

Statistiques
Master Statistique et économétrie
Notes de cours

V. Monbet

Master 1 - 2011

Table des matières

1	Introduction	4
2	Principes de l'estimation ponctuelle	6
2.1	Modèle statistique	6
2.1.1	Définitions	6
2.1.2	Identifiabilité	7
2.1.3	Familles exponentielles	8
2.1.4	Statistiques	10
2.1.5	Exhaustivité	10
2.2	Estimation ponctuelle	12
2.2.1	Biais et erreurs en moyenne quadratique	12
2.2.2	Consistance	14
3	Estimation par maximum de vraisemblance	15
3.1	La vraisemblance	15
3.2	Propriétés asymptotiques de l'EMV	16
3.2.1	Convergence en loi	17
3.2.2	Consistance	20
3.3	Estimateur efficace	22
4	Tests d'hypothèses	25
4.1	Principe et définitions	25
4.1.1	Hypothèses de test	25
4.1.2	Principe général	26
4.1.3	Erreurs et puissance	26
4.1.4	Principe de Neyman	27
4.2	Exemples	28
4.2.1	Test du signe	28
4.2.2	Test pour la moyenne d'une loi de Gauss	29
4.3	Principe de Neyman et optimalité	31
4.3.1	Test randomisé	31
4.3.2	Tests uniformément plus puissants	32
4.4	Tests UPP pour les hypothèses composites	37
4.4.1	Test unilatéral $H_0 : \theta \leq \theta_0$ contre $H_1 : \theta > \theta_0$	37
4.4.2	Test bilatéral $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$	39
4.5	Généralisation	41

4.6	Tests classiques	45
4.6.1	Tests paramétriques pour des moyennes, des variances ou des corrélations	45
4.6.2	Tests non paramétriques pour des moyennes , des variances ou des corrélations	45
4.6.3	Test d'adéquation	45
5	Estimation par intervalles	46
5.1	Exemple	46
5.2	Méthode générale pour construire des intervalles de confiance	47
5.3	Lien avec les tests	47

Chapitre 1

Introduction

Généralités

En probabilité on suppose (implicitement ou explicitement) que tous les paramètres nécessaires pour calculer des probabilités associées à un modèle particulier sont connus. Par exemple, on peut calculer la probabilité d'occurrence d'un événement donné de façon exacte ou approchée (à l'aide des théorèmes limite). En statistique, le rôle des paramètres (des modèles de probabilité) et des événements/résultats (liés à une expérience) sont en quelques sortes inversés. Le résultat d'une expérience est observé par l'expérimentateur tandis que la vraie valeur du paramètre (et plus généralement le modèle de probabilité) est inconnue de l'expérimentateur. Ainsi, l'objectif de la statistique est d'utiliser les résultats d'une expérience (c'est à dire les données) pour inférer la valeur des paramètres inconnus du modèle de probabilité supposé être sous-jacent.

Le paragraphe précédent suggère qu'il n'y a pas d'ambiguïté dans le choix du modèle sous-jacent à une expérience donnée. Cependant, en réalité dans les problèmes statistiques, on verra qu'il y a de nombreuses incertitudes quand au choix du modèle et ce choix est souvent fait essentiellement sur la base des observations. Et dans la grande majorité des cas, le modèle est seulement une approximation de la réalité ; et il est important pour un statisticien de vérifier que les modèles supposés sont plus ou moins proches de la réalité et d'être conscient des conséquences du choix d'un "mauvais" modèle.

Une philosophie reconnue en statistique est qu'un modèle doit être aussi simple que possible. On préférera toujours un modèle ayant peu de paramètres à un modèle caractérisé par un grand nombre de paramètres.

Notations

Dans la mesure du possible, nous utiliserons les notations suivantes dans ce cours :

- Variables aléatoires : lettres majuscules (ex : X_1, \dots, X_n)
- Observations : lettres minuscules (ex : x_1, \dots, x_n)
- Paramètres : lettres grecques (ex : θ, μ, σ)

Plan du cours

Estimation ponctuelle

Dans la première partie du cours, nous aborderons le problème de l'estimation ponctuelle de paramètres.

Exemple : On suppose que l'on a deux candidats A et B lors d'une élection. On cherche à prédire la proportion de votes pour A à partir d'un échantillon représentatif de taille n sélectionné par un institut de sondage. Chaque individu de l'échantillon donne son intention de vote. On modélise le choix A pour l'individu i par une variable aléatoire

$$X_i = 1 \text{ si } i \text{ vote A, } X_i = 0 \text{ sinon}$$

Les X_i suivent une loi de Bernoulli de paramètre π inconnu. On dispose de n observations x_1, \dots, x_n de X_1, \dots, X_n . On cherche à inférer π à partir de l'échantillon x_1, \dots, x_n . Une estimation naturelle est

$$\hat{\pi}_n = \frac{1}{n} \sum_{i=1}^n x_i$$

On se pose alors des questions : Cette quantité estime t'elle bien le paramètre p ? Peut-on lui associer une marge d'erreur ? Que se passe t'il si on dispose d'un échantillon plus grand (c'est à dire si n grandit) ? Existe t'il d'autres quantités qui donnerait une meilleure estimation de p ? En existe t'il une qui est optimale pour un critère bien choisi ?

Théorie de la décision : tests statistiques

Dans une seconde partie, nous introduirons le concept de test statistique.

Exemple : On teste l'efficacité d'un médicament contre le cholestérol. On dispose pour cela de n individus pour lesquels on a effectué deux mesures du taux de cholestérol, l'une avant et l'autre après le traitement. On note

- X : le taux de cholestérol avant le traitement,
- Y : le taux de cholestérol après le traitement.

On dispose donc des couples d'observations $(x_1, y_1), \dots, (x_n, y_n)$ et on veut déterminer à l'aide de ces observations si

$$D = Y - X$$

est positif. Etant donnés les caractères aléatoires de D et de l'expérience, on décidera de l'efficacité du traitement si la moyenne observée

$$\bar{d}_n = \frac{1}{n} (y_i - x_i)$$

est plus grande qu'un seuil. Ce seuil est défini en fonction du risque de se tromper qui est fixé par l'expérimentateur et de la taille n de l'échantillon.

Chapitre 2

Principes de l'estimation ponctuelle

2.1 Modèle statistique

2.1.1 Définitions

Soient X_1, \dots, X_n des variables aléatoires définies sur (Ω, \mathcal{A}, P) dans un espace mesurable (E_n, \mathcal{E}_n) . On suppose qu'on observe x_1, \dots, x_n qui sont telles que pour tout i , x_i est une *réalisation* (un tirage) de la variable aléatoire X_i . Supposons que la distribution jointe de $\mathbf{X} = (X_1, \dots, X_n)$ est inconnue mais qu'elle appartient à une famille particulière de distributions. Le couple formé par l'espace d'observation E_n et cette famille de distributions \mathcal{P}_n est appelé *modèle statistique*. On note (E_n, \mathcal{P}_n) .

Remarque - Bien qu'on suppose ici que \mathbf{X} est observée, on peut parler de modèle statistique pour \mathbf{X} même si certaines variables X_i ne sont pas observables.

En général, les distributions appartenant à un modèle statistique sont indexées par un paramètre $\theta \in \Theta$; θ représente typiquement la partie inconnue ou non spécifiée du modèle. On peut alors écrire

$$\mathbf{X} = (X_1, \dots, X_n) \sim F_\theta \text{ pour } \theta \in \Theta$$

où F_θ est la distribution jointe de \mathbf{X} et Θ l'ensemble des toutes les valeurs possibles pour θ . Le plus souvent, $\Theta \subset \mathbb{R}^p$. Si $p > 1$, $\theta = (\theta_1, \dots, \theta_p)$ est un vecteur de paramètres. Le modèle statistique est ici $(E_n, F_\theta)_{\{\theta \in \Theta\}}$. On notera parfois abusivement $P_\theta(A)$, $E_\theta(X)$ et $Var_\theta(X)$ pour les probabilités, espérance et variance qui dépendent de θ .

Si les variables X_1, \dots, X_n sont indépendantes et identiquement distribuées (i.i.d.), on a

$$P_{(X_1, \dots, X_n)} = P_{X_1} \times \dots \times P_{X_n} = P_{X_1}^n$$

Et on note alors le modèle statistique $(E, P_{X_1})^n$ où E est l'espace de définition de X_1 .

Quand les distributions d'un modèle peuvent être indexées par un paramètre de dimension finie, on parlera de *modèle paramétrique*. Il existe des cas où le paramètre est de dimension infinie. On parle alors, à tort, de *modèle non paramétrique*. Dans la suite du cours non nous intéressons principalement aux modèles paramétriques. On notera $f(x, \theta)$ la densité de P_θ relativement à une mesure dominante et σ -finie, μ . On va se restreindre au cas où

- au cas où μ est la mesure de Lebesgue (variables aléatoires de loi absolument continue) et on retrouve la densité $f_\theta(x)$ ou,
- au cas où μ est la mesure de comptage (variables aléatoires de loi discrète) et on retrouve le système $P_\theta(X = x)$. On note \mathbf{X} l'échantillon (X_1, \dots, X_n) issu du même modèle (E^n, P_θ) .

Exemple - Dans l'exemple des votes, X_1, \dots, X_n sont des variables indépendantes de même loi de Bernoulli de paramètre $\pi \in [0, 1]$. Chaque variable X_i est définie sur $E = \{0, 1\}$.

Exemple - Supposons que X_1, \dots, X_n sont des variables aléatoires indépendantes et identiquement distribuées suivant une loi de Poisson de moyenne λ .

La probabilité jointe de $\mathbf{X} = (X_1, \dots, X_n)$ a pour densité

$$f(\mathbf{x}; \lambda) = \prod_{i=1}^n \frac{\exp(-\lambda) \lambda^{x_i}}{x_i!}$$

pour une réalisation $\mathbf{x} = (x_1, \dots, x_n)$ de \mathbf{X} . L'espace des paramètres pour ce modèle est l'ensemble $\{\lambda : \lambda > 0\}$.

2.1.2 Identifiabilité

Pour un modèle statistique donné, un paramètre donné θ correspond à une unique distribution F_θ . Cependant, il peut exister des valeurs distinctes du paramètre, θ_1 et θ_2 telles que $F_{\theta_1} = F_{\theta_2}$. Pour éviter cette difficulté, on requiert qu'un modèle, ou plus précisément sa paramétrisation, soit *identifiable*. En effet une paramétrisation non identifiable pose souvent des problèmes d'estimation.

On dit qu'un modèle a une paramétrisation identifiable si $F_{\theta_1} = F_{\theta_2}$ implique que $\theta_1 = \theta_2$.

Exemple - Supposons que X_1, \dots, X_n sont des variables aléatoires indépendantes gaussiennes avec

$$E(X_i) = \beta_0 + \beta_1 t_i + \beta_2 s_i$$

où t_1, \dots, t_n et s_1, \dots, s_n sont des constantes connues, et $Var(X_i) = \sigma^2$. L'espace des paramètres est

$$\{(\beta_0, \beta_1, \beta_2, \sigma) : -\infty < \beta_k < \infty, \sigma > 0\}$$

La paramétrisation de ce modèle est identifiable si et seulement si les vecteurs

$$z_0 = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \quad z_1 = \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix} \quad \text{et} \quad z_2 = \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix}$$

sont linéairement indépendants, c'est à dire que $a_0 z_0 + a_1 z_1 + a_2 z_2 = 0$ implique que $a_0 = a_1 = a_2 = 0$. Notons

$$\mu = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix}$$

et remarquons que la paramétrisation est identifiable si il y a une bijection entre les valeurs possibles de μ et les paramètres $\beta_0, \beta_1, \beta_2$. Supposons maintenant que z_0, z_1 et z_2 sont linéairement dépendants; alors on peut avoir $a_0 z_0 + a_1 z_1 + a_2 z_2 = 0$ avec au moins un des coefficients a_0, a_1 ou a_2 non nul. Dans ce cas, nous aurons

$$\begin{aligned}\mu &= \beta_0 z_0 + \beta_1 z_1 + \beta_2 z_2 \\ &= (\beta_0 + a_0) z_0 + (\beta_1 + a_1) z_1 + (\beta_2 + a_2) z_2\end{aligned}$$

et par conséquent il n'y a pas bijection entre μ et $(\beta_0, \beta_1, \beta_2)$. Cependant, quand z_0, z_1 et z_2 sont linéairement dépendants, il est possible d'obtenir une paramétrisation identifiable en restreignant l'espace des paramètres; ceci est réalisé en contraignant les paramètres $\beta_0, \beta_1, \beta_2$.

Dans la suite, nous supposons implicitement que les modèles statistiques sont identifiables; sauf mention contraire.

2.1.3 Familles exponentielles

Une classe importante de modèles statistiques est la classe des modèles de la famille exponentielle.

Définition 1 *Supposons que X_1, \dots, X_n a une distribution jointe F_θ avec $\theta = (\theta_1, \dots, \theta_p)$ un paramètre (inconnu). On dit que la famille de distribution $\{F_\theta\}$ est une famille exponentielle à k paramètres si la densité de probabilité jointe de (X_1, \dots, X_n) est de la forme*

$$f(\mathbf{x}; \theta) = \exp \left[\sum_{i=1}^k c_i(\theta) T_i(\mathbf{x}) - d(\theta) + S(\mathbf{x}) \right]$$

pour $\mathbf{x} \in A$ avec A un ensemble qui ne dépend pas de θ .

Remarque - k n'est pas forcément égal à p , même si c'est souvent le cas.

Exemple - Loi binomiale.

Supposons que X suit une loi binomiale de paramètres n et θ avec θ inconnu. Alors

$$\begin{aligned}f(x, \theta) &= \binom{n}{x} \theta^x (1 - \theta)^{1-x} \\ &= \exp \left[\ln \left(\frac{\theta}{1 - \theta} \right) x + n \ln(1 - \theta) + \ln \left(\binom{n}{x} \right) \right]\end{aligned}$$

pour $x \in A = \{0, 1, \dots, n\}$ et ainsi la distribution de X est dans une famille exponentielle à 1 paramètre.

Exemple - Loi de Gauss.

Supposons que X_1, \dots, X_n sont des variables aléatoires normales i.i.d. de moyenne θ et de variance θ^2 avec $\theta > 0$. La densité de probabilité jointe de (X_1, \dots, X_n) est

$$\begin{aligned}f(\mathbf{x}; \theta) &= \prod_{i=1}^n \left[\frac{1}{\theta \sqrt{2\pi}} \exp \left(-\frac{1}{2\theta^2} (x_i - \theta)^2 \right) \right] \\ &= \exp \left[-\frac{1}{2\theta^2} \sum_{i=1}^n x_i^2 + \frac{1}{\theta} \sum_{i=1}^n x_i - \frac{n}{2} (1 + \ln(\theta^2)) + \ln(2\pi) \right]\end{aligned}$$

et on a $A = \mathbb{R}^n$. On conclut que c'est une loi de la famille exponentielle à 2 paramètres en dépit du fait que l'espace des paramètres est de dimension 1.

Proposition 1 Si X_1, \dots, X_n sont i.i.d. de loi appartenant à la famille exponentielle alors la loi jointe du vecteur (X_1, \dots, X_n) appartient à la famille exponentielle.

Preuve - La densité de (X_1, \dots, X_n) est

$$\begin{aligned} f(\mathbf{x}; \theta) &= \prod_{i=1}^n f(x_i; \theta) \text{ car les } x_i \text{ sont i.i.d.} \\ &= \prod_{i=1}^n \exp \left[\sum_{j=1}^k c_j(\theta) T_j(x_i) - d(\theta) + S(x_i) \right] \\ &= \exp \left[\sum_{j=1}^k c_j(\theta) \sum_{i=1}^n T_j(x_i) - nd(\theta) + \sum_{i=1}^n S(x_i) \right] \\ &= \exp \left[\sum_{j=1}^k c_j(\theta) \tilde{T}_j(\mathbf{x}) - nd(\theta) + \tilde{S}(\mathbf{x}) \right] \end{aligned}$$

□

Proposition 2 Supposons que $\mathbf{X} = (X_1, \dots, X_n)$ est distribué suivant une loi d'une famille exponentielle à un paramètre avec une densité de la forme

$$f(\mathbf{x}; \theta) = \exp [c(\theta)T(\mathbf{x}) - d(\theta) + S(\mathbf{x})]$$

pour $\mathbf{x} \in A$ où

- (a) l'espace Θ du paramètre est ouvert,
- (b) $c(\theta)$ est une fonction bijective sur Θ
- (c) $c(\theta)$ et $d(\theta)$ sont deux fois différentiables sur Θ .

Alors

$$\begin{aligned} E_\theta[T(\mathbf{X})] &= \frac{d'(\theta)}{c'(\theta)} \\ \text{Var}_\theta[T(\mathbf{X})] &= \frac{d''(\theta)c'(\theta) - d'(\theta)c''(\theta)}{(c'(\theta))^3} \end{aligned}$$

Preuve - Définissons $\phi = c(\theta)$; ϕ est appelé paramètre naturel de la famille exponentielle. Soit

$$d_0(\phi) = d(c^{-1}(\phi))$$

c^{-1} est bien défini car c est une bijection sur Θ . Alors pour s assez petit (tel que $\phi + s$ reste dans l'espace des paramètres), on a

$$E_\phi[\exp(sT(\mathbf{X}))] = \exp[d_0(\phi + s) - d_0(\phi)]$$

qui est la fonction génératrice des moments de $T(\mathbf{X})$. En différenciant, et en posant $s = 0$, on obtient

$$E_\phi[T(\mathbf{X})] = d'_0(\phi) \text{ et } Var_\phi[T(\mathbf{X})] = d''_0(\phi)$$

Maintenant notons que

$$d'_0(\phi) = \frac{d'(\theta)}{c'(\theta)}$$

$$d''_0(\phi) = \frac{d''(\theta)c'(\theta) - d'(\theta)c''(\theta)}{(c'(\theta))^3}$$

et qui conclut la preuve. \square

2.1.4 Statistiques

Supposons que le modèle statistique pour $\mathbf{X} = (X_1, \dots, X_n)$ a un espace de paramètres Θ . Comme le paramètre θ est inconnu, nous cherchons à extraire de l'information le concernant dans \mathbf{X} , sans perdre trop d'information.

