'améliore pas $\hat{\theta}_n^{RB}$ en appliquant à nouveau le théorème de Rao Blackwell

Application

Soit $X_1, \ldots X_n$ iid suivant la loi de Poisson de paramètre θ

$S_n = \sum_{i=1}^{n} X_i$ est une statistique exhaustive

Estimation de θ : $\hat{\theta}_n = X_n$ est un estimateur sans biais de θ

En effet $E_\theta(X_n) = \theta$ pour tout $\theta \in \Theta$

Amélioré de Rao Blackwell : $E_\theta(X_n\,|\,S_n) = \bar{X}_n$

Peut on trouver un estimateur sans biais de θ meilleur que $\bar{X}_n$ au sens L^2?

Statistique totale

Soit $X = (X_1, .. X_n) \sim P_\theta^{(n)}$ avec $\theta \in \Theta$

Définition On dit que la statistique $S_n = S_n(X_1, .. X_n)$ est totale si pour toute fonction h telle que $h(S_n) \in L^1$

$$(E_\theta(h(S_n)) = 0, \quad \forall \theta \in \Theta) \quad \Rightarrow \quad (h(x) = 0, \quad \forall x \in \mathcal{S}_\theta \; \forall \theta \in \Theta)$$

où $\mathcal{S}_\theta$ vérifie $P_\theta(S_n \in \mathcal{S}_\theta^c) = 0$

Exemple -1-

Modèle idd suivant la loi de Poisson : la statistique exhaustive $S_n = \sum_{i=1}^{n} X_i$ est aussi une statistique totale

En effet soit h telle que $h(S_n) \in L^1$ et $E_\theta(h(S_n)) = 0$.

Comme la loi de S_n est la loi de Poisson de paramètre $n\theta$

On a $E_\theta(h(S_n)) = \sum_{i=0}^{\infty} h(i) \frac{e^{-n\theta} n^i \theta^i}{i!} = 0 \; \forall \theta > 0 \Leftrightarrow \sum_{i=0}^{\infty} \frac{h(i) n^i}{i!} \theta^i = 0 \; \forall \theta > 0$

Cette série entière est nulle pour tout $\theta > 0$ donc tous ses coefficients sont nuls $\frac{h(i) n^i}{i!} = \; \forall i \in N$ autrement dit la fonction h est nulle sur N,

Comme le support de la loi de S_n est N pour tout $\theta > 0$, S_n est totale

Exemple -2-

Modèle idd suivant la loi exponentielle : la statistique $S_n = \sum_{i=1}^{n} X_i$ est exhaustive et totale

On a $f_\theta^{(n)}(x_1, \ldots, x_n) = \theta^n e^{-n \sum_{i=1}^n x_i}$, donc d'après le théorème de factorisation $S_n = \sum_{i=1}^{n} X_i$ est exhaustive

Soit h telle que $h(S_n) \in L^1$ et $E_\theta(h(S_n)) = 0$. La loi de S_n est la loi $\Gamma(n, \theta)$

$$E_\theta(h(S_n)) = \int_0^\infty h(x) \frac{e^{-x\theta} x^{n-1} \theta^n}{\Gamma(n)} dx = 0 \; \forall \theta > 0 \Leftrightarrow \int_0^\infty h(x) x^{n-1} e^{-x\theta} dx = 0 \; \forall \theta > 0$$

La transformation de Laplace de $x \mapsto h(x)x^{n-1}$ est nulle sur R_+^* Donc $x \mapsto h(x)x^{n-1}$ est nulle sur $\mathcal{S} \subset \mathcal{R}^+$ tel que $\int_{\mathcal{S}} dx = 1$ et $P_\theta(S_n \in \mathcal{S}^c) = \int_{\mathcal{S}^c} \frac{e^{-x\theta} x^{n-1} \theta^n}{\Gamma(n)} dx = 0$

$S_n = \sum_{i=1}^{n} X_i$ est bien une statistique totale

Famille exponentielle (suite)

- Soit $\{f_\theta(x) = e^{\eta(\theta) \bullet K(x) + H(x) + q(\theta)} \; : \; \theta \in \Theta\}$ une famille exponentielle et $X_1, \ldots X_n$ un n-échantillon iid suivant f_θ
- (Admis) La statistique exhaustive $K_n(X) = \sum_{i=1}^{n} K(x_i)$ est aussi une statistique totale

Théorème d'unicité

Soit $X = (X_1, ..X_n) \sim P_\theta^{(n)}$ avec $\theta \in \Theta$ et S_n une statistique exhaustive totale

L'estimateur amélioré de Rao Blackwell est unique presque surement

Autrement dit si $\theta_n^{(1)}$ et $\theta_n^{(2)}$ sont des estimateurs sans biais de $\theta \in \Theta$ alors $E(\theta_n^{(1)} \mid S_n) = E(\theta_n^{(2)} \mid S_n)$ ps

Théorème d'Optimalité

Soit $X = (X_1, ..X_n) \sim P_\theta^{(n)}$ avec $\theta \in \Theta$ et S_n une statistique exhaustive totale

Si θ_n est un estimateur sans biais de $\theta \in \Theta$ alors parmi les estimateurs sans biais de θ, $E(\theta_n \mid S_n)$ est le meilleur estimateur au sens L^2

-2-
Optimalité L^2

Théorème d'Optimalité

Soit $X = (X_1, ..X_n) \sim P_\theta^{(n)}$ avec $\theta \in \Theta$

On note $\mathscr{D}$ l'ensemble des estimateurs sans biais de $\theta \in \Theta$

et $V_\theta = \inf_{\hat{\theta}_n \in \mathscr{D}} var_\theta(\hat{\theta}_n)$, V_θ est appelée variance minimale.

Proposition

S'il existe un estimateur sans biais de variance V_θ alors il est unique presque surement

Conséquence

S'il existe une statistique exhaustive S_n

et

Si θ_n est un estimateur sans biais de variance minimale V_θ alors

$$\theta_n = E(\theta_n \mid S_n)$$

Borne de Cramer Rao

Théorème

Soit $X = (X_1, .. X_n) \sim P_\theta^{(n)}$ avec $\theta \in \Theta \subset R$. On suppose que le modèle est régulier.

Si $\hat{g}_n$ est un estimateur sans biais de $g(\theta)$ tel que

1) [H] $\frac{\partial}{\partial \theta} E_\theta(h(X)) = \int h(x) \frac{\partial}{\partial \theta} f_\theta^{(n)}(x)\, dx_1 \dots dx_n$ pour $h \equiv 1$ et $h(x) = \hat{g}_n(x) I_n(\theta) > 0$

Alors pour tout $\theta \in \Theta$ on a

$$var_\theta(\hat{g}_n) \geq \frac{g'(\theta)^2}{I_n(\theta)}$$

Remarque L'hypothèse [H] est vérifiée par les familles exponentielles

Définition

1) $\frac{g'(\theta)^2}{I_n(\theta)}$ est la borne de Cramer Rao

2) Un estimateur $\hat{g}_n$ sans biais de $g(\theta)$ est efficace si sa variance atteint la borne de Cramer Rao

C'est à dire $var_\theta(\hat{g}_n) = \frac{g'(\theta)^2}{I_n(\theta)}$ pour tout $\theta \in \Theta$

Extension dans $g : R^p \mapsto R^k$. On note $Dg = (\nabla g_1, \dots, \nabla g_k)$ avec ∇ le gradient

La matrice $var(\hat{g}_n) - {}^tDg(\theta) I_n(\theta)^{-1}\, Dg(\theta)$ est définie positive

Remarque Soit θ_n un estimateur sans biais de θ

Si S_n est exhaustive totale et si $var(\theta_n^{RB}) > I_n^{-1}(\theta)$

Alors il n'existe pas d'estimateur efficace du paramètre θ

Quelles fonctions de $g(\theta)$ peuvent être estimées efficacement ?

