

Cours de Statistiques Inférentielles

P. Ribereau

6 janvier 2016

Table des matières

1	Introduction Générale	7
I	Intro générale	7
II	Le recueil des données.	9
III	La statistique exploratoire ou descriptive.	9
IV	La statistique inférentielle.	9
V	La modélisation statistique.	10
VI	Un exemple simple en assurance automobile	10
2	Statistiques descriptives	13
I	Généralités / définitions de base	13
II	Résumés numériques pour des variables quantitatives	14
III	Résumés graphiques	17
IV	Exercices	20
3	Echantillonnage	23
I	Modèle statistique	23
1	Généralités	23
2	Modèle d'échantillonnage.	23
3	Modèle dominé, vraisemblance	25
II	Définition d'une statistique	25
1	Notions d'estimation et de test d'hypothèses	26
III	Quelques notions de base sur les estimateurs	27
1	Définition d'un estimateur	27
2	Notion de biais	27
3	Convergence d'un estimateur	27
4	Comparaisons des estimateurs	28
5	Moyenne aléatoire, variance aléatoire	28
IV	Rappels sur les vecteurs gaussiens	29
1	Vecteurs gaussiens et lois du χ^2	31
V	Application à l'estimation dans un cadre normal	32
4	Estimation	35
I	Hypothèses fondamentales sur la densité $f(x, \theta)$	35
II	Information	36
1	Information de Fisher	36
2	Propriétés.	37

III	Inégalité de Cramer-Rao	37
1	Hypothèses supplémentaires	37
2	Relation entre estimateurs efficaces	39
3	Dégradation de l'information	39
IV	Notion d'exhaustivité	40
V	Exhaustivité et estimateurs efficaces ; la famille exponentielle	41
1	Le modèle exponentiel	41
2	Théorème sur l'efficacité	42
VI	Quelques méthodes usuelles d'estimation	43
1	Méthode empirique	43
2	Méthode des moindres carrés	44
3	Méthode des moments	44
4	Méthode du maximum de vraisemblance : principe	45
VII	Exercices	46
VIII	Généralisation au cas d'un paramètre multidimensionnel	47
1	Généralisation des définitions sur les estimateurs	47
2	Généralisation de l'inégalité de Cramer-Rao	48
3	Généralisation de la méthode du maximum de vraisemblance	50
5	Comportement asymptotique des estimateurs	53
I	Propriétés asymptotiques de l'EMV	53
1	En dimension 1	53
2	En dimension supérieure	54
II	Définitions / outils	54
1	Normalité et efficacité asymptotique	54
2	Propriétés de convergence	55
3	Méthode Delta	55
III	Exercices	55
6	Estimation par intervalle de confiance	57
I	Introduction	57
II	I.C. pour les paramètres de la loi normale	59
III	I.C. pour une proportion (paramètre de la loi binomiale)	60
IV	Construction d'I.C. asymptotiques	61
1	Utilisation du théorème central limite	61
2	Application à la loi binomiale.	62
3	Utilisation de la convergence de l'EMV	62
4	Remarque sur l'intervalle de confiance pour une variance hors du cadre normal	62
V	Recherche de régions de confiance	64
VI	Exercices	64
7	Généralités sur les tests	67
I	Problèmes de test	67
II	Tests uniformément plus puissants	68
III	Tests fondés sur le rapport du maximum de vraisemblance	71
IV	Exemples	72

1	Adéquation d'une moyenne pour un échantillon gaussien . . .	72
2	Comparaison de deux moyennes	73
3	Un exemple avec une loi discrète	74
V	Tests asymptotiques	75
1	Propriétés asymptotiques des tests du maximum de vraisem- blance	76
2	Tests de Wald et du score	77
8	Tests paramétriques classiques	79
I	Tests gaussiens	79
II	Tests asymptotiques	81
9	Quelques tests non paramétriques	83
I	Tests du χ^2	83
1	Loi multinômiale	83
2	Loi asymptotique	83
3	Test du χ^2 d'adéquation à une loi	85
4	Test du χ^2 d'indépendance	86
II	Test de Kolmogorov-Smirnov	88
III	Test de Shapiro-Wilk	89
1	Droite de Henry	90
2	Test de Shapiro-Wilk	90
IV	Tests de rang	92
1	Statistiques de l'ordre, de rang	93
2	Le test de Wilcoxon	93
10	Exemples d'estimation non paramétrique	95
I	Estimation d'une densité de probabilité	95
1	Histogramme empirique	95
2	Fenêtres mobiles	95
3	Versions lisses	96
4	Un exemple	96
II	Estimation des quantiles	99
1	Quantiles empiriques	99
2	Lien avec les statistiques d'ordre	99
3	Résultats asymptotiques	99

Chapitre 1

Introduction Générale

I Intro générale

Définition 1.1 (Première définition d'un problème statistique) *On dira qu'on se trouve devant un problème statistique si*

- on est confronté à des éventualités (en nombre fini ou infini) dont on sait que certaines sont vraies sans savoir lesquelles,
- on doit choisir une de ces éventualités,
- en s'appuyant sur le résultat d'une expérience aléatoire, éventuellement à définir.

La dernière partie de la définition par le d'expérience aléatoire. Cette partie là (qui concerne la majorité du cours) est la statistique inférentielle, mathématique ou inductive. On va encore restreindre la définition pour définir le champs du cours et introduire la notion de modèle statistique. Pour pouvoir tirer quelque chose du résultat de l'expérience aléatoire, il faut qu'à chaque éventualité, on fasse correspondre une famille de probabilité à laquelle la proba qui régit l'expérience appartient (si cette éventualité est vraie).

On note \mathcal{P} l'ensemble de toutes les probas possibles.

Définition 1.2 *On appelle modèle statistique le triplet $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ où :*

- \mathcal{X} est l'ensemble appelé ensemble fondamental ou espace des résultats,
- \mathcal{A} est une tribu de partie de \mathcal{X} ,
- \mathcal{P} est une famille de probabilités sur l'espace mesurable $(\mathcal{X}, \mathcal{A})$.

Définition 1.3 *On appelle modèle statistique paramétrique un modèle statistique $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ tel que*

$$\exists p \in \mathbb{N} : \mathcal{P} = \{\mathbb{P}_\theta ; \theta \in \Theta \subset \mathbb{R}^p\}$$

Θ est appelé l'espace des paramètres. Le modèle est aussi noté $(\mathcal{X}, \mathcal{A}, \mathcal{P}_\Theta, \Theta)$

Les grands problèmes statistiques les plus fréquents sont :

- dans le modèle statistique paramétrique, les "éventualités" sont représentées par le paramètre θ lui même. Comment choisir le θ ? C'est l'estimation ponctuelle. Problème d'identifiabilité (si θ_1 et θ_2 sont tels que $\mathbb{P}_{\theta_1} = \mathbb{P}_{\theta_2}$ alors on ne pourra jamais choisir entre les deux) : il faut que l'application $\theta \rightarrow \mathbb{P}_\theta$ soit injective.

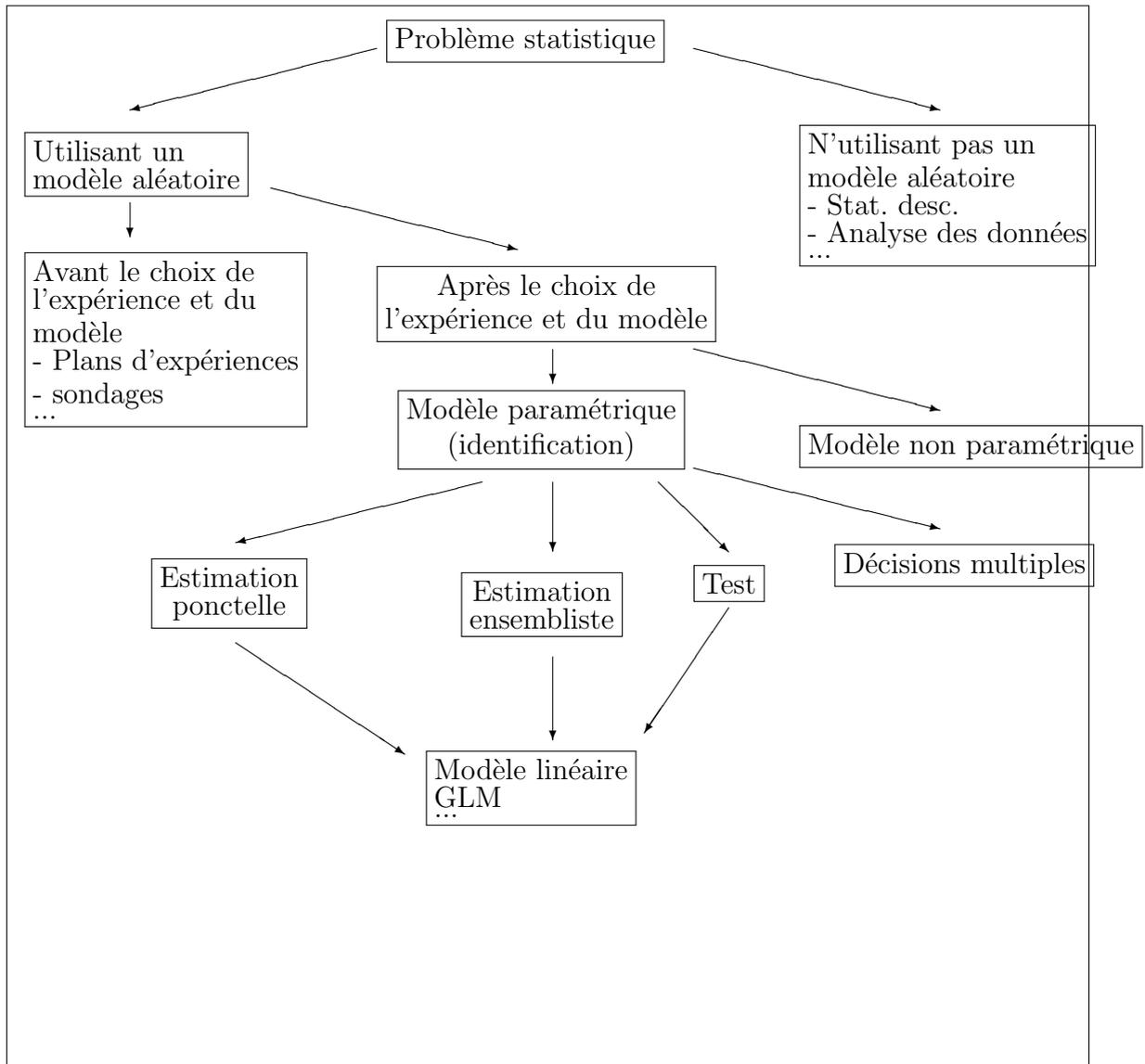


FIGURE 1.1 – Approche d'un problème statistique

- Toujours dans ce modèle, l'ensemble des éventualités est l'ensemble des parties de Θ . Estimation ensembliste.
- on a deux éventualités dont une seule est vraie. Problème de tests!

Pour en revenir au problème statistique, commentons la Figure 1.1 :

La démarche statistique consiste à traiter et à interpréter les informations recueillies par le biais de données. Elle comporte quatre grands aspects : le recueil des données, l'aspect descriptif ou exploratoire, l'aspect inférentiel ou décisionnel et la modélisation statistique.

II Le recueil des données.

Cette étape est importante car elle doit permettre d'obtenir des données de "bonne qualité" en un certain sens. Contrairement à ce qu'indique le vocabulaire, les informations dont a besoin le statisticien ne sont pourtant pas "données" et la qualité des résultats obtenus dépendra autant de la manière dont les données ont été collectées que de la méthode statistique utilisée ensuite. La théorie des sondages et celle des plans d'expériences fournissent un cadre théorique pour une collecte optimale de données.

III La statistique exploratoire ou descriptive.

Une fois les données collectées, il convient de synthétiser et de résumer l'information contenue dans ces données. On utilise pour cela des représentations des données sous forme de tableaux, de graphiques ou d'indicateurs numériques (tels que la moyenne, la variance, la corrélation linéaire, . . . pour des variables quantitatives).

Cette phase est connue sous le nom de **statistique descriptive**. On parle de statistique descriptive *univariée* lorsque l'on regarde une seule variable, de statistique descriptive *bivariée* lorsque l'on regarde simultanément deux variables, et de statistique descriptive *multidimensionnelle* lorsque l'on regarde simultanément p variables. Dans ce dernier cas, on parle aussi d'**analyse des données**.

IV La statistique inférentielle.

Son but est d'étendre (d'inférer) les propriétés constatées sur l'échantillon (grâce l'analyse exploratoire par exemple) à la population toute entière, et de valider ou d'infirmer des hypothèses.

Contrairement à la statistique exploratoire, des hypothèses probabilistes sont ici nécessaires : elle suppose un modèle probabiliste. L'estimation ponctuelle ou par intervalle de confiance et la théorie des tests d'hypothèses constituent une partie principale de la statistique inférentielle.

V La modélisation statistique.

Elle consiste en général à rechercher une relation *approximative* entre une variable et plusieurs autres variables, la forme de cette relation est le plus souvent linéaire. Lorsque la variable à expliquer est quantitative et que les variables explicatives sont aussi quantitatives, on parle de *régression linéaire*.

Si les variables explicatives sont qualitatives, on parle alors d'*analyse de la variance*. Le *modèle linéaire général* englobe une grande partie de tous les cas de figures possibles.

VI Un exemple simple en assurance automobile

Un assureur souhaite évaluer s'il est pertinent de tarifer ses contrats d'assurance auto en fonction de la région d'habitation de l'assuré. Il dispose de deux échantillons d'assurés : 1 500 assurés de la région Nord - Pas de Calais et 1 249 assurés de la région Provence - Alpes - CÔte d'Azur. Sur l'année 2008, les 1 500 assurés de la région Nord - Pas de Calais considérés ont déclaré 89 sinistres pour un montant total de remboursement de 57 227€ . Sur l'année 2008, les 1 249 assurés de la région Nord - Pas de Calais considérés ont déclaré 78 sinistres pour un montant total de remboursement de 57 954€ .

Au vu des ces données, peut-on conclure qu'une des deux régions est plus risquée du point de vue de l'assureur et prendre en compte l'origine géographique dans la tarification ? Réponse à la fin du cours sur les tests paramétriques.

L'objet de ce cours est de présenter les principes de la statistique inférentielle : échantillonnage, estimation, tests. Au delà des outils théoriques indispensables à la maîtrise de l'outil statistique, on s'attachera aussi à le mettre œuvre sur des exemples concrets, en utilisant les logiciels R et/ou SAS.

Plusieurs exemples du cours sont utilisent les données que vous pouvez télécharger à l'adresse :

<http://isfaserveur.univ-lyon1.fr/vmaume/donnees/bab.csv>

Ce fichier contient des informations sur le suivi des grossesses de 1236 mères américaines :

- **poidb** est le poids du bébé à la naissance (en onces - 1 onces \sim 28,35 grammes),
- **grossess** est le nombre de jours d'aménorée (en moyenne 9 mois et 2 semaines),
- **nbgrossesses** est le nombre de grossesses de la mère (avant celle qui est le sujet de l'étude),
- **age** est l'âge de la mère,
- **taillem** est la taille de la mère (en pouces),
- **poidm** est le poids de la mère avant la grossesse (en livres),

- `fumeuse` indique si la mère fume : 0=jamais, 1= fume maintenat, 2=a fumé jusqu'à la grossesse, 3=occasionnellement mais pas maintenant, 9=non renseigné,
- `ed` donne le niveau d'étude.

Pour travailler avec ce fichier, l'enregistre dans un répertoire courant, démarrer R, charger les données avec la commande : `bebes<-read.table('bab.csv',h=T)`

Quelques références bibliographiques :

G. Saporta *Probabilités, Statistique et Analyse des données*. Contient l'essentiel mais avec peu de preuves mathématiques.

D. Fourdrinier *Statistique Inférentielle*. Plus théorique.

On commence par un court chapitre sur la statistique descriptive unidimensionnelle.

Chapitre 2

Statistiques descriptives

Le but de la statistique descriptive - unidimensionnelle - est de fournir une synthèse des données sous forme de tableaux / graphiques / résumés numériques. Il s'agit de méthodes exploratoires et descriptives.

I Généralités / définitions de base

On étudie un ensemble d'objets de même nature : *la population* dont les éléments sont appelés *individus*, sur lesquels on observe des caractéristiques : *les variables*.

Exemple : un ensemble de pièces produites par une même machine sur lesquels on mesure le poids, le diamètre ect ...

Il s'agit d'étudier les caractéristiques de la population concernée (par exemple la proportion de pièces défectueuses) et non les caractéristiques d'un individu donné.

L'étude de tous les individus d'une population s'appelle un *recensement*. Une telle étude n'est généralement pas réalisable en pratique (nombre d'individu de la population trop important, coût élevé du recensement, destruction de l'objet que l'on étudie ect ...). On est ainsi amené à n'observer qu'une (petite) partie de la population : *un échantillon*. Le nombre d'individu d'une population ou d'un échantillon s'appelle la *taille*. Chaque individu est décrit par un ensemble de caractéristiques appelées *variables*. Ces variables peuvent être *quantitatives* : caractéristiques numériques ou *qualitatives*.

Dans la suite, on considère un échantillon \mathcal{E} de taille n sur lequel on étudie une variable X , c'est ainsi une application $X : \mathcal{E} \rightarrow A$, A est un sous-ensemble de \mathbb{R} si X est quantitative. On note a_1, \dots, a_n les valeurs prises par X : $a_i = X(e_i)$ si e_i est le $i^{\text{ème}}$ élément de \mathcal{E} . Un premier résumé de cette série statistique (a_1, \dots, a_n) consiste à regrouper les individus pour lesquels X prend la même valeur : soient x_1, \dots, x_k les valeurs distinctes prises par X , dans le cas d'une variable quantitative, $x_1 < \dots < x_k$. Les x_i s'appellent les *modalités* de la variable X . Le nombre n_i

d'individu de \mathcal{E} pour lesquels X vaut x_i s'appelle l'effectif de x_i . On a évidemment $n = \sum_{i=1}^k n_i$. Soit $f_i = \frac{n_i}{n}$ la *fréquence* de x_i , il s'agit de la proportion d'individus de l'échantillon pour lesquels X vaut x_i . On résume alors la série statistique sous la forme du tableau :

Modalité	Effectif	Fréquence
x_1	n_1	f_1
\vdots	\vdots	\vdots
x_k	n_k	f_k

II Résumés numériques pour des variables quantitatives

On considère une variable quantitative sur un échantillon de taille n . La variable prend des valeurs x_1, \dots, x_n . On note n_i l'effectif de x_i .

moyenne

La valeur moyenne de la variable sur cet échantillon est :

$$m(x) = m_1 = \bar{x} = \frac{1}{n} \sum_{i=1}^n n_i x_i.$$

Avec R, on obtient la moyenne avec la fonction `mean`. Faites l'essai avec `bebes [, 1]` puis `bebes [, 2]`, quel est le problème ? pour y remédier, on utilise la commande `mean(bebes [, 2], na.rm=T)`.

écart type, variance

La variance et l'écart-type mesurent l'écart de la variable à la valeur moyenne :

$$Var(x) = \frac{1}{n} \sum_{i=1}^n n_i (x_i - m(x))^2 = m(x^2) - [m(x)]^2.$$

L'écart-type est la racine de la variance : $\sigma(x) = \sigma_n = \sqrt{Var(x)}$.

Variance estimée, écart-type estimé

La variance estimée est l'estimateur sans biais de la variance (voir paragraphe *Échantillonnage*) :

$$\frac{1}{n-1} \sum_{i=1}^n n_i (x_i - m(x))^2 = \frac{n}{n-1} \text{Var}(x).$$

L'écart-type estimé est la racine de la variance estimée.

$$s_n(x) = \sqrt{s^2(x)}.$$

Sous R, la variance et l'écart-type **estimés** sont donnés par les fonctions `var` et `sd`. Lancer la commande `apply(bebes, 2, var, na.rm=T)`.

Remarque : un échantillon définit une mesure sur \mathbb{R} appelée *mesure empirique* de l'échantillon :

$$\mu_n = \frac{1}{n} \sum_{i=1}^n n_i \delta_{x_i}.$$

La moyenne de l'échantillon est l'espérance de loi μ_n , la variance de l'échantillon est la variance de la loi μ_n .

L'idée sous-jacente à la théorie statistique est que les données de l'échantillon sont des réalisations d'un échantillon aléatoire indépendant de loi μ et que l'on peut retrouver des informations sur μ (qui est à priori inconnue) à partir de la connaissance de μ_n .

Autres moments

Plus généralement, on définit les moments et les moments centrés d'ordre k d'une variable statistique par :

$$m_k = \frac{1}{n} \sum_{i=1}^k n_i (x_i)^k \quad mc_k = \frac{1}{n} \sum_{i=1}^k n_i (x_i - m(x))^k.$$

Le *coefficient d'asymétrie* $\gamma_1 = \frac{mc_3}{\sigma^3}$ et le *coefficient d'aplatissement* $\gamma_2 = \frac{mc_4}{\sigma^4}$ sont des caractéristiques de forme de la série statistique.

Mode / étendue

Le *mode* mo est la valeur la plus fréquente de l'échantillon. Autrement dit, $mo = x_i$ si et seulement si $n_i = \max\{n_j, j = 1, \dots, k\}$. L'*étendue* $w = x_k - x_1$ est la différence entre la plus grande et la plus petite valeur prise par les individus

de l'échantillon. C'est un indicateur très dépendant des valeurs extrêmes.

Fonction de répartition, médiane, quartiles

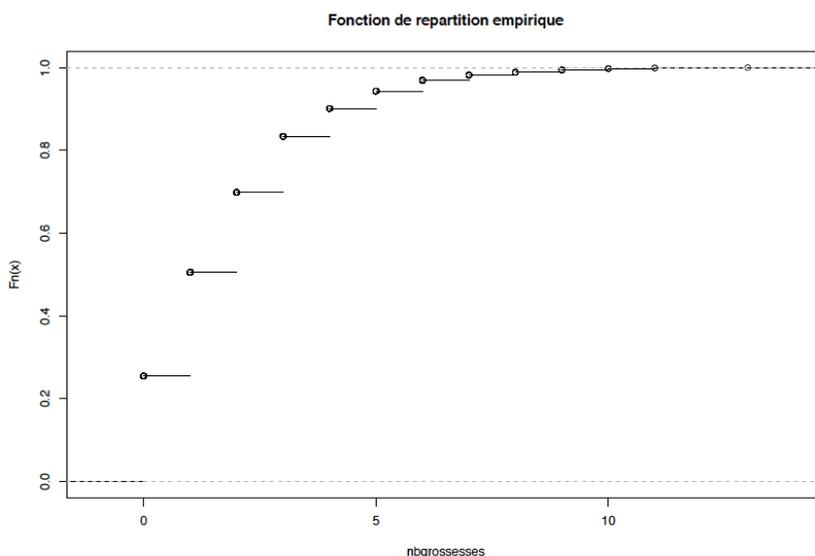
La fonction de répartition est la fonction de répartition de la loi empirique :

$$F_n(x) = \mu_n([-\infty, x]).$$

Cette fonction est constante par morceaux. Si on note f_i la fréquence de x_i ($f_i = \frac{n_i}{n}$) et F_i la fréquence cumulée ($F_i = f_1 + \dots + f_i$), on a

$$F_n(x) = F_i \text{ si } x_i \leq x < x_{i+1}.$$

Sous R, la fonction `ecdf` permet de tracer la fonction de répartition empirique (ecdf pour *empirical cumulative distribution function*) : `plot(ecdf(bebes[,3]))`.



Pour $0 \leq \alpha \leq 1$, on appelle α -quantile, le nombre x_α tel que $F_n(x_\alpha) \geq \alpha$ et $F_n(x) < \alpha$ pour $x < x_\alpha$.

Les quantiles correspondants à $\alpha = \frac{1}{2}$, $\alpha = \frac{1}{4}$, $\alpha = \frac{3}{4}$ s'appellent respectivement *médiane* (Me), *premier quartile* q_1 et *troisième quartile* q_3 . Sous R, on dispose de la fonction `quantile`.

L'écart inter quartile $E = q_3 - q_1$ mesure la dispersion de la variable autour de la médiane.

Remarque 2.1 Dans le cas d'une variable quantitative continue, les données peuvent être regroupées par classes : $[h_i, h_{i+1}[$, dans ce cas la fonction de répartition est la fonction de répartition de la mesure

$$\mu_n = \frac{1}{n} \sum \frac{n_i}{h_{i+1} - h_i} \mathbb{I}_{[h_i, h_{i+1}[} \lambda,$$

où λ désigne la mesure de Lebesgue. Cela revient à considérer que dans chaque classe, la répartition est uniforme. Le α -quantile est alors le nombre x_α tel que $F_n(x_\alpha) = \alpha$.

Si on ne connaît que la répartition en classe, pour calculer les moyenne, variance ..., on remplace les x_i par les centres des classes $c_i = \frac{h_{i+1} + h_i}{2}$.

III Résumés graphiques

Variables qualitatives

On parle de variables qualitatives lorsque les x_i ne sont pas des nombres (catégories socio-professionnelles, couleurs des yeux ...), lorsque les x_i sont des nombres mais que le nombre de valeurs différentes est très faible (moins de 5 ou 6 valeurs différentes), on peut aussi traiter la variable concernée comme une variable qualitative.

Les fréquences sont définies comme ci-dessus. On peut représenter ces variables de différentes façons : diagrammes en secteurs, diagrammes en colonnes, diagrammes en barres. Le principe de ces représentations est que l'aire est proportionnelle à l'effectif.