Définition 2 Une statistique est une fonction $T : E^n \rightarrow \mathbb{R}^p$ qui à \mathbf{X} associe $T(\mathbf{X})$ et qui ne dépend d'aucun paramètre inconnu ; autrement dit, T ne dépend que de variables aléatoires observables et de constantes connues.

Une statistique peut être scalaire ou vectorielle.

Exemple - Moyenne empirique :

$$T(\mathbf{X}) = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

La taille n est connue et T est bien une statistique.

Il est important de bien comprendre qu'une statistique est elle même une variable aléatoire et qu'elle a sa propre loi de probabilité ; cette distribution peut éventuellement dépendre de θ .

2.1.5 Exhaustivité

Afin de faire de l'inférence statistique, le statisticien va devoir extraire de l'information de la suite de variables aléatoires X_1, \dots, X_n dont il dispose. Lorsque la taille de l'échantillon n est grande, il est naturel de tenter de réduire l'échantillon et de résumer l'information qui y est contenue. Lorsque il est possible de "remplacer" (X_1, \dots, X_n) par une statistique $T = T(X_1, \dots, X_n)$, on optera bien sûr pour cette solution. Cependant, une question se pose : Comment savoir si la réduction des données opérée par la statistique T ne conduit pas à une perte d'information ? C'est ce type de problèmes que cherche à résoudre la notion d'exhaustivité.

L'idée est basée sur la remarque suivante : si la loi conditionnelle de \mathbf{X} sachant T ne dépend pas de la loi P_θ de \mathbf{X} , alors T est suffisamment informative pour P_θ . En effet, dans cette situation

la loi conditionnelle de \mathbf{X} sachant S peut être spécifiée indépendamment de P_θ , lorsqu'on donne $S = s$, on peut générer une variable aléatoire X' de même loi que X . Donc les informations données par $\mathbf{X} = (X_1, \dots, X_n)$ ne donnent pas plus sur P_θ que T ne le fait. T est dite alors statistique exhaustive (ou suffisante).

Définition 3 - La statistique T sera dite exhaustive pour θ si la loi conditionnelle de \mathbf{X} sachant $T(\mathbf{X}) = t$ n'est pas une fonction du paramètre θ :

$$P_\theta(\mathbf{X}|T(\mathbf{X}) = t)$$

ne dépend pas de θ .

Exemple - Supposons que X_1, \dots, X_k soient des variables indépendantes de loi binomiale de paramètres n_i connus et θ (inconnu). Soit $T = X_1 + \dots + X_k$; T a aussi une loi binomiale de paramètres $m = n_1 + \dots + n_k$ et θ . Pour montrer que T est une statistique exhaustive pour θ nous devons montrer que

$$P_\theta[\mathbf{X} = \mathbf{x}|T = t]$$

est indépendant de θ pour tout t et tout x_1, \dots, x_k . Tout d'abord notons que si $t \neq x_1 + \dots + x_k$ alors sa probabilité conditionnelle est 0. Si $t = x_1 + \dots + x_k$ alors

$$\begin{aligned} P_\theta[\mathbf{X} = \mathbf{x}|T = t] &= \frac{P_\theta[\mathbf{X} = \mathbf{x}]}{P_\theta[T = t]} \\ &= \frac{\prod_{i=1}^k \binom{n_i}{x_i} \theta^{x_i} (1 - \theta)^{(n_i - x_i)}}{\binom{m}{t} \theta^t (1 - \theta)^{(m - t)}} \\ &= \frac{\prod_{i=1}^k \binom{n_i}{x_i}}{\binom{m}{t}} \end{aligned}$$

ce qui est indépendant de θ . Ainsi T est une statistique exhaustive pour θ .

Le problème est que, dans la plupart des lois (en particulier les lois continues), il est difficile d'utiliser directement la définition pour montrer qu'une statistique est exhaustive. De plus, cette définition ne permet pas d'identifier ou de construire des statistiques exhaustives. Mais il existe un critère simple de Jerzy Neyman qui donne une condition nécessaire et suffisante pour que T soit une statistique exhaustive quand la loi de \mathbf{X} admet une densité.

Théorème 1 (Théorème de factorisation) - Supposons que $\mathbf{X} = (X_1, \dots, X_n)$ admet une densité jointe $f(\mathbf{x}; \theta)$ pour $\theta \in \Theta$. Alors, $T = T(\mathbf{X})$ est une statistique exhaustive pour θ si et seulement s'il existe deux fonctions mesurables $g : \mathbb{R}^p \times \Theta \rightarrow \mathbb{R}^+$ et $h : E \rightarrow \mathbb{R}^+$ telles que $f(x, ; \theta)$ se met sous la forme

$$f(\mathbf{x}; \theta) = h(x)g(T(x), \theta)$$

(T et θ peuvent être des vecteurs).

La preuve rigoureuse de ce théorème est difficile dans le cas des variables continues.
Faire le schéma de la preuve ?

Exemple - Supposons que X_1, \dots, X_n sont des variables i.i.d. avec pour densité

$$f(x; \theta) = \frac{1}{\theta} \text{ pour } 0 \leq x \leq \theta$$

avec $\theta > 0$. La densité jointe de $\mathbf{X} = (X_1, \dots, X_n)$ est

$$\begin{aligned} f(\mathbf{x}; \theta) &= \frac{1}{\theta^n} \text{ pour } 0 \leq x_1, \dots, x_n \leq \theta \\ &= \frac{1}{\theta^n} I(0 \leq x_1, \dots, x_n \leq \theta) \\ &= \frac{1}{\theta^n} I\left(\max_{1 \leq i \leq n} x_i \leq \theta\right) I\left(\min_{1 \leq i \leq n} x_i \geq 0\right) \\ &= g(\max_i x_i; \theta) h(\mathbf{x}) \end{aligned}$$

et $X_{(n)} = \max_{i=1, \dots, n} (X_i)$ est une statistique exhaustive pour θ .

Exemple - Supposons que X_1, \dots, X_n sont des variables i.i.d. issues d'une loi exponentielle à k paramètres de densité

$$f(\mathbf{x}; \theta) = \exp \left[\sum_{i=1}^k c_i(\theta) T_i(\mathbf{x}) - d(\theta) + S(\mathbf{x}) \right] I(\mathbf{x} \in A)$$

en prenant $h(\mathbf{x}) = \exp[S(\mathbf{x})] I(\mathbf{x} \in A)$, on obtient par le théorème de factorisation que, $T = (T_1(\mathbf{X}), \dots, T_k(\mathbf{X}))$ est une statistique exhaustive pour θ .

2.2 Estimation ponctuelle

Un *estimateur* est une statistique qui a pour vocation d'estimer la vraie valeur d'un paramètre θ . Ainsi si

$$\mathbf{X} \sim F_\theta \text{ pour } \theta \in \Theta$$

alors un estimateur $\hat{\theta}$ est égal à une statistique $T(\mathbf{X})$.

Supposons que θ est un paramètre réel et que $\hat{\theta}$ est un estimateur de θ . La distribution de l'estimateur $\hat{\theta}$ est souvent appelée, distribution empirique de $\hat{\theta}$. Idéalement, nous souhaitons que la distribution empirique de $\hat{\theta}$ soit centrée sur θ et de variance faible. Plusieurs mesures de qualité d'un estimateur sont basées sur sa distribution.

2.2.1 Biais et erreurs en moyenne quadratique

Définition 4 - Le biais¹ d'un estimateur $\hat{\theta}$ est défini par

$$b_{\theta}(\hat{\theta}) = E_{\theta}(\hat{\theta}) - \theta$$

On dit qu'un estimateur est non biaisé ou sans biais si $b_{\theta}(\hat{\theta}) = 0$.

Définition 5 - L'erreur en moyenne absolue² (EMA) de θ est définie par

$$EMA = E_{\theta}(|\hat{\theta} - \theta|)$$

Définition 6 - L'erreur en moyenne quadratique³ (EMQ) de θ est définie par

$$EMQ = E_{\theta}((\hat{\theta} - \theta)^2)$$

Le biais indique si la distribution empirique de l'estimateur $\hat{\theta}$ est centrée ou non sur la vraie valeur du paramètre θ tandis que l'erreur en moyenne absolue et l'erreur en moyenne quadratique donnent des informations sur la dispersion de la distribution autour de θ . Les erreurs EMA et EMQ sont de bonnes mesures pour comparer plusieurs estimateurs d'un paramètre θ . On utilise plus souvent l'EMQ que l'EMA en particulier parce que l'EMQ se décompose en une somme du biais au carré et de la variance de l'estimateur :

$$EMQ = Var_{\theta}(\hat{\theta}) + b_{\theta}(\hat{\theta})^2$$

Preuve à faire en TD.

Cette décomposition permet notamment de calculer ou d'approcher facilement l'EMQ alors que l'EMA est plus difficile à calculer. Notamment, si la variance de l'estimateur est bien plus grande que le biais alors $EQM \simeq Var_{\theta}(\theta)$.

Exemple - Supposons que X_1, \dots, X_n sont des variables i.i.d. de loi uniforme sur $[0, \theta]$. Et choisissons $\hat{\theta} = \max_{i=1, \dots, n} X_i$. On montre (voir TD) que la densité de $\hat{\theta}$ est

$$f(x; \theta) = \frac{n}{\theta^n} x^{n-1} \text{ pour } 0 \leq x \leq \theta$$

On obtient alors

$$E_{\theta}(\hat{\theta}) = \frac{1}{n+1} \theta$$

D'où on déduit facilement un estimateur sans biais de θ :

$$\tilde{\theta} = \frac{n+1}{n} \max_{i=1, \dots, n} X_i$$

¹bias

²mean absolute error

³mean square error

2.2.2 Consistance

Supposons que $\hat{\theta}_n$ soit un estimateur d'un paramètre θ construit à partir de n variables aléatoires X_1, \dots, X_n . Quand n croît, il est raisonnable d'attendre que la loi empirique de $\hat{\theta}$ soit de plus en plus concentrée autour de la vraie valeur du paramètre θ . Cette propriété de la suite $\{\hat{\theta}_n\}$ est la consistance.

Définition 7 - Une suite d'estimateurs $\{\hat{\theta}_n\}$ est dite consistante pour θ si $\{\hat{\theta}_n\}$ converge en probabilité vers θ , c'est à dire si,

$$\lim_{n \rightarrow +\infty} P_\theta[|\hat{\theta}_n - \theta| > \epsilon] = 0$$

pour tout $\epsilon > 0$ et pour toute valeur de θ .

On dit abusivement que " $\hat{\theta}_n$ est un estimateur consistant de θ ".

Exemple - Supposons que X_1, \dots, X_n sont des variables i.i.d. de moyenne μ , alors \bar{X}_n est un estimateur consistant de μ . En effet par la loi forte des grands nombres, on a que \bar{X}_n tend presque sûrement vers μ et on en déduit la consistance.

Chapitre 3

Estimation par maximum de vraisemblance

Un des estimateurs les plus utilisés en statistique est l'estimateur du maximum de vraisemblance. La vraisemblance est une fonction qui contient toute l'information des données sur un paramètre inconnu. Elle joue un rôle important dans de nombreuses méthodes statistiques. Et l'estimateur du maximum de vraisemblance a de très bonnes propriétés d'optimalité.

3.1 La vraisemblance

Soient X_1, \dots, X_n des variables aléatoires de densité $f(\mathbf{x}; \theta)$ avec $\theta \in \Theta$. Et considérons $\mathbf{x} = (x_1, \dots, x_n)$ une réalisation du vecteur $\mathbf{X} = (X_1, \dots, X_n)$. La fonction de vraisemblance¹ est définie par

$$\mathcal{L}(\theta) = f(\mathbf{x}; \theta)$$

C'est une fonction réelle définie sur l'espace des paramètres Θ . La vraisemblance donne, en quelque sorte, la probabilité que la réalisation $\mathbf{x} = (x_1, \dots, x_n)$ soit émise par le modèle associé à la valeur θ du paramètre. Ainsi plus il est probable que le modèle émette cette réalisation pour la valeur θ , plus la vraisemblance sera grande.

Définition 8 - Soit $\mathbf{X} = (X_1, \dots, X_n)$ un vecteur aléatoire suivant le modèle F_θ , $\theta \in \Theta$. Pour une réalisation $\mathbf{x} = (x_1, \dots, x_n)$, l'estimateur du maximum de vraisemblance (EMV) est l'estimateur $\hat{\theta} = S((X))$ avec S telle que

$$\mathcal{L}(S(x)) \geq \mathcal{L}(\theta) \text{ pour tout } \theta \in \Theta$$

Exemple pour une loi discrète - Supposons que X_1, \dots, X_n sont des variables aléatoires i.i.d de loi de Bernoulli de paramètre inconnu $\theta \in]0, 1[$. Et notons $f(., \theta)$ la densité de X_i pour tout $i = 1, \dots, n$. On a

$$f(x; \theta) = \theta^x (1 - \theta)^{1-x}$$

¹likelihood function

Soit (x_1, \dots, x_n) une réalisation de (X_1, \dots, X_n) . Par définition, la vraisemblance est

$$\mathcal{L}(\theta; x_1, \dots, x_n) = \prod_{i=1}^n f_\theta(x_i)$$

car les variables X_i sont indépendantes. D'où

$$\mathcal{L}(\theta; x_1, \dots, x_n) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i}$$

On en déduit la log-vraisemblance :

$$\ln \mathcal{L}(\theta; \mathbf{x}) = \sum_{i=1}^n x_i \ln \theta + (n - \sum_{i=1}^n x_i) \ln(1 - \theta)$$

Le maximum est la racine de la dérivée de cette expression dont la dérivée seconde est négative.

$$\frac{\partial \ln \mathcal{L}(\theta; \mathbf{x})}{\partial \theta} = \frac{\sum_{i=1}^n x_i}{\theta} + \frac{n - \sum_{i=1}^n x_i}{1 - \theta}$$

s'annule en $\theta^* = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}_n$. On vérifie que

$$\frac{\partial^2 \ln \mathcal{L}(\theta; \mathbf{x})}{\partial \theta^2} < 0$$

au voisinage de $\theta^* = \bar{x}_n$. Et on conclut que l'estimateur du maximum de vraisemblance pour le paramètre de la loi de Bernoulli est $\hat{\theta}_n = \bar{X}_n$.

Propriétés 1 - Soit ϕ telle que $\phi = g(\theta)$ pour une fonction bijective g . Si $\hat{\theta}$ est un EMV pour θ alors $\hat{\phi} = g(\hat{\theta})$ est un EMV pour ϕ .

3.2 Propriétés asymptotiques de l'EMV

Nous allons voir dans cette section, que sous des conditions faibles de régularité, on peut montrer que l'EMV est consistant et asymptotiquement normal. Dans la suite de cette section, nous supposons que X_1, \dots, X_n sont des v.a.i.i.d. de densité $f(x; \theta)$ avec $\theta \in \Theta \subset \mathbb{R}$ avec $\ell(x; \theta) = \ln f(x; \theta)$ trois fois différentiable par rapport à θ . Et nous ajoutons les hypothèses suivantes :

- (A1) Θ est un sous ensemble ouvert de \mathbb{R} .
- (A2) L'ensemble $A = \{x : f(x; \theta) > 0\}$ ne dépend pas de θ .
- (A3) $f(x; \theta)$ est trois fois continûment différentiable par rapport à θ sur A .
- (A4) $E_\theta[\ell'(X_i; \theta)] = 0$ pour tout θ et $Var_\theta[\ell'(X_i; \theta)] = I(\theta)$ où $0 < I(\theta) < \infty$ pour tout θ .
- (A5) $E_\theta[\ell''(X_i; \theta)] = -J(\theta)$ où $0 < J(\theta) < \infty$ pour tout θ .
- (A6) Pour θ et $\delta > 0$, $|\ell'''(x; t)| < M$ pour $|\theta - t| \leq \delta$ où $E_\theta[M(X_i)] < \infty$.

Définition 9 Si le modèle (E, P_θ) vérifie les hypothèses (A1) à (A3) et que l'intégrale $\int_B f(x; \theta) d\mu(x)$ est au moins deux fois dérivable sous le signe d'intégration pour tout borélien B , alors on dit que le modèle est un modèle régulier.

Exemples - Les modèles basés sur la loi exponentielle et sur la loi de Gauss sont réguliers mais pas celui basé sur la loi uniforme car dans ce dernier cas, le domaine de définition du modèle dépend du paramètre de la loi.

3.2.1 Convergence en loi

Proposition 3 - Sous l'hypothèse (A2), on a $I(\theta) = J(\theta)$ soit $\text{Var}_\theta[\ell'(X_i; \theta)] = -E_\theta[\ell''(X_i; \theta)]$. $I(\theta)$ est appelée information de Fisher.