Soit $\hat{g}_n = \hat{g}_n(X)$ est un estimateur sans biais de $g(\theta)$

$\hat{g}_n$ est efficace ssi $g'(\theta)^2 = var_\theta(\hat{g}_n) I_n(\theta)$

D'après la preuve pour obtenir un estimateur efficace, il faut avoir une égalité dans l'inégalité de Cauchy-Schwarz C'est à dire

$$\hat{g}_n - g(\theta) = \lambda(\theta) \frac{\partial}{\partial \theta} \log(f_\theta^n(X))$$

Conclusion si g ne vérifie pas la condition ci dessus il n'existe pas d'estimateur efficace

Applications

1) $X_1, \ldots, X_n$ sont iid suivant la loi exponentielle
Pour obtenir des estimateurs efficaces de g il faut que $g(\theta) = C\theta^{-1}$ où C est une constante

2) $X_1, \ldots, X_n$ sont iid suivant la loi Poisson
Pour obtenir des estimateurs efficaces de g il faut que $g(\theta) = C\theta$ où C est une constante

Intervalle de confiance

Chapitre 4

Définitions

Soit $X = (X_1, .. X_n) \sim P_\theta^{(n)}$ avec $\theta \in \Theta$

On fixe $\alpha \in (0,1/2)$

Définition : Un intervalle (ou une région) de confiance de niveau $1 - \alpha$

est un sous ensemble aléatoire $R_n(X_1, \ldots X_n)$ de Θ qui vérifie la propriété suivante, pour tout $\theta \in \Theta$

$$P_\theta(R_n(X_1, \ldots X_n) \ni \theta) \geq 1 - \alpha \qquad \forall \theta \in \Theta$$

Remarques :

- On contrôle la probabilité que la région $R_n(X_1, \ldots X_n)$ contienne θ

$$[R_n(X_1, \ldots X_n) \ni \theta] = \{(X_1 \ldots X_n) \ : \ R_n(X_1, \ldots X_n) \ni \theta\} = S(\theta) \subset E^n$$

Autrement dit

$$R_n(X_1, \ldots X_n) \ni \theta \quad \Leftrightarrow (X_1 \ldots X_n) \in S(\theta)$$

Définitions (suite)

- Le coefficient de confiance de $R_n(X_1, \ldots X_n)$ est défini par

$$\beta_n = \inf_{\theta \in \Theta} P_\theta(R_n(X_1, \ldots X_n) \ni \theta)$$

- On dit que $R_n(X_1, \ldots X_n)$ est de niveau exact $1 - \alpha$

si $P_\theta(R_n(X_1, \ldots X_n) \ni \theta) = 1 - \alpha \qquad \forall \theta \in \Theta$

Remarque : Dans les applications la valeur typique du niveau $1 - \alpha$ est 95%

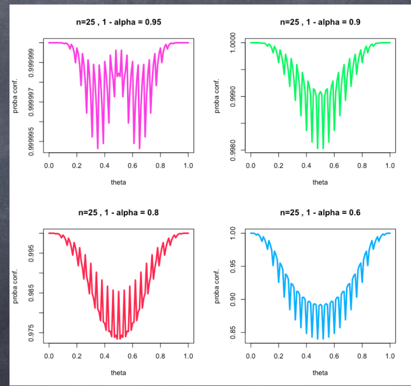
Modèle binomiale

- $X = (X_1, .. X_n)$ iid suivant la loi de Bernoulli de paramètre θ
- D'après l'inégalité de Bienaymé-Tchebychev :

$$P_\theta(|\bar{X}_n - \theta| > \epsilon) \leq \frac{\theta(1-\theta)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}$$

- En prenant $\epsilon = \sqrt{\frac{1}{4n\alpha}}$, on a

$$P_\theta(\theta \in [\bar{X}_n - \epsilon \,;\, \bar{X}_n + \epsilon]) \geq 1 - \alpha$$



Modèle gaussien

- $X = (X_1, .. X_n)$ iid suivant la loi de gaussienne de moyenne θ et de variance 1
- On a $\bar{X}_n \sim N(\theta, 1/n)$ c'est à dire $\sqrt{n}(\bar{X}_n - \theta) \sim N(0,1)$
- Pour tout α on note q_α le quantile d'ordre α de la loi $N(0,1)$
- On a $P_\theta\left(\left[\bar{X}_n - \frac{q_{1-\alpha/2}}{\sqrt{n}} \,;\, \bar{X}_n + \frac{q_{1-\alpha/2}}{\sqrt{n}}\right] \ni \theta\right) = 1 - \alpha$
- $\left[\bar{X}_n - \frac{q_{1-\alpha/2}}{\sqrt{n}} \,;\, \bar{X}_n + \frac{q_{1-\alpha/2}}{\sqrt{n}}\right]$ est un intervalle de niveau exact $1 - \alpha$

Fonction pivotale

Définition : On dit que la fonction $h : E^n \times \Theta \mapsto R$ est une fonction pivotale si la loi de $h(X, \theta)$ ne dépend pas de θ

Pour tout $\theta \in \Theta \quad P_\theta(h(X, \theta) \leq y) = H_n(y)$ où H_n ne dépend pas du paramètre θ

Il existe c_n^-, c_n^+ tel que pour tout $\theta \in \Theta$

$$P_\theta(h(X, \theta) \in [c_n^-, c_n^+]) \geq 1 - \alpha$$

c_n^-, c_n^+ dépendent ni de X ni de θ

Construction de la région de confiance

On construit $\ell_n(\alpha) < u_n(\alpha)$ tels que

$$H_n(u_n(\alpha)) - H_n(\ell_n(\alpha)) \geq 1 - \alpha$$

On définit une région de confiance de niveau $1-\alpha$ en prenant

$$\{(X_1, \ldots, X_n) \ : \ h_n(X_1, \ldots, X_n, \theta) \in [\ell_n(\alpha) \, ; \, u_n(\alpha)]\} \Leftrightarrow \theta \in R_n(X, \ell_n(\alpha), u_n(\alpha))$$

Remarques

$R_n(X, \ell_n(\alpha), u_n(\alpha)) \subset \Theta$ n'est pas nécessairement un intervalle

Il n'y a pas unicité de $\ell_n(\alpha), u_n(\alpha)$. On sélectionne le couple qui donne la plus petite région.

Si $h(X, \theta)$ est une va continue

H_n est continue strictement croissante alors H_n^{-1} existe

Les couples $\ell_n(\alpha), u_n(\alpha)$ s'exprime de la forme

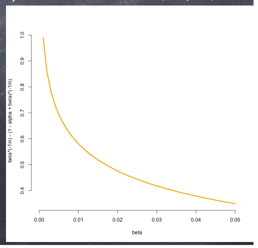
$$\begin{cases} \ell_n(\alpha_\beta) = H_n^{-1}(\beta) \\ u_n(\alpha_\beta) = H_n^{-1}(1 - \alpha + \beta) \end{cases} \text{avec } \beta \in [0, \alpha]$$

1) $H_n(u_n(\alpha_\beta)) - H_n(\ell_n(\alpha_\beta)) = 1 - \alpha$
donc la région de confiance est de niveau exact $1 - \alpha$

2) On sélectionne la valeur de β qui fournit la plus petite région de confiance

Application : modèle uniforme

- $X = (X_1, \ldots X_n)$ iid suivant la loi uniforme sur $[0, \theta]$
- On pose $M_n = \max(X_1, \ldots X_n)$
- $h_n(X, \theta) = \dfrac{M_n}{\theta} = \max(X_1/\theta, \ldots, X_n/\theta)$ est une fonction pivotale.
  En effet $(X_1/\theta, \ldots, X_n/\theta)$ sont iid suivant la loi uniforme sur [0,1]
- Les intervalles $\left[M_n(1 - \alpha + \beta)^{-1/n} \, ; \, M_n \beta^{-1/n}\right]$ avec $\beta \in [0, \alpha]$ sont des intervalles de niveau $1 - \alpha$
- La valeur de β optimale minimise $\beta^{-1/n} - (1 - \alpha + \beta)^{-1/n}$



Construction d'une fonction pivotale

- Soit $T_n(X_1, \ldots X_n)$ une statistique.
- On note G_θ^T la fonction de répartition de la loi de $T_n(X_1, \ldots X_n)$
- Si G_θ^T est inversible alors $h_n(X, \theta) = G_\theta^T(T_n(X_1, \ldots, X_n))$ est une fonction pivotale
- De plus la loi de $h_n(X, \theta)$ est la loi uniforme sur [0,1]

