Université Mohammed Sedik Benyahia-Jijel

Faculté des Sciences Exactes et Informatique

Département de Mathématiques

STATISTIQUE INFERENTIELLE

Cours et TD

Niveau : Troisième Anneé

Spécialité : Mathématiques

Année universitaire 2019/2020

Enseignant : GHERDA Mebrouk

Comme tout polycopié, il y a très probablement des coquilles qui subsistent. Merci de m'en faire part si vous en trouvez.

Le présent cours est en grande partie inspiré des documents suivant :

1-Cours de Statistique Mathématique Modèles, Méthodes, Applications à l'usage des étudiants de DEUG, Licence et Master BORDEAUX M. Nikulin V. Bagdonavi ˇ cius C. Huber V. Nikoulina

2-Support du cours de statistique inférentielle ENSAI 2004-2005 Pierre Druilhet

3-Introduction 'a la statistique inférentielle Didier Concordet Unité de Biométrie Ecole Vétérinaire de Toulouse

4-CTU, Licence de Mathématiques Statistique Inférentielle Jean-Yves DAUXOIS Université de Franche-Comté Année scolaire 2011-2012

Table des matières

Première partie

Modes de Convergence et Approximations

Chapitre 1

MODES DE CONVERGENCES

Soit X une varible alétoire et (X_n) une suite de variables alétoires définies sue le même espace probabilisé $(\Omega,\ \Lambda,\ P)$.

section Convergence en probabilité : $X_n \longrightarrow_P X$

Définition 1.1. *la suite* (X_n) *Converge en probabilité vers* X *si* $\forall \varepsilon > 0\ \lim_{n\longrightarrow\infty} P\left(|X_n - X| > \varepsilon\right) = 0$

La loi faible des grand nombres

Si les variables aléatoires X_n sont deux à deux non covariées, de même loi, d'espérance μ de variance σ^2, alors $\frac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow_P \mu$.

1.1 Convergence en moyenne quadratique : $X_n \longrightarrow_{mq} X$

Définition 1.2. *la suite* (X_n) *Converge en moyenne quadratique vers* X *si* $\lim_{n\longrightarrow\infty} P\left((X_n - X)^2\right) = 0$

2.2 Propriétés : (X_n) Converge en moyenne quadratique vers X si et seulement si $\lim_{n\longrightarrow\infty} E\left(X_n\right) = E\left(X\right)$ et $\lim_{n\longrightarrow\infty} var\left(X_n - X\right) = 0$.

1.2 Convergence presque sûre : $X_n \longrightarrow_{ps} X$.

Définition 1.3. *la suite* (X_n) *Converge presque sû rement vers* X *si* $P\left(w/\lim_{0\to\infty} X_n\left(w\right) = X\left(w\right)\right) = 1$

Loi forte des grands nombres

Si les variables aléatoires X_n sont mutuellement indépendantes de même loi, d'espérance μ, alors $\frac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow_{Ps} \mu$.

1.3 Convergence en loi : $X_n \longrightarrow_l X$

Soit F_n la fonction de répartition de X_n et F celle de X.

Définition 1.4. *la suite* (X_n) *Converge en loi vers* X *si por tout* x *où* F *est continue,* $\lim_{n\to\infty} F_n(x) = F(x)$.

Distance de Kolmogorov entre deux fonction de répartition G_1 et G_2. Elle est définit par $\Delta(G_1, G_2) = \sup_{x\in\mathbb{R}} |G_1(x) - G_2(x)|$.

Proprités de la convergence en loi -Si $\lim_{n\to\infty} \Delta(F_n, F) = 0$ alors $X_n \longrightarrow_l X$.

-Si F est continue alors : $X_n \longrightarrow_l X$ si et seulement si $\lim_{n\to\infty} \Delta(F_n, F) = 0$.

-Si X_n et X sot des variables aléatoires à valeurs dans $\mathbb{N}$ alors $X_n \longrightarrow_l X$ si et seulement si : $\forall k \in \mathbb{N}$, $\lim_{n\to\infty} P(X_n = k) = P(X = k)$.

-Soit a et b deux réels. Si $X_n \longrightarrow_l X$ alors $aX_n + b \longrightarrow_l aX + b$.

Théorème limite centrale

Si les variables aléatoires X_n sont mutuellement indépendantes de même loi, d'espérance μ et d'écart-type σ diff érent de 0 alors :

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right) \longrightarrow_l X$$

où X est une variable aléatoire de loi de Laplace-Gauss centré e réduite.

1.4 Liens entre les différents type de convergence.

Ils se résume de la façon suivante :

$X_n \longrightarrow_{Ps} X \Longrightarrow X_n \longrightarrow_P X \Longrightarrow X_n \longrightarrow_l X$

$X_n \longrightarrow_{mq} X \Longrightarrow X_n \longrightarrow_P X \Longrightarrow X_n \longrightarrow_l X$

La convergence en loi est la seule qui ne fait intervenir que les lois des variables aléatoires.

Dans le cas où X est une variables aléatoire égale à a ou presque sûrement égale à a :

$$X_n \longrightarrow_P X \Leftrightarrow X_n \longrightarrow_l X$$

Chapitre 2

APPROXIMATION

2.1 Approximation d'une loi binomiale par une loi de Poisson

Considéronons une loi binomiale de paramètres n et p ; Si n et grand et p assez petit, la loi de Poisson est une bonne approximation de la loi binomiale à condition que le produit np reste fini, et dans ce cas la loi binomiale $B(n,p)$ tend vers la loi de Poisson $P(\lambda = np)$

En pratique, nous utiliserons l'approximation de la loi binomiale par la loi de Poisson dans les conditions suivantes :

a) $n > 50$, $p < 0.1$ $n > 50$, $p > 0.9$ car alors $q < 0.1$ ce qui nous ramène au cas précédent compte tenu du rôle symétrique que jouent p et q dans le cas d'une loi binomiale.

2.2 Approximation d'une loi binomiale par une loi normale

Soit une variable aléatoire discrète X suivant une loi binomiale $B(n,p)$ telle que : $P(X = k) = C_n^k p^k q^{n-k}$.

Si n est suffisamment grand et p pas trop proche de 0 ni de 1 avec $np \geqslant 5$ et $nq \geqslant 5$ alors la loi normale de paramètres $m = np$ et $\sigma = \sqrt{npq}$ constitue une bonne approximation de la loi binomiale.

Remarque 1. *Il y a nécessité de remplacer $P(X = k)$ par $P(k - 0.5 < X < k + 0.5)$ (correction de continuité) car dans le cas d'une loi discrète les probabilités de type $P(X = k)$ sont nulles.*

Les conditions pratique de l'approximation sont :

$n \geqslant 30$ $p \in [0.1, 0.9]$ car sinon la loi de Poisson réalise une meilleure approximation

$np \geqslant 5, nq \geqslant 5$.

2.3 Approximation de loi de Poisson par une loi normale

Soit une variable aléatoire discrète X suivant la loi de Poisson $P(\lambda)$ telle que : $P(X = k) = \frac{\lambda^k}{k!}\varrho^{-\lambda}$

Si λ est suffisamment grand, la loi normale de paramètres $m = \lambda$ et $\sigma = \sqrt{\lambda}$ constitue une bonne approximation de la loi de Poisson..la correction de continuité citée ci-dessus s'applique ici aussi et $P(X = k) = P(k - 0.5 < X < k + 0.5)$

APPLICATION

On sait que la probabilité qu'une personne soit allergique à un certain médicament est égale à $(10)^{-3}$, On s'int éresse à un échantillon de 1000 personnes. On appelle X la variable aléatoire dont la valeur est le nombre de personne allergique dans l'échantillon.

1-Déterminer, on la justifiant, la loiu de probabilité de X.

2-En utilisant une approximation que l'on justifiera, calculer les probabilit és des événements suivants :

a-Il y a exactement deux personnes allergique dans l'echantillon

b- Il y a au moins deux personnes allergiques dans l'échantillon.

Que peut-on dire si 30% de la population d'où provient cet é chantillon sont allergiques à ce médicament.

2.4 Lois dérivées de la loi normale

2.4.1 Loi du khi carré : χ^2_n

Construction de la loi de distribution :

Répéter les opérations suivantes pour un grand nombre de points i :

• Tirer au hasard des valeurs de plusieurs (n) lois N(0, 1).

• Mettre chacune de ces valeurs au carré.

• La valeur du point i est la somme de ces valeurs au carré.

Continuer avec le point suivant

La distribution des valeurs obtenues obéit à une loi $\chi^2_n : z_1^2 + z_2^2 + ... + z_n^2 = \chi^2_n$

En particulier, $\chi^2_1 = [N(0,1)]^2$.

2.4.2 Limite de la loi du khi-carré

Quand n devient grand [en pratique, quand $n \geq 30$], la loi du χ^2 tend vers une nouvelle loi normale $N(m, s)$ de moyenne $m = n$ et d'é cart type $\sigma = \sqrt{2n}$. Il suffit de centrer et réduire pour passer de la loi du χ^2 à une loi normale centrée ré duite z.

Par conséquent, $\dfrac{\chi^2_{n,\alpha} - n}{\sqrt{2n}} = z_\alpha$ ou inversement, $\chi^2_{n,\alpha} = n + z_\alpha\sqrt{2n}$

pour toute valeur de probabilité α.

Remarque 2. *L'analyse des données biologiques utilise abondamment la loi du* χ^2

2.4.3 Loi de Fisher-Snedecor :

Définition 2.1. $F_{(v_1,v_2)} = \dfrac{\chi^2_{v_1}/v_1}{\chi^2_{v_2}/v_2}$

Le rapport de deux variables aléatoires distribuées comme khi-carr é, chacune divisée par ses degrés de liberté, est une variable aléatoire distribuée comme F.

Il existe autant de courbes de densité de probabilité de F que de

combinaisons possibles de n_1 et n_2.

Applications

- Test F de rapport de variances.
- Analyse de variance.

2.5 Loi de Student : t(n)

Loi décrite en 1908 par William Sealy Gosset sous le pseudonyme "Student". Le premier article de Student, publié en 1907, avait établi que la distribution des dénombrements de cellules dans les carrés d'un hémacytomètre suivaient la loi de Poisson (répartition aléatoire).

Deux définitions équivalentes de la loi de t :

1) $t(n) = \dfrac{Z}{\sqrt{\chi^2/n}}$

2) $t(n) = F_{(v_1,v_2)}$ lorsque $n_1 = 1$. Le nombre de degr és de liberté de la loi de t est alors $n = n_2$.

Applications

- Estimation des paramètres d'une population à partir de renseignements portant sur un échantillon.
- Test de comparaison des moyennes.
- Calcul de la probabilité d'observer un écart donné à la moyenne, en particulier dans le cas de petits échantillons :

Pour un écart observé , la probabilité d'une telle observation x_i est donnée par la variable aléatoire $t = (x_i - \overline{x})/s_x$.

Il existe autant de courbes de densité de probabilité de t que de valeurs possibles de n. Voir la table de la distribution de t

Thérorème de la limite centrée Soit (X_n) une suite de variables aléatoires mutuellement indé pendantes de même loi de moyenne μ et d'écart-type σ et soit $\overline{X} = \dfrac{1}{n}\left(\sum_{i=1}^{n} X_i\right)$. Pour $n \geq 30$, la variable aléatoire $\overline{X}$ suit, approximativement, la loi normale de moyenne μ et d'écart-type $\frac{\sigma}{\sqrt{n}}$.

Deuxième partie

INFERENCE STATISTIQUE

Introduction

La science des statistiques comporte 2 aspects :

1. Les statistiques descriptives qui consistent à synthétiser, résumer, structurer l'information contenue dans les données.

2. La statistique mathématique qui consiste à traduire en langage mathématique la démarche d'inférence statistique.

L'inférence statistique :

L'inférence statistique est le fait de fournir à partir d'une propri été observée dans des cas particuliers des caractéristiques de la propriété en général.

Chapitre 3

Le modèle statistique

3.1 Notions et définitions

3.1.1 Le modèle statistique

Un modèle statistique est un objet mathématique associé à l'observation de données issues d'un phénomène aléatoire.

Une expérience statistique consiste a recueillir une observation x d'un elément aléatoire X, à valeurs dans un espace χ et dont on ne connait pas exactement la loi de probabilité P. Des consid érations de modélisation du phénomène observé amè nent à admettre que P appartient a une famille P de lois de probabilit é possibles.

Définition 3.1. *Le modèle statistique (ou la structure statistique) associé a cette expérience est le triplet* $(\chi; A; P)$*, où :*

X est l'espace des observations, ensemble de toutes les observations possibles.

A est la tribu des evènements observables associée.

P est une famille de lois de probabilit es possibles défnie sur A.

L'int erêt de cette notion de modèle statistique est qu'elle permet de traiter avec le même formalisme tous les types d'observations possibles.

On dit que le modèle est discret quand X est fini ou dénombrable. Dans ce cas, la tribu A est l'ensemble des parties de X : $A = P(X)$.

On dit que le modèle est continu quand X$\subset$ IR^p et $\forall$ P $\in P$, P admet une densité (par rapport à la mesure de Lebesgue) dans IR^p. Dans ce cas, A est la tribu des boréliens de X (tribu engendrée par les ouverts de X) : $A = B(X)$.

On peut aussi envisager des modèles ni continus ni discrets, par exemple si l'observation

a certains eléments continus et d'autres discrets. X et A sont alors plus complexes.

Le cas le plus fréquent, est celui o u l'élément aléatoire observé est constitué de variables aléatoires indépendantes et de même loi ($i.i.d.$) : $X = (X_1, ..., X_n)$, où les X$_i$ sont $i.i.d.$ On dit que l'on a alors un modèle d'échantillon.

Dans ce cas, par convention, si on note $(X; A; P)$ le modèle correspondant a un echantillon de taille 1, on notera $(X; A; P)^n$ le mod éle correspondant a un echantillon de taille n.

Exemple 3.2. *l'expérience consiste à recueillir les durées de vie, supposé es indépendantes et de même loi exponentielle, de n ampoules electriques. L'observation est de la forme $x = (x_1, ..., x_n)$, où les x_i sont des réalisations de variables al eatoires X_i indé pendantes et de même loi exponentielle de paramètre inconnu.*

Pour tout i, $x_i \in IR+$, donc l'espace des observations est $X = IR_+^n$. Alors la tribuassociée est $A = B(IR_+^n)$. Le modè le est continu. Comme on admet que la loi est exponentielle mais que son paramètre est inconnu, l'ensemble des lois de probabilités possibles pour chaque X_i est $exp(\lambda)$; $\lambda \in IR_+$, Comme les X_i sont ind ependantes, la loi de probabilité du vecteur $(X_1, ..., X_n)$ est la loi produit $P = \{exp(\lambda)^{xn}; \lambda \in IR_+\}$, ensemble des lois de probabilité des vecteurs al éatoires de taille n dont les composantes sont indépendantes et de même loi exponentielle de paramètre inconnu.

Finalement, le modèle statistique associé est :

$(IR_+^n; B(IR_+^n); exp(\lambda)^{xn}; \lambda \in IR_+)$ qu'on peut aussi écrire : $(IR_+^n; B(IR_+^n); exp(\lambda); \lambda \in IR_+)^n$

Modèle paramétrique ou non paramétrique

Un modèle param etrique est un modèle où l'on suppose que le type de loi de X est connu, mais qu'il dépend d'un paramètre inconnu, de dimension d. Alors, la famille de lois de probabilité possibles pour X peut s'écrire $P = \left\{p_\theta; \theta \in IR^d\right\}$.

Un modèle non paramétrique est un modèle où P ne peut pas se mettre sous la forme ci-dessus. Par exemple, P peut être :

l'ensemble des lois de probabilité continues sur IR,

l'ensemble des lois de probabilit e sur IR symétriques par rapport a l'origine, etc...

Dans ce cadre, il est possible de déterminer des estimations, des intervalles de confiance, d'effectuer des tests d'hypothèses. Mais les objets sur lesquels portent ces procédures statistiques ne sont plus des paramètres de lois de probabilité. On peut vouloir estimer des quantités réelles comme l'espérance et la variance des observations. On peut aussi vouloir estimer des fonctions, comme la fonction de répartition et la densité des observations.

3.1.2 Fonction de vraisemblance

Dans un modèle paramétrique, la fonction de vraisemblance joue un rôle fondamental.

Pour un modèle d'échantillon discret, l'élément aléatoire observé est $X = (X_1, ..., X_n)$, où les X_i sont ind ependantes et de même loi discrète. Alors la fonction de vraisemblance est :

$L(\theta; x_1, ..., x_n) = P(X_1 = x_1, ..., X_n = x_n; \theta) = \Pi_{i=1}^{n} P(X_i = x_i; \theta)$

Pour un modèle d'échantillon continu, l'élément aléatoire observé est $X = (X_1, ..., X_n)$, où les X_i sont ind ependantes et de même loi continue. Alors la fonction de vraisemblance est :

$L(\theta; x_1, ..., x_n) = f_{(X_1,...,X_n)}(x_1, ..., x_n; \theta) = \Pi_{i=1}^{n} f_{X_i}(xi; \theta)$

Pour définir une fonction de vraisemblance valable dans n'importe quel modèle statistique, pas forcément d'échantillon et pas forcément discret ou continu, il faut utiliser des notions de théorie de la mesure.

Une mesure sur $(\chi; A)$ est σ-finie si et seulement si il existe une suite $\{A_n\}_{n\geq 1}$d'évènements de A telle que $\cup_{n\geq 1} A_n = \chi$ et $\forall n \geq 1$, $(A_n) < +\infty$ (χ est une union dénombrable d'évènements de mesure finie).

P est absolument continue par rapport a σ si et seulement si : $\forall A \in A$; $\sigma(A) = 0) \Rightarrow P(A) = 0$:

On considère un modèle paramétrique quelconque $(X; A; \{P_\theta\}; \theta \in \Theta)$. On supposera qu'il existe une mesure $\sigma-$finie u sur $(X; A)$ telle que $\forall\theta \in \Theta$, la loi de P est absolument continue par rapport a u (on dit que u est la mesure dominante du modèle). Alors le théorème de Radon-Nikodyn assure que P admet une densitéu par rapport a . Cette densité est appelée fonction de vraisemblance du modèle.

Définition 3.3. *La fonction de vraisemblance du modèle* $(X; A; \{P_\theta\}; \theta \in \Theta)$ *est la fonction de définie par :*

$\forall A \in A;\ P_\theta(A) = P(X \in A;\theta) = \int_A L(\theta;x)du(x)$:

Plus généralement, pour toute fonction φ intégrable, on a : $E[\varphi(X)] =_\chi \varphi(x)L(\theta;x)du(x)$

Cas des modèles continus. Si X est un vecteur alèatoire admettant une densitè $f_X(x;\theta)$ (par rapport a la mesure de Lebesgue), on sait bien que $P(X \in A;\theta) = \int_A f_X(x;\theta)dx$.

Donc la mesure dominante est la mesure de Lebesgue et la fonction de vraisemblance est $L(\theta;x) = f_X(x;\theta)$.

Cas des modèles discrets. Si X est un vecteur aléatoire de loi discrète, définie par les probabilités elémentaires $P(X = x;\theta)$, alors :$P(X \in A;\theta) = \sum_{x\in A} P(X = x;\theta) = \int_A P(X = x;\theta)du_d(x)$

où u_d est la mesure de dénombrement sur X : $u_d(A) = card(A)$ et $\int_A f(x)du_d(x) = \sum_{x\in A} f(x)$. Donc la fonction de vraisemblance est bien $L(\theta;x) = P(X = x;\theta)$.

