Statistique en Sciences Sociales

(SESP 1240)

Michel Mouchart

4 janvier 2005

Table des matières

Préface

0.1 Présentation des Notes de cours

Le texte qui suit reproduit la majorité des transparents utilisés lors de l'enseignement du cours "Statistique en Sciences Sociales" (SESP1240). Il donne surtout des formules: *l'essentiel de la matière, c'est-à-dire les raisonnements, les exemples et les motivations, n'y sont pas rédigés.* La rédaction de cet élément essentiel de la matière est donc du ressort de chaque étudiant qui peut ainsi adapter ses propres notes à son mode personnel de compréhension.

Ce document vise, en ordre principal, à faciliter une participation plus active à l'enseignement magistral. En réduisant quelque peu le travail de prises de note, l'étudiant pourra ainsi consacrer, durant cet exposé, plus d'attention au travail de compréhension active, en particulier en posant des questions durant le cours et les interruptions de cours. Accessoirement ces notes ont aussi un rôle de sécurisation, en assurant les étudiants de la correction de certaines formules.

Instrument de travail, la présentation aérée de ce texte invite les annotations tant sur la page de gauche que sur la page de droite. *Ces notes, brochées et sans feuille ajoutée (ni collée, ni volante), constituent le seul document dont vous pourrez disposer pendant l'exament écrit; toute annotation manuscrite de votre main y sera évidemment bienvenue.*

0.2 Présentation générale du cours: Organisation et Remarques pédagogiques

0.2.1 Orientation du cours

Du point de vue de la structure du programme des cours de candidatures, ce cours constitue une suite au cours "SESP 1148 - Statistique", de première candidature, et un accompagnement du cours "SESP 1242 - Méthodes et démarches de recherche en sciences sociales" de seconde candidature. Il est aussi une préparation aux méthodes statistiques utilisées et développées, en second cycle, dans les différentes licences auxquelles ce programme donne accès.

L'objet de ce cours est "de compléter la formation de base en probabilités (modèles continus et bivariés) puis d'introduire aux concepts de base de l'inférence statistique (théorie de l'estimation ponctuelle, intervalles de confiance, tests d'hypothèses). Ces concepts sont ensuite appliqués dans une série de modèles particulièrement utiles en sciences sociales. La formation obtenue devrait permettre aux étudiants d'utiliser de manière critique les techniques statistiques et d'éviter ainsi de tomber dans la routine d'application de recettes."(cahier des charges)

0.2.2 Structure pédagogique du cours

Au point de vue pédagogique, insistons sur une pédagogie basée sur un *appentissage en trois vagues* : enseignement magistral, séances d'exercices et travail de semestre. Chaque vague donne un éclairage différent et complémentaire sur la matière.

Il importe de remarquer que la compréhension de la matière ne peut être que progressive, et non du type tout ou rien. Chaque vague apporte ainsi un supplément de compréhension: ne pas participer activement au cours ou être passif aux séances d'exercices prive l'étudiant d'une des phases de l'apprentissage.

Le cours magistral

Le cours magistral se donnera à raison de 4h par semaine durant les 6 premières semaines du quadrimestre, et de 2h par semaine les 2 semaines suivantes, selon l'horaire annoncé par la faculté (le lundi- 8 semaines de 10h45 à 12h45- et le vendredi- 6 semaines de 8h30 à 10h30- , local Doyen 31).

Le cours magistral fournit à l'étudiant un premier contact avec la matière; l'accent y est mis sur la structure des raisonnements et de la matière. Ce premier contact peut être assez ardu, mais une bonne façon de faciliter ce premier contact est de lire le syllabus AVANT le cours. L'étudiant sait alors quel sera l'objet général du prochain cours et peut repérer les moments qui demanderont un effort plus soutenu d'attention et les questions qu'il pourra envisager de poser.

Le texte projeté au cours reproduit exactement le texte du syllabus et ses pages ont la même numérotation. Il est donc possible à tout moment de savoir où le professeur en est dans l'exposé de la matière.

Sur le plan pédagogique, le jeu des questions et réponses durant le cours constitue une occasion privilégiée pour revenir sur des points plus délicats vus précédemment: ce qui n'a pas été compris aujourd'hui pourra l'être lors d'un cours suivant. On insistera donc sur la nécessaire interaction "enseignant-enseigné" réalisable, notamment, par un jeu continu de questions-réponses; celui-ci suppose la révision continue de la matière par l'étudiant.

Les séances d'exercices

Les séances d'exercices sont organisées à partir de la seconde semaine du quadrimestre. La séance de la seconde semaine se déroulera en salle informatique et assurera la continuité vis à vis du cours d'informatique du premier quadrimestre. Les séances des semaines 3 à 7 et 9, 10 et 11 constitueront un second contact avec la matière avec un décalage de 1 à 3 semaines par rapport à l'enseignement magistral. Elles mettront surtout l'accent sur l'aspect opérationnel de la matière. Les séances des semaines 8 et 12 serviront pour l'accompagnement et la guidance du travail de quadrimestre réalisé par les étudiants (voir point suivant). La semaine 13 sera utilisée pour un test de connaissance de logiciel. La semaine 14 est consacrée à la présentation orale des travaux de semestre.

Il est heureux que le contact avec la matière lors des séances d'exercices soit souvent perçu comme moins aride. Ces séances ont une finalité complémentaire à celle du cours, et non une fonction d'élucidation systématique de la matière. Certains aspects du cours n'y sont donc pas revus ; ils ne peuvent dès lors être maîtrisés que par un jeu de questions et réponses.

Le travail de quadrimestre

Les étudiants réaliseront, par groupe de deux, un travail de quadrimestre. Ce travail constituera une troisième phase dans l'assimilation de la matière, au moyen de la réalisation d'un travail d'analyse et d'interprétation de données réelles. Il aura pour buts de
(i) réaliser une synthèse de la matière,
(ii) permettre une première expérience d'un travail statistique complet, alors que les séances d'exercices n'en donneront qu'une expérience parcellaire,

(iii) initier les étudiants à une pratique critique de la statistique,
(iv) auto-contrôler, avant l'échéance des examens, la compréhension effective de la matière.

Ce travail supposera l'utilisation du logiciel statistique SAS. Les modalités pratiques de ce travail seront données tout au long du quadrimestre. En remettant le travail au plus tard le jeudi 12 mai à 13h, les étudiants devront s'inscrire par binôme pour la présentation orale qui aura lieu la semaine suivante (entre le 17 et le 20 mai). Cette présentation orale requiert une révision préalable de la matière afin de permettre à l' étudiant non seulement d'expliquer mais aussi de justifier les différentes étapes de son travail.

0.2.3 L'apprentissage par "vagues successives"

Cette organisation met l'accent sur un apprentissage par "vagues successives" et sur un glissement progressif d'une activité essentiellement magistrale (4h de cours les premières semaines) à une activité essentiellement personnelle (mais sous guidance...) durant les dernières semaines. La charge de travail est prévue pour être relativement constante durant tout le quadrimestre.

Le "contrat pédagogique" sous-jacent vise à permettre à chaque étudiant d'acquérir une compréhension en profondeur de la matière (afin qu'elle ne soit pas "trop vite" oubliée après avoir passé l'examen...) et d'aborder la session de juin sans anxiété excessive (pour la statistique...), dans la mesure où il aura suivi durant tout le quadrimestre le rythme de travail proposé.

Il convient d'insister sur le fait qu'un bon apprentissage ne peut se faire que par un QUESTIONNEMENT CONTINU : apprendre n'est pas enrégistrer ni répéter. Un étudiant ne peut espérer comprendre que par les questions qu'il se sera posées. Les questions sont d'abord adressées à soi-même, ensuite aux compagnons du cours, au professeur ou aux assistants. Ces questions peuvent être exprimées oralement ou par écrit, avant, pendant ou après le cours ou les séances d'exercices, lors des monitorats ou des heures de réception du professeur.

Au bout de la quatorzième semaine du semestre, ce travail de compréhension est terminé. Le travail de semestre à remettre à ce moment est un outil qui permet à l'étudiant d'auto-évaluer son apprentissage effectif de la matière et, en cas de besoin, de se remettre à jour pour la session de juin.

0.2.4 L'encadrement

Deux assistants, Mademoiselle R.Rousseau et Monsieur C. Almeida, participeront à l'encadrement de ce cours et assureront un service de réception des étudiants selon un horaire qui sera annoncé d'ici peu.

Le professeur recevra les étudiants le lundi de 14h. à 16h (Institut de Statistique, 20 Voie du Roman Pays, LLN, bureau d.382), moyennant rendez-vous pris, de préférence, lors d'une des interruptions de cours.

0.2.5 Le matériel pédagogique

Le matériel pédagogique accompagnant ce cours est disponible à la DUC et comprend : un syllabus reproduisant les transparents utilisés pour l'enseignement magistral et un fascicule d'exercices.

L'enseignement magistral, soutenu par le syllabus, est conçu pour être suffisant pour assurer une première compréhension de la matière. Aux étudiants qui, pour une raison quelconque, éprouveraient des difficultés à assimiler la matière de cette façon,

il est suggéré de consulter un texte de référence reprenant un exposé détaillé de la matière. Une référence possible est la suivante : WONNACOTT,Th. et R.J. WONNACOTT, *Introductory Statistics for Business and Economics*, New York : John Wiley and Sons.
Il appartiendra aux étudiants, qui décideraient d'en faire usage, de repérer les chapitres de ce livre qui correspondent à la matière du cours. Il existe aussi une version française de cet ouvrage. Il convient de remarquer que le style et l'orientation de ce livre sont différents de ceux du cours magistral.

0.3 L'évaluation du cours

L'évaluation du cours sera basée sur la somme des points obtenus pour le travail de quadrimestre (3 points pour le rapport écrit et 3 points pour l'oral), l'évaluation des connaissances en SAS (2 points) et l'examen de session (12 points de théorie et d'exercices).

L' examen de session comportera des questions pour évaluer la compréhension de la matière et des questions pour évaluer la qualité de l' opérationalisation. Les questions n'étant pas encore rédigées, le professeur ne peut pas encore les décrire mais toute suggestion est bienvenue...

Un travail remis pour la session de juin doit obligatoirement être défendu oralement par les deux étudiants auteurs du rapport. Pour la session de septembre, les étudiants qui le désirent peuvent remettre en jeu la côte du travail obtenue en juin. Dans ce cas, ils doivent remettre un nouveau rapport écrit et le défendre oralement.

Louvain-la-Neuve, janvier 2005.

Chapitre 1

Buts et limitations de la statistique

1.1 Question de méthode

1.1.1 Sciences sociales comme science

$$\text{Interaction entre} \left[\begin{array}{l} \text{faits} \\ \text{théorie} : \left[\begin{array}{l} \text{hypothèse} \\ \text{déduction} \end{array} \right. \end{array} \right.$$

1.1.2 Rôle de la statistique:

- Confrontation $\left[\begin{array}{l}\text{hypothèse}\\ \text{réalité}\end{array}\right.$

  Méthodologie de l'apprentissage par l'observation.

●● Mesure en Sciences sociales

c'est-à-dire relation entre $\left[\begin{array}{l}\text{concept abstrait}\\ \text{mesure de ce concept}\end{array}\right.$

ex : intelligence $\left[\begin{array}{l}\text{"forme"}\\ \text{"quantité"}\end{array}\right.$

● ● ● Modèle structurel : relation entre concepts

Modèle statistique : population - échantillonnage

● ● ●● Cadre général : Méthodes et démarches de recherche en Sciences Sociales.

1.2 Modèle statistique

1.2.1 Le problème

$$\text{Toute inférence statistique} \left\{ \begin{array}{l} \text{incertaine} \\ \text{basée} \left[\begin{array}{l} \text{observations} \\ \text{hypothèses} \rightarrow \quad \text{“Modèle statistique”} \end{array} \right. \end{array} \right.$$

1.2.2 Modèle statistique: trois idées de base

- Ensemble d'hypothèses

- Rôle d'une hypothèse: “utile”

- Modèle statistique “ensemble de distributions d'échantillonnage”

1.3 Modèle statistique: Exemples

1.3.1 Jeu de pile ou face: jeter une pièce 1 fois

observation

$$X = \begin{cases} 1 & \text{pile} \\ 0 & \text{face} \end{cases}$$

1er modèle statistique : "Modèle de Bernoulli"

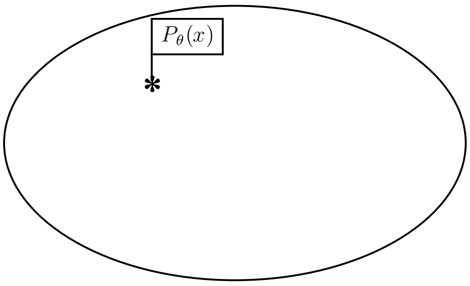
x	$P(X = x)$	
1	θ	(theta)
0	$1 - \theta$	

$$0 \leq \theta \leq 1$$

Exemple $\theta = 1/2$ pour une pièce équilibrée

$$P_\theta(x): \qquad P(X = x) = \theta^x \, (1-\theta)^{1-x} \, \mathbb{1}_{\{0,1\}^{(x)}}$$

c'est-à-dire: $x = 0{,}1$



"distribution d'échantillonnage"

1.3.2 Jeter la pièce n fois

Observations $(X_1\ X_2 \cdots X_n)$ avec $X_i \in \{0{,}1\}$

par exemple : lorsque $n = 6$, on pourrait observer $(0\ 1\ 1\ 0\ 0\ 1)$

2ème modèle statistique : Modèle binomial

Hypothèses

- H_1 : X_i indépendants
- H_2 : X_i identiquement distribués
- H_3 : X_i de Bernoulli

A partir de ces hypothèses, on déduira:

$$
\begin{aligned}
P_\theta(X_1 \cdots X_n) &= P^1_\theta(X_1) \cdot P^2_\theta(X_2) \cdots P^n_\theta(X_n) && (H_1)\\
&= P_\theta(X_1)\ P_\theta(X_2) \cdots P_\theta(X_n) && (H_2)\\
&= \theta^{X_1}(1-\theta)^{(1-X_1)}\theta^{X_2}(1-\theta)^{1-X_2} \cdots \theta^{X_n}(1-\theta)^{1-X_n} && (H_3)\\
&= \prod_{1 \le j \le n} \theta^{X_j}(1-\theta)^{1-X_j} && \text{(manipulation)}\\
&= \theta^{\sum_j X_j}(1-\theta)^{\sum_j (1-X_j)} && \text{(manipulation)}
\end{aligned}
$$

On écrira donc:

$$P_\theta(X_1 \cdots X_n) = \theta^S \quad (1-\theta)^{n-S} \qquad\qquad 0 \leq \theta \leq 1$$

où

$$S = \sum_j X_j = X_1 + X_2 + \cdots + X_n$$

en effet:

$$\begin{aligned}\sum_{1\leq j\leq n} (1-X_j) &= (1-X_1) + (1-X_2) + \cdots + (1-X_n) \\ &= n - \sum_j X_j \\ &= n - S\end{aligned}$$

Remarque : Sur les fonctions indicatrices

Soit U un "univers de référence", c'est-à-dire l'ensemble des valeurs possibles de la variable X.