Échantillon Gaussien

- $X = (X_1, .. X_n)$ iid suivant la loi de gaussienne de moyenne μ et de variance σ^2
- On pose $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$ et $S_n^2 = \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$

Théorème

- Les variables aléatoires $\bar{X}_n$ et S_n^2 sont indépendantes
- La loi de $\bar{X}_n$ est la loi gaussienne $\bar{X}_n \sim N\left(\mu\, ; \frac{\sigma^2}{n}\right)$
- La loi de $R_n(\sigma^2) = \frac{S_n^2}{\sigma^2}$ est la loi du χ^2 à (n-1) degrés de liberté (ddl)

Notation

Loi	N(0,1)	Student à n ddl	χ^2 à n ddl
Quantile d'ordre γ	$q(\gamma)$	$t_n(\gamma)$	$x_n(\gamma)$

IC pour la moyenne

- $T_n(\mu)$ est une fonction pivotale pour la moyenne d'un l'échantillon gaussien
- Pour tout $\beta \in [0,\alpha]$

$$\mu \in \left[\bar{X}_n - t_{n-1}(1-\alpha+\beta)\frac{1}{\sqrt{n}}\sqrt{\frac{S_n^2}{n-1}}\, ; \bar{X}_n - t_{n-1}(\beta)\frac{1}{\sqrt{n}}\sqrt{\frac{S_n^2}{n-1}}\right]$$

est un intervalle de confiance de niveau exact $1-\alpha$

- Le choix $\beta = \frac{\alpha}{2}$ est optimal et par symétrie on a

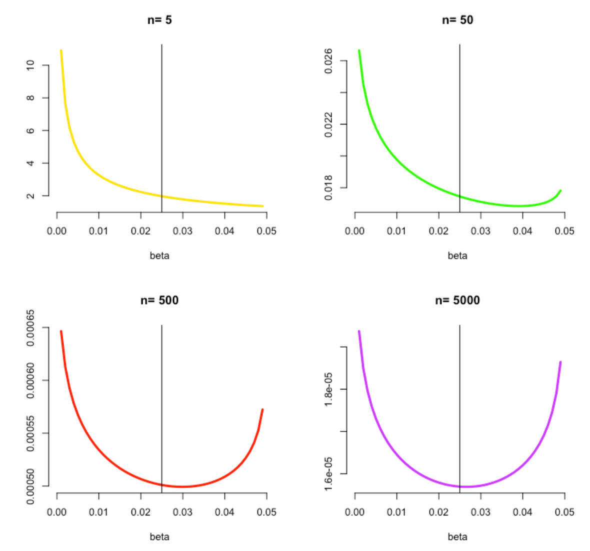
$$\mu \in \left[\bar{X}_n - t_{n-1}(1-\alpha/2)\frac{1}{\sqrt{n}}\sqrt{\frac{S_n^2}{n-1}}\, ; \bar{X}_n + t_{n-1}(1-\alpha/2)\frac{1}{\sqrt{n}}\sqrt{\frac{S_n^2}{n-1}}\right]$$

IC pour la variance

- $R_n(\sigma^2)$ est une fonction pivotale pour la variance d'un l'échantillon gaussien
- Pour tout $\beta \in [0,\alpha]$ $\sigma^2 \in \left[\frac{S_n^2}{x_{n-1}(1-\alpha+\beta)}\, ; \frac{S_n^2}{x_{n-1}(\beta)}\right]$

  est un intervalle de confiance de niveau exact $1-\alpha$
- Le choix $\beta = \alpha/2$ est généralement utilisé en pratique mais ce choix n'est pas optimal

Représentation de la longueur de IC de niveau 95%
en fonction de β



On observe que le choix optimal de β est différent de $\alpha/2 = 2.5\,\%$ (trait en noir) mais il converge vers cette valeur

Échantillon Gaussien

Région de confiance pour (μ, σ^2)

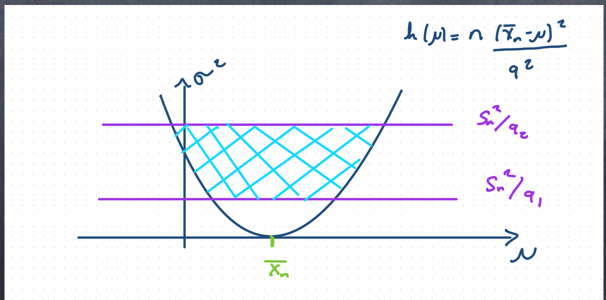
- Soit $q>0$ tel que $P_\theta\left(\dfrac{\bar{X}_n - \mu}{\sigma} \in [-q; q]\right) = \sqrt{1-\alpha}$
- q_1, q_2 tels que $P_\theta\left(\dfrac{S_n^2}{\sigma^2} \in [q_1; q_2]\right) = \sqrt{1-\alpha}$
- Comme $\bar{X}_n$ et S_n^2 sont indépendante on a

$$P_\theta\left(\frac{\bar{X}_n - \mu}{\sigma} \in [-q; q] \quad \frac{S_n^2}{\sigma^2} \in [q_1; q_2]\right) = 1-\alpha$$

La région définit par les contraintes

$$\sigma^2 \in \left[\frac{S_n^2}{q_2}; \frac{S_n^2}{q_1}\right] \quad \text{et} \quad \sigma^2 \geq \frac{n(\bar{X}_n - \mu)^2}{q^2}$$

est une région de confiance de niveau exact $1-\alpha$



Niveau asymptotique

- On dit que $R_n(X)$ est une région de confiance asymptotique de niveau $1-\alpha$ si

$$\forall \theta \in \Theta \quad P_\theta(R_n(X) \ni \theta) \xrightarrow{n\to\infty} 1-\alpha$$

- Une fonction $h_n(X, \theta)$ est asymptotiquement pivotale si

$$\forall \theta \in \Theta \quad h_n(X, \theta) \underset{loi}{\xrightarrow{n\to\infty}} Z \quad \text{et la loi de } Z \text{ ne dépend pas de } \theta$$

Exemple

Soit $X_1, \ldots X_n$ iid L^2 On note $\mu = E(X_1)$

On a $\quad h_n(X,\mu) = \sqrt{n}(\bar{X}_n - \mu)\frac{1}{\hat{\sigma}_n} \overset{loi}{\to} N(0,1)$ avec $\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$

$h_n(X,\mu)$ est une fonction asymptotiquement pivotale

$\mu \in \left[\bar{X}_n - q(1-\alpha/2)\frac{\hat{\sigma}_n}{\sqrt{n}} ; \bar{X}_n + q(1-\alpha/2)\frac{\hat{\sigma}_n}{\sqrt{n}}\right]$ est un intervalle de confiance de niveau asymptotique $1-\alpha$

Niveau exact

Soit $R_n(X)$ une région de confiance pour le paramètre θ de niveau asymptotique $1-\alpha$

Définition le niveau exact de cet région de confiance est

$$\beta_n = \inf_{\theta\in\Theta} \tilde{\beta}_n(\theta) \quad \text{avec } \tilde{\beta}_n(\theta) = P_\theta(R_n(X) \ni \theta)$$

Par construction $\tilde{\beta}_n(\theta) \overset{n\to\infty}{\longrightarrow} 1-\alpha$

Modèle régulier -1-

L'estimateur du maximum de vraisemblance $\hat{\theta}_n^{MV}$ est asymptotiquement efficace : $\sqrt{n}(\hat{\theta}_n^{MV} - \theta) \overset{loi}{\to} N(0,I^{-1}(\theta))$

$\sqrt{n}\sqrt{I(\theta)}(\hat{\theta}_n^{MV} - \theta)$ est asymptotiquement pivotale

Donc $\sqrt{n}\sqrt{I(\theta)}(\hat{\theta}_n^{MV} - \theta) \in [-q(1-\alpha/2] ; q(1-\alpha/2]]$ définit une région de confiance de niveau asymptotique $1-\alpha$