3.1.3 Statistique

*Dénition*3 Dans un modèle statistique $(X;A;P)$, une statistique est une application mesurable t de $(X;A)$ dans un espace Y muni d'une tribu B.

Rappel : une application t de $(X;A)$ dans $(Y;B)$ est mesurable si et seulement si $\forall B \in B$, l'évènement $t^{-1}(B) = [t(X) \in B]$ est dans A, c'est- a-dire $\forall\ A;\ t(A) = B \Rightarrow A \in A$.

Concrètement, cela signifie que l'on peut calculer la probabiliti de tout evènement de la forme $[t(X) \in B]$, donc t ne doit pas dépendre de paramètres inconnus.

Puisque x est une réalisation de l'élément aléatoire X, $t(x)$ est une réalisation de l'élément aléatoire $T = t(X)$.

Définition 3.4. *La loi de probabilité P_T de T est appelée loi image par t et le modèle $(Y;B;\{P_T;P \in P\})$ est le modèle image par t de $(X;A;P)$.*

Exemple des ampoules. Le modèle est $(IR+;B(IR+)\,;\{exp(\lambda);\lambda \in IR+\}^n$. $X = (X_1,...,X_n)$,

où les X_i sont des variables aléatoires indépendantes et de même loi $exp(\lambda)$. On sait qu'alors $T = \sum_{i=1}^n X_i$ est de loi gamma $G(n;\lambda)$. Donc la loi image par $t(x) = \sum_{i=1}^n x_i$ est la loi $G(n;\lambda)$ et le modèle image est le mod ele $(IR+;B(IR+)\,;\{G(n;\lambda);\lambda \in IR+\}$.

Remarquons que le modèle image est de dimension 1 alors que le modèle initial etait de dimension n. Autrement dit, la statistique $t(x) = \sum_{i=1}^{n} x_i$ est un résumé des observations $x = (x_1, ..., x_n)$.

3.2 Modèle statistique. Fonction de vraisemblance

Soient (W ;A ;P) un espace probabilisé et (Rn ;Bn) un espace borélien.

3.2.1 Modèle statistique

Soient $(W; A; P)$ un espace probabilisé et $(\mathbb{R}n^n; B_n)$ un espace borélien. Une application

$$X = X(w) = (X_1(w); X_2(w); ::::; X_n(w))^T : \Omega \longrightarrow \mathbb{R}^n$$

de l'ensemble $\Omega = \{w\}$ de tous les événements élémentaires dans $\mathbb{R}^n$ est appelée un vecteur aléatoire si

$$X^{-1}(B) \in A; pour\ tout\ B \in B^n : (1)$$

$D\acute{e}finition2$. Soit P_X une mesure sur $(\mathbb{R}^n; B_n)$, déterminée par la formule suivante :

$$P_X(B) = P\{w : X(w) \in B\} = P\left\{X^{-1}(B)\right\} = P\{X \in B\} : (2)$$

La mesure P_X, déterminée sur la $\sigma - alg\grave{e}bre$ bor élienne B_n par l'égalité (2), s'appelle la distribution (la répartition) de X dans $\mathbb{R}^n$.

Supposons que la distribution P_X de X appartienne à une famille $P = \{P_\theta; \theta \in \Theta\}$:

Définition 3.5. *On appelle modèle statistique le triplet* $(\mathbb{R}^n; B_n; P)$. *Souvent au lieu de* $\mathbb{R}^n; B_n; P)$ *on écrit* $(\mathbb{R}^n; B_n; P_\theta; \theta \in \Theta)$ *pour indiquer l'espace des paramètres* Θ.

Définition 3.6. *Un modèle* $(\mathbb{R}^n; B_n; P_\theta; \theta \in \Theta)$ *est dit dominé par une mesure* σ*-finie* μ *dans* $\mathbb{R}^n$, *si la famille* $P = \{P_\theta; \theta \in \Theta\}$ *est absolumemt continue par rapport à* $\mu : P_\theta \ll \mu\ \forall\theta \in \Theta$.

Autrement dit, le modèle $(\mathbb{R}^n; B_n; P_\theta; \theta \in \Theta)$ *est dominé par* μ, *si pour tout* $\theta \in \Theta$ *il existe une fonction non négative* B_n*- mesurable* $p(x; \theta)$ *telle que*

$$P_\theta(B) = \int_B p(x; \theta) d\mu(x)$$

pour tout $B \in B_n$. La fonction $p(x;\theta) = p_\theta(x)$ est appelée la dérivée de Radon-Nikodym de la mesure P_θ par rapport à la σ-mesure μ, et on note souvent

$$p(x;\theta) = \frac{dP_\theta(x)}{d\mu} \text{ ou } dP_\theta(x) = p(x;\theta)d\mu(x)$$

Considérons le modèle : $H_0 : X \sim p(x;\theta); \theta \in \Theta; x \in \mathbb{R}^n$;

d'après lequel la densité d'un vecteur aléatoire $X = X(w)$ de dimension n appartient à une famille des densités

$$\{P(x;\theta); \theta \in \Theta\}; x = (x_1; x_2; ...; x_n)^T \in \mathbb{R}^n :$$

Remarque 3. *. Si Θ est un ensemble de $\mathbb{R}^m$, on dit que le modèle H_0 est paramétrique, sinon le modèle H_0 s'appelle non paramétrique.*

Définition 3.7. *La variable aléatoire*

$$L(\theta) = L(X;\theta) = p(X;\theta); \theta \in \Theta \subset \mathbb{R}^m; (3)$$

est appelée la fonction de vraisemblance de X.

Remarque 4. *On appelle $L(\theta)$ ainsi car la fonction de vraisemblence $L(\theta)$, sachant la réalisation x du vecteur alé atoire X, nous permet de comparer les paramètres $\theta_1 \in \Theta$ et $\theta_2 \in \Theta$.*

Si

$$L(\theta_1) > L(\theta_1);$$

il est plus probable que $X = x$ pour $\theta = \theta_1$.

Avec cette optique il est très naturel de considérer

$$\widehat{\theta}_n = \widehat{\theta}_n(X) = arg_\theta max L(\theta); c - \grave{a} - d.L(\widehat{\theta}_n) = max L(\theta)\theta \in \Theta;$$

comme un estimateur de θ, appelé l'estimateur de maximum de vraisemblance.

3.2.2 Statistique. Échantillon. Loi empirique.

Définition 3.8. *Soit $T = T(x)$ une application de $(\mathbb{R}^n; B_n)$ dans un espace E muni d'une $\sigma-$algèbre boré lienne Ξ ; $T : \mathbb{R}^n \longrightarrow E$. On dit que T est une application borélienne si pour tout ensemble borélien B de l'espace $(E; \Xi)$; $\in \Xi; T^{-1}(B)$ est un ensemble borélien dans $(\mathbb{R}^n; B_n)$, c-à-d $\{x : T(x) \in B\} = T^{-1}(B) \in B_n$; pour tout $B \in \Xi$:*

Définition 3.9. *Soient $X = X(w)$ un vecteur aléatoire sur $(\Omega; A; P)$, $X : \Omega \longrightarrow \mathbb{R}^n$, et $T(x)$, une application borélienne de $\mathbb{R}^n$ dans un espace mesurable $(E; \Xi)$, $T : \sigma_n \longrightarrow E$*

Dans ce cas on dit que $T(X) = T(X(w))$ est une statistique et l'application T elle-même s'appelle une fonction de décision.

En d'autres termes n'importe quelle transformation du vecteur d'observations X ne dépendant pas du paramètre inconnu θ est une statistique.

Définition 3.10. *Soit $X(w) = (X_1(w); X_2(w); :::; X_n(w))^T$ un vecteur aléatoire. Considérons un modèle H_0 d'après lequel les variables aléatoires $X_1; ...; X_n$ sont indépendantes et suivent la même loi. Dans ce cas on dit que X est un é chantillon de taille n et on écrit $\mathbf{X}$ au lieu de X.*

Remarque 5. *. Soit $X = (X_1; ...; X_n)^T$ un échantillon de taille n, $X : \Omega \longrightarrow \mathbb{R}^n$. Considérons un modèle paramétrique $H_0 : \mathbf{X} \sim p(x; \theta)$; $\theta \in \Theta$; $x \in \mathbb{R}^n$:*

Soit $f(x_i; \theta)$ la densité de $X_i : \mathbb{R}^1 \times \Theta \longrightarrow \mathbb{R}^1$. Dans ce cas pour tout $x \in \mathbb{R}^n$

$$p(x; \theta) = \Pi_{i=1}^n f(x_i; \theta); \theta \in \Theta;$$

et la fonction de vraisemblance de l'échantillon $\mathbf{X}$ est

$$L(\theta) = p(\mathbf{X}; \theta) = \Pi_{i=1}^n f(X_i; \theta); \theta \in \Theta :$$

Exemple 3.11. *Statistiques d'ordre. Vecteur des rangs. Soit $\mathbf{X} = (X_1; ...; X_n)^T$ un échantillon, $\mathbf{X} \in X \subset \mathbb{R}^n$. A toute réalisation $x = (x_1; ...; x_n)^T \in \mathbf{X}$ de X on peut associer le vecteur $x^{(n)} = (x_{(1)}; ...; x_{(n)})^T$ obtenu en ordonnant les x_i par ordre croissant $x_{(1)} \leq x_{(2)\leq} ... \leq x_{(n)}$:*

La statistique correspondante $X_{(n)} = (X_{(1)}; ...; X_{(n)})^T$ est appel ée le vecteur des statistiques d'ordre et $X_{(i)}$ est la $i-$ème statistique d'ordre dans $A \subset \mathbb{R}^n$:

$$A = \left\{x = (x_1; ...; x_n)^T \in \mathbb{R}^n : x_{(1)} \leq x_{(2)\leq} ... \leq x_{(n)}\right\}$$

Si de plus on associe à $\mathbf{X}$ *le vecteur* $R = (R_1; :::; R_n)^T$ *des rangs* R_i *des* X_i *(i = 1 ;... ;n), dans* $X^{(n)}$, *avec* $R_i = \sum_{j=1}^{n} 1_{\{X_j \leq \cdot X_i\}}$ *et on suppose que* $P\left\{X_{(1)} < X_{(2)} < ... < X_{(n)}\right\} = 1;$

alors dans ce cas la correspondance entre $\mathbf{X}$ *et la statistique* $(X^{(n)}; R)$ *est bijective. En général,* R *est à valeurs dans l'ensemble* σ_n *des permutations des n premier entiers, avec r épétition car il peut y avoir des ex aequo parmi les composantes de* $\mathbf{X}$. *Cependant, si la probabilité pour qu'au moins deux des composants de* $\mathbf{X}$ *soient égales est nulle,* R *est à valeurs dans l'ensemble* σ_n *des permutations de* $\{1; ...n\}$ *: Cela se produit en particulier si la loi de* $\mathbf{X}$ *admet une densité* $p(x)$ *par rapport à la mesure de Lebesgue sur* σ_n. *Parfois, au lieu de* $X^{(n)}$ *on utilise le signe* $X^{(.)}$.

La statistique $J_n = (J_1; :::; J_n)^T$, *oú* $J_k = \sum_{j=1}^{n} 1_{\{R_J + k\};}\ k = 1; 2; ...; h;$ *est connue comme le vecteur des antirangs.*

Soit $F(x) = P\{X_1 \leq x\}\cdot$ *la fonction de répartition de* X_1. *Dans ce cas on a, par exemple,*

$$P\{X(n) \leq x\} = F^n(x); \quad P\{X_1 \leq x\} = 1 - [1 - F(x)]^n;$$

$$P\{X_r \leq x\} = n! \sum_{k=r}^{n} \frac{F^k(x)\,(1 - F(x))^{n-k}}{k!(n-k)!}$$

puisque

$$P\left\{X_{(r)} \leq x < X_{(r+1)}\right\} = \frac{n!}{r!(n\text{-}r)!}(F(x))^r[1 - F(x)]^{n-r} :$$

Donc si la loi F *de* X_1 *est absolument continue,c-à-d. s'il existe la densité* $f(x)$ *telle que* $F(x) = \int_{-\infty}^{x} xf(u)du;\ x \in \mathbb{R};$

alors la loi de $X_{(r)}$ *est absolument continue aussi et sa densité est donnée par la formule*

$$f_{X(r)}(x) = \frac{n!}{(r-1)!(n\text{-}r)!}(F(x)^{r-1}[1 - F(x)]^{n-r}; r = 1; ...; n :$$

Exemple 3.12. *Soit* $X = (X_1; ...; X_n)^T$ *un échantillon. Dans ce cas les statistiques*

$T_1 = \sum_{i=1}^{n} X_i;\ T_2 = \sum_{i=1}^{n} X_i^2;\ \overline{X}_n = \frac{T_1}{n};\ S_n^2 = \sum_{i=1}^{n} \left(X_i - \overline{X}_n\right)^2;\ T_3 = X_{(1)};\ T_4 = X_{(n)};\ T_3 = X_{(n)} - X_{(1)};$ $V_n = \frac{S_n}{\overline{X}_n}$

donnent des exemples simples de statistiques scalaires, tandis que $T = (T_1; T_2)^T$ *et* $U = (\overline{X}_n\ ;\ S_n^2)^T$

sont deux statistiques vectorielles de dimension deux. La statistique V_n *s'appelle le coefficient de variabilité,* T_5 *est l'étendu de l' échantillon,* T_3 *et* T_4 *sont les statistiques extrémales.*

Exemple 3.13. *La loi empirique. Soit* $X = (X_1; ...; X_n)^T$ *un é chantillon,* $F(x) = P\{X_i \leq x\}$ *est la fonction de ré partition de* X_i*. Ayant la réalisation* $x = (x_1; :::; x_n)^T$ *de la statistique* $\mathbf{X} = (X_1; ...; X_n)^T$ *, nous pouvons construire la fonction*

$$F_n(x) = F_n(x; x_1; ...; x_n) = \frac{1}{n}\sum_{i=1}^{n} 1_{]-\infty;x]}(x_{(i)}) = \frac{1}{n}\sum_{i=1}^{n} 1_{]-\infty;x]}(x_i);\ x \in n\mathbb{R};$$

dont la valeur$F_n(x)$ *en n'importe quel point* x *;* $x \in \mathbb{R}$*;, représente la réalisation de la statistique*

$$\mathbf{F}_n(x) = \mathbf{F}_{nn}(x; x_1; ...; x_n) = \frac{1}{n}\sum_{i=1}^{n} 1_{]-\infty;x]}(X_i) = \frac{1}{n}\sum_{i=1}^{n} 1_{]-\infty;x]}(X_i)$$

calculée au point choisi x*.*

Par construction, la fonction $F_n(x)$ *;* $x \in \mathbb{R}$*, a toutes les propriétés d'une fonction de répartition, car elle est croissante de* 0 *à* 1 *et continue à droite, et pour cette raison nous pouvons introduire une variable aléatoire discrète, disons* X*, dont la loi conditionnelle, conditionnée par* $\mathbf{X} = x$*, est donnée par la fonction* $F_n(x)$*, c'est-à-dire*

$$F_n(x) = P\{\mathbf{X} \leq x/X = x\} \cdot = P\{X \leq x/X_n + x_1; ...; X_n = x_n\};\quad x \in \mathbb{R}.$$

et par conséquent

$$\mathbf{F}_n(x) = P\{\mathbf{X} \leq x/\mathbf{X} = x\} : x \in \mathbb{R}$$

Cette formule détermine la fonction de répartition aléatoire et, par tradition, on l'appelle la fonction de répartition empirique. Par conséquent, la loi conditionnelle de la variable aléatoire X*, conditionnée par* $\mathbf{X}$*, s'appelle la loi empirique. La loi empirique est la loi discrète de* X *telle que* $P\{\mathbf{X} \leq X_i/\mathbf{X}\} = \frac{1}{n}$ *pour tout* $i = 1; 2; ...; n$ *et* $F_n(x)$ *est la fonction de répartition de cette loi.*

Les statistiques $\overline{X_n}$ *et* S_n^2 *représentent la moyenne et la variance de la loi empirique. Par définition la statistique* $\widehat{x}_P = X_{([nP]+1)}$ *représente* $P-$ *quantile de la loi empirique, et par conséquant,* $\widehat{x}_{0.5} = X_{([\frac{n}{2}]+1)}$ *est la médiane de la loi empirique.*

Remarque 6. *. Soit* $X = (X_1; ...; X_n)^T$ *un vecteur aléatoire,* $X \in \mathbb{R}^n$*, dont la densité est* $p_X(x)$*,* $x = (x_1; ...; x_n)^T$*.*

Considérons une statistique $Y = f(X)$*, où* $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ *est une application dérivable.*

Notons $y = f(x)$*; c-à-d.* $y = (y_1; :::; y_n)^T$ *; où* $y_j = f_j(x)$*;* $x \in \mathbb{R}^n$ *:*

Le Jacobien de f est une application donnée par la formule :

$$Df : \mathbb{R}^n \longrightarrow \in \mathbb{R}; Df(x) = \left\| \frac{\partial f_j(x)}{\partial x_i} \right\|$$

c-à-d. $Df(x)$ est le déterminant de la matrice Jacobienne.

Si $Df(x) \neq 0$ au voisinage d'un point x, $x \in \mathbb{R}^n$, dans ce cas $f^{-1}(y)$ existe au voisinage du point $y = f(x)$ avec

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

ou

$$Df^{-1}(y))Df(x) = 1; y = f(x)$$

Si f^{-1} existe, alors d'après une propriété connue en analyse, pour toute fonction integrable φ de $\mathbb{R}^n$ on a

$$\int_A \varphi(y)dy = \int_{f^{-1}(A)} (f(x)) \left|Df(x)\right| dx \qquad (2)$$

3.3 ECHANTILLONAGE

Population de taille finie

Soit E un ensemble, que nous appellerons population mère, contenant un nombre fini N d'éléments. Nous supposerons que l'on veut étudier une propriété X de cette population. L'objectif serait donc de déterminer les principales caractéristiques de

la loi de X.

S'il est possible d'effectuer un recensement, c'est-à-dire interroger ou inspecter

tous les éléments de E, les caractéristiques de X seront parfaitement connues. Si on écrit $E = \{e_1, ..., e_N\}$ et si X est une propriété mesurable, on observe alors $(x_1, ..., x_N)$ l'ensemble des valeurs prises par X correspondant aux éléments de E. Remarquons que ce sont des valeurs déterministes. Dans ce cas précis on peut par exemple calculer les vraies moyenne μ et variance σ^2 de X :

$\mu = \frac{1}{N} \sum_{j=1}^{N} .x_j \; ; \; \sigma^2 = \frac{1}{N} \sum_{j=1}^{N} . (x_j - \mu)^2$

Une telle situation est très rare, et l'étude de X sera fréquemment réalisée à partir d'observations partielles de X, ceci pour des considérations de coût, de rapidité de collecte et d'exploitation.

Soit E_n un échantillon de E de taille n. E_n est tout simplement un sous-ensemble quelconque de E de n éléments ; $E_n = \{e_{i_1}, ..., e_{i_n}\}$ où $1 \leq i_k \leq N$ et $1 \leq k \leq n$. Il est clair qu'il existe dans ce cas-là C_N^n différentes possibilités pour E_n. Nous supposons ici avoir procédé à la sélection de l'échantillon E_n de manière aléatoire.

On est alors dans le cas d'un tirage aléatoire. Tout calcul statistique sera effectué à partir des valeurs de la propriété X sur l'échantillon choisit aléatoirement E_n.

On note $X_1, ..., Xn$ les valeurs de X correspondant aux éléments de E_n. Ce sont des variables aléatoires car E_n a été tiré aléatoirement.

