



الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire

وزارة التعليم العالي والبحث العلمي
Ministère de L'Enseignement Supérieur et de la Recherche Scientifique

جامعة سعد دحلب البلدية-1-
Université Saad Dahlab Blida -1-



Mémoire de fin D'études

En vue de l'obtention du diplôme Master

Faculté de sciences
Département : Mathématique

Spécialité : Modélisation stochastique et statistique

THEME

**Méthode de vraisemblance empirique,
application sur la régression linéaire**

Présenté par :
CHERGUI Sakina
CHENAH Nadjat

Soutenu le : 23/ 09 /2020

Devant le Jury :

Président : O.TAMI MAA Université de Blida 1

Promoteur: A. RASSOUL MCA ENSH de Blida

Examineur : R.FRIHI MAA Université de Blida 1

Promotion : 2019/2020

Remerciements

*Nous tenons tout d'abord à remercier Dieu le tout puissant et
miséricordieux,
qui nous a donné la force et la patience d'accomplir ce modeste travail.*

*En second lieu, notre encadreur **Mr Abdelaziz RASSOUL** pour ses
précieux conseil set son orientation ficelée durant toute la période du
travail.*

*Nos vifs remerciements vont également aux membres du jury, **Mr
O.TAMI**
et **Mr R.FRIHI** pour l'intérêt qu'ils ont porté à notre recherche en
acceptant d'examiner notre travail et de l'enrichir par leurs
propositions.*

*On n'oublie pas nos familles et nos amis qui par leurs prières et leurs
encouragements, on a pu surmonter tous les obstacles.*

*Enfin, nous tenons également à remercier toutes les personnes qui ont
participé
de près ou de loin à la réalisation de ce travail.*

المخلص :

تستخدم طرق الاحتمالية التجريبية على نطاق واسع من السياقات البيانية المختلفة لتقدير فترات الثقة وفقا للقيود و معلومات سابقة على المعلمات، في حالات العينات الصغيرة و متعددة الأبعاد فترات الثقة تكون غير دقيقة لهذا نقترح طريقة جديدة متوسط الاحتمال التجريبي يعتمد على بناء مجموعة جديدة من البيانات الافتراضية باستخدام قيم الملاحظات لتحديد نسبة الاحتمالية التجريبية وفق نظرية ولكس.

محاكاة بأمتلة ذات عينات محدودة و تقيمها، مما يتضح أن فترات الثقة التي تم إنشائها من خلال متوسط الاحتمال التجريبي أكثر دقة بكثير من طرق الاحتمالات التجريبية الأخرى.

Résumé :

Les méthodes empiriques de vraisemblance sont largement utilisées dans différents contextes pour construire les régions de confiance des paramétrées qui satisfont aux contraintes de moment. Cependant, les régions de confiance du rapport de vraisemblance empirique peuvent être peu précises, en particulier pour les échantillons de petite taille et les situations multidimensionnelles. Dans ce travail nous proposons une nouvelle méthode de probabilité empirique moyenne (MEL). Cette nouvelle méthode construit un nouvel ensemble de pseudo-données en utilisant les valeurs des moyennes d'observation pour définir le rapport de vraisemblance empirique et nous prouvons que ce rapport MEL satisfait au théorème de WILKS. Des simulations avec différents exemples sont données pour évaluer sa performance en matière des échantillons finis, ce qui montre que les régions de confiance construites par la probabilité empirique moyenne sont beaucoup plus précises que celles des autres méthodes de probabilité empirique.

ABSTRACT :

Empirical likelihood methods are widely used in diferen contexts to construct condolence regions of parameters that satisfy timing constraints. However, the condolence regions of the empirical likelihood ratio can be imprecise, especially for small sample sizes and multidimensional situations. In this work we propose a new method of mean empirical probability (MEL). This new method constructs a new set of pseudo-data using the values of the observation means to define the empirical likelihood ratio and we prove that this MEL ratio satifietes WILKS'theorem. Simulations with diferents examples are given to assess its finite sample performance, showing that the confidence regions constructed by the mean empirical probability are much more accurate than those of other empirical probability method

Table des matières

1	Introduction générale	5
2	Généralités sur la statistiques inférentielle	7
2.1	Généralités	7
2.2	Modèle d'échantillonnage	7
2.3	Vraisemblance	9
2.4	Famille exponentielle	9
2.5	Estimation ponctuelle	10
2.5.1	Définition d'une statistique	10
2.5.2	Estimation	10
2.5.3	Qualités d'un estimateur	11
2.5.4	Information de Fisher	12
2.5.5	Borne de Fréchet-Damois-Cramer-Rao (FDCR)	13
2.6	Estimateur efficace	14
2.7	Méthodes d'estimation	15
2.7.1	La méthode des moments	15
2.7.2	Méthode du maximum de vraisemblance	15
2.7.3	Estimation par intervalle de confiance	18
2.8	Tests statistiques	22
2.8.1	Introduction	22
2.9	Notions générales	23
2.9.1	Types d'erreurs	24
2.9.2	Etapes de construction d'un test statistique	24
2.9.3	Tests uniformément plus puissants	25
2.9.4	Test paramétrique usuelles	26
2.9.5	Test et statistique exhaustive	27
2.9.6	Test entre deux hypothèses multiples (test de Lehman)	27
2.9.7	Famille à rapport de vraisemblance monotone	27
2.9.8	Tests bilatérales	28
2.9.9	Tests unilatérales	29

2.9.10	Tests fondés sur le rapport du maximum de vraisemblance	30
2.9.11	Test de comparaison d'échantillons	30
2.9.12	Test de Student et F.S pour des échantillons indépendants	30
3	Vraisemblance empirique	35
3.1	Introduction	35
3.2	Vraisemblance empirique et bootstrap	35
3.3	Pour quoi les méthodes conventionnelles ne s'appliquent-elles pas?	36
3.3.1	Estimation de γ et κ	37
3.4	Introduction à la vraisemblance empirique	38
3.5	Inférence de vraisemblance empirique pour les moyennes	39
3.5.1	Ratios de vraisemblance empirique (ELR)	39
3.6	Intervalles de confiance pour μ	45
3.7	Vraisemblance empirique pour des vecteurs aléatoires	46
3.7.1	EL pour les moyennes multivariées	47
3.7.2	EL pour des fonctions régulières des moyennes	50
3.8	Estimation des équations	53
3.8.1	Généralisation	54
3.8.2	El pour estimer les équations	55
4	Applications sur la régression linéaire	58
4.1	Introduction	58
4.2	Régression paramétrique	59
4.2.1	Estimation par Moindres carrées (LS)	60
4.2.2	Estimation par vraisemblance empirique (EL)	60
4.3	Régression non paramétrique	66
4.3.1	Estimateur de Nadaraya-Watson	66
4.3.2	Estimateur de Buckley-James	67
4.3.3	Estimation par vraisemblance empirique (EL)	69
4.4	Simulations et applications	72
4.4.1	Introduction	72
4.4.2	Présentation du package emplik	72
4.4.3	Simulation des modèles linéaires	73
4.4.4	conclusion	78
5	Conclusion générale	79

Liste des tableaux

4.1	Résultats du testes sur un modèle de régression par la méthode de Backley James	74
4.2	Les résultats du test d'un modèle de régression par l'estimateur de régression pondéré par rapport la vraisemblance empirique.	74
4.3	Estimation du paramètres du modèle de régression par la méthode de Backley James.	75
4.4	l'intervalle de confiance du modèle de régression AFT	75
4.5	les estimations médianes de régression	77

Table des figures

3.1	Profile de vraisemblance empirique	41
3.2	Le graphique de gauche montre la valeur de l'indice S&P 500 pendant 256 jours de négociation. celles-ci les valeurs ont été converties en 255 retours logarithmiques. Le graphique de droite montre un graphique QQ de ces retours sur l'indice S&P 500	52
4.1	QQplot de-2ELR pour n=100	73

Chapitre 1

Introduction générale

Le travail présenté dans ce mémoire entre dans le cadre de l'estimation des paramètres inconnus dans des modèles statistiques, ce qui constitue un thème important pour l'inférence statistique. Nous présentons particulièrement une étude basée sur la vraisemblance empirique et aussi sur la méthode du maximum de vraisemblance paramétrique.

Depuis les travaux fondateurs d'**OWEN** ([Owen,(1988)], [Owen, (1990)]), ils ont été de nombreuses avancées dans la méthode de probabilité empirique qui ont permis d'appliquer cette méthode pratiquement tous les domaines de la recherche statistique. Cette méthode de type non-paramétrique consiste à maximiser la vraisemblance d'une loi ne chargeant que les données, sous des contraintes satisfaites par le modèle (des contraintes de marges en sondage). [Owen,(1988)] [Owen, (1990)] et de nombreux auteurs (voir [Owen, (2001)] pour de nombreuses références) ont montré que l'on pouvait en effet obtenir dans ce cadre une version non-paramétrique du théorème de **WILKS**, à savoir la convergence du rapport de vraisemblance correctement normalisée vers une loi du χ^2 , permettant ainsi de réaliser des tests ou de construire des régions de confiance non-paramétriques pour certains paramètres du modèle. [voir [Hall et Scala, (1990)], [Qin et Lawless, (1994)], [Corcoran.SA and Spady(1995)], [Owen, (2001)], ils ont observés que les régions de confiance du rapport de vraisemblance empirique peuvent avoir une précision médiocre, en particulier dans les situations à petit échantillon et multidimensionnelles. En particulier, il existe un problème persistant de sous-couverture dans la mesure où les probabilités de couverture des régions de confiance du rapport de vraisemblance empirique ont tendance à être inférieures aux niveaux nominaux. Dans ce travail, nous abordons un problème fondamental qui sous-tend la faible précision et la sous-couverture, à savoir que la probabilité empirique n'est définie que sur une partie de l'espace des paramètres. C'est ce que nous appelons le problème de décalage entre le do-

maine de la probabilité empirique et l'espace des paramètres. Nous résolvons ce problème par une approche géométrique qui étend le domaine à tout l'espace des paramètres. Notre solution apporte des améliorations substantielles dans la précision de l'inférence de la probabilité empirique et est particulièrement utile pour les petits échantillons et les situations multidimensionnelles.

Le chapitre introductif présenterons quelques rappel et définitions de base de la statistique inférentielle comme les méthodes d'estimation, nous concentrons particulièrement sur la méthode du maximum de vraisemblance, c'est une méthode qui donne assez souvent de bonnes estimations. Elle a été introduite dans le cadre des variables aléatoires indépendantes et identiquement distribuées (i.i.d), on rappelle comment on obtient les estimateurs du maximum de vraisemblance (dit ML) et comment on les calcule en pratique, et en déduit les statistiques classiques des tests asymptotiques. **Dans le chapitre 2**, on utilise les études et les résultats précédents pour présenter la vraisemblance empirique (EL) parce qu'il est considéré une méthode d'inférence non paramétrique basée sur une fonction de rapport de vraisemblance, nous proposerons un échantillonnage biaisé et des données censurées pour traiter les hypothèses et les intervalles d'estimation, Notre travail s'applique aux distributions discrètes, distributions continues. Nous démontrons convergence de la distribution de l'estimateur dans certains cas, et nous illustrons nos résultats par des expériences numériques, et nous parlons des méthodes empiriques de vraisemblance (MEL).

Un autre but de ce mémoire qui se trouve dans le **chapitre 3**, nous présentons la méthode de vraisemblance empirique pour le type de régression problèmes d'inférence. Les modèles de régression examinés dans cette étude comprennent les modèles paramétriques et non paramétriques aussi on parlons sur des packages en R comme le package `emplik`, pour mettre en œuvre la vraisemblance empirique pour l'inférence sur une moyenne multi variée. Ce package contient des fonctions qui utilisent différents algorithmes d'optimisation. Ces fonctions sont basées sur la théorie de la convexité. Enfin, nous présentons quelque exemple d'application avec des données réelles.

Chapitre 2

Généralités sur la statistique inférentielle

2.1 Généralités

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

$$X_i : (\Omega, \mathcal{B}) \longrightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$$

$$X : (\Omega, \mathcal{B}, \mathbb{P}) \longrightarrow (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}}, \mathbb{Q}_X)$$

\mathbb{Q}_X est la loi de probabilité image du vecteur aléatoire X .

Généralement, on ne connaît pas la loi \mathbb{Q}_X . On souhaite ici étudier le plus précisément possible cette loi de probabilité.

Hypothèse fondamentale.

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

2.2 Modèle d'échantillonnage

Définition 2.1 *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 \mathbb{Q}_X à appartenir.*

Remarque 2.1 *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

$$X : (x_1, \dots, x_n) = (X_1(\omega), \dots, X_n(\omega)), \omega \in \Omega.$$

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 suivantes :

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

$(\mathbb{R}; \mathcal{B}_{\mathbb{R}}; P)^n$ si la loi P appartient à la famille de mesures de probabilité 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 P de lois de probabilité est indicée par un paramètre θ .

On note alors

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

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 $\Theta \subset \mathbb{R}^k \times U$ ou U est non paramétrique.

Exemple 2.1 *Si on considère un modèle où P peut être n'importe quelle loi de probabilité absolument continue par rapport à la mesure de Lebesgue, P est dans un cadre non paramétrique. On a*

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

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,

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

2.3 Vraisemblance

Définition 2.2 Dans un modèle statistique paramétrique $(\mathbb{R}_n, B_{\mathbb{R}_n}, \mathbb{Q})$, on appelle vraisemblance de l'observation x la fonction

$$\begin{aligned} \mathcal{L}(x; \cdot) & : \Theta \longrightarrow \mathbb{R}^+ \\ \theta & \longrightarrow \mathcal{L}(x; \theta) = f_\theta(x). \end{aligned}$$

Bien sur, dans le cas d'un modèle d'échantillonnage, la vraisemblance de l'échantillon observé $x = (x_1, \dots, x_n)$ s'écrit sous la forme

$$\mathcal{L}(x_1, \dots, x_n, \theta) = \prod_{i=1}^n f_\theta(x_i). \quad (2.1)$$

C'est la loi conjointe du n -échantillon évalué aux valeurs observées et considérée comme fonction du paramètre θ .

2.4 Famille exponentielle

Définition 2.3 Soit X une variable aléatoire réelle, dont la loi de probabilité dépend d'un paramètre $\theta \in \mathbb{R}^d$. On dit que la loi de X **appartient à la famille exponentielle**, si seulement si $P(X = x; \theta)$ (cas discret) ou $f_X(x; \theta)$ (cas continu) est de la forme :

$$\frac{\sum_{j=1}^d a_j(x) \alpha_j(\theta) + b(x) + \beta(\theta)}{\exp} \quad (2.2)$$

Remarque 2.2 la plupart des lois usuelles appartiennent à la famille exponentielle : loi exponentielle $\exp(\lambda)$:

$$f_X(x; \lambda) = \lambda \exp^{-\lambda x} = \exp^{-\lambda x + \ln \lambda}.$$

Qui est de la forme souhaitée avec

$$d = 1, a(x) = x, \alpha(\lambda) = -\lambda, b(x) = 0 \text{ et } \beta(\lambda) = \ln \lambda.$$

2.5 Estimation ponctuelle

2.5.1 Définition d'une statistique

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

Définition 2.4 *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 permettant de "préciser" le paramètre $\theta \in \mathbb{R}^p$. Le plus souvent, on aura $k = p$.*

Définition 2.5 *On utilisera la notation T_n pour une statistique. Elle ne dépend pas de la famille de lois de probabilité \mathbb{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 pas de θ .*

Exemple 2.2 *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 "la moyenne empirique" :*

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

2.5.2 Estimation

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

Définition d'un estimateur

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

Définition 2.6 – *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 $T_n(x_1, \dots, x_n)$.*

2.5.3 Qualités d'un estimateur

Notion de biais

Définition 2.7 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 $g(\theta)$.

Convergence d'un estimateur

Définition 2.8 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 $V(T_n) \rightarrow 0$ 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 $V(T_n) \rightarrow 0$ quand $n \rightarrow 0$, alors T_n est un estimateur convergent pour $g(\theta)$.

Perte quadratique

C'est l'écart au carré entre le paramètre et son estimateur :

$$l(\theta, T) = (\theta - T)^2.$$

Risque d'un estimateur

C'est la moyenne des pertes :

$$R(T, \theta) = \mathbb{E}[\mathcal{L}(\theta, T)].$$

$R(T, \theta) = \mathbb{E}[(\theta - T)^2]$ est le **risque quadratique moyen**. Un estimateur T est dit convergent si

$$\lim_{n \rightarrow \infty} R(T, \theta) = 0.$$

2.5.4 Information de Fisher

Définition 2.9 On appelle quantité d'information de Fisher $I_n(\theta)$ apportée par un échantillon sur le paramètre θ , la quantité suivante positive ou nulle (si elle existe) :

$$\mathcal{I}_n(\theta) = \mathbb{E} \left[\left(\frac{\partial \ln \mathcal{L}(X, \theta)}{\partial \theta} \right)^2 \right]. \quad (2.3)$$

Avec

$$\mathcal{L}(X, \theta) = \prod_{i=1}^n f(X_i, \theta).$$

Théorème 2.1 Si le domaine de définition de X ne dépend pas de θ alors :

$$\mathcal{I}_n(\theta) = -\mathbb{E} \left[\left(\frac{\partial \ln \mathcal{L}(X, \theta)}{\partial \theta^2} \right) \right].$$

Preuve: On a $\mathcal{L}(x, \theta)$ est une densité de probabilité :

$$\int_{\mathbb{R}^n} \mathcal{L}(x, \theta) dx = 1,$$

On

$$\frac{\partial \ln \mathcal{L}(x, \theta)}{\partial \theta} = \frac{\partial \mathcal{L}(x, \theta)}{\partial \theta \mathcal{L}(x, \theta)},$$

Donc

$$\frac{\partial \mathcal{L}(x, \theta)}{\partial \theta} = \mathcal{L}(x, \theta) \cdot \frac{\partial \ln \mathcal{L}(x, \theta)}{\partial \theta}$$

$$\int_{\mathbb{R}^n} \frac{\partial \ln \mathcal{L}(x, \theta)}{\partial \theta} dx = 0 \implies \int_{\mathbb{R}^n} \mathcal{L}(x, \theta) \cdot \frac{\partial \ln \mathcal{L}(x, \theta)}{\partial \theta} dx = 0.$$

Ce qui prouve que la variable aléatoire $\frac{\partial \ln \mathcal{L}(x, \theta)}{\partial \theta}$ est centrée i.e :

$$\mathbb{E} \left[\frac{\partial \ln \mathcal{L}(X, \theta)}{\partial \theta} \right] = 0.$$

Donc

$$\mathcal{I}_n(\theta) = \mathbb{E} \left[\left(\frac{\partial \ln \mathcal{L}(X, \theta)}{\partial \theta} \right)^2 \right] = \text{Var} \left(\frac{\partial \ln \mathcal{L}(X, \theta)}{\partial \theta} \right).$$

En dérivant la deuxième fois :

$$\int_{\mathbb{R}^n} \mathcal{L}(x, \theta) \cdot \frac{\partial^2 \ln \mathcal{L}(x, \theta)}{\partial \theta^2} dx + \int_{\mathbb{R}^n} \frac{\partial \ln \mathcal{L}(x, \theta)}{\partial \theta} \frac{\partial \mathcal{L}(x, \theta)}{\partial \theta} dx = 0.$$

$$\int_{\mathbb{R}^n} \mathcal{L}(x, \theta) \cdot \frac{\partial^2 \ln \mathcal{L}(x, \theta)}{\partial \theta^2} dx + \int_{\mathbb{R}^n} \left(\frac{\partial \ln \mathcal{L}(x, \theta)}{\partial \theta} \right)^2 \mathcal{L}(x, \theta) dx = 0.$$

$$I_n(\theta) = - \int_{\mathbb{R}^n} \frac{\partial^2 \ln \mathcal{L}(x, \theta)}{\partial \theta^2} \mathcal{L}(x, \theta) d\theta.$$

$$I_n(\theta) = -\mathbb{E} \left[\frac{\partial^2 \ln \mathcal{L}(x, \theta)}{\partial \theta^2} \right].$$

■

Conséquence : Si le domaine de définition ne dépend pas de θ alors :

$$\mathcal{I}_n(\theta) = n\mathcal{I}_1(\theta). \quad (2.4)$$

2.5.5 Borne de Fréchet-Damois-Cramer-Rao (FDCCR)

Théorème 2.2 (Inégalité de FDCCR) *Si le domaine de définition de X ne dépend pas de θ , alors pour tout estimateur sans biais :*

$$\text{Var}(T(X)) \geq \frac{1}{\mathcal{I}_n(\theta)}$$

Si T est un estimateur sans biais de $g(\theta)$:

$$\text{Var}(T(X)) \geq \frac{(g'(\theta))^2}{\mathcal{I}_n(\theta)}.$$

La borne $(g'(\theta))^2 / \mathcal{I}(\theta)$ est appelée borne de **Cramer-Rao** (ou **Fréchet-Darmois-Cramer-Rao**)

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

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

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

Ce qui termine la preuve.

2.6 Estimateur efficace

Définition 2.10 *Un estimateur T est dit efficace si variance est égale à la borne de **FDCR**.*

Propriété : Un estimateur sans biais efficace est convergent. En effet :
 T efficace $\implies Var(T) = \frac{[g'(\theta)]^2}{I_n(\theta)}$, or $I_n(\theta) = nI_1(\theta)$ et T est sans biais

$$\lim_{n \rightarrow \infty} R(T, \theta) = \lim_{n \rightarrow \infty} Var(T) = \lim_{n \rightarrow \infty} \frac{[g'(\theta)]^2}{I_n(\theta)} = \lim_{n \rightarrow \infty} \frac{[h'(\theta)]^2}{nI_1(\theta)} = 0.$$

Donc T convergent.