Remarque : il peut être difficile d'exprimer la région de la forme $R_n(X) \ni \theta$

Modèle régulier -2-

L'estimateur du maximum de vraisemblance $\hat{\theta}_n^{MV}$ est consistant donc si I est une fonction continue alors

$\sqrt{n}\sqrt{I(\hat{\theta}_n^{MV})}(\hat{\theta}_n^{MV} - \theta)$ est asymptotiquement pivotale

Donc $\theta \in \left[\hat{\theta}_n^{MV} - \frac{q(1-\alpha/2)}{\sqrt{nI(\hat{\theta}_n^{MV})}} ; \hat{\theta}_n^{MV} + \frac{q(1-\alpha/2)}{\sqrt{nI(\hat{\theta}_n^{MV})}}\right]$ définit un intervalle de confiance de niveau asymptotique $1-\alpha$

méthode -1- Estimation de la variance

On considère $\hat{\theta}_n$ un estimateur de θ tel que

$$\sqrt{n}(\hat{\theta}_n - \theta) \overset{loi}{\to} N(0,\sigma^2(\theta))$$

Méthode 1 : si σ^2 est une fonction continue alors $\dfrac{\sqrt{n}}{\sigma(\hat{\theta}_n)}(\hat{\theta}_n - \theta)$ est une fonction asymptotiquement pivotale et

$$\theta \in \left[\hat{\theta}_n - \frac{q(1-\alpha/2)\sigma(\hat{\theta}_n)}{\sqrt{n}} ; \hat{\theta}_n + \frac{q(1-\alpha/2)\sigma(\hat{\theta}_n)}{\sqrt{n}}\right]$$

définit un intervalle de confiance de niveau asymptotique $1-\alpha$

méthode -2- Delta méthode

On considère $\hat{\theta}_n$ un estimateur de θ tel que

$$\sqrt{n}(\hat{\theta}_n - \theta) \overset{loi}{\to} N(0,\sigma^2(\theta))$$

Méthode 2 : on considère une fonction g de classe C^1 telle que

$$g'(\theta)^2\sigma^2(\theta) = \text{constante}$$

On applique la Δ méthode $\sqrt{n}(g(\hat{\theta}_n) - g(\theta))$ est une fonction asymptotiquement pivotale et

$$g(\theta) \in \left[g(\hat{\theta}_n) - \frac{q(1-\alpha/2)}{\sqrt{n}} ; g(\hat{\theta}_n) + \frac{q(1-\alpha/2)}{\sqrt{n}}\right] \text{ ou } \theta \in g^{-1}\left(\left[g(\hat{\theta}_n) - \frac{q(1-\alpha/2)}{\sqrt{n}} ; g(\hat{\theta}_n) + \frac{q(1-\alpha/2)}{\sqrt{n}}\right]\right)$$

définit une région de confiance de niveau asymptotique $1-\alpha$

Tests d'hypothèses

Chapitre 5

Définition

Objectif

Soit $X = (X_1, ..X_n) \sim P_\theta^{(n)}$ avec $\theta \in \Theta$

- On considère Θ_0 et Θ_1 deux sous ensembles disjoints $\Theta_0 \cap \Theta_1 = \emptyset$ de Θ
- On formule deux hypothèses sur la position du paramètre θ

  H0 : $\theta \in \Theta_0$ contre H1 : $\theta \in \Theta_1$
- H0 est désignée comme l'hypothèse nulle
- H1 est désignée comme l'hypothèse alternative
- A partir des observations $X = (X_1, ..X_n)$ on construit une règle de décision, c'est à dire on décide si $\theta \in \Theta_0$ ou $\theta \in \Theta_1$

  Si on décide que $\theta \in \Theta_0$, on dit que l'on accepte H0 (on rejette H1)

  Si on décide que $\theta \in \Theta_1$; on dit que l'on rejette H0 (on accepte H1)

Erreurs de décision

La décision qui repose sur un échantillon X peut être erronée.

Il y a deux types d'erreur :

- Erreur de type I : on décide que $\theta \in \Theta_1$ alors qu'en réalité $\theta \in \Theta_0$
- Erreur de type II : On décide que $\theta \in \Theta_0$ alors qu'en réalité $\theta \in \Theta_1$

Décision	Réalité H0 est vraie	Réalité H1 est vraie
On rejette H0	Erreur type 1	Décision correcte
On accepte H0	Décision correcte	Erreur type 2

Règle de décision

- On teste les hypothèses H0 contre H1
- On a $X = (X_1, ..X_n) \sim P_\theta^{(n)}$ un échantillon à valeurs dans E^n
- Un test de H0 contre H1 est une région $C \subset E^n$ appelée région critique telle que
  - On rejette H0 / on accepte H1 si $X = (X_1, ..X_n) \in C$
  - On accepte H0 / on rejette H1 si $X = (X_1, ..X_n) \notin C$
- On définit aussi la fonction de test $\phi(X) = 1_C(X) = \begin{cases} 1 & \text{on accepte H1} \\ 0 & \text{on accepte H0} \end{cases}$

Remarque

- L'objectif serait de sélectionner une région critique qui minimise les probabilités de ces erreurs.
- Cela est généralement impossible car les probabilités de ces erreurs ont souvent un effet de balancier.
  - $C = \emptyset$ on rejette jamais H0 donc la proba de l'erreur de type I est nulle et la proba de l'erreur de type II est égale à 1
  - $C = E^n$ on rejette toujours H0 donc la proba de l'erreur de type I est 1 et la proba de l'erreur de type II est nulle
- On considère que l'erreur de type I est la pire des deux erreurs. On sélectionne des régions critiques qui limitent la probabilité d'une erreur de type I

Niveau

Définition :

On dit que la région critique C est de niveau α si

$$\alpha = \sup_{\theta \in \Theta_0} P_\theta(X \in C)$$

- Stratégie : Parmi toutes les régions critiques de niveau α, on cherche les régions qui ont les plus petites probabilités d'erreur de type II.

  Pour tout $\theta \in \Theta_1$ on veut minimiser la probabilité $P_\theta(X \in C^c)$ ce qui est équivalent à maximiser $P_\theta(X \in C)$

Puissance

Définition :

La puissance de la région critique C est une fonction définie sur Θ_1 par $p_C(\theta) = P_\theta(X \in C)$ pour tout $\theta \in \Theta_1$.

- Si C_1, C_2 sont des régions critiques de niveau au plus α alors C_1 est meilleure que C_2 si pour tout $\theta \in \Theta_1$ on a

$$p_{C_1}(\theta) \geq p_{C_2}(\theta)$$

On dit que C_1 est plus puissant que C_2

Objectif : Déterminer le test le plus puissant parmi les tests de niveau au plus α.

Version asymptotique

- On dit que la région critique C est de niveau asymptotique α si

$$\alpha_n = \sup_{\theta \in \Theta_0} P_\theta(X \in C) \to \alpha \quad (n \to \infty)$$

- On dit que le test est consistant si la puissance converge vers 1 :

  pour tout $\theta \in \Theta_1$

$$p_n(\theta) = P_\theta(X \in C) \to 1$$

Notation

On utilise la notation $P_\theta(\cdot \,|H_0)$ pour indiquer que $\theta \in \Theta_O$

et $P_\theta(\cdot \,|H_1)$ pour $\theta \in \Theta_1$

H0 et H1 ne sont pas de évènements donc ceci n'est pas une probabilité conditionnelle

On utilise la terminologie $P_\theta(\cdot \,|H_0)$ est la probabilité sous H0

Exemple

$X = (X_1, ..X_n) \sim B(\theta)$ la loi binomiale de paramètre θ la proba de succès

On veut tester si la probabilité de succès est 1/2 ou inférieure à 1/2

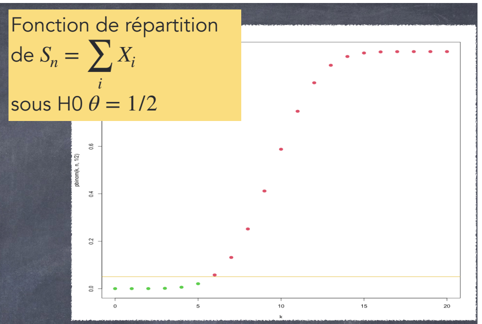
On définit les hypothèses H0 $\theta = 1/2$ contre H1 $\theta < 1/2$