3.3.1 Modèle d'échantillonnage

Définition 3.14. *Soit une propriété définie par la v.a. X à valeur dans $\mathbf{X}$, application mesurable de $(\Omega, A, P) \to (X, B, P^X)$, B étant ici la tribu des Boréliens. Le modèle d'échantillonnage de taille n est l'espace produit $(\mathbf{X}, B, P)^n = (\mathbf{X}^n, B_n, P_n^X)$*

où - $\mathbf{X}^n = \mathbf{X} \times .... \times \mathbf{X}$ n fois est le produit cartésien de l'espace $\mathbf{X}$,

– B_n est la tribu produit des événements de $\mathbf{X}^n$,

– P_n^X est la loi ou la distribution jointe des observations.

On notera X_i la ième observation, v.a. de même loi que X et l'ensemble des observations $(X_1, ..., X_n)$ est l'échantillon aléatoire.

On notera que

$X_1, ..., X_n$ iid de loi P^X ou $(iid) \rightsquigarrow F_X$, F_X étant la fonction de répartition de X.

dans le cas où

la loi P^X est une loi discrète :

$$P(X_1 = x_1, ..., X_n = x_n) = \Pi_{j=1}^n P(X_j = x_j) = \Pi_{j=1}^n p_X(x_j)$$

ou la densité jointe dans le cas continu (P^X admet une densité f_X relativement à la mesure de Lebesgue) :

$$f_{(X_1,...,X_n)}(x_1,...,x_n) = \Pi_{j=1}^{n} f_{X_j}(x_j) = \Pi_{j=1}^{n} f_X(x_j)$$

Cas de la population finie

On se place dans le cas d'une population E de taille finie N pour laquelle la propriété X n'est observée que sur un ensemble E_n de taille $n \leq N$. On note $(x_1, ..., x_N)$ l'ensemble des valeurs prises par la propriété X sur l'ensemble de la population $E = \{e_1, ..., e_N\}$. Ces valeurs sont déterministes, elles appartiennent à $\mathbf{X}$. On a alors les vraies moyenne μ et variance σ^2 de X :

$$\mu = \frac{1}{N}\sum_{j=1}^{N} .x_j; \qquad \sigma^2 = \frac{1}{N}\sum_{j=1}^{N} .(x_j - \mu)^2$$

La moyenne empirique

La moyenne empirique de l'échantillon est donnée par l'expression

$\overline{X_N} = \mu = \frac{1}{N}\sum_{j=1}^{N} .X_j$

Pour calculer $E(\overline{X_N})$ et $Var(\overline{X_N})$ dans le cas d'une population finie E de taille N, il faut distinguer le mode de tirage.

a. Tirage avec remise On a

$$E(\overline{X_n}) = \frac{1}{n}\sum_{j=1}^{n} .E(X_j)$$

Chacune des variables X_j est tirée de l'ensemble $\{x_1, ..., x_N\}$ avec la probabilité $1/N$, c'est-à-dire $P(X_j = x_l) = 1/N$, $\forall l = 1, ..., N$. D'où

$$E(X_j) = \frac{1}{n}\sum_{j=1}^{n} .x_l = \mu$$

(la vraie moyenne de la population) et $E(\overline{X_N}) = \mu$.

Pour calculer la variance, notons que les Xj sont des variables aléatoires indépendantes, et donc

$Var(\overline{X_n})$=$\frac{1}{n^2}\sum_{j=1}^{n} .Var(X_j)$

3.3.2 Familles Exponentielles

Définition 3.15. *Un modèle paramétrique important en Statistique est celui des familles exponentielles. Il recouvre de nombreux modèle paramétriques classiques : normal, binomial, poisson, gamma etc...*

. Un modèle statistique $(E; E; P)$ sur un espace des observations E est dit famille exponentielle générale s'il existe un entier p, des fonctions η,T ,C et h tels que les densités puisse s'écrire, pour tout θ de Θ, sous la forme : $f_\theta(x) = e^{<\eta(\theta),T(x)>}C(\theta)h(x)$;

avec les contraintes que T soit une fonction mesurable à valeurs dans R^p ; η soit une fonction à valeurs dans R^p ; C soit une fonction réelle positive qui ne dépend pas x ; h soit une fonction borélienne positive qui ne dépend pas de θ. Le vecteur aléatoire $T(X)$ est appelé statistique canonique du modèle. Si la fonction T est l'identité, la famille exponentielle est dite naturelle. On parle de forme canonique d'une famille exponentielle générale quand les densités de probabilités ont la forme $f_\theta(x) = e^{<\theta,T(x)>}C(\theta)h(x)$; pour tout θ de Θ, ce qu'il est toujours possible d'obtenir quitte à reparamétriser la famille par $\theta' = \eta^\theta$. Dans ce cas le paramètre θ de la famille exponentielle est appelé paramètre canonique.

Exemple 3.16. *Revenons sur le modèle de Bernoulli. La densité s'écrit :* $f_p(x) = p^x(1-p)^{1-x} = (\frac{p}{1-p}^x)(1-p) = e^{xln(\frac{p}{1-p})}(1-p) = e^{<\eta(p),T(x)>}C(p)h(x)$;

avec $\eta(p) = \frac{p}{1-p}; T(x) = x; C(p) = (1p)eth(x) = 1.$

3.3.3 Modèle position-échelle

Considérons un vecteur aléatoire X de loi P connue sur $(\mathbb{R}^n; B_{\mathbb{R}^n})$net A un sous espace de $\mathbb{R}^n$. Pour tout a dans A et tout b dans $\mathbb{R}_+$, on note $P_{a;b}$ la loin du vecteur $Y = a + bX$.

$P_{A;b} = \{P_{a;b} : a \in A; b \in \mathbb{R}_+\}$ est appelé modèle position-échelle engendré par P (ou par X). Le paramètre a est appelé paramètre de position et b paramètre d'échelle.

Si b est fixé (par exemple à 1) on parle de modèle de position. Dans le cas où A ne contient que le vecteur nul de $\mathbb{R}^n$, on parle de modèle échelle. **Exemple :Le Modèle gaussien unidimensionne**

Le modèle $P = \{N(u; \sigma^2); u \in \mathbb{R}\}$ est un modèle position engendré par la loi $N(0; \sigma^2)$.

Le modèle $P = \{N(u; \sigma^2); u \in \mathbb{R}, \sigma^2 > 0\}$ est un modèle position-échelle engendré par la loi $N(0; 1)$.

3.4 Exhaustivité

Statistique

Soit X une v.a. à valeurs dans $(\mathbf{X}, B)$ et soit $(\mathbf{Y}, C)$ un espace mesurable auxiliaire quelconque.

Définition 3.17. *On appelle statistique toute application T mesurable de $\mathbf{X}^n$ dans $\mathbf{Y}$, $\forall\ n\ T : \mathbf{X}^n \to \mathbf{Y}$*

Par exemple, $\mathbf{X} = \mathbf{Y} = \mathbb{R}$ et

$$T(X_1, ..., X_n) = \frac{1}{n}\sum_{j=1}^{n} X_j = \overline{X_n}$$

$$T(X_1, ..., X_n) = \frac{1}{n}\sum_{j=1}^{n} \left(X_j - \overline{X_n}\right)^2$$

ou $\mathbf{X} = \mathbb{R}$, $\mathbf{Y}= \mathbb{R}^n$ et $T(X_1, ..., X_n) = (X_{(1)}, ..., X_{(n)})$, où $X_{(1)} \leq X_{(2)}...X_{(n)}$

(cette statistique porte le nom de statistique d'ordre.

3.4.1 Statistique exhaustive

Définition 3.18. *On appelle modèle statistique paramétrique de paramètre $\theta \in \Theta$ pour un certain espace de dimension fini le couple $(\mathbf{X}, P_\theta)$, où $\mathbf{X}$ est l'espace des valeurs de X, v.a. du modèle, et P_θ la loi de probabilité de X*

La statistique T sera dite **exhaustive pour** θ si la loi conditionnelle de X sachant $T(X) = t$ n'est pas une fonction du paramètre θ : $P_\theta(X|T(X) = t)$ne dépend pas de θ.

On notera $f(x, \theta)$ la densité de P_θ relativement à une mesure dominante et
$sigma-$finie, μ. On va se restreindre au cas où μ est la mesure de Lebesgue (variables aléatoires de loi absolument continue) et on retrouve la densité $f_\theta(x)$ ou la mesure de comptage (variables aléatoires de loi discrète) et on retrouve le système $P_\theta(X = x)$. On note X l'échantillon $(X_1, ..., X_n)$ issu du même modèle (X, P_θ).

Exemple la vraisemblance d'un échantillon $X = (X_1; ...; X_n)$ dans un tel modèle est :

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

On peut écrire :

$L(x_1, ..., x_n; p) = g_p(T(\underline{x}))h(\underline{x})$;

avec $g_p(x) = p^x(1-p)^{n-x}$ et h égale à 1. Grâce au théorème de factorisation on retrouve que la Statistique $T(X) = \sum_{i=1}^{n} X_i$ est bien exhaustive pour le paramètre p dans ce modèle.

Théorème : (Théorème de factorisation) Soit le modèle $(X, P_\theta$ et T une statistique $(\mathbf{X}^n, B_n) \to (\mathbf{Y}, C)$. T est exhaustive pour θ si et seulement s'il existe deux fonctions mesurables $g : \mathbf{X}\mathbb{R}_+$ et $h : \mathbf{Y} \to \mathbb{R}_+$ telles que $f(x, \theta)$ se met sous la forme $f(x, \theta) = h(x)g(T(x), \theta)$ où $x = (x_1, ..., x_n)$.

Exemples :

– Soit $X \sim U[0, \theta]$. On a $f(x_1, ..., x_n, \theta) = \frac{1}{\theta^n}\mathbf{1}_{sup1\leq j\leq nxj\leq\theta}$

En posant $h(x) = 1$ et $g(T(x), \theta) = \frac{1}{\theta^n}\mathbf{1}_{T(x)\leq\theta}$

on déduit que $T : x \longmapsto sup_{1\leq j\leq n}x_j$ est une statistique exhaustive pour θ.

Soit $X \sim Exp(\theta)$. On a $f(x_1, ..., x_n, \theta) = \frac{1}{\theta^n} \exp\left(-\theta \sum_{j=1}^{n} X_j\right.$

et donc $T(X_1, ..., X_n) = \sum_{j=1}^{n} X_j$ est bien une statistique exhaustive.

– Soit $X \sim P(\theta)$. On a $f(x_1, ..., x_n, \theta) = e^{-n\theta}\theta^{\sum_{j=1}^{n} x_j}\Pi_{j=1}^{n}x_j!$

et donc $T(X_1, ..., X_n) = \sum_{j=1}^{n} X_j$ est bien une statistique exhaustive.

– Soit $X \sim N(\mu, \sigma^2)$. Alors la statistique $T(X_1, ..., X_n) = \left(\frac{1}{n}\sum_{j=1}^{n} X_j; \frac{1}{n}\sum_{j=1}^{n} X_j^2\right)$ est une statistique exhaustive pour $\theta = (\mu, \sigma^2)$.

3.4.2 Statistique exhaustive minimale

Le théorème de factorisation implique que si T_1 est une statistique exhaustive pour θ alors la statistique T_2 telle que $T_1 = \varphi \circ T_2$, où φ est une application mesurable, est aussi une statistique exhaustive pour θ. Une statistique exhaustive n'est donc pas unique car il suffit de choisir n'importe quelle application mesurable et bijective, φ, et de considérer $T_2 = \varphi^{-1} \circ T_1$. Ces remarques nous conduisent à la définition suivante.

Définition 3.19. *Une statistique T est dite **exhaustive minimale** pour θ si elle est exhaustive et si pour toute autre statisti que exhaustive S pour θ, il existe une application φ telle que $T = \varphi \circ S$.*

definition

Lemme Deux statistiques exhaustives minimales pour θ sont en liaison bijective.

Soit $f(x,\theta)$ la densité de P_θ par rapport à la mesure dominante μ. Le théorème suivant nous donne une condition suffisante pour qu'une statistique soit exhaustive minimale.

Montrons d'abord que T est une statistique exhaustive pour θ.

Pour tout $t \in T(\mathbf{X}^n)$, considérons l'ensemble $[T=t] = \{x \in X^n : T(x) = t\}$.à tout élément $x \in \mathbf{X}^n$, on associe x_t dans $[T=t]$. On a donc par construction $T(x) = T(x_{T(x)})$ et par conséquent, par hypothèse, le rapport $h(x) = \frac{f(x,\theta)}{f(x_{T(x)},\theta)}$ est indépendant de θ). Définissons maintenant la fonction $g(T(x),\theta) = f(x_{T(x)},\theta)$. On peut écrire $f(x,\theta) = h(x)g(T(x),\theta))$ et ainsi le théorème de factorisation assure que T est une statistique exhaustive pour θ).

Montrons maintenant que T est minimale. Soit $T\prime$ une autre statistique exhaustive.

Par le théorème de factorisation, il existe deux fonctions $h\prime$ et $g\prime$ telles que $f(x,\theta)) = h\prime(x)g\prime(T\prime(x),\theta))$.

Alors, pour tout x_1 et x_2 tels que $T\prime(x_1) = T\prime(x_2)$, il vient que $\frac{f(x_1,\theta))}{f(x_2,\theta))} = \frac{h\prime(x_1)}{h\prime(x_2)}$

Puisque ce rapport ne dépend pas de θ), l'hypothèse du théorème assure que $T(x_1) = T(x_2)$. On en déduit que T doit être une fonction de $T\prime$ et que T est donc une statistique exhaustive minimale.

Soient $X_1,...,X_n$ $(i.i.d)$. $N(\mu,\sigma^2)$ où μ et σ sont inconnus. Montrons que $T(X_1,...,X_n) = \overline{X_n}$, la moyenne empirique, est une statistique exhaustive minimale pour μ. Notons tout d'abord que

$$\mathrm{f}(\mathrm{x}_1,...,x_n,\mu,\sigma^2)=(2\Pi\sigma^2)^{-n/2}\exp(\frac{-n\overline{x}_n-\mu^2+ns_n^2}{2\sigma^2})$$

où $\overline{x}_n = 1/n\sum_{j=1}^{n} x_j$ et $\mathrm{s}^2 = \sum_{j=1}^{n}(x_j-\overline{x}_n)^2$ Il s'en suit que $\frac{f(x_1,...,x_n,\theta))}{f(y_1,...,y_n,\theta))} = \exp(-\frac{n(\overline{x}_n-\mu)^2-n(\overline{y}_n-\mu)^2+ns_x^2-ns_y^2}{2\sigma^2})$

Le rapport ne dépend pas deμ $sietseulementsi\overline{x_n} = \overline{y_n}$

3.4.3 Statistique libre, complète et notion d'identifiabilité

Statistique

Qu'elle serait une sorte d'opposée de la notion de statistique exhaustive minimale ? Ce devrait être une statistique ne dépendant pas du paramètre, soit :

Une statistique libre n'apporte donc aucune information pour l'estimation du paramètre θ). C'est ce qu'on appelle un paramètre de nuisance. Or, de façon assez surprenante il peut arriver qu'une statistique exhaustive minimale comprenne une statistique libre, qui intuitivement ne devrait pas être prise en compte pour donner toute l'information sur
theta). Par exemple la loi $P_{\theta)}$ uniforme sur$[\theta), \theta) + 1]$; pour un échantillon de taille 2, la statistique $(X_{(2)} - X_{(1)}, X_1 + X_2)$ est exhaustive minimale, mais $X_{(2)} - X_{(1)}$ est libre.

Aussi peut-on rajouter une autre caractérisation des statistiques exhaustives pour pouvoir atteindre une forme d'optimalité pour ces statistiques, qui serait qu'aucune fonctionnelle non constante de la statistique ne peut être libre. C'est la notion de statistique complète.

Statistique complète

Définition 3.20. *Une statistique exhaustive T d'un modèle statistique paramétrique avec T à valeur dans $\mathbb{R}^d$ est dite complète si pour toute fonction borélienne $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ telle que $f(T)$ soit intégrable, on ait : $\forall \theta) \in \Theta$, $E_{\theta)}(f(T)) = 0 \Rightarrow f(T) = 0$ $P_\theta - p.s.$*

Propriété Soit un modèle statistique paramétrique dominé.

1. si est une statistique exhaustive complète alors pour toute fonction borélienne φ bijective $\varphi(T)$ est une statistique exhaustive complète.

2. si T est une statistique exhaustive complète alors T est une statistique exhaustive minimale.

3. (Théorème de Basu) si T est une statistique exhaustive complète alors T est indépendante de toute statistique libre sur le modèle.

Démonstration : Nous allons prouver le troisième point à savoir le Théorème de Basu. Soit S une statistique libre pour le modèle et soit f une fonction telle que $E_{\theta)}(f(S))$ existe. On peut noter e l'application linéaire qui à f associe $e(f) = E_{\theta)}(f(S))$. Comme S est libre, e ne dépend pas de θ). Par suite, et pour tout $\theta)\prime \in \Theta$, la statistique $E_{\theta)\prime}(f(S)|T) - e(f)$ est une fonction de T mesurable telle que $E_{\theta)}\left(E_{\theta)\prime}(f(S)|T) - e(f)\right) = 0$ pour tout $\theta) \in \Theta$. Comme on a supposé que T est exhaustive complète, alors $E_{\theta)\prime}(f(S)|T) = e(f)$ presque-sûrement. Autrement dit l'espérance conditionnelle de $f(S)$ par rapport à T est une fonction constante de T. Elle n'est pas aléatoire et les statistiques S et T sont indépendantes.

Dans un modèle statistique paramétrique, il existe toujours une statistique exhaustive minimale

mais pas toujours de statistique exhaustive complète.

Notion d'identifiabilité

Soit $(\mathbf{X}, P_{\theta)}), \theta) \in \Theta$ un modèle statistique paramétrique.

Définition 3.21. *Une valeur du paramètre* $\theta)_0 \in \Theta estidentifiablesi \forall \theta) \neq \theta)_0$, $P_{\theta)} \neq P_{\theta_0}$. *Le modèle* $(\mathbf{X}, P_{\theta)}), \theta) \in \Theta$ *est dit identifiable si tous les paramètres sont identifiables ; c-à-d., si l'application* $\theta) \longmapsto P_\theta$ *est injective.*

On peut affaiblir la notion précédente à une notion locale.

Définition 3.22. *Une valeur du paramètre* $\theta)_0 \in \Theta$ *est localement identifiable s'il existe un voisinage* ω_0 *de* $\theta)_0$ *tel que* $\forall \theta \in \omega_0 : \theta \neq \theta_0$ *on a* $P_\theta \neq P_{\theta 0}$ *. Le modèle* $(\mathbf{X}, P_\theta)$, $\theta \in \Theta$ *est dit localement identifiable si tous les paramètres sont localement identifiables.*

3.5 Éléments de théorie de l'information

On définira dans cette section différentes quantités mesurant l'information contenue dans un modèle statistique.

Information au sens de Fisher

Soit le modèle statistique $(\mathbf{X}, P_\theta)$, $\theta \in \Theta$ tel que P_θ admet une densité $f(x,\theta)$ relativement à la mesure dominante μ. On appellera hypothèses usuelles les 4 hypothèses suivantes :

$H1$: Θ est un ouvert de $\mathbb{R}^d$ pour un certain d fini.

$H2$: Le support $\{x : f(x,\theta) > 0\}$ ne dépend pas de θ.

$H3$: Pour tout $x \in \mathbf{X}$ la fonction $f(x,\theta)$ est au moins deux fois dérivable par rapport à θ pour tout $\theta \in \Theta$ et que les dérivées première et seconde sont continues. On dit que $\theta \longmapsto f(x,$
$theta)$ est C^2.