Preuve - La condition (A2) implique

$$\int_A f(x; \theta) dx = 1 \text{ pour tout } \theta \in \Theta$$

Si on peut échanger le signe somme et la dérivée, on a

$$\begin{aligned} 0 &= \int_A \frac{\partial}{\partial \theta} f(x; \theta) dx \\ &= \int_A \ell'(x; \theta) f(x; \theta) dx \\ &= E_\theta[\ell'(X_i; \theta)] \end{aligned}$$

De plus, si on dérive deux fois sous le signe \int , on a

$$\begin{aligned} 0 &= \int_A \frac{\partial}{\partial \theta} (\ell'(x; \theta) f(x; \theta)) dx \\ &= \int_A \ell''(x; \theta) f(x; \theta) dx + \int_A (\ell'(x; \theta))^2 f(x; \theta) dx \\ &= -J(\theta) + I(\theta) \end{aligned}$$

◇

Théorème 2 - Sous les hypothèses (A1) à (A5), on a un théorème de limite centrale pour l'EMV $\hat{\theta}_n$ de θ :

$$\sqrt{n}(\hat{\theta}_n - \theta) \rightarrow_d \frac{Z}{J(\theta)} \sim \mathcal{N}(0, I(\theta)/J^2(\theta))$$

avec \rightarrow_d la convergence en loi.

Lemme 1 - lemme de Slutsky - Soit $\{X_n\}_{n=0, \dots, \infty}$ une suite de variables aléatoires qui tend en loi vers X et soit $\{Y_n\}_{n=0, \dots, \infty}$ une suite de variables aléatoires qui tend en probabilité vers une constante $c \in \mathbb{R}$, alors $X_n + Y_n$ tend en loi vers $X + c$ et $X_n Y_n$ tend en loi vers cX .

Preuve du théorème - Sous les conditions (A1) à (A3), si $\hat{\theta}_n$ maximise la vraisemblance, on a

$$\sum_{i=1}^n \ell'(X_i; \hat{\theta}_n) = 0$$

et en écrivant un développement de Taylor de cette expression, on obtient

$$\begin{aligned} 0 &= \sum_{i=1}^n \ell'(X_i; \hat{\theta}_n) = \sum_{i=1}^n \ell'(X_i; \theta) \\ &+ (\hat{\theta}_n - \theta) \sum_{i=1}^n \ell''(X_i; \theta) + \frac{1}{2} (\hat{\theta}_n - \theta)^2 \sum_{i=1}^n \ell'''(X_i; \theta_n^*) \end{aligned}$$

où θ_n^* appartient à l'intervall $[\min(\hat{\theta}_n, \theta), \max(\hat{\theta}_n, \theta)]$. En divisant les termes de gauche et de droite par \sqrt{n} , on peut encore écrire

$$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{-(1/\sqrt{n}) \sum_{i=1}^n \ell'(X_i; \theta)}{n^{-1} \sum_{i=1}^n \ell''(X_i; \theta) + (\hat{\theta}_n - \theta)(2n^{-1}) \sum_{i=1}^n \ell'''(X_i; \theta_n^*)}$$

Par le théorème de limite centrale et la condition (A4), on obtient

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \ell'(X_i; \theta) \rightarrow_d Z \sim \mathcal{N}(0, I(\theta))$$

et par la loi faible des grands nombres et la condition (A5), on a

$$\frac{1}{n} \sum_{i=1}^n \ell''(X_i; \theta) \rightarrow_P -J(\theta)$$

On conclut la preuve en utilisant le lemme de Slutsky. \diamond

Théorème 3 - Delta méthode - *Supposons que* that

$$a_n(X_n - \theta) \rightarrow_d Z$$

où θ est une contante et $\{a_n\}$ une suite de constantes telle que $a_n \rightarrow +\infty$. Si $g(x)$ est une fonction dérivable de dérivée $g'(\theta)$ en θ alors

$$n(g(X_n) - g(\theta)) \rightarrow_d g'(\theta)Z.$$

Preuve de la Delta méthode - Développement de Taylor + lemme de Slutsky. Notons d'abord que $X_n \rightarrow_P \theta$ (d'après le lemme de Slutsky). Si g est continuellement différentiable en θ ,

$$g(X_n) = g(\theta) + g(\theta_n^*)(X_n - \theta)$$

avec θ_n^* entre X_n et θ ; ainsi $|\theta_n^* - \theta| \leq |X_n - \theta|$ et $\theta_n^* \rightarrow_P \theta$. Comme $g'(x)$ est continue en θ , il suit que $g'(\theta_n^*) \rightarrow_P g'(\theta)$. Maintenant,

$$a_n(g(X_n) - g(\theta)) = g'(\theta_n^*)a_n(X_n - \theta) \rightarrow_d g'(\theta)Z$$

par le lemme de Slutsky. \diamond

Estimation de l'écart-type des estimateurs

Dans les cas où $I(\theta) = J(\theta)$, le résultat du théorème 2 suggère que pour n assez grand, l'EMV $\hat{\theta}_n$ suit approximativement une loi de Gauss de moyenne θ et de variance $1/(nI(\theta))$. Ce résultat peut être utilisé pour approcher l'écart-type de $\hat{\theta}_n$ par $[nI(\theta)]^{-1/2}$. Comme $I(\theta)$ dépend de θ il est nécessaire de l'estimer pour obtenir une approximation de l'écart-type de $\hat{\theta}$. Il y a deux approches pour faire ceci :

- Si $I(\theta)$ une expression analytique, on peut substituer $\hat{\theta}_n$ à θ dans cette expression. On obtient alors

$$\widehat{se}(\hat{\theta}_n) = \frac{1}{\sqrt{nI(\hat{\theta}_n)}}$$

$nI(\hat{\theta}_n)$ est appelée *information de Fisher attendue* pour θ .

- Comme $I(\theta) = -E[\ell''(X_i; \theta)]$, on peut estimer $I(\theta)$ par

$$\widehat{I(\theta)} = -\frac{1}{n} \sum_{i=1}^n \ell''(X_i; \hat{\theta}_n)$$

ce qui conduit à l'écart-type

$$\widehat{se}(\hat{\theta}_n) = \frac{1}{\sqrt{n\widehat{I(\theta)}}} = \left(-\frac{1}{n} \sum_{i=1}^n \ell''(X_i; \hat{\theta}_n) \right)^{-1/2}$$

$n\widehat{I(\theta)}$ est appelée *information de Fisher observée* pour θ .

Exemple - Supposons que X_1, \dots, X_n sont des variables aléatoires i.i.d. de loi géométrique de densité de probabilité

$$f(x; \theta) = \theta(1 - \theta)^x \text{ pour } \theta = 0, 1, \dots$$

L'estimateur du maximum de vraisemblance de θ basé sur X_1, \dots, X_n est

$$\hat{\theta}_n = \frac{1}{\bar{X}_n + 1}$$

Par le théorème de limite centrale, on a $\sqrt{\bar{X}_n - (\theta^{-1} - 1)}$ tend en loi vers une v.a. Z de loi de Gauss de moyenne nulle et de variance $\theta^{-2}(1 - \theta)$. Ainsi nous on obtient que

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n - \theta) &= \sqrt{n}(g(\bar{X}_n) - g(\theta^{-1} - 1)) \\ &\rightarrow_d (0, \theta^2(1 - \theta)) \end{aligned}$$

en appliquant la Delta-méthode avec $g(s) = 1/(1 + s)$ et $g'(x) = -1/(1 + x)^2$.

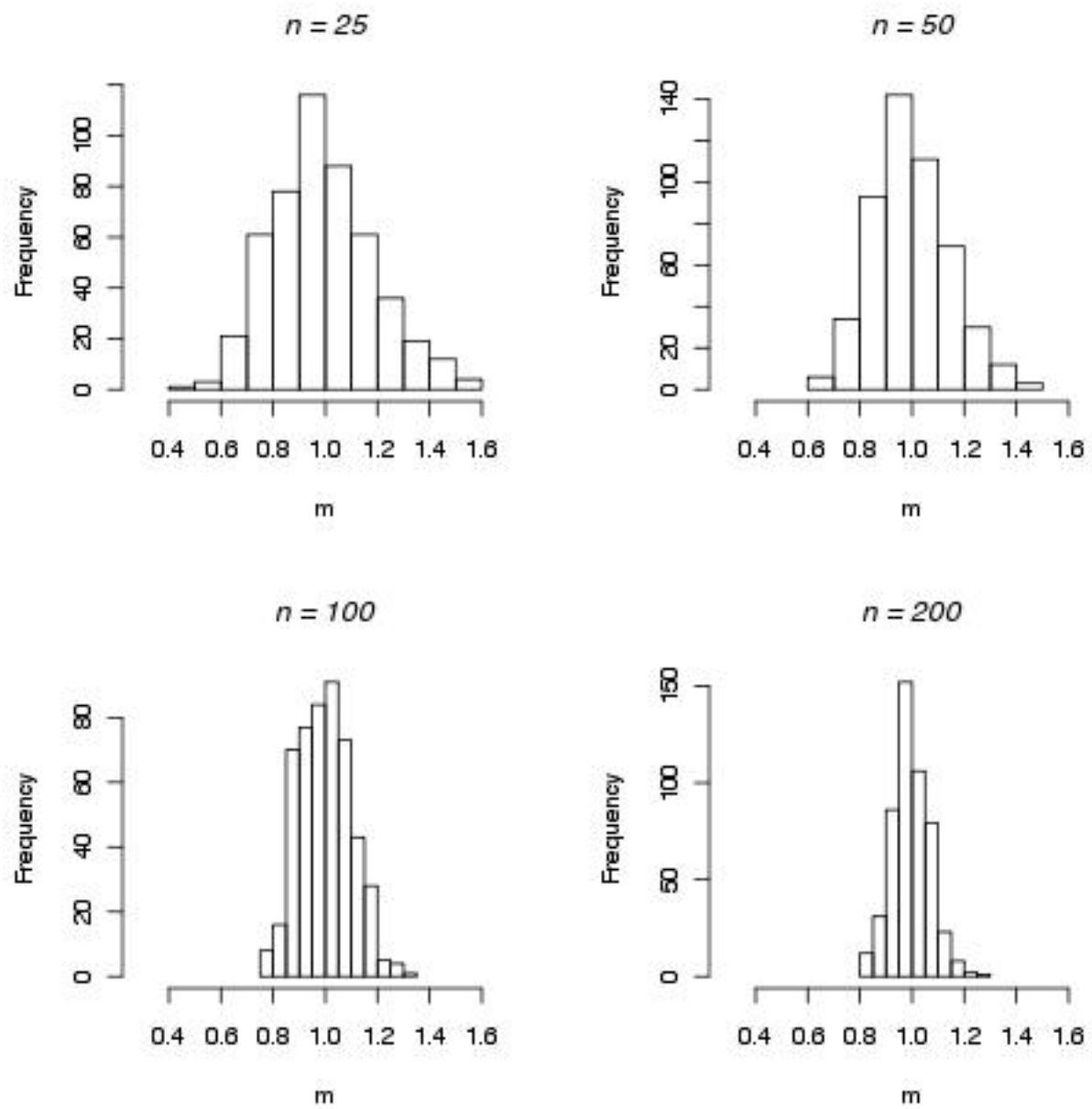


FIG. 3.1 – TCL pour l'EMV de la moyenne d'une variable aléatoire de loi exponentielle de paramètre 1.

Exemple - Illustration du TCL - Considérons une suite de variables aléatoires i.i.d. X_1, \dots, X_n de loi exponentielle. On estime la moyenne de X_i par l'estimateur empirique $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ pour différentes valeurs de n . Cet estimateur est l'estimateur du maximum de vraisemblance (voir TD). Puis on trace l'histogramme de $\bar{x}_{n,1}, \dots, \bar{x}_{n,B}$ pour $B = 500$.

3.2.2 Consistance

Il existe des résultats de consistance asymptotique pour le maximum de vraisemblance global, mais sous des hypothèses restrictives et difficilement vérifiables. On s'intéresse plutôt aux maxima locaux, plus précisément aux solutions de

$$\frac{\partial \ln \mathcal{L}(\theta; x_1, \dots, x_n)}{\partial \theta} = 0$$

Théorème 4 - Soient X_1, \dots, X_n des variables aléatoires i.i.d. suivant le modèle paramétrique (E^n, P_θ) avec $\theta \in \Theta$ ouvert. On suppose que le modèle est identifiable (i.e. $\theta \mapsto P_\theta$ est injective). Alors il existe une suite $\{\hat{\theta}_n\}_{n=1}^\infty$ solution de

$$\frac{\partial \ln \mathcal{L}(\theta; x_1, \dots, x_n)}{\partial \theta_k} = 0 \quad k = 1, \dots, p$$

telle que $\hat{\theta}_n \rightarrow \theta$

Remarque - Si le système d'équations

$$\frac{\partial \ln \mathcal{L}(\theta; x_1, \dots, x_n)}{\partial \theta_k} = 0 \quad k = 1, \dots, p$$

admet une unique solution alors c'est forcément l'EMV. S'il y a plusieurs solutions, le théorème ne précise pas laquelle choisir ; la suite de solutions qui converge peut ne pas correspondre au maximum global mais peut être un maximum local.

Preuve - Nous proposons une preuve pour $p = 1$ c'est à dire $\theta \in \mathbb{R}$. On s'intéresse aux solutions de

$$\frac{\partial \ln \mathcal{L}(\theta; x_1, \dots, x_n)}{\partial \theta} = 0$$

La suite X_1, \dots, X_n étant constituée de variables aléatoires i.i.d. ceci revient à étudier

$$\frac{\sum_{i=1}^n \partial \ln \mathcal{L}(\theta; x_i)}{\partial \theta} = 0$$

On note θ_0 la vraie valeur du paramètre θ (inconnu). Résoudre l'équation précédente revient à résoudre

$$\frac{\partial}{\partial \theta} \frac{1}{n} \ln \left(\frac{\mathcal{L}(\theta; X_i)}{\mathcal{L}(\theta_0; X_i)} \right) = 0$$

On note

$$S_n(\theta) = \frac{1}{n} \sum_{i=1}^n \ln \left(\frac{\mathcal{L}(\theta; X_i)}{\mathcal{L}(\theta_0; X_i)} \right)$$

On va montrer qu'il existe θ_0^- et θ_0^+ autour de θ distants de $\epsilon > 0$ tels que

$$S_n(\theta_0^-) < 0, S_n(\theta_0^+) < 0, S_n(\theta_0) = 0$$

donc il existe un maximum $\hat{\theta}_n$ dans $]\theta_0^-, \theta_0^+[$ vérifiant $|\hat{\theta}_n - \theta_0| < \epsilon$.

Si

$$\ln \left(\frac{\mathcal{L}(\theta; X_i)}{\mathcal{L}(\theta_0; X_i)} \right) \in L^1(P_{\theta_0})$$

alors par la loi forte des grands nombres $S_n(\theta)$ tend presque sûrement vers

$$E_{\theta_0} \left[\ln \left(\frac{\mathcal{L}(\theta; X)}{\mathcal{L}(\theta_0; X)} \right) \right].$$

On montre que la limite est strictement négative si $\theta \neq \theta_0$.

Inégalité de Jensen - Si ψ est strictement convexe et $Y \in L^1(P)$, $\psi(E(Y)) \leq E(\psi(Y))$. Et si on a égalité, $Y = \text{cte}$ p.s..

La fonction \ln est strictement concave, donc $-\ln$ est strictement convexe. Donc par Jensen

$$E_{\theta_0} \left[\ln \left(\frac{\mathcal{L}(\theta; X)}{\mathcal{L}(\theta_0; X)} \right) \right] \leq \ln \left(E_{\theta_0} \left[\left(\frac{\mathcal{L}(\theta; X)}{\mathcal{L}(\theta_0; X)} \right) \right] \right)$$

or

$$E_{\theta_0} \left[\left(\frac{\mathcal{L}(\theta; X)}{\mathcal{L}(\theta_0; X)} \right) \right] = \int \frac{\mathcal{L}(\theta; x)}{\mathcal{L}(\theta_0; x)} \mathcal{L}(\theta_0; x) d\mu(x) = 1$$

car $x \mapsto \mathcal{L}(\theta; x)$ est une densité de probabilité par rapport à la mesure μ . Donc

$$E_{\theta_0} \left[\ln \left(\frac{\mathcal{L}(\theta; X)}{\mathcal{L}(\theta_0; X)} \right) \right] \leq 0$$

Quand on a égalité, c'est équivalent à

$$\ln \left(\frac{\mathcal{L}(\theta; x)}{\mathcal{L}(\theta_0; x)} \right) = \text{cte p.s.}$$

ou encore à

$$\frac{\mathcal{L}(\theta; x)}{\mathcal{L}(\theta_0; x)} = \text{cte p.s.}$$

On en déduit que $\mathcal{L}(\theta; x) = \mathcal{L}(\theta_0; x)$ car $\mathcal{L}(\theta; x)$ et $\mathcal{L}(\theta_0; x)$ sont des densités de probabilité. On a alors $\theta = \theta_0$ par l'hypothèse d'identifiabilité.

Si $\theta \neq \theta_0$,

$$E_{\theta_0} \left[\ln \left(\frac{\mathcal{L}(\theta; X)}{\mathcal{L}(\theta_0; X)} \right) \right] < 0$$

donc pour tout $\theta \neq \theta_0$ et pour tout $\omega \in \Omega_\theta$ avec $P(\Omega_\theta) = 1$, il existe un n_0 tel que pour $n \geq n_0$ on a $S_n(\theta, \omega) < 0$ p.s.