A événement sur la v.a. X, c'est-à-dire $A \subset U$

Par exemple:

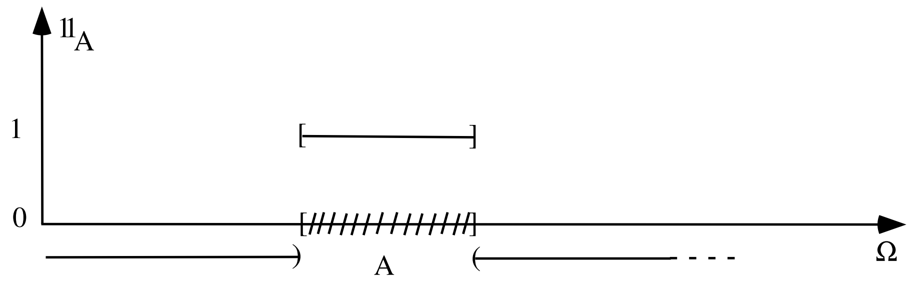
$U = \mathbb{R}$ ou $U = \mathbb{R}_+$

$A = \{a \leq X \leq b\}$ ou $A = \{X \geq 10\}$ ou $A = \{X \leq -5\}$

Fonction caractéristique (ou indicatrice): $\mathbb{1}_A : U \to \{0,1\}$ telle que :

$$\begin{aligned}\mathbb{1}_A(x) &= 1 \quad x \in A \\ &= 0 \quad x \notin A\end{aligned}$$

Graphiquement :



Exercices : Vérifiez

- $\mathbb{1}_{A^c}(x) = 1 - \mathbb{1}_A(x)$ (où $A^c = \overline{A}$ "complémentaire de A")
- en général : $\mathbb{1}_{A\cap B}(X) = \mathbb{1}_A(X).\mathbb{1}_B(X)$
- $A \cap B = \phi \Rightarrow \mathbb{1}_{A\cup B}(X) = \mathbb{1}_A(X) + \mathbb{1}_B(X)$
- en général : $\mathbb{1}_{A\cup B}(x) = \mathbb{1}_A(x) + \mathbb{1}_B(x) - \mathbb{1}_{A\cap B}(x)$

1.4 Inférence statistique

1.4.1 Trois étapes dans l'analyse statistique

- analyse descriptive

•• inférence

• • • planification expérimentale

1.4.2 Trois formes d'inférence statistique

- estimation ponctuelle

$$\theta \text{ estimé par:} \qquad \hat{\theta}(X_1 \cdots X_n)$$

par exemple $\hat{\theta} = \dfrac{\sum X_j}{n}$

•• intervalle de confiance: $\theta \in \left[\hat{\theta}_{\min} \;\; \hat{\theta}_{\max}\right]$ avec une probabilité contrôlée

• • • tests d'hypothèse

Chapitre 2

Probabilités

Remarque préliminaire : chapitre de rappel de 1ère candi!

2.1 Espace de probabilité

$$(\Omega,\mathcal{A},P)$$

- Ω univers:par exemple, espace d'échantillonnage

- $\mathcal{A}$ ensemble événements ("A ronde")

$$\mathcal{A} = \{A,B,C\cdots\} \text{ tels que } A \subset \Omega, B \subset \Omega, \cdots$$

- $P : \mathcal{A} \to [0,1]$ donc : $0 \leq P(A) \leq 1 \quad \forall\, A \in \mathcal{A}$

2.2 Propriétés (rappel)

1. $P(\Omega) = 1 \qquad P(\emptyset) = 0$

2. $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

2.bis (dans le cas dénombrable)

Si $\{A_i : \ i \in I\!N\}$ tels que $A_i \in \mathcal{A} \quad A_i \cap A_j = \emptyset \quad i \neq j$

alors $P(\bigcup_{1 \leq i \leq \infty} A_i) = \sum_{1 \leq i \leq \infty} P(A_i)$

En particulier
$P(A^c) = 1 - P(A)$

2.3 Exemple fondamental (cas équiprobable)

$$\Omega \text{ fini} \qquad |\Omega| = N$$

$$P(\{\omega\}) = \frac{1}{N}$$

Dans ce cas, pour tout:

$$A \subset \Omega$$

notons le nombre de ses éléments (c'est-à-dire, le cardinal de l'ensemble)

$$|A| = a$$

On vérifiera alors

$$\begin{aligned} P(A) &= \frac{|A|}{|\Omega|} = \frac{a}{N} \\ &= \frac{\text{nombre de cas "favorables" (à l'événement } A)}{\text{nombre de cas "possibles"}} \end{aligned}$$

Attention

Ceci est un *exemple* de quantification de la probabilité, ce n'est *pas* une définition de la probabilité.

Exercice

Commentez cette remarque

2.4 Probabilité conditionnelle

2.4.1 Notions de base

Idée de base: restreindre l'univers

Exemple:

Ω tous les étudiants UCL

P $\forall\, \omega : P(\{\omega\}) = \frac{1}{N}$ (cas équiprobable)

$A \subset \Omega$ tous les étudiants ESPO

$B \subset \Omega$ tous les étudiants filles

$P(B)$: % de filles.

Définition

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \qquad \text{si} \qquad P(A) > 0$$

Propriétés

(i) pour A fixé: $P(B_i|A)$ $B_i \in \mathcal{A}$: toutes les propriétés d'une probabilité
(ii) $B \supset A \Rightarrow P(B|A) = 1$
(iii) $A \cap B = \emptyset \Rightarrow P(B|A) = 0$

Exercice: vérifier ces propriétés

En particulier:

$$P(A|A) = 1$$

$$P(B|\Omega) = P(B)$$

2.4.2 Formules des probabilités composées

$$\begin{aligned}
P(B|A) &= \tfrac{P(A\cap B)}{P(A)} \\
P(A \cap B) &= P(A)P(B|A) \\
&= P(B)P(A|B) \\
P(\underbrace{A \cap B}_{D} \cap C) &= \underbrace{P(A)P(B|A)}_{P(D)}\, P(C|A \cap B) \\
&= \cdots
\end{aligned}$$

Exercice : Trouver d'autres décompositions de $P(A \cap B \cap C)$.

Remarque :
$P(A_1 \cap A_2 \cap \cdots \cap A_p) \rightarrow p$:! décompositions différentes

Rappel : $p! = 1 \cdot 2 \cdot 3 \ldots \cdot p$

2.4.3 "Partition" de Ω ou "Système contradictoire d'événements"

Définition : $\{A_1\ A_2\ \cdots A_n\}$ tels que

$$A_i \cap A_j = \emptyset \quad i \neq j \quad \text{"mutuellement exclusifs"}$$

$$\bigcup_{1 \leq i \leq n} A_i = \Omega \qquad \text{"collectivement exhaustifs."}$$

Exemples

Ω: ensemble des étudiants de l'U.C.L. pour une année donnée
A_i: faculté d'inscription principale d'un étudiant

Exercice
Pourquogi la spécification d'inscription principale est-elle importante pour s'assurer que les A_i forment une partition de Ω?

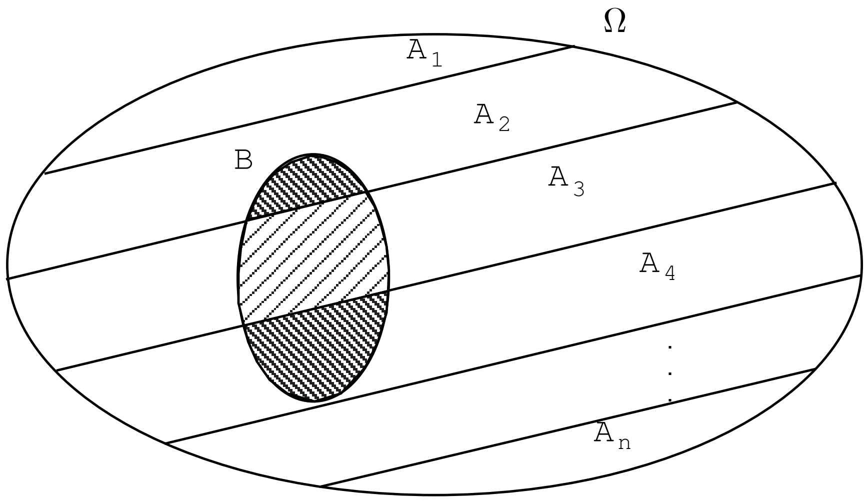
2.4.4 Formule des probabilités totales

Exemple:

Ω	:	tous les éléments de l'U.C.L.
A_i	:	étudiants appartenant à la faculté i
B	:	tous les étudiants filles

Si $\{A_i :\ 1 \leq i \leq n\}$ partition de Ω alors

$$\begin{aligned} P(B) &= \sum_{1\leq i\leq n} P(B \cap A_i) \\ &= \sum_{1\leq i\leq n} P(A_i)P(B|A_i) \end{aligned}$$



En effet :

$B = (A_1 \cap B) \cup (A_2 \cap B) \cup (A_3 \cap B) \cdots \cup (A_n \cap B)$

2.4.5 Formules de Bayes

a) **Formule de calcul**

$$\begin{aligned} P(B|A) &= \frac{P(A\cap B)}{P(A)} \\ &= \frac{P(B)P(A|B)}{P(A\cap B)+P(A\cap B^c)} \\ &= \frac{P(B)P(A|B)}{P(B)P(A|B)+P(B^c)P(A|B^c)} \end{aligned}$$

$\{B_1 \cdots B_n\}$ partition de Ω.

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum\limits_{1\le j\le n} P(B_j)P(A|B_j)}$$

b) **interprétation**

- "Probabilité inverse"
- Formule d'apprentissage

1° étape : "a priori" $P(B_i)$

2° étape : nouvelle information : A

question : comment "réviser" $P(B_i)$?

réponse : "a posteriori" : $P(B_i|A)$

2.5 Indépendance en probabilité

Soit A, B: 2 événements

Définition: "A et B indépendants en probabilité"

$A \perp\!\!\!\perp B$

$$\Leftrightarrow \quad P(A \cap B) = P(A)P(B)$$

$$\Leftrightarrow \quad P(A|B) = P(A) \qquad \text{si } P(B) > 0$$

$$\Leftrightarrow \quad P(B|A) = P(B) \qquad \text{si } P(A) > 0$$

Remarque : Relation symétrique entre les événements A et B.

Soit $A_1 \cdots A_n$

Définition. "A_i mutuellement indépendants" : $\underset{i}{\perp\!\!\!\perp} A_i$

$$\Leftrightarrow \quad \forall \quad J \subset \{1,2, \cdots n\} : \; P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i)$$

Remarque:

$2^n - n - 1$ relations = nombre de sous-ensembles de $\{1, \ldots, n\}$ dont le nombre d'éléments est au moins égal à 2.

Chapitre 3

Variables aléatoires.

3.1 Idées de base

3.1.1 Expérience aléatoire

Par exemple: extraire 1 échantillon

Exercice : décrire: $(\Omega,\mathcal{A},P)$

3.1.2 Variable aléatoire

$$X : \Omega \to I\!R$$

Par exemple:
$\Omega = \mathcal{N}$ et $k \to X(k) = \xi_k$ =le poids de l'individu k

UNE VARIABLE ALEATOIRE EST DONC UNE FONCTION

Remarque

Au point de vue mathématique, c'est une "bonne" fonction :

$P\{a \leq X \leq b\}$ "bien définie" pour tout $-\infty < a \leq b < +\infty$

en particulier, lorsque $a \to -\infty$

$P[X \leq b] = F_X(b)$ "bien définie" pour tout $-\infty < b < +\infty$

3.1.3 Distinction au sujet de "l'ensemble $\mathcal{X}$ des valeurs possibles de X"

2 cas
- $\mathcal{X}$ "discret"
  - fini
  - dénombrable ($I\!N$: entiers naturels)
- $\mathcal{X}$ "continu" intervalle de $I\!R$ (éventuellement $I\!R$ ou $I\!R_+$)

Cas discret: déjà abordé au cours de première candidature !

3.2 Variables aléatoires discrètes

3.2.1 Fonction de probabilité

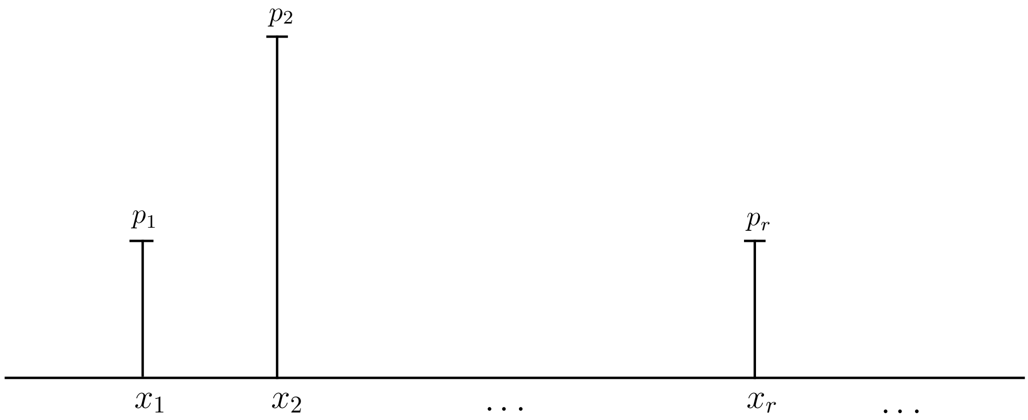
$$\begin{cases} x_1\ x_2\ \cdots x_i \cdots : & \text{valeurs possibles} \\ \\ p_1\ p_2\ \cdots p_i \cdots : & \text{probabilités de chacune des valeurs possibles.} \end{cases}$$

ou $p_i = P(X = x_i) \quad i = 1,2,...$ "fonction de probabilité"

Soit: I intervalle de $\mathbb{R}$

alors: $P(X \in I) = \sum_{\{i : x_i \in I\}} p_i$

Graphiquement: diagramme à barres



Propriétés

1. $p_i > 0 \qquad \forall i$
2. $\sum_i p_i = 1$

Remarque:

La hauteur de chaque barre est proportionnelle à la probabilité p_i.