Exemple 2.3 Soit (X_1, X_2, \dots, X_n) un n -échantillon issu de $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$,

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

Estimateur de μ . Calculer l'information de Fisher et la borne de **FDCR**. On a

$$\mathcal{I}_1(\mu) = -\mathbb{E} \left[\frac{\partial^2 \ln f(X, \mu)}{\partial \mu^2} \right],$$

Et

$$f(x, \mu) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\},$$

Alors

$$\ln f(x, \mu) = -\ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}(x-\mu)^2,$$

$$\frac{\partial \ln L}{\partial \mu} = \frac{-1}{\sigma^2} \implies \mathcal{I}_n(\mu) = \frac{n}{\sigma^2}.$$

$$\mathbb{E}(\bar{X}) = \mu \text{ et } Var(\bar{X}) = \frac{\sigma^2}{n}, \text{ et on a : } \mathcal{I}_n(\mu) = \frac{n}{\sigma^2}.$$

La borne de **FDCR** est

$$\frac{1}{\mathcal{I}_n(\mu)} = \frac{\sigma^2}{n} Var(\bar{X}).$$

Dans la prochaine section, nous ne nous intéressons qu'aux deux méthodes d'estimateur les plus usuelles, la méthode des moments et la méthode du maximum de vraisemblance.

2.7 Méthodes d'estimation

2.7.1 La méthode des moments

C'est la méthode la plus naturelle, que nous avons déjà utilisé sans la formaliser.

L'idée de base est d'estimer une espérance mathématique par une moyenne empirique, une variance par une variance empirique, etc...

Si le paramètre à estimer est l'espérance de la loi des X_i , alors on peut l'estimer par la moyenne empirique de l'échantillon. Autrement dit, si $\theta = E(X)$, alors l'estimateur de θ par **la méthode des moments (EMM)** est

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

Plus généralement, pour $\theta \in \mathbb{R}$, si $E(X) = \varphi(\theta)$, où φ est une fonction inversible, alors l'estimateur de θ par la méthode des moments est $\tilde{\theta}_n = \varphi^{-1}(\bar{X}_n)$.

De la même manière, on estime la variance de la lois des X_i par la variance empirique de l'échantillon

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

Ce principe peut naturellement se généraliser aux moments de tous ordres, centrés ou non centrés :

$$E \left[(X - E(X))^k \right] \text{ et } E(X^k), k \geq 1.$$

Exemple 2.4 Loi exponentielle : Si X_1, \dots, X_n sont indépendantes et de même loi exponentielle $\exp(\lambda)$, $E(X) = \frac{1}{\lambda}$. Donc l'estimateur de λ par la méthode des moments est : $\tilde{\lambda}_n = \frac{1}{\bar{X}}$.

2.7.2 Méthode du maximum de vraisemblance

La fonction de vraisemblance

Définition 2.11 Quand les observations sont toutes discrètes ou toutes continues, on appelle **fonction de vraisemblance** (ou plus simplement vraisemblance) pour l'échantillon x_1, \dots, x_n , la fonction du paramètre θ .

$$\mathcal{L}(\theta; x_1, \dots, x_n) = \begin{cases} P(X_1 = x_1, \dots, X_n = x_n; \theta) & \text{si les } X_i \text{ sont discrètes} \\ f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) & \text{si les } X_i \text{ sont continus} \end{cases}$$

Dans tous les exemples que nous traiterons ici, les X_i sont indépendantes et de même loi. Dans ce cas, la fonction de vraisemblance s'écrit :

$$\mathcal{L}(\theta; x_1, \dots, x_n) = \begin{cases} \prod_{i=1}^n P(X_i = x_i; \theta) = \prod_{i=1}^n P(X = x_i; \theta) & \text{si les } X_i \text{ sont discrètes} \\ \prod_{i=1}^n f_{X_i}(x_i; \theta) = \prod_{i=1}^n f(x_i; \theta) & \text{si les } X_i \text{ sont continues} \end{cases}$$

Remarque 2.3 La probabilité et la densité utilisée dans cette définition sont des fonctions des observations x_1, \dots, x_n , dépendant du paramètre θ . À l'inverse, la fonction de vraisemblance est considérée comme une fonction de θ dépendant des observations x_1, \dots, x_n , ce qui permet par exemple de dériver cette fonction par rapport à θ .

L'estimateur de maximum de vraisemblance (EMV)

En suivant le raisonnement précédent, pour n quelconque, il est logique de dire que la valeur la plus vraisemblable de θ est la valeur pour laquelle la probabilité d'observer x_1, \dots, x_n est la plus forte possible. Cela revient à faire comme si c'était l'éventualité la plus probable qui s'était produite au cours de l'expérience.

Définition 2.12 On appelle *estimation de maximum de vraisemblance (EMV) de paramètre θ* la statistique $\tilde{\theta}_n$ rendant maximale, selon θ , la fonction de vraisemblance du modèle $\mathcal{L}(X_1, \dots, X_n; \theta)$, soit :

$$\tilde{\theta}_n = \arg \max_{\theta \in \Theta} \ln \mathcal{L}(X; \theta)$$

Cette méthode consiste à résoudre :

$$\begin{cases} \frac{\partial \mathcal{L}(x; \theta)}{\partial \theta} = 0, & \text{pour trouver } \tilde{\theta} \\ \frac{\partial^2 \mathcal{L}(x; \theta)}{\partial \theta^2} \Big|_{\theta = \tilde{\theta}} < 0, & \text{pour assurer l'existence du } \max_{\theta \in \Theta} \mathcal{L}(x; \theta) \end{cases} ,$$

Quand $\theta = (\theta_1, \dots, \theta_n) \in \mathbb{R}^d$ et que toutes les dérivées partielles ci-dessous existes, $\tilde{\theta}_n$ est solution du système d'équations appelées **équation de vraisemblance** :

$$\forall j \in \{1, \dots, d\}, \frac{\partial}{\partial \theta_j} \ln \mathcal{L}(\theta; x_1, \dots, x_n) = 0.$$

Remarque 2.4 Maximum $\mathcal{L}(x, \theta)$ revient à maximiser $\ln \mathcal{L}(x, \theta)$. Il est plus commode de maximiser $\ln \mathcal{L}(x, \theta)$.

Exemple 2.5 Loi de Bernoulli : Si les X_i sont de loi $\mathcal{B}(p)$, on a :

$$P(X_i = x_i; p) = \begin{cases} p & \text{si } x_i = 1 \\ 1 - p & \text{si } x_i = 0 \end{cases} = p^{x_i} (1 - p)^{1 - x_i}$$

Donc la fonction de vraisemblance est :

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

D'où

$$\ln \mathcal{L}(p; x_1, \dots, x_n) = \left(\sum_{i=1}^n x_i \right) \ln p + \left(n - \sum_{i=1}^n x_i \right) \ln (1 - p).$$

Alors

$$\frac{\partial}{\partial p} \ln \mathcal{L}(P; x_1, \dots, x_n) = \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{1 - p} = \frac{\sum_{i=1}^n x_i - np}{p(1 - p)},$$

Qui s'annule pour $p = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}_n$. Par conséquent, l'EMV de p est $\tilde{p}_n = \bar{X}_n$.

Remarque 2.5 L'EMV ne peut s'obtenir que si la vraisemblance \mathcal{L} est explicite, donc si le modèle est bien explicité (loi normale, loi Poisson, ...). La méthode des moindres carrés et celle des moments ne nécessitent pas cette spécification.

Propriétés de l'estimation de maximum de vraisemblance

Un estimateur de maximum de vraisemblance n'est pas forcément unique (la vraisemblance peut avoir plusieurs maxima), ni sans biais, ni de variance minimale, ni efficace. Mais il possède d'excellentes propriétés asymptotiques, pour que la loi des observations vérifie les conditions de régularité déjà évoquées pour la quantité d'information.

Propriété Si les X_i sont indépendantes et de même loi dépendant d'un paramètre réel θ , cette loi vérifiant les conditions de régularité, on a :

- $\tilde{\theta}_n$ converge presque sûrement vers θ .
- $\sqrt{I_n(\theta)} (\tilde{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$, ce qui signifie que, quand n est grand, $\tilde{\theta}_n$ est approximativement de loi $\mathcal{N}\left(\theta, \frac{1}{I_n(\theta)}\right)$. On en déduit que $\tilde{\theta}_n$ est

asymptotiquement gaussien, sans biais et efficace. Cette propriété peut aussi s'écrire :

$$\sqrt{n} (\tilde{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{1}{I_n(\theta)} \right).$$

- Si $\tilde{\theta}_n$ est l'EMV de θ alors $\varphi(\tilde{\theta}_n)$ est l'EMV de $\varphi(\theta)$. De plus, si φ est dérivable, on a

$$\sqrt{n} [\varphi(\tilde{\theta}_n) - \varphi(\theta)] \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{\dot{\varphi}(\theta)^2}{I_n(\theta)} \right).$$

Ce résultat est connu sous le nom de méthode delta. Quand, $\varphi(\tilde{\theta}_n)$ est donc approximativement de loi $\mathcal{N} \left(\varphi(\theta), \frac{\dot{\varphi}(\theta)^2}{I_n(\theta)} \right)$

- En général, l'EMV est meilleur que l'EMM au sens où $\text{Var}(\hat{\theta}_n) \leq \text{Var}(\tilde{\theta}_n)$. C'est au moins vrai asymptotiquement.

2.7.3 Estimation par intervalle de confiance

Soit X une variable aléatoire dont la loi dépend d'un paramètre θ inconnu. Soit (X_1, X_2, \dots, X_n) un n -échantillon issu de X et $\alpha \in]0, 1[$.

Définition 2.13 *Un intervalle de confiance pour le paramètre θ , de niveau de confiance $1 - \alpha \in]0, 1[$, est l'intervalle qui a la probabilité $1 - \alpha$ de contenir la vraie valeur du paramètre θ .*

Remarque 2.6 *Un intervalle de confiance indique la précision d'une estimation car pour un risque α donné, l'intervalle est d'autant plus grand que la précision est faible.*

Principe de construction

Le principe de la méthode d'estimation par intervalle de confiance est le suivant :

Soit T un estimateur de θ (meilleur estimateur possible) dont on connaît la loi en fonction de θ .

On détermine par la suite un intervalle de probabilité de niveau $1 - \alpha$ pour T i.e :

$$\mathbb{P}(t_1(\theta) \leq T \leq t_2(\theta)) = 1 - \alpha.$$

Il faut ensuite inverser cet intervalle pour T , dont les bornes dépendent de θ pour obtenir un intervalle pour θ , i.e

$$\mathbb{P}_\theta(a(T) \leq \theta \leq b(T)) = 1 - \alpha,$$

Ou bien

$$\mathbb{P}_\theta(t_2^{-1}(t) \leq \theta \leq t_1^{-1}(t)) = 1 - \alpha.$$

Intervalle de confiance classiques

Intervalle de l'espérance μ d'une loi $\mathcal{N}(\mu, \sigma^2)$:

a) Cas où est σ^2 connue : On sait que \bar{X} est le meilleur estimateur de μ

$$\mathbb{E}(\bar{X}) = \mu, \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}.$$

$$\bar{X} \rightsquigarrow \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \implies \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \rightsquigarrow \mathcal{N}(0, 1).$$

$$\mathbb{P}\left(-\mu_{\alpha/2} \leq \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \leq \mu_{\alpha/2}\right) = 1 - \alpha.$$

$$\mathbb{P}\left(\mu - \frac{\sigma}{\sqrt{n}} \mu_{\alpha/2} \leq \bar{X} \leq \mu + \frac{\sigma}{\sqrt{n}} \mu_{\alpha/2}\right) = 1 - \alpha.$$

D'où

$$IC_{(1-\alpha)}(\mu) = \left[\bar{X} - \frac{\sigma}{\sqrt{n}} \mu_{\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}} \mu_{\alpha/2} \right].$$

Où : $\mu_{\alpha/2}$ est le fractile d'ordre

$(1 - \frac{\alpha}{2})$ de $\mathcal{N}(0, 1)$, i.e :

$$\mathbb{P}\left(\sqrt{n} \frac{\bar{X} - \mu}{\sigma} \leq \mu_{\alpha/2}\right) = 1 - \frac{\alpha}{2}.$$

D'où :

$$\mu_{\alpha/2} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right).$$

Pour $1 - \alpha = 0.95$,

$$\mathbb{P}\left(\sqrt{n} \frac{\bar{X} - \mu}{\sigma} \leq \mu_{\alpha/2}\right) = 1 - \frac{\alpha}{2} = 0.975$$

Sur la table de $\mathcal{N}(0, 1)$ on aura $\mu_{\alpha/2} = 1.96$

$$IC(\mu) = \left[\bar{X} - \frac{\sigma}{\sqrt{n}} 1.96, \bar{X} + \frac{\sigma}{\sqrt{n}} 1.96 \right].$$

b) Cas où σ^2 est inconnue : On a $S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ est un estimateur sans biais de σ^2 et :

$$\frac{X_i - \bar{X}}{\sigma} \rightsquigarrow \mathcal{N}(0, 1) \implies \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2} \rightsquigarrow \mathcal{X}_{n-1}^2.$$

D'après la décomposition de S^2 :

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

en dérivant par σ^2 :

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

On a $\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 \rightsquigarrow \mathcal{X}_n^2$, comme somme de n carrés de variables aléatoires centrées réduites. $\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2 \rightsquigarrow \mathcal{X}_1^2$, on en déduit que $\frac{nS^2}{\sigma^2} \rightsquigarrow \mathcal{X}_{n-1}^2$

$$\frac{nS^2}{\sigma^2} = \frac{(n-1)S'^2}{\sigma^2} \rightsquigarrow \mathcal{X}_{n-1}^2.$$

Alors la variable aléatoire :

$$T_{n-1} = \frac{\sqrt{n} \frac{\bar{X} - \mu}{\sigma}}{\sqrt{\frac{(n-1)S'^2}{\sigma^2(n-1)}}} = \frac{\bar{X} - \mu}{S'} \sqrt{n} \rightsquigarrow T_{(n-1)} \text{ Student.}$$

L'intervalle de probabilité est :

$$\mathbb{P} \left(-t_{\alpha/2} \leq \frac{\bar{X} - \mu}{S'} \leq t_{\alpha/2} \right) = 1 - \alpha$$

$$\mathbb{P} (|T_{n-1}| \leq t_{\alpha}) = 1 - \alpha.$$

ou bien $\mathbb{P} (|T_{n-1}| > t_{\alpha}) = \alpha$ (t_{α} est tiré de la table Student). On trouve l'intervalle de confiance pour μ :

$$\bar{X} - t_{\alpha/2} \frac{S'}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2} \frac{S'}{\sqrt{n}}$$

si $\alpha = 0.05$, $n = 10$ on trouve $t = 2.262$

$$IC(\mu) = \left[\bar{X} - 2.262 \frac{S'}{\sqrt{10}} \leq \mu \leq \bar{X} + 2.262 \frac{S'}{\sqrt{10}} \right].$$

Intervalle de confiance pour σ^2 d'une loi $\mathcal{N}(\mu, \sigma^2)$.

a) μ connu : On l'estime par $S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ et on a $\frac{(n-1)S'^2}{\sigma^2} \rightsquigarrow \mathcal{X}_{n-1}^2$

$$\mathbb{P} \left(k_1 \leq \frac{(n-1)S'^2}{\sigma^2} \leq k_2 \right) = 1 - \alpha$$

D'où :

$$IC(\sigma^2) = \left[\frac{(n-1)S'^2}{k_1}; \frac{(n-1)S'^2}{k_2} \right].$$

Exemple 2.6 Si $1 - \alpha = 0.95$, on cherche :

$$k_1 \text{ tel que } \mathbb{P} \left(\frac{(n-1)S'^2}{\sigma^2} < k_1 \right) = \alpha/2 = 0.025.$$

$$k_2 \text{ tel que } \mathbb{P} \left(\frac{(n-1)S'^2}{\sigma^2} < k_2 \right) = 1 - \alpha = 0.975.$$

Estimation et intervalle de confiance d'une proportion

Soit une population comportant deux modalités A et B . Soit p la proportion d'individus possédant la modalité A . $1 - p$ est donc la proportion des individus possédant la modalité B .

On extrait de la population un échantillon de taille n . Soit la variable aléatoire K_n , nombre d'individus dans l'échantillon ayant la modalité A .

Définition 2.14 La variable aléatoire $F = \frac{K_n}{n}$ s'appelle fréquence empirique sa réalisation f est la proportion d'individus dans l'échantillon ayant la modalité A .

Propriété :

$$K_n \rightsquigarrow \mathcal{B}(n, p), \mathbb{E}(K_n) = np, \text{Var}(K_n) = npq.$$

$$\mathbb{E}(F) = p \text{ et } \text{Var}(F) = \frac{p(1-p)}{n}$$

Loi de probabilité de F : Dès que $np > 5$ et $n(1-p) > 5$.

$$F \rightsquigarrow \mathcal{N}\left(p, \frac{p(1-p)}{n}\right) \text{ i.e. : } \frac{F-p}{\sqrt{\frac{p(1-p)}{n}}} \rightsquigarrow \mathcal{N}(0,1).$$

Approximativement.

Théorème 2.3 *La fréquence empirique $F = \frac{K_n}{n}$ est l'estimateur efficace de p . F est l'estimateur sans biais et convergent de p , on a :*

$$\begin{aligned} F &\rightsquigarrow \mathcal{N}\left(p, \sqrt{\frac{p(1-p)}{n}}\right), \\ \frac{F-p}{\sqrt{\frac{p(1-p)}{n}}} &\rightsquigarrow \mathcal{N}(0,1), \\ \mathbb{P}\left(-\mu \leq \frac{F-p}{\sqrt{\frac{p(1-p)}{n}}}\right) &= 1-\alpha, \\ \mathbb{P}\left(\frac{F-p}{\sqrt{\frac{p(1-p)}{n}}} \leq \mu\right) &= 1-\alpha/2, \quad \mu = \phi_{\mathcal{N}(0,1)}^{-1}(1-\alpha/2). \end{aligned}$$

Problème : $p(1-p)$ inconnu.

On remplace $\sqrt{\frac{p(1-p)}{n}}$ par $\sqrt{\frac{f(1-f)}{n}}$, f étant l'estimation de p . D'où

$$IC \left[f - \mu_{\alpha/2} \sqrt{\frac{f(1-f)}{n}}, f + \mu_{\alpha/2} \sqrt{\frac{f(1-f)}{n}} \right].$$

2.8 Tests statistiques

2.8.1 Introduction

Un test statistique est une procédure de décision qui permet de choisir entre deux hypothèses antagonistes (contraires) faites sur une population ou sur un plusieurs paramètres au vu d'un échantillon aléatoire. Un test statistique est basé sur les données d'une expérience aléatoire et l'essence d'un test est une règle qui dit au statisticien si les données qu'il a recueillies le conduisent à accepter ou rejeter l'hypothèse.

On distingue deux catégories de tests :

- a) **Test paramétriques** : tester certaine hypothèse relative à un plusieurs paramètres d'une variable aléatoire de densité de probabilité connue.
- b) **Test non paramétriques** : utilisée lorsqu'on a un échantillon aléatoire dont la loi est complètement inconnue.

2.9 Notions générales

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 2.15 *Un test est donné par une fonction $\Phi : E^n \longrightarrow \{0, 1\}$, 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 $\mathcal{R} = \{\Phi(X_1, \dots, X_n) = 1\}$. Évidemment, étant donnée une zone de rejet $\mathcal{R} \subset E^n$, on définit un test en posant $\Phi = \mathbf{1}_{\mathcal{R}}$.

Il existe deux sorte d'hypothèses : hypothèse simple et composite. Si l'hypothèse est réduite à un seul paramètre de l'espace des paramètres, on parlera d'hypothèse simple sinon on parlera d'hypothèse composite (multiple).

- $H_0 : \theta = \theta_0$ hypothèse simple.
- $H_1 : \theta > \theta_0$, $H_1 : \theta < \theta_0$, $H_1 : \theta \neq \theta_0$ sont composites (multiples).

Remarque 2.7 *La nature de H_0 détermine la façon de formuler H_1 par conséquent la nature unilatérale ou bilatérale du test.*

- ♣ **Test bilatérale** : $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$ ($\theta > \theta_0$ ou $\theta < \theta_0$).
- ♣ **Test unilatérale** : $H_0 : \theta = \theta_0$ contre $H_1 : \theta > \theta_1$ ou bien $H_0 : \theta = \theta_0$ contre $H_1 : \theta < \theta_0$.

Exemple 2.7 *Test sur le paramètre p de la loi Bernoulli. Soit $X \rightsquigarrow (\mathcal{B}p)$:*

$$H_0 : p = 1/2 \text{ contre } H_1 : p > 1/2.$$

On peut aussi tester $H_0 : p \geq 1/2$ contre $H_1 : p < 1/2$.

Remarque 2.8 Le niveau du test, ou sa sensibilité, est la probabilité de rejeter H_0 à tort :

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

La puissance du test est la fonction $\beta : \Theta_0 \rightarrow [0, 1]$ définie par $\beta(\theta) = \mathbb{P}((X_1, \dots, X_n) \in \mathcal{R})$. 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 une région critique \mathcal{R} telle que pour $\theta \in \Theta_0$,

$$\mathbb{P}((X_1, \dots, X_n) \in \mathcal{R}) = \alpha.$$

La fonction de puissance est souvent difficile à déterminer explicitement.

2.9.1 Types d'erreurs

Lorsqu'on effectue un test statistique quatre situations peuvent se présenter :

1. Rejeter H_0 alors qu'elle est vraie : on dit qu'on commet une erreur de premier espèce (erreur de type 1).
2. Accepter H_0 alors qu'elle est fautive (erreur de deuxième espèce ou bien de type 2).
3. Accepter H_0 alors qu'elle est vraie.
4. Rejeter H_0 alors qu'elle fautive.

Vérité Décision	H_0	H_1
H_0	$1 - \alpha$	β erreur de deuxième espèce
H_1	α erreur de première espèce	$1 - \beta$

2.9.2 Étapes de construction d'un test statistique

1. Choix de H_0 et H_1 .
2. Détermination de la variable de décision.
3. Calcul de la région critique W en fonction de α .
4. Calcul de la puissance $1 - \beta$.
5. Calcul de la valeur expérimentale de la décision.
6. Conclusion : rejet ou non rejet de H_0 .