On considère les ensembles $\Theta_0 = \{\theta \in \Theta, \theta = \theta_0 \pm \frac{1}{k}, k \in \mathbb{N}^*\}$ et $\Omega = \cap_{\theta \in \Theta_0} \Omega_\theta$ qui est mesurable car Θ_0 est dénombrable. On remarque que Ω est fixé quelque soit θ . Soit $\omega \in \Omega$, soit $\epsilon > 0$ on choisit $\theta_0^- = \theta_0 - \frac{1}{k}$ et $\theta_0^+ = \theta_0 + \frac{1}{k}$ tels que $|\theta_0^+ - \theta_0^-| < \epsilon$. On a $S_n(\theta_0^-, \omega) < 0$, $S_n(\theta_0^+, \omega) < 0$ et $S_n(\theta_0, \omega) = 0$. Donc il existe $\hat{\theta}_n \in]\theta_0^-, \theta_0^+[$ tel que

$$\frac{\partial S_n(\theta, \omega)}{\partial \theta} \Big|_{\theta=\hat{\theta}_n} = 0 \quad (3.1)$$

Finalement, $\forall \omega \in \Omega$, avec $P(\Omega) = 1$, pour tout $\epsilon > 0$, $\exists n_0$ tel que $\forall n \geq n_0$, on peut trouver $\hat{\theta}_n$ vérifiant l'équation (3.1), c'est à dire $\hat{\theta}_n$ tend presque sûrement vers θ_0 . \diamond

3.3 Etimateur efficace

On étudie l'optimalité des estimateurs sans biais. Parmi les estimateurs sans biais, on cherche ceux dont la variance est minimale (ou de manière équivalente l'EQM est minimale). On suppose que l'on est dans un modèle paramétrique (E, P_θ) et qu'on veut estimer une fonction quelconque $g(\theta)$ de θ . Dans un premier temps, on se restreint au cas $\theta \in \mathbb{R}$.

Théorème 5 - Soit $\mathbf{X} = (X_1, \dots, X_n)$ défini sur $(E, P_\theta)_{\theta \in \Theta}$ de densité jointe $f(\mathbf{x}; \theta)$ et vérifiant les hypothèses (1) L'ensemble $A = \{x \in E : f(\mathbf{x}; \theta) > 0\}$ ne dépend pas de θ

(2) Pour tout $\mathbf{x} \in A$, $f(\mathbf{x}; \theta)$ est différentiable par rapport à θ .

(3) $E_\theta[\frac{\partial \ln f(\mathbf{X}; \theta)}{\partial \theta}] = 0$

Soit $T \in L^1(P_\theta)$ une statistique telle que $g(\theta) = E_\theta[T(\mathbf{X})]$ différentiable,

$$g'(\theta) = E_\theta \left[T(\mathbf{X}) \frac{\partial \ln f(\mathbf{X}; \theta)}{\partial \theta} \right]$$

Alors,

$$Var_\theta(T(\mathbf{X})) \geq \frac{g'(\theta)^2}{I_n(\theta)}$$

Le minorant est appelé *borne de Cramer-Rao*. La borne de Cramer-Rao est atteinte si

$$Var_\theta(T) = \frac{(g'(\theta))^2}{I(\theta)}$$

donc si

$$Var_\theta(T) = \frac{Cov_\theta^2(T, U_\theta)}{I(\theta)}$$

avec $U_\theta = \frac{\partial \ln f(\mathbf{X}; \theta)}{\partial \theta}$. Or ceci est vérifié si et seulement si U_θ est une fonction affine de T ; c'est à dire si, avec probabilité 1,

$$U_\theta = C(\theta)T + \delta(\theta)$$

pour tout $\mathbf{x} \in A$. Ainsi

$$\ln f(\mathbf{x}; \theta) = C^T(\theta)T(\mathbf{x}) + \delta^T(\theta) + S((x))$$

Autrement dit, $Var_\theta(T)$ n'atteint la borne de Cramér-Rao si et seulement si la fonction de densité de (X_1, \dots, X_n) appartient à la famille exponentielle à un paramètre. En particulier, T doit être une statistique exhaustive.

Preuve du théorème 5 - Par l'inégalité de Cauchy-Schwarz,

$$Var_\theta(T) \geq \frac{Cov(T, U_\theta)}{Var_\theta(U_\theta)}$$

Comme $E_\theta[U_\theta] = 0$, on a $Var_\theta(U_\theta) = I(\theta)$. De plus,

$$\begin{aligned} Cov_\theta(T, U_\theta) &= E_\theta(TU_\theta) - E_\theta(T)E_\theta(U_\theta) \\ &= E_\theta(TU_\theta) \\ &= \frac{d}{d\theta} E_\theta(T) = g'(\theta) \end{aligned}$$

ce qui conclut la preuve. \diamond

Définition 10 - On dit qu'un estimateur est efficace s'il est sans biais et que sa variance atteint la borne de Cramér-Rao.

Définition 11 - Soient T_n et T'_n deux estimateurs sans biais de $g(\theta)$. T'_n est dit plus efficace que T_n s'il est préférable au sens de la variance :

$$Var_\theta(T'_n) \leq Var_\theta(T_n) \text{ pour tout } \theta \in \Theta$$

On dit que l'estimateur sans biais T'_n est uniformément plus efficace si il est plus efficace que tous les estimateurs sans biais. On dit aussi qu'il est de variance minimale.

Définition 12 Une statistique complète est une statistique exhaustive minimale ne contient plus que de l'information utile à l'estimation de du paramètre θ . Ceci est formalisé ainsi : T est une statistique complète si pour toute fonction intégrale g :

$$E_\theta(g(T)) = 0 \quad \forall \theta \in \Theta \text{ implique que } g(T) = 0 \text{ p.s.}$$

On rappelle que pour deux matrices A et B on a $A \leq B$ si et seulement si $B - A$ est une matrice symétrique positive.

Le critère d'efficacité n'a de sens que pour discriminer les estimateurs sans biais.

Théorème 6 (Théorème de Lehmann-Scheffé). Si T_n est un estimateur sans biais de $g(\theta)$ et si S_n est une statistique exhaustive et complète, alors l'unique estimateur de $g(\theta)$ sans biais uniformément de variance minimale est $T'_n = E_\theta(T_n | S_n)$.

Théorème 7 *Pour toute statistique T , on a*

$$I_T(\theta) \leq I_n(\theta)$$

*et $I_T(\theta) = I_n(\theta)$ si et seulement si T est exhaustive,
 $I_T(\theta) = 0$ si et seulement si T est libre (c'est à dire que sa loi ne dépend pas de θ).*

Remarques

- Un estimateur efficace est de variance minimale.
- Un estimateur peut être sans biais, de variance minimale, mais ne pas atteindre la borne de Cramer-Rao, donc ne pas être efficace. Dans ce cas-là, la borne Cramer-Rao est "trop petite" pour être atteinte.

Exemple - modèle de Poisson (voir TD).

Chapitre 4

Tests d'hypothèses

4.1 Principe et définitions

On se place dans un modèle paramétrique $\mathbf{X} = (X_1, \dots, X_n)$ i.i.d suivant $(P_\theta, \theta \in \Theta)$. Supposons que $\Theta = \Theta_0 \cup \Theta_1$ où Θ_0 et Θ_1 sont deux ensembles disjoints. Connaissant une réalisation (x_1, \dots, x_n) de \mathbf{X} , on voudrait décider si θ est dans Θ_0 ou Θ_1 . En pratique, on choisira toujours pour Θ_0 le plus petit des deux sous-espaces Θ_0, Θ_1 . Ainsi, $\theta \in \Theta_0$ correspond à la version la plus simple du modèle.

4.1.1 Hypothèses de test

On pose une *hypothèse nulle* notée

$$H_0 : \theta \in \Theta_0$$

contre une *hypothèse alternative* notée

$$H_1 : \theta \in \Theta_1$$

Exemple - Supposons que X_1, \dots, X_m et Y_1, \dots, Y_n sont des variables indépendantes telles que $X_i \sim \mathcal{N}(\mu_1, \sigma_2)$ et $Y_i \sim \mathcal{N}(\mu_2, \sigma_2)$. L'espace des paramètres est

$$\Theta = \{(\mu_1, \mu_2, \sigma) : -\infty < \mu_1, \mu_2 < \infty, \sigma > 0\}$$

Dans les applications, on cherche à déterminer si les X_i et les Y_i ont la même distribution (c'est à dire si $\mu_1 = \mu_2$). Par exemple, on administre 2 somminifères différents à 2 groupes de patients et on cherche à savoir si la durée moyenne du sommeil est la même dans les deux groupes. On définit alors Θ_0 par

$$\Theta_0 = \{(\mu_1, \mu_2, \sigma) : -\infty < \mu_1 = \mu_2 < \infty, \sigma > 0\}$$

et Θ_1 est le complémentaire de Θ_0 . On remarque ici de Θ et de dimension 3 alors que Θ_0 n'est plus que de dimension 2.

4.1.2 Principe général

On décide que θ est dans Θ_0 ou Θ_1 à l'aide des observations (x_1, \dots, x_n) . Pour cela on cherche une règle de décision qui prend la forme suivante.

- Si $S(x_1, \dots, x_n) \in \text{RC}$ alors on rejette H_0
- Si $S(x_1, \dots, x_n) \notin \text{RC}$ alors on ne rejette pas H_0

où S est une *statistique* (ou fonction) de *test* et RC est une *région critique*. Le plus souvent, S est une fonction d'un estimateur de θ .

Exemple - Supposons que X_1, \dots, X_n sont des variables i.i.d telles que $E(X_i) = \theta$. On souhaite tester

$$H_0 : \theta = 0 \text{ contre } H_1 : \theta \neq 0$$

On peut alors choisir $S(X_1, \dots, X_n) = \bar{X}$. Et on rejette H_0 si \bar{x} est éloigné de 0 c'est à dire si $\text{RC} =]\infty, -c] \cup [c, \infty[$ avec c une constante positive. On verra plus loin comment on détermine c .

4.1.3 Erreurs et puissance

Il est peu probable qu'une règle de décision soit parfaite. Ainsi, quand on définit une règle de décision, on doit regarder qu'elle est la probabilité de faire des erreurs quand on prend l'une ou l'autre décision en fonction de la valeur de $\theta \in \Theta$. On peut faire deux types d'erreur

- On fait une *erreur de première espèce*¹ quand on rejette H_0 à tort c'est à dire alors que $\theta \in \Theta_0$. On peut associer une probabilité à cette erreur :

$$P(S(X_1, \dots, X_n) \in \text{RC} | \theta \in \Theta_0)$$

On parle aussi parfois de *risque de première espèce*.

- On fait une *erreur de seconde espèce*² quand on accepte H_0 à tort c'est à dire alors que $\theta \in \Theta_1$. La probabilité associée est alors

$$P(S(X_1, \dots, X_n) \notin \text{RC} | \theta \in \Theta_1)$$

On pourrait être tenté de chercher à définir des statistiques de tests telles que ces deux erreurs soient uniformément petites pour tout $\theta \in \Theta$. On verra que c'est généralement impossible.

Définition 13 - On appelle *fonction de risque de première espèce* la fonction

$$\begin{aligned} \alpha & : \Theta_0 \rightarrow [0, 1] \\ \theta & \mapsto P_\theta(S(X_1, \dots, X_n) \in \text{RC}) \end{aligned}$$

On appelle *fonction de risque de seconde espèce* la fonction

$$\begin{aligned} \beta & : \Theta_1 \rightarrow [0, 1] \\ \theta & \mapsto P_\theta(S(X_1, \dots, X_n) \notin \text{RC}) \end{aligned}$$

¹ en anglais : type I error

² en anglais : type II error

Ainsi, α représente la probabilité de se tromper quand on est sous H_0 . On note parfois abusivement $\alpha(\theta) = P_\theta(H_1|H_0)$ et $\beta(\theta) = P_\theta(H_0|H_1)$.

Définition 14 - On appelle *fonction puissance* la fonction

$$\begin{aligned}\Pi &: \Theta_1 \rightarrow [0, 1] \\ \theta &\mapsto P_\theta(S(X_1, \dots, X_n) \in RC)\end{aligned}$$

Propriété - On a pour tout $\theta \in \Theta_1$, $\Pi(\theta) = 1 - \beta(\theta)$

Définition 15 - $\alpha = \sup_{\theta \in \Theta_0} \alpha(\theta)$ est appelé *niveau du test*. C'est le *risque de première espèce maximal*.

4.1.4 Principe de Neyman

On voudrait trouver des procédures de test qui minimisent les 2 erreurs. Or il est facile de voir, que le plus souvent, si α diminue alors β augmente (voir par exemple le graphique de la puissance du test du signe ci-dessous).

Le principe de Neyman consiste à fixer le niveau α à une valeur petite (typiquement 5% ou 1 %) et à chercher une région critique qui minimise $\beta(\theta)$ à α fixé.

En pratique :

1. On fixe le niveau α .
2. On en déduit une région critique : $RC(\alpha)$. Si plusieurs régions sont possibles, on choisit celle qui minimise $\beta(\theta)$.
3. On conclut : si $S(x_1, \dots, x_n) \in RC(\alpha)$, on rejette H_0 .

On utilise parfois une alternative pour conclure. Au lieu de fixer α et de comparer la valeur de la statistique de test observée à la région critique $RC(\alpha)$, on estime un *degré de significativité* ou (p-value) :

$$\hat{\alpha}(X_1, \dots, X_n) = \inf\{\alpha \text{ tel que } S(X_1, \dots, X_n) \in RC(\alpha)\}$$

Ainsi le degré de significativité est le niveau le plus faible qu'on peut choisir pour conclure au rejet de H_0 . On dit parfois que c'est l'erreur que l'on fait quand on rejette H_0 . Concrètement, on compare $\hat{\alpha}$ au niveau α fixé.

4.2 Exemples

4.2.1 Test du signe

Un aquaculteur a 114 poissons d'une certaine espèce dans un de ses bassins. On extrait un échantillon de 12 poissons afin de vérifier l'hypothèse selon laquelle que la médiane de la longueur des poissons est de 220 mm. On observe les longueurs suivantes :

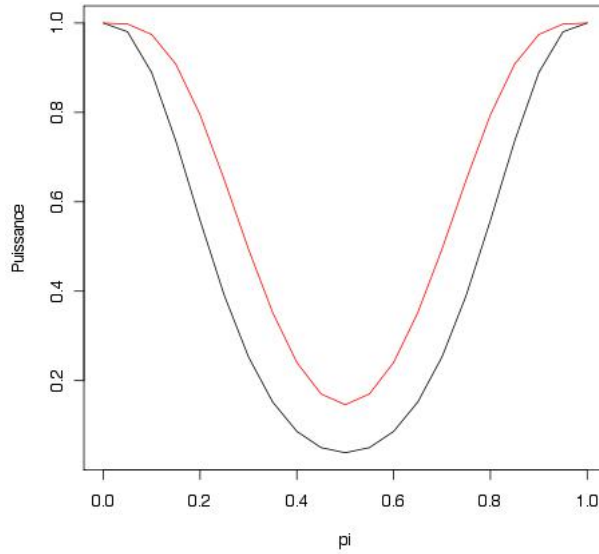
126 142 156 228 245 246 370 419 433 454 478 503

1. Hypothèses de test : en notant μ la médiane, $H_0 : \mu = 220$ contre $H_1 : \mu \neq 220$.
2. Risque de première espèce : on fixe, arbitrairement, $\alpha = 0.05$.
3. Statistique de test. On choisit ici un test naïf appelé *test du signe*. La statistique de test est construite de la façon suivante. Si la médiane est 220, il est également probable pour chaque poisson sélectionné d'être plus ou moins long que 220 mm. Puis on calcule S égale aux nombre d'individus plus long que 220. (Implicitement, on associe à chaque individu un signe - si sa longueur est inférieure à 220 et un signe + sinon ; S est alors la somme des signes +).
4. Loi de la statistique de test. Il est facile de voir que la loi de la statistique de test est une loi binomiale de paramètres $n = 12$ et $\pi = 1/2$.
5. Région critique. On remarque aisément que l'on va rejeter H_0 si S est trop grande ou trop petite (dominance de signes +, ou de signes -) ; la région critique est donc de la forme $\{s \text{ tels que } s \notin [s_{\text{inf}}, s_{\text{sup}}]\}$. Les bornes s_{inf} et s_{sup} sont déterminées à l'aide de la loi binomiale de telle sorte que

$$P_{H_0}(S < s_{\text{inf}}) = \frac{\alpha}{2} \text{ et } P_{H_0}(S > s_{\text{sup}}) = \frac{\alpha}{2}$$

En s'aidant de la table 4.1, on trouve que $s_{\text{inf}} = 2$ et $s_{\text{sup}} = 10$ (car $P(x = 0) + P(x = 1) + P(x = 2) < .025$).

6. Puissance du test. On ne peut calculer $P_{H_1}(S \notin [s_{\text{inf}}, s_{\text{sup}}])$ que si on choisit une alternative ponctuel pour H_1 c'est à dire si on fixe la valeur de la médiane sous H_1 . En pratique, on dispose généralement pas d'une telle information. On peut alors regarder comment varie la puissance pour différentes valeurs de la médiane. Par exemple, le graphique ci-dessous montre la puissance du test du signe quand π varie. La courbe en noir correspond à un risque de première espèce de 5% et la courbe en rouge à un risque de première espèce de 20%.



r	0	1	2	3	4	5	6
P	0.000	0.003	0.016	0.054	0.121	0.193	0.226
		7	8	9	10	11	12
		0.193	0.121	0.054	0.016	0.003	0.000

TAB. 4.1 – Probabilités binomiales $P(S = k)$, $n = 12$, $\pi = \frac{1}{2}$

4.2.2 Test pour la moyenne d'une loi de Gauss

Un contrôle anti-dopage a été effectué sur 16 sportifs. On a mesuré la variable X de moyenne m , qui est le taux (dans le sang) d'une certaine substance interdite. Voici les données obtenues :

0.35	0.4	0.65	0.27	0.14	0.59	0.73	0.13
0.24	0.48	0.12	0.70	0.21	0.13	0.74	0.18

La variable X est supposée gaussienne et de variance $\sigma^2 = 0.04$. On veut tester, au niveau 5% l'hypothèse selon laquelle le taux moyen dans le sang de la population des sportifs est égal à 0.4.