:

Exemple 1: V.A. de Bernoulli

$\mathcal{X} = \{0,1\}$

x	$P(X = x)$	
0	$1-p$	
1	p	$0 \leq p \leq 1$

$$P(X = x) = p^x(1-p)^{1-x}$$

Exemple 2: V.A. binomiale

Par exemple: probabilité du nombre de "succès" dans n répétitions indépendantes d'une V.A. de Bernoulli.

$\mathcal{X} = \{0,1,2,\cdots n\}$

$$P[X = x] = \binom{n}{x} p^x(1-p)^{n-x}$$

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

à revoir: cours de première candi. !

3.2.2 Quelques rappels sur $\sum$ et $\prod$

$$\sum_{j=1}^{3} a_j = \sum_{i=1}^{3} a_i = a_1 + a_2 + a_3$$

$$\prod_{j=1}^{3} a_j = \prod_{i=1}^{3} a_i = a_1 \times a_2 \times a_3$$

$$a^b a^c = a^c a^b = a^{b+c}$$

$$\prod_{j=1}^{3} a^{x_j} = a^{\sum\limits_{j=1}^{3} x_j} = a^{x_1+x_2+x_3}$$

3.3 Variables aléatoires continues

3.3.1 Présentation générale

$\mathcal{X}$ intervalle de $\mathbb{R}$

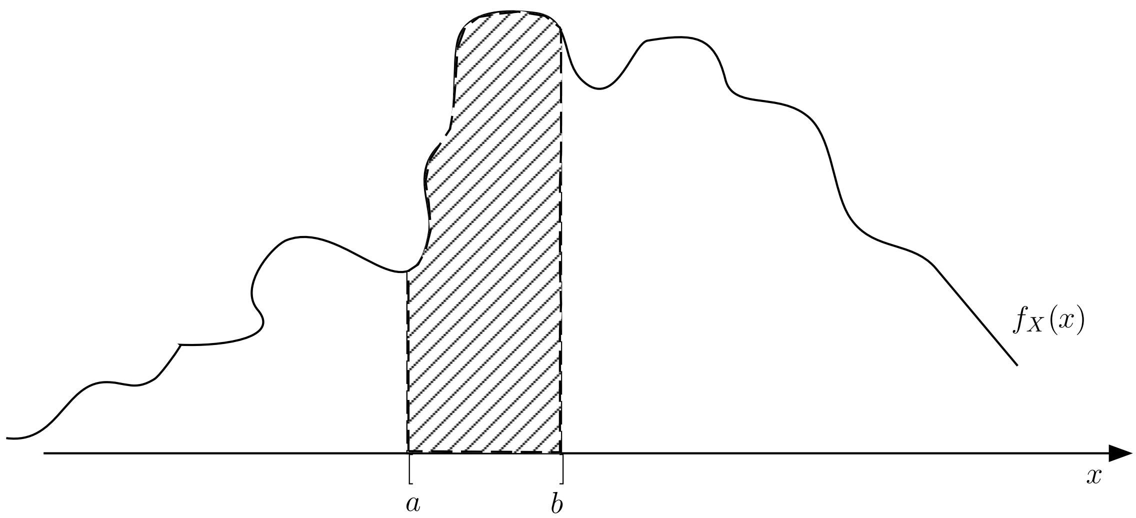
Probabilité d'intervalle: seul intérêt

c'est-à-dire $\forall x \in \mathbb{R} \qquad P(X = x) = 0$

$$\Rightarrow \begin{cases} P(X \leq a) = P(X < a) \\ \\ P(X \geq a) = P(X > a) \end{cases}$$

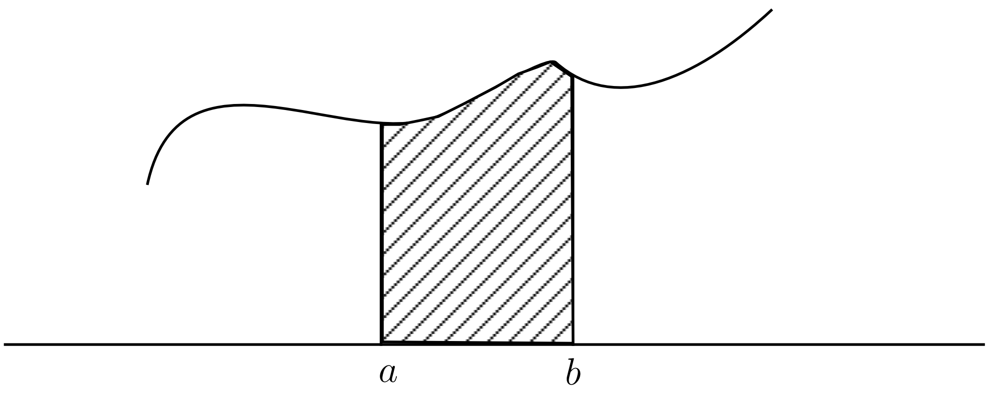
Exercice : Justifiez.

Comment la calculer ?



$P(X \in [a,b]) = ////$

$f_X(x)$ "densité de probabilité"
"fonction de densité"



"Surface sur $[a,b]$"

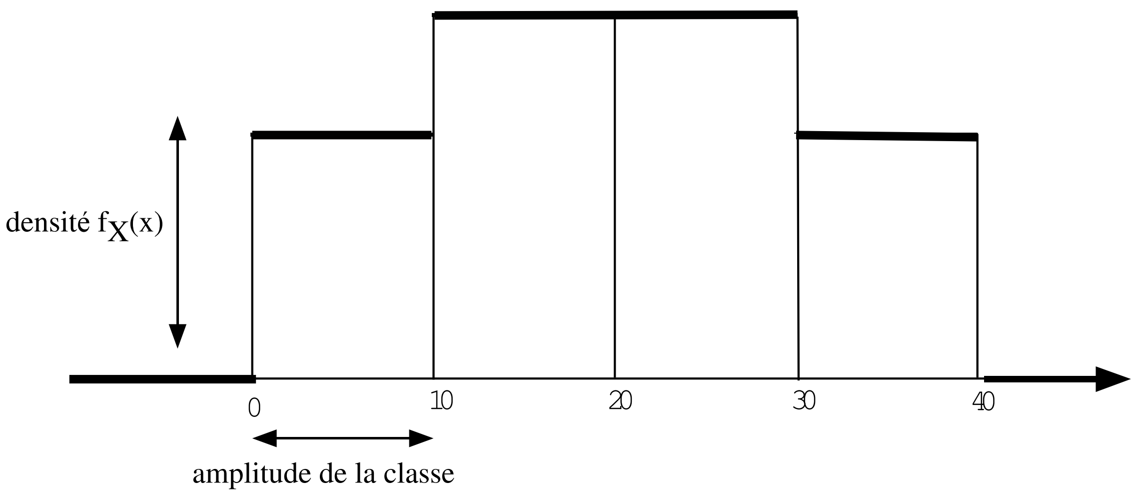
$$= \int_a^b f_X(x)dx : \text{“intégrale de la fonction } f_X(x) \text{ de } a \text{ à } b\text{”}.$$

Propriétés

1. $f_X(x) \geq 0 \qquad \forall x \in \mathbb{R}$
2. $\displaystyle\int_{-\infty}^{+\infty} f_X(u)du \equiv 1$

3.3.2 Rappel: histogramme pour données groupées

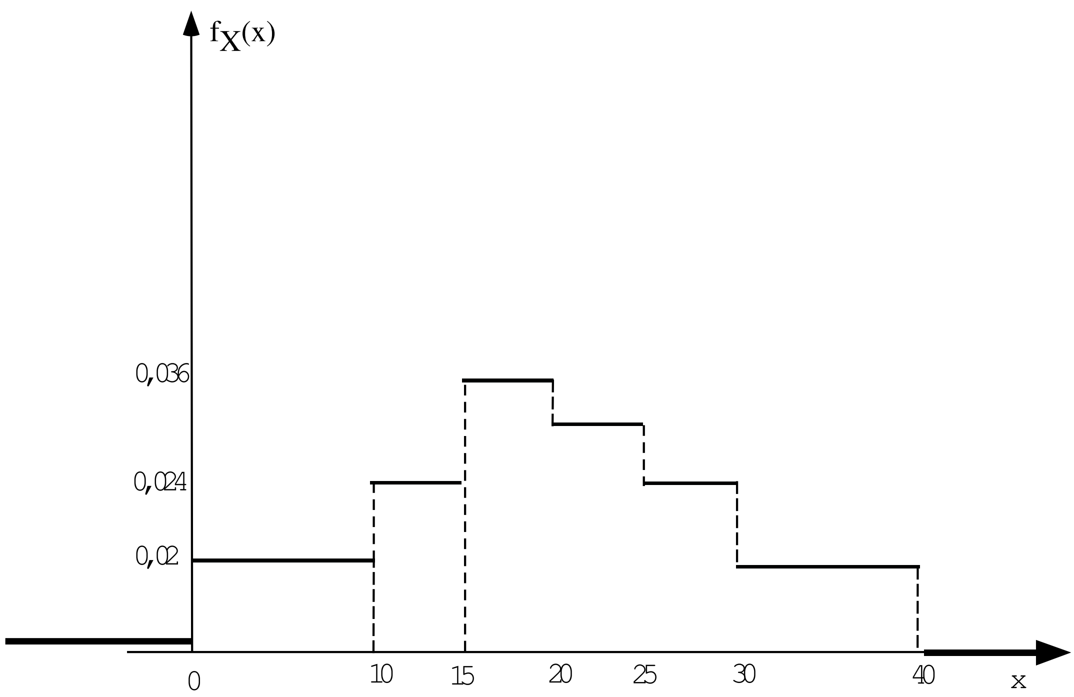
$[x_{i-1}, x_i)$	p_i	a_i	$d_i = \dfrac{p_i}{a_i}$
[0, 10)	0,2	10	0,02
[10, 20)	0,3	10	0,03
[20, 30)	0,3	10	0,03
[30, 40)	0,2	10	0,02



$$\begin{aligned} \text{densité} \quad &= \frac{\text{fréquence}}{\text{amplitude}} \\ \Rightarrow \text{fréquence} \quad &= \text{densité} \times \text{amplitude} \\ &= \text{surface du rectangle} \end{aligned}$$

Exercice : modifions les intervalles de classe

$[x_{i-1} \quad x_i[$	p_i	a_i	$d_i = \frac{p_i}{a_i}$
[0, 10)	0,2	10	0,02
[10, 15)	0,12	5	0,024
[15, 20)	0,18	5	0,036
[20, 25)	0,17	5	0,034
[25, 30)	0,13	5	0,026
[30, 40)	0,2	10	0,02



On peut ainsi calculer :

$$\begin{aligned}
P[10 \leq X < 15] &= 0{,}024 \times 5 = 0{,}12 \\
P[12{,}5 \leq X \leq 17.5] &= P[12{,}5 \leq X \leq 15] + P[15 < X < 17.5] \\
&= (0{,}024 \times [15 - 12.5]) + (0{,}036 \times [17.5 - 15]) \\
&= 0{,}06 + 0{,}09 = 0{,}15
\end{aligned}$$

3.4 Fonction de distribution ou de répartition d'une v.a.

3.4.1 En général

C'est-à-dire v.a. $\begin{cases} \text{continue} \\ \\ \text{discrète} \end{cases}$

$F_X(x) = P(X \leq x) \qquad\qquad x \in \mathbb{R}$

Propriétés générales:

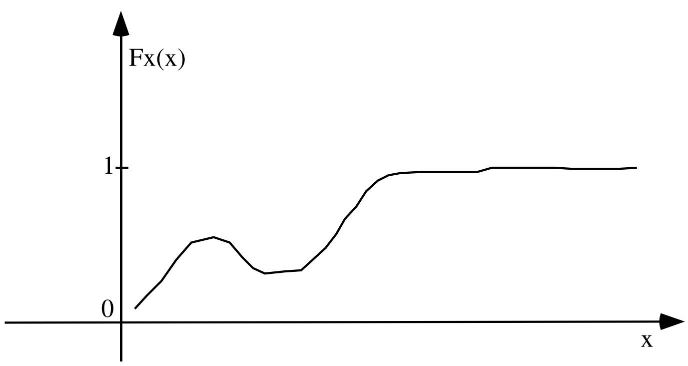
$F_X : \mathbb{R} \to [0{,}1]$

$\lim_{x\to-\infty} F_X(x) = 0$

$\lim_{x\to+\infty} F_X(x) = +1$

$a_2 > a_1 \Rightarrow F_X(a_2) \geq F_X(a_1)$

Exercice : Expliquez pourquoi la définition de la fonction de distribution exclut une fonction du type suivant :



3.4.2 Cas discret (rappel 1ère candidature)

Définition :

$P(X \leq x) = F_X(x)$

Par exemple:

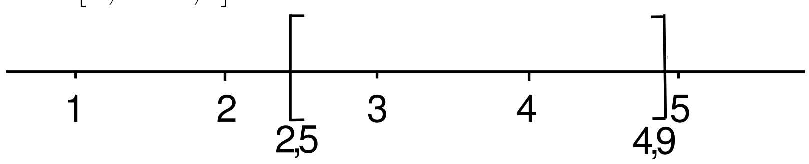
x_i	p_i	$F_X(x_i)$
1	0,2	0,2
2	0,3	0,5
3	0,3	0,8
4	0,2	1,0

Exemples du calcul $P(X \in I)$

1) $I = [-20 \quad -1] \Rightarrow P(X \in I) = 0$

En effet $\mathcal{X} = \{1,2,3,4\}$
$\downarrow$
dès lors $\mathcal{X} \cap I = \phi$