2.9.3 Tests uniformément plus puissants

Définition 2.16 *Étant donné 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 $\Theta_0 = \{\theta_2\}$, 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 2.17 *On considère des hypothèses simples $\Theta_0 = \{\theta_0\}$ et $\Theta_1 = \{\theta_1\}$. Soit*

$$V_{\theta_0, \theta_1}(x) = \frac{L(x, \theta_1)}{L(x, \theta_2)},$$

le rapport de vraisemblance. On considère la famille de tests Φ_k dont la région critique \mathcal{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 2.9 *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 2.4 (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 $k_\alpha \in \mathcal{R}$ 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 $\forall \theta \in \{\theta_0, \theta_1\}$,*

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

2.9.4 Test paramétrique usuelles

(a) Test sur la moyenne d'une loi normale $\mathcal{N}(\mu, \sigma^2)$.

σ^2 connu : $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu = \mu_1 \end{cases}$ avec $\mu_1 > \mu_0$. On utilise la variable de décision \bar{X} .

$$W = \{x \in \mathbb{R}^n, \bar{x} > k\}, \bar{X} \rightsquigarrow \mathcal{N}(\mu, \sigma^2/n).$$

$$\mathbb{P}_{H_0}(W) = \alpha \implies \mathbb{P}_{H_0}(\bar{X} > k) = \alpha$$

$$\implies \mathbb{P}\left(\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > \frac{k - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right) = \alpha$$

$$\implies \mathbb{P}\left(\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \leq \frac{k - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right) = 1 - \alpha.$$

$$k = \frac{\sigma}{\sqrt{n}} \Phi_{\mathcal{N}(0,1)}^{-1}(1 - \alpha) + \mu_0.$$

$$\underline{\text{Décision}} : \begin{cases} \bar{x} > k, & \text{on rejette } H_0; \\ \bar{x} \leq k, & \text{on ne rejette pas } H_0. \end{cases}$$

– σ^2 inconnu : σ^2 est estimé par

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

On a

$$\frac{(n-1)S'^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2 \text{ et } T_{n-1} = \frac{\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{(n-1)S'^2}{(n-1)\sigma^2}}} = \frac{\bar{X} - \mu}{S'^2} \sqrt{n} \rightsquigarrow \text{Student à } (n-1).$$

$$\underline{H_0 : \mu = \mu_0, H_1 : \mu = \mu_1} :$$

$$W = \{x \in \mathbb{R}^n, |T_{n-1}| > k\}.$$

k est déterminé par

$$\mathbb{P}_{H_0}(W) = \alpha = \mathbb{P}\left(\left|\frac{\bar{X} - \mu_0}{S'^2} \sqrt{n}\right| > k\right) = \alpha.$$

Remarque sur les tests de la moyenne : Si la variable aléatoire parente X ne suit pas une loi normale, les tests précédents s'appliquent encore dès que n est assez grand ($n > 30$), en raison du théorème centrale limite.

(b) Test sur les proportions :

Soit X une variable aléatoire observée sur une population et soit (X_1, X_2, \dots, X_n) un n -échantillon extrait de cette population. On veut savoir si cet échantillon de fréquence observée $\frac{k}{n}f$ estimateur de p , appartient à une population de fréquence p_0 (sous H_0) ou à une autre population inconnu de fréquence p (sous H_1).

$$H_0 : p = p_0 \text{ contre, } H_1 : p \neq p_0$$

On sait que $\frac{k}{n} = f \rightsquigarrow \mathcal{N}(p_0, \sqrt{\frac{p_0 q_0}{n}})$ sous H_0

$$W = \left\{ x \in \mathbb{R}^n, \left| \frac{K}{n} - p_0 \right| > k \right\},$$

$$\mathbb{P} \left(\left| \frac{K}{n} - p_0 \right| > k \right) = \alpha \implies \mathbb{P} \left(\frac{\left| \frac{K}{n} - p_0 \right|}{\sqrt{\frac{p_0 q_0}{n}}} > \frac{k}{\sqrt{\frac{p_0 q_0}{n}}} \right) = \alpha.$$

$$k = \Phi_{\mathcal{N}(0,1)}^{-1}(1 - \alpha) \sqrt{\frac{p_0 q_0}{n}} = \Phi_{\mathcal{N}(0,1)}^{-1}(1 - \alpha) \sqrt{p_0 \frac{1-p_0}{n}}.$$

2.9.5 Test et statistique exhaustive

La considération d'une statistique exhaustive simplifie la recherche de la région critique du test. S'elle existe une statistique exhaustive T pour θ , de densité $g(t, \theta)$ on a :

$$L(x, \theta) = g(t, \theta) \cdot h(x)$$

Les test de N-P se réduit alors

$$\frac{g(t, \theta_1)}{g(t, \theta_0)} > k_\alpha.$$

2.9.6 Test entre deux hypothèses multiples (test de Lehman)

2.9.7 Famille à rapport de vraisemblance monotone

Si H_0 est composite, le risque de première espèce est une fonction de θ , et le niveau du test est :

$$\alpha = \sup_{\theta \in \Theta_0} \alpha(\theta), \Theta_0 \subset \mathbb{R}, \alpha(\theta) \leq \alpha.$$

Définition 2.18 Soit X_1, X_2, \dots, X_n un n -échantillon issu de X . La loi de X est dite à rapport de vraisemblance monotone s'il existe une statistique T à valeur dans \mathbb{R} tel que :

$$\frac{L(x, \theta_1)}{L(x, \theta_0)} = f_{\theta_0, \theta_1}(t).$$

Est une fonction monotone croissante de T si $\theta_0 < \theta_1$

Cas des familles exponentielles

$$\ln f(x, \theta) = a(x)q(\theta) + A(\theta) + B(x).$$

$$\ln \frac{L(x, \theta_1)}{L(x, \theta_0)} = [q(\theta_1) - q(\theta_0)] \sum_{i=1}^n a(x_i) + nA(\theta_1) - nA(\theta_0).$$

$T = \sum_{i=1}^n a(X_i)$ est une statistique exhaustive.

$\ln \frac{L(x, \theta_1)}{L(x, \theta_0)}$ est une fonction monotone croissante en T si et seulement si $q(\theta_1) - q(\theta_0) > 0$ lorsque $\theta_1 > \theta_0$ i.e q doit être fonction croissante de θ .

Remarque 2.10 Les lois de famille exponentielle vérifiant cette propriété sont donc à rapport de vraisemblance monotone.

2.9.8 Tests bilatérales

Test de la forme $H_0 : \theta \leq \theta_1$ ou $\theta \geq \theta_2$ contre $H_1 : \theta_1 < \theta < \theta_2$.

Théorème 2.5 (Lehmann) Soit X une variable aléatoire de loi $f(x, \theta)$ et T une statistique. Si la loi de X est à rapport de vraisemblance monotone en T alors il existe un test uniformément le plus puissant de niveau α donné dont la région de rejet est de la forme :

$$W = \{x \in \mathbb{R}^n, K_1 < t(x) < K_2\}.$$

Remarque 2.11 Les constantes K_1 et K_2 sont déterminées par : $\mathbb{P}_{H_0}(W) = \alpha$.

Test de la forme : $H_0 : \theta_1 \leq \theta \leq \theta_2$ contre $H_1 : \theta > \theta_2$ ou $\theta < \theta_1$

$$W = \{x \in \mathbb{R}, t(x) < c_1 \text{ ou } t(x) > c_2\}$$

Avec

$$\mathbb{P}_{H_0}(W) = \alpha \implies \mathbb{P}_{\theta_1}(T < c_1) = \mathbb{P}_{\theta_2}(T > c_2) = \frac{\alpha}{2}.$$

2.9.9 Tests unilatérales

Test de la forme $H_0 : \theta \leq \theta_0$ contre $H_1 : \theta > \theta_0$.

Théorème 2.6 Soit X une variable aléatoire de densité $f(x, \theta)$, si la loi de X est à rapport de vraisemblance monotone en une statistique T , alors il existe un test uniformément le plus puissant dont la région critique est de la forme :

$$W = (x \in \mathbb{R}^n, t(x) > k)$$

Tel que : $\mathbb{P}_{H_0}(W) = \alpha$.

Exemple 2.8 Soit $X \rightsquigarrow \mathcal{B}(p)$, X_1, X_2, \dots, X_n un n -échantillon issu de X .
Tester :

$$H_0 : p \leq p_0 \text{ contre, } H : p > p_0,$$

on a :

$$f(x, p) = p^x (1-p)^{1-x}, x \in 0, 1$$

$$\ln f(x, p) = x \ln \frac{p}{1-p} + \ln(1-p)$$

$$q(p) = \ln \frac{p}{1-p} = \ln p - \ln(1-p)$$

$$q'(p) = \frac{1}{p} + \frac{1}{1-p} > 0 \implies q$$

est croissant en p . D'où la loi de $\mathcal{B}(p)$ est à rapport de vraisemblance monotone en

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

avec

$$W = \left\{ x \in \mathbb{R}^n, \sum_{i=1}^n X_i > k_\alpha \right\}.$$

avec sous $H_0 : \sum_{i=1}^n X_i \rightsquigarrow \mathcal{B}(n, p_0)$ tabulée, on trouve k_α par : $\mathbb{P}_{H_0}(W) = \alpha$.

L'autre test unilatéral se détermine de la même manière : $H_0 : \theta \geq \theta_0$
Test de comparaison d'échantillons $H_1 : \theta < \theta_0$

$$W = \{x \in \mathbb{R}^n, T(x) < k_\alpha\}.$$

2.9.10 Tests fondés sur le rapport du maximum de vraisemblance

Définition 2.19 On appelle test du maximum de vraisemblance tout test fondé sur la région critique

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

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 :

$$\mathcal{R} = \left\{ x \in \mathbb{R}^n \mid \frac{\sup_{\theta \in \Theta_1} L(x, \theta)}{\sup_{\theta \in \Theta_0} L(x, \theta)} \right\}.$$

2.9.11 Test de comparaison d'échantillons

2.9.12 Test de Student et F.S pour des échantillons indépendants

Etant donné deux échantillons de taille n_1 et n_2 . Peut-on admettre qu'ils ont été prélevé dans une même population relativement à la variable étudiée, ces échantillons ayant-été prélevé indépendamment l'un de l'autre ?

Exemple 2.9 Les résultats scolaires des filles et des garçons sont-ils comparables ? Les ampoules de l'entreprise A ont elles une durée de vie plus longue que celle de l'entreprise B ?

Mathématiquement le problème se formalise de la manière suivante : on observe sur le premier échantillon les réalisations d'une variable aléatoire X_1 de fonction de répartition $F_1(x)$ et sur le deuxième échantillon les réalisations d'une variable aléatoire X_2 de fonction de répartition $F_2(x)$.

On test :

$$\begin{cases} H_0 : F_1(x) = F_2(x) \\ H_1 : F_1(x) \neq F_2(x) \end{cases}$$

Remarque 2.12 En pratique on se contente de vérifier l'égalité des espérances et des variances de X_1 et X_2 .

(A) Cas des échantillons Gaussiennes : $X_1 \rightsquigarrow \mathcal{N}(m_1, \sigma_1^2)$, $X_2 \rightsquigarrow \mathcal{N}(m_2, \sigma_2^2)$.

Les hypothèses reviennent à tester :

$$\begin{cases} H_0 : m_1 = m_2 \text{ et } \sigma_1^2 = \sigma_2^2 & \text{contre;} \\ H_1 : m_2 \neq m_2 \text{ et } \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

Le test va consister à tester d'abord les variances, si elles ne sont pas significativement différentes à tester ensuite les espérances mathématiques en admettant $\sigma_1 = \sigma_2$.

(a) Test des variances de Fisher-Snédecor :

$$S_1^2 = \frac{1}{n_1} \sum_{i=1}^n (X_{1i} - \bar{X}_1)^2 \text{ et } S_2^2 = \frac{1}{n_2} \sum_{i=1}^n (X_{2i} - \bar{X}_2)^2$$

$$\frac{n_1 S_1^2}{\sigma_1^2} \rightsquigarrow \chi_{n_1-1}^2, \quad \frac{n_2 S_2^2}{\sigma_2^2} \rightsquigarrow \chi_{n_2-1}^2.$$

Sous $H_0 : \sigma_1 = \sigma_2$.

$$F = \frac{\frac{n_1 S_1^2}{n_1 - 1}}{\frac{n_2 S_2^2}{n_2 - 1}} \rightsquigarrow \mathcal{F}(n_1 - 1, n_2 - 1).$$

Si $\sigma_1 = \sigma_2$, F ne doit pas être significativement différente de 1, F sera la variable

de décision.

$$W = \{x \in \mathbb{R}^n, f > k, k > 1\},$$

f est une réalisation de F .

Remarque 2.13 *En pratique, on met au numérateur la plus grande des deux quantités. Si*

$$n_1 = n_2 = n, F_{n_1-1, n_2-1} = \frac{S_1^2}{S_2^2}$$

Si le test aboutit à $\sigma_1 = \sigma_2$, on passe au test des espérances

(b) Test sur les espérances de Student :

Supposons que $\sigma_1 = \sigma_2$ on a :

$$\left\{ \frac{n_1 S_1^2}{\sigma_1^2} \rightsquigarrow \chi_{n_1-1}^2, \right. \text{ et } \left. \left\{ \frac{n_2 S_2^2}{\sigma_2^2} \rightsquigarrow \chi_{n_2-1}^2, \right. \right. \\ \left. \left. \bar{X}_1 \rightsquigarrow \mathcal{N}\left(m_1, \frac{\sigma}{\sqrt{n_1}}\right). \right. \right. \left. \left. \bar{X}_2 \rightsquigarrow \mathcal{N}\left(m_2, \frac{\sigma}{\sqrt{n_2}}\right). \right. \right.$$

$$\frac{n_1 S_1^2 + n_2 S_2^2}{\sigma^2} \rightsquigarrow \chi_{n_1+n_2-2}^2$$

mais

$$\bar{X}_1 - \bar{X}_2 \rightsquigarrow \mathcal{N}\left(m_1 + m_2, \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

σ étant connu, on utilise la loi de Student.

$$T_{n_1+n_2-2} = \frac{\bar{X}_1 + \bar{X}_2 - (m_1 + m_2)}{\sqrt{(n_1 S_1^2 + n_2 S_2^2) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sqrt{n_1 + n_2 - 2}.$$

Sous $H_0 : m_1 = m_2$, la région critique est :

$$W = \{x \in \mathbb{R}^n : |T| > k\}.$$

(B) Cas des échantillon non Gaussiennes :

Le test des variances F n'est plus valable car $\frac{nS^2}{\sigma^2} \not\rightarrow \chi_n^2$, mais pour n_1 et n_2 assez grand, on applique la formule de **Student** que σ_1 soit différent ou non de σ_2 .

Remarque 2.14 On dit que le test de **Student** est "robuste" car il résiste au changement de loi de X_1 et X_2

Exemple 2.10 Soit X_1, X_2, \dots, X_n un n -échantillon issu de $X \rightsquigarrow \mathcal{N}(m_1, \sigma_1^2)$ tel que :

$$n = 7, \bar{x} = 14.2, S_1^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 0.0114.$$

Soit Y_1, Y_2, \dots, Y_m un m -échantillon issu de $Y \rightsquigarrow \mathcal{N}(m_1, \sigma_1^2)$ tel que : $m = 5, \bar{y} = \frac{1}{m} \sum_{i=1}^m y_i = 14.5$ et

$$S_Y^2 = \frac{1}{m-1} \sum_{i=1}^m (y_i - \bar{y})^2 = 0.05.$$

Peut-on accepter que X et Y suivant la même loi normale au niveau $\alpha = 0.05$.

Exemple 2.11 1. On test d'abord l'égalité des variance :

$H_0 : \sigma_1^2 = \sigma_2^2$ contre $H_1 : \sigma_1^2 \neq \sigma_2^2$. Ce test est basé sur la statistique de **Fisher** : on a :

$$\frac{(n-1)S_1^2}{\sigma_1^2} \rightsquigarrow \chi_{n-1}^2, \frac{(m-1)S_Y^2}{\sigma_2^2} \rightsquigarrow \chi_{m-1}^2, F = \frac{\frac{S_Y^2}{\sigma_2^2}}{\frac{S_X^2}{\sigma_1^2}}.$$

Sous

$$H_0 : F = \frac{S_y^2}{S_x^2} \rightsquigarrow \text{Fisher}(m-1, n-1).$$

$$W = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m, f > k\}.$$

$$f = \frac{S_Y^2}{S_X^2} \frac{0.05}{0.0114} = 4.38.$$

$$\mathbb{P}_{H_0}(W) = \alpha \implies \mathbb{P}_{H_0}(F > k) = \alpha \implies \mathbb{P}_{H_0}(F_{(m-1, n-1)} < k) = 1 - \alpha = 0.95.$$

Sur la table $F(4.6, 0.95)$ on trouve $k = 4.53$ ou $f = 4.38 < 4.53 \implies$ on rejette pas H_0 i.e : $\sigma_1^2 = \sigma_2^2$.

2. Test sur les moyennes : $H_0 : m_1 = m_2$ contre $H_0 : m_1 \neq m_2$ (test bilattiral) on a : $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

L'estimateur de σ^2 est :

$$S^2 = \frac{1}{n+m-2} \left[\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2 \right]$$

$$\frac{(n+m-2)S^2}{\sigma^2} = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{\sigma^2} \rightsquigarrow \chi_{n+m-2}^2.$$

On a :

$$\bar{X} \rightsquigarrow \mathcal{N}\left(m_1, \frac{\sigma^2}{n}\right), \bar{Y} \rightsquigarrow \mathcal{N}\left(m_2, \frac{\sigma^2}{m}\right).$$

$$\bar{X} - \bar{Y} \rightsquigarrow \mathcal{N}\left(m_1 - m_2, \sigma \sqrt{\frac{1}{n} + \frac{1}{m}}\right).$$

$$\frac{\frac{\bar{X} - \bar{Y}}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}}}{\sqrt{\frac{(n-1)S_X^2 + (m-1)S_Y^2}{(n+m-2)\sigma^2}}} \rightsquigarrow T_{n+m-2}.$$

$$W = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m, |T| > t\}$$

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

$$\sqrt{\frac{n+m-2}{(n+m-2)S^2}} = \frac{\bar{X} - \bar{Y}}{S \sqrt{\frac{1}{n} + \frac{1}{m}}}.$$

$$S^2 = 0.026 \implies S = 0.16, T = \frac{0.30}{0.16 \cdot 0.47} = 3.98.$$

Sur la table de Student :

$$\mathbb{P}(|T_{n+m-2}| > t) = 0.05 \implies t = 2.22 \implies \text{on rejette } H_0.$$

Chapitre 3

Vraisemblance empirique

3.1 Introduction

La vraisemblance empirique est une méthode d'estimation inspirée de la méthode du maximum de vraisemblance usuelle. La méthode de vraisemblance empirique a été principalement introduite par [Owen,(1988)], [Owen, (1990)] et [Owen, (2001)], bien qu'on puisse la considérer comme une extension des méthodes de calage utilisées depuis de nombreuses années en sondage (voir [Deville et Sarndal, (1992)] et [Hartley et Rao, (1968)]). Cette méthode de type non-paramétrique consiste à maximiser la vraisemblance d'une loi ne chargeant que les données, sous des contraintes satisfaites par le modèle. [Owen, (1990)] a montré que l'on pouvait obtenir une version non-paramétrique du théorème de **WILKS**. Ce théorème établit la convergence du rapport de vraisemblance empirique vers une loi du \mathcal{X}_2 , permettant ainsi de réaliser des tests ou de construire des régions de confiance. Cette méthode proposée a été généralisée à de nombreux modèles, lorsque le paramètre d'intérêt est défini à partir de contraintes de moments ([Qin et Lawless, (1994)], [Newey, (2004)]).

3.2 Vraisemblance empirique et bootstrap

Comparant la méthode de vraisemblance empirique avec celle de bootstrap, les avantages principaux de la vraisemblance empirique découlent de l'utilisation de la fonction de vraisemblance.

Contrairement à la méthode bootstrap, [DiCiccio et al., (1991)] a montré que la vraisemblance empirique est corrigeable au sens de [Bartlett,(1937)], ceci signifie que l'on peut corriger la statistique de test d'un facteur qui permet de faire disparaître le biais du premier ordre. Ceci assure un meilleur comportement à distance finie, puisque la différence avec le comportement

asymptotique est réduit de $O(n^{-1})$ à $O(n^{-2})$. De plus, la vraisemblance permet de combiner des données qui proviennent de plusieurs sources avec des différents plans d'échantillonnage. En général, la vraisemblance empirique est facile à calculer par un problème d'optimisation convexe pour lequel l'optimum global unique peut être facilement trouvé par les moindres carrés itérés.

Le bootstrap et la vraisemblance empirique peuvent être combinés de manière efficace. Une manière consiste à utiliser la vraisemblance empirique pour déterminer la famille imbriquée des régions de confiance, et le bootstrap pour choisir lequel utiliser pour un niveau de confiance donné. Les vraisemblances paramétriques et empiriques peuvent être également combiné afin d'avoir des solutions pour des certains problèmes.

3.3 Pour quoi les méthodes conventionelles ne s'appliquent-elles pas ?

Voici les méthodes existantes :

Vraisemblance paramétrique

On suppose une distribution normale $N(\mu, \sigma^2)$ avec $\gamma = \kappa = 0$ et ne peut donc pas être utilisée pour tirer des conclusions sur l'asymétrie et l'aplatissement. Lorsque les données ne sont pas distribuées normalement, les méthodes basées sur la distribution normale donnent des conclusions asymptotiquement fiables pour la moyenne μ , en raison de la théorème limite central, si la distribution des données a une variance finie. La vraie variance de la moyenne de l'échantillon est σ^2/n , l'estimation de la valeur de σ^2 se rapproche de la vraie, puis les intervalles de confiance dont la largeur est basée sur la variance estimée sont asymptotiquement fiable. Le comportement des méthodes théoriques normales pour la variance

$$\sigma^2 = E[X - E(X)]^2 \text{ pour } X \in R$$

est plus subtile. Contrairement à l'asymétrie et à l'aplatissement, la distribution normale a un paramètre σ^2 pour la variance. La variance est généralement estimée par une quantité non biaisée.