On pose des hypothèses de test unilatérales :

$$H_0 : m = m_0 = 0.4 \text{ contre } H_1 : m > 0.4$$

La statistique de test est la moyenne empirique. Si on note X_1, \dots, X_n l'échantillon de variables aléatoires de même loi que X , la moyenne empirique est donnée par

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Intuitivement, on comprend bien qu'on va rejeter H_0 si $\bar{X}_n - m_0$ est trop grande en valeur absolue c'est à dire si la moyenne empirique est trop éloignée de la moyenne sous H_0 .

Sous H_0 , $Z = \frac{\bar{X}_n - m_0}{\sigma/\sqrt{n}}$ suit une loi de Gauss de moyenne 0 et de variance 1. D'autre part, d'après la remarque faite plus haut on comprend qu'on rejette H_0 si $|Z| > z_0$. Pour construire la région critique, on cherche donc z_0 tel que

$$P(|Z| > z_0) = \alpha$$

soit encore

$$P(Z > z_0 \text{ ou } Z < -z_0) = P(Z > z_0) + P(Z < -z_0) = \alpha$$

or on a par symétrie de la loi de Gauss de moyenne 0 et de variance 1

$$P(Z > z_0) = P(Z < -z_0) = \Phi(-z_0) = 1 - \Phi(z_0)$$

où on note Φ la fonction de répartition de la loi Gauss de moyenne 0 et de variance 1. Ainsi z_0 est tel que

$$1 - \Phi(z_0) = \alpha/2$$

ce qui s'écrit encore

$$z_0 = \Phi^{-1}(1 - \alpha/2)$$

D'après la table de la fonction de répartition inverse de la loi normale, on en déduit que $z_0 = 1.96$ car $\alpha = 0.05$.

Finalement, on rejette donc H_0 si

$$|\bar{X}_n - m_0| > 1.96 \frac{\sigma}{\sqrt{n}}$$

.

Remarques

- Lorsque le nombre d'observations n est grand (supérieur à 30), d'après le théorème de limite centrale on a que la statistique de test

$$Z = \frac{\bar{X} - m_0}{\sigma/\sqrt{n}}$$

suit approximativement une loi de Gauss quelque soit la loi de la variable X considérée.

Si la variance est inconnue

Dans le cas où la variance n'est pas connue, on doit l'estimer en utilisant les observations. Et la statistique de test du test de la moyenne donnée par

$$Z = \frac{\bar{X} - m_0}{s/\sqrt{n}}$$

Elle ne suit plus une loi de Gauss car le dénominateur n'est plus une constante mais une réalisation de l'estimateur de la variance de la variable X . L'écart-type s est obtenu par

$$s^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2$$

Par construction, S^2 suit une loi du χ^2 . Y est donc une v.a. suivant une de Student à $(n-1)$ degrés de libertés. Et on utilise alors une table de la loi de Student pour conclure le test.

Remarque : Lorsque le nombre d'observations n est grand (supérieur à 30), on peut utiliser le théorème de limite centrale pour approcher la loi de la statistique Z .

Calcul de la puissance du test

Dans le cas d'un test de Student, on peut calculer la puissance du test si on on peut donner une valeur de la moyenne sous l'hypothèse alternative.

$$H_0 : m = m_0 \text{ contre } H_1 : m = m_1$$

La puissance est définie par

$$\mathcal{P} = P(\text{rejeter } H_0 | H_0 \text{ est fausse})$$

Ainsi la puissance est la probabilité de la la région de rejet de H_0 sous la loi de H_1 .

$$\begin{aligned} \mathcal{P} &= P\left(Z > z_0 \mid \frac{Z - m_1}{\sigma/\sqrt{n}} \text{ suit une loi } \mathcal{N}(0, 1)\right) \\ &= P\left(\tilde{Z} > \frac{z_0 - m_1}{\sigma/\sqrt{n}}\right) \\ &= 1 - \Phi\left(\frac{z_0 - m_1}{\sigma/\sqrt{n}}\right) \end{aligned}$$

4.3 Principe de Neyman et optimalité

4.3.1 Test randomisé

Affinons le test du signe réalisé plus haut. En effet, nous avons remarqué qu'on ne peut pas toujours atteindre exactement le niveau α fixé.

Dans l'exemple des poissons, si on pose des hypothèses de test unilatérales, correspondants à la question : "les poissons pêchés sont-ils trop petits (de longueur inférieure à 22 mm) ?",

$$H_0 : \mu = 22 \text{ contre } H_1 : \mu < 22$$

alors la région critique est de la forme $RC = \{S \leq c\}$ où c doit vérifier $P_{H_0}(S \leq c) = \alpha$. Or sous H_0 , la statistique de test suit une loi binomiale (définie sur un ensemble discret) et il n'existe pas de c qui permette d'obtenir l'égalité : $P(S \leq 2) = 0.019$ et $P(S \leq 3) = 0.073$.

Une solution consiste à randomiser le test, c'est à dire qu'on tire au hasard la solution du test avec une certaine probabilité. Dans l'exemple des poissons,

- si $s_{obs} \geq 4$ on ne refuse pas H_0 ;
- si $s_{obs} \leq 2$ on refuse H_0 ;
- si $s_{obs} = 3$, on tire au sort H_1 avec une probabilité $\gamma \in [0, 1]$. On choisit γ telle que

$$\begin{aligned}\alpha = P_{H_0}(\text{RC}) &= 0 \times P_{H_0}(S \geq 4) + \gamma \times P_{H_0}(S = 3) + 1 \times P_{H_0}(S \leq 2) \\ &= \gamma P_{H_0}(S = 3) + P_{H_0}(S \leq 2)\end{aligned}$$

d'où $\gamma = (0.05 - 0.019)/0.054 = 0.57$.

Donc, si $s_{obs} = 3$ on décide H_1 avec une probabilité de 57%.

Définition 16 Soit (X_1, \dots, X_n) à valeur dans E^n . Un test est une fonction aléatoire Ψ de $E^n \rightarrow [0, 1]$.

Interprétation - La fonction Ψ représente la probabilité de décider H_1 .

Si $\Psi(X_1, \dots, X_n) = 0$ on conclut à H_0 .

Si $\Psi(X_1, \dots, X_n) = 1$ on conclut à H_1 .

Si $\Psi(X_1, \dots, X_n) \in]0, 1[$ on tire au hasard la décision H_1 avec la probabilité $\Psi(X_1, \dots, X_n)$.

Lorsque Ψ est à valeurs dans $\{0, 1\}$, on parle de test pur, c'est à dire non randomisé. C'est le cas de la plupart des tests classiques. Par exemple quand on teste, pour une variable aléatoire de loi de Gauss de moyenne μ ,

$$H_0 : \mu = \mu_0 \text{ contre } H_1 : \mu \neq \mu_0$$

La région critique est

$$\text{RC} = \left\{ \sqrt{n} \left| \frac{\bar{x}_n - \mu_0}{s_n} \right| > F_{n-1}^{-1}(1 - \alpha/2) \right\}$$

au niveau α avec F_{n-1} la fonction de répartition de la loi de Student à $n - 1$ degrés de liberté. C'est un test pur pour lequel $\Psi(x_1, \dots, x_n) = \mathbb{I}_{\{\text{RC}\}}$.

Définition 17 Pour le test $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$,

- le risque de 1ère espèce est la fonction $\alpha(\theta) = E_\theta(\Psi(X_1, \dots, X_n))$, $\forall \theta \in \Theta_0$;
- le risque de 2nde espèce est la fonction $\beta(\theta) = E_\theta(1 - \Psi(X_1, \dots, X_n))$, $\forall \theta \in \Theta_1$;
- le niveau est $\alpha = \sup_{\theta \in \Theta_0} \alpha(\theta)$;
- la puissance du test est la fonction $\Pi(\theta) = 1 - \beta(\theta)$.

L'utilisation de tests randomisés permet de considérer des tests de niveau α pour tout $\alpha \in [0, 1]$. Ils existent d'après le lemme de Neyman-Pearson donné ci-dessous. On peut donc définir la notion de test le plus puissant parmi les tests de niveau α .

Définition 18 Un test associé à la fonction Ψ est un test uniformément plus puissant (UPP) au niveau α , si son niveau est inférieur ou égal à α et si pour tout test Ψ^* de niveau inférieur ou égal à α ,

$$\Pi_\theta(\Psi(X_1, \dots, X_n)) \geq \Pi_\theta(\Psi^*(X_1, \dots, X_n))$$

pour tout $\theta \in \Theta_1$.

4.3.2 Tests uniformément plus puissants

Le lemme de Neyman-Pearson est important car il suggère un principe pour trouver de "bons" tests au sens du compromis entre une puissance forte et une erreur de première espèce faible.

Théorème 8 - Lemme de Neyman-Pearson. Soit (X_1, \dots, X_n) un échantillon de vraisemblance $\mathcal{L}(\theta; X_1, \dots, X_n)$. Pour tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$, $\theta_0 \neq \theta_1$, pour tout $\alpha \in]0, 1[$ fixé, il existe $c > 0$ et $\gamma \in [0, 1[$ tels que le test

$$\Psi(X_1, \dots, X_n) = \begin{cases} 1 & \text{si } \frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} > c \\ \gamma & \text{si } \frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} = c \\ 0 & \text{si } \frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} < c \end{cases}$$

De plus ce test est uniformément plus puissant parmi les tests de niveau au plus α et c'est le seul. Ce test est appelé test de Neyman-Pearson associé à c et γ est déterminé par l'équation de test $E_{\theta_0}(\Psi(X_1, \dots, X_n)) = \alpha$. (γ n'est pas forcément unique.)

Preuve - 1. Nous montrons tout d'abord que le niveau est bien α .

$$\begin{aligned} E_{\theta_0}(\Psi(X_1, \dots, X_n)) &= 1 \times P_{\theta_0} \left(\frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} > c \right) \\ &+ \gamma \times P_{\theta_0} \left(\frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} = c \right) \\ &+ 0 \times P_{\theta_0} \left(\frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} < c \right) \end{aligned}$$

On veut trouver c et γ tels que $E_{\theta_0}(\Psi(X_1, \dots, X_n)) = \alpha$ pour tout $\alpha \in]0, 1[$. Soit F la fonction de répartition de $\frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} = \frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)}$

$$F(t) = P_{\theta_0} \left(\frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} \leq t \right).$$

- Si F est continue, alors il existe c tel que $F(c) = 1 - \alpha$. En prenant ce c et $\gamma = 0$, on a

$$\begin{aligned} &P_{\theta_0} \left(\frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} > c \right) + \gamma P_{\theta_0} \left(\frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} = c \right) \\ &= 1 - F(c) = 1 - (1 - \alpha) = \alpha \end{aligned}$$

- Si F n'est pas continue (cas discret), on note $c^+ = \min_t (F(t) \geq 1 - \alpha)$ et on a

$$F(c^+) = \lim_{x \rightarrow c^+, x > c^+} F(t), F(c^-) = \lim_{x \rightarrow c^+, x < c^+} F(t)$$

ainsi

$$F(c^+) - F(c^-) = \lim_{\substack{x \rightarrow c^+, x > c^+ \\ y \rightarrow c^+, y < c^+}} P \left(y < \frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} \leq x \right) = P \left(\frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} = c^+ \right)$$

Or on cherche c et γ tels que $E_{\theta_0}(\Psi(X_1, \dots, X_n)) = \alpha$. En choisissant $c = c^+$, on obtient

$$1 - F(c^+) + \gamma(F(c^+) - F(c^-)) = \alpha$$

ce qui est équivalent à

$$\gamma = \frac{\alpha + F(c^+) - 1}{F(c^+) - F(c^-)}$$

Il reste à vérifier que γ appartient à $[0, 1[$.

$$\gamma \geq 0 \Leftrightarrow \alpha - 1 + F(c^+) \geq 0 \Leftrightarrow F(c^+) \geq 1 - \alpha$$

ce qui est vrai par définition de c^+ . Par ailleurs,

$$\gamma < 1 \Leftrightarrow \alpha - 1 + F(c^+) < F(c^+) - F(c^-) \Leftrightarrow F(c^-) < 1 - \alpha$$

ce qui est vrai par définition.

2. Montrons maintenant que le test est UPP.

Soit Ψ^* un test de niveau au plus α . On considère l'intégrale

$$\int_{E^n} (\Psi(x_1, \dots, x_n) - \Psi^*(x_1, \dots, x_n)(\mathcal{L}(\theta_1) - c\mathcal{L}(\theta_0))) d\mu(x_1, \dots, x_n) \quad (4.1)$$

où μ est la mesure de référence par rapport à laquelle $\mathcal{L}(\theta_0)$ et $\mathcal{L}(\theta_1)$ sont définies. Cette intégrale est toujours positive car

- si $\mathcal{L}(\theta_1) - c\mathcal{L}(\theta_0) > 0$ alors $\Psi(X_1, \dots, X_n) = 1$ par définition de Ψ et, $\Psi(X_1, \dots, X_n) \geq \Psi^*(X_1, \dots, X_n)$;
- si $\mathcal{L}(\theta_1) - c\mathcal{L}(\theta_0) < 0$ alors $\Psi(X_1, \dots, X_n) = 0$ par définition de Ψ et, $\Psi(X_1, \dots, X_n) \leq \Psi^*(X_1, \dots, X_n)$;
- si $\mathcal{L}(\theta_1) - c\mathcal{L}(\theta_0) = 0$ alors l'intégrale est nulle.

On a donc

$$\begin{aligned} & \int_{E^n} \Psi(x_1, \dots, x_n) \mathcal{L}(\theta_1) d\mu(x_1, \dots, x_n) - c \int_{E^n} \Psi(x_1, \dots, x_n) \mathcal{L}(\theta_0) d\mu(x_1, \dots, x_n) \\ & - \int_{E^n} \Psi^*(x_1, \dots, x_n) \mathcal{L}(\theta_1) d\mu(x_1, \dots, x_n) + c \int_{E^n} \Psi^*(x_1, \dots, x_n) \mathcal{L}(\theta_0) d\mu(x_1, \dots, x_n) \geq 0 \end{aligned}$$

Ce qui s'écrit aussi

$$E_{\theta_1}(\Psi) - cE_{\theta_0}(\Psi) - E_{\theta_1}(\Psi^*) + cE_{\theta_0}(\Psi^*) \geq 0$$

ou encore

$$E_{\theta_1}(\Psi) - E_{\theta_1}(\Psi^*) \geq c(E_{\theta_0}(\Psi) - E_{\theta_0}(\Psi^*))$$

et on reconnaît que

- $E_{\theta_1}(\Psi)$: puissance de Ψ
- $E_{\theta_1}(\Psi^*)$: puissance de Ψ^*
- $E_{\theta_0}(\Psi)$: niveau de Ψ qui est égal à α

- $E_{\theta_0}(\Psi)$: niveau de Ψ^* qui est inférieur ou égal à α . On en déduit donc que $E_{\theta_1}(\Psi) - E_{\theta_1}(\Psi^*) \geq 0$ et que la puissance Ψ est supérieure à celle de tout test Ψ^* de niveau au plus α .

3. On montre enfin que c'est le seul test UPP, à γ près.

Soit Ψ^* un test UPP au niveau au plus α . On a donc

$$E_{\theta_1}(\Psi^*(X_1, \dots, X_n)) \geq E_{\theta_1}(\Psi(X_1, \dots, X_n)) \text{ car } \Psi^* \text{ est UPP,}$$

$$\text{et } E_{\theta_1}(\Psi(X_1, \dots, X_n)) \geq E_{\theta_1}(\Psi^*(X_1, \dots, X_n)) \text{ car } \Psi \text{ est UPP.}$$

Donc, ces deux tests sont de même puissance

$$E_{\theta_1}(\Psi(X_1, \dots, X_n)) = E_{\theta_1}(\Psi^*(X_1, \dots, X_n))$$

Reprenons l'intégrale (4.1) :

$$\int_{E^n} (\Psi(x_1, \dots, x_n) - \Psi^*(x_1, \dots, x_n)(\mathcal{L}(\theta_1) - c\mathcal{L}(\theta_0))) d\mu(x_1, \dots, x_n)$$

$$= E_{\theta_1}(\Psi) - cE_{\theta_0}(\Psi) - E_{\theta_1}(\Psi^*) + cE_{\theta_0}(\Psi^*)$$

On a noté que cette intégrale est positive donc $E_{\theta_0}(\Psi) - E_{\theta_0}(\Psi^*) \leq 0$. Or on a vu que $E_{\theta_0}(\Psi) = \alpha$ et $E_{\theta_0}(\Psi^*) \leq \alpha$. Donc $E_{\theta_0}(\Psi) = E_{\theta_0}(\Psi^*)$ et l'intégrale est nulle. Comme $\mathcal{L}(\theta_1) - c\mathcal{L}(\theta_0)$ est différent de 0, cela implique $Psi = Psi^*$, μ presque sûrement. Donc les tests coïncident (à γ près qui reste à déterminer). \diamond

Remarque - Dans le cas continu (μ est la mesure de Lebesgue), on retrouve un test pur de région critique

$$\text{RC} = \left\{ \frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} > c \right\}$$

En effet, dans ce cas,

$$E_{\theta_0}(\Psi(X)) = 1 \times P_{\theta_0} \left(\frac{\mathcal{L}(\theta_1; X)}{\mathcal{L}(\theta_0; X)} > c \right) + \gamma \times P_{\theta_0} \left(\frac{\mathcal{L}(\theta_1; X)}{\mathcal{L}(\theta_0; X)} = c \right)$$

or

$$P_{\theta_0} \left(\frac{\mathcal{L}(\theta_1; X)}{\mathcal{L}(\theta_0; X)} = c \right) = 0$$

car la loi est continue.