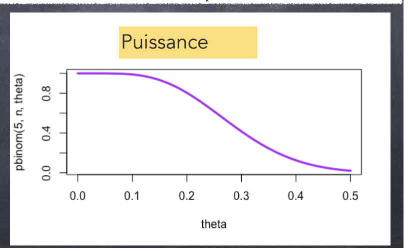
Une région critique intuitive : on rejette H0 en faveur de H1 si $\sum_{i=1}^{n} X_i \leq K$

On choisit K afin de contrôler le niveau : $P_{1/2}(\sum_{i=1}^{n} X_i \leq K) \leq \alpha$

La puissance sera $p(\theta) = P_\theta(\sum_{i=1}^{n} X_i \leq K)$ pour tout $\theta < 1/2$

- n=20
- on fixe le niveau $\alpha = 5\,\%$
- on ne peut pas trouver de test de niveau 5%
- $P(S_n \leq 5) = 2{,}06\,\%$ et $P(S_n \leq 6) > 5\,\%$
- La région critique $C = (S_n \leq 5)$ est un test de niveau 2.06 % donc au plus 5%
- La puissance est $p(\theta) = P_\theta(S_n \leq 5)$ pour tout p<1/2

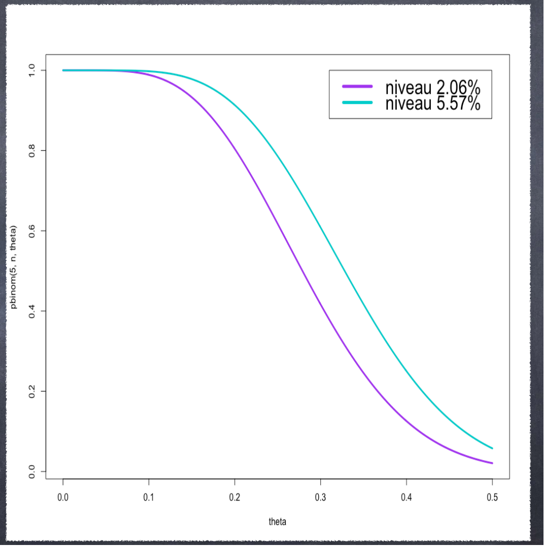




Comparaison des puissances

un test de niveau 2.06 %
$C = (S_n \leq 5)$

un test de niveau 5.57 %
$C^* = (S_n \leq 6)$



Tests du maximum de vraisemblance

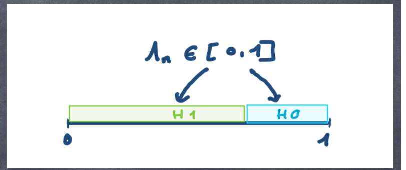
Contexte

- Dans cette partie, on suppose que
  - les observations sont iid
  - l'estimateur du MV $\hat{\theta} = \hat{\theta}^{MV}$ existe et les hypothèses qui assurent la consistance et l'efficacité asymptotique sont satisfaites
- On considère un test bilatéral H0 $\theta = \theta_0$ contre H1 $\theta \neq \theta_0$
- Pour construire le test on s'appuie sur le fait que si θ_0 est la vraie valeur de θ, alors, asymptotiquement, $V(\theta_0)$ est la valeur maximale de $V(\theta)$.

Région critique

On considère la statistique

$$\Lambda_n = \frac{V(\theta_0)}{V(\hat{\theta})}$$



On a $\Lambda_n \leq 1$ pour tout θ_0

Si H0 est vraie alors le rapport Λ_n doit être proche de 1

Si H1 est vraie alors le rapport sera petit par rapport à 1

Règle de décision : On rejette H0 en faveur de H1 si $\Lambda_n \leq K$

On fixe K tel que $P_{\theta_0}(\Lambda_n \leq K) = \alpha$ pour obtenir un test de niveau α

Exemple 1

On suppose que les observations sont iid suivant la loi exponentielle de paramètre $1/\theta$ et on teste H0 $\theta = \theta_0$ contre H1 $\theta \neq \theta_0$

On a

$$\Lambda = e^n \left(\frac{\overline{X}}{\theta_0}\right)^n \exp\{-n\overline{X}/\theta_0\}.$$

Règle de décision de niveau α :

On rejette H0 (on accepte H1) si $\frac{2}{\theta_0}\sum_{i=1}^{n} X_i \leq x_{2n}(\frac{\alpha}{2})$ ou $\frac{2}{\theta_0}\sum_{i=1}^{n} X_i \geq x_{2n}(1-\frac{\alpha}{2})$

Exemple 2

On suppose que les observations sont iid suivant la loi gaussienne $N(\theta, \sigma^2)$ où σ^2 est connue.

On teste H0 $\theta = \theta_0$ contre H1 $\theta \neq \theta_0$

$\Lambda_n = \frac{V(\theta_0)}{V(\hat{\theta})} \leq C \Leftrightarrow -2\log(\Lambda) = \left(\frac{\bar{X}-\theta_0}{\sigma/\sqrt{n}}\right)^2 \geq C'$ et $\left(\frac{\bar{X}-\theta_0}{\sigma/\sqrt{n}}\right)^2$ suit une loi du $\chi^2(1)$

Règle de décision de niveau α :

On rejette H0 (on accepte H1) si $\left(\frac{\bar{X}-\theta_0}{\sigma/\sqrt{n}}\right)^2 \geq x_1(1-\alpha)$

Version asymptotique

Théorème 1 (admis)

Si les hypothèses qui assurent la consistance et l'efficacité asymptotique de l'estimateur du MV sont satisfaites alors

$$\chi_L^2 = -2\log(\Lambda_n) \overset{loi}{\to} \chi^2(1) \quad \text{avec } \Lambda_n = \frac{V(\theta_0)}{V(\hat{\theta})}$$

on teste H0 $\theta = \theta_0$ contre H1 $\theta \neq \theta_O$

Règle de décision

On rejette H0 (on accepte H1) si $\chi_L^2 \geq x_1(1-\alpha)$

c'est un test de niveau asymptotique α

Alternative : test de Wald

Théorème 2

Si les hypothèses qui assurent la consistance et l'efficacité asymptotique de l'estimateur du MV sont satisfaites alors

$$\chi_W^2 = nI(\hat{\theta})(\hat{\theta} - \theta_0)^2 \overset{loi}{\to} \chi^2(1) \quad \text{avec I l'info de Fisher}$$

On teste H0 $\theta = \theta_0$ contre H1 $\theta \neq \theta_O$

Règle de décision

On rejette H0 (on accepte H1) si $\chi_W^2 \geq x_1(1-\alpha)$

Propriétés :

Le test est de niveau asymptotique α

Le test de Wald est consistant

tests de rapport de vraisemblance & optimalité

Définition de l'optimalité

- Un test C de niveau α est sans biais si sa puissance vérifie $p_C(\theta) = P_\theta(X \in C) \geq \alpha$ pour tout $\theta \in \Theta_1$
- On dit qu'un test C est uniformément le plus puissant parmi les tests de niveau au plus α (UPP(α)) si pour tout test C^* de niveau au plus α, $\sup_{\theta \in \Theta_0} P_\theta(X \in C^*) \leq \alpha$, on a $p_C(\theta) \geq p_{C^*}(\theta)$ pour tout $\theta \in \Theta_1$
- On dit qu'un test C est uniformément le plus puissant parmi les tests sans biais de niveau au plus α (UPPSB(α)) si pour tout test C^* sans biais de niveau au plus α on a $p_C(\theta) \geq p_{C^*}(\theta)$ pour tout $\theta \in \Theta_1$