La distribution normale $\kappa = 0$. Si la distribution de l'échantillonnage a $\kappa = 0$, alors le modèle théorique normal n'utilise pas une estimation asymptotiquement correcte de $nVar(s^2)$, $nVar(\hat{\sigma}^2)$, et les intervalles de confiance qui en résultent pour σ^2 ne s'approchent de leurs niveaux de couverture nominale à mesure que n augmente. Si la distribution normale est inadéquate, peut-être qu'une autre distribution est meilleure.

Il serait difficile de trouver une famille paramétrique qui corresponde aux données et aussi de l'asymétrie et de l'aplatissement qui peuvent varier librement. Il serait encore plus difficile de famille paramétrique finda dans laquelle les déductions pour l'asymétrie et l'aplatissement être fiable, même si les données ne proviennent pas d'un membre de cette famille paramétrique. Il serait difficile de trouver une famille paramétrique qui corresponde aux données et aussi de l'asymétrie et de l'aplatissement qui peuvent varier librement. Il serait encore plus difficile de famille paramétrique finda dans laquelle les déductions pour l'asymétrie et l'aplatissement être fiable, même si les données ne proviennent pas d'un membre de cette famille paramétrique.

Rappel

Rappelez-vous les mesures d'asymétrie (symétrie) et de kurtosis (lourdeur de la queue) :

$$\text{Asymétrie : } \quad \gamma = \frac{E \{(X - EX)^3\}}{\{Var(X)\}^{3/2}},$$

et

$$\text{kurtosis : } \quad \kappa = \frac{E \{(X - EX)^4\}}{\{Var(X)\}^2} - 3.$$

Bootstrap

difficile de choisir la région de confiance à partir d'un nuage de points composé d'un grand nombre d'estimations bootstrap pour (γ, κ) . par exemple, étant donné 1000 estimations bootstrap pour (γ, κ) , idéalement une région de confiance à 95 % devrait contenir 950 centimes points trals. en pratique, nous limitons aux régions rectangle ou ellipse afin de faciliter l'estimation.

3.3.1 Estimation de γ et κ

Soit

$$\bar{X} = n^{-1} \sum_{i=1}^n X_i,$$

et

$$\hat{\sigma}^2 = (n - 1)^{-1} \sum_{1 \leq i \leq n} (X_i - \bar{X})^2,$$

alors

$$\hat{\gamma} = \frac{1}{n \hat{\sigma}^3} \sum_{i=1}^n (X_i - \bar{X})^3, \quad \hat{\kappa} = \frac{1}{n \hat{\sigma}^4} \sum_{i=1}^n (X_i - \bar{X})^4.$$

comment trouver des ensembles de confiance pour (γ, κ) ?

Dans cette section, nous définirons $l(\gamma, \kappa)$ comme le log fonction de vraisemblance empirique de (γ, κ) , la région de confiance pour (γ, κ) est définie comme $\{(\gamma, \kappa) : l(\gamma, \kappa) > C\}$, où $C > 0$ est une constante déterminée par le niveau de confiance, c'est à dire

$$P[(\gamma, \kappa) > C] = 1 - a.$$

3.4 Introduction à la vraisemblance empirique

Définition 3.1 Soient $X_1, \dots, X_n \in \mathbb{R}$. La fonction de distribution cumulative empirique (ECDF) de X_1, \dots, X_n est

$$F(x) = \frac{1}{n} \sum_{i=1}^n 1_{X_i \leq x}, \text{ avec } -\infty < x < \infty.$$

Idée de base : supposons que F est une distribution discrète sur $\{x_1, \dots, x_n\}$ avec

$$p_i = F(x_i), \quad i = 1, \dots, n$$

Où

$$p_i \geq 0, \quad \sum_{i=1}^n p_i = 1.$$

Quelle est la fonction de vraisemblance de $\{p_i\}$ et qu'est-ce que le MLE ? depuis

$$p\{X_1 = x_1, \dots, X_n = x_n\} = p_1 \dots p_n,$$

La vraisemblance est :

$$\mathcal{L}(p_1, \dots, p_n) = \mathcal{L}(p_1, \dots, p_n; X) = \prod_{i=1}^n p_i$$

Qui est appelé une vraisemblance empirique.

Remarque 3.1 Le nombre de paramètre est le même que le nombre d'observations. Notez que

$$\left(\prod_{i=1}^n p_i\right)^{1/n} \leq \frac{1}{n} \sum_{i=1}^n p_i = \frac{1}{n}$$

l'égalité est valable si $p_1 = \dots = p_n = 1/n$ si $\hat{p}_i = 1/n$, on a

$$L(p_1, \dots, p_n; X) \leq L(\hat{p}_1, \dots, \hat{p}_n; X)$$

pour tout $p_i \geq 0$ et $\sum_{i=1}^n p_i = 1$ d'où la MLE basée sur vraisemblance empirique, qui est appelé l'estimateur empirique du maximum de vraisemblance (MELE), met une masse de probabilité égale $1/n$ sur les n valeurs observées x_1, x_2, \dots, x_n .

Exemple 3.1 Trouver le MELE pour $\mu = E[x_1]$ correspondant au EL

$$\mu = \sum_{i=1}^n p_i x_i = \mu(p_1, \dots, p_n),$$

Donc le MELE pour μ est

$$\hat{\mu} = \mu(\hat{p}_1, \dots, \hat{p}_n) = \bar{X}.$$

1. MELEs sans plus de contraintes sont simplement la méthode des estimateurs, qui ne sont pas nouveaux.
2. La vraisemblance empirique est un outil puissant pour traiter les hypothèses et les intervalles estimation de valeur dans une matière non paramétrique basée sur la tradition de vraisemblance, qui implique d'évaluer les MELE sous certaines contraintes supplémentaires.

3.5 Inférence de vraisemblance empirique pour les moyennes

Soit X_1, \dots, X_n un échantillon aléatoire d'une distribution inconnue.
objectif : tester l'hypothèse sur $\mu = E(X_1)$, ou trouver des intervalles de confiance pour μ .

3.5.1 Ratios de vraisemblance empirique (ELR)

Rappel

Théorème 3.1 (WILKS, 1938)

Si $\Theta_0 \subset \Theta$, $\dim\Theta = k$, $\dim\Theta_0 = l$, et

$$LR_0 = \frac{\sup_{\theta \in \Theta_1} L(\theta)}{\sup_{\theta \in \Theta_0} L(\theta)}. \quad (\text{eqwilks1})$$

Alors, sous $H_0 : \theta \in \Theta_0$ quand $n \rightarrow \infty$

$$-2 \log LR_0 \xrightarrow{L} \chi^2(K - l). \quad (\text{eqWilks2})$$

D'où la région de niveau asymp. α

$$\{-2 \log LR_0 \geq R\} \text{ où } \mathbb{P} [\chi^2(k-l) \geq r] = \alpha. \quad (\text{eqWilks3})$$

Corollaire 3.1 *une région de confiance pour θ (de niv asymp. $1 - \alpha$) est*

$$\{\theta \in \Theta / -2 \log LR(\theta) \leq r\} \text{ où } LR_0 = \frac{l(\theta)}{l(\hat{\theta})} \quad (\text{eqWilks4})$$

Et

$$\mathbb{P} [\chi^2(k) \geq r] = \alpha.$$

En inférence paramétrique, nous pouvons fonder les tests d'hypothèse et les régions de confiance sur le rapport de vraisemblance, si $L(\eta)$ est beaucoup plus petit que $L(\hat{\eta})$, alors nous rejetons l'hypothèse est $\eta_0 = \eta$, et exclure de notre région de confiance pour η_0 . Théorème de WILKS[Wilks,(1937)] prouvé que $2 \log (L(\eta_0)/L(\hat{\eta}))$ tend vers une distribution lorsque $n \rightarrow \infty$, dans les conditions de régularité modérées, nous permettons de décider à quel point $L(\eta)$ doit être pour que soit rejeté, les degrés de liberté dans la distribution sont généralement égales à la dimension de l'ensemble des valeurs η quand on veut un intervalle de confiance pour θ nous prenons l'image d'une confiance pour η , c'est :

$$\{\theta(\eta) \mid L(\eta) \geq cL(\hat{\eta})\},$$

Où c est le seuil qu'il choisit en utilisant le théorème de WILKS avec degrés de liberté égal à la dimension de l'ensemble des valeurs θ , nous pouvons également utiliser comme base les ratios de probabilité de la vraisemblance non paramétrique comme base d'hypothèse tests et intervalle de confiance, pour une distribution F , définir

$$R(F) = \frac{L(F)}{L(F_n)},$$

D'après la vraisemblance non paramétrique $L(F)$. Supposons que nous soyons intéressés par un paramètre $\theta = T(F)$ pour une fonction T des distributions, ce F est un membre d'un ensemble \mathcal{F} des distributions dans certains cas, nous pouvons prendre \mathcal{F} comme l'ensemble de toutes les distributions sur R , plus souvent nous utilisons un plus petit ensemble de distributions, définir la vraisemblance profile fonction de rapport

$$\mathcal{R}(\theta) = \sup \{R(F) \mid T(F) = \theta, F \in \mathcal{F}\}$$

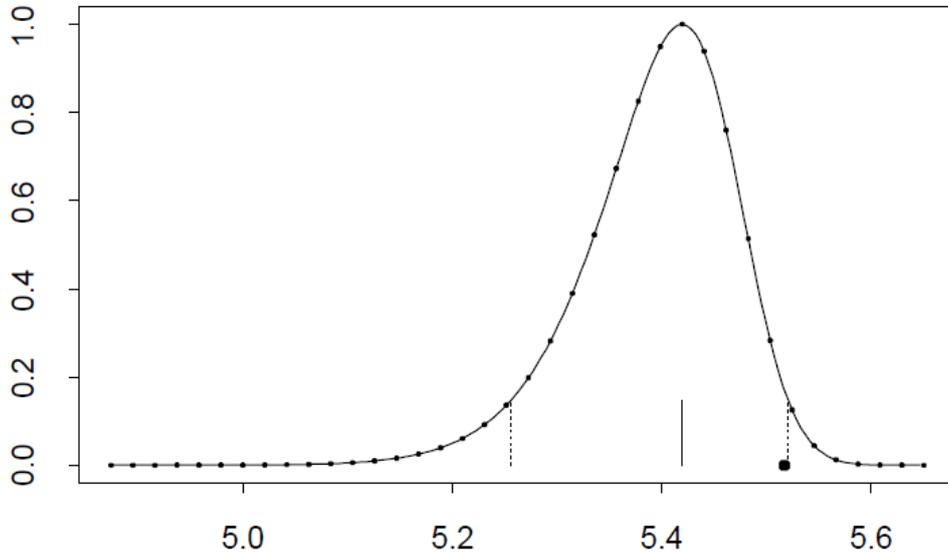


FIG. 3.1 – Profile de vraisemblance empirique

Les tests d'hypothèse de vraisemblance empirique rejettent $H_0 : T(F_0) = 0$, lorsque $R(\theta_0) < r_0$ pour une valeur seuil r_0 . Les régions de confiance de vraisemblance empirique est de la forme :

$$\{\theta \mid \mathcal{R}(\theta) \geq r_0\},$$

Dans de nombreux contextes, le seuil r_0 peut être choisi en utilisant une vraisemblance empirique théorème (ELT), un analogue non paramétrique du théorème de WILKS[Wilks,(1937)].

On considère l'hypothèse

$$H_0 : \mu = \mu_0 \quad \& \quad H_1 : \mu \neq \mu_0.$$

Soit $\mathcal{L}(p_1, \dots, p_n) = \prod_i p_i$, nous rejetons H_0 pour les grandes valeurs de l'ELR

$$T = \frac{\max \mathcal{L}(p_1, \dots, p_n)}{\max_{H_0} \mathcal{L}(p_1, \dots, p_n)} = \frac{L(n^{-1}, \dots, n^{-1})}{L(\tilde{p}_1, \dots, \tilde{p}_n)}.$$

où $\{\tilde{p}_i\}$ sont les MELE contraints pour $\{p_i\}$ sous H_0 .

1. Comment trouver $\{\tilde{p}_n\}$?
2. Quelle est la distribution de T sous H_0 ?

Les MELE contraints $\tilde{p}_i = p_i(\mu_0)$ où $\{p_i(\mu)\}$ sont la solution de la maximisation de problème soumis aux conditions

$$\max_{\{p_i\}} \sum_{i=1}^n \log p_i,$$

Si

$$p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i x_i = \mu.$$

La solution au problème ci-dessus est donnée dans le théorème ci-dessous, notez que

$$x_{(1)} = \min_i (x_i) \leq \sum_{i=1}^n p_i x_i \leq \max_i (x_i) \equiv x_{(n)}.$$

Il est donc naturel que nous ayons besoin de $x_{(1)} \leq \mu \leq x_{(n)}$.

Théorème 3.2 Pour $\mu \in (x_{(1)}, x_{(n)})$,

$$p_i(\mu) = \frac{1}{n - \lambda(x_i - \mu)} > 0, \quad 1 \leq i \leq n, \quad (3.1)$$

Où λ est la solution unique de l'équation

$$\sum_{j=1}^n \frac{x_j - \mu}{n - \lambda(x_j - \mu)} = 0, \quad (3.2)$$

Dans l'intervalle $(n/(X_{(1)} - \mu), n/(X_{(n)} - \mu))$.

Preuve: Nous utilisons la technique du multiplicateur de Lagrange pour résoudre ce problème d'optimisation, mettre

$$Q = \sum_i \log p_i + \psi \left(\sum_i p_i - 1 \right) + \lambda \left(\sum_i p_i x_i - \mu \right).$$

en laissant les dérivées partielles de Q et p_i, ψ et λ égales à 0, on a

$$p_i^{-1} + \psi + \lambda x_i = 0 \quad (3.3)$$

$$\sum_i p_i = 1, \quad (3.4)$$

$$\sum_i p_i x_i = \mu, \quad (3.5)$$

par l'équation (3.3),

$$p_i = -1/(\psi + \lambda x_i) \quad (3.6)$$

donc, $1 + \psi p_i + \lambda p_i x_i = 0$, ce qui implique $\psi = -(n + \lambda \mu)$, ceci avec l'équation (3.6) implique l'équation (3.1), ensuite par l'équation (3.1) et l'équation (3.5)

$$\sum_i \frac{x_i}{n - \lambda(x_i - \mu)} = \mu, \quad (3.7)$$

Il résulte de l'équation (3.4) que

$$\mu = \mu \sum_i p_i = \sum_i \frac{\mu}{n - \lambda(x_i - \mu)},$$

Ceci avec (3.7) implique (3.2).

Soit maintenant $g(\lambda)$ la fonction sur le LHS de l'équation (3.2), alors

$$\frac{d}{d\lambda} g(\lambda) = \sum_i \frac{(x_i - \mu)^2}{\{n - \lambda(x_i - \mu)\}^2} > 0,$$

Donc $g(\lambda)$ est une fonction strictement croissante. On remarque que

$$\lim_{\lambda \uparrow n/(X_{(n)} - \mu)} g(\lambda) = \infty, \quad \lim_{\lambda \downarrow n/(X_{(1)} - \mu)} g(\lambda) = -\infty,$$

Donc $g(\lambda) = 0$ a une solution unique dans l'intervalle,

$$\frac{1}{n - \lambda(X_{(1)} - \mu)} > 0, \quad \frac{1}{n - \lambda(X_{(n)} - \mu)} > 0$$

Et $1/\{n - \lambda(x - \mu)\}$ est une fonction monotone de X , il considère que $p_i(\mu) > 0$ pour tout $1 \leq i \leq n$. ■

Remarque 3.2 Lorsque $\mu = \bar{X}$, $\lambda = 0$, et $p_i(\mu) = 1/n$, $i = 1, \dots, n$.

a) Il peut être affiché pour μ , $E(x_i)$, et n large

$$p_i(\mu) \approx \frac{1}{n} \frac{1}{1 + \frac{\bar{x} - \mu}{s(\mu)}(x_i - \mu)},$$

Où

$$s(\mu) = (1/n) \sum_i (X_i - \mu)^2.$$

b) Nous pouvons voir

$$\mathcal{L}(\mu) = \mathcal{L}(p_1(\mu), \dots, p_n(\mu)).$$

Comme une vraisemblance empirique de profil pour μ . hypothétiquement
 $1 - 1$ un paramètre transformation de $\{p_1, \dots, p_n\}$ en $\{\mu, \theta_1, \dots, \theta_{n-1}\}$.
alors

$$\mathcal{L}(\mu) = \max \mathcal{L}(\mu, \theta_1, \dots, \theta_{n-1}) = \mathcal{L} \left\{ \mu, \hat{\theta}_1(\mu), \dots, \hat{\theta}_{n-1}(\mu) \right\}.$$

c) La fonction de vraisemblance $L(\mu)$ peut être calculé utiliser R-code, le
théorème asymptotique pour les tests classique du rapport de vrai-
semblance (c-à-d le théorème de WILKS) est valable pour les tests
ELR. Soit X_1, \dots, X_n iid et $\mu = E(X_1)$. tester

$$H_0 : \mu = \mu_0 \quad \& \quad H_1 : \mu \neq \mu_0$$

La statistique ELR est

$$\begin{aligned} T &= \frac{\max L(p_1, \dots, p_n)}{\max_{H_0} L(p_1, \dots, p_n)} = \frac{(1/n)^n}{L(\mu_0)} \\ &= \prod_{i=1}^n \frac{1}{np_i(\mu_0)} = \prod_{i=1}^n \left\{ 1 - \frac{\lambda}{n} (x_i - \mu_0) \right\}. \end{aligned}$$

Où λ est la solution unique de

$$\sum_{j=1}^n \frac{x_j - \mu_0}{n - \lambda(x_j - \mu_0)} = 0.$$

Théorème 3.3 Soit $E(X_1^2) < \infty$ puis sous H_0 ,

$$-2 \log T = 2 \sum_{i=1}^n \log \left\{ 1 - \frac{\lambda}{n} (x_i - \mu_0) \right\} \rightarrow \mathcal{X}_1^2.$$

Preuve: (*Esquisse*) Sous H_0 , $E(X_i) = \mu_0$. par conséquent, μ_0 est proche
de \bar{X} pour les grandes n , d'où le λ , ou plus précisément $\lambda_n \equiv \lambda/n$ est petit,
qui est la solution de $f(\lambda_n) = 0$, où

$$f(\lambda_n) = \frac{1}{n} \sum_{i=1}^n \frac{x_j - \mu_0}{1 - \lambda_n(x_j - \mu_0)}.$$

Par une simple expansion de Taylor $f(\lambda_n) \simeq f(0) + f'(0) \lambda_n$. Impliquant

$$\lambda_n \approx -\frac{f(0)}{f'(0)} = -(\bar{X} - \mu_0) \Big/ \left\{ \frac{1}{n} \sum_j (x_j - \mu_0)^2 \right\}.$$

Maintenant,

$$\begin{aligned}
 2 \log T &\approx 2 \sum_j \left\{ -\lambda_n (x_j - \mu_0) - \frac{\lambda_n^2}{2} (x_j - \mu_0)^2 \right\} \\
 &= -2\lambda_n n (\bar{X} - \mu_0) - \lambda_n^2 \sum_i (x_i - \mu_0)^2 \\
 &\approx \frac{n (\bar{X} - \mu_0)^2}{n^{-1} \sum_i (x_i - \mu_0)^2}
 \end{aligned}$$

Par le LLN,

$$n^{-1} \sum_i (x_i - \mu_0)^2 \longrightarrow \text{var}(X_1),$$

Par le CLT,

$$\sqrt{n} (\bar{X} - \mu_0) \longrightarrow N(0, \text{var}(X_1)).$$

Dans distribution, d'où $-2 \log T \longrightarrow \chi_1^2$ en distribution. ■

3.6 Intervalles de confiance pour μ

Définition 3.2 Pour un $\alpha \in (0, 1)$ donné, puisque nous ne rejeterons pas l'hypothèse nulle $H_0 : \mu = \mu_0$ si $2 \log T < \xi_{1, 1-\alpha}^2$

où $\{\chi_1^2 \leq \chi_{1, 1-\alpha}^2\} = 1 - \alpha$, pour $\alpha = 0.05$, $\chi_{1, 1-\alpha}^2 = 3.84$, d'où un $100(1 - \alpha)\%$, l'intervalle de confiance pour μ est

$$\begin{aligned}
 \{\mu : -2 \log \{L(\mu) n^n\} < \chi_{1, 1-\alpha}^2\} &= \left\{ \mu : \sum_i \log p_i(\mu) > -0.5 \chi_{1, 1-\alpha}^2 - n \log n \right\} \\
 &= \left\{ \mu : \sum_i \log(np_i(\mu)) > -0.5 \chi_{1, 1-\alpha}^2 \right\}.
 \end{aligned}$$

Exemple 3.2 On considère les données de Darwin : gains en hauteur des plantes grâce à la fertilisation croisée. $X =$ hauteur (Traveser-F) hauteur (Soi-F). il y a 15 observations :

6.1, -8.4, 1.0, 2.0, 0.7, 2.9, 3.5, 5.1, 1.8, 3.6, 7.0, 3.0, 9.3, 7.5, -6.0.

L'échantillon signifie $\bar{X} = 2,61$ et l'erreur standard $s = 4,71$.

Le gain est-il significatif?

Intuitivement : oui, si les observations négatives -8.4 et -6.0 n'existent pas.

Soit $\mu = E(X_i)$ et formulons les hypothèses comme

$$H_0 : \mu = 0 \text{ contre } H_1 : \mu > 0.$$

1. Approche standard : supposons que $\{X_1, \dots, X_{15}\}$ est un échantillon aléatoire de $\mathcal{N}(\mu, \sigma^2)$, le MLE est

$\hat{\mu} = \bar{X} = 2.61$, la statistique du test t est

$$T = \sqrt{n}\bar{X}/s = 2.14.$$

Puisque $T = t(14)$ sous H_0 , la valeur de p est 0.06 significative mais pas écrasante.