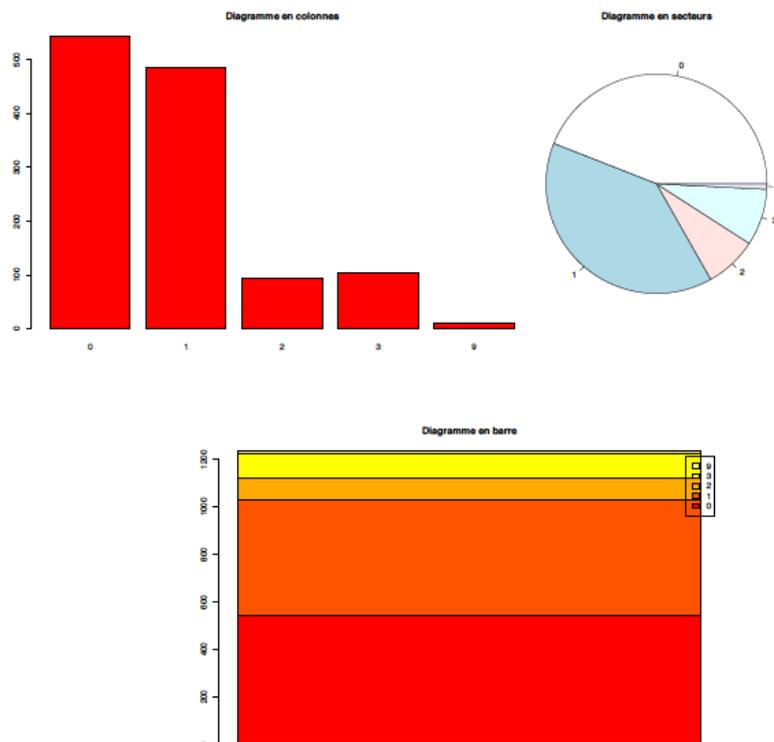
La commande R `table` permet de calculer les effectifs de chaque modalité.

```
tableb <- table(bebes[,7])
```

`barplot(tableb, main='Diagramme en colonnes')` pour le diagramme en colonnes.

`barplot(as.matrix(tableb), main='Diagramme en barres')` pour le diagramme en barres.

`pie(tableb, main='Diagramme en secteurs')` pour le diagramme en secteurs.



Variables quantitatives

Pour les variables quantitatives, on représente :

- les fréquences (diagramme en bâton ou histogramme si la variable est continue), `hist(bebes[,1],nclass=8)`.
- les fréquences cumulées : courbe représentative de la fonction de répartition ou polygone des fréquences cumulées si la variable est continue,
- la médiane et les quartiles : boîte à moustaches (commande `boxplot`).

Concentration : courbe de Lorenz et indice de Gini

La courbe de concentration ou courbe de Lorenz et l'indice de Gini sont deux indicateurs des inégalités de répartition d'une variable quantitative X à valeurs positives (revenu, chiffre d'affaire, capital patrimonial ...).

Soit $F(x)$ la proportion d'individus pour lesquels la variable X est inférieure à x (fréquence cumulée) et $G(x)$ la proportion de la masse totale correspondante. Pour une série statistique (n_i, x_i) on a :

$$F(x) = \frac{n_1 + \dots + n_i}{n} \text{ ssi } x_i \leq x < x_{i+1},$$

$$G(x) = \frac{n_1 x_1 + \dots + n_i x_i}{M} \text{ ssi } x_i \leq x < x_{i+1},$$

où $M = \sum_{i=1}^k n_i x_i$ désigne la masse totale.

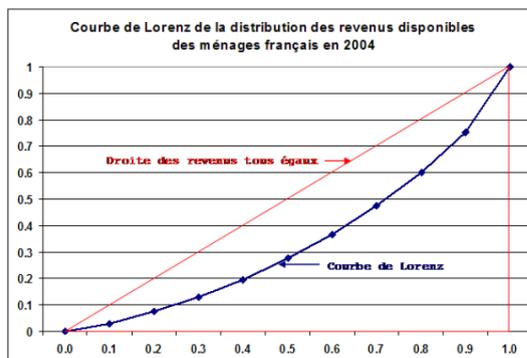
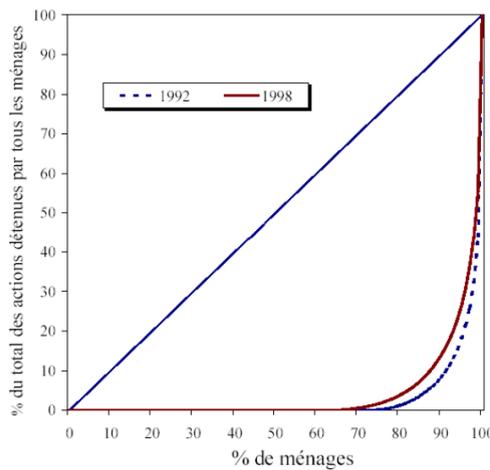
La courbe de Lorentz est la courbe des points $(F(x), G(x))$. Le fait que les x_i soient ordonnés implique que pour tout x , $F(x) \geq G(x)$. Ainsi, la courbe de Lorentz se situe en-dessous de la première bissectrice. Plus elle est proche de la première bissectrice, plus la répartition est égalitaire.

La médiane \mathcal{M} est l'inf des x_i tel que $G(x_i) \geq \frac{1}{2}$ ou l'inf des x tel que $G(x) = \frac{1}{2}$ dans le cas continu. Comme $F(x) \geq G(x)$, on a $\mathcal{M} \geq \text{Me}$.

L'indice de Gini : γ est le double de la surface comprise entre la courbe et la première bissectrice. On a :

$$0 \leq \gamma \leq 1 \text{ et}$$

$$\gamma = 1 - \sum_{i=1}^k f_i \times (G(x_i) + G(x_{i+1})).$$



IV Exercices

Exercice 1

1. Considérons une variable statistique quantitative x prenant les valeurs x_1, \dots, x_p avec fréquence f_1, \dots, f_p . Pour tout réel a , soit : $f(a) = \sum_{i=1}^p f_i(x_i - a)^2$.

Montrer que f admet un unique minimum. Pour quelle valeur de a ce minimum est-il atteint ? Exprimer $f(a)$ en fonction de $\sigma^2(x)$ et de $m(x) - a$.

2. Étant donnée $d(\cdot, \cdot)$ une fonction d'écart sur \mathbb{R}^2 (typiquement $d(x, y) = (x - y)^2$ ou $d(x, y) = |x - y|$), pour un réel a , soit $D = \sum_{i=1}^p f_i d(a, x_i)$. D définit une mesure de dispersion des observation autour de a . La fonction $d(\cdot, \cdot)$ étant choisie, on choisit a tel que D soit minimal.

Que valent a puis D dans le cas où $d(x, y) = (x - y)^2$? dans le cas où $d(x, y) = |x - y|$.

Exercice 2 Expliciter la fonction de répartition F et les quantiles dans le cas d'une variable continue répartie en classes.

Exercice 3 La série statistique suivante donne le nombre de bons de commande enregistrés chaque jour par une entreprise pendant un mois : 30, 26, 26, 32, 31, 29, 27, 27, 28, 30, 31, 27, 29, 30, 28, 26, 26, 32, 31, 30.

1. Regrouper ces données dans un tableau donnant les effectifs, fréquences et fréquences cumulées.

2. Représenter graphiquement les fréquences et les fréquences cumulées.

3. Déterminer la médiane et les premier et troisième quartiles de la série.

4. Calculer la moyenne, la variance et l'écart-type.

5. Représenter la boîte à moustache de cette distribution statistique.

Exercice 4 Une maison d'édition confie la frappe de ses manuscrits à une entreprise spécialisée. Une fois la saisie effectuée, le manuscrit est renvoyé à l'auteur pour correction des fautes de frappe. On a extrait 315 parties d'un document, toutes les parties contiennent le même nombre de caractères. Pour chacune de ces parties, on compte le nombre de fautes de frappe :

nombre de fautes de frappe	[70, 80[[80, 85[[85, 90[[90, 95[[95, 100[[100, 105[[105, 115[
Nombre de parties	16	32	63	113	47	25	19

(on a donc 32 parties dans lesquelles il y a entre 80 et 85 fautes de frappe).

1. Représenter graphiquement les fréquences et les fréquences cumulées en supposant qu'à l'intérieur de chaque classe, la distribution est uniforme.

2. Calculer la moyenne, la variance et l'écart-type.

3. Déterminer la classe médiane. En supposant qu'à l'intérieur de chaque classe, la distribution est uniforme, déterminer la médiane.

4. Même question pour les premier et troisième quartiles.

5. Représenter la boîte à moustache de cette distribution.

Chapitre 3

Echantillonnage

I Modèle statistique

1 Généralités

Soit $X = (X_1, \dots, X_n)$, un vecteur de n variables aléatoires X_i .

$$X_i : (\Omega, \mathcal{B}) \longrightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}}) \quad X : (\Omega, \mathcal{B}, \mathbb{P}) \longrightarrow (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, Q_X).$$

Q_X est la loi de probabilité image du vecteur aléatoire X .

Généralement, on ne connaît pas la loi Q_X . On souhaite ici étudier le plus précisément possible cette loi de probabilité.

Hypothèse fondamentale.

On va supposer que la loi de probabilité Q_X appartient à une famille de lois de probabilité \mathbb{Q} .

2 Modèle d'échantillonnage.

On appelle *modèle statistique* (ou structure statistique) le triplet $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, \mathbb{Q})$, c'est à dire la donnée d'une famille de lois de probabilité \mathbb{Q} à laquelle on astreint Q_X à appartenir.

Remarque.

La connaissance du phénomène étudié permet d'avoir une idée pour le choix de la famille de lois de probabilité \mathbb{Q} : cette connaissance peut provenir de l'étude de la série statistique (x_1, \dots, x_n) , de l'expérience du statisticien, ...

Une série statistique (x_1, \dots, x_n) est vue comme une réalisation du vecteur aléatoire $X : (x_1, \dots, x_n) = (X_1(\omega), \dots, X_n(\omega)), \omega \in \Omega$.

Le modèle d'échantillonnage indépendant est un cas particulier du modèle statistique général et sera utilisé presque systématiquement dans la suite de ce cours.

Il consiste à supposer que les variables aléatoires X_1, \dots, X_n sont *indépendantes* et *identiquement distribuées* (i.i.d.).

On a donc les propriétés suivante :

- les variables aléatoires X_i étant identiquement distribuées, elles ont même loi Q_{X_i} , ne dépend pas de i . On pose ainsi $\forall i = 1, \dots, n, Q_{X_i} = P$.
- Les variables aléatoires X_i étant indépendantes, la loi Q_X du vecteur aléatoire X est la loi produit $Q = P^{\otimes n}$. De manière un peu abusive, on notera le modèle d'échantillonnage indépendant réel de la manière suivante :

$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \mathbb{P})^n$ si la loi P appartient à la famille de mesures de probabilité \mathbb{P} .

Le choix de la famille de lois de probabilité se fait ainsi au niveau de P .

Le modèle d'échantillonnage paramétrique

Il consiste, dans le cadre de l'échantillonnage, à supposer que la famille \mathbb{P} de lois de probabilité est indicée par un paramètre θ . On note alors

$$\mathbb{P} = (P_\theta, \theta \in \Theta \subset \mathbb{R}^p),$$

avec :

- P_θ est la loi de probabilité correspondant à la valeur θ du paramètre.
- Θ est l'espace paramétrique (dans lequel θ peut prendre sa valeur).
- p est la dimension du paramètre (pour $p = 1$, on parle de paramètre unidimensionnel, pour $p > 1$, on parle de paramètre multidimensionnel ou vectoriel).

Un modèle d'échantillonnage paramétrique sera donc noté

$$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (P_\theta, \theta \in \Theta \subset \mathbb{R}^p))^n.$$

En dehors de ce cadre, on parle de modèle non paramétrique ou semi-paramétrique si $\Theta \subset \mathbb{R}^k \times U$ où U est non paramétrique.

Exemple Si on considère un modèle où P peut être n'importe quelle loi de probabilité absolument continue par rapport à la mesure de Lebesgue, on est dans un cadre non paramétrique. On a

$$\mathbb{P} = \{f\lambda / f \geq 0, \int_{\mathbb{R}} f d\lambda = 1\}.$$

Si P peut être une loi normale de paramètres $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_*^+$, on est dans un cadre paramétrique bi-dimensionnel,

$$\Theta = \{(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma > 0\}$$

3 Modèle dominé, vraisemblance

Considérons un modèle statistique $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, Q \in \mathcal{Q})$. Le modèle est *dominé* par une mesure σ -finie μ sur $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ si pour tout $Q \in \mathcal{Q}$, Q est absolument continue par rapport à μ . Dans ce cas, il existe (théorème de Radon-Nikodym) une fonction de densité f_Q telle que $Q = f_Q \mu$ (i.e $Q(A) = \int_A f_Q(x) d\mu(x)$ pour tout $A \in \mathcal{C}$). La fonction f_Q est la *vraisemblance* (Likelihood en anglais) du modèle notée $L_Q(x) = f_Q(x)$. On note $\ell_Q(x) = \log L_Q(x)$ la log vraisemblance.

Dans le cas d'un modèle discret

Les mesures Q sont des mesures discrètes, de la forme $Q = \sum_{x \in E} p_Q(x) \delta_{\{x\}}$ ($\delta_{\{x\}}$ est la mesure de Dirac en x et $\sum_{x \in E} p_Q(x) = 1$), E est l'ensemble (au plus dénombrable) des valeurs prises. Si μ est la mesure de comptage ($\mu(\{e\}) = 1$ pour tout $e \in E$), le modèle est dominé et on a $L_Q(x) = p_Q(x) = Q(X = x)$.

Modèle absolument continu (par rapport à la mesure de Lebesgue). Si les lois Q sont absolument continue par rapport à la mesure de Lebesgue de $\mathbb{R}^n : \lambda_{\mathbb{R}^n}$, la vraisemblance L_Q est simplement la densité de Q .

Dans le cas d'un modèle d'échantillonnage réel, on a le modèle $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (P_\theta, \theta \in \Theta \subset \mathbb{R}^p))^n$, le modèle est dominé si et seulement si le modèle unidimensionnel est dominé. Dans ce cas, si $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, P_\theta)$ est dominé par la mesure μ alors le modèle $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (P_\theta, \theta \in \Theta \subset \mathbb{R}^p))^n$ est dominé par $\mu^{\otimes n}$. La vraisemblance $L_\theta(x_1, \dots, x_n) = f_\theta(x_1) \times \dots \times f_\theta(x_n)$ est le produit des vraisemblances du modèle $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (P_\theta, \theta \in \Theta))$ (propriété des mesures produit).

Interprétation : le terme de vraisemblance s'interprète en remarquant que plus $L_\theta(x)$ est grand, plus la probabilité d'observer x est grande.

II Définition d'une statistique

C'est une notion fondamentale pour la suite. Reprenons le modèle statistique réel $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, \mathcal{Q})$.

— On appelle *statistique* toute variable aléatoire définie sur $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ à valeurs dans $(\mathbb{R}^k, \mathcal{B}_{\mathbb{R}^k})$.

L'idée est que les statistiques intervenant dans le modèle paramétrique permettront de "préciser" le paramètre $\theta \in \mathbb{R}^p$. Le plus souvent, on aura $k = p$.

— On utilisera la notation T_n pour une statistique. Elle ne dépend pas de la famille de lois de probabilité \mathcal{Q} . En particulier, dans un modèle paramétrique, une statistique T_n ne dépend pas de θ . C'est la loi de probabilité de T_n qui dépend de θ .

Exemple.

Prenons $P_\theta = \mathcal{N}(\mu, 1)$. On a ici $\theta = \mu$ (avec $p = 1$). Une statistique permettant de “préciser” le paramètre θ (moyenne d’une loi normale réduite) peut être naturellement la statistique “moyenne empirique” :

$$T_n : (\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (P_\theta, \theta \in \Theta \subset \mathbb{R}^p))^n \longrightarrow (\mathbb{R}^k, \mathcal{B}_{\mathbb{R}^k}) \\ (x_1, \dots, x_n) \longmapsto \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}_n.$$

Remarque.

On distingue :

- la statistique T_n qui est une variable aléatoire sur $(\mathbb{R}^n, \mathcal{B}, \mathbb{Q})$. Dans l’exemple précédent, c’est l’application moyenne empirique (on est ici au niveau conceptuel, mathématique).
- la variable aléatoire $T_n(X_1, \dots, X_n)$ qui est une variable aléatoire sur (Ω, \mathcal{B}) . Dans l’exemple précédent, c’est la variable aléatoire \bar{X}_n (on est ici au niveau de la statistique inférentielle).
- la valeur observée de cette variable aléatoire : $T_n(x_1, \dots, x_n) \in \mathbb{R}^k$. Dans l’exemple précédent, c’est le nombre réel \bar{x}_n , moyenne empirique de la série statistique x_1, \dots, x_n (on est au niveau de la statistique descriptive).

1 Notions d’estimation et de test d’hypothèses

Prenons le modèle le plus usuel, à savoir le modèle d’échantillonnage indépendant paramétrique réel

$$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (P_\theta, \theta \in \Theta \subset \mathbb{R}^p))^n.$$

Si le paramètre θ est connu, alors la loi de probabilité P_θ est connue.

Les méthodes standards de la statistique inférentielle (estimation et test d’hypothèse) ont pour objectif de préciser au maximum le paramètre θ quand ce dernier est inconnu.

Estimation.

Ce sont des méthodes permettant de fixer une valeur (on parle d’*estimation ponctuelle*) ou un ensemble de valeurs (on parle d’*estimation par intervalle de confiance*) pour le paramètre θ .

Test d’hypothèses.

On appelle *hypothèse statistique* l’énoncé d’une propriété relative à la nature d’une ou plusieurs distributions. On distingue généralement les tests paramétriques et les tests non paramétriques. Les premiers permettent par exemple d’étudier si une valeur fixée (ou bien un ensemble de valeurs) est acceptable ou non pour le paramètre θ, \dots

Les seconds permettent par exemple de savoir si une famille donnée de lois est acceptable ou pas (on parle de test d'ajustement), si deux variables sont indépendantes (on parle de test d'indépendance), ...

III Quelques notions de base sur les estimateurs

On considère dans cette section un modèle paramétrique réel d'échantillonnage indépendant :

$$(\mathbb{R}, \mathcal{B}, (P_\theta, \theta \in \Theta \subset \mathbb{R}^p))^n.$$

1 Définition d'un estimateur

Soit g une fonction définie sur Θ , à valeur dans \mathbb{R}^p .

- On appelle *estimateur d'un paramètre* $g(\theta)$ toute statistique à valeur dans $(\mathbb{R}^k, \mathcal{B}_{\mathbb{R}^k})$, il sera généralement noté T_n .
- On appelle *estimation* la valeur observée de T_n sur l'échantillon, elle sera notée $T_n(x_1, \dots, x_n)$.

2 Notion de biais

- On appelle *biais* de l'estimateur T_n pour le paramètre $g(\theta)$ la quantité

$$B_\theta(T_n) = \mathbb{E}_\theta[T_n] - g(\theta).$$

C'est une fonction de θ .

- On appelle *estimateur sans biais de* $g(\theta)$ un estimateur T_n tel que $B(T_n) = 0$, sinon on parle d'estimateur biaisé.
- Si l'estimateur T_n est biaisé, mais que $B(T_n) \rightarrow 0$ quand $n \rightarrow +\infty$, on dit que T_n est *asymptotiquement sans biais* pour $g(\theta)$.

3 Convergence d'un estimateur

On dit que T_n est *convergent* pour $g(\theta)$ s'il converge en probabilité vers $g(\theta)$:

$$\forall \varepsilon > 0, P(|T_n - g(\theta)| < \varepsilon) \rightarrow 1 \text{ quand } n \rightarrow +\infty.$$

Critères de convergence d'un estimateur

- (i) Si T_n est un estimateur sans biais de $g(\theta)$ et si $\mathbb{V}(T_n) \rightarrow 0$ quand $n \rightarrow +\infty$, alors T_n est un estimateur convergent pour $g(\theta)$.
- (ii) Si T_n est un estimateur asymptotiquement sans biais de $g(\theta)$ et si $\mathbb{V}(T_n) \rightarrow 0$ quand $n \rightarrow +\infty$, alors T_n est un estimateur convergent pour $g(\theta)$.

Exercice. Démontrer les résultats énoncés ci-dessus.

4 Comparaisons des estimateurs

On utilise le *risque quadratique* pour comparer deux estimateurs du même paramètre $g(\theta)$. Pour l'estimateur T_n de $g(\theta)$, il est défini par :

$$R(T_n, g(\theta)) = \mathbb{E}[(T_n - g(\theta))^2].$$

Propriété. $R(T_n, g(\theta)) = (B(T_n))^2 + \mathbb{V}(T_n)$.

Exercice. Démontrer la propriété précédente.

Commentaires.

- L'idée est de minimiser le risque quadratique $R(T_n, g(\theta))$, c'est à dire minimiser le biais (si possible l'annuler) ainsi que la variance de l'estimateur.
- La variance de T_n ne pourra pas descendre en dessous d'une certaine borne (voir l'inégalité de Cramer-Rao à la section suivante).
- La quantité $R(T_n, g(\theta))$ est aussi appelé *erreur quadratique moyenne*.
- Le risque quadratique est le critère de choix généralement utilisé entre deux estimateurs.

5 Moyenne aléatoire, variance aléatoire

Soit un modèle paramétrique réel d'échantillonnage

$$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (P_{\theta}, \theta \in \Theta \subset \mathbb{R}^p))^n$$

tel que la loi de probabilité P_{θ} admette pour espérance $\mu < \infty$ et pour variance $0 < \sigma^2 < \infty$.

Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire associé de ce modèle.

On a deux estimateurs "naturels" pour l'espérance et la variance d'une variable aléatoire. Ils sont inspirés de la moyenne et de la variance d'une série statistique. On appelle *moyenne aléatoire* la statistique

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

On appelle *variance aléatoire* la statistique

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

on verra ci-dessous que cet estimateur de la variance est biaisé, on lui préférera la *variance estimée* :

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

Proposition 3.1 1. \bar{X}_n est un estimateur sans biais et convergent pour μ .
 2. V_n est un estimateur biaisé mais asymptotiquement sans biais de σ^2 .

$$3. S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \text{ est un estimateur sans biais de } \sigma^2.$$

4. V_n et S_n^2 sont des estimateurs convergent de σ^2 .

On note mc_3 et mc_4 les moments centrés d'ordre 3 et 4 des v.a.i.i.d. X_i .

On peut calculer (c'est un peu fastidieux pour les moments moments d'ordre 2 de V_n et S_n^2) l'espérance et la variance de \bar{X}_n , V_n , S_n^2 .

Proposition 3.2 — $\mathbb{E}[\bar{X}_n] = \mu$; $\mathbb{V}(\bar{X}_n) = \frac{\sigma^2}{n}$.

$$— \mathbb{E}[S_n^2] = \sigma^2; \mathbb{V}(S_n^2) = \frac{mc_4}{n} - \frac{n-3}{n(n-1)}\sigma^4,$$

avec $mc_4 = \mathbb{E}[(X_i - \mu)^4]$ et $\sigma^4 = (\sigma^2)^2$.

$$— \mathbb{E}[V_n^2] = \frac{n-1}{n}\sigma^2; \mathbb{V}(V_n^2) = \frac{(n-1)^2}{n^3}mc_4 - \frac{(n-1)(n-3)}{n^2}\sigma^4.$$

$$— \text{cov}(\bar{X}_n, S_n^2) = \frac{mc_3}{n},$$

avec $mc_3 = \mathbb{E}[(X_i - \mu)^3]$.

$$— \text{cov}(\bar{X}_n, V_n^2) = \frac{n-1}{n^2}mc_3.$$

Si l'on suppose en plus que chaque variable aléatoire réelle X_i suit la loi normale $\mathcal{N}(\mu, \sigma^2)$, alors on a :

$$— \mathbb{V}(S_n^2) = \frac{2}{n-1}\sigma^4.$$

$$— \mathbb{V}(V_n) = \frac{2(n-1)}{n^2}\sigma^4.$$

$$— \text{cov}(\bar{X}_n, S_n^2) = \text{cov}(\bar{X}_n, V_n) = 0.$$

IV Rappels sur les vecteurs gaussiens

Définition 3.1 Un vecteur aléatoire $X = (X_1, \dots, X_n)$ de \mathbb{R}^n est un vecteur gaussien centré et réduit si c'est un vecteur aléatoire absolument continu par rapport à la mesure de Lebesgue de \mathbb{R}^n , de densité :

$$\frac{1}{\sqrt{2\pi}^n} e^{-\frac{\|x\|^2}{2}}, \text{ avec } \|x\|^2 = \sum_{i=1}^n x_i^2.$$

C'est équivalent au fait que les X_i sont indépendants de loi $\mathcal{N}(0, 1)$. On notera $X \sim \mathcal{N}(0, \mathbb{I}_n)$ où \mathbb{I}_n désigne la matrice identité sur \mathbb{R}^n .

Un vecteur aléatoire $X = (X_1, \dots, X_n)$ de \mathbb{R}^n est un vecteur gaussien s'il existe une matrice $n \times n$, A et un vecteur $b \in \mathbb{R}^n$ tels que $X = AX_0 + b$ avec $X_0 \sim \mathcal{N}(0, \mathbb{I}_n)$.

Si $X = AX_0 + b$ est un vecteur gaussien avec $X_0 \sim \mathcal{N}(0, \mathbb{I}_n)$, on a :

$$— \mathbb{E}(X) = b,$$

$$— \text{cov}(X) = AA^t$$

Proposition 3.3 Soit $X = (X_1, \dots, X_n)$ un vecteur gaussien de \mathbb{R}^n . Les X_i sont indépendants si et seulement si pour tout $i \neq j$, $\text{cov}(X_i, X_j) = 0$.

Proposition 3.4 Un vecteur aléatoire $X = (X_1, \dots, X_n)$ de \mathbb{R}^n est un vecteur gaussien si et seulement si toute combinaison linéaire de ses coordonnées $Z_a = \sum_{i=1}^n a_i X_i$ suit une loi normale ou sa loi est une mesure de Dirac.

Proposition 3.5 Si A est une matrice orthogonale et $X = AX_0$, $X_0 \sim \mathcal{N}(0, \mathbb{I}_n)$ alors $X \sim \mathcal{N}(0, \mathbb{I}_n)$.

Définition 3.2 Une variable aléatoire réelle est dite variable aléatoire gamma, de paramètres a et λ ($a > 0$, $\lambda > 0$), et notée $G(a, \lambda)$, si sa loi a la densité $\gamma_{a,\lambda}$ définie par

$$\left[\gamma_{a,\lambda}(x) = \frac{\lambda^a}{\Gamma(a)} e^{-\lambda x} x^{a-1} \mathbb{I}_{\mathbb{R}^+}(x) \right]$$

où Γ est la fonction Gamma définie pour tout $a > 0$ par

$$\Gamma(a) = \int_0^{+\infty} e^{-x} x^{a-1} dx$$

La fonction Gamma vérifie : pour tout $a > 0$, $\Gamma(a+1) = a\Gamma(a)$, en particulier, pour $n \in \mathbb{N}$, $\Gamma(n+1) = n!$.