$H4$: Pour tout $B \in \mathbf{B}$ l'intégrale ${}_B f(x,\theta)d\mu(x)$ est au moins deux fois dérivable sous le signe d'intégration et on peut permuter intégration et dérivation ; c-à-d.,

$$\frac{\partial}{\partial\theta_j}\int\limits_B f(x,theta)d\mu(x) = \int\limits_B \frac{\partial f(x,\theta)}{\partial\theta_j}d\mu(x), \quad j = 1,...,d$$
$$\frac{\partial^2}{\partial\theta i\partial\theta_j}\int\limits_B f(x,\theta)d\mu(x) = \int\limits_B \frac{\partial^2 f(x,\theta)}{\partial\theta_i\partial\theta_j}d\mu(x), \quad i,j = \{1,...,d\}$$

Lorsque ces 4 hypothèses sont vérifiées, on dit que le modèle est régulier.

Les modèles $X \rightsquigarrow P(\theta)$, $\theta > 0$, $X \rightsquigarrow Exp(\lambda)$, $\lambda > 0$ et $X \rightsquigarrow N(\mu,\sigma 2)$, $\theta = (\mu,\sigma 2) \in \mathbb{R} \times \mathbb{R}_+^*$ sont réguliers mais pas $X \rightsquigarrow U[0,\theta]$, $\theta > 0$.

On appelle score le vecteur aléatoire $S(X,\theta)$ définit par

$$S(X,\theta) = \nabla_\theta(logf(X,\theta)) = \left(\frac{\partial logf(X,\theta)}{\partial\theta 1},...,\frac{\partial logf(X,\theta)}{\partial\ theta_d}\right)^T$$

Propriété – Le score est un vecteur aléatoire centré $E(S(X,\theta)) = 0$.

– Le vecteur score est additif : Soient X et Y deux variables aléatoires indépendantes associées aux modèles statistiques (X,P_θ) et (Y,Q_θ). Alors $S(X,\theta)$ et $S(Y,\theta)$ sont indépendants

$$S((X,Y),\theta) = S(X,\theta) + S(Y,\theta), \forall\theta \in \Theta.$$

Ici (X, Y) est associé au modèle statistique $(X \times Y, P_\theta \otimes Q_\theta)$.

Définition 3.23. *On appelle information de Fisher au point θ la matrice*

$$I(\theta) = E\left(S(X,\theta)S(X,\theta)^T\right) =$$

$$\begin{pmatrix} E\left[\left(\frac{\partial logf(X,\theta)}{\partial\theta_1}\right)^2\right] & E\left[\left(\frac{\partial logf(X,\theta)}{\partial\theta_1}\frac{\partial logf(X,\theta)}{\partial\theta_2}\right)\right] & . & . & . & E\left[\left(\frac{\partial logf(X,\theta)}{\partial\theta_1}\frac{\partial logf(X,\theta)}{\partial\theta_d}\right)\right] \\ . & & & . & & . \\ . & & & . & & . \\ . & & & . & & . \\ E\left[\left(\frac{\partial logf(X,\theta)}{\partial\theta_d}\right)^2\right] & E\left[\left(\frac{\partial logf(X,\theta)}{\partial\theta_d}\frac{\partial logf(X,\theta)}{\partial\theta}{}_2\right)\right] & . & . & . & E\left[\left(\frac{\partial logf(X,\theta)}{\partial\theta_d}\right)^2\right] \end{pmatrix}$$

Pour un modèle régulier, on a la relation $I(\theta) = -E\left[\nabla_\theta(S(X,\theta)^T\right] =$

$$\begin{pmatrix} -E\left[\left(\frac{\partial^2 logf(X,\theta)}{\partial\theta_1^2}\right)\right] & -E\left[\left(\frac{\partial^2 logf(X,\theta)}{\partial\theta_1\partial\theta_2}\right)\right] & . & . & . & -E\left[\left(\frac{\partial^2 logf(X,\theta)}{\partial\theta_1\partial\theta_d}\right)\right] \\ . & & & . & & . \\ . & & & . & & . \\ . & & & . & & . \\ -E\left[\left(\frac{\partial^2 logf(X,\theta)}{\partial\theta_d^2}\right)^2\right] & E\left[\left(\frac{\partial^2 logf(X,\theta)}{\partial\theta_d\partial\theta_2}\right)\right] & . & . & . & E\left[\left(\frac{\partial^2 logf(X,\theta)}{\partial\theta_d^2}\right)\right] \end{pmatrix}$$

et donc pour tout $1 \leq i, j \leq d$: $I_{ij}(\theta) = -E\left[\left(\frac{\partial^2 logf(X,\theta)}{\partial\theta_i\partial\theta_j}\right)\right]$

Notons que pour le calcul de $I(\theta)$, l'espérance est prise par rapport à P_θ, à θ fixé.

Propriété On suppose ici que les hypothèses $H1 - H4$ sont vérifiées, donc que le modèle est régulier.

– L'information de Fischer est une matrice symétrique définie positive. En effet, étant donné que le score est centré $I(\theta) = Var(S(X,\theta)) \geq 0$.

– L'information de Fischer est additive : Si X et Y deux variables aléatoires indépendantes dans des modèles paramétriques au paramètre θ commun alors $I(X,Y)(\theta) = IX(\theta) + IY(\theta)$, $\forall\theta \in, \Theta$

car c'est la variance d'une somme de scores indépendants.

Soit $X \rightsquigarrow N(\mu\sigma^2)$, alors $I(\mu,\sigma^2) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$

En effet,

$$\log f(x,\mu,\sigma^2) = -\frac{1}{2}\log 2\Pi - \Pi 12\log\sigma^2 - \frac{1}{2\sigma^2}(x-\mu)^2$$

$$\frac{\partial^2 logf(x,\mu,\sigma^2)}{\partial\mu^2} = \frac{1}{\sigma^2} \Rightarrow -E\left[\frac{\partial^2 logf(X,\mu,\sigma}{2)}\partial\mu^2\right] = \frac{1}{\sigma^2}$$

$$\frac{\partial^2 logf(x,\mu,\sigma^2)}{(\partial\sigma^2)^2} = \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(x-\mu)^2 \Rightarrow -E\frac{\partial^2 logf(X,\mu,\sigma^2)}{(\partial\sigma^2)^2} = \frac{1}{2\sigma^4}$$

$$\frac{\partial^2 logf(x,\mu,\sigma^2)}{\partial\mu\partial\sigma^2} = 0 \Rightarrow E\frac{\partial^2 logf(X,\mu,\sigma^2)}{\partial\mu\partial\sigma^2 = 0}$$

Pour un échantillon $X_1, ..., X_n$, le vecteur score $S_n(\theta)$ et l'information de Fischer $I_n(\theta)$ associés à θ sont donnés par

$$S_n(\theta) = \nabla_{\theta}{}_{i=1}^{n}\log f(X_i,\theta) \quad et \qquad I_n(\theta) = var(S_n(\theta)).$$

On déduit de l'indépendance des X_j que

$$S_n(\theta) = \sum_{j=1}^{n} S(X_j,\theta)$$

où les scores $S(X_1,\theta), ..., S(X_n,\theta)$ sont $(i.i.d)$. (la loi de $S(X,\theta)$ est l'image de la loi de X par l'application $S : x \longmapsto S(x,\theta)$). Etant donné que $E(S(X,\theta)) = 0$, et $Var(S(X,\theta)) = I(\theta) < +\infty$, on a donc la relation

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

En vertu de la loi forte des grands nombres et du théorème central limite, on a aussi :

$$\frac{1}{n}S_n(\theta) \to 0p.s \quad et \qquad \frac{S_n(\theta)}{\sqrt{n}}\,(L) - \to N_d(0, I(\theta))$$

.

Chapitre 4

ESTIMATION

Objectif : L'estimation consiste à rechercher la valeur numérique d'un ou plusieur paramètres inconnus d'une loi de probabilité à partir d'observations (valeurs prises par la v.a. qui suit cette loi de probabilité)

4.1 Distribution d'échantillonnage

Pour résoudre les problèmes d'estimation de paramètres inconnus, il faut tout d'abord étudier les distributions d'échantillonnage, c'est à dire la loi de probabilité suivie par l'estimateur.

Remarque : En théorie de l'estimation, il s'agit de distinguer soigneusement trois concepts différents :

-les paramètres de la population comme la moyenne $\mu dontlavaleurestcertainemaisinconnuesymbolisspar$

-les résultats de l'échantillonnage comme la moyenne x dont la valeur est certaine mais connue symbolisés par des minuscules.

-les variables aléatoires des paramètres, comme la moyenne aléatoire X dont la

valeur est incertaine puisque aléatoire mais dont la loi de probabilité est souvent connue et symbolisées par des majuscules.

4.1.1 Approche empirique

Il est possible d'extraire d'une population de paramètres p, $\mu ou \sigma^2 pour unevariablealatoire X, kchantillons$

On obtient ainsi pour chaque paramètre estimé, une série statistique composée de k éléments à savoir

les k estimations du paramètre étudié. Par exemple, on aura k valeurs de moyennes observées (graphe ci-dessus).

La distribution associée à ces k estimations constitue la distribution d'échantillonnage du paramètre. On peut alors associer une variable aléatoire à chacun des paramètres. La loi de probabilité suivie par cette variable aléatoire admet comme distribution, la distribution d'échantillonnage du paramètre auquel on pourra associer une espérance et une variance.

4.1.2 Approche théorique

En pratique, les données étudiées sont relatives à un seul échantillon. C'est pourquoi, il faut rechercher les propriétés des échantillons susceptibles d'être prélevés de la population ou plus précisément les lois de probabilité de variables aléatoires associées à un échantillon aléatoire.

Ainsi les n observations $\mathrm{x}_1, x_2, \ldots, x_i, \ldots, x_n, faitessurunchantillonpeuventtreconsidrescommenvariable$

Cette valeur sera différente si l'on considère un autre échantillon. Il en est de même pour les n valeurs extraites de la population.

A partir de ces n variables aléatoires, on peut définir alors une nouvelle variable qui sera fonction de ces dernières telle que : Y = f($\mathrm{X1}_1, X_2, \ldots, X_i, \ldots, X_n)parexemple : Y = X_1 + X_2 + \ldots + X_i + \ldots . X_n$

Ainsi la loi de probabilité de la variable aléatoire Y dépendra à la fois de la loi de probabilité de la variable aléatoire X et de la nature de la fonction f.

4.1.3 Loi de probabilité de la moyenne

Soit X une variable aléatoire suivant une loi normale d'espérance $\mu etdevariance\sigma^2$ et n copies indépendantes X_1,X_2,...,X_i,...,X_n telle que X_i associe le ième élément de chacun des n échantillons avec $E(X_i) = \mu$ et V(X_i) =σ^2.

On construit alors la variable aléatoire X , telle que $\overline{X} = \frac{X_1+X_2+..X_i+...+X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$

avec pour espérance : $E(\overline{X}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}E(\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) == \frac{1}{n}\mu$ (Propriétés de l'espérance)

d'où $E(\overline{X}) = \mu$ $E(\overline{X})$ est notée également $\mu_{\overline{X}}$

et pour variance si $V(X_i) = \sigma^2$:

$V(\overline{X}) = V(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}E(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{sigma^2}{n}$ (Propriétés de la variance)

$V(\overline{X})$ est notée également $\sigma^2_{\overline{X}}$

La loi de probabilité de la variable aléatoire X , moyenne de n v.a. X de loi de probabilité $\mathrm{N}(\mu, \sigma), estuneloin$

Remarque : il est aisé de voir sur un graphe que la variance associée à une moyenne $\frac{\sigma^2}{n}$ est plus faible que la variance de la variable elle-même (σ^2).

Exemple :

Des études statistiques montrent que le taux de glucose dans le sang est une variable normale X d'espérance $\mu = 1g/letd'cart - type\sigma = 0,1g/l$.

En prenant un échantillon de 9 individus dans la population, l'espérance et l'écart-type théorique attendu de la variable aléatoire X sont alors :

$\mu_X = \mu = 1g/l \; et \; \sigma_X = \frac{\sigma}{\sqrt{n9}} = \dfrac{0,1}{\sqrt{9}} = 0,03g/l$

4.1.4 Convergence

En fonction de la nature de la variable aléatoire continue X, de la taille de l'échantillon n et dela connaissance que nous avons sur le paramètre σ^2, la variable centrée réduite construite avec $\overline{X}$ converge vers différentes lois de probabilité.

Lorsque la variance σ^2 est connue et n grand ($n \geq 30$), on se trouve dans les conditions du théorème central limite et la loi suivie par : $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \to N(0,1)$ loi normale réduite

Ceci reste vrai lorsque n $\leq 30 seulementsilaloisuivieparXsuituneloinormale.$

Lorsque la variance σ^2 est inconnue et X suit une loi normale, la loi suivie par la variable centrée réduite est alors :$\frac{X-\mu}{\sigma}/\sqrt{n} \to T_{n-1}$ loi de student à n-1 degrés de liberté.

Lorsque n $\geq 30, laloidestudenttendversuneloinormalerduite(voirconvergence)$.

Lorsque la variance σ^2 est inconnue et X ne suit pas une loi normale, la loi suivie par $\frac{X-\mu}{\widehat{\sigma}/\sqrt{n}}$ n'est pas connue.

4.1.5 Loi de probabilité d'une fréquence

Soit une population dans laquelle une proportion p des individus présente une certaine propriété.

Si k est le nombre d'individu présentant la propriété dans un échantillon de taille n, alors la variable

aléatoire K résultant de différents échantillonnages suit une loi binomiale B(n,p) avec $E(K) = np$ et $V(K) = npq$.

On construit la variable aléatoire $F = \frac{K}{n}$ avec pour espérance : $E(F) = E(\frac{K}{n}) = \frac{1}{n}E(K) = \frac{1}{n}np = p$

et pour variance : V(F) =V($\frac{K}{n}$) = $\frac{1}{n^2}V(K) = \frac{1}{n^2}npq = npq$.

La loi de probabilité d'une fréquence $\frac{K}{n}$, suit une loi normale N(p,$\sqrt{\frac{pq}{n}}$) vrai si np > 5 et nq > 5.

4.2 Estimateur

Définition 4.1. *Soient X_1 , X_2 ,. . . , X_i , . . . , X_n , n réalisations indépendantes de la variable aléatoire X (discrète ou continue) et θunparamtreassocilaloideprobabilitsuiviparX, unestimateurduparamtreθestuneva* $\Theta = f(X_1, X_2, \ldots, X_i, \ldots, X_n)$

Si on considère n observations x_1 , x_2 ,. . . , x_i , . . . , x_n, l'estimateur $\Theta fournirauneestimationde\theta notegaleme$

$\widehat{\theta} = f(x_1, x_2, \ldots, x_i, \ldots, x_n)$

L'estimation d'un paramètre inconnu, noté $\theta estfonctiondesobservationsrsultantd'unchantillonnagealato$

L'estimation de $\theta estunevariablealatoire\Theta dontladistributiondeprobabilits'appelleladistributiond'chantill$

L'estimateur $\Theta admetdoncuneespranceE(\Theta)etunevarianceV(\Theta)$.

4.2.1 Propriétés

Convergence

L'estimateur $\Theta doittendreverslavaleurrelleduparamtre\theta lorsquelenombred'individustudiaugmente.Ondi$

Si $\forall \epsilon > 0 P(|\Theta - \theta|) > \epsilon) \to 0$ lorsque $n \to \infty$

Ceci équivaut à dire qu'en limite $\Theta \to \Theta$ lorsque $n \to \infty$.

Biais d'un estimateur

Le biais d'un estimateur noté B(Θ)$estladifrencemoyenneentresavaleuretcelleduparamtrequ'ilestime.L$

$B(\Theta) = E(\Theta - \theta) = E(\Theta) - E(\theta) = E(\Theta) - \theta = 0$ (voir propriétés de l'espérance)

d'où $E(\Theta) = \theta$

Ainsi l'estimateur sera sans biais si son espérance est égale à la valeur du paramètre de la population. $E(\Theta) = \theta$

Remarque : Un estimateur est asymptotiquement sans biais si $E(\Theta) \to \theta$ lorsque n $\longrightarrow \infty$

Variance d'un estimateur

Si deux estimateurs sont convergents et sans biais, le plus efficace est celui qui a la variance la plus faible car ses valeurs sont en moyenne plus proches de la quantité estimée. $V(\Theta) = E(\Theta - E(\Theta))^2$ minimale

Remarque : Quand les estimateurs sont biaisés, en revanche, leur comparaison n'est pas simple. Ainsi un estimateur peu biaisé mais de variance très faible, pourrait même être préféré à un estimateur sans biais mais de grande variance.

Si un estimateur est asymptotiquement sans biais et si sa variance tend vers 0 lorsque n $\longrightarrow \infty, ilestconvergent.$

$P(|\Theta - \theta| \geq \epsilon) \leq \frac{V(\Theta)}{\epsilon^2}$ avec ϵ>0.$(IngalitdeBienaym - Tchbycheff)$

Cette inégalité exprime que si $|\Theta - \theta|$ tend vers 0 quand n augmente, V(Θ)$doitaussitendrevers$0.

4.2.2 Estimation ponctuelle et par intervalle

L'estimation d'un paramètre quelconque $\theta estponctuellesil'onassocieuneseulevaleurl'estimateur\widehat{\theta}$ à partir des données observables sur un échantillon aléatoire.

L'estimation par intervalle associe à un échantillon aléatoire, un intervalle $[\widehat{\theta}_1, \widehat{\theta}_2]$ qui recouvre $\theta avecunecertaineprobabilit.$

Estimation ponctuelle

Si la distribution de la variable aléatoire X est connue, on utilise la méthode du maximum de vraisemblance pour estimer les paramètres de la loi de probabilité. En revanche si la distribution n'est pas connue, on utilise la méthode des moindres carrés.

Espérance

Soit X une variable aléatoire continue suivant une loi normale $\mathrm{N}(\mu,\sigma) dont la valeur des paramtres n'est pas co$

Soient X_1 , X_2 ,. . . , X_i , . . . , X_n , n réalisations indépendantes de la variable aléatoire X, un estimateur du paramètre $\mu est une suite de variable aléatoire \Theta fonctions des X_i : \Theta = f(X_1, X_2, \ldots, X_i, \ldots, X_n)$

La méthode des moindres carrés consiste à rechercher les coefficients de la combinaison linéaire $\Theta = a_1X_1 + a_2X_2 + \ldots + a_iX_i + \ldots + a_nX_n$ telle que $E(\Theta) = \mu$ et $V(\Theta)$ soit minimale.

La moyenne arithmétique constitue le meilleur estimateur de $\mu, esprance de la loi de probabilit de la variable ala$

$\widehat{u} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Voici pourquoi :

Estimateur sans biais : $E(\overline{X}) = \mu$ (voir loi de la moyenne)

Estimateur convergent : si l'on pose l'inégalité de Biénaymé-Tchébycheff :

$P(\left|\overline{X} - \mu\right| \geq \epsilon) \leq \frac{V(\overline{X})}{\epsilon^2}$ avec ϵ>0

lorsque $n \to \infty$ $\frac{V(\overline{X})}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0$ et ceci $\forall\epsilon$>0.

ainsi en limite,$P(\left|\overline{X} - \mu\right| \geq \epsilon) = 0$, ce qui indique que X $\longrightarrow \mu en probabilit.$

Variance

Soit X une variable aléatoire continue suivant une loi normale $\mathrm{N}\ (\mu,\sigma) pour laquelle on souhaite estimer la var$

Soient X_1 , X_2 ,. . . , X_i , . . . , X_n , n réalisations indépendantes de la variable aléatoire X, un estimateur du paramètre σ^2 est une suite de variable aléatoire $\Theta fonctions des X_i : \Theta = f(X_1, X_2, \ldots, X_i, \ldots, X_n)$

• Cas où l'espérance $\mu est connue$

La méthode des moindres carrés consiste à rechercher les coefficients de la combinaison linéaire $\Theta = a_1(X_1 - \mu)^2 + a_2(X_2 - \mu)^2 + \ldots + a_i(X_i - \mu)^2 + \ldots + a_n(X_n - \mu)^2$

telle que $E(\Theta) = \sigma^2$ et $V(\Theta)$ soit minimale.