$\mathcal{N}(\mu, \sigma^2)$ est-il une hypothèse appropriée? Car les données ne semblent pas être normales (avec une queue gauche lourde)

2. Considérons une famille normale généralisée

$$f_k(x|\mu, \sigma) = \frac{2^{-1-1/k}}{T(1+1/k)\sigma} \exp\left\{-\frac{1}{2}\left|\frac{X-\mu}{\sigma}\right|^k\right\},$$

qui a la moyenne μ , lorsque $k = 2$, c'est $\mathcal{N}(\mu, \sigma^2)$, pour trouver la vraisemblance de profil de μ , le "MLE" pour σ est

$$\hat{\sigma}^k \equiv \hat{\sigma}(\mu)^k = \frac{k}{2n} \sum_{i=1}^n |X_i - \mu|^k.$$

3.7 Vraisemblance empirique pour des vecteurs aléatoires

Nous considérons les vecteurs aléatoires indépendants $X_i \in \mathbb{R}^d$ pour $d \geq 1$, en supposant qu'ils ont une distribution commune F_0 . Il n'est plus pratique pour décrire F_0 par une fonction de distribution cumulative, les dimensions d étant le long de laquelle il faut cumuler. Nous décrivons plutôt des distributions par les probabilités qu'ils attachent aux décors. Ainsi, $F(A)$ signifie $P(X \in A)$ pour $X \rightsquigarrow F$ et $A \subseteq \mathbb{R}^d$. Nous laissons δ_x indiquer la distribution sous laquelle $X = x$ avec une probabilité de 1. Ainsi $\delta_x(A) = 1_{x \in A}$. Nous constatons toujours que la distribution plaçant la probabilité $1/n$ sur chaque observation est un NPMLE, sous de légers changements de notation.

Définition 3.3 Soit $X_1, \dots, X_n \in \mathbb{R}^d$. La fonction de distribution empirique (EDF) de X_1, \dots, X_n est

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

Définition 3.4 Étant donné $X_1, \dots, X_n \in \mathbb{R}^d$, supposé indépendant avec DF cumulée F_0 , la vraisemblance non paramétrique de la DF F est

$$\mathcal{L}(F) = \prod_{i=1}^n F(\{X_i\}).$$

Ici, $F(\{X_i\})$ est la probabilité d'obtenir la valeur X_i dans un échantillon de F .

Il s'agit d'une interprétation très littérale de la vraisemblance.

Théorème 3.4 Soit $X_1, \dots, X_n \in \mathbb{R}^d$ des variables aléatoires indépendantes avec un commun DF F_0 . Que F_n soit leur EDF et que F soit n'importe quel DF. Si $F = F_n$, alors $\mathcal{L}(F) < \mathcal{L}(F_n)$.

3.7.1 EL pour les moyennes multivariées

Définition 3.5 La vraisemblance empirique du profil pour $\mu = E(X_1)$ est

$$\mathcal{L}(\mu) = \left\{ \max \prod_{i=1}^n p_i : p_i \geq 0, \sum_i p_i = 1, \sum_i p_i X_i = \mu \right\},$$

Où $p_i(\mu)$ est le MELE de p_i avec la contrainte supplémentaire $E(X_i) = \mu$. définir l'ELR

$$T \equiv T(\mu) = \frac{\mathcal{L}(1/n, \dots, 1/n)}{\mathcal{L}(\mu)} = 1 / \prod_{i=1}^n \{np_i(\mu)\}.$$

Théorème 3.5 Soit X_1, \dots, X_n et $d \times 1$ iid avec μ moyenne et matrice de covariance finie Σ avec $|\Sigma| \neq 0$. Quand $n \rightarrow \infty$,

$$-2 \log T \{(\mu)\} = -2 \sum_{i=1}^n \log \{np_i(\mu)\} \rightarrow \chi_d^2.$$

En distribution.

Remarque 3.3 1. Le cas où $|\sum| = 0$, il existe un entier $q < d$ pour lequel, $X_i = AY_i$ où Y_i est un $q \times 1$ variable aléatoire telle que $|\text{Var}(Y_i)| = 0$, et A est une $d \times q$ matrice constante. le théorème ci-dessus tient toujours avec la distribution limite remplacée par χ_q^2 .

2. L'hypothèse nulle $H_0 : \mu = \mu_0$ sera rejetée au niveau de signification α ssi :

$$\sum_{i=1}^n \log \{np_i(\mu_0)\} \leq -0.5\chi_d^2(1 - \alpha),$$

Où

$$p \{ \chi_d^2 \leq \chi_d^2(1 - \alpha) \} = 1 - \alpha.$$

3. Une région de confiance de $100(1 - \alpha)\%$ pour μ est :

$$\left\{ \mu : \sum_{i=1}^n \log \{np_i(\mu)\} \geq -0.5\chi_d^2(1 - \alpha) \right\}.$$

4. Calibration bootstrap : puisque (1) et (2) sont basées sur un résultat asymptotique quand n est petit et grand $\chi_{d, 1-\alpha}^2$ peut être remplacé par la valeur $[B\alpha]$ -ème parmi $2 \log T_1^*, \dots, 2 \log T_B^*$ qui sont calculés comme suit :

a. Dessinez l'échantillon iid X_1^*, \dots, X_n^* à partir de la distribution uniforme sur $\{X_1, \dots, X_n\}$. Laisser

$$T^* = 1 / \prod_{i=1}^n \{np_i^*(\bar{X})\}.$$

où $\bar{X} = (1/n) \sum_{i=1}^n X_i$ et $p_i^*(\mu)$ est obtenu de la même manière que $p_i(\mu)$ avec $\{X_1, \dots, X_n\}$ remplacé par $\{X_1^*, \dots, X_n^*\}$

b. Répéter (a) B fois, notez les valeurs B de T^* comme T_1^*, \dots, T_B^* .

c. Calcul de $p_i(\mu)$: hypothèse : $|\text{Var}(X_i)| = 0$ et μ est un point intérieur de la coque convexe enjambé par les observations, à savoir,

$$\mu \in \left\{ \sum_{i=1}^n p_i x_i : p_i > 0, \sum_{i=1}^n p_i = 1 \right\}.$$

cela garantit que les solutions $p_i(\mu)$ existent, nous résolvons le problème en 3 étapes.

i. Transformer le problème à n dimensions contraint en une dimension à contraint problème.

ii. Transformer le problème contraint en un problème non contraint.

iii. Appliquer un algorithme de [Newton MA et Raftery(1994)], laisser

$$\begin{aligned} l(\mu) &= \log L(\mu) = \sum_{i=1}^n \log p_i(\mu) \\ &= \max \left\{ \sum_{i=1}^n \log p_i : p_i > 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i X_i = \mu \right\}. \end{aligned}$$

Étape1 : similaire un précédent théorème 1, la méthode du multiplicateur lagrangien implique

$$p_i(\mu) = \frac{1}{n - \lambda_T (X_i - \mu)} \quad i = 1, \dots, n$$

Où λ est la solution de

$$\sum_{i=1}^n \frac{X_j - \mu}{n - \lambda_T (X_i - \mu)} = 0, \quad (3.8)$$

Par conséquent

$$l(\mu) = - \sum_{i=1}^n \log \{n - \lambda_T (X_i - \mu)\} \equiv M(\lambda).$$

$\frac{\partial}{\partial \lambda} M(\lambda) = 0$ conduit à (3.8), et

$$\frac{\partial^2 M(\lambda)}{\partial \lambda \lambda_T} = \sum_{i=1}^n \frac{(X_i - \mu) X_i - \mu)^T}{n - \lambda_T (X_i - \mu)} > 0.$$

Ainsi $M(\cdot)$ est une fonction convexe sur tout les ensembles connectés satisfaisant

$$n - \lambda_T (X_i - \mu) > 0 \quad i = 1, \dots, n. \quad (3.9)$$

de (3.9) et (3.8) on trouve $\sum_{i=1}^n p_i(\mu) = 1$ la dimension n d'origine le problème d'optimisation est équivalent à un problème d-dimensionnel de minimise $M(\cdot)$ soumis aux contraintes (3.9). Soit H_λ l'ensemble composé de toutes les valeurs de λ satisfaisant

$$n - \lambda_T (X_i - \mu) > 1, \quad i = 1, \dots, n.$$

alors \mathcal{H}_λ est un ensemble convexe dans \mathbb{R}^d , qui minimise de la fonction convexe $M(\lambda)$. malheureusement $M(\lambda)$ n'est pas défini sur les ensembles :

$$\{\lambda : n - \lambda_T (X_i - \mu) = 0 \quad i = 1, 2, \dots, n\}$$

Etape 2 : On étend $M(\lambda)$ en dehors de \mathcal{H}_λ telle sorte que c'est toujours une fonction convexe sur la ensemble \mathbb{R}^d définir

$$\log_*(z) = \begin{cases} \log z, & z \geq 1 \\ -1.5 + 2z - 0.5z^2, & z < 1 \end{cases}$$

Il est facile de voir que $\log_*(z)$ a deux dérivées continues sur \mathbb{R} .

$$M_*(\lambda) = - \sum \log_* \{n - \lambda_T (X_i - \mu)\}.$$

Alors

$$M_*(\lambda) = M(\lambda) \text{ sur } \mathcal{H}_\lambda, \text{ et } M_*(\lambda) \text{ et } M(\lambda).$$

Sont le même minimiseur qui est la solution de (3.8).

Etape 3 : nous appliquons un algorithme de [Newton MA et Raftery(1994)], pour calculer λ de manière itérative :

$$\lambda_{k+1} = \lambda_k - \left\{ \ddot{M}_*(\lambda_k) \right\}^{-1} \dot{M}_*(\lambda_k).$$

une valeur initial $\lambda_0 = 0$, correspondant à $p_i = 1/n$.

3.7.2 EL pour des fonctions régulières des moyennes

Les méthodes empiriques de vraisemblance pour la moyenne nécessitent quelques modifications du travail pour d'autres paramètres comme la variance. La variance de X est la moyenne de $(X - \mu)^2$ dans laquelle μ , la moyenne de X , est généralement inconnue. Si nous connaissons μ , nous pourrions alors former $(X_i - \mu)^2$ et construire une fonction de rapport de vraisemblance empirique pour leur moyen commun. Nous envisageons ici d'étendre les inférences empiriques de vraisemblance aux statistiques qui peuvent être écrites comme des fonctions lisses d'un vecteur à dimensions finies de moyens.

L'idée de base : Soit Y_1, \dots, Y_n iid des variables aléatoires de variance σ^2 , notez que

$$\sigma^2 = E(Y_i^2) - E^2(Y_i) = h(\mu).$$

où $\mu = E(X_i)$ et $X_i = (Y_i, Y_i^2)$, on peut déduire un intervalle de confiance pour σ^2 de cette de μ .

Théorème 3.6 Soit X_1, \dots, X_n des variables aléatoires $d \times 1$ iid de moyenne μ_0 et $(\text{Var}(X_1)) = 0$, soit $\theta = h(\mu)$ une fonction lisse de $\mathbb{R}^d \rightarrow \mathbb{R}^q$ où $q \leq d$, et $\theta_0 = h(\mu_0)$. nous supposons que :

$$|GG^T| \neq 0, G = \frac{\partial \theta}{\partial \mu^T},$$

Pour tout $r > 0$, soit

$$C_{1,r} = \left\{ \mu : \sum_{i=1}^n (np_i(\mu) \geq -0.5r) \right\},$$

et

$$C_{3,r} = \{\theta_0 + G(\mu - \mu_0) : \mu \in C_{1,r}\},$$

Alors comme $n \rightarrow \infty$,

$$p(\theta \in C_{3,r}) \rightarrow P(\chi_q^2 \leq r).$$

Remarque 3.4 l'idée de l'étalonnage bootstrap peut également être appropriée ici :

1. Dans plus de condition,

$$p(\theta \in C_{2,r}) \rightarrow p(\chi_q^2 \leq r) \text{ où } C_{2,r} = \{h(\mu) : \mu \in C_{1,r}\}.$$

2. $C_{2,r}$ est un ensemble de confiance réalisable, tandis que $C_{3,r}$ ne l'est pas puisque μ_0 et θ_0 sont inconnus en pratique. Notez que μ proche de μ_0 ,

$$\theta_0 + G(\mu - \mu_0) \approx h(\mu).$$

3. En général

$$p(\mu \in C_{1,r}) \leq p(\theta \in C_{2,r}).$$

4. Par le théorème,

$$p(\theta \in C_{1,r}) \rightarrow p(\chi_d^2 \leq r).$$

5. La fonction de vraisemblance empirique du profil de θ est :

$$L(\theta) = \max \left\{ \prod_{i=1}^n p_i(\mu) : h(\mu) = \theta \right\} = \max \left\{ \prod_{i=1}^n p_i : h \left(\sum_{i=1}^n p_i X_i \right) = \theta, \sum_{i=1}^n p_i = 1 \right\}.$$

Qui peut être calculé directement en utilisant la méthode du multiplicateur de Lagrange, l'ordinateur et plus impliquée pour h non linéaire.

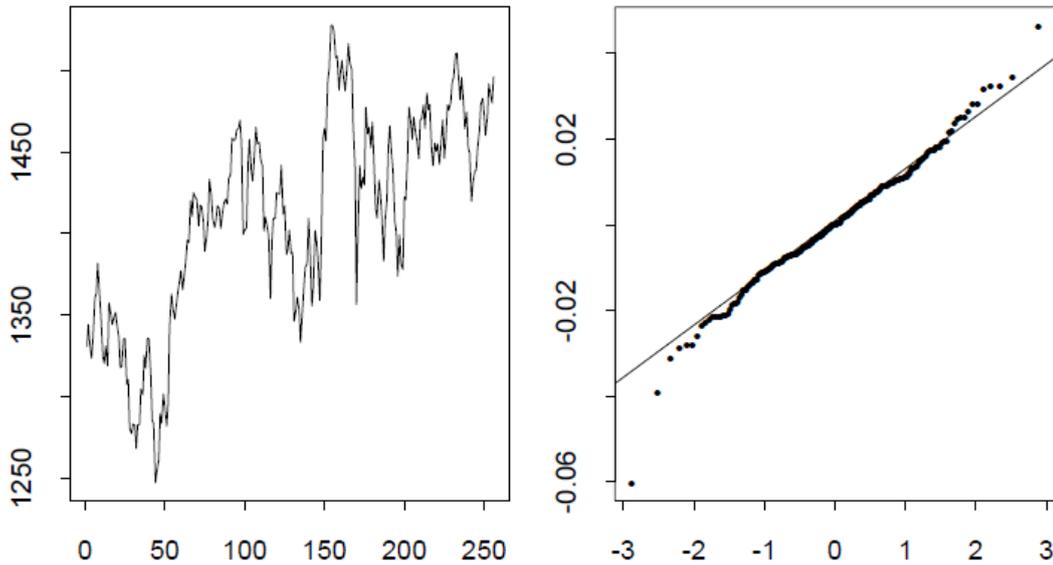


FIG. 3.2 – Le graphique de gauche montre la valeur de l'indice S&P 500 pendant 256 jours de négociation. celles-ci les valeurs ont été converties en 255 retours logarithmiques. Le graphique de droite montre un graphique QQ de ces retours sur l'indice S&P 500

Exemple 3.3 *Indice boursier S&P 500 du 17.8.1999 au 17.8.2000 (256 jours de bourse). Soit Y_i le prix le i -ème jour*

$$X_i = \log(Y_i/Y_{i-1}) \approx (Y_i - Y_{i-1})/Y_{i-1},$$

qui est le rendement, c'est - à - dire le pourcentage de variation au i ème jours, en évaluant X_i iid , nous construisons des intervalles de confiance pour la volatilité annuelle $\sigma = \{255\text{Var}(X_i)\}^{1/2}$. L'estimateur ponctuel simple est

$$\hat{\sigma} = \left\{ \frac{254}{256} \sum_{i=1}^{255} (X_i - \bar{X})^2 \right\}^{1/2} = 0.2116.$$

Les intervalles de confiance à 95% pour σ l'approche d'approximation normale est [0.1950; 0.2322] et pour la méthode EL est [0.1895; 0.2422] . L'intervalle de confiance EL est 41.67% plus large que l'intervalle basé sur une distribution normale, ce qui reflète le fait que les rendements ont une queue plus lourde.

3.8 Estimation des équations

Les équations d'estimation constituent un moyen extrêmement souple de décrire les paramètres et les statistiques correspondantes. Pour une variable aléatoire $X \in \mathbb{R}^d$, un paramètre $\theta \in \mathbb{R}^d$, et une fonction à valeur vectorielle $m(X, \theta) \in \mathbb{R}^s$ suppose que

$$E[m(X, \theta)] = 0. \quad (3.10)$$

Le réglage habituel a $p = s$ et ensuite, dans les conditions on $m(X, \theta)$ et éventuellement sur F , il y a une solution unique θ . Dans ce cas juste déterminé, la vraie valeur θ_0 peut être estimée en résolvant pour $\hat{\theta}$.

$$\frac{1}{n} \sum_{i=1}^n m(X_i, \hat{\theta}) = 0. \quad (3.11)$$

Pour écrire une équation vectorielle (3.10), nous prenons $m(X, \theta) = X - \theta$, et l'équation (3.11) donne $\hat{\theta} = \bar{X}$.

Pour $p(X \in A)$ prendre $m(X, \theta) = 1_{X \in A} - \theta$. Pour un scalaire distribué en continu X et $\theta \in \mathbb{R}$, la fonction $m(X, \theta) = 1_{X \leq \theta} - \alpha$

définit θ comme le quantile α de X . L'équation (3.11) est connue sous le nom d'équation d'estimation, et $m(X, \theta)$ est appelée fonction d'estimation. La plupart des estimateurs du maximum de vraisemblance sont définis par l'estimation des équations, ainsi que certaines statistiques robustes connus sous le nom de M-estimateurs.

En économétrie, un intérêt considérable est attaché au cas surdéterminé avec $s > p$. Entrées avec $s > p$ le fait que (3.10) tienne est une particularité de F et constitue une information secondaire importante. Même lorsque (3.10) est valable pour la vrai F_0 , il ne tiendra pas normalement pour le NPMLE \hat{F} , dans lequel le cas (3.11) n'a pas solution. La méthode généralisée des moments cherche une valeur $\hat{\theta}$ qui s'en rapproche à la résolution (3.11). Une approche de probabilité empirique de ce problème, décrite dans le cas sous-déterminé $s < p$ canaux où être utile. Les points (3.10) et (3.11) pourraient alors ont chacune un ensemble de solutions dimensionnelles $s - p$ de valeurs θ . Certaines fonctions de θ peuvent être déterminés avec précision à partir des données, alors que d'autres ne le seront pas, la probabilité empirique et les équations d'estimation sont bien adaptées l'une à l'autre.

La fonction de rapport de vraisemblance empirique pour θ est définie par

$$R(\theta) = \max \left\{ \prod_{i=1}^n n \omega_i / \sum_{i=1}^n \omega_i m(X_i, \theta) = 0, \omega_i \geq 0, \sum_{i=1}^n \omega_i = 1 \right\}. \quad (3.12)$$

3.8.1 Généralisation

Soit X_1, \dots, X_n iid à partir d'une distribution F . Nous nous intéressons à une caractéristique $\theta \equiv \theta(F)$, qui est déterminé par l'équation

$$E[m(X_1, \theta)] = 0,$$

Où θ est un vecteur $q \times 1$, m est une fonction à valeur vectorielle X_1 . Par exemple

1. $\theta = E[X_1]$ si $m(X, \theta) = X - \theta$.
2. $\theta = E[X_1^k]$ si $m(X, \theta) = X^k - \theta$.
3. $\theta = p(X_1 \in A)$ si $m(X, \theta) = I(X \in A) - \theta$.
4. θ est le quantile si $m(X, \theta) = I(X \leq \theta) - a$.

Un estimateur naturel pour θ est déterminé par l'équation d'estimation

$$\frac{1}{n} \sum_{i=1}^n m(x_i, \hat{\theta}) = 0. \quad (3.13)$$

De toute évidence, dans le cas où F est dans une famille paramétrique et m est la fonction de score, $\hat{\theta}$ est l'ordinaire MLE.

Le cas déterminé $q = s$: $\hat{\theta}$ peut être déterminé uniquement par (3.13).

Cas déterminé $q > s$: les solutions de (3.13) peuvent former un ensemble $(q - s)$ dimensionnel.

Cas déterminé $q > s$: (3.13) peut ne pas avoir de solution exacte, solutions approximatives sont recherchés, un tel exemple est ce qu'on appelle la méthode généralisée d'estimation des moments, qui est très populaire en économétrie.

Exemple 3.4 Soit $\{(X_i, Y_i), i = 1, \dots, n\}$ un échantillon aléatoire. trouver un ensemble d'équations d'estimation pour estimer $\gamma \equiv [var(X_1)/var(Y_1)]$. Pour estimer γ , nous devons estimer $\mu_X = E(X_1)$, $\mu_Y = E(Y_1)$ et σ^2 mettre $\theta_T = (\mu_X, \mu_Y, \sigma_Y^2, \gamma)$ et

$$m_1(X, Y, \theta) = X - \mu_X, \quad m_2(X, Y, \theta) = Y - \mu_Y$$

$$m_3(X, Y, \theta) = [(Y - \mu_Y)^2 - \sigma_Y^2]$$

$$m_4(X, Y, \theta) = [(Y - \mu_X)^2 - \sigma_Y^2 \gamma].$$

et $m = (m_1, m_2, m_3, m_4)^t$ alors $E\{m(X_i, Y_i, \theta)\} = 0$ conduisant à l'équation d'estimation.

Remarque 3.5 L'estimation de la méthode d'équation ne facilite pas les tests d'hypothèse de l'intervalle estimation pour θ .