Les lois Gamma vérifient les propriétés suivantes.

Proposition 3.6

1. Si X suit une loi $G(a, \lambda)$, $\mathbb{E}[X] = \frac{a}{\lambda}$ et $\mathbb{V}(X) = \frac{a}{\lambda^2}$,
2. Soit X et Y deux variables aléatoires indépendantes de lois respectives $G(a, \lambda)$ et $G(b, \lambda)$ alors les variables aléatoires $S = X + Y$ et $T = \frac{X}{X+Y}$ sont indépendantes. S suit une loi $G(a+b, \lambda)$. La loi de T est appelée loi bêta sur $]0, 1[$ de paramètres a et b .

Définition 3.3 On dit qu'une variable aléatoire Z suit une loi du χ^2 à n degrés de liberté notée $\chi^2(n)$, si

$$Z = \sum_{i=1}^n X_i^2$$

où les X_i sont des variables aléatoires indépendantes de loi $\mathcal{N}(0, 1)$.

Proposition 3.7 Une loi du χ^2 à n degrés de liberté est une loi $G(\frac{n}{2}, \frac{1}{2})$.

Définition 3.4 Une variable aléatoire T suit une loi de Student à n degrés de liberté si :

$$T = \frac{X}{\sqrt{\frac{Z}{n}}}$$

où X et Z sont des variables aléatoires indépendantes de loi $\mathcal{N}(0, 1)$ et $\chi^2(n)$ respectivement.

Une variable aléatoire Z suit une loi de Fisher-Snedecor à (n_1, n_2) degré de liberté notée $F(n_1, n_2)$ si

$$Z = \frac{\frac{Z_1}{n_1}}{\frac{Z_2}{n_2}}$$

où Z_1 et Z_2 sont des variables aléatoires indépendantes de loi $\chi^2(n_1)$ et $\chi^2(n_2)$ respectivement.

Théorème 3.1 (Théorème de Cochran) Soit $X \sim \mathcal{N}(0, \mathbb{I}_n)$, A une matrice orthogonale et $Y = AX$. Pour $k = 1, \dots, n$, soit

$$Z_k = \sum_{i=1}^n X_i^2 - \sum_{i=1}^k Y_i^2.$$

Alors Z_k est indépendant de (Y_1, \dots, Y_k) et suit une loi du $\chi^2(n - k)$.

Preuve:

Comme A est une matrice orthogonale, Y est un vecteur gaussien centré et réduit et on a $\|X\|^2 = \|Y\|^2$, autrement dit,

$$\sum_{i=1}^n X_i^2 = \sum_{i=1}^n Y_i^2$$

ainsi,

$$Z_k = \sum_{i=k+1}^n Y_i^2$$

d'où le résultat annoncé. □

1 Vecteurs gaussiens et lois du χ^2 .

Le résultat suivant relie les vecteurs gaussiens aux lois du χ^2 , ce genre de résultat est essentiel en statistique, notamment dans des problèmes de décomposition de la variance.

Théorème 3.2 Soit X un vecteur gaussien non dégénéré de \mathbb{R}^d , $X \rightsquigarrow \mathcal{N}_d(\mu, \Sigma)$, alors

$$D^2 = (X - \mu)' \Sigma^{-1} (X - \mu)$$

suit une loi du χ^2 à d degrés de liberté.

Preuve:

$D^2 = \langle (X - \mu), \Sigma^{-1}(X - \mu) \rangle$. X est un vecteur gaussien donc il existe $X_0 \rightsquigarrow \mathcal{N}_d(0, \mathbb{I})$ et une matrice A tels que $\Sigma = AA'$ et $X = AX_0 + \mu$. On a alors

$$D^2 = \|A^{-1}(X - \mu)\|^2 = \|X_0\|^2.$$

Autrement dit, D^2 apparaît comme la somme de d carrés de loi normales centrées et réduites et indépendantes, et donc $D^2 \rightsquigarrow \chi^2(d)$. □

V Application à l'estimation dans un cadre normal

Le théorème de Cochran permet de montrer les propriétés fondamentales des estimateurs de la moyenne et de la variance dans le cas gaussien. On suppose que chaque variable aléatoire réelle X_i suit la loi normale $\mathcal{N}(\mu, \sigma^2)$, alors on a :

Théorème 3.3

$$\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2(n-1).$$

Les statistiques \bar{X}_n et S_n^2 sont indépendantes.

Preuve:

Il est clair que $\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ (propriété élémentaire d'additivité des lois normales). Les deux autres propriétés sont une applications directe du théorème de Cochran.

On a

$$\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X}_n - \mu)^2,$$

ainsi,

$$\frac{(n-1)S_n^2}{\sigma^2} = \frac{nV_n}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 - \left(\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}\right)^2.$$

On a $Z_1 = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0, 1)$, $Y_i = \frac{X_i - \mu}{\sigma} \sim \mathcal{N}(0, 1)$ et $Z_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i$.

Les Y_i sont indépendantes, ainsi $Y = (Y_1, \dots, Y_n) \sim \mathcal{N}(0, \mathbb{I}_n)$. Le vecteur $u_1 = (\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}})^t$ de \mathbb{R}^n est de norme 1. On le complète en une base orthonormée (u_1, \dots, u_n) de \mathbb{R}^n . On note A la matrice de passage de (u_1, \dots, u_n) à la base canonique de \mathbb{R}^n . On peut ainsi appliquer le théorème de Cochran.

□

Chapitre 4

Estimation

- On reste dans le cadre d'un modèle paramétrique réel d'échantillonnage, on suppose que ce modèle est dominé par une mesure μ , on considère le cas d'un paramètre unidimensionnel ($\Theta \subset \mathbb{R}$).

$$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (P_{\theta}, \theta \in \Theta \subset \mathbb{R}))^n.$$

- Soit $f(x, \theta)$ la densité de la loi des X_i (ou la loi de probabilité discrète dans le cas discret).
- Soit L la vraisemblance. On a :
 $L(x_1, \dots, x_n, \theta) = \prod_{i=1}^n f(x_i, \theta)$ car on est dans le cadre d'un modèle d'échantillonnage.
- Soit $\mathcal{X} \subset \mathbb{R}$ le support de $f(x, \theta)$ où

$$\mathcal{X} = \{x \in \mathbb{R} \mid f(x, \theta) > 0\}.$$

Notons que \mathcal{X} peut dépendre ou non de θ .

Le support de L est alors \mathcal{X}^n .

I Hypothèses fondamentales sur la densité $f(x, \theta)$

On est amené à faire un certain nombre d'*hypothèses de régularité* sur f pour pouvoir établir certaines propriétés.

(H1) \mathcal{X} est indépendant de θ .

(H2) Θ est un ouvert.

Le fait d'avoir aussi $f(x, \theta) > 0, \forall (x, \theta) \in \mathcal{X} \times \Theta$, va permettre de ne pas avoir de problèmes de différentiabilité aux frontières.

(H3) $\frac{\partial}{\partial \theta} f(x, \theta)$ et $\frac{\partial^2}{\partial \theta^2} f(x, \theta)$ sont définies $\forall (x, \theta) \in \mathcal{X} \times \Theta$.

(H4) On suppose que les dérivés et dérivés secondes de f par rapport à θ sont dominées par des fonctions μ -intégrables sur tout compact inclu dans Θ : pour tout compact $K \subset \Theta$, il existe deux fonctions positive μ -intégrables $\phi(x)$ et $\psi(x)$ telles que pour tout $\theta \in K$ et presque tout $x \in \mathcal{X}$,

$$\left| \frac{\partial}{\partial \theta} f(x, \theta) \right| \leq \phi(x) \text{ et } \left| \frac{\partial^2}{\partial \theta^2} f(x, \theta) \right| \leq \psi(x).$$

Remarque.

L'hypothèse (H_4) permettra de dériver $f(x, \theta)$ au moins deux fois par rapport à θ sous le signe \int .

II Information

1 Information de Fisher

Définition 4.1 On appelle fonction score ou score la fonction S définie par :

$$\begin{aligned} S &: \mathcal{X} \times \Theta \longrightarrow \mathbb{R} \\ (x, \theta) &\longmapsto S(x, \theta) = \frac{\partial}{\partial \theta} \ln(f(x, \theta)) \end{aligned}$$

Remarques.

- Le score n'est défini que si **(H1)**, **(H2)** et **(H3)** sont vraies.
- Le score intervient très souvent en estimation.
- On peut définir de même le score à partir de la vraisemblance. On le notera alors :

$$S_n(x, \theta) = \frac{\partial}{\partial \theta} \ln(L(x, \theta)) \quad \text{avec ici } x \in \mathbb{R}^n.$$

Dans le modèle d'échantillonnage, on a : $S_n(x, \theta) = \sum_{i=1}^n S(x_i, \theta)$.

Propriétés.

Supposons **(H4)** vraie, on a alors :

$$\mathbb{E}[S(X, \theta)] = 0 \quad \text{et} \quad \mathbb{E}[S_n(X, \theta)] = 0.$$

Preuve:

C'est une conséquence du fait que $f(x, \theta)$ est une fonction de densité. Ainsi,

$$\int f(x, \theta) d\mu = 1,$$

on dérive par rapport à θ :

$$0 = \int \frac{\partial}{\partial \theta} (f(x, \theta)) d\mu = \int \frac{\partial}{\partial \theta} \ln(f(x, \theta)) \cdot f(x, \theta) d\mu = \mathbb{E}[S(X, \theta)].$$

□

Information au sens de Fisher

Définition 4.2 On appelle information de Fisher la fonction I définie par :

$$\begin{aligned} I &: \Theta \longrightarrow \mathbb{R}^+ \\ \theta &\longmapsto I(\theta) = \mathbb{E} [(S(X, \theta))^2] \end{aligned}$$

Remarques.

- L'information de Fisher I n'est définie que si les hypothèses **(H1)**, **(H2)** et **(H3)** sont vérifiées.
- L'information de Fisher I ne dépend que de θ et du modèle choisi. C'est une information contenue dans le modèle sur le paramètre θ .
- On peut aussi poser (en terme de vraisemblance) :

$$I_n(\theta) = \mathbb{E} [(S_n(X, \theta))^2].$$

2 Propriétés.

Si **(H4)** est vraie, alors

$$I(\theta) = \mathbb{V}(S(X, \theta)) = -\mathbb{E} \left[\frac{\partial^2}{\partial \theta^2} \ln(f(X, \theta)) \right] \quad \text{et} \quad I_n(\theta) = \mathbb{V}(S_n(X, \theta)) = -\mathbb{E} \left[\frac{\partial^2}{\partial \theta^2} \ln(L(X, \theta)) \right].$$

Dans le modèle d'échantillonnage, on a donc :

$$I_n(\theta) = nI(\theta).$$

Preuve:

On reprend la preuve précédente :

$$0 = \int \frac{\partial}{\partial \theta} \ln(f(x, \theta)) \cdot f(x, \theta) d\mu,$$

on dérive à nouveau par rapport à θ :

$$\begin{aligned} 0 &= \int \frac{\partial^2}{\partial \theta^2} \ln(f(x, \theta)) \cdot f(x, \theta) d\mu + \int \frac{\partial}{\partial \theta} \ln(f(x, \theta)) \cdot \frac{\partial}{\partial \theta} f(x, \theta) d\mu \\ &= \int \frac{\partial^2}{\partial \theta^2} \ln(f(x, \theta)) \cdot f(x, \theta) d\mu + \int \left(\frac{\partial}{\partial \theta} \ln(f(x, \theta)) \right)^2 \cdot f(x, \theta) d\mu. \end{aligned}$$

Ce qui donne le résultat pour $I(\theta)$.

□

III Inégalité de Cramer-Rao

1 Hypothèses supplémentaires

On a ici besoin d'une hypothèse supplémentaire :

(H5) $0 < I_n(\theta) < +\infty$ ou $0 < I(\theta) < +\infty$.

Soit T_n un estimateur. Posons $\mathbb{E}_\theta[T_n] = g(\theta)$. Le résultat fondamental suivant donne une borne sur la variance de T_n .

Théorème 4.1 (Inégalité de Cramer-Rao.) *Si les hypothèses (H1), (H2), (H3), (H4) et (H5) sont vérifiées, si de plus la fonction g est dérivable, alors on a :*

$$\mathbb{V}(T_n) \geq \frac{[g'(\theta)]^2}{I_n(\theta)}.$$

Preuve:

L'ingrédient essentiel de la preuve de l'inégalité de Cramer-Rao est l'inégalité de Cauchy-Schwartz. Comme le score est une variable centrée,

$$\begin{aligned} \text{cov} \left(T, \frac{\partial}{\partial \theta} \ln L \right) &= \mathbb{E} \left(T \cdot \frac{\partial}{\partial \theta} \ln L \right) \\ &= \int T \cdot \frac{\partial}{\partial \theta} \ln L \cdot L d\mu^{\otimes n} \\ &= \int \frac{\partial}{\partial \theta} T L d\mu^{\otimes n} \\ &= \frac{\partial}{\partial \theta} \int T L d\mu^{\otimes n} = \frac{\partial}{\partial \theta} \mathbb{E}(T) = g'(\theta). \end{aligned}$$

Par l'inégalité de Cauchy-Schwarz, on a :

$$\left[\text{cov} \left(T, \frac{\partial}{\partial \theta} \ln L \right) \right]^2 \leq V(T) V \left(\frac{\partial}{\partial \theta} \ln L \right),$$

ce qui termine la preuve. □

Commentaires.

- La partie de droite est appelée *borne inférieure de l'inégalité de Cramer-Rao*. Nous la noterons $K_{T_n}(\theta)$.
- L'estimateur T_n intervient au numérateur.

Au dénominateur, seul le modèle intervient.

- Dans le cas particulier des estimateurs sans biais (soit $\mathbb{E}(T_n) = \theta$), on a :

$$\mathbb{V}(T_n) \geq I_n^{-1}(\theta).$$

Définition 4.3 — *On dit que l'estimateur T_n est efficace s'il vérifie $\mathbb{V}(T_n) = K_{T_n}(\theta)$.*

Notons que cette notion dépend de l'estimateur et du modèle.

- *Si T_n n'est pas efficace mais que $\frac{K_{T_n}(\theta)}{\mathbb{V}(T_n)} \rightarrow 1$ quand $n \rightarrow +\infty$, on dit que l'estimateur T_n est asymptotiquement efficace.*

2 Relation entre estimateurs efficaces

Propriétés.

(P1) Si T_n est un estimateur efficace de θ , alors $kT_n + b$ est aussi un estimateur efficace de θ , $\forall k \in \mathbb{R}^*$, $\forall b \in \mathbb{R}$.

(P2) Soient T_{1n} et T_{2n} deux estimateurs sans biais du paramètre θ . S'ils sont tous les deux efficaces, alors $T_{1n} = T_{2n}$ presque sûrement.

Remarques.

- Si T_n est un estimateur efficace de θ , il apparaît alors intéressant de chercher dans les estimateurs de la forme kT_n celui qui minimise (ou annule) le biais. Par contre, on prend toujours $b = 0$: en effet, la constante b ne dépend pas de X (donc de l'expérience), ni de θ (sinon cela ne serait plus un estimateur).
- Pour l'instant, on ne peut rien dire de deux estimateurs efficaces d'un même paramètre qui n'auraient pas le même biais.

Exercice. Démontrer la propriété **(P1)**.

Preuve de la propriété (P2): Considérons deux estimateurs T_1 et T_2 sans biais et efficaces d'un paramètre θ . Soit $T_3 = \frac{T_1 + T_2}{2}$, c'est aussi un estimateur sans biais de θ . De plus, $V(T_3) = \frac{1}{4}[V(T_1) + V(T_2) + 2\rho\sigma_1\sigma_2]$ où ρ le coefficient de corrélation linéaire, σ_i est l'écart-type de T_i . T_1 et T_2 ont la même variance V , ainsi $V(T_3) = \frac{V}{2}(1 + \rho)$. On ne peut pas avoir $\rho < 1$ (sinon $V(T_3) < V$), donc $\rho = 1$, ce qui implique $(T_1 - \mathbb{E}(T_1)) = \alpha(T_2 - \mathbb{E}(T_2))$ presque partout, pour un $\alpha \in \mathbb{R}_*^+$. Comme $V(T_1) = V(T_2)$, il vient $\alpha = 1$ puis comme $\mathbb{E}(T_1) = \mathbb{E}(T_2)$, $T_1 = T_2$ presque sûrement. □

La recherche d'estimateurs sans biais de variance minimale est intimement liée à l'existence de statistiques exhaustives.

3 Dégradation de l'information

Si $h(t, \theta)$ est la vraisemblance de T , on note $I_T(\theta)$ l'information relative à T :

$$I_T(\theta) = \mathbb{E}(S(T, \theta)^2) \text{ où } S(T, \theta) = \frac{\partial \ln h(t, \theta)}{\partial \theta}.$$

$I_T(\theta)$ vérifie les mêmes propriétés (centrée, relation avec $\frac{\partial^2}{\partial \theta^2} h$, inégalité de Cramer-Rao) que $I(\theta)$.

Proposition 4.1 *Soit T une statistique, on a $I_T(\theta) \leq I_n(\theta)$ avec égalité si et seulement si la statistique T est exhaustive pour le paramètre θ .*

Preuve:

La preuve repose sur le fait que :

$$L(x_1, \dots, x_n, \theta) = h(t, \theta) \cdot k(x_1, \dots, x_n, \theta|t),$$

où $h(t, \theta)$ est la vraisemblance de T et $k(x_1, \dots, x_n, \theta|t)$ la vraisemblance conditionnelle de (X_1, \dots, X_n) sachant T .

□

On voit dans le résultat ci-dessus que l'information est préservée si la loi conditionnelle de (X_1, \dots, X_n) sachant T ne dépend pas de θ .

IV Notion d'exhaustivité

Dans un problème statistique où figure un paramètre θ inconnu, un échantillon apporte une certaine information sur ce paramètre. Lorsque l'on résume cet échantillon avec une statistique, il s'agit de ne pas perdre d'information. Une statistique qui conserve l'information est une statistique exhaustive. Autrement dit, si la loi conditionnelle de (X_1, \dots, X_n) à T ne dépend plus de θ , cela signifie que toute l'information sur θ est contenue dans T .

Formellement, on est dans le cadre d'un modèle paramétrique réel d'échantillonnage

$$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (P_{\theta}, \theta \in \Theta \subset \mathbb{R}^p))^n.$$

Définition 4.4 (Principe de factorisation) La statistique T_n est ici une variable aléatoire définie sur $(\mathcal{X}, \mathcal{B}_{\mathcal{X}}, (P_{\theta}, \theta \in \Theta \subset \mathbb{R}))^n$ à valeurs dans $(\Theta, \mathcal{B}_{\Theta})$.

On considère :

- la vraisemblance (fonction de densité ou probabilité) de T_n que l'on va noter $h(t, \theta)$,
- la densité conjointe de l'échantillon (X_1, \dots, X_n) notée $L(x_1, \dots, x_n, \theta)$,

On dira que la statistique T_n est exhaustive pour θ s'il existe une fonction k sur \mathcal{X}^n telle que

$$L(x_1, \dots, x_n, \theta) = h(t, \theta)k(x_1, \dots, x_n).$$

Remarque. k détermine (c'est la densité ou la probabilité conditionnelle) la loi conditionnelle de (X_1, \dots, X_n) sachant T . La définition d'exhaustivité exprime que la loi conditionnelle est indépendante de θ .

Exemple. On considère un modèle d'échantillonnage avec $n > 1$ et pour lequel P_{θ} est une loi de Poisson de paramètre θ . On a donc les X_i i.i.d. avec $\forall i = 1, \dots, n, X_i \sim \text{Poisson}(\theta)$. On pose

$$T_n = \sum_{i=1}^n X_i \quad \text{et} \quad T'_n = X_1.$$

Montrer que la statistique T_n est exhaustive pour le paramètre λ , alors que la statistique T'_n ne l'est pas.

Théorème 4.2 (de factorisation (Admis)) Pour qu'une statistique T soit exhaustive, il suffit que la vraisemblance s'écrive :

$$L(x_1, \dots, x_n, \theta) = \phi(t, \theta)\psi(x_1, \dots, x_n).$$

V Exhaustivité et estimateurs efficaces ; la famille exponentielle

1 Le modèle exponentiel

Définition 4.5 On appelle famille exponentielle à paramètre unidimensionnel θ toute loi de probabilité (discrète ou continue) dont la vraisemblance peut se mettre sous la forme :

$$f(x, \theta) = \begin{cases} \exp[\alpha(\theta)\beta(x) + \gamma(\theta) + \delta(x)] & \text{si } x \in \mathcal{X} \\ 0 & \text{si } x \notin \mathcal{X} \end{cases},$$

avec, α et γ des fonctions deux fois différentiables.

Les lois Binômiales, de Poisson, normales font partie de la famille exponentielle.

Remarques.

— Dans la définition de la famille exponentielle, la définition des fonctions $\alpha(\cdot)$, $\beta(\cdot)$, $\gamma(\cdot)$ et $\delta(\cdot)$ n'est pas unique. En particulier, on peut poser $\alpha'(\cdot) = \frac{1}{a}\alpha(\cdot)$, $\beta'(\cdot) = a\beta(\cdot)$, $\gamma'(\cdot) = \gamma(\cdot) - b$ et $\delta'(\cdot) = \delta(\cdot) + b$ pour $a \in \mathbb{R}^*$ et $b \in \mathbb{R}$, et on obtiendra la même densité.

Proposition 4.2 Dans la famille exponentielle, toute statistique de la forme $T_n = k \sum_{i=1}^n \beta(X_i)$ est exhaustive pour θ .

Cette proposition résulte du théorème de factorisation.

Théorème 4.3 (Théorème de Darmois) Lorsque \mathcal{X} est indépendant de θ , le modèle admet une statistique exhaustive ssi le modèle est exponentiel.

Idée de la preuve:

Pour un modèle exponentiel, la statistique $T(X_1, \dots, X_n) = \sum_{i=1}^n \beta(X_i)$ est exhaustive pour θ .

Réciproquement, supposons que $T = \phi(X_1, \dots, X_n)$ soit une statistique exhaustive. On a :

$$L(x, \theta) = h(t, \theta)h(x) = \prod_{i=1}^n f(x_i, \theta),$$

on a

$$\frac{\partial^2 \ln f(x_i, \theta)}{\partial x_i \partial \theta} = \frac{\partial^2 \ln h(t, \theta)}{\partial x_i \partial \theta} = \frac{\partial^2 \ln h(t, \theta)}{\partial t \partial \theta} \frac{\partial \phi}{\partial x_i}. \quad (4.1)$$

Soit $k(\xi, \theta) = \frac{\partial \ln f(\xi, \theta)}{\partial \theta}$, soit $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, (4.1) donne

$$\frac{\partial_1 k(x_i, \theta)}{\partial_1 k(x_j, \theta)} = \frac{\partial_i \phi(x)}{\partial_j \phi(x)},$$

∂_j désigne la dérivée partielle par rapport à la j ème coordonnée.

Le second membre est indépendant de θ , on a donc pour $\xi \in \mathbb{R}$

$$\partial_1 k(\xi, \theta) = u(\xi)v(\theta),$$

en intégrant par rapport à ξ puis à θ , on obtient que $f(\xi, \theta)$ est de la forme :

$$\ln f(\xi, \theta) = \beta(\xi)\alpha(\theta) + \gamma(\theta) + \delta(\xi).$$

□

Remarque.

Lorsque \mathcal{X} dépend de θ , le théorème de Darmois ne s'applique pas. Il peut néanmoins exister des statistiques exhaustives. Par exemple, montrer que $T = \max X_i$ est une statistique exhaustive pour θ dans le modèle $\mathcal{U}([0, \theta])$.

2 Théorème sur l'efficacité

Dans le cas où \mathcal{X} ne dépend pas de θ , les seuls modèles qui admettent des estimateurs efficaces sont les modèles exponentiels. C'est l'objet du théorème suivant.

Théorème 4.4 *Si \mathcal{X} ne dépend pas de θ et que le modèle admette un estimateur efficace alors le modèle est exponentiel car l'estimateur est nécessairement exhaustif.*

Dans un modèle exponentiel, alors à une transformation linéaire près, il existe une unique estimateur efficace qui vérifie :

$$T = \frac{1}{n} \sum_{i=1}^n \beta(X_i),$$

$$g(\theta) = -\frac{\gamma'(\theta)}{\alpha'(\theta)},$$

$$g(\theta) = \mathbb{E}_\theta(T), \text{ et } \mathbb{V}_\theta(T) = \frac{g'(\theta)}{n\alpha'(\theta)}.$$

Preuve:

Montrons qu'un estimateur efficace est nécessairement exhaustif. On a :

$$V(T) \geq \frac{h'(\theta)^2}{I_T(\theta)} \text{ et } V(T) = \frac{h'(\theta)^2}{I_n(\theta)},$$

on en déduit que $I_T(\theta) = I_n(\theta)$ est donc que T est exhaustive.

Déterminons maintenant quelles sont les expressions possibles pour T . L'inégalité de Cramer-Rao découle d'une inégalité de Cauchy-Schwartz sur la covariance de T et $\frac{\partial}{\partial \theta} \ln L$, il y a donc égalité ssi les deux v.a. sont liées par une relation affine du type :

$$\frac{\partial}{\partial \theta} \ln L = \lambda(\theta)(T - h(\theta)).$$

Dans le cas d'un modèle exponentiel, on a :

$$\frac{\partial}{\partial \theta} \ln L = \alpha'(\theta) \left[\sum_{i=1}^n \beta(x_i) + \frac{n\gamma'(\theta)}{\alpha'(\theta)} \right]$$

ce qui conduit au résultat (après calculs).

□

Exemple.

Soit le modèle paramétrique réel d'échantillonnage suivant

$$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (\mathcal{N}(\mu, \sigma^2), \mu \in \Theta = \mathbb{R}))^n.$$

Soit l'estimateur $T_n = \bar{X}_n$.

- On a vu que, dans ce cadre-là, $I_n(\mu) = \frac{n}{\sigma^2}$.
- La statistique \bar{X}_n a pour distribution $\mathcal{N}(\mu, \sigma^2/n)$.

Le modèle associé à T est :

$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (\mathcal{N}(\mu, \sigma^2/n), \mu \in \mathbb{R}))$, l'information est $I_{\bar{X}_n}(\mu) = \frac{n}{\sigma^2}$ (information de dimension 1 d'un modèle normal).

On a $I_n(\mu) = I_{\bar{X}_n}(\mu) = \frac{n}{\sigma^2}$, donc la statistique \bar{X}_n est exhaustive dans le cadre du modèle considéré.