2) $I = [2{,}5 \quad 4{,}9]$

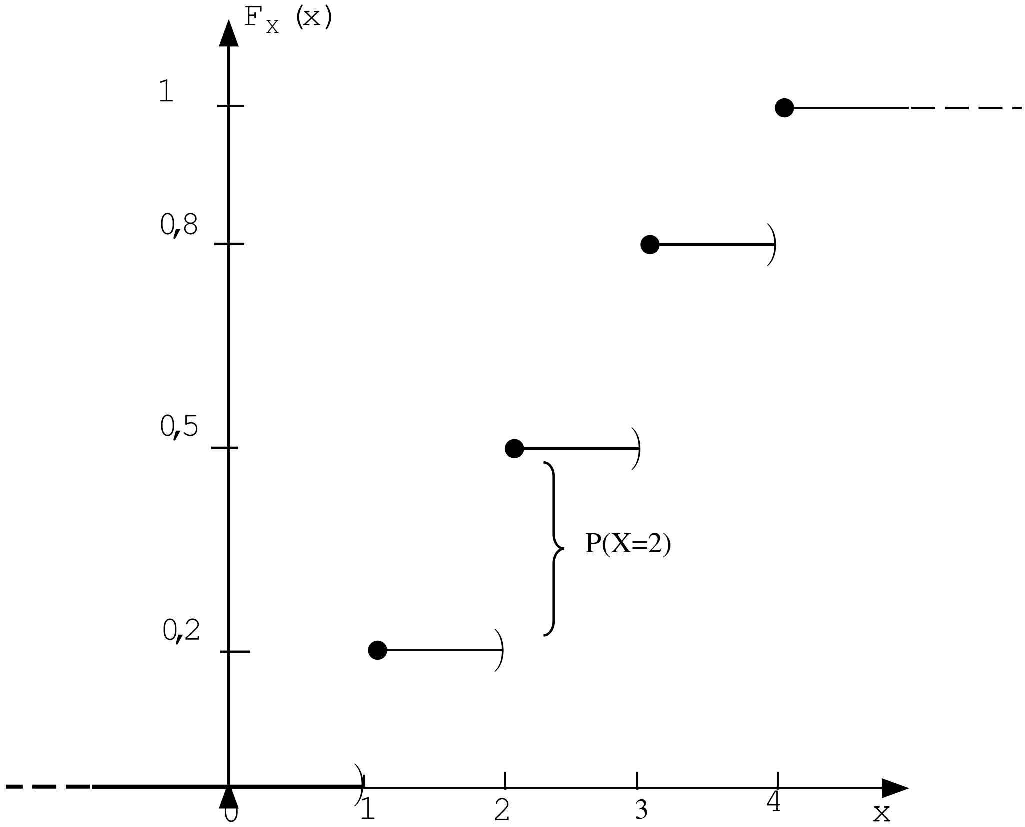


$$P(X \in I) = P(X = 3) + P(X = 4) \qquad \text{car } \mathcal{X} \cap I = \{3{,}4\}$$

$$= 0{,}3 + 0{,}2 = 0{,}5$$

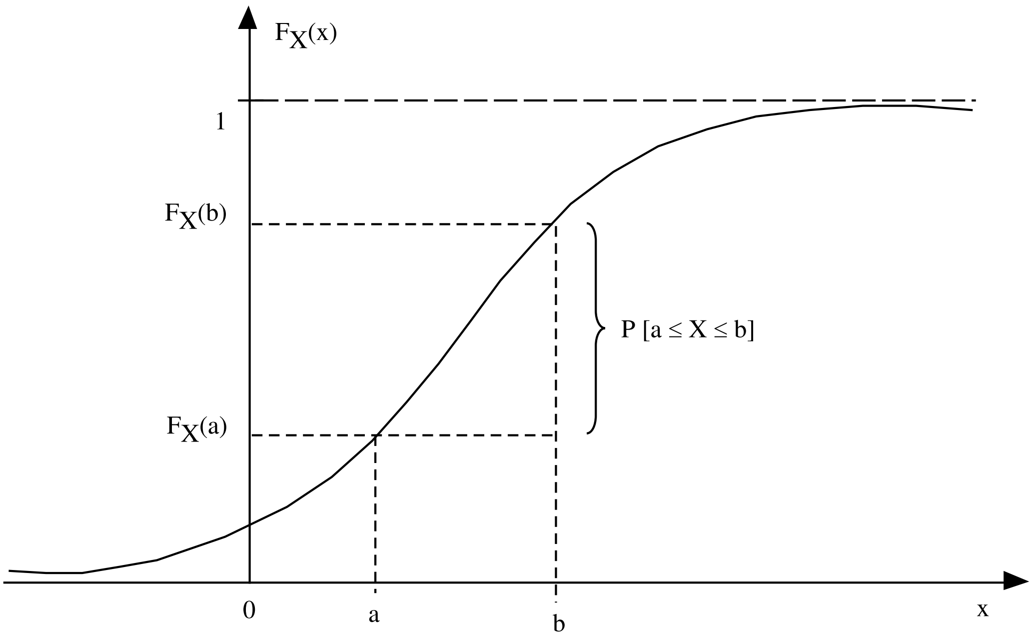
mais aussi

$$
\begin{aligned}
P(X \in I) &= P(X \leq 4{,}9) - P(X \leq 2{,}5) \\
&= F_X(4{,}9) - F_X(2{,}5) \\
&= 1 - 0{,}5 = 0{,}5
\end{aligned}
$$



Remarque (pour les esprits curieux!)
Une fonction de distribution est partout "continue à droite avec limite à gauche". Dès lors, lorsque la fonction F_X n'est pas continue au point a_1, sa valeur en a_1 est la valeur supérieure du "saut" en a_1. On comprend ainsi la convention [—–) dans la représentation graphique.

3.4.3 Cas continu

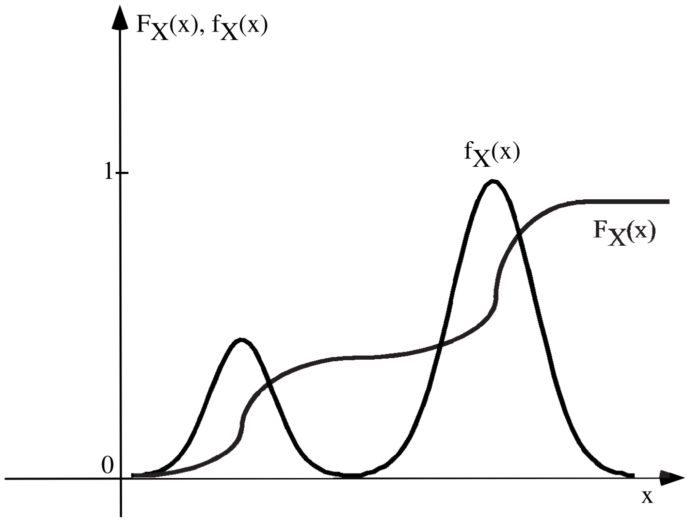


Remarque (pour les esprits curieux!)
Lorsque la fonction de distribution F_X est non seulement continue mais aussi différentiable, sa dérivée n'est autre que la fonction de densité f_X. On a ainsi la double relation:

$$f_X(x) = \frac{d\,F_X\,(x)}{dx} = \lim_{\Delta\downarrow 0} \frac{F_X(x+\Delta)-F_X(x)}{\Delta}$$
$$F_X(x) = \int_{-\infty}^{x} f_X(u)du$$

Si $f_X(x) > 0$, F_X est strictement croissante: $a_1 > a_2 \Rightarrow F_X(a_1) > F_X(a_2)$

Exercice : Commentez le graphique suivant :

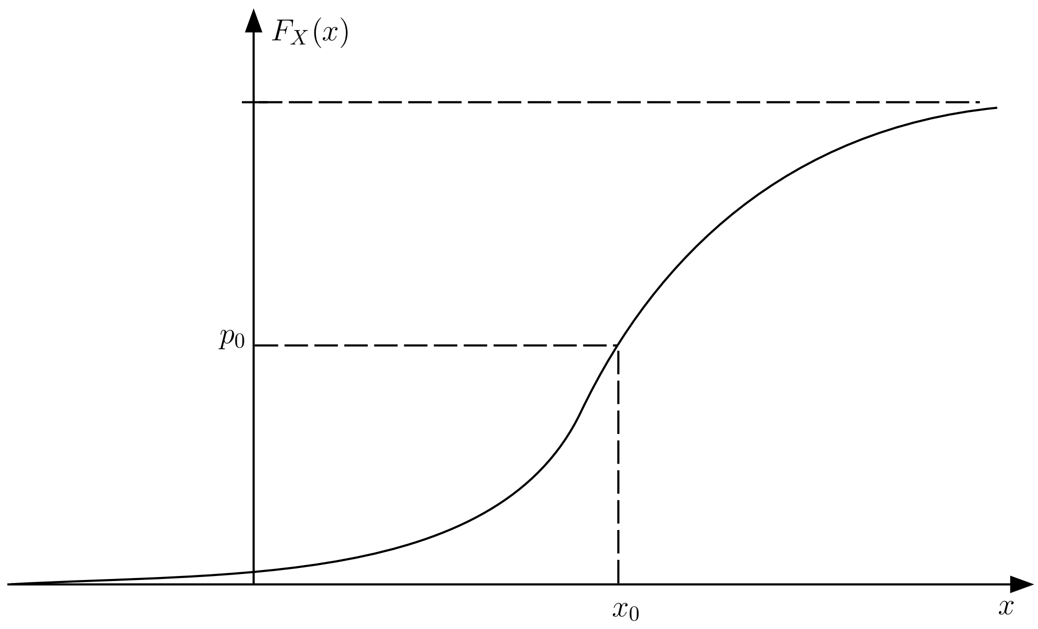


3.5 Fonction quantile d'une v.a.

3.5.1 Idées de base

1. Fonction inverse de la fonction de distribution
2. plus simple dans le cas continu avec F_X strictement croissante
3. déjà vus : $\begin{cases} \text{médiane} \\ \\ \text{quartile} \end{cases}$

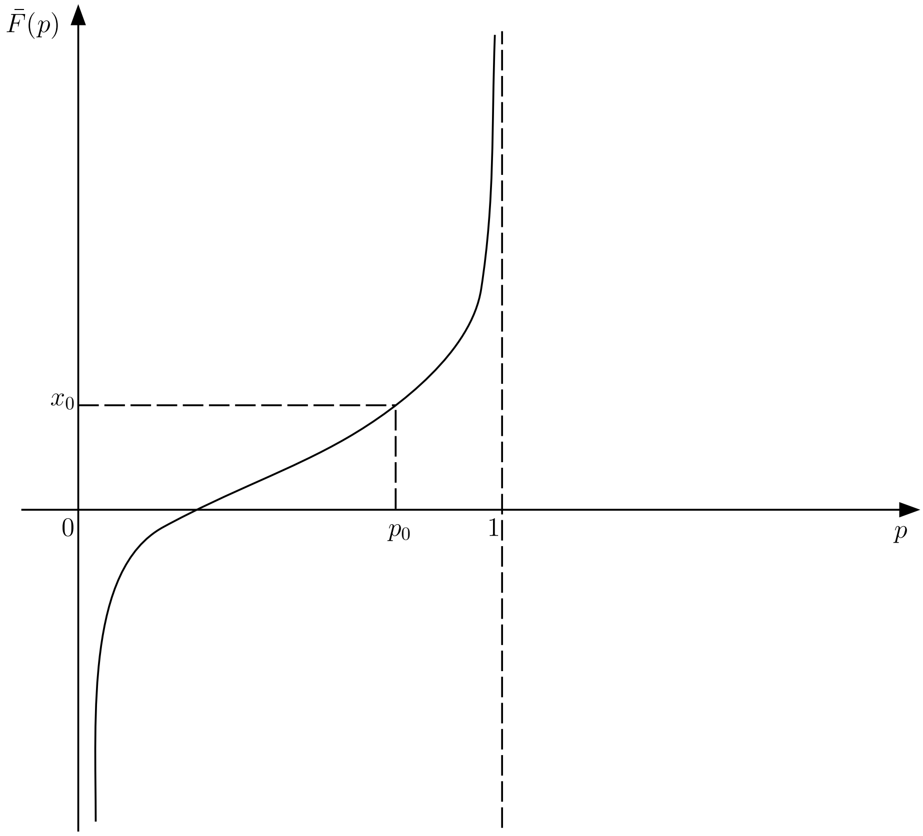
3.5.2 Définition



$F_X(x_0) = p_0 \Leftrightarrow P(X \leq x_0) = p_0$

$\Leftrightarrow x_0$ est tel que $P(X \leq x_0) = p_0$

on écrit $\overline{F}(p_0) = x_0$



On a donc:

$\overline{F}_X : [0,1] \to \mathbb{R}$
$F_X : \mathbb{R} \to [0,1]$

Dans le cas continu: $F_X(x_0) = p_0 \Leftrightarrow \overline{F}_X(p_0) = x_0$

Cas particuliers (importants!)

Médiane $= \overline{F}_X(0{,}5)$
1er quartile $= \overline{F}_X(0{,}25)$
3ème quartile $= \overline{F}_X(0{,}75)$

De plus

$p_0 > p_1 \Rightarrow \overline{F}_X(p_0) \geq \overline{F}_X(p_1)$

3.6 Espérance d'une v.a.

3.6.1 Idée de base:

équivalent de la moyenne arithmétique pour une v.a.

3.6.2 Notation

$E(X)$

Cas discret

$x_1 \cdots x_r$

$p_1 \cdots p_r$

$$E(X) = \sum_j x_j p_j$$

Cas continu

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx$$

Remarque: fonction indicatrice

Lorsque X est une v.a., $\mathbb{1}_A(X)$ est aussi une v.a. et on aura (toujours), quelque soit l'événement A:

$$P_X(A) = E[\mathbb{1}_A(X)]$$

Exercice: y réfléchir.

3.6.3 Propriété essentielle (rappel)

$$E[a + bX] = a + bE(X)$$

ou encore

$$\begin{aligned} Y &= a + bX \\ \underbrace{\mu_Y}_{\substack{\| \\ E(Y)}} &= a + b\,\underbrace{\mu_X}_{\substack{\| \\ E(X)}} \end{aligned}$$

En particulier

Soit: $\mu = E(X)$ alors: $E(X - \mu) = 0$ TOUJOURS

Exercice : Vérifiez ce cas particulier.

3.7 Variance

$$V(X) = E[(X - E(X))^2]$$

On utilisera aussi la notation:

$$\sigma_X^2 = E[(X - \mu_X)^2] \qquad (\mu_X = E(X))$$

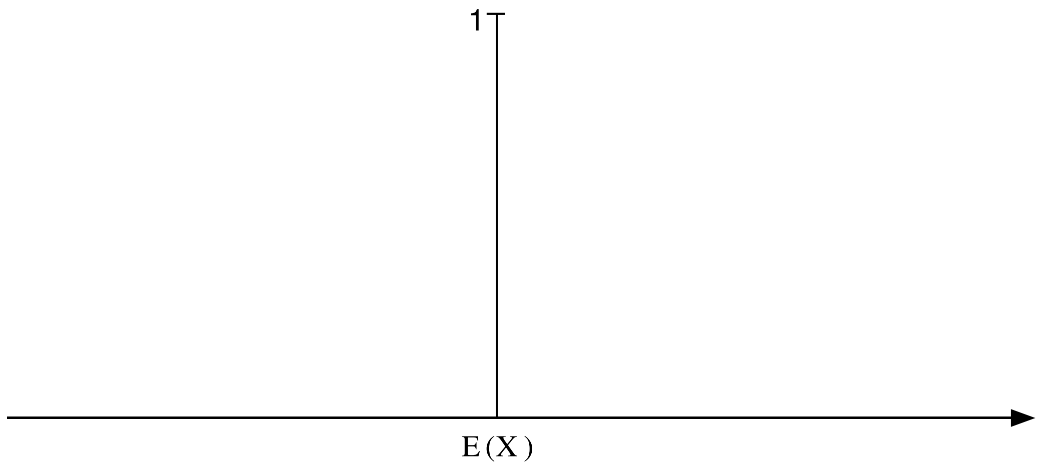
Remarques : • unités de $V(X)$ = carré des unités de X

• $\sigma_X = \sqrt{V(X)}$: mêmes unités que X

Propriétés

1. $V(X) \geq 0$

   $V(X) = 0 \Leftrightarrow P[X = E(X)] = 1$



2. $V(X) = E[X^2] - [E(X)]^2$
3. $V(a + bX) = b^2 V(X)$
   et donc: $\sigma_{a+bX} = |b|\sigma_X$

3.8 Standardisation d'une v.a.

Soit

une v.a. X telle que $E(X) = \mu \qquad V(X) = \sigma^2$

$Z = \dfrac{X - \mu}{\sigma}$

alors

$E(Z) = 0 \qquad V(Z) = 1$

Exercice : Vérifiez.