La variance observée constitue le meilleur estimateur de σ^2, variance de la loi de probabilité de la variable aléatoire X lorsque l'espérance $\mu est connue$:

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

Remarque : Cette estimation de la variance de la population est rarement utilisée dans la mesure où si la variance σ^2 n'est pas connue, l'espérance $\mu nel'estpasnonplus$.

• Cas où l'espérance $\mu estinconnue$

Dans ce cas, nous allons estimer $\mu avec\widehat{\mu} = \overline{X}$ et dans ce cas $\sum_{i=1}^{n}(X_i - \mu)^2 \neq \sum_{i=1}^{n}(X_i - \overline{X})^2$

Nous allons étudier la relation entre ces deux termes à partir de la variance observée :

$s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}\left[(X_i - u) - (\overline{X} - u)\right]^2 = \sigma^2 - \frac{\sigma^2}{n}$

en effet $\sigma^2_{\overline{X}} = \frac{1}{n}\sum_{i=1}^{n}(\overline{X} - u)^2 = (\overline{X} - u)^2 = \frac{\sigma^2}{n}$

ainsi $s^2 = \frac{(n-1)\sigma^2}{n}$

Le meilleur estimateur de σ^2, variance de la loi de probabilité de la variable aléatoire X lorsque l'espérance $\mu estinconnueest$:$\widehat{\sigma}^2 = \frac{ns^2}{n-1} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

Remarque : Lorsque n augmente, la variance observée s^2 tend vers la variance de la population σ^2.
( ) 2 2 2 1

$lim_{n\to\infty}s^2 = lim_{n\to\infty}\frac{(n-1)\sigma^2}{n} = \sigma^2$.

4.2.3 Fréquence

Soit le schéma de Bernoulli dans lequel le caractère A correspond au succès. On note p la fréquence des individus de la population possédant le caractère A. La valeur de ce paramètre étant inconnu, on cherche à estimer la fréquence p à partir des données observables sur un échantillon.

A chaque échantillon non exhaustif de taille n, on associe l'entier k, nombre d'individus possédant le caractère A.

Soit K une variable aléatoire discrète suivant une loi binomiale $\mathbf{B}(\mathbf{n}, \mathbf{p})$ et pour laquelle on souhaite estimer la fréquence **p**.La fréquence observée du nombre de succès observé dans un échantillon de taille n constitue le meilleur estimateur de p : $\widehat{p} = \frac{K}{n}$

Voici pourquoi :

Estimateur sans biais : $E(\frac{K}{n}) = p$ (voir loi de fréquence)

Estimateur convergent : si l'on pose l'inégalité de Biénaymé-Tchébycheff

$P(\left|\frac{K}{n} - p\right| \geq \epsilon) \leq \frac{V(\frac{K}{n})}{\epsilon^2}$ avec ϵ>0

alors lorsque n $\longrightarrow$V($\infty_{\overline{Kn)}\ \overline{\epsilon^2)=\frac{pq}{n\epsilon^2}\to 0}}$ et ceci $\forall\epsilon$>0

ainsi en limite $P(\left|\frac{K}{n}-p\right|\geq\epsilon)=0$ ce qui indique que $\frac{K}{n}\longrightarrow penprobabilit.$

Exemple :

On a prélevé au hasard, dans une population de lapin, 100 individus. Sur ces 100 lapins, 20 sont atteints par la myxomatose. Le pourcentage de lapins atteints par la myxomatose dans la population est donc :$\widehat{p}=\frac{K}{n}=\frac{20}{100}=0,2$ soit 20% de lapins atteins dans la population

Ce résultat n'aura de signification que s'il est associé à un intervalle de confiance.

Estimation par intervalle

Définition 4.2. *L'estimation par intervalle associe à un échantillon aléatoire, un intervalle* $\left[\widehat{\theta}_1,\widehat{\theta}_1\right]$ *qui recouvre* $\theta avecunecertaineprobabilit.$

Cet intervalle est appelé l'intervalle de confiance du paramètre $\theta carlaprobabilitque\theta dontlavaleurestinconn$ *et* $\widehat{\theta}_1 est\ égale\ à\ 1\text{-}\alpha, lecoefficientdeconfiance P(\widehat{\theta}_1<\theta<\widehat{\theta}_1)=1-\alpha$

Son complément $\alpha correspondaucoefficientderisque. P(\theta\notin\left[\widehat{\theta}_1,\widehat{\theta}_1\right])=\alpha$

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 comme l'indiquent les graphes ci-dessous. Pour chaque graphe, l'aire hachurée correspond au coefficient de risque $\alpha. Ainsidepartetd'autredeladistribution, lav$ $\frac{\alpha}{2}$.

.

Voici pourquoi :

Si $P(\overline{X}-i<\mu<\overline{X}+i)=1-\alpha$ alors $P(\mu-i<\overline{X}<\mu+i)=1-\alpha$

Connaissant la loi suivie par la v. a. X et d'après le théorème central limite, nous pouvons établir que $P(\frac{-i}{\sigma/\sqrt{n}}<\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}<\frac{i}{\sigma/\sqrt{n}})=1-\alpha$ sachant que $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\to N(0,1)$

par conséquent $\left|\frac{i}{\sigma/\sqrt{n}}\right|$ correspond à la valeur de la variable normale réduite pour la probabilité $\alpha donnenote\epsilon_\alpha$ ou écart réduit

ainsi $\frac{X-\mu}{\sigma/\sqrt{n}}=\epsilon_\alpha$ implique $i=\epsilon_\alpha\times\sigma/\sqrt{n}$

L'intervalle de confiance de la moyenne $\mu pouruncoefficientderisque\alpha estdonc\overline{X}-\epsilon_\alpha\frac{\sigma}{\sqrt{n}}<\mu<\overline{X}+\epsilon_\alpha\frac{\sigma}{\sqrt{n}}$ quelque soit la valeur de n si $X\to N(\mu,\sigma)$ et la variance σ^2 est connue

Exemple :

Pour des masses comprises entre 50g et 200g, une balance donne une pesée avec une variance de 0,0015. Les résultats des trois pesées d'un même corps sont : 64,32 ; 64,27 ; 64 ,39.

On veut connaître le poids moyen de ce corps dans la population avec un coefficient de confiance de 99%. avec $\overline{X} = 64{,}33$g et $\epsilon_\alpha = 2,576$ alors $\varepsilon_\alpha \frac{\sigma}{\sqrt{n}} = 2,576 \times \frac{0{,}039}{1{,}732} = 0,058$

et donc $\mu = X \pm \epsilon_\alpha \frac{\sigma}{\sqrt{n}} = 64,33g \pm 0,058$

d'où le poids moyen de ce corps est compris dans l'intervalle [64,27 ; 64,39] avec une probabilité de 0,99.

Remarque : La valeur de ϵ_α est donnée par la table de l'écart-réduit pour une valeur $\alpha donne$.

• Quelque soit la valeur de n, si $X \to N(\mu, \sigma)$ et σ^2 est inconnue,

Le raisonnement reste le même mais la variance de la population σ^2 doit être estimée par $\widehat{\sigma}^2 = \frac{n}{n-1} s^2$

Si $P(\overline{X} - i < \mu < \overline{X} + i) = 1 - \alpha$ alors $P(\mu - i < \overline{X} < \mu + i) = 1 - \alpha$

Connaissant la loi suivie par la v. a. $\overline{X}$ et celle suivie par la variable centrée réduite, on peut établir que $P(\frac{-i}{\widehat{\sigma}/\sqrt{n}} < \frac{\overline{X}-\mu}{\widehat{\sigma}/\sqrt{n}} < \frac{i}{\widehat{\sigma}/\sqrt{n}}) = 1 - \alpha$ sachant que $\frac{\overline{X}-\mu}{\widehat{\sigma}/\sqrt{n}} \to T(n-1 \ ddl)$

par conséquent

$\left|\frac{i}{\widehat{\sigma}/\sqrt{n}}\right|$ correspond à la valeur de la variable de student pour une valeur de probabilité $\alpha donnenotet_\alpha$ pour $n-1$ degrés de liberté.

Ainsi $\left|\frac{i}{\widehat{\sigma}/\sqrt{n}}\right| = \mathrm{t}_\alpha$ implique $i = t_\alpha x \widehat{\sigma}/\sqrt{n} i$

L'intervalle de confiance de l'espérance μ pour un coefficient de risque α est donc

$\overline{X} - t_\alpha \widehat{\sigma}/\sqrt{n} < \mu < \overline{X} + t_\alpha \frac{\widehat{\sigma}}{\sqrt{n}}$ quelque soit la valeur de n si $X \to N(\mu, \sigma)$ et σ^2 est inconnue

Remarque :

Lorsque $n > 30$, la loi de student converge vers une loi normale réduite. Ainsi la valeur de $t_\alpha(n-1)$ est égale à ϵ_α . Ci-dessous, un exemple pour un risque $\alpha = 0,05$.

4.2.4 quelques méthodes d'estimation

Les diverses méthodes permettent d'obtenir des estimateurs de qualités différentes

La méthode de maximum de vraisemblance

Définition 4.3. *La statistique* $w \longmapsto \arg max(\theta \longmapsto \prod_{i=1}^{n} f_\theta(X_i(w)))$ *s'appelle l'estimateur de maximum de vraisemblance de* θ.
$L : \theta \longmapsto \prod_{i=1}^{n} f_\theta(x_i)$ *s'appelle la fonction vraisemblance du modèle.*

$l : \theta \longmapsto \sum_{i=1}^{n} \log f_\theta(x_i)$ *s'appelle la fonction log-vraisemblance du modèle.*

En pratique, on fait l'étude de l'une des fonctions L *ou* l. *Il n'y a pas forcément unicité. Ces fonctions ne sont pas nécessairement dérivables ce qui annule le gradient ne réalise pas forcément un maximum .*

Remarque 7. *L'estimateur de maximum de vraisemblance n'existe pas toujours et n'est pas toujours unique.*

Exemple 4.4. *Le modèle de la loi exponentielle*
$\Theta = \mathbb{R}^+,\ f_\theta(x) = \theta e^{-\theta x}$ *on a*

$$L(\theta) = \prod_{1}^{n} f_\theta(x) = \theta^n e^{-\theta \sum_{i=1}^{n} x_i}$$

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

$$\frac{\partial l(\theta)}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^{n} x_i = 0 \Longleftrightarrow \hat{\theta} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{X}_n}$$

L est une application concave car on a

$$\frac{\partial^2 L(\theta)}{\partial^2 \theta} = -\frac{n}{\theta}$$

Donc, $\hat{\theta} = \frac{1}{\bar{X}}$ *est l'estimateur de maximum de vraisemblance dans le cas d'un modèle de la loi exponentielle .*

La méthode des moments

L'idée de base est d'estimer une espérance mathématique par une moyenne empirique, une variance par une variance empirique, etc...
Si la loi des X_i a deux paramètres θ_1 et θ_2 tels que $(\mathbb{E}(X), Var(X)) = \varphi(\theta_1, \theta_2)$, où φ est une fonction inversible, alors les estimateurs de θ_1 et θ_2 par la méthode des moments sont : $(\hat{\theta}_{1n}, \hat{\theta}_{2n}) = \varphi^{-1}(\bar{X}_n, S_n^2)$.
Ce principe peut naturellement se généraliser aux moments de tous ordres, centrés ou non centrés : $\mathbb{E}[(X - \mathbb{E}(X))^k]$, et $\mathbb{E}(X^k), k \geq 1$.

Exemple 4.5. *La loi Gamma*
Si $X_1, ..., X_n$ *sont indépendantes et de même loi gamma* $G(\alpha, \lambda),\ \mathbb{E}(X) = \frac{\alpha}{\lambda}$ *et* $Var(X) = \frac{\alpha}{\lambda}^2$. *On en déduit facilement que*

$$\lambda = \frac{\mathbb{E}(X)}{Var(X)} \text{ et } \alpha = \frac{[\mathbb{E}(X)]^2}{Var(X)}$$

Donc les EMM de α et λ sont

$$\hat{\lambda}_n = \frac{\bar{X}_n}{S_n^2} \text{ et } \hat{\alpha}_n = \frac{\bar{X}_n^2}{S_n^2}$$

Remarque 8. *Dans certains cas, l'estimation par la méthode des moments est moins bonne que l'estimation par maximum de vraisemblance. Néanmoins, dans le cas de la loi Gamma par exemple, le calcul de la fonction de vraisemblance peut poser des problèmes (l'utilisation de l'ordinateur et d'algorithmes numériques est indispensable) tandis que l'estimation des moments est très facilement accessible.*
Lorsque la taille de l'échantillon n'est pas suffisamment grande, la loi des grands nombres ne s'applique pas et par conséquent, les moments empiriques n'approchent pas suffisamment les moments théoriques.

Chapitre 5

LES TESTS STATISTIQUES

5.1 Introduction

Un test statistique est appelé à dégager un résultat significatif au milieu d'un ensemble de données expérimentales aléatoires.

La méthodologie des tests consiste à répondre à l'aide de résultats expérimentaux à une question concernant les paramétres de la loi de probabilité des variables aléatoires.

Quatre conditions préalables au calcul d'un test doivent être réunies :

– la question doit être posée de telle sorte qu'il n'y ait que deux réponses possibles : oui et non ;

– on doit avoir des données chiffrées résultant d'un échantillon ou d'une expérimentation ;

– ces données doivent pouvoir être considérées comme la réalisation de variables aléatoires dont la forme de la loi de probabilité est connue ;

– la question doit concerner un ou plusieurs paramétres de cette loi.

Une fois posée cette dernière, la réponse du test est :

– soit l'acceptation de l'hypothèse, ce qui signifie que les donneés ne sont pas en contradiction avec l'hypothèse ;

– soit le rejet de cette hypothèse, ce qui signifie qu'il est très peu probable d'obtenir les résultats que l'on a trouvés si l'hypothèse est vraie, ou encore que les données sont en contradiction avec elle.

En un sens, le test d'hypothèse est une généralisation probabiliste du raisonnement par l'absurde, mais alors que ce dernier met en contradiction logique deux affirmations formelles,

le premier oppose une affirmation formelle (l'hypothèse) avec des résultats du monde réel (les résultats de l'expérience).

De plus, le premier ne donne pas une certitude logique (l'hypothèse est fausse), mais seulement une forte présomption mesurée par une probabilité.

Enfin les deux formes du raisonnement ont en commun qu'elles ne peuvent que prouver (ou donner une présomption de preuve de) la fausseté de l'hypothèse et non sa vérité : ce n'est que parce qu'une expérience ne conduit pas au rejet de l'hypothèse que cette dernière est vraie : on peut imaginer d'autres expériences qui pourraient peut-être la rejeter.

Remarque 9. *Il s'agira de prendre une décision ; Elle consistera à accipter ou non une hypothèse de départ, formulée soit à partir de connaissances théhoriques soit à partir de présomptions suggérées par le ou les échantillons étudiés. Cependant, on ne pourra jamais conclure avec une certitude absolue, puisque la base de l'information qui perment de mener un test statistique provient de sous-ensembles de la population sur laquelle sont formulées les hypothèses.*

Remarque 10. *Il faudra donc se fixer un certain risque d'erreur qui n'est autre que la probabilité de se tromper en prenant la décision retenue.*

5.2 Principe général

5.2.1 L'interprétation statistique

Dans les chapitres précédent on a pu tirer un certains nombre de conclusins à partir d'un nombre limité d'observations. Ces conclusions ont permis d'estimer certains caractéristiques inconnues de la population.

Dautres méthodes, regroupées sous la dénomination de tests statistiques qui constituent la théorie de la décision, vont permettre de résoudre des problèmes pratiques tels que :

les tests de nouvelles thérapeutiques,

les comparaisons de méthodes de cultures,

les tests d'emploi de tel ou tel engrais.

5.2.2 La formulation des hypothèses

Un test statistique est un mécanisme qui permet de trancher entre deux hypothèses à partir de résultats observés sur un ou plusieurs échantillons.

Soit H_0 et H_1 ces deux hypothèses. La première appelée **hypothèse nulle**, joue un rôle particulier ; elle prétendra que les différences observées entre valeurs calculées et valeurs

théoriques sont dûes au hasard. Si on doit rejeter l'hypothèse nulle H_0, on dira que les écarts obsevés sont significatifs et on choisira H_1 appelée **hypothèse alternative.** Les tests statistiques permettent de retenir ou de rejeter H_0 qui est la seule hypothèse testée et celle qui permet les calculs pour conduire à la conclusion.

On a

H_0 vraie et H_1 fausse

ou

H_0 fausse et H_1 vraie

Il ya 4 solutions dont seulement les deux premières son justes :

a)-H_0 est vraie et on a choisi H_0

b)-H_0 est fausse et on a rejeté H_0

c)-H_0 est vraie et on a rejeté H_0

d)-H_0 est fausse est on a choisi H_0

5.2.3 Le risque d'erreur

Soit un test qui aboutit à chgoisir H_0 ou H_1. Seule une de ces deux hypothèses est vraie et on peut résumer les différents cas de décision et de validité de cette décision par le tableau suivant :

Hypothèse retenueHypothèse vraie	H_0	H_1
H_0	1-α	β
H_1	α	1-β

De ce tableau on tire les définitions suivantes :

Le risque de première espèce α

On appelle risque de première espèce et on note α la probabilité de rejeter l'hypothèse nulle H_0 alors qu'elle sest vraie.

Dans la pratique des tests statistiques, il est d'usage de choisir α a priori $\alpha = 1\%$ ou 5% dans la plupart des cas, cette probabilité est aussi appelée seuil de signification du test.

Le risque de deuxième espèce

β

On appelle risque de seconde espèce et on note β la probabilité d'accipter l'hypothèse nulle H_0 alors qu'elle sest fausse.

α étant fixé, β est déterminé par un calcul de pribabilité si H_1 est précisément définie.

On appelle puissance du test la probabilité $(1 - \beta)$ de rejeter H_0 en ayant raison.

5.3 Les différents types de tests

1-Les tests de conformité

2-les tests de comparaison

3-Les tests d'ajustement à une loi théorique

4-les tests d'indépendance

5.3.1 Les tests de conformité

Dans cette partie nous traiterons un premier typpe de test d'hypothèse en nous limitant au cas des grands échantillons (en pratique des échantillons de taille $n \geq 50$).

Nous disposons d'une distribution statistique expérimentale se présentant sous la forme d'un tableau d'effectifs ou des fréquences du caractère étudié.

Nous voulons savoir si ces effectifs ou ces fréquences sont compatible avec une distribution théorique connue. Il s'agit de déterminer si les différences constatées entre la distribution théorique et la

distribution expérimentale sont liées à la constitution de l'échantillon ou si elle sont trop importantes.

1.3.1.1 - Etude des moyennes

-Test bilatéral

Nous nous proposon d'étudier la conformité d'un échantillon par rapport à une norme préalablement définie.

a- position du problème

Dans un laboratoire pharmaceutique, une machine automatique fabrique en grande quantité des suppositoires contenant du paracétamol.