Exercice. Soit le modèle paramétrique réel d'échantillonnage suivant

$$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (\mathcal{N}(\mu, \sigma^2), \sigma^2 \in \Theta = \mathbb{R}_+^*))^n.$$

Soit l'estimateur $T_n = S_n^2$. Calculer $I_{S_n^2}(\sigma^2)$ et en déduire que S_n^2 n'est pas une statistique exhaustive pour σ^2 .

VI Quelques méthodes usuelles d'estimation

On va présenter brièvement ici des méthodes permettant de calculer de manière systématique un estimateur pour le paramètre θ d'un modèle statistique.

1 Méthode empirique

Si le paramètre θ considéré représente une quantité particulière pour le modèle (par exemple, l'espérance ou la variance), on peut naturellement choisir comme estimateur la quantité empirique correspondante pour l'échantillon X_1, \dots, X_n .

Exemple. Lorsque $\theta = \mathbb{E}[X]$, on peut choisir \bar{X}_n comme estimateur de θ .

Lorsque $\theta = \mathbb{V}(X)$, on peut choisir V_n^2 comme estimateur de θ ; on peut ensuite modifier l'estimateur pour améliorer ses qualités (V_n^2 asymptotiquement sans biais $\rightarrow S_n^2$ sans biais).

Remarque. Cette méthode très naturelle est relativement limitée.

2 Méthode des moindres carrés

Elle est utilisable lorsque l'espérance de la loi est une fonction inversible du paramètre θ , c'est à dire $\mathbb{E}[X] = h(\theta)$ où $h(\cdot)$ est bijective.

Définition 4.6 On appelle estimateur des moindres carrés de θ la statistique

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} \sum_{i=1}^n (X_i - h(\theta))^2.$$

Proposition 4.3 $\hat{\theta}_n = h^{-1}(\bar{X}_n)$.

Preuve:

En effet, $\sum_{i=1}^n (X_i - h(\theta))^2$ est un polynôme de degré 2 en $h(\theta)$ qui admet un unique minimum pour $h(\theta) = \bar{X}_n$. □

Exemple. Lorsque $\theta = \mathbb{E}[X]$, on a alors $h = Id = h^{-1}$. On en déduit que l'estimateur des moindres carrés $\hat{\theta}_n$ est : $\hat{\theta}_n = \bar{X}_n$.

Remarques.

- Cette méthode s'adapte au cas d'un paramètre θ multidimensionnel.
- Cette méthode est essentiellement utilisée dans le cadre du modèle linéaire.

3 Méthode des moments

Rappels.

- Soit X une variable aléatoire réelle.

On appelle *moment (théorique) d'ordre r* : $M_r = \mathbb{E}[X^r]$.

On appelle *moment (théorique) centré d'ordre r* : $\bar{M}_r = \mathbb{E}[(X - \mathbb{E}[X])^r]$.

- Soit (x_1, \dots, x_n) les valeurs observées d'un échantillon de taille n .

On appelle *moment empirique d'ordre r* : $m_r = \frac{1}{n} \sum_{i=1}^n (x_i)^r$.

On appelle *moment empirique centré d'ordre r* : $\bar{m}_r = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^r$.

Principe. Supposons le paramètre θ de dimension p . La méthode consiste à poser un système d'équations en égalant moments théoriques (centrés ou non) et moments empiriques :

$$\begin{cases} M_1(\theta) = m_1 \\ \vdots \\ M_p(\theta) = m_p \end{cases}$$

Chaque $M_j(\theta)$ dépend de θ , alors que les m_j n'en dépendent pas. On a donc un système à p équations et p inconnues pour obtenir les estimateurs des p composantes de θ .

Plus généralement, cette méthode s'applique si l'on souhaite estimer un paramètre θ qui s'exprime en fonction des moments. Exemple : paramètre de la loi Gamma, Béta (voir TD).

Exemple.

Soit une variable aléatoire réelle admettant une espérance μ et une variance σ^2 . Posons $\theta = (\mu, \sigma^2)$. Par les méthodes des moments, on obtient le système :

$$\begin{cases} M_1(\theta) = m_1 \\ \overline{M}_2(\theta) = \overline{m}_2 \end{cases} \quad \text{soit} \quad \begin{cases} \mu = \bar{x}_n \\ \sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 \end{cases}$$

d'où $\hat{\mu}_n = \bar{X}_n$ et $\hat{\sigma}_n^2 = V_n^2$.

4 Méthode du maximum de vraisemblance : principe

Définition 4.7 On appelle estimateur du maximum de vraisemblance (EMV) du paramètre θ la statistique $\hat{\theta}_n$ rendant maximale, selon θ , la fonction de vraisemblance du modèle $L(X_1, \dots, X_n, \theta)$, soit :

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} L(X, \theta).$$

Commentaires.

- L'idée de chercher la valeur de θ qui rend maximale la vraisemblance est naturelle : en effet, cette valeur particulière de θ permet de maximiser la probabilité d'obtenir les observations réalisées.
- La technique du maximum de vraisemblance est utilisable quel que soit le modèle utilisé. Notons que si le modèle a de bonnes propriétés de régularité, la détermination de l'EMV $\hat{\theta}_n$ sera simplifiée.

Propriété.

Si le modèle vérifie les propriétés **(H1)**, **(H2)** et **(H3)**, alors pour que $\hat{\theta}_n$ soit un EMV de θ il est nécessaire que

- $\left. \frac{\partial \ln L(X, \theta)}{\partial \theta} \right|_{\theta=\hat{\theta}_n} = 0$ soit $S_n(X, \hat{\theta}_n) = 0$ (équation de vraisemblance),
- $\left. \frac{\partial^2 \ln L(X, \theta)}{\partial \theta^2} \right|_{\theta=\hat{\theta}_n} < 0$.

Exemple. Soit le vecteur (X_1, \dots, X_n) dont les composantes X_i sont i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$. Posons $\theta = \mu$ et cherchons l'EMV de θ . On a :

$$\bullet \frac{\partial \ln L}{\partial \theta} \Big|_{\theta=\hat{\theta}_n} = \frac{\partial \ln L}{\partial \mu} \Big|_{\mu=\hat{\mu}_n} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \hat{\mu}_n) = 0 \text{ (équation de vraisem-}$$

blance)

d'où $\hat{\mu}_n = \bar{X}_n$.

$$\bullet \frac{\partial^2 \ln L}{\partial \theta^2} \Big|_{\theta=\hat{\theta}_n} = \frac{\partial^2 \ln L}{\partial \mu^2} \Big|_{\mu=\hat{\mu}_n} = -n/\sigma^2 < 0.$$

Donc l'EMV de μ est ici $\hat{\mu}_n$.

Remarque.

L'EMV ne peut s'obtenir que si la vraisemblance L est explicite, donc si le modèle est bien explicité (loi normale, loi de Poisson, ...). La méthode des moindres carrés et celle des moments ne nécessitent pas cette spécification.

Propriété (lien avec l'exhaustivité).

S'il existe une statistique exhaustive T_n pour θ , alors l'EMV de θ ne dépend que de T_n .

VII Exercices

Exercice 1. On considère la loi normale $\mathcal{N}(\mu, \sigma^2)$.

a) On suppose σ^2 connue et l'on considère le modèle paramétrique réel d'échantillonnage suivant

$$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (\mathcal{N}(\mu, \sigma^2), \mu \in \Theta = \mathbb{R}))^n.$$

L'estimateur \bar{X}_n est-il un estimateur efficace de μ ?

b) Sans supposer μ connue, on pose $\theta = \sigma^2$ et l'on considère le modèle paramétrique réel d'échantillonnage suivant

$$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (\mathcal{N}(\mu, \sigma^2), \sigma^2 \in \Theta = \mathbb{R}_+^*))^n.$$

L'estimateur S_n^2 est-il un estimateur efficace de σ^2 ?

Exercice 2.

a) Montrer que la loi de Poisson appartient à la famille exponentielle.

b) Montrer que la loi de Cauchy n'appartient pas à la famille exponentielle.

Exercice 3. Déterminer l'EMV du paramètre λ d'une loi de Poisson. En étudier les propriétés (biais, convergence, efficacité, exhaustivité).

Exercice 4. Ecrire la vraisemblance d'une loi de Bernoulli de paramètre $p \in]0, 1[$.

Déterminer l'EMV de p . Etudier ses propriétés (biais, convergence, efficacité, exhaustivité).

Exercice 5.

a) Déterminer l'EMV $\hat{\theta}_n$ du paramètre θ de la loi uniforme sur $[0, \theta]$ avec $\theta \in \mathbb{R}_+^*$.

b) Déterminer la densité de probabilité de $\hat{\theta}_n$.

c) Calculer $\mathbb{E}[\hat{\theta}_n]$ et $\mathbb{V}(\hat{\theta}_n)$.

- d) Etudier les propriétés de $\hat{\theta}_n$ (biais, convergence, efficacité).
- e) Proposer un estimateur T_n de θ sans biais et convergent.
- f) Choisir entre $\hat{\theta}_n$ et T_n au moyen du risque quadratique.
- g) Montrer que l'estimateur de θ obtenu par la méthode des moindres carrés est identique à l'estimateur des moments. On notera U_n cet estimateur.
- h) Etudier les propriétés de U_n (biais et convergence) et le comparer à T_n .
- i) Commenter.

VIII Généralisation au cas d'un paramètre multidimensionnel

On considère dans ce chapitre un modèle paramétrique réel d'échantillonnage :

$$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (P_{\theta}, \theta \in \Theta \subset \mathbb{R}^p))^n, \quad \text{avec } p \geq 2.$$

1 Généralisation des définitions sur les estimateurs

a. Estimateurs

Un estimateur est une statistique T_n définie sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (P_{\theta}, \theta \in \Theta \subset \mathbb{R}^p))^n$ à valeurs dans $(\Theta, \mathcal{B}_{\Theta})$. C'est donc un vecteur aléatoire de dimension p : $T_n = (T_{n,1}, \dots, T_{n,p})$.

Estimateur sans biais.

- **Biais de T_n** : $B(T_n) = \mathbb{E}(T_n) - \theta \in \mathbb{R}^p$.
- On dit que T_n est *sans biais pour θ* si $B(T_n) = 0_p$,

sinon T_n est biaisé.

- On dit que T_n est *asymptotiquement sans biais pour θ* si $B(T_n) \rightarrow 0_p$ pour $n \rightarrow +\infty$,
autrement dit si $\forall j = 1, \dots, p, \mathbb{E}[T_{n,j}] \rightarrow \theta_j$ pour $n \rightarrow +\infty$.

b. Estimateur convergent.

- T_n est *convergent pour θ* si et seulement si $T_n \xrightarrow{\text{proba}} \theta$ pour $n \rightarrow +\infty$,

c'est à dire : $\forall j = 1, \dots, p, T_{n,j} \xrightarrow{\text{proba}} \theta_j$.

- **Conditions nécessaires et suffisantes de convergence :**

$$\|T_n - \theta\| \xrightarrow{\text{proba}} 0 \text{ pour } n \rightarrow +\infty \iff T_n \text{ est convergent pour } \theta$$

où $\|\cdot\|$ désigne toute norme de \mathbb{R}^p .

c. Risque quadratique.

- Il est défini par : $R(T_n, \theta) = \mathbb{E}[(T_n - \theta)'(T_n - \theta)] = \sum_{j=1}^p \mathbb{E}[(T_{n,j} - \theta_j)^2]$.

— **Propriété :**

$$R(T_n, \theta) = B(T_n)'B(T_n) + \sum_{j=1}^p \mathbb{V}(T_{n,j}).$$

On peut réécrire ceci sous la forme :

$$R(T_n, \theta) = \sum_{j=1}^p (\mathbb{E}[T_{n,j}] - \theta_j)^2 + \sum_{j=1}^p \mathbb{V}(T_{n,j}).$$

2 Généralisation de l'inégalité de Cramer-Rao

On continue à noter

- $f(x, \theta)$ la vraisemblance du modèle de dimension 1, ici $x \in \mathbb{R}$ et $\theta \in \mathbb{R}^p$;
- $L(x, \theta)$ la vraisemblance du modèle de dimension n , ici $x \in \mathbb{R}^n$ et $\theta \in \mathbb{R}^p$;
- \mathcal{X} le support de f et \mathcal{X}^n celui de L .

a. Généralisation des hypothèses de régularité

Les hypothèses **(H1)** et **(H2)** ne sont pas modifiées mais seront ici notées **(H1')** et **(H2')**.

(H1') \mathcal{X} est indépendant de θ .

(H2') Θ est un ouvert.

On a $f(x, \theta) > 0, \forall (x, \theta) \in \mathcal{X} \times \Theta$.

(H3') $\forall j = 1, \dots, p, \frac{\partial}{\partial \theta_j} f(x, \theta)$ est définie $\forall (x, \theta) \in \mathcal{X} \times \Theta$.

$\forall (j, k) \in \{1, \dots, p\}, \frac{\partial^2}{\partial \theta_j \partial \theta_k} f(x, \theta)$ est définie $\forall (x, \theta) \in \mathcal{X} \times \Theta$.

(H4') $\forall (j, k) \in \{1, \dots, p\}$

$$\frac{\partial}{\partial \theta_j} f(x, \theta) \text{ et } \frac{\partial^2}{\partial \theta_j \partial \theta_k} f(x, \theta)$$

vérifient la propriété de domination sur tout compact de Θ (par des fonctions de x μ -intégrables).

b. Fonction de score (ou score)

On suppose les hypothèses **(H1')**, **(H2')** et **(H3')** vérifiées.

Définition. La fonction *score* est définie par :

$$S : \mathcal{X} \times \Theta \longrightarrow \mathbb{R}^p$$

$$(x, \theta) \longmapsto S(x, \theta) = \text{grad}_\theta \ln(f(x, \theta)) = \begin{pmatrix} \frac{\partial}{\partial \theta_1} \ln(f(x, \theta)) \\ \vdots \\ \frac{\partial}{\partial \theta_p} \ln(f(x, \theta)) \end{pmatrix}$$

Remarques et propriétés.

— On peut aussi définir le score du modèle de dimension n : $S_n(x, \theta) = \text{grad}_\theta \ln(L(x, \theta))$.

— Dans un modèle d'échantillonnage, on a :

$$S_n(x, \theta) = \sum_{i=1}^n S(x_i, \theta).$$

— Sous **(H4')**, on peut montrer que : $\mathbb{E}[S(X, \theta)] = 0_p = \mathbb{E}[S_n(x, \theta)]$.

c. Matrice d'information de Fisher

Définition. La *matrice d'information de Fisher* est une matrice carrée $p \times p$ définie par :

$$I(\theta) = \mathbb{E}[S(X, \theta)(S(X, \theta))'],$$

l'élément (j, k) de la matrice $I(\theta)$ est donnée par

$$\mathbb{E} \left[\frac{\partial}{\partial \theta_j} \ln(f(x, \theta)) \frac{\partial}{\partial \theta_k} \ln(f(x, \theta)) \right].$$

Remarques et propriétés.

— On peut aussi définir la matrice d'information de Fisher par rapport du modèle de dimension n :

$$I_n(x, \theta) = \mathbb{E}[S_n(X, \theta)(S_n(X, \theta))'].$$

— Sous **(H4')**, la matrice d'information de Fisher $I(\theta)$ est la matrice de covariance de $S(X, \theta)$. Ainsi, la matrice $I(\theta)$ est symétrique et semi-définie positive. Sous **(H4')**, l'élément (j, k) de la matrice $I(\theta)$ est donnée par $-\mathbb{E} \left[\frac{\partial^2}{\partial \theta_j \partial \theta_k} \ln(f(x, \theta)) \right]$.

— Dans un modèle d'échantillonnage, on a :

$$I_n(x, \theta) = nI(\theta).$$

d. Généralisation de l'inégalité de Cramer-Rao

— Soit T_n un estimateur de θ . On pose $\mathbb{E}[T_n] = g(\theta)$.

La fonction g définie sur Θ est à valeurs dans \mathbb{R}^p , sa j ème coordonnée est $g_j(\theta) = \mathbb{E}[T_{n,j}]$.

— Soit $D_g(\theta)$ la matrice jacobienne de g .

Cette matrice $D_g(\theta)$ est carrée d'ordre p , son terme général (j, k) est $\frac{\partial}{\partial \theta_k} g_j(\theta)$.

— Notons $V_{T_n}(\theta)$ la matrice de variances-covariances de T_n .

Cette matrice $V_{T_n}(\theta)$ est carrée d'ordre p et de terme général $\text{cov}(T_{n,j}, T_{n,k})$.

Considérons l'hypothèse supplémentaire suivante :

(H5') $I_n(\theta)$ est une matrice définie positive.

Inégalité de Cramer-Rao. Sous les hypothèses **(H1')** à **(H5')**, la matrice

$$V_{T_n}(\theta) - D_g(\theta) [I_n(\theta)]^{-1} (D_g(\theta))'$$

est semi-définie positive.

Définitions.

— La matrice $D_g(\theta) [I_n(\theta)]^{-1} (D_g(\theta))'$ s'appelle la *borne inférieure de l'inégalité de Cramer-Rao*.

— On dit que T_n est *efficace pour θ* s'il vérifie $V_{T_n}(\theta) = D_g(\theta) [I_n(\theta)]^{-1} (D_g(\theta))'$.

— La notion d'efficacité asymptotique sera définie ultérieurement.

- Dans le cas d'un estimateur sans biais, on a $D_g(\theta) = I_p$. L'inégalité de Cramer-Rao est alors : la matrice $V_{T_n}(\theta) - [I_n(\theta)]^{-1}$ est semi-définie positive.

e. Forme générale de la famille exponentielle

Définition. On dit qu'une loi de probabilité appartient à la famille exponentielle (à paramètre multidimensionnel) si sa vraisemblance peut s'écrire sous la forme :

$$f(x, \theta) = \begin{cases} \exp \left[\sum_{j=1}^p \alpha_j(\theta) \beta_j(x) + \gamma(\theta) + \delta(x) \right] & \text{si } x \in \mathcal{X}, \\ 0 & \text{sinon,} \end{cases}$$

avec \mathcal{X} indépendant de θ . Les applications α_j et γ vont de \mathbb{R}^p dans \mathbb{R} . Les applications β_j et δ vont de \mathbb{R} dans \mathbb{R} .

Vocabulaire.

- Si $\alpha_j(\theta) = \theta_j$, $\forall j = 1, \dots, p$, alors on dit que l'on a la *forme naturelle* de la famille exponentielle.
 — Si $\beta_j(x) = x$, $\forall j = 1, \dots, p$, alors on dit que l'on a la *forme canonique* de la famille exponentielle.

Exercice. On considère la loi normale $\mathcal{N}(\mu, \sigma^2)$ et on prend comme paramètre le vecteur $\theta = (\mu, \sigma^2)$. Montrer que la loi normale appartient bien à la famille exponentielle (préciser les diverses fonctions α_j , β_j , γ et δ).

Catalogue (non exhaustif).

- Les lois suivantes appartiennent à la famille exponentielle :
- lois discrètes : loi binomiale (donc loi de Bernoulli), binomiale négative, loi de Poisson, loi multinomiale.
 - lois continues : loi normale, loi log-normale, loi Gamma (donc loi du χ^2 et loi exponentielle), loi Béta.
- Les lois suivantes n'appartiennent pas à la famille exponentielle : loi de Cauchy, loi uniforme sur $[0, \theta]$, loi uniforme sur $[\theta_1, \theta_2]$.

Hypothèses de régularité et famille exponentielle.

Les hypothèses de régularité posées jusqu'à présent **(H1)** à **(H5)** dans le cas d'un paramètre θ unidimensionnel et **(H1')** à **(H5')** pour un paramètre multidimensionnel sont en général vérifiées pour la famille exponentielle.

3 Généralisation de la méthode du maximum de vraisemblance

Définition. L'estimateur du maximum de vraisemblance (EMV) de θ est définie par :

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} L(X_1, \dots, X_n, \theta).$$

Caractérisation de l'EMV $\hat{\theta}_n$. Si les hypothèses **(H1')**, **(H2')** et **(H3')** sont vérifiées, alors pour déterminer $\hat{\theta}_n$,

- i*) on résoud $S_n(X, \hat{\theta}_n) = 0$ (équations de vraisemblance),
- ii*) on vérifie que la matrice hessienne de $\ln L$ (matrice carrée d'ordre p de terme général $\frac{\partial^2}{\partial \theta_j \partial \theta_k} \ln(L(x, \theta))$ calculée en $\hat{\theta}_n$ est définie négative,
- iii*) on vérifie que le maximum local est un maximum.

Chapitre 5

Comportement asymptotique des estimateurs

I Propriétés asymptotiques de l'EMV

1 En dimension 1

On donne ci-après, sous forme de théorèmes, deux propriétés asymptotiques de l'EMV $\hat{\theta}_n$. Deux hypothèses supplémentaires sont nécessaires :

(H6) $\theta \neq \theta' \implies P_\theta \neq P_{\theta'}$.

On dit que *le modèle est identifiable*.

(H7) $\frac{\partial^2}{\partial \theta^2} \ln f(x, \theta)$ est continue en θ , uniformément en x .

Théorème 5.1 *Si les hypothèses (H1), (H2), (H3), (H4) et (H6) sont vérifiées, alors il existe une suite $\hat{\theta}_n$ d'estimateurs du maximum de vraisemblance qui converge presque sûrement vers θ .*

Théorème 5.2 *Sous les hypothèses (H1) à (H7), on a :*

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\text{loi}} \mathcal{N}(0, I^{-1}(\theta)) \quad \text{quand } n \rightarrow +\infty.$$

Preuve:

On considère :

$$h_n(X_1, \dots, X_n, \theta) = \frac{1}{n} \sum_{j=1}^n \frac{\partial \ln f(X_j, \theta)}{\partial \theta}.$$

C'est une suite de v.a.i.d centrée et de variance $I(\theta)$, ainsi $\sqrt{n}h_n$ converge en loi vers $\mathcal{N}(0, I(\theta))$. Notons,

$$K_n(X_1, \dots, X_n, \theta) = \frac{1}{n} \sum_{j=1}^n \frac{\partial^2 \ln f(X_j, \theta)}{\partial \theta^2}.$$

$$h_n(X_1, \dots, X_n, t) = h_n(X_1, \dots, X_n, \theta) + (t - \theta)K_n(X_1, \dots, X_n, \theta^*),$$

avec $\theta^* \in]\min(t, \theta), \max(t, \theta)[$. Pour $t = \hat{\theta}_n$,

$$0 = h_n(X_1, \dots, X_n, \theta) + (\hat{\theta}_n - \theta)K_n(X_1, \dots, X_n, \theta^*),$$

avec $\theta^* \in]\min(\hat{\theta}_n, \theta), \max(\hat{\theta}_n, \theta)[$. On utilise que $\hat{\theta}_n$ converge p.s. vers θ et la continuité uniforme de $\frac{\partial^2 \ln f(x, \theta)}{\partial \theta^2}$ pour montrer que $K_n(X_1, \dots, X_n, \theta^*)$ converge p.s. vers $-I(\theta)$, ce qui permet de conclure. \square

Commentaires.

- La propriété asymptotique donnée au Théorème 5.1 explique que, dans la pratique, on cherche une valeur annulant l'équation de vraisemblance et on prend cette valeur comme estimation pour $\hat{\theta}_n$. En particulier, dans les cas compliqués (où il n'est pas facile d'obtenir une expression analytique de $\hat{\theta}_n$), les logiciels de statistique utilisent des algorithmes (de type Newton-Raphson) pour obtenir l'estimation du maximum de vraisemblance.
- On peut écrire aussi le résultat du Théorème 5.2 sous la forme :

$$\sqrt{I_n(\theta)} (\hat{\theta}_n - \theta) \xrightarrow{\text{loi}} \mathcal{N}(0, 1) \quad \text{quand } n \rightarrow +\infty.$$

2 En dimension supérieure

Les résultats de convergence pour l'EMV en dimension supérieure restent valables. En particulier, si le modèle est identifiable, on a l'existence d'une suite d'EMV qui converge presque sûrement et la convergence en loi :

Théorème 5.3 *Si les hypothèses (H1') à (H7') sont toutes vérifiées, alors on a :*

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\text{loi}} \mathcal{N}_p(0_n, I^{-1}(\theta)).$$

II Définitions / outils

1 Normalité et efficacité asymptotique

Soit T_n un estimateur de θ .

- Si $\sqrt{n}(T_n - \theta) \xrightarrow{\text{loi}} \mathcal{N}_p(0_p, \Sigma)$,
alors on dit que T_n est *asymptotiquement normal*.

La matrice Σ est appelée matrice de variances-covariances asymptotique de T_n .

(Cela n'implique pas que $nV(T_n) \rightarrow \Sigma$.)

- Si $\sqrt{n}(T_n - \theta) \xrightarrow{\text{loi}} \mathcal{N}_p(0_p, I^{-1}(\theta))$,
alors on dit que T_n est *asymptotiquement efficace*.

Remarque : cette définition d'efficacité asymptotique en dimension > 1 ne coïncide pas avec la définition donnée en dimension 1 (dans le cas sans biais : $nV(T_n) \rightarrow \frac{1}{I(\theta)}$).

— L'EMV est asymptotiquement normal et efficace.

2 Propriétés de convergence

Les résultats de convergence suivants sont souvent utiles en statistique. X_n et Y_n sont des vecteurs aléatoires de \mathbb{R}^k .

- Si $X_n - Y_n \xrightarrow{\mathbb{P}} 0$ et que $X_n \xrightarrow{\mathcal{L}} X$ alors $Y_n \xrightarrow{\mathcal{L}} X$.
- Si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathbb{P}} 0$ alors $X_n \cdot Y_n \xrightarrow{\mathbb{P}} 0$.
- Si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathbb{P}} b$ alors $X_n + Y_n \xrightarrow{\mathcal{L}} X + b$ et $X_n \cdot Y_n \xrightarrow{\mathcal{L}} bX$.
- Si $a_n(X_n - \mu) \xrightarrow{\mathcal{L}} X$ avec $a_n \rightarrow \infty$ alors $X_n \xrightarrow{\mathbb{P}} \mu$.