Rapport de vraisemblance

- On suppose que pour tout $\theta \in \Theta$ la vraisemblance existe $V(\theta) = f_\theta^{(n)}(X_1, \dots, X_n) \quad \forall \theta \in \Theta$
- On veut tester $\theta \in \Theta_0$ contre $\theta \in \Theta_1$
- On pose $Z = \dfrac{\sup_{\theta \in \Theta_1} V(\theta)}{\sup_{\theta \in \Theta_0} V(\theta)}$ c'est la statistique du rapport de vraisemblance.
  - Si H1 est vraie alors Z est grand
  - Si H0 est vraie alors Z est petit (proche de 0)

125

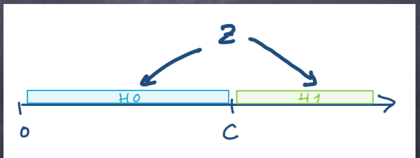
Définition du test

Le test du rapport de vraisemblable pour tester $\theta \in \Theta_0$ contre $\theta \in \Theta_1$ admet une région critique de la forme

$$R = \{Z > C\}$$

On fixe un niveau α

Puis on détermine C pour que le test soit de niveau exact α ou au plus α



126

Théorème de Neyman Pearson

On suppose que $\Theta_i = \{\theta_i\}$ pour i=0,1

On teste $\theta = \theta_0$ contre $\theta = \theta_1$

S'il existe un test de niveau α dont la région critique est de la forme $R = \{Z > K\}$ alors ce test est uniformément le plus puissant parmi les test de niveau au plus α (UPP(α))

Autrement dit

S'il existe K_α tel que $P_{\theta_0}(Z > K_\alpha) = \alpha$ alors pour tout test $C \subset E^n$ tq $P_{\theta_0}(C) \leq \alpha$ on a $P_{\theta_1}(Z > K_\alpha) \geq P_{\theta_1}(C)$

127

Corollaire

On teste $\theta = \theta_0$ contre $\theta = \theta_1$

1. Si la loi de $Z = \dfrac{V(\theta_1)}{V(\theta_0)}$ est continue alors il existe un test du rapport de vraisemblance de niveau α et UPP(α)
2. Si la loi de $Z = \dfrac{V(\theta_1)}{V(\theta_0)}$ est discrète à valeur dans F et si $\alpha \in \{P_{\theta_0}(Z > z), z \in F\}$ alors il existe un test du rapport de vraisemblance de niveau α et UPP(α)
   On dit que $\{P_{\theta_0}(Z > z), z \in F\}$ est l'ensemble des niveaux admissibles

128

Famille exponentielle

- On suppose que la loi de X appartient à une famille exponentielle canonique $\{e^{\theta \bullet K(x) + H(x) + q(\theta)} \ : \ \theta \in \Theta\}$

On teste $\theta = \theta_0$ contre $\theta = \theta_1$

- Si $\theta_1 - \theta_0 > 0$ alors le test du rapport de vraisemblable s'écrit $\{K(X) > C\}$
  - Si la loi de $K(X)$ est continue alors il existe C_α tel que $P_{\theta_0}(K(X) > C_\alpha) = \alpha$ et ce test est UPP(α)
- Si $\theta_1 - \theta_0 < 0$ alors le test du rapport de vraisemblable s'écrit $\{K(X) < C\}$
  - Si la loi de $K(X)$ est continue alors il existe D_α tel que $P_{\theta_0}(K(X) < D_\alpha) = \alpha$ et ce test est UPP(α)

En pratique

On cherche la forme de la région critique du test de Neyman Pearson $Z = \frac{V(\theta_1)}{V(\theta_0)} > K$

On cherche à exprimer la région critique de la forme

$$\textit{une fonction de } (X_1, \ldots X_n) \ : T(X_1, \ldots X_n) < \ et/ou \ > constantes$$

On détermine la constante K pour obtenir niveau α

Cette étape nécessite la connaissance de la loi de $T(X_1, \ldots X_n)$ sous H0 c'est a dire si le paramètre est égal à θ_0

3 situations :

1. On connait explicitement la loi de $T(X_1, \ldots X_n)$ est ses quantiles
2. Approximation par la simulation : On simule un échantillon suivant la loi de T_n et on approche K par les quantiles empiriques
3. [n grand] Approximation de la loi de $T(X_1, \ldots X_n)$ grâce à un théorème limite

Exemple : loi beta($\theta; \theta$)

On observe $X_1, \ldots, X_n$ iid suivant la loi beta de paramètre $(\theta; \theta) \in R_+^* \times R_+^*$

La vraisemblance s'écrit $V(\theta) = \prod_{i=1}^{n} \frac{\Gamma(2\theta)}{\Gamma(\theta)^2} X_i^{\theta-1}(1 - X_i)^{\theta-1}$

On teste $\theta = \theta_0$ contre $\theta = \theta_1$ on suppose que $\theta_0 < \theta_1$

La région critique est de la forme $\sum_{i=1}^{n} \log(X_i(1 - X_i)) > K$

K est le quantile d'ordre $1 - \alpha$ de la loi de $T_n = \sum_{i=1}^{n} \log(X_i(1 - X_i))$ lorsque $\theta = \theta_0$

Difficulté pratique : la loi de T_n et ses quantiles ne sont pas connus explicitement

En pratique

On observe un échantillon de taille n = 20

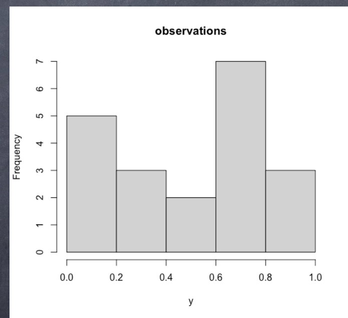
On teste $\theta = 1 \, contre \, \theta = 2$ au niveau 5%

Remarque Pour $\theta = 1$ on retrouve la loi uniforme sur [0,1]

On observe $T_n = -42.08$

Question :

Est ce que l on rejette l'hypothèse nulle au niveau 5% ?



Approximation par simulation : Méthode de Monte Carlo

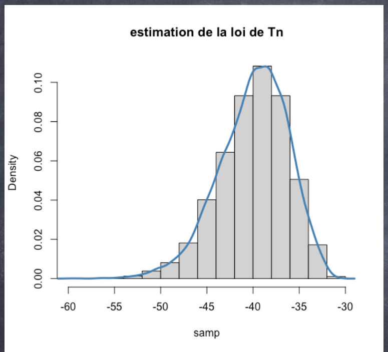
- On répète B (grand) les deux états suivantes :

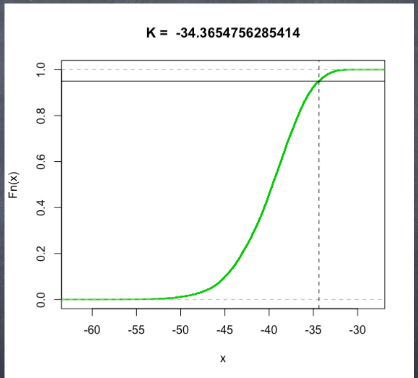
  [1] On simule un échantillon $X_1^{sim}, \ldots, X_n^{sim}$ iid suivant la loi beta(1,1) (H0)
  [2] On calcule $T_n^{sim} = \sum \log(X_i^{sim}(1 - X_i^{sim}))$

- On approche K par le quantile empirique d'ordre 95% de l'échantillon $T_n^{sim_1}, \ldots, T_n^{sim_B}$ simulé suivant la loi de la statistique T_n sous H0


```
theta0 = 1
n=20

randT <- function()
 { x = rbeta(n,theta0,theta0 )
  Tn = sum(log(x*(1-x)))
  return(Tn)
}

samp.T = replicate(10000, randT())

hist(samp.T, proba = "TRUE" , main ="estimation de la loi de Tn")
lines(density(samp.T), lwd =3 ,col="steelblue")
plot(ecdf(samp.T), col= "green3", lwd= 2, main = paste("K = ",quantile(samp, .95)))
abline(h=.95)
quantile(samp, .95)
abline(v =quantile(samp, .95) , lty =2 )
```


Approximation de Monte Carlo





On a observé $T_n = -42.08 < K$
Décision : on accepte H0 au niveau 5%

Approximation asymptotique (si n grand) de K par une approximation gaussienne (TCL)