On désigne par X la variable aléatoire, qui à tout suppositoire pris au hasard dans la production, associe la masse (en mg) de paracitamol qu'il contient.

On admet que X suit la loi normale de moyenne m et d'écart-type $\sigma = 0.8$.

On veut contrôler la qualité de fabrication sur une pé riode donnée. Dans ce but, pendant le fonctionnement de la machine, on prélève d'un temps à l'autre un suppositoire dont on mesure la masse du paracétamole. On constitue ainsi un échantillon de 100 suppositoires. Les tirages sont supposéss indépendants.

On se propose de construire un test bilatéral permettant d'accepter ou de refuser, au seuil de signification de 5%, l'hypothèse selon laquelle la masse moyenne de paracétamol contenue dans un suppositoire est égale à 170 mg.

l'hypothèse nulle H_{00} est m=170 mg et lhypothèse alternative est H_1 $m \neq 170$ mg.

1- Sou H_0 quelle est la loi de la variable alétoire $\overline{X}$? préciser ces paramètres.

2- Enoncer clairement la règle de décision du test

3- Les résultats des mesures de l'échantion prélevé sont donnés dans le tableau :

Masse (mg)	$[145; 155[$	$[155; 165[$	$[165; 175[$	$[175; 185[$	$[185; 195[$
Effectifs	7	30	43	16	4

Peut-on accepter l'hypothèse H_0 au seuil de signification de 5% ?.

b - Lois d'échantillonnage

Puisque $n \geq 30$, le théhrème de la limite centrée nous permet de dire que la variable aléatoire $\overline{X}$ qui à chaque échantillon de taille n associe sa moyenne, suit approximativement la loi normale $N(u; \frac{\sigma}{\sqrt{n}})$. Alors la variable aléatoire $T = \frac{\overline{X} - u}{\frac{\sigma}{\sqrt{n}}}$ suit la loi normale centrée réduite.

c- Construction d'un test bilatéral :

l'hypothèse nulle $(H)_0$ est m=170 mg et l'hypothèse alternative est (H_1) m$\neq$ 170mg

e- Règle de décision :

Fixon, 'a priori, le risque maximal que nous acceptons de prendre en refusant H_0 alors qu'elle est vraie. Ce risque dit de première espèce, et noté α.

Puisque T suit la loi normale centrée réduite, il existe un unique réel strictyement positif t_α tel que :$p\,|T| > t_\alpha = \alpha . t_\alpha = \Pi^{-1}\left(1 - \frac{\alpha}{2}\right)$

Si $|T| > t_\alpha$ on rejette H_0 avec le risque α de se tromper.

Si $|T| \leq t_\alpha$ on accepte H_0 avec le risque de se tromper.(risque β de seconde espèce non quantifié).

Application numérique

Sous l'hypothèse H_0 la variable aléatoire $\overline{X}$ suit la loi normale $N(170; 0,8)$ donc la variable aléatoire $T = \frac{\overline{X} - 170}{0,8}$ suit la loi normale centrée réduite.

au seuil de risque $\alpha = 0,05$ on rejette H_0E si $|T| > 1,96$.

Pour l'échantillon proposé, en utilisant les centres des classes, on trouve $\overline{x}$= 168

On en déduit $t = -2,5$ donc $|t| > 1,96$ et on rejette H_0 au risque de 5% de se tromper.

Test unilatéral droit l'hypothèse nulle (H_0) est $m = 170\ mg$ et l'hypothèse alternative est (H_1) $m \geq 170\ mg$

La démarche ne diffère du précédente que sur deux points :

Hypothèse alternative H_1 : est selon le problème posé $m > 170$ ou $m < 170$.

Le risque α n'est plus symétriquement répartie.

Pour fixer les idées, supposons que l'hypothèse alternative H_1 : est $m > 170$ alors T est nécessairement positve.

Il existe un unique réel strictement positif u_α tel que $P(t > u_\alpha) = \alpha$ ou, ce qui équivalent tel que $P(t > u_\alpha) = 1 - \Pi(u_\alpha)$.

On a donc $\Pi(u_\alpha) = 1 - \alpha$ soit $u_\alpha = \Pi^{-1}(1 - \alpha)$

La règle de décision en résulte :

Si T$> u_\alpha$ on rejette H_0 avec un risque α de se tromper

Si T$\leq u_\alpha$ on accepte H_0 avec un risque β (non quantifié) de se tromper..

1.3.1.2 Etude des fréquences

a - Position du problème On étudie ici un caractère quantitatif C et on dispose de deux grands échantillons indépendants

A d'effectif n_A , où la fréquence du caractère est f_A.

B d'effectif n_B , où la fréquence du caractère est f_B.

A quelles condition peut-on conclure, qu'à un risque doné, ces deux échantillons proviennent de la même population ?

b - Lois d'échantillonnage Supposons que l'échantillon A provienne de la population P , où la fréquence du caractère C et p.

Supposons que l'échantillon B provienne de la population P′, où la fréquence du caractère C et p′.

On sait que si $N_A \geq 30$, La variable aléatoire F_A qui à tout échantillon de taiile n_A associe la fréquence f_A du caractère C dans cette échantillon suit apprximativement la loi normale $N(p; \sqrt{\frac{p(1-p)}{n_A}})$

Même si $N_B \geq 30$, La variable aléatoire F_B qui à tout échantillon de taiile n_B a fréquence f_B du caractère C dans cette échantillon suit apprximativement la loi normale $N(p'; \sqrt{\frac{p'(1-p')}{n_B}})$

Les variables alétoires F_A et F_B etant indépendantes et La variable aléatoire $F_A - F_B$ suit apprxi-

mativement la loi normale $N(p-p';\sqrt{\frac{p(1-p)}{n_A}+\frac{p'(1-p')}{n_B}})$.

c- Tests d'hypothèse bilatéral

1)- Hypotèse à tester Nous nous proposons de tester l'hypothèse nulle, notée H_0 "p et p' ne sont pas significativement différentes"

2)- Hypothèse alternative H_1 : le test étant bilateral H_1 est p et p' sont significativement différentes"

d - Règle de décision : Sous l'hypothèse H_0, la variable alétoire $F_A - F_B$ suit apprximativement la loi normale $N(p-p';\sqrt{\frac{p(1-p)}{n_A}+\frac{p'(1-p')}{n_B}})$.

Donc la variable alétoire $T=\dfrac{F_A-F_B}{\sqrt{p(1-p)(\frac{1}{n_A}+\frac{1}{n_B})}}$ suit apprximativement la loi normale $N(0;1)$.

Fixons alors un seuil de risque α (donc un seil de confiance $1-\alpha$), on sait qu'il existe un réel unique t_α strictement positif tel que $P(|T|\leq t_\alpha)=1-\alpha$

$P(|T|\leq t_\alpha)=1-\alpha$ é quivaut à $t_\alpha=\Pi^{-1}\left(1-\dfrac{\alpha}{2}\right)$

La règle de décision du test en résulte :

Si $|T|\leq t_\alpha$ on a aucune raison de rejeter H_0 donc on l'accepte.avec un risque β (non contifié) de se tromper

Si $|T|>t_\alpha$ on rejette H_0 un risque α de se tromper

e -Mise on oeuvre du test : $t=\dfrac{|f_A-f_B|}{\sqrt{p(1-p)(\frac{1}{n_A}+\frac{1}{n_B})}}$

On compare le nombre t avec t_α et on utilise la règle de décision pur conclure.

En général p est inconnu et, sous l'hypothèse (H_0) on réunit les deux échantillons.

Alors en estime p par $\widehat{p}=\dfrac{n_Af_A-n_Bf_B}{n_A+n_B}$.

e - Tests d'hypothèse unilatéral

La démarche ne diffère du précédente que sur deux points :

Hypothèse alternative H_1 : est selon le problème posé $p > p'$ ou $p < p'$.

Le risque α n'est plus symétriquement répartie.

Pour fixer les idées, supposons que l'hypothèse alternative H_1 : est $p < p'$ alors T est nécessairement négative.

Il existe un unique réel strictement positif v_α tel que $P(t < -v_\alpha) = \alpha$ ou, ce qui équivalent tel que .

$1 - \Pi(v_\alpha) = \alpha$ On a donc $v_\alpha = \Pi^{-1}(1 - \alpha)$

La règle de décision en résulte :

Si $T < -v_\alpha$ on rejette H_0 avec un risque α de se tromper.

Si $T \geq u_\alpha$ on accepte H_0 avec un risque β (non quantifié) de se tromper.

Test unilatéral gauche l'hypothèse nulle (H_0) est $m = 170\ mg$ et l'hypothèse alternative est (H_1) $m \leq 170\ mg$

Règle de décision : Fixon, à priori, le risque maximal que nous acceptons de prendre en refusant H_0 alors qu'elle est vraie. Ce risque dit de première espèce, et noté α.

Puisque T suit la loi normale centrée réduite, il existe un unique réel strictyement positif t_α tel que : $p(|T| > t_\alpha) = \alpha$. $t_\alpha = \Pi^{-1}\left(1 - \frac{\alpha}{2}\right)$

Si $|T| > t_\alpha$ on rejette H_0 avec le risque α de se tromper.

Si $|T| \leq t_\alpha$ on accepte H_0 avec le risque de se tromper.(risque β de seconde espèce non quantifié.

Application numérique

Sous l'hypothèse H_0 la variable aléatoire $\overline{X}$ suit la loi normale $N(170; 0,8)$ donc la variable aléatoire $T = \dfrac{\overline{X} - 170}{0,8}$ suit la loi normale centrée réduite.

au seuil de risque $\alpha = 0,05$ on rejette H_0 si $|T| > 1,96$.

Pour l'échantillon proposé, en utilisant les centres des classes, on trouve $\overline{x} = 168$

On en déduit $t = -2,5$ donc $|t| > 1,96$ et on rejette H_0 au risque de 5% de se tromper.

2- Test de conformité d'une fréquence

a - Position du problème

Dans la population algerienne, 15 personnes sur 100 ont un facreur rhésus négatif. Un laboratoire d'analyses médicales a contrôlé le facteur rhésus de 459 personnes. Il aconstaté que 75 d'entre elles avaient un facteur rhésus négatif.

Construire un test bilatéral permittant de dire, aua risque de 5%, si ce résultat est compatible, ou non, avac la norme dans la population algérienne.

b - Construction d'un test l'hypothèse nulle (H_0) est : le résultat est compatible, ou non, avac la norme dans la population algérienne.

Le test etant bilatéral :

l'hypothèse alternative est (H_1) est : "le résultat est significativement différent de la norme habituelle".

c - Lois d'échantillonnage Puisque $n \geq 30.$, le théhrème de la limite centrée nous permet de dire que la variable aléatoire F qui à chaque échantillon de taille n associe la fréquence du caractère dans cet échantillon, suit approximativement la loi normale $N(0,15; 0,017)$. Alors la variable aléatoire $T = \dfrac{F - 0,15}{0,017}$ suit la loi normale centrée réduite.

d - Rgle de dcision (condition de rejet) on rejette H_0 au risque $\alpha = 0,05$ si $|T| > 1,96$

Mise on oeuvre du test : $t = \dfrac{f - p}{\sqrt{\dfrac{p(1-p)}{n}}} = \dfrac{1,63 - 0,15}{0,017} \approx 0,76 < 1,96$ (on ne peut pas rejeter H_0).

Conclusion : on constat que lerésultat obtenu est conforme à la norme dans la population avec un risque de de 5% de se tromper.

5.3.2 les tests d'homogénéité (grands échantillons)

Nous disposons de deux échantillons indépendants donnés sous la forme d'un tableau d'éffectifs ou de fréquencesdu caractère étudié.

Nous désirons savoir si les différences observées sur la moyenne ou sur le fréquence sont dues u'niquement au hasard de l'échantillonage ou si elle sont trop importantes est doivent être atribuées à d'autres causes.

1.3.2.1.Etude des moyennes

a - Position du problème On étudie ici un caractère quantitatif C et on dispose de deux grands échantillons indépendants

A d'effectif n_A , de moyenne m_A et d'écrt-type σ_A

B d'effectif n_B , de moyenne m_B et d'écrt-type σ_B

A quelles condition peut-on conclure, qu'à un risque doné, ces deux échantillons proviennent de la même population ?

b - Lois d'échantillonnage Supposons que l'échantillon A provienne de la population P , d'effectif N, de moyenne u et d'écart-type σ.

Supposons que l'échantillon B provienne de la population P$'$, d'effectif N$'$, de moyenne u$'$ et d'écart-type σ'.

On sait que si $N_A \geq 30$, La variable aléatoire $\overline{X}$ qui à tout échantillon de taiile n_A associe sa moyenne m_A suit apprximativement la loi normale $N(u; \frac{\sigma}{\sqrt{n_A}})$.

Même si $N_B \geq 30$, La variable aléatoire $\overline{X}$ qui à tout échantillon de taiile n_B associe sa moyenne m_B suit apprximativement la loi normale $N(u'; \frac{\sigma'}{\sqrt{n_B}})$

Les variables alétoires $\overline{X}_A$ et $\overline{X}_B$ etant indépendantes et La variable aléatoire $\overline{X}_A - \overline{X}_B$ suit approximativement la loi normale $N(u - u'; \sqrt{\frac{\sigma^2}{n_A} + \frac{\sigma'^2}{n_B}})$.

Tests d/hypothèse bilatéral

a - Hypotèse à tester Nous nous proposons de tester l'hypothèse nulle, notée H_0 "u et u' ne sont pas significativement différentes"

b - Hypothèse alternative H_1 : le test étant bilateral H_1 est u et u' sont significativement différentes"

c - Règle de dcision : Sous l'hypothèse H_0, la variable alétoire $\overline{X}_A - \overline{X}_B$ suit approximativement la loi normale $N(0; \sqrt{\frac{\sigma^2}{n_A} + \frac{\sigma'^2}{n_B}})$.

Donc la variable alétoire $\dfrac{\overline{X}_A - \overline{X}_B}{\sqrt{\frac{\sigma^2}{n_A} + \frac{\sigma'^2}{n_B}}}$ suit approximativement la loi normale $N(0;1)$.

Fixons alors un seuil de risque α (donc un seil de confiance $1-\alpha$), on sait qu'il existe un réel unique t_α strictement positif tel que $P(|T| \leq t_\alpha) = 1 - \alpha$

$P(|T| \leq t_\alpha) = 1 - \alpha$ é quivaut à $t_\alpha = \Pi^{-1}\left(1 - \dfrac{\alpha}{2}\right)$

La règle de décision du test en résulte :

Si $|T| \leq t_\alpha$ on a aucune raison de rejeter H_0 donc on l'accepte.avec un risque β (non contifié) de se tromper

Si $|T| > t_\alpha$ on rejette H_0 un risque α de se tromper

d - Mise on oeuvre du test : $t = \dfrac{m_A - m_B}{\sqrt{\frac{\sigma^2}{n_A} + \frac{\sigma'^2}{n_B}}}$

On compare alors $|t|$ avec t_α et on utilise la règle de décision pur conclure.

En général σ et σ' sont inconnus et remplacés dans cette formule par $\widehat{\sigma}_A = \sigma_A\sqrt{\frac{n_A}{n_A - 1}}$ et $\widehat{\sigma}_B = \sigma_A\sqrt{\frac{n_B}{n_B - 1}}$

Tests d/hypothèse unilatéral

La démarche ne diffère du précédente que sur deux points :

Hypothèse alternative H_1 : est selon le problème posé $u > u'$ ou $u < u'$.

Le risque α n'est plus symétriquement répartie.

Pour fixer les idées, supposons que l'hypothèse alternative H_1 : est $u > u'$ alors T est nécessairement positve.

Il existe un unique réel strictement positif u_α tel que $P(t > u_\alpha) = \alpha$ ou, ce qui équivalent tel que $P(t \leq u_\alpha) = 1 - \alpha$.

On a donc $\Pi(u_\alpha) = 1 - \alpha$ soit $u_\alpha = \Pi^{-1}(1 - \alpha)$

La règle de décision en résulte :

Si T$\leq u_\alpha$ on accepte H_0 avec un risque β (non quantifié) de se tromper.

Si T$> u_\alpha$ on rejette H_0 avec un risque α de se tromper.

Etude des fréquences

Position du problème On étudie ici un caractère quantitatif C et on dispose de deux grands échantillons indépendants

A d'effectif n_A , où la fréquence du caractère est f_A.

B d'effectif n_B , où la fréquence du caractère est f_B.

A quelles condition peut-on conclure, qu'à un risque doné, ces deux échantillons proviennent de la même population ?

Lois d/échantillonnage

Supposons que l'échantillon A provienne de la population P , où la fréquence du caractère C et p.

Supposons que l'échantillon B provienne de la population P′, où la fréquence du caractère C et p′.

On sait que si $N_A \geq 30$, La variable aléatoire F_A qui à tout échantillon de taiile n_A associe la fréquence f_A du caractère C dans cette échantillon suit apprximativement la loi normale $N(p; \sqrt{\frac{p(1-p)}{n_A}})$

Même si $N_B \geq 30$, La variable aléatoire F_B qui à tout échantillon de taiile n_B a fréquence f_B du caractère C dans cette échantillon suit apprximativement la loi normale $N(p'; \sqrt{\frac{p'(1-p')}{n_B}})$

Les variables alétoires F_A et F_B etant indépendantes et La variable aléatoire $F_A - F_B$ suit apprximativement la loi normale $N(p - p'; \sqrt{\frac{p(1-p)}{n_A} + \frac{p'(1-p')}{n_B}})$.

Tests d/hypothèse bilatéral

Hypotèse à tester Nous nous proposons de tester l'hypothèse nulle, notée H_0 "p et p' ne sont pas significativement différentes"

Hypothèse alternative H_1 : le test étant bilateral H_1 est p et p' sont significativement différentes"

Règle de décision :

Sous l'hypothèse H_0, la variable alétoire $F_A - F_B$ suit apprximativement la loi normale $N(p - p'; \sqrt{\frac{p(1-p)}{n_A} + \frac{p'(1-p')}{n_B}})$.

Donc la variable alétoire $T = \frac{F_A - F_B}{\sqrt{p(1-p)(\frac{1}{n_A} + \frac{1}{n_B})}}$ suit apprximativement la loi normale $N(0;1)$.

Fixons alors un seuil de risque α (donc un seil de confiance $1-\alpha$), on sait qu'il existe un réel unique t_α strictement positif tel que $P(|T| \leq t_\alpha) = 1 - \alpha$

$P(|T| \leq t_\alpha) = 1 - \alpha$ é quivaut à $t_\alpha = \Pi^{-1}\left(1 - \frac{\alpha}{2}\right)$

La règle de décision du test en résulte :

Si $|T| \leq t_\alpha$ on a aucune raison de rejeter H_0 donc on l'accepte.avec un risque β (non contifié) de se tromper

Si $|T| > t_\alpha$ on rejette H_0 un risque α de se tromper

Mise on oeuvre du test : $t = \frac{|f_A - f_B|}{\sqrt{p(1-p)(\frac{1}{n_A} + \frac{1}{n_B})}}$ On compare le nombre t avec t_α et on utilise la règle de décision pur conclure.

En général p est inconnu et, sous l'hypothèse (H_0) on réunit les deux échantillons.

Alors en estime p par $\widehat{p} = \frac{n_A f_A - n_B f_B}{n_A + n_B}$.

Tests d/hypothèse unilatéral

La démarche ne diffère du précédente que sur deux points :

Hypothèse alternative H_1 : est selon le problème posé $p > p'$ ou $p < p'$.

Le risque α n'est plus symétriquement répartie.