3 Méthode Delta

Cette méthode est utile lorsqu'on dispose d'un estimateur asymptotiquement normal d'un paramètre θ . Soit g une fonction C^1 . On suppose que T_n est un estimateur de θ tel que

$$a_n(T_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2(\theta))$$

avec $a_n \rightarrow \infty$. Alors, $g(T_n)$ converge en probabilité vers $g(\theta)$ et

$$a_n(g(T_n) - g(\theta)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, g'(\theta)^2 \sigma^2(\theta)).$$

En dimension supérieure, on a le résultat suivant. On considère T_n un vecteur aléatoire de \mathbb{R}^k , Σ une matrice de covariance. On suppose que

$$a_n(T_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma)$$

avec $a_n \rightarrow \infty$. Alors, pour toute fonction g de classe C^1 , $g(T_n)$ converge en probabilité vers $g(\theta)$ et

$$a_n(g(T_n) - g(\theta)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, D_g \Sigma D_g^t)$$

où D_g est la matrice Jacobienne de g calculée en θ . En particulier, si T_n est un estimateur asymptotiquement efficace de θ alors, $g(T_n)$ est un estimateur asymptotiquement efficace de $g(\theta)$.

III Exercices

Exercice 1. On considère le modèle d'échantillonnage normal avec $P_\theta = \mathcal{N}(\mu, \sigma^2)$.

- a) Déterminer l'EMV de $\theta = (\mu, \sigma^2)$.
- b) Etudier ses propriétés (biais, convergence, efficacité).
- c) Quelle fonction $h(\theta)$ peut-on estimer par un estimateur sans biais et efficace ?

Exercice 2. On considère le modèle d'échantillonnage multinomial à $k \geq 3$ catégories :

$$\left(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{B}(p_1, \dots, p_k), p_i \in [0, 1], \sum_{i=1}^k p_i = 1 \right)^n$$

avec $X \rightsquigarrow \mathcal{B}(p_1, \dots, p_k)$, $\mathbb{P}(X = a_i) = p_i$.

- a) Déterminer l'EMV de $\theta = (p_1, \dots, p_{k-1})$.
- b) Montrer qu'il est sans biais, convergent et efficace.

Chapitre 6

Estimation par intervalle de confiance

I Introduction

On va considérer dans ce chapitre un modèle statistique réel paramétrique (avec un paramètre unidimensionnel) :

$$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (P_{\theta}, \theta \in \Theta \subset \mathbb{R}))^n.$$

On va construire des intervalles de confiance du paramètre θ . Lorsque θ est multidimensionnel, on parle de régions de confiance.

Définition.

Soit $\alpha \in [0, 1]$. On appelle *intervalle de confiance du paramètre θ* de niveau (de confiance) $1 - \alpha$ la donnée de deux statistiques A_n et B_n vérifiant

$$P(A_n \leq \theta \leq B_n) = 1 - \alpha.$$

Commentaires.

1. θ représente la valeur (inconnue) du paramètre.
2. A_n et B_n sont deux statistiques réelles, plus précisément il s'agit généralement de deux estimateurs de θ , donc à valeurs dans $(\Theta, \mathcal{B}_{\Theta})$.
3. A_n et B_n sont supposées tels que $P(A_n \leq B_n) = 1$.
4. $\alpha \in [0, 1]$ est un risque, appelé aussi seuil.
La valeur de α est choisie a priori par le statisticien (très souvent, $\alpha = 1\%$, 5% ou 10%).
5. Le niveau de confiance $1 - \alpha$ est aussi parfois appelé *coefficient de sécurité* ou *coefficient de confiance*.
6. Soient x_1, \dots, x_n les valeurs observées des variables aléatoires de l'échantillon X_1, \dots, X_n .
Posons $a_n = A_n(x_1, \dots, x_n)$ et $b_n = B_n(x_1, \dots, x_n)$. L'intervalle $[a_n, b_n]$ est un intervalle réel inclu dans Θ .

7. Un tel intervalle est aussi parfois appelé *fourchette*, en particulier dans le cadre des sondages.

Remarques

1. Estimer le paramètre θ par intervalle de confiance est plus raisonnable que de l'estimer ponctuellement.
En plus de l'intervalle en lui-même, on a une probabilité associée, le niveau de confiance $1 - \alpha$.
2. La longueur de l'intervalle de confiance $b_n - a_n$ nous renseigne sur la précision de l'estimation.
3. Il n'existe pas de méthode systématique de construction d'intervalles de confiance.
4. La notion d'intervalle de confiance est liée à celle de test d'hypothèses, le coefficient α étant alors le risque de première espèce.

Liens entre les bornes A_n et B_n

Commençons par remarquer que :

$$P(A_n \leq \theta \leq B_n) = 1 - \alpha \iff \alpha = P(A_n > \theta \text{ ou } B_n < \theta) = P(A_n > \theta) + P(B_n < \theta)$$

On peut être amené à construire des intervalles de confiance de trois types différents.

Cas d'un intervalle du type $[a_n, +\infty[$.

Le statisticien cherche ici à assurer une valeur minimale au paramètre θ . C'est par exemple le cas lorsque l'on s'intéresse à la durée de vie minimum d'un composant électronique.

On concentre ici le risque α entièrement sur $P(A_n > \theta)$, soit $\alpha = P(A_n > \theta)$. On a en général dans ce cas une solution unique.

Cas d'un intervalle du type $] - \infty, b_n]$.

Le statisticien cherche ici à assurer une valeur maximale au paramètre θ . C'est par exemple le cas lorsque l'on désire avoir une concentration en sucre inférieure à un certain pourcentage fixé dans un aliment.

On concentre ici le risque α entièrement sur $P(B_n < \theta)$, soit $\alpha = P(B_n < \theta)$. On a en général dans ce cas une solution unique.

Cas d'un intervalle du type $[a_n, b_n]$.

Le statisticien cherche ici à encadrer la valeur du paramètre θ . C'est par exemple le cas lorsque l'on s'intéresse au poids d'un paquet de café sur la chaîne de production.

On répartit ici le risque α des deux cotés : $\alpha = P(A_n > \theta) + P(B_n < \theta)$. Généralement, on choisit A_n et B_n de manière à ce que $P(A_n > \theta) = P(B_n < \theta) = \frac{\alpha}{2}$.

Quelques rappels sur des lois de probabilité

$T(n)$ désigne une loi de Student à n degrés de liberté, $F(n_1, n_2)$ désigne une loi de Fisher-Snedecor à (n_1, n_2) degré de liberté.

- $X \sim \mathcal{N}(0, 1) \implies X^2 \sim \chi^2(1)$

- $X \sim \mathcal{N}(0, 1), Y \sim \chi^2(n), X \text{ et } Y \text{ indépendantes} \implies T = \frac{X}{\sqrt{Y/n}} \sim T(n)$

- $X_1 \sim \chi^2(n_1)$, $X_2 \sim \chi^2(n_2)$, X_1 et X_2 indépendantes $\implies F = \frac{X_1/n_1}{X_2/n_2} \sim F(n_1, n_2)$
- $T \sim T(n)$ $\implies F = T^2 \sim F(1, n)$
- *Combinaisons linéaires de variables aléatoires indépendantes :*

$$X_i \sim B(n_i, p), \quad X_i \text{ indépendantes} \implies \sum_i X_i \sim B\left(\sum_i n_i, p\right)$$

$$X_i \sim \text{Poisson}(\lambda_i), \quad X_i \text{ indépendantes} \implies \sum_i X_i \sim \text{Poisson}\left(\sum_i \lambda_i\right)$$

$$X_i \sim \mathcal{N}(\mu_i, \sigma_i^2), \quad X_i \text{ indépendantes} \implies \sum_i a_i X_i \sim \mathcal{N}\left(\sum_i a_i \mu_i, \sum_i a_i^2 \sigma_i^2\right)$$

$$X_i \sim \chi^2(n_i), \quad X_i \text{ indépendantes} \implies \sum_i X_i \sim \chi^2\left(\sum_i n_i\right)$$

II Intervalles de confiance pour les paramètres de la loi normale

On suppose ici que l'on dispose d'un échantillon (X_1, \dots, X_n) où les X_i sont indépendants et identiquement distribués selon la loi $\mathcal{N}(\mu, \sigma^2)$.

On s'intéressera à la moyenne μ et à la variance σ^2 successivement. On travaille avec un niveau de confiance $(1 - \alpha)$ fixé.

Intervalle de confiance pour μ lorsque σ^2 est connue

L'intervalle de confiance pour μ de niveau de confiance $1 - \alpha$ lorsque σ^2 est connue est :

$$\bar{X}_n - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

où $z_{1-\alpha/2}$ est le fractile d'ordre $1 - \alpha/2$ de la loi normale centrée réduite $\mathcal{N}(0, 1)$.

Intervalle de confiance pour μ lorsque σ^2 est inconnue

L'intervalle de confiance pour μ de niveau de confiance $1 - \alpha$ lorsque σ^2 est inconnue est :

$$\bar{X}_n - t_{1-\alpha/2} \frac{S_n}{\sqrt{n}} \leq \mu \leq \bar{X}_n + t_{1-\alpha/2} \frac{S_n}{\sqrt{n}}$$

où $t_{1-\alpha/2}$ est le fractile d'ordre $1 - \alpha/2$ de la loi de Student $T(n-1)$ et $S_n = \sqrt{S_n^2}$.

Intervalle de confiance pour σ^2 lorsque μ est connue

On se donne ici $\alpha_1 > 0$ et $\alpha_2 > 0$ vérifiant $\alpha_1 + \alpha_2 = \alpha$.

L'intervalle de confiance pour σ^2 de niveau de confiance $1 - \alpha$ lorsque μ est connue est :

$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\tilde{k}_2} \leq \sigma^2 \leq \frac{\sum_{i=1}^n (X_i - \mu)^2}{\tilde{k}_1}$$

où \tilde{k}_1 (resp. \tilde{k}_2) est le fractile d'ordre α_1 (resp. $1 - \alpha_2$) de la loi du chi-deux $\chi^2(n)$.

Intervalle de confiance pour σ^2 lorsque μ est inconnue

On se donne ici à nouveau $\alpha_1 > 0$ et $\alpha_2 > 0$ vérifiant $\alpha_1 + \alpha_2 = \alpha$.

L'intervalle de confiance pour σ^2 de niveau de confiance $1 - \alpha$ lorsque μ est inconnue est :

$$\frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{k_2} \leq \sigma^2 \leq \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{k_1}$$

où k_1 (resp. k_2) est le fractile d'ordre α_1 (resp. $1 - \alpha_2$) de la loi du chi-deux $\chi^2(n - 1)$.

Remarques

- En général, les cas où σ^2 (ou μ) est connue sont rares. Les deux autres cas sont les plus usuels en pratique.
- Les quantités (bornes) intervenant dans les intervalles de confiance de la variance σ^2 sont strictement positives, on peut donc en déduire un intervalle de confiance pour l'écart-type σ au niveau de confiance $1 - \alpha$: par exemple, lorsque l'on suppose μ connue, on a

$$\sqrt{\frac{\sum_{i=1}^n (X_i - \mu)^2}{\tilde{k}_2}} \leq \sigma \leq \sqrt{\frac{\sum_{i=1}^n (X_i - \mu)^2}{\tilde{k}_1}}.$$

III Intervalles de confiance pour une proportion (paramètre de la loi binomiale)

On considère n épreuves indépendantes au cours desquelles on s'intéresse à un événement E de probabilité p (avec $0 < p < 1$). A la i ème épreuve, la variable aléatoire X_i va prendre la valeur 0 si E n'est pas réalisé (avec la probabilité $1 - p$) et 1 si E est réalisé (avec la probabilité p).

Posons $R_n = \sum_{i=1}^n X_i$. La variable aléatoire R_n suit la loi $B(n, p)$.

Le modèle est ici $(\Omega = \{0, 1\}, \mathbb{P}(\Omega), P_\theta)^n$ où $P_\theta = \text{Bernoulli}(p)$, $\theta = p$ et $\Theta =]0, 1[$.

On rappelle que $\bar{X}_n = \frac{R_n}{n} =: \hat{p}_n$ est l'estimateur du maximum de vraisemblance (EMV) de p .

L'inégalité de Bienaimé-Tchebitchev permet d'obtenir un intervalle de confiance de risque $\leq \alpha$. En effet, on a pour tout $\varepsilon > 0$,

$$\mathbb{P}(|\bar{X}_n - p| > \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2},$$

soit :

$$\mathbb{P}(|\bar{X}_n - p| > \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}.$$

Fixons un risque $\alpha \in]0, 1[$, alors pour $\varepsilon = \sqrt{\frac{1}{\alpha 4n}}$, l'intervalle

$$\left[\bar{X}_n - \sqrt{\frac{1}{\alpha 4n}}, \bar{X}_n + \sqrt{\frac{1}{\alpha 4n}} \right]$$

est de niveau de confiance supérieur ou égal à $1 - \alpha$.

Exemple voir exercice 2, comparer les résultats fournis par l'inégalité de Bienaimé Tchebitchev et les intervalles de confiance asymptotiques fournis par le TLC.

Le théorème de la limite centrale permet d'obtenir des intervalles de confiance asymptotiques dans le cas des grands échantillons.

IV Construction d'intervalles de confiance asymptotiques

On va utiliser ici des théorèmes de convergence pour construire des intervalles de confiance dont le niveau de confiance est asymptotiquement égal à $1 - \alpha$.

Définition 6.1 *Un intervalle de confiance $[A_n, B_n]$ est de niveau asymptotique $1 - \alpha$ si*

$$\mathbb{P}(A_n \leq \theta \leq B_n) \xrightarrow{n \rightarrow \infty} 1 - \alpha.$$

1 Utilisation du théorème central limite

Rappels du théorème central limite. Soient n variables aléatoires réelles X_1, \dots, X_n indépendantes et identiquement distribuées avec $\mathbb{E}[X_i] = \mu$ et $\mathbb{V}(X_i) = \sigma^2$, alors

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{\text{Loi}} \mathcal{N}(0, 1) \quad \text{pour } n \rightarrow +\infty.$$

On en déduit que

$$P \left(-z_{1-\alpha/2} \leq \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq z_{1-\alpha/2} \right) \xrightarrow{n \rightarrow \infty} 1 - \alpha.$$

Ce résultat peut servir à construire des intervalles de confiance asymptotique de niveau $1 - \alpha$ dans les cas suivants :

- pour μ lorsque σ est connue,
- pour σ lorsque μ est connue,
- pour un paramètre θ dépendant à la fois de μ et de σ (loi de Poisson ou loi binomiale par exemple).

2 Application à la loi binomiale.

Utiliser le résultat précédent pour déterminer l'intervalle de confiance asymptotique de niveau $(1 - \alpha)$ du paramètre $\theta = p$.

3 Utilisation de la convergence de l'EMV

Rappels (voir Chapitre 2). Supposons vérifiées les hypothèses de régularité **(H1)** à **(H7)**.

Soit $I_n(\theta)$ l'information de Fisher du modèle considéré. Soit $\hat{\theta}_n$ l'EMV de θ . On a vu que :

$$\sqrt{I_n(\theta)} (\hat{\theta}_n - \theta) \xrightarrow{Loi} \mathcal{N}(0, 1) \quad \text{pour } n \rightarrow +\infty.$$

On en déduit que

$$P \left(-z_{1-\alpha/2} \leq \sqrt{I_n(\theta)} (\hat{\theta}_n - \theta) \leq z_{1-\alpha/2} \right) \simeq 1 - \alpha.$$

Ce résultat peut servir à construire des intervalles de confiance asymptotique de niveau $1 - \alpha$ pour le paramètre θ (à condition de connaître $I_n(\theta)$).

La méthode Δ (stabilisation de la variance) permet aussi d'obtenir des intervalles de confiance (retour sur l'exemple d'une loi de Poisson de paramètre λ).

4 Remarque sur l'intervalle de confiance pour une variance hors du cadre normal

Dans le cadre normal, on a l'intervalle de confiance pour σ^2 :

$$\left[\frac{(n-1)S_n^2}{k_1}, \frac{(n-1)S_n^2}{k_2} \right],$$

où k_1 et k_2 sont respectivement le $(1 - \frac{\alpha}{2})$ et le $\frac{\alpha}{2}$ quantile d'une loi $\chi^2(n-1)$.

On se demande si cet intervalle de confiance est de niveau asymptotique $1 - \alpha$ lorsque l'on n'est plus dans un cadre normal. Considérons l'intervalle de confiance unilatéral :

$$\left[\frac{(n-1)S_n^2}{\tilde{k}_1}, \infty \right[$$

avec \tilde{k}_1 le $1 - \alpha$ quantile d'une loi $\chi^2(n-1)$, est-il de niveau asymptotique $1 - \alpha$? Nous allons voir que la réponse est **NON** en général.

Rappelons que pour une variable aléatoire admettant un moment d'ordre 4, on appelle *kurtosis* ou *coefficient d'applatissement* la quantité :

$$\kappa = \frac{mc_4}{\sigma^4},$$

où mc_4 est le moment centré d'ordre 4. On notera $\gamma = \kappa - 3$ l'*excès d'applatissement*. Remarquons que pour une loi normale, $\kappa = 3$.

Soit $V_n = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$. Le TLC implique que

$$\sqrt{n}(V_n - \sigma^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, mc_4 - \sigma^4),$$

autrement dit :

$$\sqrt{n} \left(\frac{V_n}{\sigma^2} - 1 \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \gamma + 2).$$

Le théorème de la limite centrale implique aussi que si Z_n suit une loi $\chi^2(n-1)$ alors

$$\frac{Z_n - (n-1)}{\sqrt{2(n-1)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Notons k_n le $(1-\alpha)$ quantile d'une loi $\chi^2(n-1)$. On cherche la limite de

$$\mathbb{P} \left(\frac{nV_n}{\sigma^2} \leq k_n \right).$$

On a :

$$\mathbb{P} \left(\frac{nV_n}{\sigma^2} \leq k_n \right) = \mathbb{P} \left(\sqrt{n} \left(\frac{V_n}{\sigma^2} - 1 \right) \leq \frac{k_n - n}{\sqrt{n}} \right)$$

k_n est le $(1-\alpha)$ quantile d'une loi $\chi^2(n-1)$, comme

$$\frac{Z_n - (n-1)}{\sqrt{2(n-1)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1),$$

avec des arguments d'uniformité dans le TCL, on peut montrer que :

$$\frac{k_n - (n-1)}{\sqrt{2(n-1)}} \longrightarrow z_{1-\alpha},$$

où $z_{1-\alpha}$ est le $1-\alpha$ quantile d'une loi $\mathcal{N}(0, 1)$. Finalement, notons U une variable aléatoire de loi $\mathcal{N}(0, 1)$, on a :

$$\mathbb{P} \left(\frac{nV_n}{\sigma^2} \leq k_n \right) \xrightarrow{n \rightarrow \infty} \mathbb{P} \left(U \leq z_{1-\alpha} \frac{\sqrt{2}}{\sqrt{\gamma + 2}} \right)$$

cette dernière quantité ne vaut $1-\alpha$ que si $\gamma = 0$ ou de manière équivalente si $\kappa = 3$.

Sous réserve de justifier précisément les convergences ci-dessus, on a montré :

Théorème 6.1 *L'intervalle de confiance unilatéral :*

$$\left[\frac{nV_n}{k_n}, \infty[$$

est de niveau asymptotique $1 - \alpha$ si et seulement si le coefficient d'aplatissement κ vaut 3.

V Recherche de régions de confiance

Montrons sur un exemple comment on peut construire des régions de confiance pour des paramètres multi dimensionnels.

On cherche une région de confiance de niveau $(1 - \alpha)$ pour le couple (μ, σ^2) dans le cadre du modèle paramétrique gaussien. On considère le vecteur aléatoire :

$$\begin{pmatrix} \sqrt{n} \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \\ \frac{(n-1)S_n^2}{\sigma^2} \end{pmatrix} = \begin{pmatrix} U \\ V \end{pmatrix}.$$

On sait que $U \rightsquigarrow \mathcal{N}(0, 1)$, $V \rightsquigarrow \chi^2(n-1)$ et U et V sont indépendantes. Soit α_1 et α_2 tels que $(1 - \alpha_1)(1 - \alpha_2) = (1 - \alpha)$, la région du plan (σ^2, μ) :

$$\left\{ \bar{X} - \frac{z_{1-\frac{\alpha_2}{2}}\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{z_{1-\frac{\alpha_2}{2}}\sigma}{\sqrt{n}}; \frac{(n-1)S_n^2}{k_1} \leq \sigma^2 \leq \frac{(n-1)S_n^2}{k_2} \right\}.$$

est une région de confiance de niveau $(1 - \alpha)$.

On pourrait aussi considérer

$$D^2 = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2$$

qui suit une loi du $\chi^2(n)$ et considérer la région de confiance :

$$\left\{ \sigma^2 k_2 \leq \sum_{i=1}^n (X_i - \mu)^2 \leq \sigma^2 k_1 \right\},$$

où k_1 et k_2 sont respectivement le $1 - \frac{\alpha}{2}$ et le $\frac{\alpha}{2}$ quantile d'une loi $\chi^2(n)$.

VI Exercices

Exercice 1. Soient X_1, \dots, X_{10} dix variables aléatoires i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$. On dispose des observations suivantes :

6 8 1 5 6 7 6 6 5 9

Calculer les intervalles de confiance de niveau 95% suivants :

- pour μ , sachant que $\sigma^2 = 4$;
- pour μ , ne connaissant pas σ^2 ;
- pour σ^2 , puis pour σ , ne connaissant pas μ .

Exercice 2.

Dans une fabrication en série, on cherche à estimer le taux de pièces défectueuses. Pour cela, on a réalisé, à quatre périodes différentes, quatre prélèvements. Les résultats sont les suivants :

- 6 pièces défectueuses sur 30,
- 10 pièces défectueuses sur 50,
- 20 pièces défectueuses sur 100,
- 40 pièces défectueuses sur 200.

Déterminer, dans chaque cas, l'intervalle de confiance de niveau 95% de ce taux.

Exercice 3.

Déterminer l'intervalle de confiance de niveau 95% de la proportion p d'un événement E , lorsque sur 80 expériences (indépendantes), l'événement s'est produit 45 fois.

Exercice 4.

En utilisant le théorème central limite, construire un intervalle de confiance de niveau asymptotiquement égal à $1 - \alpha$ pour le paramètre λ d'une loi de Poisson.

APPLICATION NUMÉRIQUE : On compte le nombre de parasites par fruit dans un lot de fruits parasités et on obtient :

x_i : nombre de parasites par fruit	0	1	2	3	4	5
n_i : nombre de fruits contenant x_i parasites	11	29	27	19	10	4

Si l'on suppose que le nombre de parasites suit une loi de Poisson de paramètre λ ,

donner l'intervalle de confiance de niveau asymptotiquement égal à 99% pour le paramètre λ .

Exercice 5.

On considère une variable aléatoire réelle continue de densité :

$$f(x) = \begin{cases} 0 & \text{si } x < 2, \\ \theta \exp(-\theta(x - 2)) & \text{si } x \geq 2, \end{cases}$$

avec $\theta > 0$.

1. Vérifier que cette loi appartient à la famille exponentielle.
2. En utilisant les propriétés de l'EMV de θ , construire un intervalle de confiance pour θ de niveau asymptotiquement égal à $1 - \alpha$.
3. APPLICATION NUMÉRIQUE : Calculer cet intervalle de confiance pour $n = 200$, $\bar{x}_n = 6,68$ et $\alpha = 5\%$.

Exercice 6.

On considère n_1 variables aléatoires réelles $X_{1,1}, \dots, X_{1,n_1}$ i.i.d. de loi $\mathcal{N}(\mu_1, \sigma^2)$ et

n_2 variables aléatoires réelles $X_{2,1}, \dots, X_{2,n_2}$ i.i.d. de loi $\mathcal{N}(\mu_2, \sigma^2)$. On suppose de plus les variables $X_{k,i}$ ($k = 1, 2$ et $i = 1, \dots, n_k$) mutuellement indépendantes.

1. Soient $\bar{X}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} X_{1,i}$ et $\bar{X}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} X_{2,i}$. Quelle est la loi de $\bar{X}_1 - \bar{X}_2$?
2. Soit $S^2 = \frac{1}{n_1 + n_2 - 2} \left[\sum_{i=1}^{n_1} (X_{1,i} - \bar{X}_1)^2 + \sum_{i=1}^{n_2} (X_{2,i} - \bar{X}_2)^2 \right]$.
Quelle est la loi de $(n_1 + n_2 - 2) \frac{S^2}{\sigma^2}$?
3. En déduire un intervalle de confiance de niveau $1 - \alpha$ pour le paramètre $\mu_1 - \mu_2$.
4. APPLICATION NUMÉRIQUE : On a observé $n_1 = 10$ et $n_2 = 8$ observations dans chacune des deux populations considérées. Les données obtenues sont les suivantes :
 $x_{1,i} : 1,36 \quad 2,66 \quad 2,05 \quad 1,85 \quad 2,28 \quad 1,71 \quad 0,75 \quad 1,97 \quad 1,70 \quad 1,68$
 $x_{2,i} : 1,91 \quad 2,03 \quad 1,31 \quad 1,33 \quad 2,68 \quad 2,04 \quad 0,40 \quad 3,31$
Calculer l'intervalle de confiance de niveau 95% pour le paramètre $\mu_1 - \mu_2$.

Chapitre 7

Généralités sur les tests

Le but d'un test statistique est d'aider à la décision. Par exemple : on veut s'assurer qu'un médicament n'influe pas sur le taux d'une certaine hormone, on observe sur un nombre n d'individus la différence du taux d'hormone avant et après la prise du médicament. Comment décider si la différence observée est significative ?

I Problèmes de test

Nous nous plaçons dans le cadre d'un modèle statistique d'échantillonnage paramétrique.

Le but d'un test statistique est de donner un critère permettant de retenir l'hypothèse $H_0 : \theta \in \Theta_0$ ou de retenir une hypothèse alternative $H_1 : \theta \in \Theta_1$, avec $\Theta_1 \subset \Theta_0^c$.

Dans l'exemple ci-dessus, le paramètre θ est la variation du taux d'hormone, le modèle statistique est $(\mathbb{R}^n, (\mathbb{P}_\theta^{\otimes n})_{\theta \in \mathbb{R}})$ où \mathbb{P}_θ est une loi symétrique autour de θ (par exemple une loi normale de moyenne θ). La question posée consiste à déterminer si, au vue des différences observées, on peut considérer que $\theta = 0$. L'hypothèse H_0 s'écrit $\theta = 0$, l'hypothèse alternative H_1 s'écrit $\theta \neq 0$.