Inversément, soit

une v.a. Z telle que $E(Z) = 0 \qquad V(Z) = 1$

$X \; = \; \mu \; + \; \sigma \, Z$

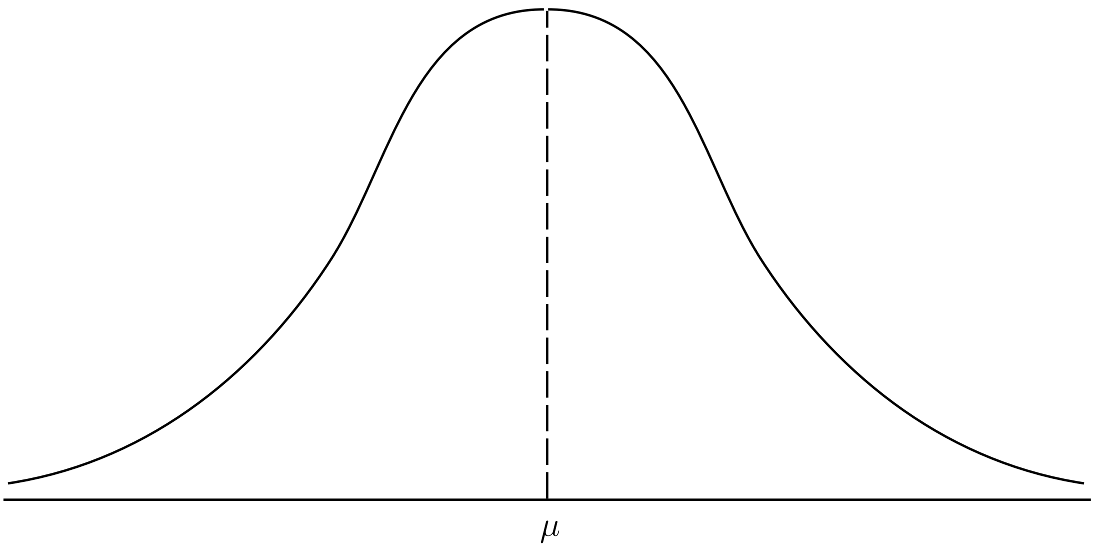
alors

$E(X) = \mu \qquad V(X) = \sigma^2$

3.9 Exemple: distribution normale

ou encore: distribution de Gauss ou de Gauss-Laplace

3.9.1 Fonction de densité



$$f_X(x) = (2\pi)^{-\frac{1}{2}}\sigma^{-1} \underbrace{\exp -\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}_{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}$$

$$\pi = 3{,}14159\cdots \qquad \mu \in \mathbb{R} \qquad \sigma > 0$$

A remarquer:

- symétrique autour de μ (nous y reviendrons)
- forme de cloche
- fines queues

On dira:

X est distribué selon $N(\mu,\sigma^2)$

Notation:

$X \sim N(\mu,\sigma^2)$

On peut vérifier

$\mu = E(X)$ "espérance mathématique"

$\sigma^2 = V(X)$

3.9.2 Première propriété importante

Si

$Y = a + bX$

$X \sim N(\mu_X,\sigma_X^2)$

alors

$Y \sim N(\mu_Y,\sigma_Y^2)$

Rappel de 1ère Candi.:

$$\mu_Y = a + b\mu_X \qquad \longleftrightarrow E(Y) = a + bE(X)$$

$$\sigma_Y^2 = b^2\sigma_X^2 \qquad \longleftrightarrow V(Y) = b^2V(X)$$

Cas particulier

$$Z = \frac{1}{\sigma}(X - \mu) = \underbrace{-\frac{\mu}{\sigma}}_{\substack{\uparrow \\ a}} + \underbrace{\frac{1}{\sigma}}_{\substack{\uparrow \\ b}} X$$

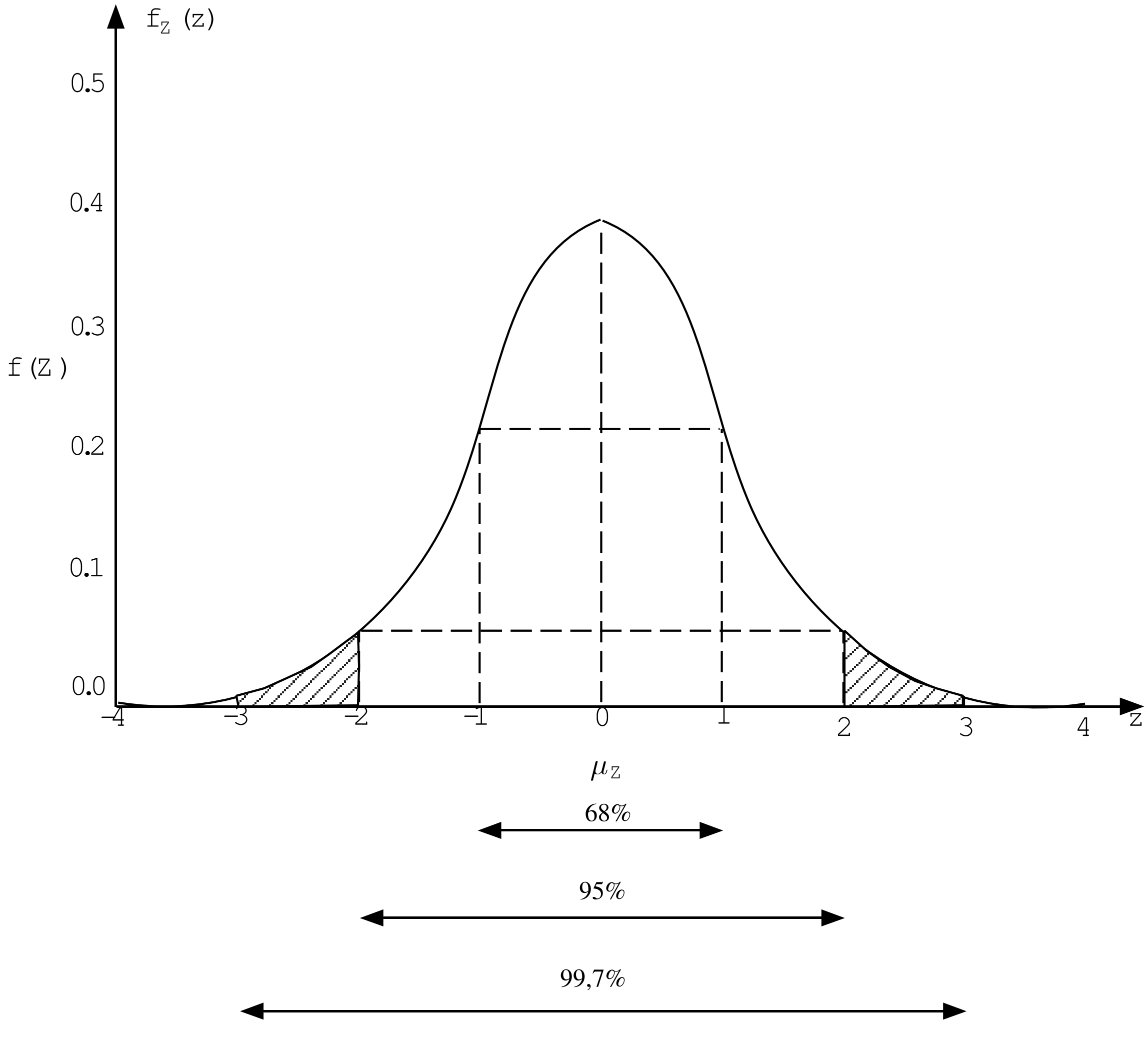
$\mu_Z = 0 \qquad \sigma_Z^2 = 1 \qquad$ "Z: V.A. standardisée"

inversément:

$$X = \mu + \sigma Z$$

Exercice: vérifier

Densité normale réduite $\mu_Z = 0$ et $\sigma_Z = 1$



3.9.3 Deuxième propriété importante

La distribution normale est "**symétrique**" autour de μ.

Au point de vue de la fonction de densité:

si $X \sim N(\mu,\sigma^2)$

alors $f_X(\mu + a) = f_X(\mu - a) \qquad \forall a \in \mathbb{R}$

Au point de vue de la fonction de distribution:

$$F_X(\mu - a) + F_X(\mu + a) \equiv 1 \qquad \forall a$$

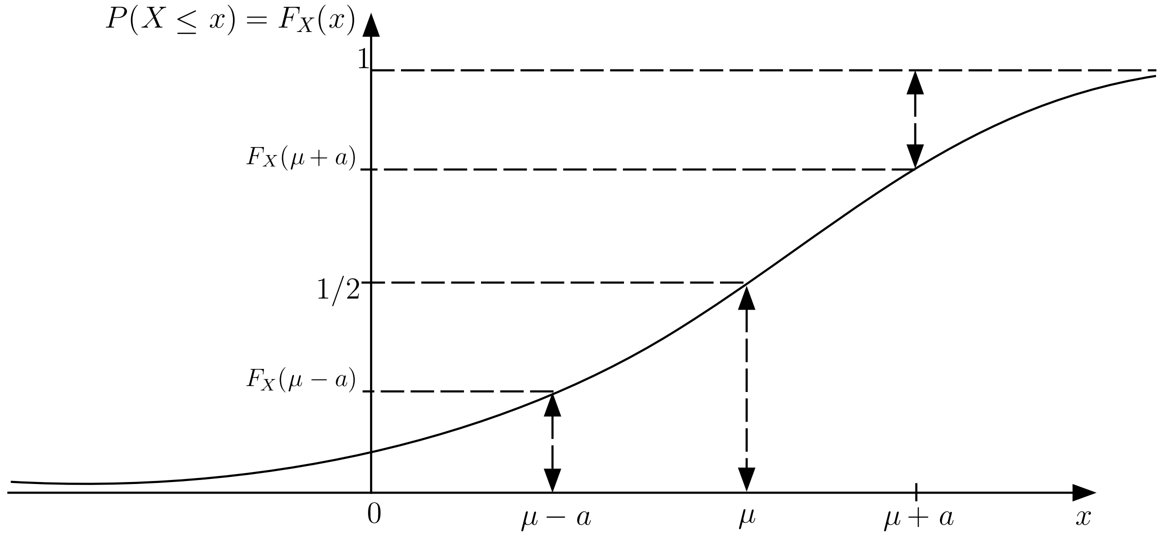
C'est-à-dire:

$$F_X(\mu - a) = \underbrace{1 - F_X(\mu + a)}$$

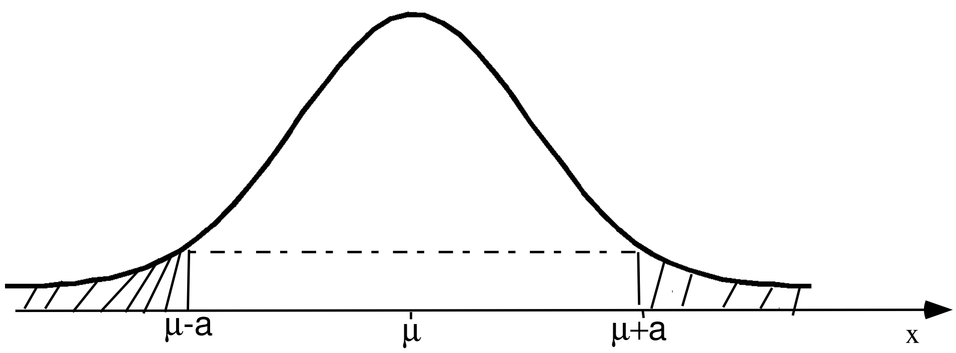
ou encore:

$$P(X \leq \mu - a) = P(X > \mu + a)$$

Donc: $\mu = E(X) = \text{Mode} = \text{Médiane}$



$P(X \leq x) = F_X(x)$
1
$F_X(\mu + a)$
1/2
$F_X(\mu - a)$
0
$\mu - a$
μ
$\mu + a$
x



μ-a
μ
μ+a
x

3.9.4 Emploi de la Table

(i) directement dans la table:

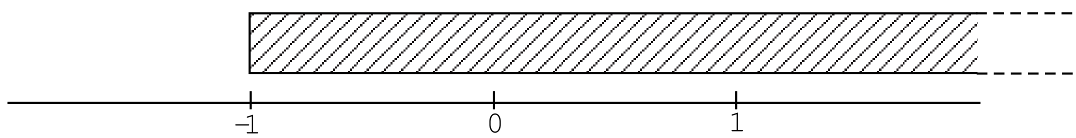
$$F_Z(a) = P(Z \leq a) \qquad \forall a \in \mathbb{R}_+$$

$$\forall a \geq 0$$

Exemple: $P(Z \leq 1) = 0{,}8413$

(ii) $a < 0$

Exemple: $P(Z \geq -1)$



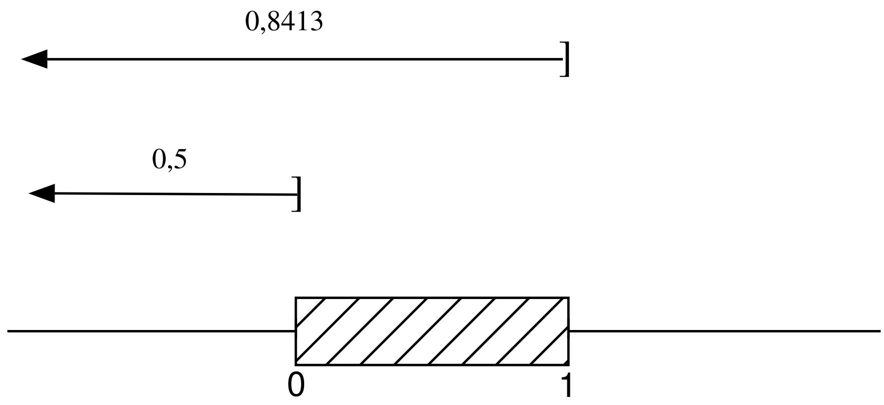
$P(Z \geq -1) = P(Z \geq 0) + P[-1 \leq Z \leq 0]$

$P(Z \geq 0) = P(Z \leq 0) = \dfrac{1}{2}$

$$\begin{aligned} P(-1 \leq Z \leq 0) = P(0 \leq Z \leq +1) &= P(Z \leq 1) - P(Z \leq 0) \\ &= 0{,}8413 - 0{,}5 \\ &= 0{,}3413 \end{aligned}$$

Donc:

$P(Z \geq -1) = 0{,}5 + 0{,}3413 = 0{,}8413$

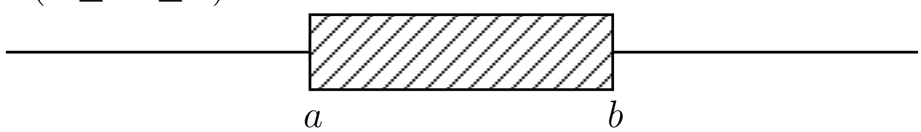


$$P(Z < 1) = 0{,}8413$$

Remarque: On pouvait aussi utiliser la propriété de symétrie pour le cas $\mu_Z = 0$:

$$P(Z \geq -1) = P(Z \leq +1) = 0{,}8413$$

(iii) $P(a \leq Z \leq b)$



$$\begin{aligned} P(a \leq Z \leq b) &= P(Z \leq b) - P(Z \leq a) \\ &= F_Z(b) - F_Z(a) \end{aligned}$$

Remarque: on a utilisé la propriété de v.a. continues $P(Z = a) = 0$ et donc $P(Z \leq a) = P(Z < a)$.