3.8.2 El pour estimer les équations

La façon la plus répandue de définir les équations d'estimation est peut-être des vraisemblances paramétriques. Supposons que :

$$m(X, \theta) = \frac{\partial}{\partial \theta} \log f(X, \theta) = \frac{g(X, \theta)}{m(X, \theta)}$$

pour une fonction de densité de probabilité $f(X, \theta)$, avec le paramètre θ , et le gradient $g(X, \theta)$ par rapport à θ . Ensuite, l'équation d'estimation (3.13) est appelée équation de score et m est la fonction de score. La résolution de l'équation de score (3.13) est la méthode habituelle que le MLE paramétrique $\hat{\theta}$ est trouvé. Il existe parfois une solution unique $\hat{\theta}$ d'autres fois, il y en a plusieurs. Les inférences paramétriques de vraisemblance sont alors centrées sur une racine cohérente, mais pas nécessairement unique $\hat{\theta}$ de l'équation de score. En particulier il n'est pas nécessaire que X_i ait une densité de probabilité de la forme $f(X; \theta)$ pour tout θ . La quantité que $\hat{\theta}$ estime peut conserver une certaine interprétabilité, malgré les petites des écarts (et parfois malgré des écarts importants) par rapport au modèle paramétrique le motivant.

objectif : construire des tests statistique et des intervalles de confiance pour θ .

La fonction de vraisemblance empirique du profil de θ :

$$L(\theta) = \max \left\{ \prod_{i=1}^n p_i : \sum_{i=1}^n p_i m(X_i, \theta) = 0, p_i \geq 0, \sum_{i=1}^n p_i = 1 \right\}.$$

Théorème 3.7 Soit X_1, \dots, X_n iid, $m(X, \theta)$ une fonction de valeur vectorielle $s \times 1$, supposons que

$$E[m(X_1, \theta_0)] = 0, \quad | \text{var}(m(X_1, \theta_0)) | = 0.$$

Quand $n \rightarrow \infty$,

$$-2 \log \{L(\theta_0)\} - 2n \log n \rightarrow \chi_s^2.$$

En distribution.

Le théorème ci-dessus s'applique dans tout les cas déterminés, sous-déterminés et surdéterminés.

1. En général, $L(\theta)$ peut être calculé en utilisant la méthode pour EL pour signifie traiter $m(X_i, \theta)$ comme un vecteur aléatoire.

2. Pour $\theta = \hat{\theta}$ qui est la solution de

$$\frac{1}{n} \sum_{i=1}^n m(X_i, \theta) = 0, \quad L(\hat{\theta}) = (1/n)^n.$$

3. Pour θ déterminé par $E\{m(X_1, \theta)\} = 0$, nous rejettons l'hypothèse nulle $H_0 : \theta = \theta_0$ ssi :

$$\log\{L(\theta_0)\} + n \log n \leq -0.5\chi_{s, 1-\alpha}^2.$$

4. Tout $(1 - \alpha)$ ensemble de confiance pour θ déterminé par $E[m(X_1, \theta)] = 0$ est :

$$\theta : \log\{L(\theta)\} + n \log n > -0.5\chi_{s, 1-\alpha}^2.$$

Exemple 3.5 (Intervalles de confiance pour les quantiles) X_1, \dots, X_n soit iid pour un $\alpha \in (0, 1)$ donné, Laisser

$$m(X, \theta_\alpha) = I(X \leq \theta_\alpha) - \alpha.$$

Alors $E\{m(X_i, \theta_\alpha)\} = 0$ implique que θ_α est le q-quantile de la distribution de X_i . Nous supposons la vraie valeur de θ_α est comprise entre $X_{(1)}$ et $X_{(n)}$. L'équation d'estimation

$$\sum_{i=1}^n m(X_i, \hat{\theta}_\alpha) = \sum_{i=1}^n I(X_i \leq \theta_\alpha) - n\alpha = 0$$

implique $\hat{\theta}_\alpha = X_{(n\alpha)}$, où $X_{(i)}$ représente la i-ème plus petite valeur parmi X_1, \dots, X_n . Soit

$$L(\theta_\alpha) = \max \left\{ \prod_{i=1}^n p_i : \sum_{i=1}^n p_i I(X_i \leq \theta_\alpha) = \alpha, p_i \geq 0, \sum_{i=1}^n p_i = 1 \right\}.$$

Un intervalle de confiance $(1 - \beta)$ pour le quantile α est

$$\Theta_\alpha = \{\theta_\alpha : \log(L(\theta_\alpha)) > -n \log n - 0.5\chi_{1, 1-\beta}^2\}.$$

Remarque 3.6 $L(\hat{\theta}_\alpha) = (1/n)^n \geq L(\theta_\alpha)$ pour tout θ_α . Il est toujours vrai que $\hat{\theta}_\alpha \in \Theta_\alpha$. En fait $L(\theta_\alpha)$ peut être calculé explicitement comme suit $r = r(\theta_\alpha)$ l'entier pour lequel

$$X_{(i)} \leq \theta_\alpha, \text{ pour } i = 1, \dots, r.$$

$$X_{(i)} > \theta_\alpha, \text{ pour } i = r + 1, \dots, n.$$

Donc

$$\begin{aligned} L(\theta_\alpha) &= \max \left\{ \prod_{i=1}^n p_i : p_i \geq 0, \sum_{i=1}^r p_i = a, \sum_{i=1}^n p_i = 1 - a \right\} \\ &= (a/r)^r [\{(1-a)/(n-r)\}^{n-r}]. \end{aligned}$$

Par conséquent

$$\begin{aligned} \Theta_\alpha &= \{ \theta_\alpha : \log(L(\theta_\alpha)) \geq -n \log n - 0.5 \chi_{1,1-\beta}^2 \} \\ &= \left\{ \theta_a : r \log \frac{na}{r} + (n-r) \log \frac{n(1-a)}{n-r} \succ -0.5 \chi_{1,1-\beta}^2 \right\}. \end{aligned}$$

qui peut également être dérivée directement sur la base d'un test de rapport de vraisemblance pour une distribution binomiale .

Chapitre 4

Applications sur la régression linéaire

4.1 Introduction

Cela fait vingt ans que [Owen, (2001)] a publié son article fondateur [Owen,(1988)] qui a introduit la notion de vraisemblance empirique (EL). Depuis lors, il y a un riche corpus de littérature sur la nouvelle idée de la formulation de versions non paramétriques vraisemblance dans diverses définitions d'inférence statistique, Il y a eu deux grands points de vue sur la empirique. Le premier avis a été donné par [Hall et al., (1999)] dans les premières années de la méthode EL, qui résumait certaines propriétés clés de la méthode. Le second était le livre de l'inventeur de la méthodologie [Owen, (2001)], qui fournissait un aperçu complet jusqu'à cette date. Le corpus de littérature empirique sur les résultats augmente rapidement, et ce serait une tâche ardue de passer en revue le champ entier dans un document de revue comme celui-ci. Nous avons donc décidé de concentrer notre examen sur la régression en raison de son importance dans référence. Les modèles de régression examinés dans cette revue couvrent des paramètres paramétriques, modèles de régression métrique et semi-paramétrique. En plus du cas de complètement données observées, nous tenons également compte des données manquantes et censurées dans cette revue. La méthode EL [Owen,(1988), Owen, (1990)] est propriétaire de son large usage et de ses recherches rapidement à un certain nombre d'avantages importants. D'une manière générale, il combine la responsabilité des méthodes non paramétriques avec l'efficacité de l'approche de la vraisemblance. Il donne des régions de confiance qui respectent les limites du soutien du pays cible rameuter. Les régions sont invariantes sous les transformations et se présentent souvent mieux que régions de confiance

reconnues sur la normalité asymptotique lorsque la taille de l'échantillon est petite. De plus, ils sont de forme et d'orientation naturelle depuis les régions sont obtenues par contourner un rapport de vraisemblance logarithmique, et ils ne nécessitent pas souvent de l'attention de la variance, depuis l'étudiantisation est réalisée en interne via la procédure d'optimisation. La méthode EL est intéressante non seulement pour obtenir des régions de confiance, mais elle a ses attractions uniques dans les paramètres et la formulation de la qualité de l'ajustement tests.

4.2 Régression paramétrique

Supposons que nous observons un échantillon d'observations indépendantes $\{(X_i^T, Y)^T\}_{i=1}^n$, où chaque Y_i est considéré comme la réponse d'un plan d -dimensionnel (covariable) variable X_i . L'intérêt préalable est ici dans la fonction moyenne conditionnelle (régression fonction) de Y_i étant donné X_i . On distingue le design X_i soit fixe ou aléatoire. Malgré la régression est généralement associée à des plans fixes, pour faciliter de présentation, nous concentrons sur des designs aléatoires. La vraisemblance empirique l'analyse de la hotte pour les conceptions fixes peut généralement être étendue en régularisant dessins.

Considérons d'abord le modèle de régression paramétrique suivant :

$$Y_i = m(X_i, \beta) + \varepsilon_i, \quad \text{pour } i = 1, \dots, n, \quad (4.1)$$

où $m(X_i, \beta)$ est la fonction de régression connue avec une p -dimension inconnue ($p < n$) paramètre $\beta \in \mathbb{R}^p$, et les erreurs ε_i sont des variables aléatoires indépendantes telles que $E(\varepsilon_i | X_i) = 0$ et $Var(\varepsilon_i | X_i) = \sigma^2(X_i)$ pour une fonction $\sigma(\cdot)$. Par conséquent, les erreurs peuvent être hétéroscédastique. Nous exigeons, comme dans toutes les formulations empiriques de vraisemblance, que les erreurs ε_i ont une variance conditionnelle finie, qui est une condition minimale requise par la méthode de vraisemblance empirique pour assurer une distribution limite du chi-carré pour la rapport de vraisemblance empirique.

La fonction de régression paramétrique inclut comme cas particuliers :

- i) la régression linéaire avec $m(X_i, \beta) = x^T \beta$.
- ii) le modèle linéaire généralisé [McCullagh P.Nelder(1983)] avec $m(X_i, \beta) = G(x^T \beta)$ et $\sigma^2(x) = [\sigma_0^2 V \{G(x^T \beta)\}]$ pour une fonction de liaison connue G , une fonction de variance connue $V(\cdot)$ et une constante inconnue $\sigma_0^2 > 0$.

Notez que pour ces deux cas particuliers, $p = d$, en l'absence d'informations de modèle sur la variance conditionnelle, les moindres carrés (LS).

4.2.1 Estimation par Moindres carrées (LS)

L'estimateur de régression de β est obtenu en minimisant la somme des moindres carrés

$$S_n(\beta) = \sum_{i=1}^n \{Y_i - m(X_i, \beta)\}^2.$$

L'estimateur LS de β est

$$\hat{\beta}_{LS} = \arg \inf_{\beta} S_n(\beta).$$

Lorsque la fonction de régression $m(X_i, \beta)$ est suffisamment lisse par rapport à β , $\hat{\beta}_{LS}$ sera une solution de l'estimation suivante équation :

$$\sum_{i=1}^n \frac{\partial m(X_i, \beta)}{\partial \beta} \{Y_i - m(X_i, \beta)\} = 0. \quad (4.2)$$

Supposons que β_0 est la vraie valeur du paramètre telle que c'est la valeur unique à faire

$$E \left[\frac{\partial m(X_i, \beta)}{\partial \beta} \{Y_i - m(X_i, \beta)\} / X_i \right] = 0.$$

4.2.2 Estimation par vraisemblance empirique (EL)

Soit p_1, \dots, p_n un ensemble de poids de probabilité allouées aux données. La vraisemblance empirique (EL) pour β , dans l'esprit de [Owen,(1988)] [Owen, (1990)], est

$$L_n(\beta) = \max \prod_{i=1}^n p_i, \quad (4.3)$$

où la maximisation est soumise aux contraintes

$$\sum_{i=1}^n p_i = 1, \quad (4.4)$$

$$\sum_{i=1}^n p_i \frac{\partial m(X_i, \beta)}{\partial \beta} \{Y_i - m(X_i, \beta)\}. \quad (4.5)$$

La vraisemblance empirique, qui est présentée par l'équation (4.3), avec une contrainte triviale (4.4) indiquant que les p_i sont des poids de probabilité la contrainte (4.5) est la plus importante car elle définit la nature des paramètres. Cette formulation est similaire à la formulation originale donnée dans [Owen,(1988)] [Owen, (1990)] pour le paramètre moyen, disons

μ de X_i . La deuxième contrainte reflétant la nature de μ a été donné par $\sum_{i=1}^n p_i (X_i - \mu) = 0$.

En obtenant la probabilité empirique à chaque valeur de paramètre candidat β , ce qui précède problème d'optimisation comme indiqué dans (4.3), (4.4) et (4.5) doit être résolu pour optimiser p_i . Il peut être surprenant en premier lieu que le problème d'optimisation ci-dessus puisse admettre une solution car il y a np_i à déterminer avec seulement $p + 1$ contraintes. Comme la fonction objectif $L_n(\beta)$ est concave, et les contraintes sont linéaires dans les p_i , les problèmes d'optimisation admet des solutions uniques.

L'algorithme de calcul de $L_n(\beta)$ à un candidat β est le suivant. Si le convexe de l'ensemble des points (selon β)

$$\left\{ \frac{\partial m(X_i, \beta)}{\partial \beta} \{Y_i - m(X_i, \beta)\} \right\}_{i=1}^n .$$

Dans \mathbb{R}^p contient l'origine (zéro) dans \mathbb{R}^p , puis le problème d'optimisation EL pour $L_n(\beta)$ admet une solution. Si le zéro de \mathbb{R}^p n'est pas contenu dans la coque convexe des points pour le β donné, alors $L_n(\beta)$ n'admet pas de solution finie car certains poids p_i sont forcés de prendre des valeurs négatives, voir [Owen,(1988)] [Owen, (1990)] pour une discussion sur cet aspect. [Tsao M(2004)] ont étudié la probabilité que l'EL n'admette pas une solution finie et la dépendance de cette probabilité à la dimensionnalité.

En introduisant les multiplicateurs de Lagrange $\lambda_0 \in \mathbb{R}$ et $\lambda_1 \in \mathbb{R}^p$, les contraintes problème de optimisation (4.3)-(4.5) peut être traduit en un modèle sans contraintes avec une fonction objective

$$T(p, \lambda_0, \lambda_1) = \sum_{i=1}^n \log(p_i) + \lambda_0 \left(\sum_{i=1}^n p_i - 1 \right) + \lambda_1^t \sum_{i=1}^n p_i \frac{\partial m(X_i, \beta)}{\partial \beta} \{Y_i - m(X_i, \beta)\}, \quad (4.6)$$

où $p = (p_1, \dots, p_n)^t$. Différencier $T(p, \lambda_0, \lambda_1)$ par rapport à chaque p_i et en mettant la dérivée à zéro, on peut montrer après une algèbre que $\lambda_0 = -n$, et en définissant $\lambda = -n\lambda_1$, on trouve que les p_i optimaux sont donnés par

$$p_i = \frac{1}{n} \frac{1}{1 + \lambda^T \frac{\partial m(X_i, \beta)}{\partial \beta} \{Y_i - m(X_i, \beta)\}},$$

Où, de la contrainte structurelle (4.5), λ satisfait

$$\sum_{i=1}^n \frac{\frac{\partial m(X_i, \beta)}{\partial \beta} \{Y_i - m(X_i, \beta)\}}{1 + \lambda^T \frac{\partial m(X_i, \beta)}{\partial \beta} \{Y_i - m(X_i, \beta)\}} = 0. \quad (4.7)$$

Substitution des poids optimaux à la vraisemblance empirique dans (4.3), nous obtenons

$$L_n(\beta) = \prod_{i=1}^n \frac{1}{n} \frac{1}{1 + \lambda^T \frac{\partial m(X_i, \beta)}{\partial \beta} \{Y_i - m(X_i, \beta)\}},$$

et la vraisemblance empirique log est

$$I_n(\beta) = \log\{L_n(\beta)\} = - \sum_{i=1}^n \log \left\{ 1 + \lambda^T \frac{\partial m(X_i, \beta)}{\partial \beta} \{Y_i - m(X_i, \beta)\} \right\} - n \log(n). \quad (4.8)$$

Le caractère intensif en informatique de la probabilité empirique ressort clairement de ce qui précède discussions. En effet, pour évaluer l'EL à un β , il faut résoudre l'équation non linéaire (4.7) pour le λ qui dépend de β . Une approche informatique alternative, comme donné à [Owen, (1990)], est de traduire le problème d'optimisation (4.3) - (4.5) avec respect aux poids EL $\{p_i\}_{i=1}^n$ à son double problème par rapport à λ . Le double problème de (4.3) - (4.5) implique de minimiser la fonction objectif

$$1 + \lambda^T \frac{\partial m(X_i, \beta)}{\partial \beta} \{Y_i - m(X_i, \beta)\} \geq 1/n \quad \text{pour } i = 1, \dots, n. \quad (4.9)$$

La contrainte (4.9) vient de $0 \leq p_i \leq 1$ pour chaque i , alors que le gradient de $Q(\lambda)$ est la fonction sur le côté gauche de (4.7). Soit

$$D = \left\{ \lambda : 1 + \lambda^T \frac{\partial m(X_i, \beta)}{\partial \beta} \{Y_i - m(X_i, \beta)\} \geq 1/n \quad \text{pour } i = 1, \dots, n \right\}.$$

Ensuite, le double problème devient le problème de minimisation $Q(\lambda)$ sur l'ensemble D . On peut vérifier que D est convexe, fermé et compact. Par conséquent, il existe un minimum dans D . Comme suggéré dans [Owen, (1990)], l'ensemble D peut être supprimé par modification de la fonction $\log(x)$ dans $Q(\lambda)$ par un $\log^*(x)$ de telle sorte que $\log^*(x) = \log(x)$ pour $x \geq 1/n$ et $\log^*(x) = -n^2 x^2/2 + 2nx - 3/2 - \log(n)$ pour $x < 1/n$, qui est la fonction quadratique qui correspond à $\log(x)$ et à ses deux premières dérivées à $x = 1/n$.

Nous notons que la probabilité de profil $\prod_{i=1}^n p_i$ atteint son maximum n^{-n} quand tout les poids p_i égal n^{-1} pour $i = 1, \dots, n$. Ainsi, s'il existe un β , par exemple β , qui résolve (4.7) avec $\lambda = 0$, à savoir

$$\sum_{i=1}^n \frac{\partial m(X_i, \beta)}{\partial \beta} \{Y_i - m(X_i, \beta)\} = 0, \quad (4.10)$$

puis l'EL atteint son maximum $L_n(\hat{\beta}) = n^{-n}$ à β . Dans la régression paramétrique, nous envisageons, le nombre de paramètres et le nombre d'équations dans (4.10) sont le même. Par conséquent, (4.10) a une solution β avec une probabilité approchant une en grande échantillons. Il existe des situations d'inférence où le nombre d'équations d'estimation est supérieur au nombre de paramètres (à proprement parler, dimension du paramètre l'espace), par exemple, la méthode généralisée des moments en économétrie [Hansen LP(1982)]. Ici, plus d'informations sur le modèle sont prises en compte en imposant plus de moment restrictions, conduisant à plus d'équations d'estimation que le nombre de paramètres le modèle. Dans les statistiques, elles apparaissent sous la forme d'informations supplémentaires sur le modèle. Dans ces soi-disant situations identifiées, le maximum EL, toujours en utilisant la notation $L_n(\hat{\beta})$, peut être différent de n^{-n} . Voir [Qin et Lawless, (1994)] pour une discussion sur cette question.

En suivant la convention de la vraisemblance paramétrique standard, nous pouvons définir de (4.8) le rapport log EL

$$r_n(\beta) = -2 \log\{L_n(\beta) / L_n(\hat{\beta})\} = 2 \sum_{i=1}^n \log\{1 + \lambda^T \frac{\partial m(X_i, \beta)}{\partial \beta} \{Y_i - m(X_i, \beta)\}\}. \quad (4.11)$$

Théorème de Wilks [Wilks,(1937)] est une propriété clé du rapport de vraisemblance paramétrique. Si on remplace EL $L_n(\beta)$ par la vraisemblance paramétrique correspondante, disons $L_{pn}(\beta)$, et utiliser $r_{pn}(\beta)$ pour désigner le rapport de vraisemblance paramétrique, selon la théorème de Wilks [Wilks,(1937)], sous certaines conditions de régularité,

$$r_{pn}(\beta_0) \xrightarrow{d} \chi_p^2 \text{ as } n \rightarrow \infty. \quad (4.12)$$

Cette propriété est entretenue par l'EL, comme le montre [Owen, (1990)] pour le paramètre moyen, [Owen, (1990)] pour la régression linéaire et de nombreuses autres situations [Qin et Lawless, (1994)]. Dans le cadre de la régression paramétrique

$$r_n(\beta_0) \xrightarrow{d} \chi_p^2 \text{ quand } n \rightarrow \infty. \quad (4.13)$$

Cela peut être considéré comme une version non paramétrique du théorème de WILKS [Wilks,(1937)], et il est tout à fait remarquable pour la probabilité empirique de réaliser une telle propriété dans un cadre non réglage métrique avec des hypothèses de distribution beaucoup moins paramétriques. Nous appelons cela analogue de partager le théorème de WILKS l'analogue du premier ordre entre le paramétrique et la vraisemblance empirique.