La mise en œuvre du critère du test détermine une zone de rejet ou zone critique W , W^c est la zone d'acceptation ou zone de confiance. Le test est basé sur un modèle probabiliste et comporte un certain risque. On appelle *risque de première espèce*, notée α , la probabilité de rejeter l'hypothèse H_0 alors qu'elle est vraie, $\alpha = \mathbb{P}(W|H_0)$. La probabilité, notée $1 - \beta$, de retenir l'hypothèse H_0 alors qu'elle est fautive s'appelle *risque de deuxième espèce*, $1 - \beta = \mathbb{P}(W^c|H_1)$. La probabilité β s'appelle *puissance du test*. Pour déterminer une région critique convenable, il existe différentes méthodes : méthode de Neymann et Pearson et test les plus puissants, méthode du maximum de vraisemblance, ... La méthode de Neyman et Pearson permet de construire, pour α fixé, des tests maximisant la puissance β .

On considère un modèle d'échantillonnage paramétrique dominé :

$$(E, \mathcal{B}, \mathbb{P}_\theta \theta \in \Theta)^n,$$

(X_1, \dots, X_n) est un vecteur aléatoire issu de ce modèle.

Définition 7.1 *Un test est donné par une fonction $\Phi : E^n \rightarrow \{0, 1\}$, on retiendra H_0 si $\Phi(X_1, \dots, X_n) = 0$, on rejette H_0 si $\Phi(X_1, \dots, X_n) = 1$. On appelle zone de rejet l'ensemble $R = \{\Phi(X_1, \dots, X_n) = 1\}$. Évidemment, étant donnée une zone de rejet $R \subset E^n$, on définit un test en posant $\Phi = \mathbb{I}_R$.*

Lorsque $\Theta_0 = \{\theta_0\}$ et $\Theta_1 = \{\theta_1\}$, on parle d'hypothèses simples.

Définition 7.2 *Le niveau du test - ou sa sensibilité - est la probabilité de rejeter H_0 à tort :*

$$\alpha = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta((X_1, \dots, X_n) \in R).$$

La puissance du test est la fonction $\beta : \Theta_1 \rightarrow [0, 1]$ définie par $\beta(\theta) = \mathbb{P}_\theta((X_1, \dots, X_n) \in R)$. Le test est dit sans biais si $\beta(\theta) \geq \alpha \forall \theta \in \Theta_1$.

Dans la pratique, on va fixer un seuil α et on cherche un région critique R telle que pour $\theta \in \Theta_0$, $\mathbb{P}_\theta((X_1, \dots, X_n) \in R) = \alpha$. La fonction de puissance est souvent difficile à déterminer explicitement.

II Tests uniformément plus puissants

Définition 7.3 *Étant donnés deux tests Φ_1 et Φ_2 de niveau $\leq \alpha$ pour tester l'hypothèse $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$. Le test Φ_1 est uniformément plus puissant (u.p.p.) que Φ_2 ssi $\forall \theta \in \Theta_1, \beta_1(\theta) \geq \beta_2(\theta)$. Dans le cas où $\Theta_1 = \{\theta_1\}$ et $\Theta_0 = \{\theta_0\}$, on parle de test plus puissant (p.p.).*

Dans un premier temps, on supposera que H_0 et H_1 sont des hypothèses simples :

$$H_0 : \theta = \theta_0, H_1 : \theta = \theta_1.$$

Soit $L(x_1, \dots, x_n, \theta)$ la fonction de vraisemblance de (X_1, \dots, X_n) .

Définition 7.4 *On considère des hypothèses simples $\Theta_0 = \{\theta_0\}$ et $\Theta_1 = \{\theta_1\}$. Soit*

$$V_{\theta_0, \theta_1}(x) = \frac{L(x, \theta_1)}{L(x, \theta_0)}$$

le rapport de vraisemblance. On considère la famille de tests Φ_k dont la région critique R_k est de la forme :

$$V_{\theta_0, \theta_1}(x) > k.$$

On appellera test de Neyman et Pearson tout test de cette forme.

Dans un test de Neyman et Pearson, on rejette H_0 si $L(x, \theta_1) > kL(x, \theta_0)$, autrement dit, “ θ_1 est plus vraisemblable que θ_0 ”.

Remarque.

Pour toute fonction positive mesurable de \mathbb{R}^n , on a

$$\mathbb{E}_{\theta_1}(f(X)) = \mathbb{E}_{\theta_0}(f(X)V_{\theta_0, \theta_1}(X)) \text{ et } \mathbb{E}_{\theta_0}(f(X)) = \mathbb{E}_{\theta_1}\left(f \frac{1}{V_{\theta_0, \theta_1}(X)}\right).$$

Théorème 7.1 (Lemme de Neyman-Pearson)

1. Soit $\alpha > 0$, si Φ_k est un test de niveau α alors il est p.p. que tout autre test de niveau $\leq \alpha$, de plus il est sans biais.
2. Si $\alpha \in]0, 1[$, si les lois \mathbb{P}_θ sont absolument continues, il existe $k_\alpha \in \mathbb{R}$ tel que Φ_{k_α} est de niveau α . Si les lois \mathbb{P}_θ sont discrètes alors il existe un plus petit k_α tel que Φ_{k_α} est de niveau $\leq \alpha$.
3. Soit Φ un test p.p. de niveau α alors $\forall \theta \in \{\theta_0, \theta_1\}$,

$$\mathbb{P}_\theta(\Phi(X) \neq \Phi_{k_\alpha}(X) \text{ et } V(X) \neq k) = 0.$$

Preuve:

On suppose que $\mathbb{P}_{\theta_0}(\Phi_k = 1) = \alpha$ et on veut montrer que pour tout autre test Φ tel que $\mathbb{P}_{\theta_0}(\Phi = 1) \leq \alpha$ alors $\mathbb{P}_{\theta_1}(\Phi_k = 1) \geq \mathbb{P}_{\theta_1}(\Phi = 1)$, ce que l'on peut formuler de la façon suivante :

$$\mathbb{E}_{\theta_0}(\Phi_k - \Phi) \geq 0 \Rightarrow \mathbb{E}_{\theta_1}(\Phi_k - \Phi) \geq 0.$$

Supposons donc que $\mathbb{E}_{\theta_0}(\Phi_k - \Phi) \geq 0$, soit

$$\Delta_k = \mathbb{E}_{\theta_1}(\Phi_k - \Phi) - k\mathbb{E}_{\theta_0}(\Phi_k - \Phi),$$

et montrons que $\Delta_k \geq 0$. On a :

$$\Delta_k = \mathbb{E}_{\theta_0}((\Phi_k - \Phi)(V - k)).$$

$\Phi_k = 0$ ssi $V \leq k$, dans ce cas $((\Phi_k - \Phi)(V - k)) \geq 0$ et donc $\Delta_k \geq 0$.

$\Phi_k = 1$ ssi $V > k$, dans ce cas $((\Phi_k - \Phi)(V - k)) \geq 0$ et donc $\Delta_k \geq 0$.

Montrons maintenant que le test de N-P est sans biais. On veut montrer que

$$\beta(\theta_1) = \mathbb{P}_{\theta_1}(\Phi_k = 1) \geq \alpha = \mathbb{P}_{\theta_0}(\Phi_k = 1).$$

On a

$$\beta(\theta_1) = \mathbb{E}_{\theta_0}(\mathbb{1}_{\{V > k\}} \cdot V) \geq k\mathbb{P}_{\theta_0}(V > k) = k\alpha.$$

Ainsi, si $k \geq 1$, on a montré que $\beta(\theta_1) \geq \alpha$. Dans le cas où $k < 1$, on montre que $1 - \beta(\theta_1) \leq 1 - \alpha$. On a :

$$\begin{aligned} 1 - \beta(\theta_1) &= \mathbb{P}_{\theta_1}(V \leq k) = \mathbb{E}_{\theta_0}(\mathbb{1}_{V \leq k} \cdot V) \\ &\leq k \mathbb{P}_{\theta_0}(V \leq k) = k(1 - \alpha). \end{aligned}$$

Ceci termine la preuve du point 1.

Fixons $\alpha \in]0, 1[$. Notons $A(k) = \mathbb{P}_{\theta_0}(V > k)$, c'est une fonction décroissante de k qui vérifie

$$\lim_{k \rightarrow \infty} A(k) = 0 \text{ et } \lim_{k \rightarrow 0} A(k) = 1.$$

Si $\mathbb{P}_{\theta_0}(V = k) = 0$ pour tout k , ce qui est vérifié si \mathbb{P}_{θ_0} est absolument continue, alors $A(k)$ est continue. Ainsi, il existe k_α tel que $A(k_\alpha) = \alpha$. Dans le cas discret, il existe k_α tel que $A(k_\alpha) \leq \alpha$ et $A(k) > \alpha$ pour $k < k_\alpha$.

Pour démontrer le point 3., on reprend la preuve du point 1. Si Φ est un test p.p. de niveau α alors $\Delta_k = 0$, or

$$(\Phi_k - \Phi)(V - k) \geq 0$$

donc presque sûrement, $(\Phi_k - \Phi)(V - k) = 0$, autrement dit

$$\mathbb{P}_{\theta_0}(\Phi_k \neq \Phi \text{ et } V \neq k) = 0.$$

Pour obtenir le résultat pour θ_1 , on remarque que

$$\Delta_k = \mathbb{E}_{\theta_1} \left((\Phi_k - \Phi) \left(1 - \frac{k}{V} \right) \right).$$

□

Ce résultat permet de construire des tests optimisant la puissance dans le cas d'hypothèses simples. Le point 3. montre que les tests de Neymann et Pearson sont les seuls tests p.p. (dans le cas absolument continu). Dans le cas d'hypothèses composites (par exemple H_1 n'est pas réduite à un point), le lemme de Neyman et Pearson permet d'obtenir des tests u.p.p., si les rapports de vraisemblance sont monotones.

Proposition 7.1 *Si T est une statistique exhaustive dont la fonction de vraisemblance $g(t, \theta)$ vérifie : pour $\theta > \theta_0$,*

$$\frac{g(t, \theta)}{g(t, \theta_0)} \text{ est une fonction croissante de } t.$$

Alors le test de zone de rejet $\widetilde{R}_k = \{T > k\}$ est u.p.p. pour tester $\theta = \theta_0$ contre $\theta > \theta_0$.

Preuve:

On sait que le test de Neymann-Pearson est p.p pour tester $\theta = \theta_0$ contre $\theta = \theta_1$. Soit $\theta > \theta_0$, comme T est une statistique exhaustive, on a :

$$V_{\theta_0, \theta}(x) = \frac{L(x, \theta)}{L(x, \theta_0)} = \frac{g(t, \theta)}{g(t, \theta_0)}.$$

Ainsi, pour $\theta > \theta_0$, $V_{\theta_0, \theta}$ est une fonction croissante de t et donc, $V > k$ ssi $T > k'$ détermine un test p.p pour tester $\theta = \theta_0$ contre $\theta = \theta_1$ pour tout $\theta_1 > \theta_0$, c'est donc un test u.p.p. □

Remarque.

En général, la zone de rejet que l'on obtient peut s'écrire sous la forme : $R = \{Z \notin I\}$ où Z est une statistique et I un intervalle de la forme $[k, \infty[$ ou $] - \infty, k]$ ou encore $[k_1, k_2]$. I est la région de confiance et Z s'appelle la statistique du test.

Lorsqu'il n'existe pas de tests uniformément plus puissant, on peut avoir recours aux tests du maximum de vraisemblance.

III Tests fondés sur le rapport du maximum de vraisemblance

Définition 7.5 On appellera test du maximum de vraisemblance tout test fondé sur la région critique

$$W = \left\{ x \in \mathbb{R}^n / \frac{\sup_{\theta \in \Theta_1} L(x, \theta)}{\sup_{\theta \in \Theta_0} L(x, \theta)} > k_\alpha \right\},$$

où k_α est choisit tel que $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(W) = \alpha$.

Dans le cas où $\Theta_1 = \Theta_0^c$, on considèrera le test de région critique :

$$R = \left\{ x \in \mathbb{R}^n / \lambda(x) = \frac{\sup_{\theta \in \Theta} L(x, \theta)}{\sup_{\theta \in \Theta_0} L(x, \theta)} > k_\alpha \right\}.$$

Remarque.

En général, pour des hypothèses alternatives unilatérales (c'est à dire du type $\theta > \theta_0$ ou $\theta < \theta_0$), il existe un test uniformément plus puissant, ce n'est pas le cas pour les hypothèses alternatives bilatérales (par exemple du type $\theta \neq \theta_0$).

Décision et "p-value"

D'un point de vue pratique, lorsqu'on procède à un test, on fixe l'hypothèse H_0 , par exemple $\theta = \theta_0$, on choisit une hypothèse alternative H_1 :

$H_1 : \theta \neq \theta_0$ hypothèse bilatérale

$H_1 : \theta < \theta_0$ ou $H_1 : \theta > \theta_0$ hypothèses unilatérales.

On se fixe un risque de première espèce α , on détermine la région critique (i.e. pour un test basé sur le rapport de vraisemblance, la valeur de k_α), on calcule la valeur expérimentale de la statistique du test Z_{exp} , si Z_{exp} est dans la région critique, on rejette H_0 (et on retient H_1), sinon on retient H_0 .

La plupart des logiciels de statistique permettent de calculer la valeur de k_α mais fournissent aussi un autre renseignement : la “p-value” p . Si le test est de région critique $f(Z) > k$ où Z est la statistique du test ($f(x) = x$ ou $f(x) = |x|$), la p-value est la probabilité : $p_{value} = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(f(T) > Z_{exp})$, on remarque que $Z_{exp} \geq k \Leftrightarrow p_{value} \leq \alpha$.

Si $p < \alpha$, on rejette H_0 (et on retient H_1), sinon, on retient H_0 . On prendra garde que certains logiciels ne fournissent la valeur de p que pour des hypothèses alternatives bilatérales. Dans le cas de distributions symétriques (normale, Student), on passe du $p_{bilatéral}$ au $p_{unilatéral}$ en divisant par 2.

IV Exemples

1 Adéquation d’une moyenne pour un échantillon gaussien

On considère un modèle paramétrique gaussien $\mathcal{N}(\mu, \sigma^2)$. On veut tester $H_0 : \mu = \mu_0$.

Test unilatère

Dans ce cas $H_1 : \mu > \mu_0$ ou $H_1 : \mu < \mu_0$. On va déterminer un test u.p.p en utilisant le Lemme de Neymann-Pearson. Le rapport des vraisemblances s’écrit :

$$V_{\mu_0, \mu} = \exp \left[-\frac{1}{2\sigma^2} (n(\mu^2 - \mu_0^2) - 2n\bar{x}(\mu - \mu_0)) \right].$$

Si $\mu > \mu_0$, ce rapport est une fonction croissante de \bar{x} . Ainsi le test de Neymann et Pearson a pour région critique : $\{\bar{X} > k\}$. Ce test est u.p.p. pour $H_1 : \mu > \mu_0$.

Pour tester $H_1 : \mu < \mu_0$, le test de zone de rejet $\{\bar{X} < k'\}$ est u.p.p.

Dans le cas où σ est connu, pour déterminer k_α tel que le test soit de niveau α , on utilise que sous H_0 ,

$$Z = \sqrt{n} \frac{\bar{X} - \mu_0}{\sigma}$$

suit une loi $\mathcal{N}(0, 1)$. Ainsi, le test de région de rejet :

$\{Z > q_{1-\alpha}\}$ avec $q_{1-\alpha}$ le $1 - \alpha$ quantile d'une loi $\mathcal{N}(0, 1)$

est de sensibilité α .

Dans le cas où σ n'est pas connu, on considère le test de région de rejet $\{Z > k\}$ avec

$$Z = \sqrt{n} \frac{\bar{X} - \mu_0}{S}$$

suit une loi de Student à $n - 1$ degrés de liberté.

Test bilatère

On considère les hypothèses du test : $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$. Dans ce cas, il n'existe pas de test u.p.p. On peut néanmoins construire un test basé sur le rapport de vraisemblance. On a $\Theta_0 = \{(\mu_0, \sigma^2) / \sigma > 0\}$. Déterminons l'estimateur du maximum de vraisemblance sur Θ_0 :

$$\ln L(x_1, \dots, x_n, \mu_0, \sigma^2) = -\frac{n}{2} \ln \sqrt{2\pi} - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_0)^2.$$

L'EMV sur Θ_0 est $\hat{\sigma}_0 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_0)^2$. Finalement,

$$\lambda(x) = \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} \right)^{\frac{n}{2}} = \left(1 + \frac{(\bar{x} - \mu_0)^2}{\hat{\sigma}^2} \right)^{\frac{n}{2}}.$$

Ainsi, λ est une fonction croissante de $|T|$, avec

$$T = \sqrt{n-1} \frac{\bar{X} - \mu_0}{\hat{\sigma}}.$$

On utilise que T suit une loi de Student à $n - 1$ d.d.l. pour déterminer la zone de rejet de la forme $\{|T| > k\}$ et de sensibilité α .

2 Comparaison de deux moyennes

Échantillons indépendants On considère deux échantillons aléatoires indépendants et indépendants entre eux : $(X_1, \dots, X_{n_1}), (Y_1, \dots, Y_{n_2})$; les X_i suivent une loi $\mathcal{N}(\mu_X, \sigma^2)$, les Y_j suivent une loi $\mathcal{N}(\mu_Y, \sigma^2)$. L'espace des paramètres est alors

$$\Theta = \{\theta = (\mu_X, \mu_Y, \sigma^2) / \mu_X, \mu_Y \in \mathbb{R}, \sigma^2 > 0\} = \mathbb{R}^2 \times \mathbb{R}^{+*}.$$

Montrer que l'estimateur du maximum de vraisemblance est $(\bar{X}, \bar{Y}, \delta^2)$ avec

$$\delta^2 = \frac{1}{n} \left(\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2 \right) \quad n = n_1 + n_2.$$

On veut tester :

$$H_0 : \mu_X = \mu_Y$$

$$H_1 : \mu_X \neq \mu_Y.$$

Sur $\Theta_0 = \{\theta \in \Theta / \mu_X = \mu_Y = \mu \in \mathbb{R}\}$, montrer que l'estimateur du maximum de vraisemblance vaut $(\hat{\mu}_0, \hat{\mu}_0, \delta_0^2)$ avec

$$\hat{\mu}_0 = \frac{1}{n} \left(\sum_{i=1}^{n_1} X_i + \sum_{i=1}^{n_2} Y_i \right) \quad \delta_0^2 = \frac{1}{n} \left(\sum_{i=1}^{n_1} (X_i - \hat{\mu}_0)^2 + \sum_{i=1}^{n_2} (Y_i - \hat{\mu}_0)^2 \right).$$

Montrer que $\log \lambda(x) = \frac{n}{2} \log \frac{\delta_0^2}{\tilde{\delta}^2}$. On pose

$$T_n = \sqrt{\frac{n_1 \cdot n_2}{n} \frac{\bar{X} - \bar{Y}}{\tilde{S}}},$$

où $\tilde{S}^2 = \frac{n}{n-2} \delta^2$. On note t_n le réel correspondant à T_n . Montrer que sous H_0 , T_n suit une loi de Student à $n - 2$ degrés de liberté.

Montrer que λ est une fonction croissante de $|t_n|$. La région critique du test de sensibilité α est donc donnée par $|T_n| > t_{1-\alpha/2}$.

Proposer un test pour $H_1 : \mu_X > \mu_Y$ et $H_1 : \mu_X < \mu_Y$.

Application On veut mettre en évidence l'influence d'un régime particulier sur la croissance de certains rats. Pour cela on soumet un lot de 5 rats à ce régime, on obtient les poids suivants :

177, 195, 159, 164, 172.

Les poids de 8 rats soumis à un régime normal sont :

145, 146, 169, 151, 142, 170, 126, 153.

Le régime a-t-il une influence sur le poids des rats ?

3 Un exemple avec une loi discrète

On considère X_1, \dots, X_n des variables indépendantes de Bernoulli de paramètre $\theta \in [0, 1]$, on note $S = X_1 + \dots + X_n$.

1. Quelle est la fonction de vraisemblance $L_\theta(x)$, $x = (x_1, \dots, x_n)$, de l'échantillon aléatoire (X_1, \dots, X_n) ? (on pourra l'exprimer en fonction de $s = x_1 + \dots + x_n$).

2. Montrer que si $\theta_1 > \theta_0$, le rapport de vraisemblance $\frac{L_{\theta_1}(x)}{L_{\theta_0}(x)}$ est une fonction croissante de s .

3. Pour tester $\theta = \theta_0$ contre $\theta = \theta_1$ avec $\theta_0 < \theta_1$, on considère la zone de rejet $S \geq k$. Montrer que ce test est sans biais et p.p. Déterminer un test u.p.p. pour tester $\theta = \theta_0$ contre $\theta > \theta_0$.

4. Montrer que pour tout $k \in \{0, \dots, n\}$, l'application $\theta \mapsto \mathbb{P}_\theta(S \geq k)$ de $[0, 1]$ dans $[0, 1]$ est croissante. Pour tester $\theta \leq \theta_0$ contre $\theta = \theta_1$, avec $\theta_1 > \theta_0$, on considère la zone de rejet $S \geq k$. Montrer que la sensibilité de ce test s'écrit :

$$\alpha = \sum_{i=k}^n C_n^i \theta_0^i (1 - \theta_0)^{n-i}.$$

5. Exemple : on étudie le risque de rechute après une angine. Sur 30 patients, 11 on rechuté. On note θ le taux de rechute. Modéliser le problème à l'aide d'un modèle de Bernouilli. Utiliser les questions précédentes pour tester l'hypothèse : *le taux de rechute est inférieur à 30%* contre *le taux de rechute est > à 30%*.

V Tests asymptotiques

Il est parfois difficile de déterminer explicitement la région critique d'un test de Neyman-Pearson ou d'un test basé sur le rapport de vraisemblance. Dans ce cas, on peut avoir recours à des tests asymptotiques.

Définition 7.6 *On considère une suite $(E^{(N)}, \mathcal{B}^{(N)}, (\mathbb{P}_\theta^{(N)})_{\theta \in \Theta})$ de modèles d'échantillonnages paramétriques ayant le même espace de paramètres Θ . On note $X^{(N)}$ le vecteur aléatoire correspondant.*

Le niveau asymptotique d'une suite de tests de Θ_0 contre Θ_1 , de région de rejet $R^{(N)}$ est la limite (lorsqu'elle existe)

$$\alpha = \lim_{N \rightarrow \infty} \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(X^{(N)} \in R_{(N)}).$$

On dit que la suite de tests est convergente si

$$\forall \theta \in \Theta_1 \lim_{N \rightarrow \infty} \mathbb{P}_\theta(X^{(N)} \in R_{(N)}) = 1.$$

Le plus souvent, on obtient des tests asymptotiques à l'aide de statistiques dont on connaît la loi asymptotique.

Exercice 5 *On reprend l'exemple ci-dessus sur la loi de Bernouilli. Montrer que le test de région de confiance :*

$$\left\{ \frac{S - n\theta_0}{\sqrt{n\hat{\sigma}}} > z_{1-\alpha} \right\}$$

avec $z_{1-\alpha}$ le $1 - \alpha$ quantile d'une loi $\mathcal{N}(0, 1)$ est de sensibilité asymptotique α et convergent pour tester $\theta = \theta_0$ contre $\theta > \theta_0$.

Proposer un test de sensibilité asymptotique α et convergent pour tester $\theta = \theta_0$ contre $\theta \neq \theta_0$.

1 Propriétés asymptotiques des tests du maximum de vraisemblance

On se place dans le cas où $\Theta_0 = \{\theta_0\}$ et $\Theta_1 = \{\theta \neq \theta_0\}$. On considère alors

$$\lambda_n(X) = \frac{\sup_{\theta \in \Theta} L_n(X, \theta)}{L_n(X, \theta_0)}.$$

Théorème 7.2 *On suppose que les hypothèses de régularité du modèle (H1) – (H7) sont satisfaites et que $\Theta \subset \mathbb{R}$ (paramètre de dimension 1). La suite de test de région critique :*

$$R_n = \{2 \ln \lambda_n > K\},$$

où K est le $1 - \alpha$ quantile d'une loi $\chi^2(1)$ est de sensibilité asymptotique α et convergente.

Preuve:

On a

$$\begin{aligned} 2 \ln \lambda_n &= 2[\ln L(X, \hat{\theta}) - \ln L(X, \theta_0)] \\ &= 2[(\hat{\theta} - \theta_0) \frac{\partial}{\partial \theta} \ln L(X, \hat{\theta}) + \frac{1}{2}(\hat{\theta} - \theta_0)^2 \frac{\partial^2}{\partial \theta^2} \ln L(X, \theta^*)] \\ &= (\hat{\theta} - \theta_0)^2 \frac{\partial^2}{\partial \theta^2} \ln L(X, \theta^*). \end{aligned}$$

En reprenant les arguments de la preuve de la normalité asymptotique de l'EMV, on montre que si $\theta = \theta_0$

$$\frac{1}{n} \frac{\partial^2}{\partial \theta^2} \ln L(X, \theta^*) \xrightarrow{\mathbb{P}} I(\theta_0).$$

Comme, de plus $\sqrt{I_n(\theta_0)}(\theta_0 - \hat{\theta}) \longrightarrow \mathcal{N}(0, 1)$ et donc $I_n(\theta_0)(\theta_0 - \hat{\theta})^2 \longrightarrow \chi_1^2$, on en déduit que $2 \ln \lambda \longrightarrow \chi_1^2$.

La convergence du test est plus délicate à démontrer.

□

2 Tests de Wald et du score

Ces tests asymptotiques sont basés sur les propriétés asymptotiques d'estimateurs.

Proposition 7.2 *On considère $\tilde{\theta}_n$ une suite d'estimateurs asymptotiquement efficace d'un paramètre $\theta \in \mathbb{R}^d$, on suppose de plus que $I(\tilde{\theta}_n) \xrightarrow{\mathbb{P}_{\theta}} I(\theta)$. Soit $g : \Theta \rightarrow \mathbb{R}^k$ une fonction C^1 de matrice jacobienne $D(\theta)$ de rang k . On considère $\Theta_0 = \{\theta \in \Theta / g(\theta) = 0\}$. Le test de Wald de région de rejet :*

$$R_n : \xi_n > \chi_{1-\alpha}^2(k) \text{ avec } \xi_n = ng(\tilde{\theta}_n)^t \left(D(\tilde{\theta}_n)I(\tilde{\theta}_n)^{-1}D(\tilde{\theta}_n)^t \right)^{-1} g(\tilde{\theta}_n).$$

est de sensibilité asymptotique α et convergent pour $\Theta_1 = \{\theta \in \Theta / g(\theta) \neq 0\}$.