(iv) $P(a \leq X \leq b)$ lorsque $X \sim N(\mu,\sigma^2)$

$$\begin{aligned}
a \leq X \leq b \qquad & \Leftrightarrow \quad \frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma} \\
P(a \leq X \leq b) \quad & = \quad P\left[\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right]
\end{aligned}$$

Exemple:

$X \sim N(100,5^2)$

$$\begin{aligned}
P[95 \leq X \leq 110] \; & = P\left[\frac{95-100}{5} \leq Z \leq \frac{110-100}{5}\right] \\
& = P[-1 \leq Z \leq 2] \\
& = F_Z(2) - F_Z(-1) \\
& = 0{,}9772 - 0{,}1587 \\
& = 0{,}8185
\end{aligned}$$

Chapitre 4

Vecteur aléatoire

4.1 Introduction

Problème: souvent, dans une enquête, on s'intéresse à **plusieurs** caractéristiques

$\Rightarrow$ définir plusieurs V.A. sur les résultats d'une **même** expérience.

"un vecteur de variables aléatoires"

Vecteur: ensemble $\begin{cases} \text{fini} \\ \text{ordonné} \end{cases}$

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$$

Cas particulier $p = 2$ :: (X,Y)

par exemple X poids

Y taille

4.2 Distribution conjointe de X et Y

4.2.1 Motivation

Objectif : étudier la loi de probabilité des événements A et B (intervalles):

$$P(X \in A \text{ et } Y \in B)$$

on écrira aussi :

$$P(X \in A, Y \in B) \quad \text{ou} \quad P(\{X \in A\} \cap \{Y \in B\})$$

Cas particulier : lorsque A et B sont des intervalles:

Exemple (X: poids, Y: taille)

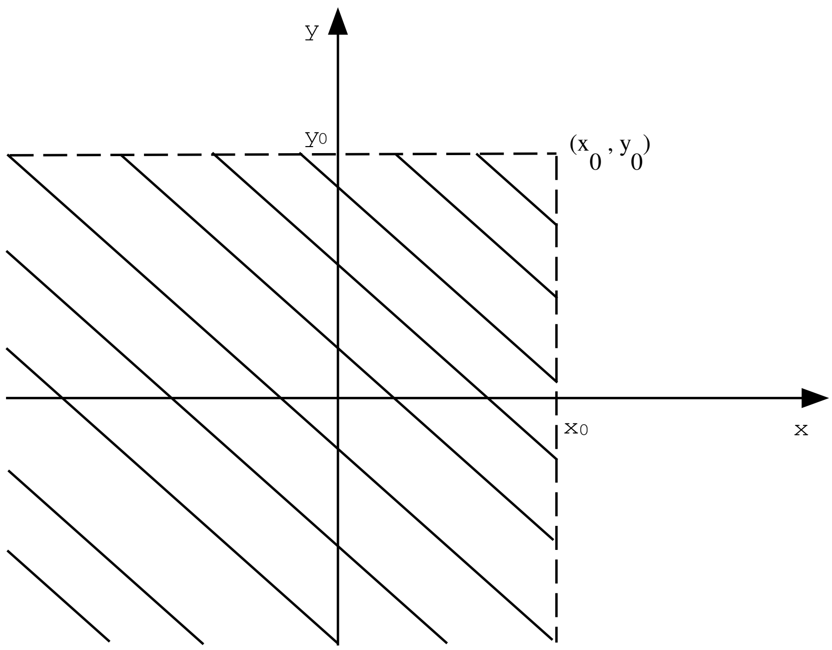
$$A = \{X \geq 55 \text{ kg }\} \qquad B = \{Y \leq 1 \text{ m } 80\}$$

4.2.2 Fonction de distribution (de répartition)

$$F_{X,Y}(x,y) = P[X \leq x, Y \leq y] \qquad \forall\, x,y \in \mathbb{R}$$

Par exemple

$$x = 60 \text{ kg} \qquad y = 1 \text{ m } 60$$



$$P(X \leq x_0, Y \leq y_0) = F_{X,Y}(x_0,y_0)$$

Propriétés

(i) $\forall x,y \qquad 0 \leq F_{X,Y}(x,y) \leq 1$

(ii) non décroissante

$\forall x$ fixé $,y_1 < y_2$: $\qquad F_{X,Y}(x,y_1) \leq F_{X,Y}(x,y_2)$

$\forall y$ fixé $,x_1 < x_2$: $\qquad F_{X,Y}(x_1,y) \leq F_{X,Y}(x_2,y)$

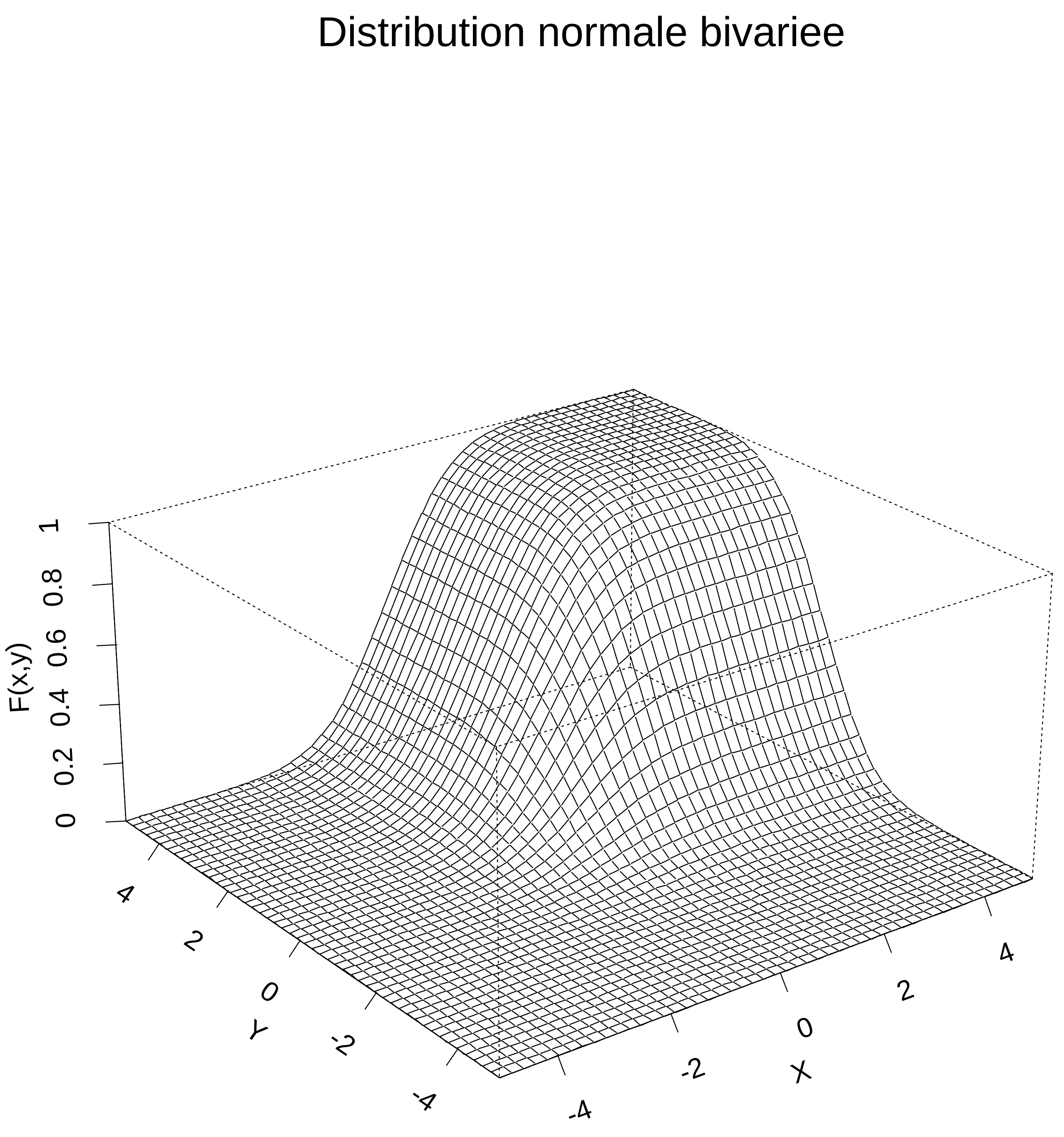
(iii) $\forall x$ fixé , $\lim\limits_{y \to -\infty} \qquad F_{X,Y}(x,y) = 0$

$\forall\ x$ fixé: $\lim\limits_{y \to +\infty} \qquad F_{X,Y}(x,y) = F_X(x)$: donc bien définie

$\forall\ y$ fixé: $\lim\limits_{x \to -\infty} \qquad F_{X,Y}(x,y) = 0$

$\forall\ y$ fixé: $\lim\limits_{x \to +\infty} \qquad F_{X,Y}(x,y) = F_Y(y)$: donc bien définie

F_X,F_Y : "distributions marginales"



Distribution normale bivariee
F(x,y)
0
0.2
0.4
0.6
0.8
1
Y
4
2
0
-2
-4
X
-4
-2
0
2
4

4.2.3 Fonction de probabilité et Fonction de densité.

Cas discret: Fonction de probabilité

Valeurs possibles $x_1 \cdots x_r \cdots x_g \qquad 1 \leq r \leq g$

$y_1 \cdots y_s \cdots y_h \qquad 1 \leq s \leq h$

Fonction de probabilité: $p_{r,s} = P[X = x_r, Y = y_s]$

Propriétés

$$p_{rs} \geq 0$$

$$\sum_r \sum_s p_{rs} = 1$$

Dès lors: $0 \leq p_{rs} \leq 1$

Fonction de distribution: $F_{X,Y}(x,y) = \sum_{\{r:x_r \leq x\}} \sum_{\{s:y_s \leq y\}} p_{rs}$

Cas continu: Fonction de densité

Fonction de densité bivariée $f_{X,Y}(x,y)$

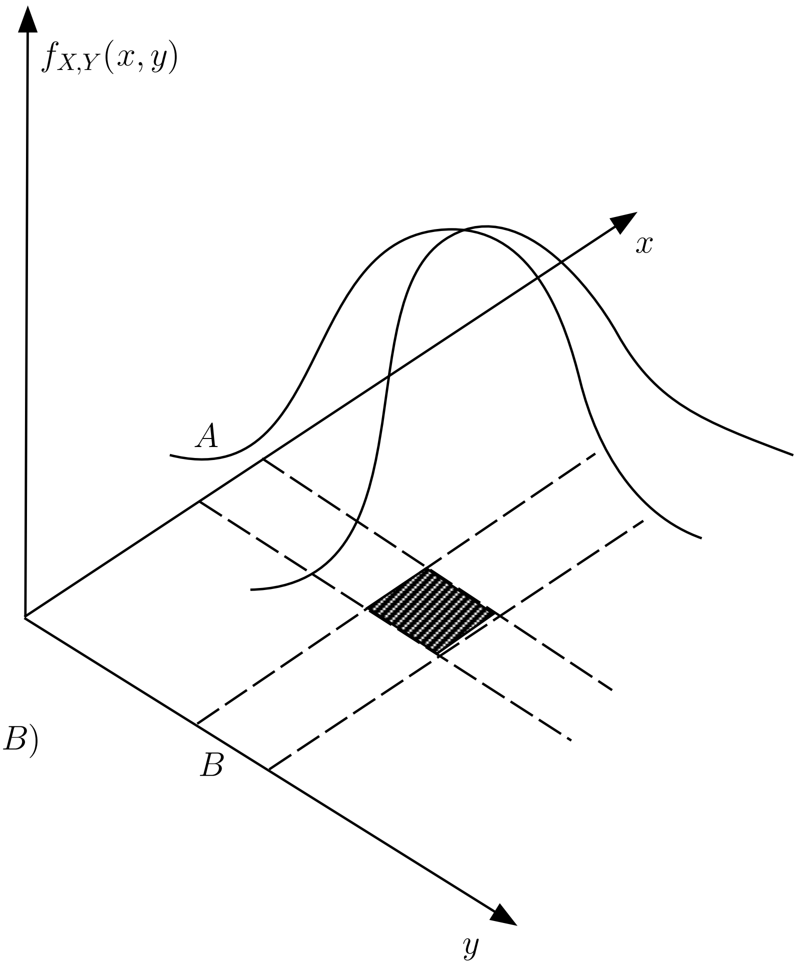
Idée de base: représenter, et calculer la fonction de distribution, et les probabilités conjointes, par une intégrale :

$$\begin{aligned} P[X \leq x, Y \leq y] &= F_{X,Y}(x,y) \\ &= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v) du\, dv \end{aligned}$$

Propriétés

$$f_{X,Y}(x,y) \geq 0$$

$$\lim_{\substack{x \to +\infty \\ y \to +\infty}} : F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)dudv \equiv 1$$



$P(X \in A \cap Y \in B)$

4.3 Distributions marginales

4.3.1 Cas discret:

Exemple simple

$X \in \{50, \quad 55, \quad 60\}$

$Y \in \{1m45, \quad 1m55\}$

p_{rs}	y_s		
	1m45	1m55	totaux : $p_{r.}$
50	0,15	0,10	0,25=$P(X = 50)$
x_r 55	0,35	0,15	0,50=$P(X = 55)$
60	0,10	0,15	0,25=$P(X = 60)$
totaux $p_{.s}$	0,60 $\Vert$ $P(Y = 1m45)$	0,40 $\Vert$ $P(Y = 1m55)$	1.