Pour comprendre pourquoi la version non paramétrique du théorème de WILKS est valide, nous serions à présenter quelques étapes de dérivation qui offrent un aperçu de la non paramétrique probabilité. En règle générale, la première étape d'une étude sur l'EL envisage une extension λ défini dans (4.7) à β_0 , la vraie valeur de β , et déterminer son ordre de grandeur. Il peut être montré que pour la régression paramétrique actuelle,

$$\lambda = O_p(n^{-1/2}). \quad (4.14)$$

Un tel taux pour λ est obtenu dans les articles originaux d'[Owen,(1988)] [Owen, (1990)] pour la moyenne (qui peut être traité comme un cas trivial de régression sans covariables), dans [Owen, (1990)] pour la régression linéaire, et dans [Qin et Lawless, (1994)] pour le cas plus général d'estimation d'équations. Avec (4.14), (4.7) peuvent être inversés voir [DiCiccio et al., (1991)]. Pour simplifier la notation, définir

$$Z_{in} = \frac{\partial m(X_i, \beta)}{\partial \beta} \{Y_i - m(X_i, \beta)\}$$

Alors, (4.7) peut être inversé comme

$$\frac{1}{n} \sum z_{ni} (1 - \lambda^T Z_{ni}) + \frac{1}{n} \sum Z_{ni} \frac{\lambda^T Z_{ni} Z_{ni}^t \lambda}{1 + \lambda^T Z_{ni}} = 0.$$

Le dernier terme sur le côté gauche (LHS) est $O_p(n^{-1})$, ce qui est négligeable par rapport à le premier terme sur le LHS. Donc,

$$\lambda = S_n^{-1} n^{-1} \sum Z_{ni} + O_p(n^{1/2}).$$

Où $S_n = n^{-1} \sum Z_{ni} Z_{ni}^t$ Appliquer une expansion de Taylor de $\log(\cdot)$ autour de 1 et en remplaçant cette expansion d'un terme dans le rapport EL ration $r_n(\beta_0)$ dans (4.11), nous avons pour certains γ_i compris entre 1 et $1 + \lambda^T Z_{in}$ ($i = 1, \dots, n$)

$$\begin{aligned} r_n(\beta_0) &= 2 \sum_{i=1}^n \log(1 + \lambda^T Z_{in}) \\ &= 2 \sum_1^n \left\{ \lambda^T Z_{in} - \frac{1}{2} (\lambda^T Z_{in})^2 + \frac{1}{3} \frac{(\lambda^T Z_{in})^3}{(1 + \gamma_i)^3} \right\} \\ &= 2\lambda^t \sum_1^n Z_{in} - \lambda^t \sum_1^n Z_{in} Z_{in}^T \lambda + O_p(n^{-1/2}) \\ &= \left(n^{-1} \sum Z_{ni} \right)^T s_n^{-1} \left(n^{-1} \sum z_{ni} \right) + O_p(1). \end{aligned} \quad (4.15)$$

ce qui conduit au théorème de WILKS comme

$$S_n \xrightarrow{p} \sum (\beta_0) = E\{Z_{in}Z_{in}^T\}$$

et

$$n^{-1/2} \sum Z_{in} \xrightarrow{d} N\left(0, \sum (\beta_0)\right) \text{ lorsque } n \rightarrow \infty.$$

Comme couramment pratiqué en vraisemblance paramétrique, la version non paramétrique ci-dessus du théorème de WILKS peut être utilisée pour construire des régions de confiance du rapport de vraisemblance pour β_0 . Une région de confiance EL avec un niveau de confiance nominal $1 - \alpha$ est

$$I_{1-\alpha} = \{\beta : r_n(\beta) \leq \chi_{p, 1-\alpha}^2\},$$

où $\chi_{p, 1-\alpha}^2$ est le $(1 - \alpha)$ -quantile du χ_p^2 distribution. Théorème de WILKS dans (4.13) s'assure que

$$p\{\beta_0 \in I_{1-\alpha}\} \rightarrow 1 - \alpha \text{ quand } n \rightarrow \infty.$$

Cette construction reflète les régions de confiance conventionnelles du rapport de vraisemblance, sauf que le rapport EL est utilisé ici au lieu du rapport de vraisemblance paramétrique.

Notez que (4.15) montre également que la méthode EL est (de premier ordre) asymptotiquement proche de la méthode d'approximation normale. Cependant, la méthode normale nécessite l'estimation de la variance $\sum (\beta_0)$, alors que la méthode EL ne nécessite aucune estimation de variance explicite. En effet, l'étudentisation se fait en interne via la procédure d'optimisation. En plus de l'analogie de premier ordre entre le paramétrique et la vraisemblance empirique, il existe entre eux un analogue de second ordre sous la Correction de [?]. La correction de Bartlett est une élégante propriété de second ordre des rapports de vraisemblance métriques, qui a été conjecturé et proposé dans [?]. Il a été officiellement établie et étudiée dans une série d'articles, [Lawley DN(1956)], [Hayakawa T(1977)], [Barndorff, (1984)].

Soit

$$w_i = \sum (\beta_0)^{-1/2} Z_{ni} = (w_i^1, \dots, w_i^p)^T$$

pour $j_l \in \{1, \dots, p\}, l = 1, \dots, k, \alpha^{j_1, \dots, j_k} = E(w_i^{j_1}, \dots, w_i^{j_k})$ pour un k -ième moment de croisement multivarié de w_i . Par résumant l'existence de moments d'ordre supérieur de Z_{ni} , il peut être démontré par selon les expansions d'Edgeworth, la distribution du rapport de vraisemblance empirique admet l'extension suivante :

$$p\{r_n(\beta_0) \leq \chi_{p, 1-\alpha}^2\} = 1 - \alpha - a\chi_{p, 1-\alpha}^2 g_p(\chi_{p, 1-\alpha}^2) n^{-1} + O(n^{-3/2}), \quad (4.16)$$

où g_p est la densité du χ_p^2 distribution, et

$$a = p^{-1} \left(\frac{1}{2} \sum_{j,m=1}^p \alpha^{jjmm} - \frac{1}{3} \sum_{k,j,m=1}^p \alpha^{jkm} \alpha^{jkm} \right). \quad (4.17)$$

Cela signifie que pour la régression paramétrique, à la fois paramétrique et empirique les régions de confiance $I_{1-\alpha}$ ont un écart de couverture d'ordre n^{-1} . Partie de l'erreur de couverture est due au fait que la moyenne de $r_n(\beta)$ n'est pas d'accord avec p , la moyenne de χ_p^2 , c'est-à-dire $E\{r_n(\beta)\} \neq p$, mais plutôt

$$E\{r_n(\beta)\} = p(1 + an^{-1}) + O(n^{-2}).$$

Où α a été indiqué ci-dessus.

L'idée de la correction de Bartlett[Bartlett,(1937)] est d'ajuster le rapport EL $r_n(\beta_0)$ à $r_n^*(\beta_0) = r_n(\beta) / (1 + an^{-1})$ de sorte que $E[r_n^*(\beta_0)] = p + O(n^{-2})$. Et incroyablement simple l'ajustement de la moyenne conduit à une amélioration de (4.16) d'un ordre de grandeur [DiCiccio et al., (1991)]; [Chen.SX(1993)]; et de sorte que

$$P\{r_n^*(\beta_0) \leq \chi_{p,1-\alpha}^2\} = 1 - \alpha + O(n^{-2}). \quad (4.18)$$

4.3 Régression non paramétrique

Considérons dans cette section le modèle de régression non paramétrique

$$Y_i = m(X_i) + \varepsilon_i, \quad (4.19)$$

où la fonction de régression $m(x) = E[Y_i/X_i = x]$ est non paramétrique et X_i est d-dimensionnelle. Nous supposons que la régression peut être hétéroscédastique en ce que $\sigma^2(x) = \text{var}(Y_i/X_i = x)$, la variance conditionnelle de Y_i étant donné $X_i = x$, peut dépendre de x . La méthode de lissage du noyau est une méthode populaire pour estimer $m(x)$ non paramétriquement, voir [Härdle W(1990)] et [Fan J et Gijbels(1996)] pour plus de discussions. D'autres méthodes non paramétriques pour estimer $m(x)$ incluent les splines, ou séries orthogonales et méthodes d'ondelettes.

4.3.1 Estimateur de Nadaraya-Watson

L'estimateur de régression du noyau le plus simple pour $m(x)$ est l'estimateur [Nadaraya,(1965)] est donné par la formule suivante :

$$\hat{m}(x) = \frac{\sum_{i=1}^n k_h(x - X_i) Y_i}{\sum_{i=1}^n k_h(x - X_i)}, \quad (4.20)$$

où $K_h(t) = K(t/h) / h^d$, K est une fonction de noyau d -dimensionnelle, et h est une bande-largeur. L'estimateur de noyau ci-dessus peut être obtenu en minimisant les localisations suivantes somme pondérée des moindres carrés :

$$\sum_{i=1}^n K_h(x - X_i) \{Y_i - m(x)\}^2,$$

Par rapport à $m(x)$. C'est effectivement la solution des équations d'estimation suivantes :

$$\sum_{i=1}^n K_h(x - X_i) \{Y_i - m(x)\} = 0. \quad (4.21)$$

Dans le modèle de régression non paramétrique, le «paramètre» inconnu est la fonction de régression $m(x)$ elle-même. La probabilité empirique pour $m(x)$ à un x fixe peut être formulée d'une manière similaire au paramètre de régression paramétrique considéré dans la section précédente. Comme la probabilité empirique est également appliquée à la moyenne pondérée $\sum_{i=1}^n K_h(x - X_i) m(x)$, il est également similaire à l'EL d'une moyenne.

4.3.2 Estimateur de Buckley-James

c'est un Modèles AFT (Accelerated Failure Time) : modèle de régression linéaire à l'analyse des données de survie. L'hypothèse du modèle est que pour les temps de survie

$$Y_i = \log(T_i) = X_i^T \beta + \varepsilon_i, \quad (\text{eqBJ1})$$

où la distribution $F\varepsilon(\cdot)$ du terme d'erreur ε_i n'est pas spécifiée (sauf pour un emplacement exigeance d'identification telle que zéro moyen ou zéro médian). Les X_i sont les covariables observées, fixes ou aléatoires.

Les réponses T_i , ou $\log(T_i) = Y_i$ sont soumises à droit censure. $Q_i = \min(T_i, C_i)$; $\delta_i = I[T_i \leq C_i]$ où C_i sont les temps de censure.

L'estimateur[Buckley and James (1979)] c'est un estimateur des moindres carrés et inférence de β .

Estimateur Buckley-James[Buckley and James (1979)] pour le modèle de régression AFT :

Hypothèses supplémentaires :

1. Erreur ε_i sont iid.
2. Les covariables X_i ; $i = 1, \dots, n$ observable k -dimensionnel vecteurs constants.

3. B-J doit estimer de manière non paramétrique la distribution des erreurs F_ε .

4. Résidus(censuré) :

$$e_i(b) = Q_i - X_i^T b. \quad (\text{eqBJ2})$$

L'estimateur B-J de β est la solution de l'estimation équation :

1.

$$0 = \sum_{i=1}^n X_i^T \left\{ \delta_i e_i(b) + (1 - \delta_i) \sum_{j:e_j > e_i} \frac{e_j(b) \Delta \hat{F}(e_j)}{1 - \hat{F}(e_i)} \right\}, \quad (\text{eqBJ3})$$

où $\hat{F}(\cdot)$ est l'estimateur de Kaplan-Meier calculé à partir de $(e_i(b), \delta_i)$.

L'expression $\sum_{j:e_j > e_i} \frac{e_j(b) \Delta \hat{F}(e_j)}{1 - \hat{F}(e_i)}$ est une estimation du conditionnel espérance, $\hat{E}(Y_i - X_i b | Q_i, \beta = b, \delta_i = 0)$.

on Compare cela à l'équation d'estimation des moindres carrés, (cas non censuré) :

$$0 = \sum X_i^T (Y_i - X_i^T \beta). \quad (\text{eqBJ5})$$

L'équation d'estimation B-J [Buckley and James (1979)] conduit naturellement à une itérative solution de β .

Echo LJ Wei ont dit que il est plus facile d'obtenir la moyenne erreurs de prédiction dans le modèle AFT. (Comparer à Cox modèle des risques proportionnels) Il y a deux signes de sommation dans l'équation B-J. (Une pour l'index i un pour l'index j). Échangez la commande des sommations que nous avons Cette deuxième forme d'équation d'estimation B-J sera utilisée à plusieurs endroits plus tard.

$$0 = \sum_{j=1}^n \delta_j e_j(b) \left\{ X_j + \sum_{i:e_i < e_j} \frac{e_i(b) \Delta \hat{F}(e_i)}{1 - \hat{F}(e_j)} \right\}. \quad (\text{eqBJ5})$$

Inférence de l'estimateur BJ :

1. Wald [Buckley and James (1979)].
2. Score([Liu RYet Singh K(1981)], + nouveau).
3. Rapport de vraisemblance ([Mai Zhou,(2004)] **et** [Mai Zhou(2016)]).

4.3.3 Estimation par vraisemblance empirique (EL)

Soit p_1, \dots, p_n des poids de probabilité s'ajoutant à un. La vraisemblance empirique évaluée à $\theta(x)$, une valeur candidate de $m(x)$, est

$$L_n\{\theta(x)\} = \max \prod_{i=1}^n p_i, \quad (4.22)$$

Où la maximisation est soumise sous la condition triviale

$$\sum_{i=1}^n p_i = 1,$$

et la contrainte

$$\sum_{i=1}^n p_i k_h(x - X_i) \{Y_i - \theta(x)\} = 0. \quad (4.23)$$

En comparant cette formulation de l'EL à celle de la régression paramétrique, on voit que les deux formulations sont largement similaires sauf que (4.23) est utilisé comme contrainte structurelle au lieu de (4.5). Cette comparaison met en évidence le rôle joué par la contrainte structurelle dans la formulation EL. En effet, différentes contraintes structurelles donnent lieu à EL pour différents «paramètres» (quantité d'intérêt), tout comme donnent lieu à différentes probabilités paramétriques. En général, la probabilité empirique est formulée à partir des paramètres d'intérêt via les contraintes structurelles, et la vraisemblance paramétrique est entièrement basée sur un modèle paramétrique. L'algorithme de résolution du problème d'optimisation ci-dessus (4.22) - (4.23) est similaire à l'algorithme EL pour la régression paramétrique donnée sous (4.4) et (4.5), sauf qu'il peut être considéré plus facilement car le "paramètre" est unidimensionnel si nous ignorons la problème de sélection de la bande passante pour la régression non paramétrique. En présentant les multiplicateurs de Lagrange comme nous l'avons fait dans (4.6) dans la section précédente, nous avons que l'EL optimal les poids pour le problème d'optimisation ci-dessus à $\theta(x)$ sont donnés par :

$$p_i = \frac{1}{n} \frac{1}{1 + \lambda(x) k_h(x - X_i) \{Y_i - \theta(x)\}},$$

Où $\lambda(x)$ est un multiplicateur de Lagrange univarié qui satisfait

$$\sum_{i=1}^n \frac{k_h(x - X_i) \{Y_i - \theta(x)\}}{1 + \lambda(x) k_h(x - X_i) \{Y_i - \theta(x)\}}. \quad (4.24)$$

En substituant les poids optimaux à la vraisemblance empirique dans (4.22), la probabilité évaluée à $\theta(x)$ est

$$L_n\{\theta(x)\} = \prod_{i=1}^n \frac{1}{n} \frac{1}{1 + \lambda(x) k_h(x - X_i) \{Y_i - \theta(x)\}},$$

et la vraisemblance empirique log est

$$\begin{aligned} l_n\{\theta(x)\} &= \log L_n\{\theta(x)\} \\ &= -\sum \log[1 + \lambda(x) k_h(x - X_i) \{Y_i - \theta(x)\}] - n \log(\pi) \end{aligned} \quad (4.25)$$

L'EL global est maximisé à $p_i = n^{-1}$, ce qui correspond à $\theta(x)$ étant l'estimateur [Nadaraya,(1965)] $\hat{m}(x)$ à 4.20. Par conséquent, nous pouvons définir le rapport log EL à $\theta(x)$:

$$\begin{aligned} r_n\{\theta(x)\} &= -2 \log[L_n\{\theta(x)\}/n^{-n}] \\ &= 2 \sum_{i=1}^n \log[1 + \lambda(x) k_h(x - X_i) \{Y_i - \theta(x)\}]. \end{aligned} \quad (4.26)$$

L'EL ci-dessus n'est pas réellement pour $m(x)$, la vraie valeur de la fonction sous-jacente à x , mais plutôt pour $E\{m(x)\}$. Cela peut être effectivement détecté par la forme de la structure contrainte 4.23. Il est bien connu dans l'estimation du noyau que $m(x)$ est pas impartiale estimateur de $m(x)$, comme c'est le cas pour presque tous les estimateurs non paramétriques. Pour l'Estimateur NADARAYA – WATSON,

$$E\{\hat{m}(x)\} = m(x) + b(x) + o(h^2)$$

où $b(x) = \frac{1}{2}h^2\{m''(x) + 2m'(x)f'(x)/f(x)\}$ est le principal biais du noyau estimateur, et f est la densité de X_i . Ensuite, l'EL est en fait évalué à un $\theta(x)$, c'est une valeur candidate de $m(x) + b(x)$ au lieu de $m(x)$. Il existe deux stratégies pour réduire l'effet du biais [Hall et Scala, (1990)]. L'une consiste à sous-lisser avec une bande passante $h = o(n^{-1/(4+d)})$, l'ordre optimal de bande passante qui minimise le carré moyen erreur d'estimation avec un noyau de second ordre (d est la dimension de X). Un autre est d'estimer explicitement le biais, puis de le soustraire de l'estimation du noyau. nous considérons ici la première approche du sous-lissage pour des raisons de simplicité.

En cas de sous-lissage de telle sorte que $n^{2/(4+d)}h^2 \rightarrow 0$, le théorème de WILKS est valable pour l'EL sous la régression non paramétrique actuelle en ce que

$$r_n\{m(x)\} \rightarrow \chi_1^2, \quad \text{comme } n \rightarrow \infty.$$

Cela signifie qu'un intervalle de confiance de vraisemblance empirique avec une couverture nominale égal à $1 - \alpha$, noté $I_{1-\alpha,el}$ est donné par

$$I_{1-\alpha,el} = \{\theta(x) : r_n\{\theta(x)\} \leq \chi_{1,1-\alpha}^2\}. \quad (4.27)$$

Une particularité de l'intervalle de confiance de vraisemblance empirique est qu'aucun estimateur de variance est nécessaire dans sa construction au

fur et à mesure que l'étudiantisation est effectuée en interne via la procédure d'optimisation.

Définissons $\omega_i = k_h(x - X_i) \{Y_i - m(x)\}$ et, pour les entiers positifs j ,

$$\bar{\omega}_j = n^{-1} \sum_{i=1}^n \omega_i^j, \quad \mu_j = E(\varpi_j) \quad \text{et} \quad R_j(k) = \int k^j(u) du.$$

On note ici que le biais dans le lissage du noyau fait $\mu_1 = O(h^2)$, alors que dans le cas de régression paramétrique $\mu_1 = 0$. Il est montré dans [Chen SX et Qin Y-S(2003)] que la probabilité de couverture $I_{1-\alpha,el}$ admet l'extension Edgeworth suivante :

$$p\{m \in I_{1-\alpha,el}\} = 1 - \alpha - \left\{ nh^{-d} \mu_1^2 \mu_2^{-1} + \left(\frac{1}{2} \mu_2^{-2} \mu_4 - \frac{1}{3} \mu_2^{-3} \mu_3^2 \right) (nh^d)^{-1} \right\} z_{1-\frac{\alpha}{2}} \phi(z_{1-\frac{\alpha}{2}}) + O\{nh^{d+6} + h^4 + (nh)^{-1} h^2 + (nh^d)^{-2}\}. \quad (4.28)$$

Où ϕ et $Z_{1-\frac{\alpha}{2}}$ sont la densité et le $(1 - \frac{\alpha}{2})$ quantile d'une normale standard Variable aléatoire.

L'expansion ci-dessus n'est pas standard en ce que l'erreur de couverture principale consiste de deux termes. Le premier terme, $nh^d \mu_1 \mu_2^{-1}$, d'ordre nh^{d+4} est dû au biais de la lissage du noyau. Le second mandat $(nh^d)^{-1}$ est largement similaire à la principale erreur de couverture pour la régression paramétrique en (4.16). Nous notons qu'au deuxième mandat, la taille réelle de l'échantillon dans l'estimation non paramétrique près de x est nh^d au lieu de n , taille effective de l'échantillon dans la régression paramétrique.

La question suivante est de savoir si la correction de [Bartlett,(1937)] est toujours valable dans le cadre du régression. La réponse est oui. On peut montrer que :

$$E\{r_n(m(x))\} = 1 + (nh^d)^{-1} \gamma + o\{nh^{d+4} + (nh^d)^{-1}\},$$

Où

$$\gamma = \mu_2^{-1} (nh^d \mu_1)^2 + \frac{1}{2} \mu_2^{-2} \mu_4 - \frac{1}{3} \mu_2^{-3} \mu_3^2. \quad (4.29)$$

Notez que γ apparaît dans le principal terme d'erreur de couverture dans (4.28). Basé sur (4.28) et choisir

$$h = O\left(n^{-\frac{1}{d+2}}\right) \text{ on a, avec } c_\alpha = \chi_{1,1-\alpha}^2$$

$$\begin{aligned} & p[r_n\{m(x)\} \leq c_\alpha \{1 + \gamma(nh^d)^{-1}\}] \\ &= p[\chi_1^2 \leq c_\alpha \{1 + \gamma(nh^d)^{-1}\}] - (nh^d)^{-1} \gamma c_\alpha^{1/2} \{1 + \gamma(nh^d)^{-1}\}^{1/2} \phi[c_\alpha^{-1/2} \{1 + \gamma(nh^d)^{-1}\}^{1/2}] + O \\ &= p(\chi_1^2 \leq c_\alpha) + (nh^d)^{-1} \gamma z_{1-\frac{\alpha}{2}} \phi(z_{1-\frac{\alpha}{2}}) - (nh^d)^{-1} z_{1-\frac{\alpha}{2}} \phi(z_{1-\frac{\alpha}{2}}) + O\{(nh^d)^{-2}\} \\ &= 1 - \alpha + O(n^{-\frac{4}{d+2}}). \end{aligned}$$

Par conséquent, la probabilité empirique peut être corrigée par [Bartlett,(1937)] dans le contexte actuel de régression non paramétrique. En pratique, le facteur de BARTLETT γ doit être estimé, disons par une cohérente γ . [Chen SX et Qin Y-S(2003)] a donné plus de détails sur la mise en œuvre pratique, voir aussi [Chen.SX(1993)] pour une implémentation dans le cas de l'estimation de densité.