La preuve de cette proposition utilise les deux lemmes suivants.

Lemme 7.3 *Soit X un vecteur aléatoire gaussien de \mathbb{R}^d , d'espérance μ et de matrice de covariance Σ . On suppose que Σ est inversible. Alors*

$$D^2 = (X - \mu)^t \Sigma^{-1} (X - \mu)$$

suit une loi du $\chi^2(d)$.

Preuve:

Il suffit de remarquer que la loi de X est la même que celle de $AX_0 + \mu$ avec A une matrice $d \times d$ telle que $AA^t = \Sigma$ et X_0 un vecteur gaussien centré et réduit. On montre ainsi que la loi de D^2 est la même que celle de $X_0^t X_0$ qui est une somme de d carrés de lois normales centrée et réduites et indépendantes.

□

Lemme 7.4 *Soit X_n une suite de vecteurs aléatoires de \mathbb{R}^d telle que $X_n \xrightarrow{\mathcal{L}} X$ et A une matrice $d \times k$. Alors la suite de vecteurs aléatoires de \mathbb{R}^k AX_n converge en loi vers AX .*

Preuve:

Considérons φ_{AX_n} la fonction caractéristique de AX_n , pour $t \in \mathbb{R}^k$,

$$\varphi_{AX_n}(t) = \mathbb{E}(\exp(i \langle t, AX_n \rangle)) = \mathbb{E}(\exp(i \langle A^t t, X_n \rangle)) = \varphi_{X_n}(A^t t).$$

□

Preuve de la proposition:

La suite d'estimateurs $\tilde{\theta}_n$ est asymptotiquement efficace, autrement dit :

$$\sqrt{n}(\tilde{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I^{-1}(\theta)),$$

en particulier, $\tilde{\theta}_n \xrightarrow{\mathbb{P}_\theta} \theta$. On applique la méthode Δ pour obtenir :

$$\sqrt{n}(g(\tilde{\theta}_n) - g(\theta)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, A_\theta)$$

avec $A_\theta = D(\theta)I^{-1}(\theta)D(\theta)^t$. Si $\theta \in \Theta_0$ alors $g(\theta) = 0$ et

$$\sqrt{n}g(\tilde{\theta}_n) \xrightarrow{\mathcal{L}} \mathcal{N}(0, A_\theta).$$

Comme $D(\theta)$ est de rang k , la matrice A_θ est inversible est on a donc :

$$ng(\tilde{\theta}_n)^t A_\theta^{-1} g(\tilde{\theta}_n) \xrightarrow{\mathcal{L}} \chi^2(k).$$

De plus, $I(\tilde{\theta}_n) \xrightarrow{\mathbb{P}} I(\theta)$, ainsi :

$$D(\tilde{\theta}_n)I^{-1}(\tilde{\theta}_n)D(\tilde{\theta}_n)^t \xrightarrow{\mathbb{P}} A_\theta$$

et finalement, $\xi_n \longrightarrow \chi^2(k)$ pour $\theta \in \Theta_0$. Ceci montre que le test est de sensibilité asymptotique α .

Si $\theta \in \Theta_1$,

$$Z_n = g(\tilde{\theta}_n)^t \left[D(\tilde{\theta}_n)I^{-1}(\tilde{\theta}_n)D(\tilde{\theta}_n)^t \right]^{-1} g(\tilde{\theta}_n) \xrightarrow{\mathbb{P}_\theta} g(\theta)^t A_\theta^{-1} g(\theta) = u > 0.$$

Ainsi, $\mathbb{P}(\xi_n > \chi_{1-\alpha}^2(k)) \geq \mathbb{P}(Z > \frac{\chi_{1-\alpha}^2(k)}{n}) \geq \mathbb{P}(Z > \frac{u}{2})$

lorsque n est suffisamment grand. Ceci montre la convergence du test. □

Les tests du score sont basés sur la région de rejet :

$$\xi_n^S > \chi_{1-\alpha}^2(k) \text{ avec } \xi_n^S = \frac{1}{n} DL_n(\hat{\theta}_{0,n})^t I(\hat{\theta}_{0,n})^{-1} DL_n(\hat{\theta}_{0,n}),$$

où DL_n est le gradient de la vraisemblance et $\hat{\theta}_{0,n}$ l'EMV de θ sur Θ_0 .

Chapitre 8

Tests paramétriques classiques

Notations :

P : proportion sur un échantillon aléatoire pour une variable de Bernoulli : X prend les valeurs 0 ou 1 avec probabilité $1 - \pi$ et π ,

$$P = \frac{X_1 + \dots + X_n}{n}.$$

\bar{X} : moyenne sur un échantillon aléatoire de taille n d'espérance μ .

S^2 est la variance estimée définie par :

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

Lorsque $\mathbb{E}(X) = \mu$ est connue, on utilise aussi :

$$D^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2.$$

I Tests gaussiens

Test	Hypothèses	Stat. du test (Z) et cond. d'appl.	Loi de Z sous H_0	Remarques
conform. d'une moy.	$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$ ou $H_1 : \mu > \mu_0$ ou $H_1 : \mu < \mu_0$	$\frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$, la variable étudiée doit suivre une loi normale	Student à $n - 1$ d.d.l.	Si σ est connu, on peut utiliser directement \bar{X} qui suit une loi normale de paramètre $\mathcal{N}(\mu_0, \sigma)$ sous H_0 .
comp. de deux moy. pour des ech. indép.	$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \neq \mu_2$ ou $H_1 : \mu_1 < \mu_2$ ou $H_1 : \mu_1 > \mu_2$	$\frac{\bar{X}^1 - \bar{X}^2}{\frac{1}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}} \times \frac{1}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}}$, la variable étudiée doit suivre une loi normale. Il faut que les écart-types σ_1 et σ_2 soient les mêmes.	Student à $n_1 + n_2 - 2$ d.d.l.	Avant de faire ce test, on doit tester l'égalité des variances avec un test de Fisher-Snedecor. Si on accepte l'hypothèse $\sigma_1 = \sigma_2$, on estime la valeur commune $\sigma_1 = \sigma_2 = \sigma$ (voir ci-dessous). Si on refuse l'hypothèse $\sigma_1 = \sigma_2$, on ne peut pas faire le test.
comp. de deux moy. pour des ech. appariés, on pose $Y = X_1 - X_2$	$H_0 : \mu_Y = 0$ $H_1 : \mu_Y \neq 0$ ou $H_1 : \mu_Y < 0$ ou $H_1 : \mu_Y > 0$	$\frac{\bar{X}_Y}{\frac{S_Y}{\sqrt{n}}}$, la variable étudiée doit suivre une loi normale.	Student à $n - 1$ d.d.l.	
conf. d'une variance	$H_0 : \sigma = \sigma_0$ $H_1 : \sigma \neq \sigma_0$ ou $H_1 : \sigma > \sigma_0$ ou $H_1 : \sigma < \sigma_0$	$\frac{(n-1)S^2}{\sigma_0^2}$ La variable étudiée doit suivre une loi normale	χ^2 à $(n-1)$ d.d.l.	Il faut tester la normalité. Si μ est connue, on peut remplacer S^2 par D^2 et on a une loi $\chi^2(n)$.
comp de deux variances	$H_0 : \sigma_1 = \sigma_2$ $H_1 : \sigma_1 \neq \sigma_2$ ou $H_1 : \sigma_1 > \sigma_2$ ou $H_1 : \sigma_1 < \sigma_2$	$\frac{S_1^2}{S_2^2}$. La variable étudiée doit suivre une loi normale	Fisher Snedecor à $(n_1 - 1, n_2 - 1)$ d.d.l.	Lorsqu'on accepte H_0 , on estime l'écart-type commun par $\sigma = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$

II Tests asymptotiques

Test	Hypothèses	Stat. du test (Z) et cond. d'appl.	Loi de Z sous H_0	Remarques
comp. de deux prop. pour des ech. indep.	$H_0 : \pi_1 = \pi_2 = \pi_0$ $H_1 : \pi_1 \neq \pi_2$ ou $H_1 : \pi_1 < \pi_2$ ou $H_1 : \pi_1 > \pi_2$	$\frac{P_1 - P_2}{\sqrt{\hat{\pi}(\hat{\pi} - 1) \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ $n_1 \geq 30$ et $n_2 \geq 30$, avec $\hat{\pi} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$	$\mathcal{N}(0, 1)$	Il s'agit d'un test asymptotique.
comp. de deux moy. pour des ech. indep.	$H_0 : \mu_1 = \mu_2 = \mu_0$ $H_1 : \mu_1 \neq \mu_2$ ou $H_1 : \mu_1 < \mu_2$ ou $H_1 : \mu_1 > \mu_2$	$\frac{\overline{X^1} - \overline{X^2}}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}}$ $\times \frac{1}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$, $n_1 \geq 30$ et $n_2 \geq 30$	$\mathcal{N}(0, 1)$	Pour les grands échantillons, il n'est pas nécessaire d'avoir l'égalité des variances. Il s'agit d'un test asymptotique.
comp. de deux moy. pour des ech. appariés, on pose $Y = X_1 - X_2$	$H_0 : \mu_Y = 0$ $H_1 : \mu_Y \neq 0$ ou $H_1 : \mu_Y < 0$ ou $H_1 : \mu_Y > 0$	$\frac{m_Y}{\frac{s_Y}{\sqrt{n}}}$, $n \geq 30$	$\mathcal{N}(0, 1)$	Il s'agit d'un test asymptotique.
conform. d'une prop.	$H_0 : \pi = \pi_0$ $H_1 : \pi \neq \pi_0$ ou $H_1 : \pi > \pi_0$ ou $H_1 : \pi < \pi_0$	$\frac{P - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$, $n \geq 30$	$\mathcal{N}(0, 1)$	Il s'agit d'un test asymptotique
conform. d'une moy.	$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$ ou $H_1 : \mu > \mu_0$ ou $H_1 : \mu < \mu_0$	$\frac{\overline{X} - \mu_0}{\frac{s_n}{\sqrt{n}}}$, $n \geq 30$	$\mathcal{N}(0, 1)$	Il s'agit d'un test asymptotique

Chapitre 9

Quelques tests non paramétriques

I Tests du χ^2 .

1 Loi multinômiale

On considère r évènements A_1, \dots, A_r de probabilité p_1, \dots, p_r . On suppose que les A_i forment un système complet d'évènement (i.e. ils sont disjoints et leur union est Ω), en particulier, $\sum_{i=1}^r p_i = 1$.

On répète n fois, de manière indépendante, l'expérience aléatoire dont le résultat est l'un des A_i (penser à un tirage avec remise de n boules dans une urne qui contient des boules de r couleurs différentes, p_i est alors la proportion de boules de couleur i).

On note N_i la variable aléatoire qui donne le nombre de fois (parmi les n expériences) où l'évènement A_i se produit.

N_i suit une loi Binômiale $\mathcal{B}(n, p_i)$.

La loi du vecteur (N_1, \dots, N_r) est donnée par :

$$\mathbb{P}(N_1 = n_1, \dots, N_r = n_r) = \frac{n!}{n_1! \dots n_r!} p_1^{n_1} \times \dots \times p_r^{n_r},$$

pour $(n_1, \dots, n_r) \in \mathbb{N}^r$ avec $\sum_{i=1}^r n_i = n$. En particulier, les N_i **ne** sont **pas** indépendants.

2 Loi asymptotique

Théorème 9.1 *Soit*

$$D^2 = \sum_{i=1}^r \frac{(N_i - np_i)^2}{np_i}.$$

Alors, D^2 converge en loi (quand $n \rightarrow \infty$) vers une loi du χ^2 à $r - 1$ degrés de liberté.

Preuve:

Chaque variable N_i est la somme de vauid : $N_i = \sum_{j=1}^n X_j^i$ avec $X_j^i = 1$ si et seulement si le $j^{\text{ème}}$ tirage donne A_i , $X_j^i = 0$ sinon. Les X_j^i sont des variables de Bernouilli $\mathcal{B}(p_i)$ vérifiant : pour $k \neq \ell$,

$$\text{cov}(X_j^k, X_j^\ell) = -p_k p_\ell.$$

Ainsi, la matrice de covariance du vecteur $X_j = (X_j^1, \dots, X_j^r)$ est

$$\Sigma = \begin{pmatrix} p_1(1-p_1) & -p_1 p_2 & \dots & -p_1 p_r \\ -p_1 p_2 & \ddots & \ddots & \vdots \\ \vdots & & & -p_{r-1} p_r \\ -p_1 p_r & \dots & -p_{r-1} p_r & p_r(1-p_r) \end{pmatrix}$$

Comme $\sum_{i=1}^r p_i = 1$, cette matrice n'est pas inversible. Néanmoins, si on considère le vecteur $\widetilde{X}_j = (X_j^1, \dots, X_j^{r-1})$, sa matrice de covariance :

$$\widetilde{\Sigma} = \begin{pmatrix} p_1(1-p_1) & -p_1 p_2 & \dots & -p_1 p_{r-1} \\ -p_1 p_2 & \ddots & \ddots & \vdots \\ \vdots & & & -p_{r-2} p_{r-1} \\ -p_1 p_{r-1} & \dots & -p_{r-2} p_{r-1} & p_{r-1}(1-p_{r-1}) \end{pmatrix}$$

est inversible et

$$\left(\widetilde{\Sigma}\right)^{-1} = \begin{pmatrix} \frac{1}{p_1} + \frac{1}{p_r} & \frac{1}{p_r} & \dots & \frac{1}{p_r} \\ \frac{1}{p_r} & \ddots & \ddots & \dots \\ \vdots & & & \frac{1}{p_r} \\ \frac{1}{p_r} & \dots & \frac{1}{p_r} & \frac{1}{p_{r-1}} + \frac{1}{p_r} \end{pmatrix}.$$

Le théorème de la limite centrale multidimensionnel donne pour $\widetilde{N} = \sum_{j=1}^n \widetilde{X}_j - (np_1, \dots, np_{r-1})$:

$$\frac{1}{\sqrt{n}} \widetilde{N} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \widetilde{\Sigma}).$$

Ainsi, $\frac{1}{n} \widetilde{N}^t (\widetilde{\Sigma})^{-1} \widetilde{N} \xrightarrow{\mathcal{L}} \chi^2(r-1)$. Or

$$\begin{aligned}
\frac{1}{n} \tilde{N}^t (\tilde{\Sigma})^{-1} \tilde{N} &= \sum_{j=1}^{r-1} \frac{(N_j - np_j)^2}{np_j} + \sum_{j=1}^{r-1} \sum_{k=1}^{r-1} \frac{(N_j - np_j)(N_k - np_k)}{np_r} \\
&= \sum_{j=1}^{r-1} \frac{(N_j - np_j)^2}{np_j} + \frac{1}{np_r} \left(\sum_{j=1}^{r-1} N_j - np_j \right)^2 \\
&= \sum_{j=1}^r \frac{(N_j - np_j)^2}{np_j}
\end{aligned}$$

$$\text{car } \sum_{j=1}^{r-1} N_j = n - N_r \text{ et } \sum_{j=1}^{r-1} p_j = 1 - p_r.$$

□

Ce test permet la mise en œuvre des tests dits du χ^2 .

3 Test du χ^2 d'adéquation à une loi

On présente ici le test du χ^2 d'adéquation à une loi théorique.

On se demande si une variable aléatoire Y suit une loi donnée notée \mathbb{P}_0 .

Soit Y_1, \dots, Y_n un échantillon aléatoire indépendant de la loi de Y . On fixe une partition de \mathbb{R} à r éléments $\mathbb{R} = C_1 \cup \dots \cup C_r$. On note N_1 le nombre d'indices i tels que $Y_i \in C_1$, ... N_r le nombre d'indices i tels que $Y_i \in C_r$. Soient $p_i^0 = \mathbb{P}_0(Y \in C_i)$ les probabilités théoriques, pour une loi \mathbb{P} , on note $p_i = \mathbb{P}(Y \in C_i)$ et on teste :

H_0 : pour tout $i = 1, \dots, r$, $p_i = p_i^0$

H_1 : il existe i tel que $p_i \neq p_i^0$.

Si on retient H_0 , on conclura que la loi de Y est P_0 .

La statistique du test est

$$D^2 = \sum_{i=1}^r \frac{(N_i - np_i^0)^2}{np_i^0}.$$

On utilise alors le théorème précédent et on rejette H_0 si $D^2 > \chi_{r-1, 1-\alpha}^2$. Ce test est de sensibilité asymptotique α et convergent. La convergence du test provient du fait que si $p_i \neq p_i^0$ pour un i alors en utilisant la loi des grands nombres,

$$\begin{aligned} \frac{1}{n} D^2 &= \sum_{j=1}^r \frac{\left(\frac{N_j}{n} - p_j^0\right)^2}{np_j^0} \\ &\xrightarrow{\mathbb{P}} \sum_{j=1}^r \frac{(p_j - p_j^0)^2}{p_j^0} > 0 \end{aligned}$$

et donc D^2 tend en probabilité vers ∞ .

Si la loi \mathbb{P}_0 appartient à une famille paramétrique, $\mathbb{P}_0 = \mathbb{P}_{\theta_0}$, $\theta \in \mathbb{R}^d$, si on connaît θ_0 , il n'y a pas de différence avec le cas considéré ci-dessus. Si on ne connaît pas θ_0 - par exemple, on se demande si la loi de Y est normale - on doit alors estimer θ_0 . Soit $\hat{\theta}$ un estimateur du maximum de vraisemblance de θ ,

Théorème 9.2 *Soit*

$$\tilde{D}^2 = \sum_{i=1}^r \frac{(N_i - np_i(\hat{\theta}))^2}{np_i(\hat{\theta})}.$$

Alors, sous \mathbb{P}_θ , \tilde{D}^2 converge en loi (quand $n \rightarrow \infty$) vers une loi du χ^2 à $r - d - 1$ degrés de liberté.

On rejette alors H_0 si $\tilde{D}^2 > \chi_{r-d-1, 1-\alpha}^2$. Ce test est de sensibilité asymptotique α et convergent.

D'un point de vue pratique, on considère que l'approximation donnée par le théorème limite ci-dessus est bonne si $n \geq 30$ et que les effectifs théoriques $np_i(\hat{\theta})$ sont supérieures à 5, $i = 1, \dots, r$. Si cette dernière condition n'est pas vérifiée, on procède à des regroupements de classes.

4 Test du χ^2 d'indépendance

On considère $X = (Y, Z)$, et $X_i = (Y_i, Z_i)$ $i = 1, \dots, n$, un échantillon aléatoire de loi P_X . Y_i et Z_i sont des variables discrètes prenant respectivement les valeurs : $\{y_1, \dots, y_\ell\}$ et $\{z_1, \dots, z_m\}$. On veut tester l'indépendance de Y et Z . Le test se base sur le fait que Y et Z sont indépendants si et seulement si $\mathbb{P}_X = \mathbb{P}_Y \otimes \mathbb{P}_Z$.

Si l'hypothèse d'indépendance est satisfaite, $p_{i,j} = q_i r_j$ avec $p_{i,j} = \mathbb{P}(X = (y_i, z_j))$, $q_i = \mathbb{P}(Y = y_i)$, $r_j = \mathbb{P}(Z = z_j)$. On est dans le cadre ci-dessus avec le paramètre $\theta = (q_1, \dots, q_{\ell-1}, r_1, \dots, r_{m-1}) \in \mathbb{R}^{\ell+m-2}$. On estime q_i par $\frac{N_{i.}}{n}$ et r_j par $\frac{N_{.j}}{n}$. Soient

$$D_1^2 = n \sum_{i=1}^{\ell} \sum_{j=1}^m \frac{\left(N_{i,j} - \frac{N_{i.} N_{.j}}{n}\right)^2}{N_{i.} N_{.j}} \quad D_2^2 = \sum_{i=1}^{\ell} \sum_{j=1}^m \frac{\left(N_{i,j} - \frac{N_{i.} N_{.j}}{n}\right)^2}{N_{i,j}}$$

D_1^2 et D_2^2 convergent en loi vers une loi $\chi^2((\ell - 1)(m - 1))$ ($(\ell - 1)(m - 1) = \ell \times m - 1 - (\ell - 1) - (m - 1)$). Les tests associés aux régions de rejet

$$\{D_1^2 > k\} \text{ et } \{D_2^2 > k\}$$

avec k le $1 - \alpha$ quantile d'une loi $\chi^2((\ell - 1)(m - 1))$ sont de sensibilité asymptotique α et convergents pour les hypothèses

$$H_0 : p_{i,j} = q_i r_j \text{ pour tout } (i, j)$$

$$H_1 : \text{il existe } (i, j) \text{ tel que } p_{i,j} \neq q_i r_j.$$

Si on retient H_0 , on retient l'hypothèse d'indépendance.

Ce test permet aussi de tester l'indépendance de variables non catégorielles, dans ce cas il faut discrétiser l'ensemble des valeurs prises.

Exercice 6 *On souhaite procéder à un test de conformité à une loi de Poisson.*

1. *On rappelle que X suit une loi de Poisson de paramètre $\lambda > 0$ si pour tout $n \in \mathbb{N}$,*

$$\mathbb{P}(X = n) = e^{-\lambda} \frac{\lambda^n}{n!}.$$

Déterminer $E(X)$ et $Var(X)$.

2. *On considère un échantillon aléatoire X_1, \dots, X_n , indépendant de loi de Poisson de paramètre λ .*

Quel est l'estimateur du maximum de vraisemblance de λ ?

3. *Proposer un test pour tester :*

$$H_0 : X \text{ suit une loi de Poisson}$$

$$H_1 : X \text{ ne suit pas une loi de Poisson.}$$

4. *Application. Pendant 100 intervalles de 10 minutes, on a compté le nombre X d'ouvriers se présentant à un magasin pour emprunter des outils. Le tableau suivant donne les valeurs observées pour ces 100 mesures et les effectifs correspondants.*

x_i	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
n_i	1	0	1	2	1	3	5	6	9	10	11	9	8	9	7	5	4	3	1	1	1

Peut-on conclure que X suit une loi de Poisson ?

Exercice 7 On procède à un sondage téléphonique, il est demandé aux sondés s'ils sont optimistes ou non quant à leur capacité d'achat pour les années à venir. Les résultats sont présentés par catégories d'âge.

Age	Optimistes	Pas optimistes
[20, 40[237	392
[40, 60[326	298
≥ 60	362	258

Peut-on considérer que le fait d'être optimiste quand à sa capacité d'achat est indépendante de l'âge ?

Un des inconvénients du test d'adéquation du χ^2 est le choix des classes. Cet inconvénient n'est plus présent pour le test de Kolmogorov-Smirnov.

II Test de Kolmogorov-Smirnov

On considère X_1, \dots, X_n un échantillon aléatoire indépendant de même loi que X . Soit F la fonction de répartition de X . On définit la fonction de répartition empirique :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{X_i \leq x\}},$$

Attention, c 'est une variable aléatoire.

Soit

$$D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|$$

Pour tout $x \in \mathbb{R}$, $F_n(x)$ s'écrit comme une somme de variables aléatoires indépendantes de loi $\mathcal{B}(F(x))$. On en déduit :

- $F_n(x)$ converge presque sûrement (et en probabilité) vers $F(x)$,
- $\sqrt{n}(F_n(t) - F(t)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, F(t)(1 - F(t)))$.

Théorème 9.3 (Théorème de Glivenko-Cantelli) D_n converge vers 0 presque sûrement.

Proposition 9.1 La loi de D_n ne dépend pas de la loi de X . Plus précisément, dans le cas où F est strictement croissante et continue, on a les égalités en loi :

$$\begin{aligned} D_n &\stackrel{\mathcal{L}}{=} \sup_{t \in [0,1]} |H_n(t) - t| \\ &\stackrel{\mathcal{L}}{=} \max \left\{ \left| \frac{i}{n} - t \right| / U_{(i)} \leq t < U_{(i+1)} \right\}, \end{aligned}$$

où H_n est la fonction de répartition empirique d'une suite $(U_i)_{i=1, \dots, n}$ de n variables indépendantes de loi uniforme $\mathcal{U}([0, 1])$ et $(U_{(i)})_{i=1, \dots, n}$ désigne les statistiques de l'ordre associées à $(U_i)_{i=1, \dots, n}$ i.e. les $U_{(i)}$ vérifient

$$U_{(1)} < U_{(2)} \cdots < U_{(n)} \text{ et } U_{(1)} = \min_{i=1,\dots,n} U_i, \quad U_{(n)} = \max_{i=1,\dots,n} U_i.$$

Preuve:

La preuve de la proposition repose sur la remarque suivante : $F(X) \rightsquigarrow \mathcal{U}([0, 1])$ et si $U \rightsquigarrow \mathcal{U}([0, 1])$ alors $F^{-1}(U) \rightsquigarrow X$. Ceci implique que pour tout $x \in \mathbb{R}$, $F_n(x) - F(x)$ a la même loi que $H_n(F(x)) - F(x)$. □

La loi de $K_n = \sqrt{n}D_n$ (loi de Kolmogorov-Smirnov à 1 échantillon) est tabulée et converge en loi vers une variable aléatoire K elle aussi tabulée. Ce résultat permet de tester si un échantillon provient d'une loi théorique connue. Attention : le résultat n'est pas valable si les paramètres de la loi sont estimés. C'est un des gros inconvénient du test de Kolmogorov-Smirnov!! Avec R : `ks.test`.

Si Y_1, \dots, Y_m est un échantillon de la loi de Y . On note $G_m(x)$ la fonction de répartition empirique associée. Soit

$$D_{n,m} = \sup_{x \in \mathbb{R}} |F_n(x) - G_m(x)|.$$

Si la loi de X est la même que celle de Y , la loi de $D_{n,m}$ est la même que la loi de $\sup_{t \in [0,1]} |H_n(t) - I_m(t)|$ où H_n et I_m sont les fonction de répartition empiriques de suites de variables aléatoires uniformes $\mathcal{U}([0, 1])$, on a aussi l'égalité en loi :

$$D_{n,m} \stackrel{\mathcal{L}}{=} \max \left\{ \left| \frac{i}{n} - \frac{j}{m} \right|, U_{(i)} < V_{(j)} < U_{(i+1)} \right\},$$

où $(U_{(i)})_{i=1,\dots,n}$ désigne les statistiques de l'ordre associées à $(U_i)_{i=1,\dots,n}$, $U_i \rightsquigarrow \mathcal{U}([0, 1])$ et $(V_{(j)})_{j=1,\dots,m}$ désigne les statistiques de l'ordre associées à $(V_j)_{j=1,\dots,m}$, $V_j \rightsquigarrow \mathcal{U}([0, 1])$. Pour

$$c_{n,m} = \left(\frac{1}{n} + \frac{1}{m} \right)^{-\frac{1}{2}},$$

$K_{n,m} = c_{n,m}D_{n,m}$ suit une loi de Kolmogorov-Smirnov à deux échantillons et converge en loi vers une variable aléatoire K_2 elle aussi identifiée. On peut donc tester si deux échantillons proviennent de la même loi (avec R : `ks.test`). Ce test est une alternative au test paramétrique de Student de comparaison de deux moyennes.