Distributions marginales: définition

$$P(X = x_r) = \sum_s p_{rs} = p_{r\cdot}$$

$$P(Y = y_s) = \sum_r p_{rs} = p_{\cdot s}$$

4.3.2 Cas continu

au niveau des fonctions de distribution

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P(X \leq x \text{ et } Y \leq \infty) \\ &= \lim_{y \to +\infty} F_{X,Y}(x,y) \\ F_Y(y) &= \lim_{x \to +\infty} F_{X,Y}(x,y) \end{aligned}$$

au niveau des fonctions de densité

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx$$

relations (déjà vues !)

$$F_X(x) = \int_{\infty}^{x} f_x(u)du = P(X \leq x)$$

4.4 Distributions conditionnelles

4.4.1 Cas discret

Exemple:
Reprenons l'exemple des distributions conjointes

En particulier 0,35 = $P[X = 55 \cap Y = 1m45]$

Parmi les étudiants tels que $Y = 1m45$:

$$\begin{aligned}
P[X = 50 \mid Y = 1m45] &= \frac{P[X = 50 \cap Y = 1m45]}{P[Y = 1m45]} \\
&= \frac{0{,}15}{0{,}60} = 0{,}25 \\
P[X = 55 \mid Y = 1m45] &= \frac{0{,}35}{0{,}60} \\
P[X = 60 \mid Y = 1m45] &= \frac{0{,}10}{0{,}60}
\end{aligned}$$

Plus généralement

$$P[X = x_r \mid Y = y_s] = \frac{p_{rs}}{p_{\cdot s}} = p_{r|s}$$

Exercices :
- Calculer de même la distribution conditionnelle de $(X \mid Y)$ lorsque Y=1m55.
- Sont-elles les mêmes lorsque Y=1m45 et lorsque Y=1m55 ?
- Calculer les distributions conditionnelles de $(Y \mid X)$.

Propriétés

Soit $p_{r|s} = P[X = x_r \mid Y = y_s] = \dfrac{p_{rs}}{p_{\cdot s}}$ $\qquad (p_{\cdot s} > 0)$

alors

(i) $p_{r|s} \geq 0 \qquad \forall r, \forall s$

(ii) $\forall s \qquad \sum_r p_{r|s} = 1$

(iii) "distribution marginale = moyenne de distributions conditionnelles"

c'est-à-dire

$$P(X = x_r) = p_{r\cdot} = \sum_s p_{r|s}\, p_{\cdot s}$$

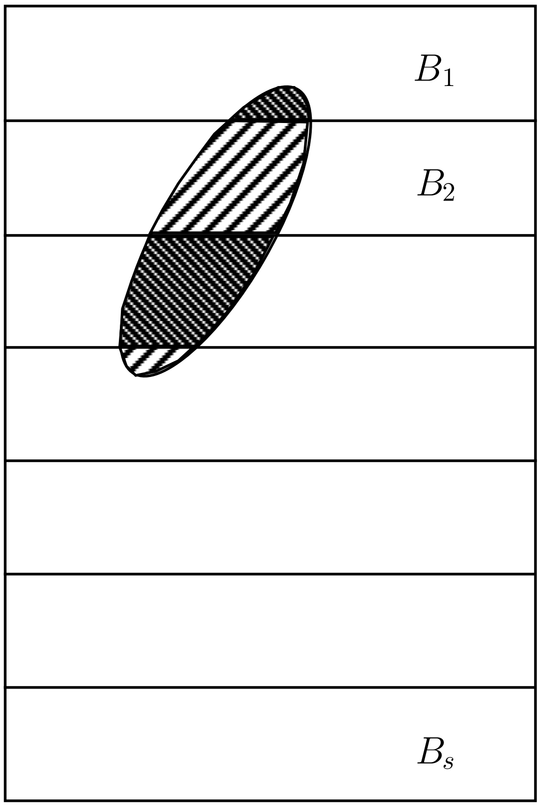
en effet

$$\begin{aligned} P(X = x_r) &= \sum_s P[X = x_r \text{ et } Y = y_s] \\ &= \sum_s P[X = x_r \mid Y = y_s] \quad P[Y = y_s] \end{aligned}$$

Exercice : Vérifiez avec les mêmes données que celles de l'exercice précédent.

Rappel : Soit $\{B_i\}$ partition,

$$\begin{aligned} P(A) &= \sum_i P(A \cap B_i) \\ &= \sum_i P(B_i) P(A|B_i) \\ P(A|B_i) &= \frac{P(A \cap B_i)}{P(B_i)} \end{aligned}$$



Remarque: Pour r fixé, la probabilité conditionnelle $P(A_r|B_i)$ est donc une V.A. dont la distribution est donnée par les probabilités $P(B_i)$ et dont l'espérance mathématique vaut $P(A_r)$.

4.4.2 Cas continu

Objectif: représenter, et calculer, les probabilités conditionnelles par une intégrale.

Soit $A = [a,b]$

$$P[X \in A \mid Y = y] = \int_a^b f_{X|Y}(x \mid y)dx$$

où $f_{X|Y}(x \mid y)$ "densité conditionnelle" :

Définition

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \qquad (f_Y(y) > 0)$$

Propriétés

i) $f_{X|Y}(x \mid y) \geq 0 \qquad \forall x, \forall y$

ii) $\displaystyle\int_{-\infty}^{+\infty} f_{X|Y}(x \mid y)dx = 1 \qquad \forall y$

iii) $\displaystyle f_X(x) = \int_{-\infty}^{+\infty} f_{X|Y}(x \mid y)f_Y(y)dy$

4.4.3 Espérance mathématique

Rappel:

$$E(X) = \sum_r x_r P(X = x_r) = \sum x_r p_r$$

$$= \int x f_X(x) dx$$

Plus généralement:

Soit (X,Y)

$$E\left[g(X,Y)\right] = \sum_r \sum_s g(x_r,y_s) p_{rs}$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y) f_{X,Y}(x,y) dx dy$$

Propriétés

- déjà vu: $E[a + bX] = a + bE(X)$

•• de plus: $E[aX + bY + c] = aE(X) + bE(Y) + c$

Espérance conditionnelle

$$\begin{aligned} E[X|Y = y_s] &= \sum_r x_r P(X = x_r|Y = y_s) \\ &= \sum_r x_r p_{r|s} \\ &= \int x f_{X|Y}(x|y)dx \end{aligned}$$

Propriétés

- linéarité $E(a + bX|Y = y_s) = a + bE(X|Y = y_s)$

•• $E[X] = E[E(X|Y)]$

"espérance marginale = moyenne des espérances conditionnelles"

$E(X|Y)$ est donc une fonction de Y, appelée "fonction de régression":

$$\mu_X(y) : \ y \longmapsto E(X|Y = y)$$

C'est donc une v.a. dont l'espérance mathématique vaut $E(X)$

Dans le cas discret:

$$\begin{aligned} E[X] &= \sum x_r p_r \\ &= \sum_s P[Y = y_s] \cdot E[X|Y = y_s] \end{aligned}$$

Exercice: Vérifier que $(0.60)E(X|Y = 1m45) + (0.4)E(X|Y = 1m55) = 55$

Exemple de calcul

$x\backslash y$	1m45	1m55	$P(X=x)$	$xP(X=x)$
50	.15	.10	.25	12.5
55	.35	.15	.50	27.5
60	.10	.15	.25	15
	.60	.40		55.0 ‖ E(X)
$E(X\|Y)$ 1) $P(X=x\|Y)$				
$x=50$	$\frac{15}{60}$	$\frac{10}{40}$		
$x=55$	$\frac{35}{60}$	$\frac{15}{40}$		
$x=60$	$\frac{10}{60}$	$\frac{15}{40}$		
2) $xP(X=x\|Y)$				
$x=50$	$\frac{50\times 15}{60}$	$\frac{50\times 10}{40}$		
$x=55$	$\frac{55\times 35}{60}$	$\frac{55\times 15}{40}$		
$x=60$	$\frac{60\times 10}{60}$	$\frac{60\times 15}{40}$		
Total	$E(X|Y=$1m45$)$	$E(X|Y=$1m55$)$		

4.4.4 Variance conditionnelle

Soit X V.A. numérique (discrète ou continue)

Y V.A. quelconque.

Pour tout y, valeur possible de Y, on a une distribution conditionnelle qui permet, entre autre, d'évaluer:

$$P(X \in A|Y = y)$$

Elle permet aussi de calculer:

- une espérance conditionnelle $E(X|Y)$
- une variance conditionnelle $V(X|Y)$

Ce sont donc des **variables aléatoires** qui sont, en fait, des fonctions d'Y. Intéressons-nous aux espérances et variances de ces variables aléatoires.

Pour l'espérance conditionnelle $E(X|Y)$, nous avons déjà vu:

$$E(X) = E[E(X|Y)]$$

De plus, la variance de $E(X|Y)$ et l'espérance de $V(X|Y)$ sont reliées par la relation fondamentale de la **décomposition de la variance** (marginale de X):

$$
\begin{array}{ccccc}
V(X) & = & E[V(X|Y)] & + & V[E(X|Y)] \\
 & & \downarrow & & \downarrow \\
\begin{bmatrix} \text{variance} \\ \text{"totale"} \end{bmatrix} & = & \begin{bmatrix} \text{variance} \\ \text{"intraclasse"} \end{bmatrix} & + & \begin{bmatrix} \text{variance} \\ \text{"interclasse"} \end{bmatrix}
\end{array}
$$

Deux cas limites :

$E\,[V(X|Y)] = 0 \quad \Rightarrow \quad X = f(Y)$
Exemple: X = nombre de frères et et soeurs d'un étudiant, Y = nombre d'enfants de la famille de ce même étudiant.

$V\,[E(X|Y)] = 0 \quad \Rightarrow \quad E[X|Y] = E[X]$
Exemple: X = nombre de coeurs dans une main de bridge, Y = nombre d'honneurs dans une main de bridge.

Chapitre 5

Indépendance et association

5.1 Introduction

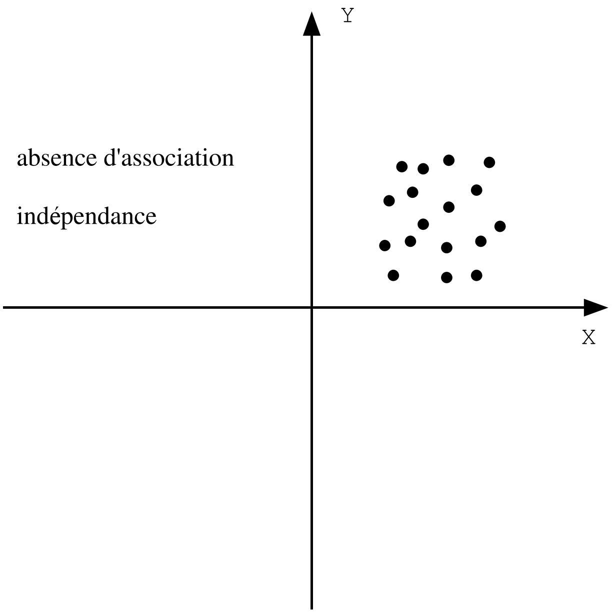
- Jusqu'à présent: échantillonnage sur 1 seule V.A.

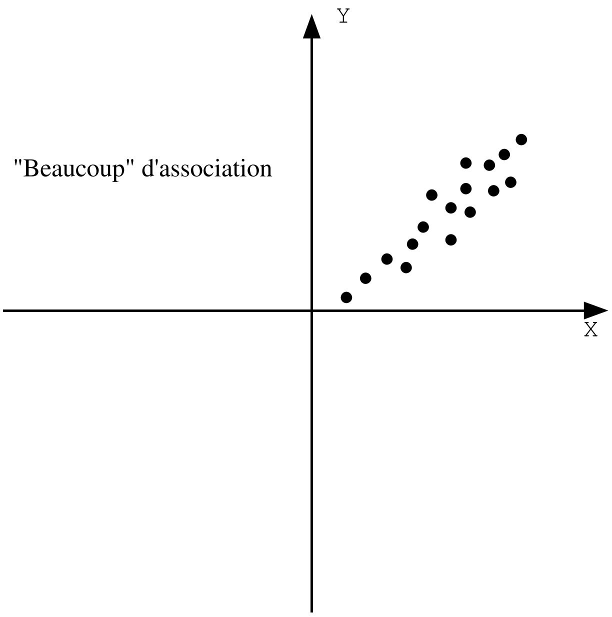
- Ce chapitre: plusieurs V.A. simultanément
  plus précisément: ici 2 V.A. (X,Y)

- Remarque: à distinguer
  - X et Y continue
  - X et Y discrète
  - X continue et Y discrète

- Dans ce chapitre: "association statistique" ou "indépendance en probabilité"?

$$P(Y|X) \stackrel{?}{=} P(Y)$$





Question: comment mesurer l'intensité de l'association?

5.2 Indépendance entre variables aléatoires

Le problème

$$\text{Axe} \left[\begin{array}{l} \text{indépendance} \\ \text{association} \end{array} \right. \quad \text{entre V.A.}$$

Définition

X et Y sont "indépendants"

Notation $X \perp\!\!\!\perp Y$

En général
pour des intervalles quelconques A et B

$\{X \in A\} \perp\!\!\!\perp \{Y \in B\}$

$A \perp\!\!\!\perp B$: revoir cours 1ére Candi.

$$X \perp\!\!\!\perp Y \Leftrightarrow F_{X,Y}(x,y) = F_X(x)F_Y(y) \qquad \forall\ x,y$$

c'est-à-dire:

$$P[X \leq x, Y \leq y] = P[X \leq x] \cdot P[Y \leq y]$$

$$A = \{X \leq x\} = \{X \in (-\infty,x)\}$$

$$B = \{Y \leq y\} = \{Y \in (-\infty,y)\}$$

V.A. discrète: $P(X = x_r, Y = y_s) = p_{rs}$

$$X \perp\!\!\!\perp Y \;\Leftrightarrow p_{rs} = p_{r\cdot} p_{\cdot s} \qquad \forall\, r,s$$

$$\Leftrightarrow p_{r|s} \;=\; p_{r.} \qquad \forall\, r,s$$

$$\Leftrightarrow p_{s|r} \;=\; p_{.s} \qquad \forall\, r,s$$

V.A. continue: soit $f_{X,Y}(x,y)$

$$X \perp\!\!\!\perp Y \;\Leftrightarrow f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) \qquad \forall\, x,y$$

$$\Leftrightarrow f_{X|Y=y}(x) = f_X(x) \qquad \forall\, x,y$$

$$\Leftrightarrow f_{Y|X=x}(y) = f_Y(y) \qquad \forall\, x,y$$

Remarques

- X et Y sont indépendants "en probabilité" ou "stochastiquement"
- •• concept symétrique entre X et Y (éviter de dire: X est indépendant de Y)

Propriétés: Si $X \perp\!\!\!\perp Y$, alors :

1. $g(X) \perp\!\!\!\perp h(Y)$ quels que soient g et h.