Pour fixer les idées, supposons que l'hypothèse alternative H_1 : est $p < p'$ alors T est nécessairement négative.

Il existe un unique réel strictement positif v_α tel que $P(t < -v_\alpha) = \alpha$ ou, ce qui équivalent tel que .

$1 - \Pi(v_\alpha) = \alpha$ On a donc $v_\alpha = \Pi^{-1}(1 - \alpha)$

La règle de décision en résulte :

Si $T < -v_\alpha$ on rejette H_0 avec un risque α de se tromper.

Si $T \geq u_\alpha$ on accepte H_0 avec un risque β (non quantifié) de se tromper.

$\frac{1}{\sqrt{6.28}} \exp(-\frac{x^2}{2})$

5.4 Le cas des petits chantillons

Définition 5.1. *Soient X et Y deux variables aléatoires indépendantes suivant respectivement* $N(0,1)$ *et* $X^2(n)$. *On appelle loi de Student à n degrés de liberté la loi suivie par le rapport :* $T = \dfrac{X}{\sqrt{Y/n}}$, *cette loi est notée* T_n. $E(T_n) = 0 \quad (n > 1)\,; \;\; Var(T_n) = \dfrac{n}{n-2} \qquad (n > 2).$

Définition 5.2. *Soient X et Y deux variables aléatoires indépendantes suivant respectivement* $X^2(n)$ *et* $X^2(m)$.

La variable aléatoire $F = \dfrac{EX/n}{EY/m}$ *suit la loi de Fisher-Snedecor à n et m degrés de liberté notée* $F_{n,m}$

$$E(F_{n,m}) = \frac{1}{m-2} \qquad (m > 2); \;\; Var(T_n) = \frac{2m^2(n+m-2)}{n\,(m-4)\,(m-2)^2} \qquad (m > 4).$$

5.4.1 Test de Student

On pratique il est rare que l'on connaisse la valeur de σ ; on n'en connait q'une estimation s, valeur calculée de l'estimateur S. Que peut-on dire alors de la variable centrée réduite $\dfrac{\overline{X} - m}{S/\sqrt{n}}$?.

Sous réserve que le caractère étudié soit distribué dans la population selon la loi normale, on peut démontrer que ce rapport suit une loi de Student 0 $(n-1)$ degré de liberté et que cette loi converge rapidement vers la loi de Gauss lorsque n augmente, peut être remplacée par elle dès que $n \geq 30$.

On voit donc que pour les petits échantillons ($n < 30$), il faut faire appel à la loi de Student.La comparaison de moyennes à partir de petits échantillons (n_1 et / ou $n_2 < 30$) va elle aussi utiliser cette loi de Student.

Faisons l'hypothèse que les deux échantillons proviennent de populations de mêmes moyennes (il s'agit de l'hypothèse $m_1 = m_2 = m$) et qu'en outre ses populations sont normales et de mêmes variances($\sigma_1^2 = \sigma_2^2 = \sigma^2$), on peut démontrer que la quantité $t = \dfrac{|\overline{x}_1 - \overline{x}_2|}{\widehat{\sigma}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

où $\begin{cases} \widehat{\sigma}^2 = \dfrac{n_1\sigma_{e1}^2 + n_2\sigma_{e2}^2}{n_1 + n_2 - 2} \\ \overline{x}_i,\ \sigma_{ei}^2 \text{ moyenne et écart-type de l'échantillon numéro i} \end{cases}$

suit la loi de Student à $v = n_1 + n_2 - 2$ degrés de liberté. Il devient alors possible de déterminer une région d'acceptation de l'hypothèse nulle H_0 d'égalité des moyennes. Cette région dépend de l'hypothèse alternative H_1, dans le cas où H_1 est "$m_1 \neq m_2$", on mène un test bilatéral et la région d'acceptation de H_0 est donnée par l'ntervalle : $\left[-t_{v;\alpha}; + - t_{v;\alpha}\right]$, avec $v = n_1 + n_2 - 2$. où $t_{v;\alpha}$ désigne la valeur de la

loi de Student ayant la probabilité α d'être dépassée en valeur absolue.

Si $t < t_{v;\alpha}$ alors on accepte H_0.

Si $t > t_{v;\alpha}$ on rejette H_0 au seuil $\alpha\%$.

Remarque : Dans le cas d'une hypothèse alternative conduisant à mener un test unilatéral du type H_0 : "$m_1 = m_2$" contre H_1 : "$m_1 > m_2$" la région d'acceptation de H_0 est de la forme : $]-\infty; t_{2\alpha;v}[$.

.On sera souvent amené à tester de façon préalable l'égalité des variance à l'aide d'un test de Fisher-Snedecor avant de comparer les moyennes à ppartir de deux petits échantillons.

5.4.2 Test de Fisher-Snedecor

Comparaisons de deux variances

Le test de comparaison de deux variances σ_1^2 et σ_2^2 est basé sur le rapport des deux estimation s_1^2 et s_1^2 calculées à partir d'échantillons, de taille respective n_1 et n_2 extraits des deux population à comparer. Il n'est pas nécessaire que n_1 et n_2 soient grand mais il est impératif que les deux populations soient normalement distribuées.

On formule l'hypothèse H_0 : "$\sigma_1^2 = \sigma_2^2$" contre l'hypothèse H_1 : "$\sigma_1^2 \neq \sigma_2^2$" ce test est donc toujours bilatéral. On calcule la quantité : $F = \dfrac{s_1^2}{s_1^2}$ si $s_1^2 \geq s_2^2$.F est toujours supérieure ou égale à un.

La règle de décision est la suivante :

Si $F < F_{\frac{\alpha}{2};v_1;v2}$ on accepte H_0.

Si $F \geq F_{\frac{\alpha}{2};v_1;v2}$ on rejette H_0.au risque α.

5.5 Le test chi-deux

5.5.1 INTRODUCTION

Problème 1 :

Partant des races pures, un sélectionneur a croisé de mufliers ivoires avec des mufliers rouges, il a obtenu en $F1$ des mufliers pâles, puis en $F2$ après autofécondation des plantes de la génération F1 : 22 mufliers rouges, 52 mufliers pâles et 23 mufliers ivoires.

La couleur des fleurs est-elle gérée par un couple d'allèles ?.

Le test chi-deux est fréquemment utilisé par les biologistes. A la différence des autres test, ce test ne s'appuie pas sur un modèle probabiliste rigoureux, mais sur une loi asymptotique ; il est donc délicat à utiliser et il est parfois préférable de le remplacer, lorsque c'est possible, par un test non parmétrique plus adapté.

Le test du χ^2 est le plus célèbre des tests dits **non paramètriques** qui n'exigent aucune condition sur la distriution de la population mère. C'est un test globale qui porte sur l'ensemble des effectifs ou fréquences observées après expérience et calculés à partir de l'hypothèse testée. On pourra comparer :

Une distribution expérimentale et une distribution théorique. Les caractéristiques de cette distribution théorique sont connues ou estimées à partir des observations. Selon le cas, on parlera de test de **conformité ou d'ajustement** à une loi théorique.

Plusieurs distributions pour savoir si on peut accépter l'hypothèse qu'elles proviennent de la même population parente, dans ce cas on mènera un test **d'homogéneité** ou

d'indépendance. On a en fait généraliser le cas précédent en comparant chaque distribution empirique à une même distribution théorique.

Le mécanisme du test du χ^2 permet de savoir si les écarts constatés entre les distributions à comparer sont imputables ou non au hasard.

Définition 5.3. *Soit X une v.a de loi.N(0 ;1), alors la v.a X^2 est dite v.a de chi-deux à 1 degé de liderté.*

Définition 5.4. *Soient X_1, X_2,...,X_n n v.a indépendantes suivent toutes loi N(0 ;1), alors la v.a $Z = X_1^2 + X_2^2 + ... + X_n^2$ est une v.a de chi-deux à n degés de liderté, avec E(Z)=n et Var(Z)=2n*

Remarque 11. *Si Z suit la loi du χ^2 à n degrés de liberté, la table du chi-deux donne pour un risque α choisi, le nombre χ^2_α tel que*

$$P(Z \geqslant \chi^2_\alpha) = \alpha.$$

5.5.2 COMPARAISON ET AJUSTEMENT A UNE LOI THEORIQUE

Construction du test

On considère une distribution expérimentale donnée par un échantillon de taille n.

Les individus de cet échantillon son classés et on a dénombré la fréquence absolue ou effectif de chaque classe. On note n_i l'effectif observé pour la classe N° i. Si on connait (ou croit connaître) la loi théorique que suit cette distribution, on est alors capable de calculer les effectifs théoriques de chaque classe. En effet la loi théorique est connue dès lors que les probabilités attachées à chaque classe le sont. On note P_i la probabilité qu'un individu tiré au hasard appartienne à la classe N° i. L'effectif théorique associé est alors nP_i.

5.5.3 Application du test chi-deux

On expliquera d'abord les principes du test sur une loi multinomiale puis dans ses applications les plus courantes, la méthode non paramétrique qui en découle.

test sur une loi multinomiale

Distribution à deux classes.

Soit une expérience aléatoire E susceptible

d'entraîner la réalisation d'un évenement E_1 de probabilité $P(E_1)$, ou d'un évenement E_2 de probabilité $P(E_2)$, E_1 et E_2 formant un système complet c-à-d $P(E_1) + P(E_2) = 1$ et $P(E_1 \cap E_2) = 0$.

Soit un ensemble de n expériences identiques à E et indépendantes. On lui associe les variables X_1 et X_2 représentant respectivement le nombre d'événement de E_1 et de E_2 que l'on peut observer $(X_1 + X_2 = n)$, la réalisatin effective des n expérience entraîne

l'observation des valeurs x_1 de X_1 et x_2 de X_2 $(x_1 + x_2 = n)$, On dit que les résultats sont reparties en deux classes. On désire tester l'hypothèse H_0 "$P(E_1) = P_1$ et $P(E_2) = P_2$ contre l'hypothèse H_1

"$P(E_1) \neq P_1$ et $P(E_2) \neq P_2$".

Compte-tenue de la relation $P_1 + P_2 = 1$, il suffit de tester "$P(E_1) = P_1$" contre

"$P(E_1) \neq P_1$". Ce que l'on peut faire à l'aide de la variable $X_1 \rightarrow B(n, P_1)$.

X_1 admet pour loi asymptotique, lorsque n augmente indéfiniment, la loi

$$N(nP_1, nP_1(1-P_1)).$$

Alors un test avec la variable

$Y = \dfrac{X_1 - nP_1}{\sqrt{nP_1(1-P_1)}}$ considéré comme pratiquement normale centrée et réduite sou H_0.

Soit maintenant la variable

$Z = \dfrac{(X_1 - nP_1)^2}{nP_1} + \dfrac{(X_2 - nP_2)^2}{nP_2}$,

on a $Z = \dfrac{(X_1 - nP_1)^2}{nP_1(1-P_1)} = Y^2$

étant donné le comportement asymptotique de Y, il est clair que Z admet pour loi asymptotique la loi de χ_1^2 sous H_0.

Pour un niveau α on peut ecrire $1 - \alpha = P\left(-y_{\frac{\alpha}{2}} \leq Y \leq y_{\frac{\alpha}{2}}\right) = P\left(0 \leq Y^2 \leq y_{\frac{\alpha}{2}}^2\right) =$

$P\left(0 \leq Z \leq z_{\frac{\alpha}{2}}\right)$ avec $z_{\frac{\alpha}{2}} = y_{\frac{\alpha}{2}}^2$,

La borne supérieur de l'intervalle d'acceptation ($3.481 = (1.96)^2$ au niveau $5\%; 6.635 = (2.576)^2$ au niveau 1%) étant lue dans les tables de χ^2..

Distribution à r classes.

Plus généralement soit une expérience aléatoire E susceptible d'entraîner la réalisation de r évenements $E_1, E_2, ..., E_r$ de probabilité $P(E_1)$, $P(E_2), ..., P(E_r)$, E_1 E_2 $...E_r$,formant un système complet c-à-d $P(E_1) + P(E_2) + ... + P(E_r) = 1$ et $P(E_i \cap E_j) = 0$ *pour* $i \neq j$.

Les résultats de n expériences identiques à E et indépendantes sont donc réparties en r classes. A un tel ensemble d'expériences, On associe les variables X_1 X_2 $...X_r$ représentant respectivement les effectifs des classes que l'on peut observer,

Le système $(X_1, X_2, ..., X_r)$, forme un sytème multinomial, on veut tester l'hypothèse

"$p(E_1) = p_1$ et $p(E_2) = p_2$ *et...* $p(E_r) = p_r$"

contre l'hypothèse H_1 :

"$P(E_1) \neq P_1$ ou $P(E_2) \neq P_2$...ou $P(E_r) \neq P_r$"

En fait il n'y a parmi r variables que $(r-1)$ variables indépendantes ; En effet les variables sont liées par la relation $X_1 + X_2 + ... + X_r = n$, dès que le hasard attribué une valeur numérique à r-1 variables, la valeur de la dernière est imposée.

APPLICATION :

Problème 1 (solutions) :

Solution : Soient p_1, p_2, p_3 les probabilités pour q'une plante de la génération F2 ait respectivement des fleurs rouges, pâles ou ivoires, soient X_1, X_2 et X_3 les variables représentant les plantes à fleurs rouges, pâles ou ivoires que l'on peut observer sur 97 plantes.

On est amené à tester, après un raisonnement génétique élémentaire, l'hypothèse H_0 :

$p_1 = \frac{1}{4}$, $p_2 = \frac{1}{2}$, $p_3 = \frac{1}{4}$ contre l'hypothèse $H_1 : p_1 \neq \frac{1}{4}$ *ou* $p_2 \neq \frac{1}{2}$ *ou* $p_3 \neq \frac{1}{4}$.

D$'$où le tableau :

phénotypes	rouge	pâle	ivoir	total
probabilité	1/4	1/2	1/4	1
effectif théorique	24.25	48.5	24.25	97
effectif observé	22	52	23	97

-Les conditions d'application de χ^2 sont s'atisfaites, à savoir :

-Les classes constituent un systèmes complet

d'événements ;

-Les 97 expériences sont identiques et

indépendantes ;

-Leur nombre est assez élvé ;

-Les effectifs théoriques sont suffisament

élevés.

Dans ces conditions, sous H_0, la variable $Z = \sum_{i=1}^{3} \frac{(X_i - 97p_i)^2}{97p_i}$ est pratiquement une variable X^2, on effectue un test. L'intervalle d'acceptation de H_0 est, au niveau

5% : $[0\ ;\ 5,991]$.

On a observé la valeur $Z_0 = \frac{(2,25)^2}{24.25} + \frac{(3,50)^2}{48,5} + \frac{(1,25)^2}{24,25} \simeq 0.52$.

Conclusion :

Au niveau 5% on peut accepter l'hypothèse que La couleur des fleurs est gérée par un couple d'allèles.

5.5.4

Tests d/hmogénéité

Principe

Le test χ^2 est également utilisé pour la comparaison de plusieurs échantillons. Le principe du test va être exposé dans un exemple à deux échantillons. on le généralise sans peine pour plusieurs échantillons.

Problème 2 :

On a étudié sur deux échantillons prevenant de deux populations différentes la répartition des quatre groupes sanguins : O, A, B,AB les résultats obtenus sont réparties dans un tableau dit tableau de contingence, à deux lignes et à quatre colonnes :

Groupe	O	A	B	AB	tot
1^{er} éch	121	120	79	33	353
2^{em} éch	118	95	121	30	364
total	239	215	200	63	717

On veut tester l'hypotèse H_0 " les quatre groupes sanguins sont réparties de la même manière sur les deux populations"

contre l'hypothèse H_1 "les répartitions sont différentes".

Sous H_0. la probabilité, pour un individu prélevé au hasard, d'être d'un groupe donné est la même dans les deux populations, on ne connaît pas cette probabilité, sinon le problème serait résolu ; on peut cependant l'estimer et, toujours sous H_0. La meilleure estimation que l'on puise en donner est la proportion des indiviudus de ce groupe observée sur l'ensemble des deux échantillons. C'est ainsi que l'on obtient les estimations :

Pour le groupe O	$p_1 = 239/717 \simeq 0,333$
Pour le groupe A	$p_2 = 215/717 \simeq 0,300$
Pour le groupe B	$p_3 = 200/717 \simeq 0,249$
Pour le groupe AB	$p_4 = 63/717 \simeq 0,088$

$p_1+p_2+p_3+p_4=1$. La relation $p_1+p_2+p_3+p_4=1$ montre qu'en fait il suffit de trois paramètres pour déterminer complètement le modèle. On déduit de l'estimation précédente les éffectifs théoriques de chaque classe pour un échantillon de taille 353 d'une part et pour un échantillon de taille 364 d'autre part. D'où le tableau :

Groupe	O	A	B	AB	total
1^{er} éch	121 (117,7)	120 (105,9)	79 (98,5)	33 (31)	353
2^{em} éch	118 (121,3)	95 (109,1)	121 (101,5)	30 (32)	364
total	239	215	200	63	717

les effectifs théhoriques sont entre parenthèses, on a par exemple, 117=0,333.353.

Soient maintenant les variables X_1, X_2, X_3, X_4 représentant les effctifs des classes du premiers échantillon et Y_1, Y_2, Y_3, Y_4 représentant les effctifs des classes du deuxième échantillon.

On pose :

$$Z=\frac{(X_1-117,7)^2}{117,7}+\frac{(X_2-105,9)^2}{105,9}+$$
$$\frac{(X_3-98,5)^2}{98,5}+\frac{(X_4-31,0)^2}{31,0}+\frac{(Y_1-121,3)^2}{121,3}+$$
$$\frac{(Y_1-109,1)^2}{109,1}+\frac{(Y_1-101,5)^2}{101,5}+\frac{(Y_4-32,0)^2}{32}.$$

Les conditions d'application du test χ^2 étant satisfaites pour chaque échantillon, sous H_0, la variable Z peut être considérée comme la somme de deux variables χ^2, l'indépendance des deux séries d'observations permet de considérer la varible Z comme une variable χ^2. On est tenté de dire qu'il s'agit d'une variable χ^2 à $2(4-1)=6$ degrés de liberté ; cependant, l'estimation, à partir des observations des trois paramètres qui déterminent coplètement le modèle probabiliste baisse le nombre de degrés de liberté de 6 à 3. D'où $Z\to\chi^2_3$.

Les valeurs élevées de Z étant plus probables sous H_1 que sous H_0.

Au niveau 5% l'intervalle d'acceptation est $[0;\ 7,815]$, et comme $Z\simeq 11.74>7,815$ donc

on peut conclure au rejet de H_0.

C'st-à-dire les quatre groupe sanguins sont réparties différemment sur les deux populations d'où proviennent les deux échantillons. Même au niveau 1% on rejetterait H_0.

Chapitre 6

EXERCICES

6.1

SERIE DE TD N 1

EXERCICE 1

On considère la variable aléatoire X_n de loi de probabilité uniforme sur $\left\{0, \frac{1}{n}, ..., \frac{n-1}{n}, 1\right\}$.

Montrer que la suite de variables aléatoires $(X_n,\ n < 1)$ converge en loi ver la variable aléatoire X de loi uniforme sur le segment $[0,\ 1]$.

EXERCICE 2

Soit une suite $(X_n,\ n < 1)$ de variables aléatoires réélles définies sur un espace probabilisé $(\Omega,\ \Lambda,\ P)$, la loi de X_n étant donnée par $P\left(X_n = 1 - \frac{1}{n}\right) = \frac{1}{2} = P\left(X_n = 1 + \frac{1}{n}\right)$.