Pour résoudre $E_{\theta_0}(\Psi(X)) = \alpha$, on peut choisir γ quelconque : on prend $\gamma = 0$ et le test devient

$$\Psi(X_1, \dots, X_n) = \mathbb{I}_{\left\{ \frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} > c \right\}}$$

autrement dit, un test pur de région critique

$$\text{RC} = \left\{ \frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} > c \right\}$$

Pour déterminer c , on résoud $E_{\theta_0}(\Psi(X_1, \dots, X_n)) = \alpha$ ce qui est équivalent à résoudre

$$P_{\theta_0} \left(\frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} > c \right) = \alpha$$

Exemple 1 - Soient X_1, \dots, X_n des v.a.i.i.d de loi de Gauss de moyenne θ et de variance un. On teste

$$H_0 : \theta = \theta_0 \text{ contre } H_1 : \theta = \theta_1$$

à partir de (x_1, \dots, x_n) une réalisation de (X_1, \dots, X_n) . La vraisemblance s'écrit

$$\mathcal{L}(\theta; x_1, \dots, x_n) = \frac{1}{(\sqrt{2\pi})^n} \exp \left(-\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2 \right)$$

Ainsi, le rapport des vraisemblances est

$$\frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} = \exp \left(-\frac{1}{2} \sum_{i=1}^n (x_i - \theta_1)^2 + \frac{1}{2} \sum_{i=1}^n (x_i - \theta_0)^2 \right)$$

en passant au log, on obtient

$$\begin{aligned} \frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} > c &\Leftrightarrow -\frac{1}{2} \sum_{i=1}^n (x_i - \theta_1)^2 + \frac{1}{2} \sum_{i=1}^n (x_i - \theta_0)^2 > \log(c) \\ &\Leftrightarrow (\theta_1 - \theta_0) \sum_{i=1}^n x_i - \frac{n\theta_1^2}{2} + \frac{n\theta_0^2}{2} > \log(c) \\ &\Leftrightarrow (\theta_1 - \theta_0) \sum_{i=1}^n x_i > \log(c) + \frac{n\theta_1^2}{2} - \frac{n\theta_0^2}{2} \end{aligned}$$

Si $\theta_1 > \theta_0$, l'inégalité devient

$$\sum_{i=1}^n x_i > \frac{1}{\theta_1 - \theta_0} \left(\log(c) + \frac{n\theta_1^2}{2} - \frac{n\theta_0^2}{2} \right)$$

et si au contraire, $\theta_1 < \theta_0$, elle s'écrit

$$\sum_{i=1}^n x_i < \frac{1}{\theta_1 - \theta_0} \left(\log(c) + \frac{n\theta_1^2}{2} - \frac{n\theta_0^2}{2} \right)$$

Notons

$$c' = \frac{1}{\theta_1 - \theta_0} \left(\log(c) + \frac{n\theta_1^2}{2} - \frac{n\theta_0^2}{2} \right)$$

Le test de Neyman-Pearson se ramène, dans le cas $\theta_1 > \theta_0$ à

$$\Psi(x_1, \dots, x_n) = \begin{cases} 1 & \text{si } \sum_{i=1}^n x_i > c' \\ \gamma & \text{si } \sum_{i=1}^n x_i = c' \\ 0 & \text{si } \sum_{i=1}^n x_i < c' \end{cases}$$

où c' et γ sont déterminés par $E_{\theta_0}(\Psi(X_1, \dots, X_n)) = \alpha$ ce qui est équivalent à

$$1 \times P_{\theta_0} \left(\sum_{i=1}^n X_i > c' \right) + \gamma \times P_{\theta_0} \left(\sum_{i=1}^n X_i = c' \right) = \alpha$$

Ici $P_{\theta_0}(\sum_{i=1}^n X_i = c') = 0$ car la loi est continue. On choisit $\gamma = 0$ et c' est obtenu en résolvant l'équation

$$P_{\theta_0}(\sum_{i=1}^n X_i > c') = \alpha$$

Or, sous H_0 , $\sum_{i=1}^n X_i$ suit une loi de Gauss de moyenne $n\theta_0$ et de variance n . On peut donc écrire,

$$P_{\theta_0}\left(\frac{\sum_{i=1}^n X_i - n\theta_0}{\sqrt{(n)}} > \frac{c' - n\theta_0}{\sqrt{(n)}}\right) = \alpha$$

et on en déduit que

$$\frac{c' - n\theta_0}{\sqrt{(n)}} = \Phi^{-1}(1 - \alpha)$$

avec Φ la fonction de répartition de la loi de Gauss centrée et réduite. Autrement dit $\frac{c' - n\theta_0}{\sqrt{(n)}} = \Phi^{-1}(1 - \alpha)$ est le quantile d'ordre $1 - \alpha$ de la loi de Gauss de moyenne 0 et de variance 1 ; on note parfois $q(1 - \alpha)$. Finalement, on conclut que pour tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$ le test de Neyman-Pearson est un test pur ($\gamma = 0$) de région critique :

$$\text{RC} = \left\{ \sum_{i=1}^n x_i > \sqrt{n}\Phi^{-1}(1 - \alpha) + n\theta_0 \right\}$$

Faut-il ajouter l'exemple sur le paramètre d'une loi binomiale ?

4.4 Tests UPP pour les hypothèses composites

Dans la partie précédente, nous avons montré des résultats (constructifs) d'existence et d'unicité pour des tests dont les hypothèses sont des singletons.

4.4.1 Test unilatéral $H_0 : \theta \leq \theta_0$ contre $H_1 : \theta > \theta_1$

Pour le test unilatéral $H_0 : \theta \leq \theta_0$ contre $H_1 : \theta > \theta_0$, on peut construire un test UPP mais en se restreignant à certaines familles de lois.

Définition 19 La famille $(P_\theta)_{\theta \in \Theta}$ est à rapport de vraisemblance monotone s'il existe une statistique $U(X_1, \dots, X_n)$ telle que pour tout $\theta < \theta_1$, le rapport

$$\frac{\mathcal{L}(\theta_1; x_1, \dots, x_n)}{\mathcal{L}(\theta_0; x_1, \dots, x_n)} = h(U(x_1, \dots, x_n))$$

avec une fonction h une fonction strictement monotone.

Remarque - On peut toujours supposer que h est strictement croissante quitte à considérer $-U(x_1, \dots, x_n)$ à la place de $U(x_1, \dots, x_n)$.

Exemple - la famille des lois exponentielles.
Considérons le cas particulier

$$\mathcal{L}(x_1, \dots, x_n) = K(\theta) \exp(c(\theta)T(x_1, \dots, x_n))$$

ainsi pour tout $\theta_1 > \theta_0$,

$$\frac{\mathcal{L}(\theta_1; x_1, \dots, x_n)}{\mathcal{L}(\theta_0; x_1, \dots, x_n)} = \frac{K(\theta_1)}{K(\theta_0)} \exp((\theta_1 - \theta_0)T(x_1, \dots, x_n))$$

est une fonction croissante de $T(x_1, \dots, x_n)$. C'est donc une famille à rapport de vraisemblance monotone avec $U(x_1, \dots, x_n) = T(x_1, \dots, x_n)$.

Théorème 9 Soient X_1, \dots, X_n v.a.i.i.d. suivant $(P_\theta)_{\theta \in \Theta}$ où $(P_\theta)_{\theta \in \Theta}$ est une famille à rapport de vraisemblance monotone. Pour tester

$$H_0 : \theta \leq \theta_0 \text{ contre } H_1 : \theta > \theta_1$$

il existe un test UPP de niveau α de la forme

$$\Psi(X_1, \dots, X_n) = \begin{cases} 1 & \text{si } U(X_1, \dots, X_n) > c \\ \gamma & \text{si } U(X_1, \dots, X_n) = c \\ 0 & \text{si } U(X_1, \dots, X_n) < c \end{cases}$$

où U est la statistique de la définition 19 et γ et c sont définies par $E_{\theta_0}(\Psi(X_1, \dots, X_n)) = \alpha$.

Preuve - a. On vérifie d'abord que le test est de niveau α .

Soient les hypothèses $H_0 : \theta = \theta'$ contre $H_1 : \theta = \theta''$ avec $\theta' < \theta''$. On considère le test du théorème. Ce test est exactement le test de Neyman-Pearson pour tester $H_0 : \theta = \theta'$ contre $H_1 : \theta = \theta''$. Son niveau vaut $E_{\theta_0}(\Psi(X_1, \dots, X_n))$ que l'on note α' . On sait d'après le lemme de Neyman-Pearson que Ψ est UPP parmi tous les tests de niveau α' .

Soit maintenant le test $\psi(X_1, \dots, X_n) = \alpha'$ pour tout (X_1, \dots, X_n) . Pour tester $H_0 : \theta = \theta'$ contre $H_1 : \theta = \theta''$, $E_{\theta}(\psi(X_1, \dots, X_n)) = \alpha'$ donc ψ est de niveau α' et Ψ est plus puissant que ψ . Donc

$$E_{\theta''}(\Psi(X_1, \dots, X_n)) \geq E_{\theta''}(\psi(X_1, \dots, X_n)) = \alpha'$$

Ainsi pour tout $\theta' < \theta''$, $\alpha' = E_{\theta'}(\Psi(X_1, \dots, X_n)) \leq E_{\theta''}(\psi(X_1, \dots, X_n))$.

En particulier pour $\theta'' = \theta_0$, pour tout $\theta' < \theta_0$, on a

$$E_{\theta'}(\Psi(X_1, \dots, X_n)) \leq E_{\theta_0}(\Psi(X_1, \dots, X_n))$$

donc $E_{\theta'}(\Psi(X_1, \dots, X_n)) = \alpha$ et le niveau de Ψ est bien α .

b. On montre que Ψ est UPP.

Le test de Neyman-Pearson pour tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$ où $\theta_1 > \theta_0$ est exactement Ψ , sa forme ne dépend pas de θ_1 . Ce test est le plus puissant pour tester $\theta = \theta_0$ contre $\theta = \theta_1$ d'après le lemme de Neyman-Pearson. Si Ψ n'était pas UPP pour tester $\theta \leq \theta_0$ contre $\theta > \theta_0$ alors il existerait un test $\Psi^{(\alpha)}$ plus puissant que Ψ au moins

pour un $\theta_1 > \theta_0$ c'est à dire $E_{\theta_1}(\Psi^{(\alpha)}(X_1, \dots, X_n)) \geq E_{\theta_1}(\Psi(X_1, \dots, X_n))$. Ce qui est impossible puisque Ψ est UPP pour tester $\theta = \theta_0$ contre $\theta = \theta_1$. \diamond

Exemple - Loi de Gauss de moyenne inconnue.

Soient X_1, \dots, X_n v.a.i.i.d. suivant une loi de Gauss de moyenne θ et de variance 1. On veut tester

$$H_0 : \theta \leq \theta_0 \text{ contre } H_1 : \theta > \theta_0$$

La vraisemblance est

$$\mathcal{L}(\theta; X_1, \dots, X_n) = \frac{1}{(\sqrt{1\pi})^n} \exp\left(-\frac{1}{2} \sum_{i=1}^n (X_i - \theta)^2\right)$$

Il faut montrer que le modèle est à rapport de vraisemblance monotone.

On peut soit montrer que la loi appartient à la famille exponentielle, soit le montrer directement. Considérons $\theta' < \theta''$,

$$\begin{aligned} \frac{\mathcal{L}(\theta''; x_1, \dots, x_n)}{\mathcal{L}(\theta'; x_1, \dots, x_n)} &= \exp\left(-\frac{1}{2} \sum_{i=1}^n (x_i - \theta'')^2 + \frac{1}{2} \sum_{i=1}^n (x_i - \theta')^2\right) \\ &= \exp\left((\theta'' - \theta') \sum_{i=1}^n x_i + \frac{n}{2}((\theta')^2 - (\theta'')^2)\right) \end{aligned}$$

C'est une fonction monotone de $U(x_1, \dots, x_n) = \sum_{i=1}^n x_i$ car $\theta'' > \theta'$. On peut appliquer le théorème précédent pour conclure que le test UPP au niveau α pour tester $H_0 : \theta \leq \theta_0$ contre $H_1 : \theta > \theta_0$ est

$$\Psi(X_1, \dots, X_n) = \begin{cases} 1 & \text{si } \sum_{i=1}^n X_i > c \\ \gamma & \text{si } \sum_{i=1}^n X_i = c \\ 0 & \text{si } \sum_{i=1}^n X_i < c \end{cases}$$

où γ et c sont définies par $E_{\theta_0}(\Psi(X_1, \dots, X_n)) = \alpha$. Comme la loi est absolument continue par rapport à la mesure de Lebesgue, on peut choisir $\gamma = 0$ et le test se ramène à un test pour la région critique

$$\text{RC} = \left\{ \sum_{i=1}^n x_i > c \right\}$$

où c est déterminée par

$$E_{\theta_0}(\Psi(X_1, \dots, X_n)) = P_{\theta_0} \left(\sum_{i=1}^n X_i > c \right) = \alpha$$

Pour $\theta = \theta_0$, $\sum_{i=1}^n X_i \sim \mathcal{N}(n\theta_0, n)$ donc $c = n\theta_0 + \sqrt{n}\Phi^{-1}(1 - \alpha)$.

4.4.2 Test bilatéral $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$

Il n'existe pas en général de test UPP pour le test bilatéral $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$. En effet, pour être UPP, un test doit être le plus puissant pour tester $H_0 : \theta = \theta_0$ contre

$H_1 : \theta = \theta_1$ pour tout $\theta_1 \neq \theta_0$. Cependant, selon le lemme de Neyman-Pearson, la forme des tests les plus puissants diffère selon que $\theta_1 > \theta_0$ ou $\theta_1 < \theta_0$.

Soient X_1, \dots, X_n des v.a.i.i.d. de loi binomiale de paramètres n et θ et supposons que l'on veuille tester

$$H_0 : \theta = \theta_0 \text{ contre } H_1 : \theta \neq \theta_0$$

à un certain niveau α . Considérons tout d'abord, les hypothèses

$$H'_0 : \theta = \theta_0 \text{ contre } H'_1 : \theta = \theta_1$$

avec $\theta_1 \neq \theta_0$. Le lemme de Neyman-Pearson indique que le test UPP de H'_0 contre H'_1 est basé sur la statistique de test :

$$T = \frac{\mathcal{L}(\theta_1; X_1, \dots, X_n)}{\mathcal{L}(\theta_0; X_1, \dots, X_n)} = \left(\frac{1 - \theta_0}{1 - \theta_1} \right)^n \left(\frac{\theta_1(1 - \theta_0)}{\theta_0(1 - \theta_1)} \right)^{\sum_{i=1}^n X_i}$$

Si $\theta_1 > \theta_0$, il est facile de vérifier que T est une fonction croissante de $\sum_{i=1}^n X_i$: donc un test plus puissant de H'_0 contre H'_1 rejettera H'_0 pour de grandes valeurs de $\sum_{i=1}^n X_i$. Mais, si $\theta_1 < \theta_0$, T est une fonction décroissante de $\sum_{i=1}^n X_i$ et un test plus puissant va rejeter H'_0 pour de petites valeurs de $\sum_{i=1}^n X_i$. On comprend donc qu'on ne pourra pas trouver de test UPP.

Une solution consiste à se restreindre à la classe des tests sans biais.

Définition 20 - Un test est dit sans biais si sa puissance est toujours supérieure à son niveau, autrement dit si pour tester $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$, pour tout $\theta_1 \in \Theta_1$,

$$E_{\theta_1}(\Psi(X_1, \dots, X_n)) \geq \sup_{\theta \in \Theta_0} E_{\theta}(\Psi(X_1, \dots, X_n))$$

c'est à dire qu'on a souvent raison quand on conclut H_1 .

Proposition 4 Un test UPP est forcément sans biais.

Preuve - Soit Ψ un test UPP de niveau α pour tester $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$. Soit $\psi(x_1, \dots, x_n) = \alpha$ pour tout (x_1, \dots, x_n) . ψ est de niveau α car

$$\sup_{\theta \in \Theta_0} E_{\theta}(\psi(X_1, \dots, X_n)) = \sup_{\theta \in \Theta_0} \alpha = \alpha$$

Comme Ψ est UPP parmi les tests de niveau α , pour tout $\theta_1 \in \Theta_1$,

$$E_{\theta_1}(\Psi(X_1, \dots, X_n)) = E_{\theta_1}(\psi(X_1, \dots, X_n)) = \alpha$$

Donc la puissance de Ψ est toujours supérieure à son niveau et donc Ψ est sans biais. \diamond

Il est parfois possible de construire des tests uniformément plus puissants parmi les tests dans biais. Supposons qu'on veuille tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$ au niveau α , on

peut construire un test UPP parmi les tests sans biais³ en combinant des tests UPP pour tester $H'_0 : \theta \leq \theta_0$ contre $H''_0 : \theta \geq \theta_0$. Plus précisément, supposons que $\Phi_1(X_1, \dots, X_n)$ est un test UPP de niveau α_1 et $\Phi_2(X_1, \dots, X_n)$ est un test UPP de niveau α_2 tels que $\alpha_1 + \alpha_2 = \alpha$. Alors $\Phi = \Phi_1 + \Phi_2$ sera un test de niveau α si

$$\Phi_1(1, \dots, X_n) + \Phi_2(1, \dots, X_n) \leq 1$$

Ainsi, en choisissant bien α_1 et α_2 il est possible d'avoir Φ un test UMPU. Le choix naturel est $\alpha_1 = \alpha_2 = \alpha/2$ mais en général, ça ne conduit pas à un test sans biais.