III Test de Shapiro-Wilk

Le test de Shapiro-Wilk permet de tester la normalité d'un échantillon, quel que soit sa taille et sans estimer les paramètres de la loi.

1 Droite de Henry

Il s'agit de représenter les quantiles théoriques d'une loi "connue" en fonction des données x_i .

Soit F_i les fréquences cumulées empiriques, on note u_i^* le quantile de la loi théorique correspondant : $\mathbb{P}(Z \leq u_i^*) = F_i$. Si le graphe (x_i, u_i^*) est quasiment une droite, alors, la loi empirique est proche d'une transformation affine de la loi théorique. En particulier, si la loi théorique considérée est une loi $\mathcal{N}(0, 1)$ alors la loi empirique est proche d'une loi normale (avec R : `qqnorm`, la commande `qqline` rajoute une droite qui passe par les premiers et troisièmes quantiles, la commande `qqplot` trace la droite de Henry pour deux échantillons).

Exemple : la distribution suivante qui donne des résultats d'essais de fatigue d'un matériau (nombre de cycles avant rupture) :

225	31	400	62	850	39	89	580	115	442	270	125	342	251	140
-----	----	-----	----	-----	----	----	-----	-----	-----	-----	-----	-----	-----	-----

2 Test de Shapiro-Wilk

Ce test est spécifique à la loi normale. Son principal avantage est qu'il ne requière pas d'estimation préalable des paramètres. L'idée est de tester la proximité du nuage de points des écarts inter-quartiles empiriques et des écarts inter-quartiles d'une loi normale centrée réduite, à la droite des moindres carrés correspondante. Si (U_1, \dots, U_n) est un échantillon aléatoire indépendant de loi $\mathcal{N}(0, 1)$, on note $V = (V_1, \dots, V_n)$ l'échantillon ordonné : $V_1 = \min_{i=1, \dots, n} U_i$, ..., $V_n = \max_{i=1, \dots, n} U_i$ (voir ci-dessous pour les détails sur les statistiques de l'ordre). μ est le vecteur d'espérance de V , $\Sigma = \mathbb{E}[(V - \mu)(V - \mu)^t]$, enfin

$$a^t = \mu^t \Sigma^{-1} (\mu^t \Sigma^{-2} \mu)^{-\frac{1}{2}},$$

$$a = (a_1, \dots, a_n).$$

Étant donné un échantillon aléatoire indépendant $X = (X_1, \dots, X_n)$ et $Y = (Y_1, \dots, Y_n)$ sa statistique de l'ordre, on définit :

$$T_n = \frac{1}{n-1} \left[\sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} a_{n-i+1} (Y_{n-i+1} - Y_i) \right]^2.$$

Sous H_0 : les X_i suivent des lois normales $\mathcal{N}(\nu, \sigma)$, T_n est un estimateur asymptotiquement sans biais de σ^2 .

La statistique du test de Shapiro-Wilk est :

$$SW = \frac{T_n}{S_n^2}.$$

La loi de SW est indépendante de ν et σ , cette statistique est aussi indépendante de \bar{X}_n et de S_n^2 .

Pour mettre en œuvre ce test, on dispose de tables qui donnent les a_i et les valeurs critiques de la statistique SW .

j	n	2	3	4	5	6	7	8	9	10	-
1		0,7071	0,7071	0,6872	0,6646	0,6431	0,6233	0,6052	0,5888	0,5739	-
2			0	0,1677	0,2413	0,2806	0,3031	0,3164	0,3244	0,3291	-
3					0	0,0875	0,1401	0,1743	0,1976	0,2141	-
4							0	0,0561	0,0947	0,1224	-
5									0	0,0399	-
j	n	11	12	13	14	15	16	17	18	19	20
1		0,5601	0,5475	0,5359	0,5251	0,515	0,5056	0,4963	0,4886	0,4808	0,4734
2		0,3315	0,3325	0,3325	0,3318	0,3306	0,329	0,3273	0,3253	0,3232	0,3211
3		0,226	0,2347	0,2412	0,246	0,2495	0,2521	0,254	0,2553	0,2561	0,2565
4		0,1429	0,1586	0,1707	0,1802	0,1878	0,1939	0,1988	0,2027	0,2059	0,2085
5		0,0695	0,0922	0,1099	0,124	0,1353	0,1447	0,1524	0,1587	0,1641	0,1686
6		0	0,0303	0,0539	0,0727	0,088	0,1005	0,1109	0,1197	0,1271	0,1334
7				0	0,024	0,0433	0,0593	0,0725	0,0837	0,0932	0,1013
8						0	0,0196	0,0359	0,0496	0,0612	0,0711
9								0	0,0163	0,0303	0,0422
10										0	0,014
j	n	21	22	23	24	25	26	27	28	29	30
1		0,4643	0,459	0,4542	0,4493	0,445	0,4407	0,4366	0,4328	0,4291	0,4254
2		0,3185	0,3156	0,3126	0,3098	0,3069	0,3043	0,3018	0,2992	0,2968	0,2944
3		0,2578	0,2571	0,2563	0,2554	0,2543	0,2533	0,2522	0,251	0,2499	0,2487
4		0,2119	0,2131	0,2139	0,2145	0,2148	0,2151	0,2152	0,2151	0,215	0,2148
5		0,1736	0,1764	0,1787	0,1807	0,1822	0,1836	0,1848	0,1857	0,1064	0,187
6		0,1399	0,1443	0,148	0,1512	0,1539	0,1563	0,1584	0,1601	0,1616	0,163
7		0,1092	0,115	0,1201	0,1245	0,1283	0,1316	0,1346	0,1372	0,1395	0,1415
8		0,0804	0,0878	0,0941	0,0997	0,1046	0,1089	0,1128	0,1162	0,1192	0,1219
9		0,053	0,0618	0,0696	0,0764	0,0823	0,0876	0,0923	0,0965	0,1002	0,1036
10		0,0263	0,0368	0,0459	0,0539	0,061	0,0672	0,0728	0,0778	0,0822	0,0862
11		0	0,0122	0,0228	0,0321	0,0403	0,0476	0,054	0,0598	0,065	0,0697
12				0	0,0107	0,02	0,0284	0,0358	0,0424	0,0483	0,0537
13						0	0,0094	0,0178	0,0253	0,032	0,0381
14								0	0,0084	0,0159	0,0227
15										0	0,0076

N	5%	1%	N	5%	1%
3	0,767	0,753	25	0,918	0,888
4	0,748	0,687	26	0,92	0,891
5	0,762	0,686	27	0,923	0,894
6	0,788	0,713	28	0,924	0,896
7	0,803	0,73	29	0,926	0,898
8	0,818	0,749	30	0,927	0,9
9	0,829	0,764	31	0,929	0,902
10	0,842	0,781	32	0,93	0,904
11	0,85	0,792	33	0,931	0,906
12	0,859	0,805	34	0,933	0,908
13	0,856	0,814	35	0,934	0,91
14	0,874	0,825	36	0,935	0,912
15	0,881	0,835	37	0,936	0,914
16	0,837	0,844	38	0,938	0,916
17	0,892	0,851	39	0,939	0,917
18	0,897	0,858	40	0,94	0,919
19	0,901	0,863	41	0,941	0,92
20	0,905	0,868	42	0,942	0,922
21	0,908	0,873	43	0,943	0,923
22	0,911	0,878	44	0,944	0,924
23	0,914	0,881	45	0,945	0,926
24	0,916	0,884	46	0,945	0,927
47	0,946	0,928	48	0,947	0,929
49	0,947	0,929	50	0,947	0,93

Tester la normalité de la distribution suivante qui donne des résultats d'essais de fatigue d'un matériau (nombre de cycles avant rupture) :

225	31	400	62	850	39	89	580	115	442	270	125	342	251	140
-----	----	-----	----	-----	----	----	-----	-----	-----	-----	-----	-----	-----	-----

Quel autre ajustement pourrait-on proposer ?

IV Tests de rang

Il s'agit de tests non paramétriques de comparaison. De manière générale, on préfère effectuer des tests paramétriques, en effet, les tests non paramétriques sont moins sensibles ; c'est à dire que, pour un test non paramétrique, la probabilité d'accepter H_0 alors que H_0 est fausse est plus importante, par contre lorsque l'on rejette H_0 , on peut être raisonnablement confiant quand à cette conclusion).

Dans les tests du rang, les valeurs observées sont remplacées par leurs rangs au sein des échantillons. L'idée du test est la suivante : on ordonne toutes les valeurs observées (i.e. les valeurs de tous les échantillons concernés), si le facteur étudié a une influence, les valeurs d'un des échantillons seront "dans les premiers" parmi les valeurs ordonnées.

1 Statistiques de l'ordre, de rang

Si (X_1, \dots, X_n) est un échantillon aléatoire indépendant i.d., on lui associe le vecteur aléatoire $X_o = (X_{(1)}, \dots, X_{(n)})$: échantillon ordonné.

$$X_{(1)} = \min_{i=1, \dots, n} X_i \leq X_{(2)} \leq \dots < X_{(n)} = \max_{i=1, \dots, n} X_i.$$

La loi de $X_{(1)}$ a pour fonction de répartition :

$$F_1(t) = 1 - [1 - F(t)]^n \text{ où } F \text{ est la fonction de répartition de } X.$$

La loi de $X_{(n)}$ a pour fonction de répartition :

$$F_n(t) = [F(t)]^n.$$

Plus généralement, on obtient :

$$\mathbb{P}(X_{(k)} < t) = \sum_{i=k}^n C_n^i [F(t)]^i [1 - F(t)]^{n-i}.$$

Définition 9.1 *Le rang de X_i dans la liste X_1, \dots, X_n est :*

$$R_i = 1 + \sum_{j \neq i} \mathbb{1}_{X_j < X_i}.$$

C'est le rang occupé par X_i dans la suite ordonnée $X_{(1)} < \dots < X_{(n)}$.

2 Le test de Wilcoxon

Il s'agit de comparer deux échantillons (X_1, \dots, X_n) et (Y_1, \dots, Y_m) , indépendants. Sont-ils issus de la même loi ? Soit $N = n + m$ et

$$(Z_1, \dots, Z_N) = (X_1, \dots, X_n, Y_1, \dots, Y_m)$$

l'échantillon concaténé. On considère les statistiques d'ordre et du rang attachées à cet échantillon :

$$Z_{(1)} < \dots < Z_{(N)}, \quad R_Z(i) = 1 + \sum_{j \neq i} \mathbb{1}_{Z_j < Z_i}.$$

Si X et Y ont même loi, alors la variable aléatoire R_Z , à valeur dans l'ensemble des permutations de $\{1, \dots, N\}$ est uniforme ($\mathbb{P}(R_Z = \sigma) = \frac{1}{N!}$), cette loi est indépendante de la loi commune de X et Y .

On note W_X la somme des rangs des X_i : $W_X = \sum_{i=1}^n R_Z(i)$. On montre que

$$\mathbb{E}(W_X) = \frac{n(N+1)}{2} \text{ et } \text{Var}(W_X) = \frac{nm(N+1)}{12}.$$

La loi de $W_X - \frac{n(n+1)}{2}$ est tabulée et permet de construire un test de comparaison de deux échantillons.

Exemple : on veut comparer les performances de deux groupes d'élèves à des tests d'habileté manuelle. Les performances en minutes sont les suivantes :

Groupe I	22	31	14	19	24	28	27	29		
Groupe II	25	13	20	11	23	16	21	18	17	26

On se demande s'il y a une différence significative entre les deux groupes (avec R : `wilcox.test`).

Chapitre 10

Exemples d'estimation non paramétrique

I Estimation d'une densité de probabilité

Lorsque l'on fait les tests de Kolmogorov-Smirnov, on utilise la convergence de la fonction de répartition empirique vers la fonction de répartition. On souhaite maintenant obtenir une estimation de la densité de probabilité f , dans le cas où la variable aléatoire X est à densité. On considère X_1, \dots, X_n un échantillon aléatoire indépendant de même loi que X .

1 Histogramme empirique

Une première approximation de la densité est fournie par l'histogramme. Pour cela, on choisit des classes : $[x_0, x_1], [x_1, x_2], \dots, [x_{k-1}, x_k]$, l'histogramme est constitué pour chaque classe d'un rectangle de hauteur $\hat{f}_i = \frac{N_i}{n(x_i - x_{i-1})}$, où

$$N_i = \sum_{j=1}^n \mathbb{I}_{]x_{i-1}, x_i]}(X_j).$$

Il s'agit d'une approximation de l'histogramme théorique (\hat{f}_i converge vers $\frac{\mathbb{P}(x_{i-1} < X \leq x_i)}{x_i - x_{i-1}}$). Si x_i et x_{i-1} convergent vers x alors ce rapport converge vers $f(x)$. Considérons des classes toutes de même taille h . Alors on considère $\hat{f}_n(x) = \frac{N_i}{nh}$ si $x_{i-1} < x \leq x_i$. Le problème est de choisir les x_i .

2 Fenêtres mobiles

Une réponse à ce problème du choix des x_i est donnée par les fenêtres mobiles : pour $x \in \mathbb{R}$, $I_x = [x - \frac{h}{2}, x + \frac{h}{2}]$, soit

$$N_x = \sum_{j=1}^n \mathbb{I}_{\{X_j \in I_x\}},$$

et

$$\widehat{f}_n(x) = \frac{1}{nh} N_x = \frac{1}{nh} \sum_{j=1}^n \mathbb{1}_I \left(\frac{x - X_j}{h} \right)$$

avec $I = [-\frac{1}{2}, \frac{1}{2}]$. On peut montrer que si $h \rightarrow 0$ et $nh \rightarrow \infty$ alors $\widehat{f}_n(x)$ converge vers $f(x)$. On a aussi un théorème de la limite centrale fonctionnel.

3 Versions lisses

L'approximation ci-dessus est assez irrégulière (à cause de la fonction $\mathbb{1}_I$). Pour obtenir un estimateur plus régulier, on peut remplacer $\mathbb{1}_I$ par une fonction régulière K appelée noyau. Par exemple :

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \text{ (noyau gaussien),}$$

$$K(x) = \frac{3}{4 \times \sqrt{5}} \left(1 - \frac{u^2}{5}\right) \text{ si } |u| < \sqrt{5} \text{ (noyau d'Epanechnikov).}$$

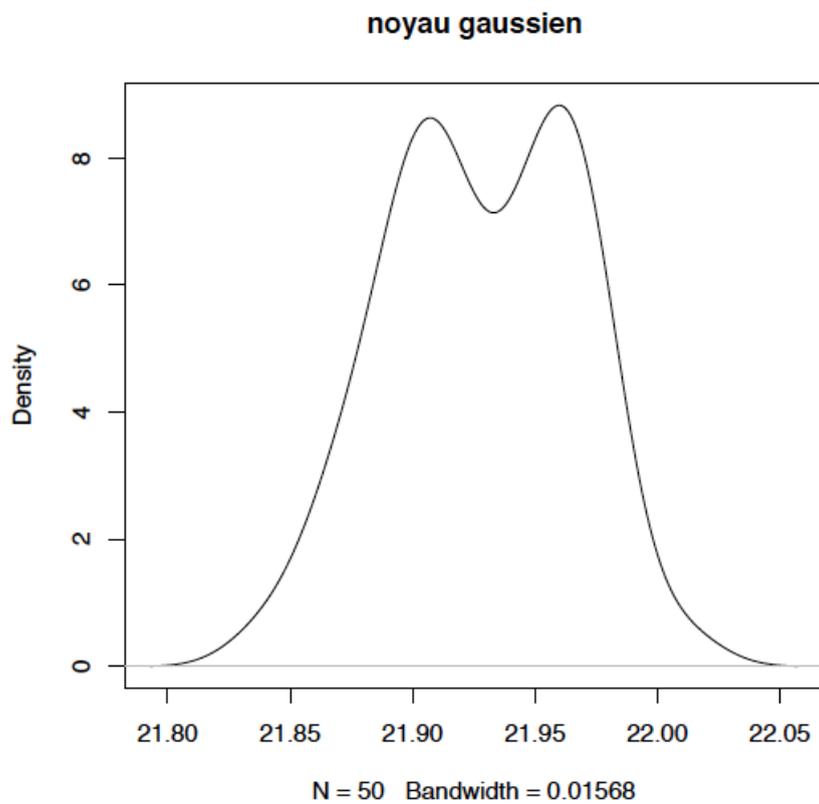
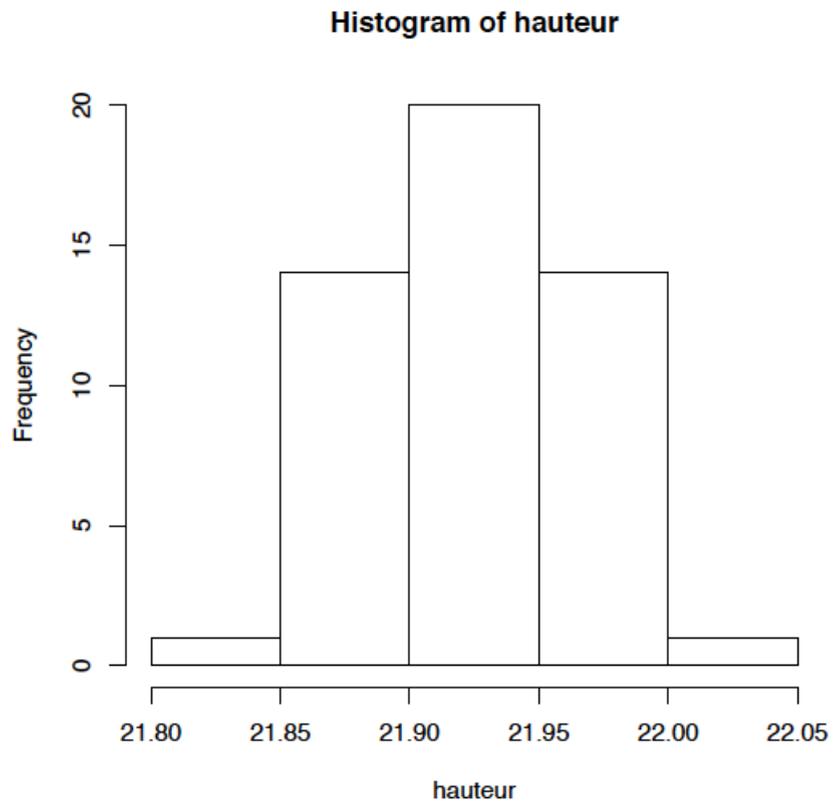
$$\widehat{f}_n(x) = \frac{1}{nh} \sum_{j=1}^n K\left(\frac{x - X_j}{h}\right).$$

4 Un exemple

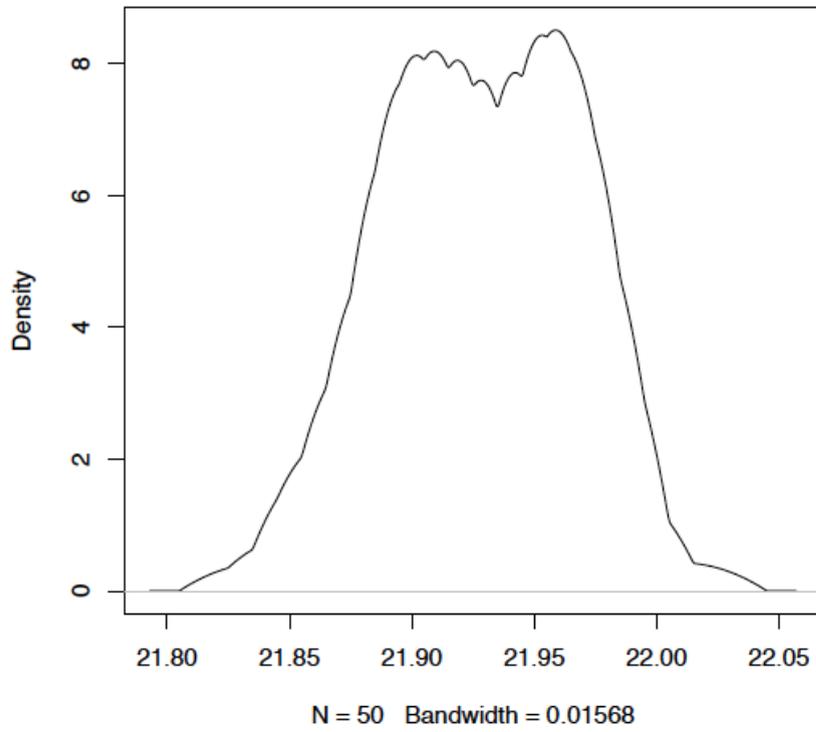
Le tableau ci-dessous donne la hauteur de 50 pièces usinées.

21.86	21.9	21.98
21.84	21.89	21.96
21.88	21.92	21.98
21.9	21.91	21.95
21.92	21.91	21.97
21.87	21.92	21.94
21.9	21.91	22.01
21.87	21.93	21.96
21.9	21.96	21.95
21.93	21.91	21.95
21.92	21.97	21.97
21.9	21.97	21.96
21.91	21.97	21.95
21.89	21.97	21.94
21.91	21.98	21.97
21.87	21.95	21.95
21.89	21.89	

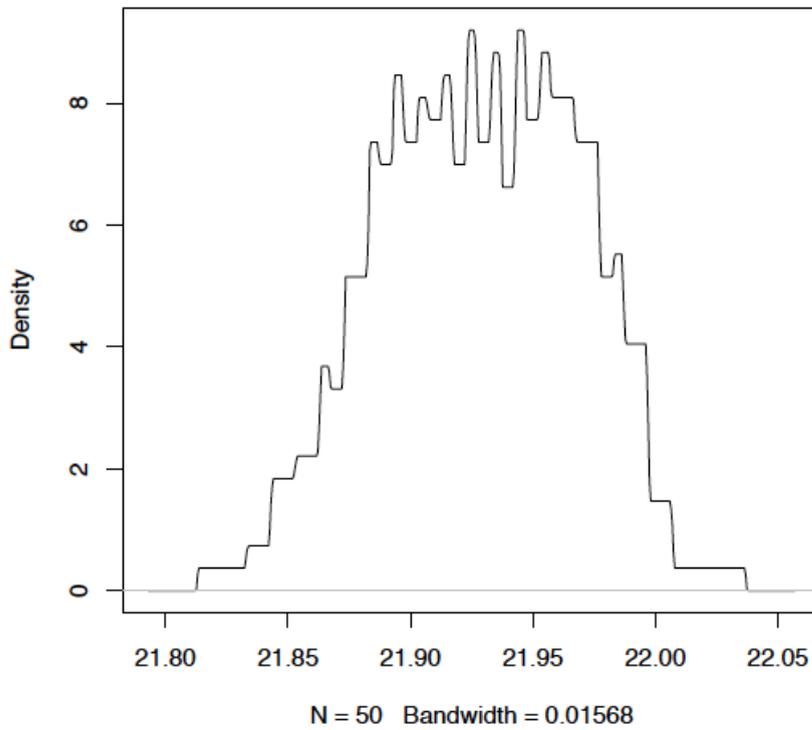
Sous R, l'estimation de la densité peut se faire avec la commande `density`.



noyau epanechnikov



noyau indicatrice



II Estimation des quantiles

L'estimation des quantiles revêt un intérêt particulier. Par exemple, la VaR (Value at Risk) utilisée comme indicateur de risque dans de nombreux domaines, n'est rien d'autre qu'un quantile.

La fonction quantile d'une distribution de probabilités est l'inverse généralisé de la fonction de distribution F :

$$F^{-1}(p) = \inf\{x \in \mathbb{R}, F(x) \geq p\}.$$

1 Quantiles empiriques

On définit alors la fonction quantile empirique F_n^{-1} comme l'inverse généralisé de la fonction de répartition empirique F_n .

On admettra que $F_n^{-1}(p)$ converge vers $F^{-1}(p)$ en tout point de continuité de F^{-1} si et seulement si $F_n(t)$ converge vers $F(t)$ en tout point de continuité de F .

2 Lien avec les statistiques d'ordre

Étant donné un échantillon aléatoire iid (X_1, \dots, X_n) , on note $(X_{(1)}, \dots, X_{(n)})$ la statistique d'ordre associée.

On a la relation suivante :

$$\forall p \in \left] \frac{i-1}{n}, \frac{i}{n} \right], F_n^{-1}(p) = X_{(i)}.$$

3 Résultats asymptotiques

On a le résultat asymptotique suivant dans le cas où la fonction de répartition F est différentiable : pour tout $p \in]0, 1[$,

$$\sqrt{n}(F_n^{-1}(p) - F^{-1}(p)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2(p))$$

avec

$$\sigma^2(p) = \frac{p(1-p)}{f(F^{-1}(p))^2}.$$

Ce résultat pose ainsi la question de l'estimation de la densité.

L'utilisation de la *transformation par quantiles* peut permettre de trouver des intervalles de confiance pour les quantiles, sans passer par l'estimation de la densité,

dans le cas d'une fonction de répartition strictement croissante et continue. On note $U_1 = F(X_1), \dots, U_n = F(X_n)$. Les U_i sont i.i.d. de loi uniforme sur $[0, 1]$. On note $U_{(1)}, \dots, U_{(n)}$ les statistiques d'ordre associées. On a alors

$$\mathbb{P}(X_{(k)} < F^{-1}(p) \leq X_{(\ell)}) = \mathbb{P}(U_{(k)} < p \leq U_{(\ell)}).$$

On admet que pour

$$\frac{k}{n} = p - z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \quad \frac{\ell}{n} = p + z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$$

on a :

$$\mathbb{P}(U_{(k)} < p \leq U_{(\ell)}) \longrightarrow 1 - \alpha.$$

On peut alors choisir $X_{(k)}$ et $X_{(\ell)}$ comme bornes d'un intervalle de confiance de niveau asymptotique $1 - \alpha$ pour $F^{-1}(p)$.