1 Montrer que la suite (X_n) converge en loi vers la variable aléatoire $X = 1$.

2 Est-ce-que pour tout x, $\lim_{n\to\infty} P(X_n = x) = P(X = x)$?.

3- Montrer que la suite (X_n) converge en probabilité vers 1.

4- (X_n) converge-t-elle en moyenne quadratique vers 1?.

5- (X_n) converge-t-elle presque sûrement vers 1?.

EXERCICE 3

Soit une suite $(X_n,\ n<1)$ de variables aléatoires réélles mutuellement indépendantes de même loi uniforme sur $[0,\ a]$, $a<0$.

1- Soit $S_n=\dfrac{X_1+X_2+...+X_n}{n}$. Etudier la suité (S_n) suivant différents modes de convergence.

2- Etudier la limite de la suité $\left(\sqrt{n}\left(S_n-\dfrac{a}{2}\right)\right)$.

3- Montrer que la suite $(M_n=\sup(X_1,,X_2,...,\ X_n))$ converge en loi, converge-t-elle enprobabilité ?.

4- Calculer la distance de Kolmogorove entre la fonction de répartition de la loi de M_n et celle de la variable aléatoire constante égale à a et déterminer sa limite quand $n\to\infty$.

EXERCICE 4 Lemme de Borel-Cantelli

Soit $(\Omega,\ \Lambda,\ P)$ un espace probabilisé et (A_n) une suite d'évènement de Λ.On définit l'évènement B " pour une infinité de n, A_n est réalisé".

1- On pose $B_n=\cup_{n\langle\rangle m}A_n$; montrer que $B=\lim_{n\to\infty}B_n$.

2- Montrer que si la série de terme général $P(A_n)$ est convergente, alors $P(B)=0$.

Exercice 5

Pour tout entier naturel n non nul, on considère la fonction f_n définie par

$f_n(x)=n^2x\exp(\frac{-n^2x^2}{2})1_{\mathbb{R}_+}(x)$.

Montrer que fnfn est la densité d'une variable aléatoire.

Soit $(X_n)_n$ une suite de variables aléatoires telle que, pour tout entier $n\geq 1$, X_n admet pour densité f_n. Démontrer que la suite (X_n) converge en probabilité vers une variable aléatoire X que l'on précisera.

Exercice 6

Soit (Un) une suite de variables aléatoires indépendantes suivant toutes la loi uniforme sur [0,1]. On note

$M_n=\max(U_1,\ldots,U_n)$ *et* $X_n=n(1-M_n)$.

Quelle est la fonction de répartition de X_n ?

Etudier la convergence en loi de la suite (X_n).

Indication : $\lim_{n\to\infty}\left(1-\dfrac{x}{n}\right)=\exp(x)$.

Exercice 7

On dit qu'une variable aléatoire Y suit une loi de Gumbel si elle admet pour densité $f(x) = e^{-x-e^{-x}}$.

Vérifier que f est une densité, et calculer la fonction de répartition de Y.

Soit $(X_n)_n$ une suite de variables aléatoires indépendantes identiquement distribuées de loi exponentielle de paramètre 1. On pose $M_n = max(X_1, \ldots, X_n)$. Démontrer que la suite $(M_n - \ln(n))$ converge en loi vers Y suivant une loi de Gumbel.

Indication : $\int_{-\infty}^{t} e^{-x-e^{-x}} dx = \left[e^{-e^{-x}}\right]_{-\infty}^{t}$.

6.2 SERIE DE TD N 2

Exercice 1 :Familles Exponentielles On considère les modè les suivants :

Modèle Binomial $\{B(m,p) : p \in [0,1]\}$;

Modèle de Poisson $\{P(\lambda) : \lambda > 0\}$;

Modèle gaussien à variance fixée $\{N(\mu, \sigma^2) : \mu \in \mathbb{R}\}$;

Modèle gaussien à paramètre bi-dimensionnel $\{N(n\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$;

Modèle Gamma

$$\{G(\alpha,\beta) : \alpha > 0,\ \beta > 0\} = \{f_{\alpha,\beta}(x) = \frac{\beta}{\alpha}\Gamma(\alpha)x^{\alpha-1}e^{-\beta x}1_{R+}(x) : \alpha > 0,\ \beta > 0\};$$

Modèle uniforme $\{U_{[0,\theta]} : \theta > 0\}$;

Modèle de Cauchy $\{f_\theta(x) = \frac{1}{\Pi(1+(x-\theta)^2)} : \theta \in R\}$; • Modèle multinomial $\{M(n, p_1, ..., p_k) : 0 < p_i < 1, \forall i = 1, ..., k \ et \sum_{i=1}^{k} p_i = 1\}$. Pour tous ces modèles, répondre aux questions suivantes.

1) Quelle est l'expression de la densité $f_{theta}(x)$?

2) Le modèle constitue-t-il une famille exponentielle générale ? Naturelle ? Quel est le paramètre canonique du modèle ?

3) Quelle est la vraisemblance d'un échantillon$x = (x_1, ..., x_n)$?

Exercice 2 : (Modèles position-échelle)

1) Construire un modèle position-échelle à partir de la loi exponentielle exp(1). Préciser la forme

des $f.d.r.$ des lois de ce modèle ainsi que leurs densités.

2) Montrer que le modèle uniforme $\{U_{[a,b]} : -\infty < a < b < +\infty\}$ est un modèle position-échelle.

Exercice 3 (Statistiques d'ordre)

Soit $X_1, ..., X_n$ des v.a.r. définies sur un même espace probabilisé (Ω, A, P), indépendantes et demême loi absolument continueparrapport à la mesure de Lebesgue de densité f. Pour tout ω dans Ω, on peut ordonner les réels $X_1(\omega), ..., X_i(\omega), ..., X_n(\omega)$ sous la forme $X_{(1)}(\omega) \leq X_{(2)}(\omega) \leq \cdots \leq X_{(i)}(\omega) \leq \cdots \leq X_{(n)}(\omega)$.

L'application $X_{(i)} : \omega \in \Omega \to X_{(i)}(\omega)$ ainsi définie pour chaque i est une v.a.r. dite ième statistique d'ordre.

1) Calculer la loi de $X_{(n)} = sup\{X_1, ..., X_n\}$ ($f.d.r.$ et densité).

2) Calculer la loi de $X_{(1)} = inf\{X1, ..., X_n\}$ ($f.d.r.$ et densité).

3) Calculer la loi du couple $(X_{(1)}, X_{(n)})$.

4) Soit N_y le nombre de X_i inférieurs à y. Quelle est la loi de N_y ? Que dire des événements $\{N_y \geq k\}$ et $\{X_{(k)} \leq y\}$? En déduire la $f.d.r.$ de $X_{(k)}$.

SERIE DE TD N 3

ÉNONCÉS

Exercice 1 (Statistiques exhaustives)

On considère les modèles suivants :
modèle de Poisson (N ; P(N) ; $P(\lambda) : \lambda > 0$) ;
modèle de la loi de exponentielle ($\mathbb{R}_+; B_{\mathbb{R}_+}, e(\lambda) : \lambda >0$);

modèle gaussien avec σ^2 positif connu : $(\mathbb{R}; B()\mathbb{R}); \mathbb{N}(\mu; \sigma^2 : \sigma^2 >0$);

modèle gaussien avec μ dans $\mathbb{R}$ connu : (R ; B(R) $\mathrm{N}(\mu; \sigma^2 : \sigma^2 >0$) :

modèle gaussien général : (R ; B(R) $\mathrm{N}(\mu; \sigma^2 : \mu \in \mathbb{R}, \sigma^2 >0$) :

1) Pour chacun de ces modèles donner l'expression d'une statistique exhaustive (éventuellement vectorielle).
2) Retrouver le résultat pour le modèle de Poisson en utilisant une autre méthode.

Exercice 2 : (Statistique exhaustive et Famille Exponentielle Générale) On considère une famille exponentielle générale de statistique canonique $T(X)$ où X est la variable générique dans ce modèle.

1) Montrer que $\sum_{i=1}^{n} X_i = \mathrm{T}(\mathrm{x}_i)$ est une statistique exhaustive pour le modèle d'échantillonnage associé.

2) En utilisant un résultat obtenu dans l'Exercice 1 série de TD N° 2, montrer que la moyenne empirique $\frac{1}{n}\sum_{i=1}^{n} X_i est une statistique exhaustive dans un modèle d'chantillonnage de la loi binomiale.$

Exercice 3 (Modèle Gamma et Méthode des moments) On considère le Modèle Statistique de la loi Gamma :
(R^+; B(R^+); $\mathrm{G}(\alpha, \beta) : \alpha >0$, $\beta >0$) :

On rappelle que la densité d'une v.a. X de loi $G(\alpha;\beta)$ est : $f_{(\alpha;\beta)}(x) = \dfrac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} 1_{\mathbb{R}^+}$.

1) Calculer $\mathrm{E}_{(\alpha;\beta)}(X)$ et $\mathrm{var}_{(\alpha;\beta)}(X)$

2) Par la méthode des moments, donner un estimateur du paramètre bidimensionnel $E_{(\alpha;\beta)}(X)$ du modèle, basé sur l'observation d'un échnatillon $X_1; ...; X_n$.

3) Déterminer des estimateurs de α et β en utilisant conjointement des estimateurs empiriques des moments et la méthode de substitution.

Exercice 4 (Modèle de la loi exponentielle et Méthode des moments) On a vu que la méthode des moments permet d'obtenir un estimateur du paramètre λ dans un modèle de la loi exponentielle :

$\lambda = 1/(\dfrac{1}{n}\sum_{i=1}^n X_i)$ basé sur la relation $\mathrm{E(X\ )} = \frac{1}{\lambda}$. L'intérêt de cet exercice est de montrer que cette méthode permet la construction de plusieurs estimateurs de ce même paramètre λ. 1) On suppose qu'une v.a.r. X suit une loi exponentielle $exp(\lambda)$. Calculer $E(X^2$

2) Soit $t_0 > 0$. Écrire la fiabilité $1 - F(t_0) = P(X > t_0)$ sous forme d'une espérance.

3) On considère le modèle de la loi exponentielle $(\mathbb{R}_+; B_{\mathbb{R}_+}, e(\lambda) : \lambda > 0)$;
En vous inspirant des résultats des deux questions précédentes et en utilisant à chaque fois la méthode des moments, proposer deux autres estimateurs du paramètre λ.

Exercice 5 (Maximum de vraisemblance pour un modèle gaussien) 1) On considère le modèle gaussien : $(\mathbb{R}; B(\mathbb{R})N(\mu;\sigma^2) : \mu \in \mathbb{R})$:
Donner l'estimateur du maximum de vraisemblance du paramètreμ basé sur une observation $x_1; ...; x_n$ d'un échantillon issu de ce modèle.
2) On considère maintenant le modèle gaussien avec paramètre bidimensionnel,i.e.
$(\mathbb{R}; B(\mathbb{R})N(\mu;\sigma^2 : \mu \in \mathbb{R}, \sigma^2 > 0\ )$: Donner l'estimateur du maximum de vraisemblance du paramètre $\theta = (\mu, \sigma^2)$ pour le modèle d'échantillonnage associé.

Exercice 6 (Maximum de vraisemblance pour un modèle de loi uniforme) On considère le modèle uniforme $U_{[0,\theta]} : \theta > 0$

1) Montrer que la vraisemblance associée à un échantillon $x_1, ..., x_n$ observé dans ce modèle est : $\mathrm{L(x_1, ..., x_n;\ \theta)} = \dfrac{1}{\theta^n} I_{X_{(1)} \geq 0} I_{x_{(n)} \leq \theta}$, où $\mathrm{x_1}$ et $x_{(n)}$ sont respectivement les observations des statistiques

d'ordre X(1) et X(n).
2) Donner l'estimateur du maximum de vraisemblance du paramètre θ.

Exercice 7 (Maximum de vraisemblance)

Pour les modèles suivants, donner l'estimateur du maximum de vraisemblance associé à l'observation d'un échantillon $X_1; ...; X_n$. 1) Modèle de la loi exponentielle décalée :
$(\mathbb{R}_+; B_{\mathbb{R}_+}, e_{t_0}(\lambda) : \lambda >0, \text{ t}_0 \in R)$;

On rappelle que la densité de la loi exponentielle décalée $E_{t_0(\lambda)}$ est :

$\text{f}_{(\lambda,t_0)}(x) = \lambda exp(-\lambda(x - t_0))I_{[0,\infty])}$

2) Modèle de la loi Bêta à un seul paramètre : $(\mathbb{R}_+; B_{\mathbb{R}_+}, Beta(1, \theta) : \theta > 1)$

On rappelle que la densité de la loi Beta(a ; b) est :

$\text{f}_{a,b} = \dfrac{1}{\beta(a,b)} x^{a-1}(1 - x)^{b-1} I_{[0,1]}(x)o$

(a ; b) est la valeur de la fonction Eulérienne Bêta prise en a et b. Ind. On pourra montrer en premier lieu que la densité pour le modèle considéré est : $\text{f}_\theta(x) = \theta(1 - x)^{\theta-1} I_{[0,1]}(x)$.

SERIE DE TD N° 4

ÉNONCÉS

exercice 1 : En désire intérpréter les résultats suivants : le nombre de guérisons du concer de la peau à été de 1712 individus sur 2015 patient pour un traitement A et de 757 individus sur 1010 patients pour un traitement B.

Tester l'hypothèse H_0"un individu à la même probabilté d'être guéri dans les deux traitements" contre l'hypothèse H_1"les deux traitements sont caractérisés par deux probabilités de guérison différentes".

exercice 2 : Une enqûete a été effectué en milieu hospitalier pour déterminer si l'usage du tabac favorise l'apparition du concer broncopulmonaire. Cette enqûete a été menée de la manière suivante :

Les individus intérrogés sont répartis en quatre catégories selon leur consommation journalière en cigarette : A (non fumeurs) , B(de 1 à 9), C (de 10 à 19), D(de 20 ou plus) ; il s'agit d'une consommation moyenne évaluée sur les deux dernières années précédant l'enqûete.

Un premier échantillon est constitué de concéreux. Un échantillon témoin a ensuite été choisi parmi les accidentés, c'est-à-dire les patients hospitalisés pour des raisons qui n'ont rien à voir avec le tabac, de plus pour éliminer tout autre facteur, àchaque concéreu correspond un témoin de même sexe, de même âge et intérrogé par le même enquiteur.

A partir des résultats ci-dessous, peut-on conclure à l'influence du tabac ?.

Ca	A	B	C	D	TOT
Co	25	66	177	334	602
T	130	136	165	171	602
TOT	155	202	342	505	1204

Ca=Catégorie, Co=Concéreux, T=Témoins.

exercice 3 : On a vacciné contre la grippe 300 personnes réparties en deux groupes A et B en fonction de l'âge :

Le groupe A comporte 120 individus de 55 ans au plus.

Le groupe B comporte 180 individus de plus de 55 ans.

On a constaté que, dans le groupe A, 38 individus ont eu la grippe l'hiver suivant la vaccination, tandis que 73 individus du groupe B ont eu la grippe ce même hiver.

Pet-on, au risque 10%, considérer qu'il existe un liaison entre l'éfficacité du vaccin et l'âge de la personne vaccinée ?.

exercice 4 : On a vacciné contre la grippe 300 personnes réparties en deux groupes A et B en fonction de l'âge :

Le groupe A comporte 120 individus de 55 ans au plus.

Le groupe B comporte 180 individus de plus de 55 ans.

On a constaté que, dans le groupe A, 38 individus ont eu la grippe l'hiver suivant la vaccination, tandis que 73 individus du groupe B ont eu la grippe ce même hiver.

Pet-on, au risque 10%, considérer qu'il existe un liaison entre l'éfficacité du vaccin et l'âge de la personne vaccinée ?.

exercice 5 : On a croisé deux races de plantes différant par deux caractères : la couleur (rouge ou blanche) et la taille (grande ou petite) des fleurs qu'elle produisent.

La première génération est homogène et donne de grandes fleurs rouges. La seconde génération fait apparaître quatre type de plantes en fonction des fleurs qu'elles produisent : grandes fleurs rouges, grandes fleurs blanches, petites fleurs rouges et petites fleurs blanches.

Sur un échantillon de 320 plantes on a observé les résultats suivants :

phénotypes	GR	GB	PR	PB
effectifs	202	59	45	14

Peut-on considérer, au risque 5% , que les deux caractères étudiés se transmettent selon les lois de MENDEL ?.

exercice 6 : Il est admet qu'en Algerie les groupes sanguins sont réparties de la façon suivante : O :40%, A :43% , B :12%, AB :5%.

Un échantillon de 300 étudiants à l'université de jijel a fourni les résultats :

Groupes	O	A	B	AB
effectifs	112	123	44	21

Peut-on affirmer, au risque 5% , que la répartition des groupes sanguins à l' université de jijel ne diffère pas sensiblement de celle de l'Algerie ?.

exercice 7 : Dan une population de 500 personnes (300 hommes et 200 femmes) on a mesuré la tension artérielle dechaque individu, ce qui a donné les résultats suivants :

	Hypert	TN	Hypot
H	72	192	36
F	38	118	44

Peut-on, au risque 5%, émettre l'hypothèse H_0 d'une liaison entre le sexe de l'individu et la tension artérielle ?.

INDICATION : Le nombre de degrés de liberté est le nombre minimum des case du tableau dont il faut connaître l'effectif pour déterminer l'ensemble du tableau où les sommes de chaque ligne et chaque conlonne sont données.

Dans l'exercice précédent le nombre de degrés de liberté est 2.

exercice 8 : Un médicament a été expérimenté sur 200 malades dévisés en deux groupes M_1 et M_2 indépendants :

-le groupe M_1 composé de 110 malades a aborbé le médicament étudié.

-le groupe M_2 composé de 90 malades a aborbé un placebo.

Les résultats sont les suivants : 60 malades guéris dans le groupe M_1, 36 malades guéris dans le groupe M_2. 1°) Calculer le pourcentage de guérisons et l'écart-type de ce pourcentage pour chacun des échantillons M_1 et M_2.

2°) En admettant que le phénomène étudié suit une loi normale, construir un test permettant d'accepter ou de rejeter l'hypothèse de l'éfficacité du médicament au risque de 5%.

exercice 9 : On veut savoir si une maladie M modifie le taux de certaines protéines dans le song. On a mesuré leurs concentrations dans un échantillon de sujets atteints pa M et dans un autre échantillons formé de sujets en bonne santé (sujets témoins). Les résultats (dans une unité convenable) sont les suivants :

	effectifs	moyenne échantillon	variance échantillon
Malades	77	141	40
Témoins	33	131	32

Tester l'hypothèse "taux identiques chez les malades et les témoins" contre l'hypothèse :

a) "taux différent chez les malades et les témoins".

b) "taux supérieur chez les malades".

exercice 10 : On a mesuré les dimensions d'une tumeur chez les souris traitées ou non par une substance anti-tumorale et on a obtenu :

Surface (cm^2)	5	5,5	6	6,5	7	7,5	8
Nombre de témoins	0	0	2	3	8	4	3
Nombre traités	4	4	8	3	0	1	0

La différence observée est-elle significative ?.

exercice 11 : Dans une maternité, on a compré les poids à la naissance des des bébés de mères primipares et multipares. On a obtenu les résultats suivants :

primipares	$n_1 = 100$	$\overline{x_1} = 3180g$	$\sigma_{e1}^2 = 214400$
multipares	$n_2 = 110$	$\overline{x_2} = 3400g$	$\sigma_{e2}^2 = 243300$

Peut-on admettre au coefficient de confiance de 99% que les enfants nés de mères multipares sont plus lourds que ceux nés de mères primipares ?.