Soit X une variable aléatoire continue de densité

$$f(x; \theta) = \theta x^{\theta-1} \text{ pour } 0 \leq x \leq 1$$

et supposons qu'on veuille tester

$$H_0 : \theta = 1 \text{ contre } H_1 : \theta_1 \neq 1$$

au niveau 5%. On va rejeter H_0 if $x \leq 0.025$ ou si $x \geq 0.975$; ce test est clairement de niveau $\alpha = 5\%$ puisque $P_{\theta=1}(X \leq 0.025) = P_{\theta=1}(X \geq 0.975) = 0.025$.

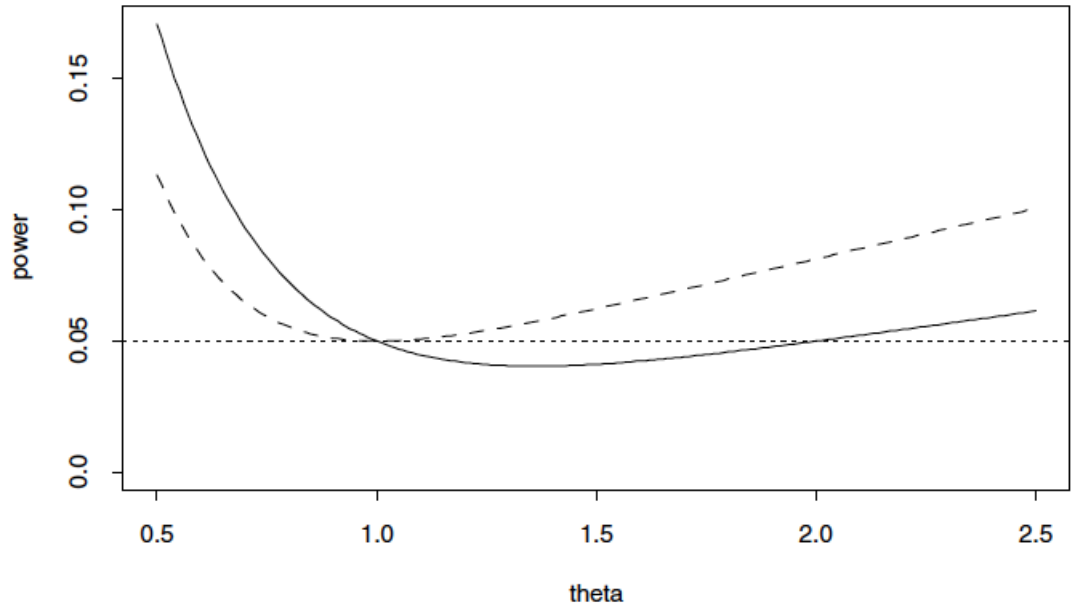
La fonction puissance est alors

$$\Pi(\theta) = \int_0^{0.025} \theta x^{\theta-1} dx + \int_{0.975}^1 \theta x^{\theta-1} dx = 1 + 0.025^\theta + 0.975^\theta$$

Si on évalue $\Pi(\theta)$ pour θ proche de 1 il est facile de voir que le test n'est pas sans biais. En effet $\Pi(\theta) < 0.05$ pour $1 < \theta < 2$.

Cependant, il est possible de trouver un test sans biais pour tester H_0 contre H_1 . Ce test rejette H_0 si $x \leq 0.0085$ ou si $x \geq 0.9585$. Les deux fonctions puissance sont représentées ci-dessous avec la fonction puissance du test sans biais en tirets :

³UMPU : unbiased most powerfull test



Dans cet exemple, le test sans biais a une puissance plus grande pour $\theta > 1$ mais plus petite pour $\theta < 1$. Ceci illustre le fait qu'en choisissant un test sans biais, on sacrifie de la puissance dans certaines régions.

4.5 Généralisation

Jusqu'à présent, nous avons proposé des méthodes constructives permettant d'obtenir des tests UPP (ou localement UPP) dans le cas d'un paramètre unique. Ce type de test optimal n'existe pas en général pour les modèles à plus d'un paramètre. Il existe une méthode qui permet de développer des tests dans des cas plus généraux.

Considérons le test

$$H_0 : \theta \in \Theta_0 \text{ contre } \theta \in \Theta_1$$

avec $\Theta_0 \cap \Theta_1 = \emptyset$. On va utiliser le même type d'idée que dans le lemme de Neyman-Pearson.

Définition 21 *La statistique Λ du rapport de vraisemblance est*

$$\Lambda = \frac{\sup_{\theta \in \Theta} \mathcal{L}(X_1, \dots, X_n; \theta)}{\sup_{\theta \in \Theta_0} \mathcal{L}(X_1, \dots, X_n; \theta)}.$$

Un test du rapport de vraisemblance pour $H_0 : \theta \in \Theta_0$ contre $\theta \in \Theta_1$ va rejeter H_0 pour de grandes valeurs de Λ .

Pour utiliser ces tests du rapport de vraisemblance, on a besoin de connaître (exactement ou approximativement) la distribution de la statistique Λ sous H_0 . Dans certains cas, Λ est fonction d'une autre statistique T dont on connaît la loi et on peut alors utiliser cette statistique. Sinon on utilise un résultat limite.

Considérons le test bilatéral

$$H_0 : \theta = \theta_0 \text{ contre } \theta \neq \theta_1$$

et X_1, \dots, X_n v.a.i.i.d. de densité (ou fonction de fréquence) f dans (P_θ) . La statistique Λ s'écrit

$$\Lambda_n = \prod_{i=1}^n \frac{f(X_i; \hat{\theta}_n)}{f(X_i; \theta)}$$

avec $\hat{\theta}_n$ l'estimateur du maximum de vraisemblance.

Théorème 10 Soient X_1, \dots, X_n v.a.i.i.d. qui admettent une densité (ou une fonction de fréquence) vérifiant les hypothèses (A1) à (A5) (du chapitre précédent) avec $I(\theta) = J(\theta)$. Si l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ satisfait un théorème de limite centrale

$$\sqrt{n}(\hat{\theta}_n - \theta) \rightarrow Z \sim \mathcal{N}(0, 1/I(\theta))$$

alors la statistique Λ satisfait

$$2 \ln(\Lambda) \rightarrow V \sim \chi^2(1)$$

quand $H_0 : \theta = \theta_0$ est vraie.

Preuve - Notons $\ell(x; \theta) = \ln(f(x; \theta))$ et $\ell'(x; \theta)$ et $\ell''(x; \theta)$ ses dérivées par rapport à θ . Sous les hypothèses du théorème, sous H_0 ,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow Z \sim \mathcal{N}(0, 1/I(\theta_0))$$

En prenant le $\log(\Lambda_n)$ et en faisant un développement de Taylor, on a

$$\begin{aligned} \ln(\Lambda_n) &= \sum_{i=1}^n \left(\ell(X_i; \hat{\theta}_n) - \ell(X_i; \theta_0) \right) \\ &= (\theta_0 - \hat{\theta}_n) \sum_{i=1}^n \ell'(X_i; \hat{\theta}_n) - \frac{1}{2}(\hat{\theta}_n - \theta_0)^2 \ell''(X_i; \theta_n^*) \\ &= -\frac{1}{2}n(\hat{\theta}_n - \theta_0)^2 \frac{1}{n} \ell''(X_i; \theta_n^*) \end{aligned}$$

où θ_n^* est entre θ_0 et $\hat{\theta}_n$. Sous les hypothèses (A4) et (A5), du chapitre précédent, on a donc sous H_0

$$\frac{1}{n} \ell''(X_i; \theta_n^*) \rightarrow_P -E_{\theta_0}[\ell''(X_i; \theta_0)] = I(\theta_0)$$

On a aussi

$$n(\hat{\theta}_n - \theta_0)^2 \rightarrow_d \frac{V}{I(\theta_0)}$$

et on conclut avec le lemme de Slutsky. \diamond

Exemple - Soient X_1, \dots, X_n v.a.i.i.d. de loi de Gauss de moyenne μ et de variance σ^2 (toutes deux inconnues) et supposons qu'on veut tester

$$H_0 : \mu = \mu_0 \text{ contre } H_1 : \mu \neq \mu_0$$

Les estimateurs du maximum de vraisemblance de μ et σ sont

$$\hat{\mu} = \bar{X} \text{ et } \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

Sous H_0 , l'estimateur du maximum de vraisemblance de σ s'écrit

$$\hat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu_0)^2$$

Et on a donc

$$\begin{aligned} \Lambda_n &= \frac{(2\pi\sigma_0^2)^{n/2}}{(2\pi\hat{\sigma}^2)^{n/2}} \exp \left(-\frac{1}{2\hat{\sigma}^2} \sum_{i=1}^n (X_i - \bar{X})^2 + \frac{1}{\hat{\sigma}_0^2} \sum_{i=1}^n (X_i - \mu_0)^2 \right) \\ &= \left(\frac{\sigma_0^2}{(\hat{\sigma}^2)} \right)^{n/2} \exp \left(\frac{1}{2\hat{\sigma}^2} n\hat{\sigma}^2 + \frac{1}{2\hat{\sigma}_0^2} n\hat{\sigma}_0^2 \right) \\ &= \left(\frac{\sigma_0^2}{(\hat{\sigma}^2)} \right)^{n/2} \end{aligned}$$

et la région critique est donc de la forme $\left\{ \left(\frac{\sigma_0^2}{\hat{\sigma}^2} \right)^{n/2} \geq c \right\}$ avec c déterminé par

$$P_\theta \left(\left(\frac{\sigma_0^2}{\hat{\sigma}^2} \right)^{n/2} \geq c \right) = \alpha$$

car le test est pur (loi continue). La distribution de Λ n'est pas triviale; cependant, on remarque que Λ est une fonction monotone de $\hat{\sigma}_0^2/\hat{\sigma}^2$ et

$$\begin{aligned} \frac{\sigma_0^2}{\hat{\sigma}^2} &= \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \\ &= 1 + \frac{n(\bar{X} - \mu_0)^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \\ &= 1 + \frac{1}{n-1} \left(\frac{n(\bar{X} - \mu_0)^2}{S^2} \right) \\ &= 1 + \frac{T^2}{n-1} \end{aligned}$$

où

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

et

$$T = \sqrt{n}(\bar{X} - \mu_0)/S$$

. Or on sait que sous H_0 , T suit une distribution de Student à $(n-1)$ degrés de liberté et donc que T^2 suit une distribution de Fisher à 1 et $(n-1)$ degrés de liberté.

On peut généraliser le théorème précédent aux de paramètres dans \mathbb{R}^p pour $p > 1$.

Théorème 11 Soient X_1, \dots, X_n v.a.i.i.d. qui admettent une densité (ou une fonction de fréquence) vérifiant les hypothèses A1 à A5 dans le cas d'un paramètre de dimension p et avec $I(\theta) = J(\theta)$ où $\theta = (\theta_1, \dots, \theta_p)$. Si l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ satisfait un théorème de limite centrale

$$\sqrt{n}(\hat{\theta}_n - \theta) \rightarrow Z \sim \mathcal{N}(0, 1/I(\theta))$$

alors la statistique Λ pour tester $H_0 : \theta_1 = \theta_{10}, \dots, \theta_r = \theta_{r0}$ satisfait

$$2 \ln(\Lambda) \rightarrow V \sim \chi^2(r)$$

quand H_0 est vraie.

La preuve du théorème repose sur le fait que la log-vraisemblance peut être approchée par une fonction quadratique près de la vraie valeur du paramètre.

Exemple - Soient X_1, \dots, X_m des variables aléatoires i.i.d. de loi exponentielle de paramètre λ et Y_1, \dots, Y_n des variables aléatoires i.i.d. de loi exponentielle de paramètre θ . On suppose que les X_i sont indépendants des Y_i . On veut tester

$$H_0 : \lambda = \theta \text{ contre } H_1 : \lambda \neq \theta$$

Les estimateurs du maximum de vraisemblance de λ et θ sont, dans le cas général,

$$\hat{\lambda} = 1/\bar{X} \text{ et } \hat{\theta} = 1/\bar{Y}$$

et sous H_0 ,

$$\hat{\lambda}_0 = \hat{\theta}_0 = \left(\frac{m\bar{X} + n\bar{Y}}{m + n} \right)^{-1}$$

On a donc

$$\Lambda = \left(\frac{m}{n+m} + \frac{n}{n+m} \frac{\bar{Y}}{\bar{X}} \right)^m \left(\frac{n}{n+m} + \frac{m}{n+m} \frac{\bar{X}}{\bar{Y}} \right)^n$$

On remarque que Λ ne dépend que de $T = \bar{X}/\bar{Y}$. On peut déduire un test de T ou construire un test asymptotique avec Λ .

Exemple - Soient $(X_1, Y_1), \dots, (X_n, Y_n)$ couples i.i.d. de variables aléatoires continues de densité jointe

$$f(x, y; \theta, \lambda, \alpha) = \frac{2\theta\lambda\alpha}{(\theta x + \lambda y + \alpha)^3} \text{ pour } x, y > 0$$

avec $\theta, \lambda, \alpha > 0$. Les densités marginales de X_i et Y_i sont

$$f_X(x; \theta, \alpha) = \frac{\theta\alpha}{(\theta x + \alpha)^2} \text{ pour } x > 0$$

$$f_Y(y; \lambda, \alpha) = \frac{\lambda\alpha}{(\lambda y + \alpha)^2} \text{ pour } y > 0$$

On veut tester $H_0 : \theta = \lambda$.

On peut reparamétriser ce problème de différentes façons. Par exemple, on peut définir $\eta_1 = \theta - \lambda$, $\eta_2 = \theta$ et $\eta_3 = \alpha$ ou $\eta_1 = \theta/\lambda$. On exprime alors H_0 en fonction de η_1 et on s'attend à ce que la statistique de test du rapport de vraisemblance suive approximativement une loi du χ^2 à 1 degré de liberté pour n grand.

4.6 Tests classiques

4.6.1 Tests paramétriques pour des moyennes, des variances ou des corrélations

- Test de Student
- Test de Fisher
- Analyse de la variance (variance inter-classe/ variance intra-classe)
- Test de Pearson

4.6.2 Tests non paramétriques pour des moyennes , des variances ou des corrélations

On les utilise quand on a de petits échantillons dont on ne connaît pas la distribution.

- Test du signe, test des signes et rangs de Wilcoxon ou mann-Whitney Wilcoxon
- Test de
- Test de Kruskal-Wallis
- test de Spearman, Test du τ de Kendall

4.6.3 Test d'adéquation⁴ ou de comparaison de distribution

- Tets du χ^2 : loi discrète
- Test de Kolmogorov, Cramer-von Mises, : loi continue quelconque
- Test de Shapiro-Wilk : loi normale

⁴En anglais : *Goodness-of-fit tests*

Chapitre 5

Estimation par intervalles

On a vu dans le chapitre sur l'estimation qu'une statistique n'est en général pas exactement égale à la valeur du paramètre qu'elle est censée estimer (si la loi de la statistique a une densité, cet événement est même de probabilité nulle). Donc, il est important que toute procédure d'estimation soit accompagnée d'une indication sur la précision de l'estimation.

5.1 Exemple

On considère un n -échantillon X_1, \dots, X_n de la loi gaussienne de paramètres $(\mu; \sigma^2)$. On suppose que σ^2 est connu. On estime μ . On a déjà vu toutes les qualités de l'estimateur $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. On sait que la variable $\sqrt{n}\sigma^{-1}(\bar{X}_n - \mu)$ est gaussienne standard. Sa loi ne dépend pas du paramètre μ , ce qu'on a déjà utilisé pour construire un test de niveau α pour tester $H_0 : \mu = \mu_0$ contre $H_1 : \mu \neq \mu_0$ ou contre $H_1 : \mu > \mu_0$. On sait par exemple que pour tout μ

$$P_\mu \left(\mu - \Phi^{-1}(1 - \alpha/2) \frac{\sigma}{\sqrt{n}} \leq \bar{X}_n \leq \mu + \Phi^{-1}(1 - \alpha/2) \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha \quad (5.1)$$

Or ceci peut aussi se lire

$$P_\mu \left(\bar{X}_n - \Phi^{-1}(1 - \alpha/2) \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \Phi^{-1}(1 - \alpha/2) \frac{\sigma}{\sqrt{n}} \right) \quad (5.2)$$

ce qui change tout car dans (5.2), l'intervalle est aléatoire (la probabilité qu'il contienne le paramètre est $1 - \alpha$), tandis que dans (5.1), l'intervalle est fixé et la variable aléatoire \bar{X}_n a une probabilité $1 - \alpha$ de se trouver dedans.

On dit que

$$I(X_1, \dots, X_n) = [\bar{X}_n - \Phi^{-1}(1 - \alpha/2) \frac{\sigma}{\sqrt{n}}, \bar{X}_n + \Phi^{-1}(1 - \alpha/2) \frac{\sigma}{\sqrt{n}}]$$

est un intervalle de confiance de niveau $1 - \alpha$ pour le paramètre μ .