4.4 Simulations et applications

4.4.1 Introduction

Les calculs de cette section sont effectués à l'aide du logiciel R avec l'ajout package `emplik`, disponible sur n'importe quel site du CRAN.

4.4.2 Présentation du package `emplik`

Le **package `emplik`** est développé par plusieurs chercheurs comme [Mai Zhou(2016)] et [Owen, (2001)] ... ect, ce package est comprend plus de 30 fonctions, à utiliser avec le logiciel R.

il peut gérer des données censurées ou non censurées, il a des fonctions pour effectuer une analyse de vraisemblance empirique pour un échantillon ou des modèles de régression. Aussi en utilise ce package pour Tester le rapport (ratio) de vraisemblance empirique pour les moyennes / quantiles / dangers à partir de données éventuellement censurées et / ou tronquées. Et parmi ces fonctions, nous mentionnons :

1. **Bjpoint** : cette fonction calcule la régression censurée de l'estimateur de Buckley-James. Cette fonction est incluse ici principalement pour produire la valeur de l'estimateur qui peut fournir des informations utiles avec la fonction `bjtest()`. Par exemple, vous pouvez vouloir tester une valeur bêta proche de l'estimateur Buckley-James.
2. **Bjtest** : cette fonction utilisé pour tester l'estimateur de [Buckley and James (1979)] par vraisemblance empirique.
3. **WRegtest** : cette fonction testé l'estimateur de régression pondéré par cas par vraisemblance empirique.

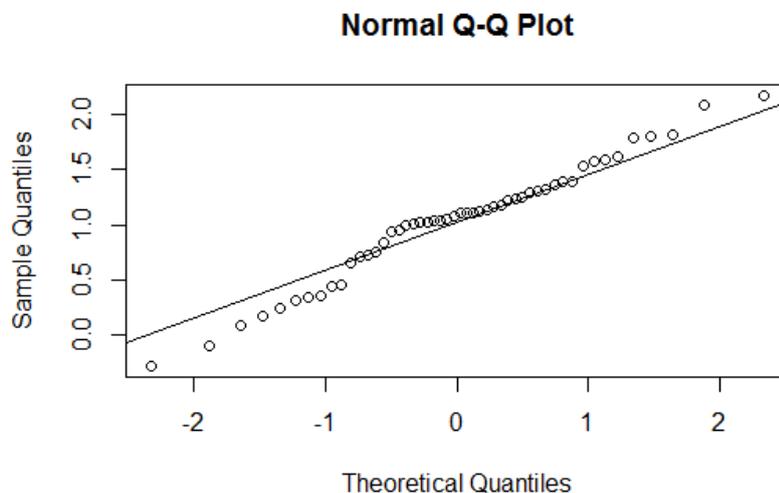


FIG. 4.1 – QQplot de -2ELR pour n=100

4.4.3 Simulation des modèles linéaires

Simulation pour une dimension β

Nous prenons le modèle de régression

$$Y_i = X_i + \varepsilon_i.$$

Où $X_i \sim N(1, 0.5)$, $\varepsilon_i \sim N(0, 0.25)$.

Nous considérons en outre que $C_i \sim N(1, 16)$.

Les tests que nous effectuons sont sur la base des observations de réponses censurées : $\min(y_i, c_i)$, la taille de l'échantillon est toujours de 50.

-2 log rapports de vraisemblance empiriques (ELR) sont calculés pour chaque simulation pour l'hypothèse $H_0 : \beta = 1$ ce qui est vrai. La courbe Q-Q résultante montre un bon ajustement à la distribution chi-carré avec 1 degré de liberté.

La simulation suivante montre les résultats de test de l'estimateur de **Buckley-James** par la vraisemblance empirique.

-2LLR : -2 ratio de log-vraisemblance est une distribution chisq approximative sous H_0 .

logel2 : le log de vraisemblance empirique, sous l'équation d'estimation.

logel : la vraisemblance empirique logarithmique de [Kaplan E.L and Meier(1958)].

prob : probabilités qui maximisent la vraisemblance empirique sous l'équation d'estimation.

Dans même modèle de régression on observe que c'est un modèle de AFT(Accelerated Failure Time).

TAB. 4.1 – Résultats du testes sur un modèle de régression par la méthode de Backley James

-2LLR	2.783049e - 05.
logl2	-54.23121.
logel	-54.23119.
prob	0.020845000.

nous montrons comment :

1. Simuler les données de ce modèle (out1).
2. Calculer l'estimateur (pondéré en fonction du cas) (out2).
3. Et en utilise le EL au cas par cas pour trouver l'intervalle de confiance à 95% pour le paramètre de régression (valeur réelle du paramètre $\beta = 1$) (out3).

TAB. 4.2 – Les résultats du test d'un modèle de régression par l'estimateur de régression pondéré par rapport la vraisemblance empirique.

\$'Est' X ₁	0.9677451
-2LLR	[1]0.1444211
conInt(intervalle de confiance)	[1]0.79527321.1357333

R code

```
library(emplik)
N=50
mu=1
x <- rnorm(N, mean=1, sd=0.5)
eps <- rnorm(N, sd=0.5)
y <- x + eps
cen <- rnorm(N, mean=mu, sd=4)
ycen <- pmin(cen, y)
d <- as.numeric( y <= cen )
qqnorm(y)
qqline(y)
BJjoint(x, y, d, beta0 = NA)
bjttest(y=y, d=d, x=x ,beta=1.18)
#####
out1 <- WRegEst(x=x, y=ycen, delta=d)
out2 <- WRegTest(x=x, y=ycen, delta=d, beta0=1)
```

```

myfunUL <- function(theta, x, y, d){
WRegTest(x=x, y=y, delta=d, beta0=theta)
}
out3 <- findUL(fun=myfunUL, MLE=out1, x=x, y=y, d=d)
list(Est=out1, "-2LLR"=out2$"-2LLR",
ConInt=c(out3$Low, out3$Up) )

```

Simulation pour deux dimensions

Nous illustrons l'analyse EL de l'estimateur de Buckley-James avec Données sur la transplantation cardiaque de Stanford. À la suite de [Miller.RG and Halpern(2007)] nous n'utilisons que 152 cas.

Le modèle de régression AFT spécifique que nous allons ajuster est :

$$Y = \log_{10}(T_i) = \beta_1 + \beta_2 age + \varepsilon_i.$$

1. On test l'estimateur de Buckley-James par la vraisemblance empirique et on obtient les résultats suivant :

TAB. 4.3 – Estimation du paramètres du modèle de régression par la méthode de Backley James.

beta (β_1, β_2)	3.52749311 : -0.01990791
iteration	31
-2LLR	0.2555586

2. Nous montrons ici comment trouver l'intervalle de confiance d'un paramètre unidimensionnel défini par $h(\beta_1, \beta_2)$.

Cette fonction ne doit pas être une fonction linéaire mais le calcul dans cet exemple utilise une fonction linéaire de β_1 et β_2 . Par exemple, la valeur moyenne pour un patient est de 50 ans, d'après le modèle de régression AFT $\beta_1 + 50\beta_2$. Donc nous prenons $h(\beta_1, \beta_2) = \beta_1 + 50\beta_2$ on obtient :

TAB. 4.4 – l'intervalle de confiance du modèle de régression AFT

intervalle de confiance ($\$Lowler, \$Upper$)	(2.395945, 2.669052)
-------------------------------------------------	----------------------

Nos intervalles de confiance sont légèrement plus larges que ceux obtenus par l'intervalle de confiance de Wald utilisant l'estimateur d'erreur standard donné par la fonction `bj()`. Nous rappelons aux lecteurs que l'estimateur d'erreur standard produit par `bj()` n'a pas de justification théorique.

R code

```

library(survival)
data(stanford2)
stanford3 <- stanford2[!is.na(stanford2[,5]),]
stanford5 <- stanford3[(stanford3[,2]>=10),]
BJnoint(x=cbind(1,stanford5$age), y=log10(stanford5$time),
delta=stanford5$status)
bjttest(y=log10(stanford5$time), d=stanford5$status,
x=cbind(1,stanford5$age), beta=c(3.5, -0.02))

#####

library(ELYP)
Bfun <- function(b1, b2){b1+50*b2}
BJLLR <- function(para, dataMat) {
bjttest(y=log10(dataMat[,2]), d=dataMat[,3],
x=cbind(1,dataMat[,4]), beta=para)
}
BJfindL2(NPmle=c(3.5275, -0.0199), ConfInt=c(1, 0.04),
LLRfn=BJLLR, Betafun=Bfun, dataMat=stanford5)
BJfindU2(NPmle=c(3.5275, -0.0199), ConfInt=c(1, 0.04),
LLRfn=BJLLR, Betafun=Bfun, dataMat=stanford5)

```

Simulation pour trois dimensions

Dans un modèle de régression quantile censuré. Nous considérons un ensemble de données sur le cancer du poumon a été analysé par [Ying.Z.Jung et Wei(1995)] utilisant la régression médiane et par [Huang.JM and Xie .H(2007)]. en utilisant une méthode de déviation la moins absolue dans le modèle AFT. Dans cette étude, 121 patients atteints d'un cancer du poumon à petites cellules de stade limité ont été affecté à l'une des deux séquences de traitement différentes A et B, avec 62 patients attribué à A et 59 patients à B. Chaque heure de décès a été observée ou administrée censuré, et la variable de censure ne dépendait pas des covariables traitement et âge.

Désignons la variable d'indicateur de traitement par X_{1i} .

l'âge d'entrée pour le ième patient par X_{2i} .

Où $X_{1i} = 1$ si le patient est dans le groupe B. on a le modèle de régression suivant. Nous supposons ce modèle est AFT modèle (temps de panne accéléré).

$$Y_i = \beta_1 + \beta_2 X_{1i} + \beta_3 X_{2i} + \varepsilon_i$$

Les valeurs estimées des paramètres obtenues à l'aide de l'approche décrite dans la présente doivent être égales à celles de [Huang.JM and Xie .H(2007)], à condition la pondération a été faite de la même manière.

La principale différence réside dans l'inférence sur les paramètres, où la vraisemblance empirique à l'avantage de ne pas être nécessaire pour estimer la variance asymptotique de l'estimateur afin de réaliser des tests d'hypothèses et de construire des régions de confiance.

Les estimations de régression médiane ont été obtenues par [Ying.Z.Jung et Wei(1995)] sont $\hat{\beta}_1 = 2.693, \hat{\beta}_2 = -0.146, \hat{\beta}_3 = 0.001$

[Huang.JM and Xie .H(2007)] n'ont pas toujours traité la plus grande observation Y comme non censuré. Il en résulte des pondérations dont la somme est inférieure à un (la somme des sans la dernière observation est de 0.85). Nous recommandons de traiter les plus grands Y comme non censurée, de sorte que les poids s'additionnent à un pour que l'estimation ne soit pas biaisée puisque les informations provenant de la plus grande observation Y sont ignorées. Traitement le plus grand Y comme non censuré, les estimations médianes de régression sont :

TAB. 4.5 – les estimations médianes de régression

coef	$\hat{\beta}_1 = 2.603, \hat{\beta}_2 = -0.263, \hat{\beta}_3 = 0.0038$
-2LLR	0.1473368

Code R :

```
library(emplik)
data(smallcell)
Z <- log10(smallcell$survival)
dd <- smallcell$indicator
temp <- WKM(x=Z, d=dd, zc=1:121)
KMweight2 <- temp$jump
sum(KMweight2)
norder <- order(smallcell$survival)
temp2 <- rq.wfit(x=cbind(1, smallcell$arm[norder],
smallcell$entry[norder]),
y=Z[norder], weights=KMweight2)
myfunq <- function(y, xmat) {
mytemp1 <- as.vector(y -( 2.6+xmat %*% c(-0.263,0.004)))
mytemp2 <- as.numeric(mytemp1 > 0) +
as.numeric(mytemp1 >= 0) - 1
return(cbind(mytemp2, mytemp2*xmat))
}
XX <- cbind(smallcell$arm, smallcell$entry)
temp3 <- el.cen.EM2(x=Z, d=dd, fun=myfunq,
mu=c(0,0,0), xmat=XX)
```

4.4.4 conclusion

Les simulations précédentes montrent les différents résultats du test des modèles de régression par l'estimateur de **Buckley-James**. Alors nous concluons que l'application de l'estimateur [Buckley and James (1979)] était limitée principalement en raison de son estimation de variance illusoire, et l'utilisation de la méthode de vraisemblance empirique [Owen, (2001)] pour dériver un test (et donc l'intervalle de confiance) basé sur l'estimateur de [Buckley and James (1979)] du coefficient de régression. Et la distribution de chi-carrée est utilisée pour calculer la construction de l'intervalle de confiance.

Enfin, les approximations du chi-carré du rapport de vraisemblance empirique fonctionnent bien.

Chapitre 5

Conclusion générale

Dans ce travail tout d'abords, nous introduison une généralisation sur la statistique inférentielle et quelque méthodes d'estimations et en déduit des tests statistiques. Ensuite, on parlons de notion de vraisemblance empirique d'**Owen** qui est une méthode d'inférence non paramétrique basée sur une fonction de rapport de vraisemblance chargeant que des données. Comme le bootstrap, l'inférence de vraisemblance empirique ne nécessite pas de spécifier une famille de distributions pour les données. Comme les méthodes de vraisemblance paramétriques, la vraisemblance empirique détermine automatiquement la forme des régions de confiance, elle intègre directement des informations secondaires exprimées par des contraintes ou des distributions préalables, elle s'étend à l'échantillonnage biaisé et aux données censurées, et elle possède des propriétés de puissance asymptotique très favorables. La vraisemblance empirique peut être considérée comme un bootstrap qui ne rééchantillonne pas et comme une vraisemblance sans hypothèses paramétriques. on voir des résultats de théorème de **Wilks** [**Wilks,(1937)**] qui permet d'établire la convergence du rapport de vraisemblance empirique vers une loi du \mathcal{X}_2 , permettant aussi de réaliser des tests ou de construire des régions de confiance, d'autre part, on fait une application sur la régression linéaire et l'estimation par la vraisemblance empirique pour ces modèles paramétriques et non paramétriques.

Nous concluons notre travail par une simulation et application de vraisemblance empirique nous utilison le logiciel **R**.

Bibliographie

- [Deville et Sarndal, (1992)] Deville, J. C., & Sarndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American statistical Association*, 87(418), 376-382.
- [DiCiccio et al., (1991)] DiCiccio, T., Hall, P., & Romano, J. (1991). Empirical likelihood is Bartlett-correctable. *the Annals of Statistics*, 19(2), 1053-1061.
- [Hall et Scala, (1990)] Hall, P., & La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *International Statistical Review/Revue Internationale de Statistique*, 109-127.
- [Hartley et Rao, (1968)] Hartley, H. O., & Rao, J. N. K. (1968). A new estimation theory for sample surveys. *Biometrika*, 55(3), 547-557.
- [Owen,(1988)] Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75 237-249.
- [Owen, (1990)] Owen, A. B. (1990). Empirical likelihood ratio confidence intervals. *Ann. Statist*, 18(1), 90-120.
- [Owen, (2001)] Owen, A. B. (2001). *Empirical likelihood*. CRC press.
- [Qin et Lawless, (1994)] Qin, J., & Lawless, J. (1994). Empirical likelihood and general estimating equations. *the Annals of Statistics*, 300-325.
- [Newey et Smith, (2004)] Newey, W. K., & Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, 72(1), 219-255.

- [Turnbull, (1976)] Turnbull, B. (1976), The empirical distribution function with arbitrarily grouped, censored and truncated data. *JRSS B*, 290-295.
- [Hall et al., (1999)] Hall, P., Wolff, R.C.L. and Yao, Q. (1999). Methods for estimating a conditional distribution function. *Journal of the American Statistical Association*, 94, 154-163.
- [McCullagh P.Nelder(1983)] McCullagh P, Nelder JA (1983) Generalized linear models. Chapman & Hall, London.
- [Hansen LP(1982)] Hansen LP (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50 :1029–1054.
- [Tsao M(2004)] Tsao M (2004) Bounds on coverage probabilities of the empirical likelihood ratio confidence regions. *Ann Stat* 32 :1215–1221.
- [Jean-Yves.D(2011)] Jean-Yves.dauxios .2011.statistique inférentielle université de franche-comté.
- [Mai Zhou(2016)] Mai Zhou 2016 ,'Empirical Likelihood Method in Survival Analysis 'University of Kentucky Lexington, USA.
- [Mai Zhou,(2004)] Mai Zhou (Art Owen for `el.test()`. Yifan Yang for some C code.),' Empirical Likelihood Ratio for Censored/Truncated Data.
- [Mme,(2013)] Mme Bouraine née Berdjoudj Louisa ,2013;,"statistique Inférentielle 'uni A. MIRA-Bejaïa.
- [P.Ribereau,(2016)] P.Ribereau 2016,'cours de statistique inférentielles'.
- [Newey, (2004)] Newey W. K., Smith R. J. Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. *Econometrica* 72(1), 219-255, 2004.
- [Wand and all1995] Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*. Chapman and Hall, London.
- [Barndorff, (1984)] Barndorff-Nielsen OE, Cox DR (1984) Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. *J R Stat Soc Ser B* 46 :483–495.

- [Bartlett,(1937)] Bartlett MS (1937) Properties of sufficiency and statistical tests. Proc R Soc A 160 :268–282.
- [Buckley and James (1979)] Buckley JJ, James IR (1979) Linear regression with censored data. Biometrika 66 :429–436.
- [Chen.SX(1993)] Chen SX (1993) On the accuracy of empirical likelihood confidence regions for linear regression model. Ann Inst Stat Math 45 :621–637.
- [Chen et Cui.H-J(2007)] Chen SX, Cui H-J (2007) On the second order properties of empirical likelihood with moment restrictions. J Econom 141 :492–516.
- [Chen SX et Qin Y-S(2003)] Chen SX, Qin Y-S (2003) Coverage accuracy of confidence intervals in nonparametric regression. Acta Math Appl Sin Engl Ser 19 :387–396.
- [Debdeep Pati(2016)] Debdeep Pati,(2016), 'Empirical likelihood.
- [Fan J et Gijbels(1996)] Fan J, Gijbels I (1996) Local polynomial modelling and its applications. Chapman & Hall, London.
- [Härdle W(1990)] Härdle W (1990) Applied Nonparametric Regression. Cambridge University Press, Cambridge.
- [Hayakawa T(1977)] Hayakawa T (1977) The likelihood ratio criterion and the asymptotic expansion of its distribution. AnnInst Stat Math 29 :359–378.
- [Liu RYet Singh K(1981)] Liu, R. Y. & Singh, K. (1992), Moving blocks jackknife and bootstrap capture weak dependence, in 'Exploring the Limits of Bootstrap' J Stat 36 :413–432tstrap'.
- [Nadaraya,(1965)] Nadaraya, E. A. (1965), 'On non-parametric estimates of density functions and regression curves', Theory of Probability and its Applications (Transl of Teorija Verovatnostei i ee Primenenija) 10, 186–190.
- [Wilks,(1937)] Wilks SS (1938) The large-sample distribution of the likelihood ratio for testing composite hypotheses. Ann Math Stat 9 :60–62.

- [Lawley DN(1956)] Lawley DN (1956) A general method for approximating the distribution of likelihood ratio criteria. *Biometrika* :295–303 procedure for Maximum Likelihood estimation in R' .Universidad Nacional de Colombia.
- [Newton MA et Raftery(1994)] Newton, M. A. & Raftery, A. E. (1994), ‘Approximate Bayesian inference with the weighted likelihood bootstrap (disc : P26-48)’, *Journal of the Royal Statistical Society, Series B, Methodological* 56, 3–26.
- [Zahraa salloum] Zahraa Salloum. Maximum de vraisemblance empirique pour la détection de changements dans un modèle avec un nombre faible ou très grand de variables. *Statistiques [math.ST]*. Université de Lyon.
- [Olivier Gaudion] Olivier Gaudion, statistique inférentielle avancée .
- [Miller.RG and Halpern(2007)] Miller, RG. and Halpern, J. (1982). Regression with censored data. *Biometrika*, 69 521-31.
- [Huang.JM and Xie .H(2007)] Huang, J., Ma, S. and Xie, H. (2007). Least absolute deviations estimation for the accelerated failure time model. *Statist. Sinica* 17, 1533-1548.
- [Kaplan E.L and Meier(1958)] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations.*J. Amer. Statist. Assoc.* 53, 282, 457-481.
- [Ying.Z.Jung et Wei(1995)] Ying, Z., Jung, S. H., and Wei, L. J. (1995). Survival analysis with median regression models.*J. Amer. Statist. Assoc.* 90, 429, 178-184.
- [Corcoran.SA and Spady(1995)] Corcoran, S. A., Davison, A. C., & Spady, R. H. (1995). *Reliable inference from empirical likelihoods*. Nuffield College Oxford.
- [Jaime Mosquera Gutiérrez] Jaime Mosquera Gutiérrez . freddy hernandez , maxlogL : A general computational.
- [Jean-Yves.Dauxios2011] Jean-Yves.dauxios .2011.statistique inférentielle université de franche-comté.

Résumé

Les méthodes empiriques de vraisemblance sont largement utilisées dans différents contextes pour construire les régions de confiance des paramètres qui satisfont aux contraintes de moment. Cependant, les régions de confiance du rapport de vraisemblance empirique peuvent être peu précises, en particulier pour les échantillons de petite taille et les situations multidimensionnelles. Dans ce travail nous proposons une nouvelle méthode de probabilité empirique moyenne (MEL). Cette nouvelle méthode construit un nouvel ensemble de pseudo-données en utilisant les valeurs des moyennes d'observation pour définir le rapport de vraisemblance empirique et nous prouvons que ce rapport MEL satisfait au théorème de WILKS. Des simulations avec différents exemples sont données pour évaluer sa performance en matière d'échantillon fini, ce qui montre que les régions de confiance construites par la probabilité empirique moyenne sont beaucoup plus précises que celles des autres méthodes de probabilité empirique.