Sous H0 c'est à dire $\theta = \theta_0$

$(\log(X_1(1 - X_1)), \ldots, \log(X_n(1 - X_n)))$ est un échantillon de variables aléatoires iid L^2. On note m_0 l'espérance de $\log(X_1(1 - X_1))$ et σ_0^2 sa variance

On applique le TCL : $\frac{\sqrt{n}}{\sigma_0}\left(\frac{T_n}{n} - m_0\right) = \frac{1}{\sqrt{n}\sigma_0}(T_n - nm_0) \overset{loi}{\to} N(0,1)$

on pose $N \sim N(0,1)$

$$P_{\theta_0}(T_n > K) = P_{\theta_0}\left(\frac{1}{\sqrt{n}\sigma_0}(T_n - nm_0)) > \frac{1}{\sqrt{n}\sigma_0}(K - nm_0)\right) \approx P\left(N > \frac{1}{\sqrt{n}\sigma_0}(K - nm_0)\right) = 1 - \alpha$$

$\Rightarrow \frac{1}{\sqrt{n}\sigma_0}(K - nm_0) \approx q_{1-\alpha}$ et donc $K \approx nm_0 + \sqrt{n}\sigma_0 q_{1-\alpha}$

2 situations

Situation 1

On sait calculer explicitement les valeurs de m_0 et σ_0

C'est à dire on sait calculer $E_{\theta_0}\left(\left(\log(X_1(1-X_1))\right)^k\right)$ pour k =1,2

Situation 2

On approche les valeurs de m_0 et σ_0 par une méthode de Monte Carlo

Suite situation 2

On simule un échantillon de taille B iid suivant la loi de $\log(X_1(1-X_1))$ avec $\theta = \theta_0$

On approche m_0 par la moyenne et σ_0 par l'écart type de l'échantillon

On calcule $K = -39.58419$

Cette approximation est valide pour n grand

Ici n=20 « petit » échantillon

On obtient une valeur assez différente de celle obtenue en approchant la loi de T_n

```
n=20
theta0 = 1
Y.beta = rbeta(10000,theta0,theta0)
Y = log(Y.beta*(1-Y.beta))
m0 = mean(Y)
sigma0 = sd(Y)
K = n*m0 +qnorm(.95) *sigma0/sqrt(n)
```


tests d'hypothèses composées

Définitions

- Test bilatéral : on teste H0 $\theta = \theta_0$ contre H1 $\theta \neq \theta_0$
- Test unilatéral : on teste
  - T1 : H0 $\theta = \theta_0$ contre H1 $\theta > \theta_0$
  - T2 : H0 $\theta \leq \theta_0$ contre H1 $\theta > \theta_0$
  - T3 : H0 $\theta = \theta_0$ contre H1 $\theta < \theta_0$
  - T4 : H0 $\theta \geq \theta_0$ contre H1 $\theta < \theta_0$

rapport de vraisemblance monotone.

Définition

Une famille paramétrique $\{f_\theta^{(n)}(X_1,\ldots,X_n), \forall\theta\in\Theta\}$ est une famille à rapport de vraisemblance monotone en $U(X_1,\ldots,X_n)$ si pour tout $\theta_1>\theta_2$ il existe h croissante tel que $\frac{V(\theta_1)}{V(\theta_2)} = h_{\theta_1,\theta_2}(U(X_1,\ldots,X_n))$

Exemple

une famille exponentielle canonique $\{e^{\theta\bullet K(x)+H(x)+q(\theta)} \ : \ \theta\in\Theta\}$ est une famille à rapport de vraisemblance monotone

Test unilatéral UPP(α)

Théorème On suppose

1. $\{f_\theta^{(n)}(X_1,\ldots,X_n), \forall\theta\in\Theta\}$ est une famille à rapport de vraisemblance monotone en $U(X_1,\ldots,X_n)$
2. La loi de $U(X_1,\ldots,X_n)$ est continue pour tout $\theta\in\Theta$

Alors

A. Pour les tests unilatéraux T1 et T2, le test $\{U(X_1,..X_n)>C_\alpha\}$ avec $P_{\theta_0}(U(X_1,..,X_n)>C_\alpha)=\alpha$ est un test de niveau α et UPP(α)

B. Pour les tests unilatéraux T3 et T4 le test $\{U(X_1,..X_n)<D_\alpha\}$ avec $P_{\theta_0}(U(X_1,..,X_n)<D_\alpha)=\alpha$ est un test de niveau α et UPP(α)

Test sur la moyenne θ d'un échantillon gaussien de variance connue

On considère des observations iid suivant la loi $N(\theta,\sigma^2)$ avec σ^2 connue

D'après le théorème précédent on a

A. Pour les tests unilatéraux T1 et T2 le test $\left\{\bar{X}_n>\theta_0+\frac{\sigma q_{1-\alpha}}{\sqrt{n}}\right\}$ est un test de niveau α et UPP(α)

B. Pour les tests unilatéraux T3 et T4 le test $\left\{\bar{X}_n<\theta_0+\frac{\sigma q_{\alpha}}{\sqrt{n}}\right\}$ est un test de niveau α et UPP(α)

Test bilatéral sur la moyenne θ d'un échantillon gaussien de variance connue

On considère des observations iid suivant la loi $N(\theta,\sigma^2)$ avec σ^2 connue

On veut tester H0 $\theta=\theta_0$ contre H1 $\theta\neq\theta_0$

Le test $\left\{|\bar{X}_n-\theta_0|>\frac{\sigma q_{1-\alpha/2}}{\sqrt{n}}\right\}$ est un test de niveau α

Ce test n'est pas un test UPP(α)

Ce test est un test UPPSB(α)

Tests avec paramètre de nuisance

Dans cette partie on suppose que $\Theta \subset R^P$ et on veut effectuer un test sur une coordonnée de θ

Les autres coordonnées sont appelés les paramètres de nuisance

Exemple pour des échantillons gaussiens

- On effectue un test sur la moyenne et la variance est le paramètre de nuisance
- Ou inversement

Dans cette partie, on suppose que la loi des observations appartient à une famille exponentielle canonique

$\{e^{\theta \bullet K(x) + H(x) + q(\theta)} \ : \ \theta \in \Theta\}$ où θ est de dimension p

On effectue un test sur la i ème coordonnée (i=1,..,p)

- T1 : H0 $\theta_i \leq \theta_0$ contre H1 $\theta_i > \theta_O$ ou $\theta_i = \theta_0$ contre $\theta_i > \theta_0$
- T2 : H0 $\theta_i \geq \theta_0$ contre H1 $\theta_i < \theta_O$ ou $\theta_i = \theta_0$ contre $\theta_i < \theta_0$
- T3 : H0 $\theta_i = \theta_0$ contre H1 $\theta_i \neq \theta_0$

Les autres coordonnées $\theta_{-i} = \{\theta_j, j \neq i\}$ sont inconnues

Théorème

On suppose qu'il existe une fonction $g : R^p \to R$ croissante par rapport à la i ème coordonnée telle que

la loi de la statistique $W = g(K_1(X), \ldots, K_p(X))$ ne dépend pas de θ_{-i} si $\theta_i = \theta_0$

1. Pour T1 : S'il existe C_α tq $P_{\theta_i=\theta_0}(W > C_\alpha) = \alpha$ alors le test $\{W > C_\alpha\}$ est de niveau α et UPPSB(α)
2. Pour T2 : S'il existe D_α tq $P_{\theta_i=\theta_0}(W < D_\alpha) = \alpha$ alors le test $\{W < D_\alpha\}$ est de niveau α et UPPSB(α)

3. Pour T3
on suppose que g est linéaire par rapport à la i ème coordonnée
S'il existe C_1, C_2 tel que

$$P_{\theta_i=\theta_0}(W < C_1) = P_{\theta_i=\theta_0}(W > C_2) = \alpha/2$$

Alors $\{W < C_1\} \cup \{W > C_2\}$ est un test de niveau α et UPPS(α)