Remarque - On a choisi un intervalle symétrique : est ce nécessaire ?

Dans une certaine mesure la réponse est oui. En effet, soit U une variable gaussienne

standard. Il est facile de vérifier que, $P(U \in [x, x + 2\Phi^{-1}(1 - \alpha/2)])$ est maximale (et vaut $1 - \alpha$) quand $x = -\Phi^{-1}(1 - \alpha/2)$. On en déduit que parmi les intervalles tels que $P(U \in [x, y]) = 1 - \alpha$, le plus court est $[-\Phi^{-1}(1 - \alpha/2), \Phi^{-1}(1 - \alpha/2)]$. Autrement dit l'intervalle proposé en (5.2) est le plus précis des intervalles de confiance de niveau $1 - \alpha$ pour le paramètre μ . Ceci dit, on peut avoir d'autres critères que la précision pour choisir l'intervalle de confiance. On peut vouloir par exemple une demi-droite de confiance si le seul souci est de garantir que le paramètre est suffisamment grand (ou suffisamment petit).

5.2 Méthode générale pour construire des intervalles de confiance

On cherche à estimer $\theta \in \Theta$ à partir de (X_1, \dots, X_n) , variables indépendantes et de même loi P_θ . Ici, $\Theta \subset \mathbb{R}^d$.

Définition 22 On appelle fonction pivotale, une fonction $h(\theta; X_1, \dots, X_n)$ à valeurs dans \mathbb{R}^k possédant les propriétés suivantes :

Propriété 1 : quelque soit $\theta \in \Theta$, la fonction $h(\theta; X_1, \dots, X_n)$ est mesurable.

Propriété 2 : la loi de $h(\theta; X_1, \dots, X_n)$ ne dépend pas de θ .

Il est facile de s'appuyer sur une fonction pivotale pour construire des intervalles de confiance de la façon suivante. Supposons qu'il existe B , un borélien de \mathbb{R}^k , tel que

$$P_\theta(h(\theta; X_1, \dots, X_n) \in B) = 1 - \alpha, \forall \theta \in \Theta \quad (5.3)$$

on défini alors la région de confiance

$$I(X_1, \dots, X_n) = \{\theta \in \Theta | h(\theta; X_1, \dots, X_n) \in B\}$$

Comme dans l'exemple introductif, l'ensemble $I(X_1, \dots, X_n)$ est aléatoire et a, sous P_θ , une probabilité $1 - \alpha$ de contenir θ . Dans l'exemple on avait

$$h(\mu; X_1, \dots, X_n) = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

et le borélien B était l'intervalle $[-\Phi^{-1}(1 - \alpha/2), \Phi^{-1}(1 - \alpha/2)]$

Remarques -

- L'égalité (5.3) s'écrit aussi $\pi(B) = 1 - \alpha$ où π désigne la loi (ne dépendant pas de θ , d'après la Propriété 2) transportée de P_θ^n par $h(\theta, x)$. Bien sûr, un tel B peut ne pas exister. On recherchera alors un borélien tel que $\pi(B) > 1 - \alpha$.
- On peut aussi bien se servir de la fonction pivotale pour construire un test de l'hypothèse $\theta = \theta_0$. Pour cela, il suffit de prendre comme région de rejet l'ensemble des (X_1, \dots, X_n) tels que $h(\theta_0; X_1, \dots, X_n) \notin B$.

5.3 Lien avec les tests

On peut construire des intervalles de confiance à l'aide des tests.

Théorème 12 Soit $\mathbf{X} = X_1, \dots, X_n$ un n -échantillon de loi $(P_\theta)_{\theta \in \Theta}$. Pour chaque $\theta_0 \in \Theta$ on considère le problème de tester $H_0 : \theta = \theta_0$ au niveau α et soit $A(\theta_0) = RC(\theta_0)^c$ sa région d'acceptation.

Pour chaque $(x_1, \dots, x_n) \in \mathbb{R}^n$ on définit $C(x)$ par

$$C(x_1, \dots, x_n) = \{\theta_0 \in \Theta : \mathbf{x} \in A(\theta_0)\}.$$

Alors $C(X_1, \dots, X_n)$ est une région de confiance pour θ de niveau $1 - \alpha$.

Réciproquement, soit $C(X)$ une région de confiance pour θ de niveau $1 - \alpha$. Pour tout $\theta_0 \in \Theta$ on définit

$$A(\theta_0) = \{(x_1, \dots, x_n) : \theta_0 \in C(\mathbf{x})\}$$

Alors $A(\theta_0)$ est la région d'acceptation d'un tests de niveau α pour $H_0 : \theta = \theta_0$.

Preuve - Comme $A(\theta_0)$ est la région d'acceptation d'une test de niveau α , on a

$$P_{\theta_0}(\mathbf{X} \notin A(\theta_0)) \leq \alpha$$

ou encore

$$P_{\theta_0}(\mathbf{X} \in A(\theta_0)) \geq 1 - \alpha$$

Comme θ_0 est quelconque, on peut écrire θ à la place de θ_0 . Alors, d'après l'inégalité précédente,

$$P_\theta(\theta \in C(\mathbf{X})) = P_\theta(\mathbf{X} \in A(\theta)) \geq 1 - \alpha$$

donc C est une région de confiance pour θ de niveau $1 - \alpha$.

Par ailleurs, on voit que l'erreur de 1ère espèce pour H_0 avec pour région d'acceptation $A(\theta_0)$ est

$$P_{\theta_0}(\mathbf{X} \notin A(\theta_0)) = P_{\theta_0}(\theta_0 \notin C(\mathbf{X})) \leq \alpha$$

donc le test est de niveau α sous H_0 . \diamond

Tables statistique

[a4paper,11pt]article
[latin1]inputenc amsmath amssymb theorem

Loi binomiale

$$X \sim \mathcal{B}(n, \pi), \pi = 1/2, P(X = k)$$

	0	1	2	3	4	5	6	7	8	9	10	11	12	...
n=6	0.015	0.094	0.234	0.312	0.234	0.094	0.016							
n=7	0.008	0.055	0.164	0.273	0.273	0.164	0.055	0.008						
n=8	0.004	0.031	0.109	0.219	0.273	0.219	0.109	0.031	0.004					
n=9	0.002	0.018	0.070	0.164	0.246	0.246	0.164	0.070	0.018	0.002				
n=10	0.001	0.010	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.010	0.001			
n=11	0.000	0.005	0.027	0.081	0.161	0.226	0.226	0.161	0.081	0.027	0.005	0.000		
n=12	0.000	0.003	0.016	0.054	0.121	0.193	0.226	0.193	0.121	0.054	0.016	0.003	0.000	
n=13	0.000	0.002	0.010	0.035	0.087	0.157	0.209	0.209	0.157	0.087	0.035	0.010	0.002	...
n=14	0.000	0.001	0.006	0.022	0.061	0.122	0.183	0.209	0.183	0.122	0.061	0.022	0.006	...
n=15	0.000	0.000	0.003	0.014	0.042	0.092	0.153	0.196	0.196	0.153	0.092	0.042	0.014	...
n=16	0.000	0.000	0.002	0.009	0.028	0.067	0.122	0.175	0.196	0.175	0.122	0.067	0.028	...
n=17	0.000	0.000	0.001	0.005	0.018	0.047	0.094	0.148	0.185	0.185	0.148	0.094	0.047	...
n=18	0.000	0.000	0.001	0.003	0.012	0.033	0.071	0.121	0.167	0.185	0.167	0.121	0.071	...
n=19	0.000	0.000	0.000	0.002	0.007	0.022	0.052	0.096	0.144	0.176	0.176	0.144	0.096	...
n=20	0.000	0.000	0.000	0.001	0.005	0.015	0.037	0.074	0.120	0.160	0.176	0.160	0.120	...

Si $n > 20$, on approche la loi binomiale par une loi de Gauss de moyenne $n\pi$ et de variance $n\pi(1-\pi)$.

$X \sim \mathcal{B}(n, \pi), \pi = 1/4, P(X = k)$

	0	1	2	3	4	5	6	7	8	9	10	11	12
n=6	0.178	0.356	0.297	0.132	0.033	0.004	0.000						
n=7	0.133	0.311	0.311	0.173	0.058	0.012	0.001	0.000					
n=8	0.100	0.267	0.311	0.208	0.087	0.023	0.004	0.000	0.000				
n=9	0.075	0.225	0.300	0.234	0.117	0.039	0.009	0.001	0.000	0.000			
n=10	0.056	0.188	0.282	0.250	0.146	0.058	0.016	0.003	0.000	0.000	0.000		
n=11	0.042	0.155	0.258	0.258	0.172	0.080	0.027	0.006	0.001	0.000	0.000	0.000	
n=12	0.032	0.127	0.232	0.258	0.194	0.103	0.040	0.011	0.002	0.000	0.000	0.000	0.000

Loi de Gauss centrée réduite

$$\Pi(t) = P(X \leq t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \quad \text{et} \quad \Pi(-t) = 1 - \Pi(t).$$

t	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Loi de Student

The t distribution

$$\Pr[T \leq t] = \int_{-\infty}^t \frac{\Gamma((r+1)/2)}{\sqrt{\pi r} \Gamma(r/2)} \frac{1}{(1+x^2/r)^{(r+1)/2}} dx$$

$$\Pr[T \leq -t] = 1 - \Pr[T \leq t]$$

		$\Pr[T \leq t]$				
(1)2-7	r	0.90	0.95	0.975	0.99	0.995
	1	3.078	6.314	12.706	31.821	63.657
	2	1.886	2.920	4.303	6.965	9.925
	3	1.638	2.353	3.182	4.541	5.841
	4	1.533	2.132	2.776	3.747	4.604
	5	1.476	2.015	2.571	3.365	4.032
	6	1.440	1.943	2.447	3.143	3.707
	7	1.415	1.895	2.365	2.998	3.499
	8	1.397	1.860	2.306	2.896	3.355
	9	1.383	1.833	2.262	2.821	3.250
	10	1.372	1.812	2.228	2.764	3.169
	11	1.363	1.796	2.201	2.718	3.106
	12	1.356	1.782	2.179	2.681	3.055
	13	1.350	1.771	2.160	2.650	3.012
	14	1.345	1.761	2.145	2.624	2.977
	15	1.341	1.753	2.131	2.602	2.947
	16	1.337	1.746	2.120	2.583	2.921
	17	1.333	1.740	2.110	2.567	2.898
	18	1.330	1.734	2.101	2.552	2.878
	19	1.328	1.729	2.093	2.539	2.861
	20	1.325	1.725	2.086	2.528	2.845
	21	1.323	1.721	2.080	2.518	2.831
	22	1.321	1.717	2.074	2.508	2.819
	23	1.319	1.714	2.069	2.500	2.807
	24	1.318	1.711	2.064	2.492	2.797
	25	1.316	1.708	2.060	2.485	2.787
	26	1.315	1.706	2.056	2.479	2.779
	27	1.314	1.703	2.052	2.473	2.771
	28	1.313	1.701	2.048	2.467	2.763
	29	1.311	1.699	2.045	2.462	2.756
	30	1.310	1.697	2.042	2.457	2.750

Loi du khi 2

The chi-square distribution

$$\Pr[X \leq x] = \int_0^x \frac{1}{\Gamma(r/2)2^{r/2}} y^{r/2-1} e^{-y/2} dy$$

		Pr[X ≤ x]					
(l)2-8	r	0.01	0.025	0.05	0.95	0.975	0.99
	1	0.000	0.001	0.004	3.841	5.024	6.635
	2	0.020	0.051	0.103	5.991	7.378	9.210
	3	0.115	0.216	0.352	7.815	9.348	11.345
	4	0.297	0.484	0.711	9.488	11.143	13.277
	5	0.554	0.831	1.145	11.070	12.833	15.086
	6	0.872	1.237	1.635	12.592	14.449	16.812
	7	1.239	1.690	2.167	14.067	16.013	18.475
	8	1.646	2.180	2.733	15.507	17.535	20.090
	9	2.088	2.700	3.325	16.919	19.023	21.666
	10	2.558	3.247	3.940	18.307	20.483	23.209
	11	3.053	3.816	4.575	19.675	21.920	24.725
	12	3.571	4.404	5.226	21.026	23.337	26.217
	13	4.107	5.009	5.892	22.362	24.736	27.688
	14	4.660	5.629	6.571	23.685	26.119	29.141
	15	5.229	6.262	7.261	24.996	27.488	30.578
	16	5.812	6.908	7.962	26.296	28.845	32.000
	17	6.408	7.564	8.672	27.587	30.191	33.409
	18	7.015	8.231	9.390	28.869	31.526	34.805
	19	7.633	8.907	10.117	30.144	32.852	36.191
	20	8.260	9.591	10.851	31.410	34.170	37.566
	21	8.897	10.283	11.591	32.671	35.479	38.932
	22	9.542	10.982	12.338	33.924	36.781	40.289
	23	10.196	11.689	13.091	35.172	38.076	41.638
	24	10.856	12.401	13.848	36.415	39.364	42.980
	25	11.524	13.120	14.611	37.652	40.646	44.314
	26	12.198	13.844	15.379	38.885	41.923	45.642
	27	12.879	14.573	16.151	40.113	43.195	46.963
	28	13.565	15.308	16.928	41.337	44.461	48.278
	29	14.256	16.047	17.708	42.557	45.722	49.588
	30	14.953	16.791	18.493	43.773	46.979	50.892

Loi de Fisher

The F distribution

$$\Pr[F_{r_1, r_2} \leq f] = \int_0^f \frac{\Gamma((r_1 + r_2)/2)(r_1/r_2)^{r_1/2} x^{(r_1/2-1)}}{\Gamma(r_1/2)\Gamma(r_2/2)(1 + r_1 x/r_2)^{(r_1+r_2)/2}} dx$$

[illegible]

647.79 799.50 864.16 899.58 921.85 937.11 948.22 956.66 963.28 968.63 976.71 984.87
0.99

4052.184999.505403.355624.585763.655858.995928.365981.076022.476055.856106.326157.28

0.95218.5119.0019.1619.2519.3019.3319.3519.3719.3819.4019.4119.43
0.975

38.51 39.00 39.17 39.25 39.30 39.33 39.36 39.37 39.39 39.40 39.41 39.43
0.99

98.50 99.00 99.17 99.25 99.30 99.33 99.36 99.37 99.39 99.40 99.42 99.43

0.95310.139.559.289.129.018.948.898.858.818.798.748.70
0.975

17.44 16.04 15.44 15.10 14.88 14.73 14.62 14.54 14.47 14.42 14.34 14.25
0.99

34.12 30.82 29.46 28.71 28.24 27.91 27.67 27.49 27.35 27.23 27.05 26.87

0.9547.716.946.596.396.266.166.096.046.005.965.915.86
0.975

12.22 10.65 9.98 9.60 9.36 9.20 9.07 8.98 8.90 8.84 8.75 8.66
0.99

21.20 18.00 16.69 15.98 15.52 15.21 14.98 14.80 14.66 14.55 14.37 14.20

0.9556.615.795.415.195.054.954.884.824.774.744.684.62
0.975

10.01 8.43 7.76 7.39 7.15 6.98 6.85 6.76 6.68 6.62 6.52 6.43
0.99

16.26 13.27 12.06 11.39 10.97 10.67 10.46 10.29 10.16 10.05 9.89 9.72

0.9565.995.144.764.534.394.284.214.154.104.064.003.94
0.975

8.81 7.26 6.60 6.23 5.99 5.82 5.70 5.60 5.52 5.46 5.37 5.27
0.99

13.75 10.92 9.78 9.15 8.75 8.47 8.26 8.10 7.98 7.87 7.72 7.56

0.9575.594.744.354.123.973.873.793.733.683.643.573.51
0.975

8.07 6.54 5.89 5.52 5.29 5.12 4.99 4.90 4.82 4.76 4.67 4.57
0.99

12.25 9.55 8.45 7.85 7.46 7.19 6.99 6.84 6.72 6.62 6.47 6.31

0.9585.324.464.073.843.693.583.503.443.393.353.283.22
0.975

7.57 6.06 5.42 5.05 4.82 4.65 4.53 4.43 4.36 4.30 4.20 4.10
0.99

11.26 8.65 7.59 7.01 6.63 6.37 6.18 6.03 5.91 5.81 5.67 5.52

0.9595.124.263.863.633.483.373.293.233.183.143.073.01
0.975

7.21 5.71 5.08 4.72 4.48 4.32 4.20 4.10 4.03 3.96 3.87 3.77
0.99

10.56 8.02 6.99 6.42 6.06 5.80 5.61 5.47 5.35 5.26 5.11 4.96

0.95104.964.103.713.483.333.223.143.073.022.982.912.85
0.975

6.94 5.46 4.83 4.47 4.24 4.07 3.95 3.85 3.78 3.72 3.62 3.52
0.99

10.04 7.56 6.55 5.99 5.64 5.39 5.20 5.06 4.94 4.85 4.71 4.56

0.95124.753.893.493.263.113.002.912.852.802.752.692.62
0.975

6.55 5.10 4.47 4.12 3.89 3.73 3.61 3.51 3.44 3.37 3.28 3.18
0.99

9.33 6.93 5.95 5.41 5.06 4.82 4.64 4.50 4.39 4.30 4.16 4.01

0.95154.543.683.293.062.902.792.712.642.592.542.482.40
0.975

6.20 4.77 4.15 3.80 3.58 3.41 3.29 3.20 3.12 3.06 2.96 2.86
0.99

8.68 6.36 5.42 4.89 4.56 4.32 4.14 4.00 3.89 3.80 3.67 3.52