Application pour les échantillons gaussiens $N(\mu, \sigma^2)$

La famille des lois gaussienne est une famille exponentielle

$$f_\theta(X) = e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n X_i^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^n X_i} \frac{1}{(\sqrt{2\pi}\sigma)^2} e^{-n\mu^2/(2\sigma^2)}$$

avec $\theta = \begin{pmatrix} \frac{\mu}{\sigma^2} \\ \frac{-1}{2\sigma^2} \end{pmatrix}$ et $K(X) = \begin{pmatrix} \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i^2 \end{pmatrix} = \begin{pmatrix} K_1 \\ K_2 \end{pmatrix}$

On peut appliquer le théorème précédent sur les coordonnées du paramètre θ

Test sur la variance

T1 : $\sigma^2 \le \sigma_0^2$ contre $\sigma^2 > \sigma_0^2$

T2 : $\sigma^2 \ge \sigma_0^2$ contre $\sigma^2 < \sigma_0^2$

T3 : $\sigma^2 = \sigma_0^2$ contre $\sigma^2 \ne \sigma_0^2$

Ces tests peuvent être reformuler sur le paramètre $\theta_2 = -\frac{1}{2\sigma^2}$ (avec les mêmes inégalités)

Si $\sigma^2 = \sigma_0^2$
alors la loi de $W = \frac{1}{\sigma_0^2}\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{\sigma_0^2}(K_2 - \frac{1}{n}K_1^2) \sim \chi^2(n-1)$ (La loi ne dépend pas de μ)

$W = g(K_1, K_2)$ est une fonction linéaire croissante par rapport à la 2 ème coordonnée

1) T1 : $\sigma^2 \le \sigma_0^2$ contre $\sigma^2 > \sigma_0^2$
Le test de région critique $\sum_{i=1}^n (X_i - \bar{X}_n)^2 > \sigma_0^2 x_{n-1}(1-\alpha)$ est de niveau α et UPPSB(α)

2) T2 : $\sigma^2 \ge \sigma_0^2$ contre $\sigma^2 < \sigma_0^2$
Le test de région critique $\sum_{i=1}^n (X_i - \bar{X}_n)^2 < \sigma_0^2 x_{n-1}(\alpha)$ est de niveau α et UPPSB(α)

3) T3 : $\sigma^2 = \sigma_0^2$ contre $\sigma^2 \ne \sigma_0^2$
Le test de région critique
$\sum_{i=1}^n (X_i - \bar{X}_n)^2 < \sigma_0^2 x_{n-1}(\alpha/2) \cup \sum_{i=1}^n (X_i - \bar{X}_n)^2 > \sigma_0^2 x_{n-1}(1-\alpha/2)$
est de niveau α et UPPSB(α)

Test sur la moyenne

T1 : $\mu \leq 0$ contre $\mu > 0$ est équivalent $\theta_1 \leq 0$ contre $\theta_1 > 0$

T2 : $\mu \geq 0$ contre $\mu < 0$ est équivalent $\theta_1 \geq 0$ contre $\theta_1 < 0$

T3 : $\mu = 0$ contre $\mu \neq 0$ est équivalent $\theta_1 = 0$ contre $\theta_1 \neq 0$

Si $\mu = 0\ (\theta_1 = 0)$ alors $T_n = \sqrt{n}\dfrac{\bar{X}_n}{\sqrt{\frac{S_n^2}{n-1}}} \sim Student(n-1)$ avec $S_n^2 = \sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = K_2 - \frac{1}{n}K_1^2$

On a $T_n = g(K_1, K_2)$ avec g croissante par rapport à la 1 er coordonnée

Le théorème précédent s'applique pour T1 et T2

Mais il ne s'applique pas pour T3 car g n 'est pas linéaire.

1) T1 : $\mu \leq 0$ contre $\mu > 0$ est équivalent $\theta_1 \leq 0$ contre $\theta_1 > 0$

La test de région critique $\bar{X}_n > \dfrac{S_n}{\sqrt{n}\sqrt{n-1}}t_{n-1}(1-\alpha)$ est de niveau α et UPPSB(α)

2) T2 : $\mu \geq 0$ contre $\mu < 0$ est équivalent $\theta_1 \geq 0$ contre $\theta_1 < 0$

La test de région critique $\bar{X}_n < \dfrac{S_n}{\sqrt{n}\sqrt{n-1}}t_{n-1}(\alpha)$ est de niveau α et UPPSB(α)

3) Tester $\mu \leq \mu_0$ contre $\mu > \mu_0$ est équivalent à T1 sur l'échantillon $X_1 - \mu_0, \ldots, X_n - \mu_0$ $\bar{X}_n \to \bar{X}_n - \mu_o$ et $S_n \to S_n$

On obtient un test de niveau α et UPPSB(α) en prenant comme région critique
$\bar{X}_n > \mu_0 + \dfrac{S_n}{\sqrt{n}\sqrt{n-1}}t_{n-1}(1-\alpha)$

4) Tester $\mu \geq \mu_0$ contre $\mu < \mu_0$ est équivalent à T2 sur l'échantillon $X_1 - \mu_0, \ldots, X_n - \mu_0$

On obtient un test de niveau α et UPPSB(α) en prenant comme région critique
$\bar{X}_n < \mu_0 + \dfrac{S_n}{\sqrt{n}\sqrt{n-1}}t_{n-1}(\alpha)$

Proposition pour le test T3 : $\mu = \mu_0$ contre $\mu \neq \mu_0$

Le test de région critique

$$\bar{X}_n > \mu_0 + \frac{S_n}{\sqrt{n}\sqrt{n-1}}t_{n-1}(1-\alpha/2) \cup \bar{X}_n < \mu_0 + \frac{S_n}{\sqrt{n}\sqrt{n-1}}t_{n-1}(\alpha/2)$$

ce qui est équivalent à $|\bar{X}_n - \mu_0| > \dfrac{S_n}{\sqrt{n}\sqrt{n-1}}t_{n-1}(1-\alpha/2)$

est un test de niveau α. Il est UPPSB(α)

Test et Pvalue

C'est une présentation alternative des résultats d'un test

Notation : pour tout α on note R_α la région critique d'un test de niveau α

Définition La Pvalue est un variable aléatoire à valeur dans [0,1] définie par

$$Pvalue = \inf\{\beta \,|\, X \in R(\beta)\}$$

A partir de la valeur de la Pvalue on peut donner la décision du test à tous niveau

Règle de décision : Si le niveau est égal à α alors

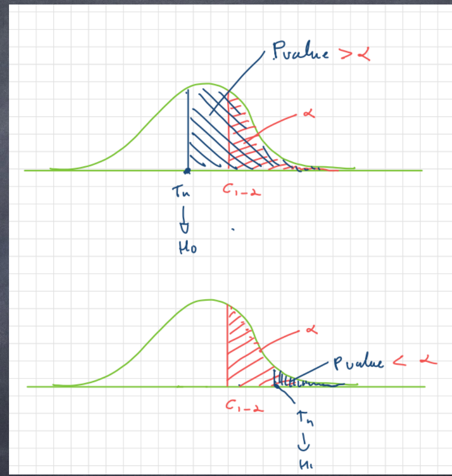
on accepte H0 / on rejette H1 au niveau α si $Pvalue > \alpha$

on accepte H1 / on rejette H0au niveau α si $Pvalue < \alpha$

Si le test s'écrit $\{T_n(X_1,\ldots,X_n) > c_{1-\alpha}\}$

avec $c_{1-\alpha}$ le quantile d'ordre $1-\alpha$ de la loi de fonction de répartition F_0

alors $Pvalue = 1 - F_0(T_n(X_1,\ldots,X_n))$



Si le test s'écrit $\{T_n(X_1,\ldots,X_n) < c_\alpha\}$ avec c_α le quantile d'ordre α de la Loi de fonction de répartition F_0

alors $Pvalue = F_0(T_n(X_1,\ldots,X_n))$

Si le test s'écrit $\{T_n(X_1,\ldots,X_n) < c_{\alpha/2}\} \cup \{T_n(X_1,\ldots,X_n) > c_{1-\alpha/2}\}$

alors $Pvalue = 2 * min(F_0(T_n(X_1,\ldots,X_n)), 1 - F_0(T_n(X_1,\ldots,X_n)))$