   En particulier: $E(X\,Y) = E(X)E(Y)$

   $$E[g(X).h(Y)] = E[g(X)].E[h(Y)]$$

2. $V(X+Y) = V(X) + V(Y)$

3. $E(Y|X) = E(Y) =$ constante

   Et donc : $V[E(Y|X)] = 0$

   Et donc : $V(Y) = E\, V(Y|X)$

Attention ! L'inverse est faux (Exercice : commentez)

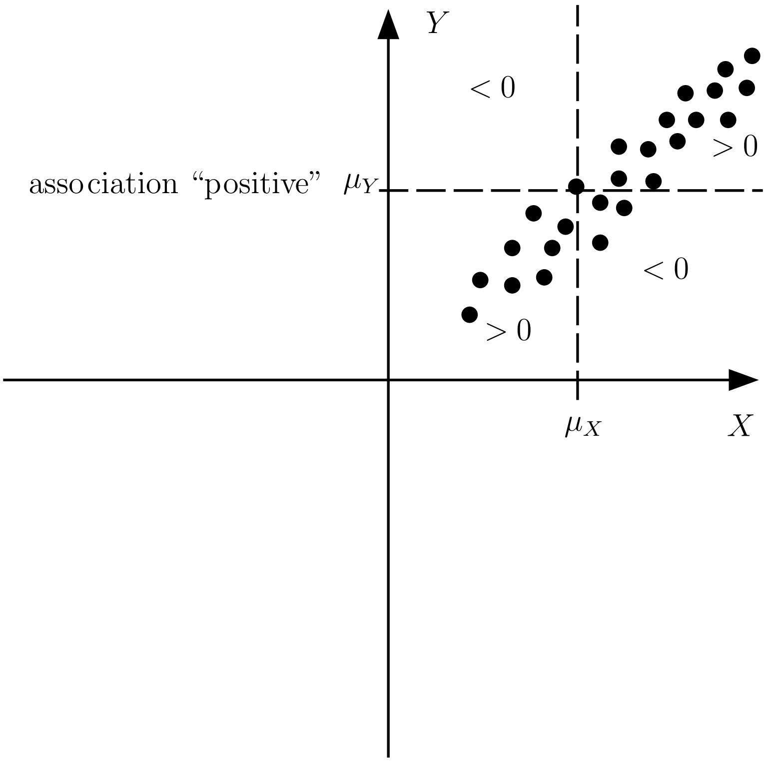
5.3 Corrélation - Covariance

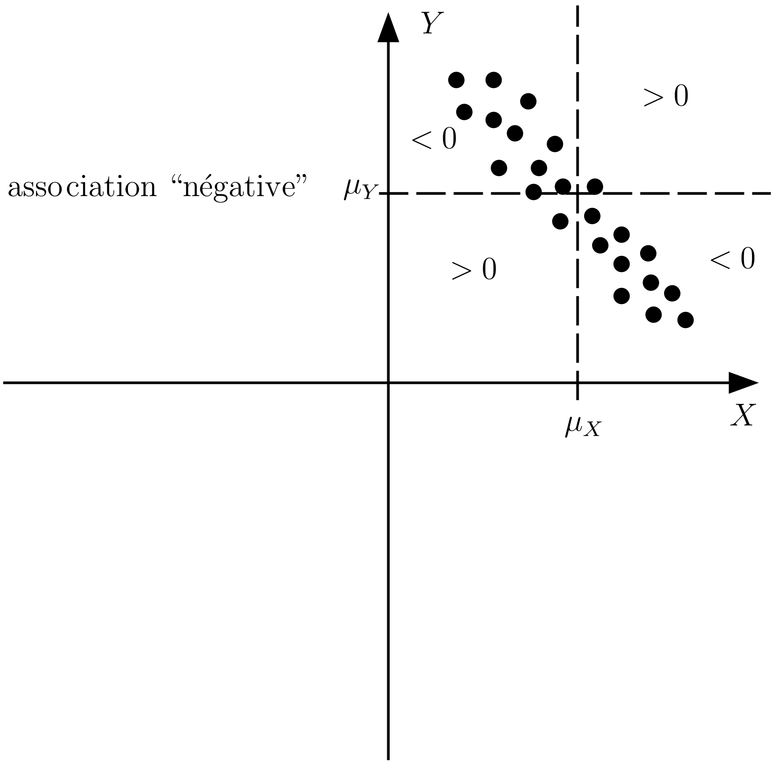
5.3.1 Introduction graphique

Soit X et Y: V.A. continue(s) ou discrète(s) mais numériques.

Etudions le signe de $(X-\mu_X)(Y-\mu_Y)$:

X,Y $\rightarrow$	$(X-\mu_X)$	$(Y-\mu_Y)$	signe du produit
	>0	$>0 \Rightarrow$	>0
	<0	$<0 \Rightarrow$	>0
	<0	$>0 \Rightarrow$	<0
	>0	$<0 \Rightarrow$	<0


Y
< 0
> 0
association “positive”
μ_Y
< 0
> 0
μ_X
X


Y
> 0
< 0
association “négative”
μ_Y
> 0
< 0
μ_X
X

5.3.2 Covariance entre X et Y

Définition:

$$\begin{aligned} C(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)] \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

Propriétés:

1. $X \perp\!\!\!\perp Y \Rightarrow C(X,Y) = 0$: toujours [*Exercice :* démontrez !]
   $X \perp\!\!\!\perp Y \Leftarrow C(X,Y) = 0$: faux

2. Symétrie: $C(X,Y) = C(Y,X)$

3. Linéarité: $C(aX + b,Y) = a\, C(X,Y)$

   $$C(X + Z,Y) = C(X,Y) + C(Y,Z)$$

4. $V(X) = C(X,X)$

5. $V(X \pm Y) = V(X) + V(Y) \pm 2\, C(X,Y)$

   en particulier (rappel !)

   $X \perp\!\!\!\perp Y \Rightarrow V(X + Y) = V(X) + V(Y)$

6. $\mathrm{COV}(X,Y) = \mathrm{COV}(X,E(Y|X)) = \mathrm{COV}(E(X|Y),Y)$
   Dès lors, si $E(Y|X) = E(Y) =$ constante, $\mathrm{COV}(X,Y) = 0$ bien que $E(Y|X) = E(Y)$ n'implique pas $X \perp\!\!\!\perp Y$ (vu en

propriété 1).

7. $[\text{COV}(X,Y)]^2 \leq V(X) \cdot V(Y)$

   Dès lors $V(Y) = 0 \qquad \Rightarrow \qquad C(X,Y) = 0 \qquad \forall : X$

   Ou encore : $C(X,a) = 0 \qquad \forall : X$

5.3.3 Corrélation (linéaire)

Problème:

Est-ce que $C(X,Y) = 1000$ indique "beaucoup" d'associations?

Difficulté: on peut interpréter le signe (> 0 ou < 0) mais pas la valeur absolue, car elle dépend des unités de X et Y.

Solution: coefficient de corrélation

$$\begin{aligned}\rho(X,Y) &= C\left(\frac{X-\mu_X}{\sigma_X},\frac{Y-\mu_Y}{\sigma_Y}\right)\\ &= \frac{C(X,Y)}{\sqrt{V(X)\,V(Y)}}\end{aligned}$$

Propriétés

1. *standardisation*: $-1 \leq \rho(X,Y) \leq +1$
   car :$|C(X,Y)| \leq \sqrt{V(X)V(Y)}$
2. *symétrie*: $\rho(X,Y) = \rho(Y,X)$
3. $\rho(X,Y) = 0 \Leftrightarrow C(X,Y) = 0$
4. *transformation linéaire*: $\rho(aX+b,Y) = \dfrac{a}{|a|}\rho(X,Y)$

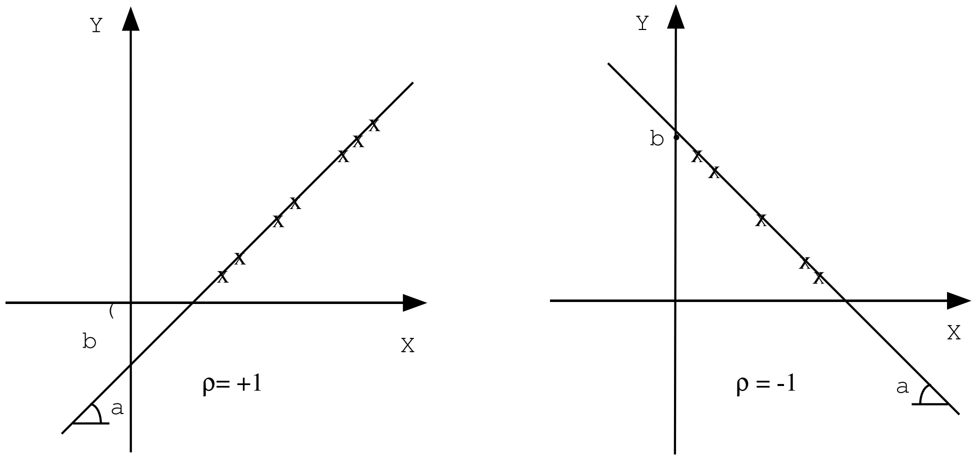
Remarque: $\dfrac{a}{|a|} = \pm 1$: indicateur du signe de a

De plus:

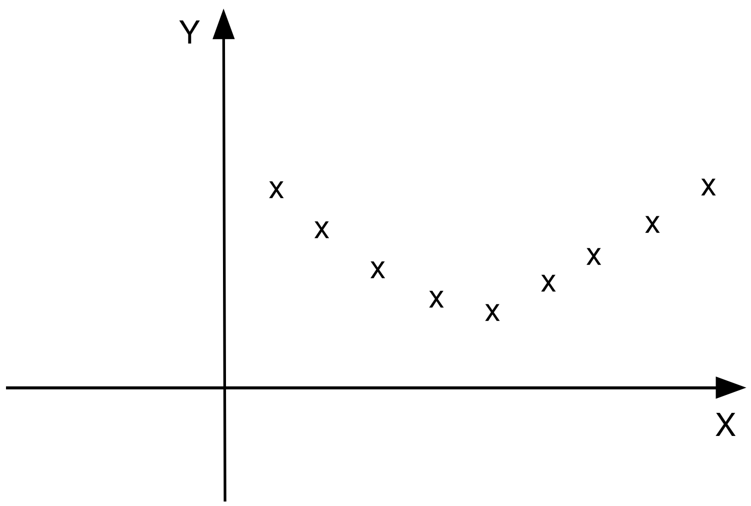
$$\rho = +1 \quad \Leftrightarrow \quad \exists\, a > 0, \exists\, b : \quad Y = aX + b$$

$$\rho = -1 \quad \Leftrightarrow \quad \exists\, a < 0, \exists\, b : \quad Y = aX + b$$

$$\rho = \pm 1 \quad \Rightarrow \quad V(X|Y) = V(Y|X) = 0$$



Remarque:



$C(X,Y) = 0 \quad \rho(X,Y) = 0$
et cependant $Y = f(X)$ ou encore $V(Y|X = x) = 0 \qquad \forall\, x$
MAIS: $E\ V(X|Y) > 0$

5.4 Indice de corrélation

5.4.1 Soit

Y numérique

X quelconque

5.4.2 Rappel

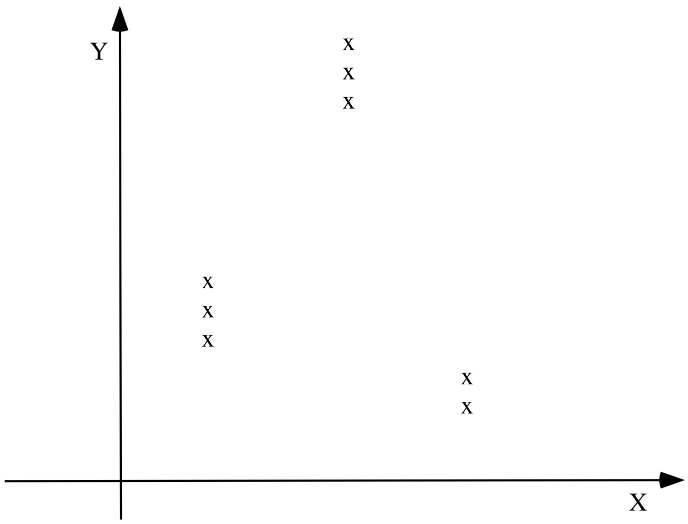
$V(Y) = E\,[V(Y|X)] + V\,[E(Y|X)]$

5.4.3 Interprétation

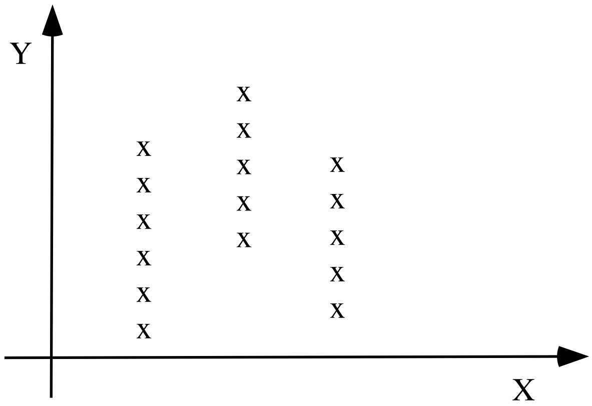
$E\;V(Y|X)$ "petit": beaucoup d'associations

$\Updownarrow$

$V\;E(Y|X)$ "grand"



Exercice : Commenter le graphique suivant, en termes d'association



5.4.4 Définition

$$\eta_{Y|X} = \frac{V\ E(Y|X)}{V(Y)} = 1 - \frac{EV(Y|X)}{V(Y)}$$

Exercice : Vérifiez la seconde égalité

Exercice : Calculez $\eta_{Y|X}$ et $\eta_{X|Y}$ pour les données numériques de la section 4.3.1.

5.4.5 Propriétés

1. **Standardisation**: $\eta_{Y|X} \in [0,1]$

   Ce coefficient ne mesure donc pas de *"direction"* dans l'association

2. **Non-symétrie**: en général: $\eta_{Y|X} \neq \eta_{X|Y}$

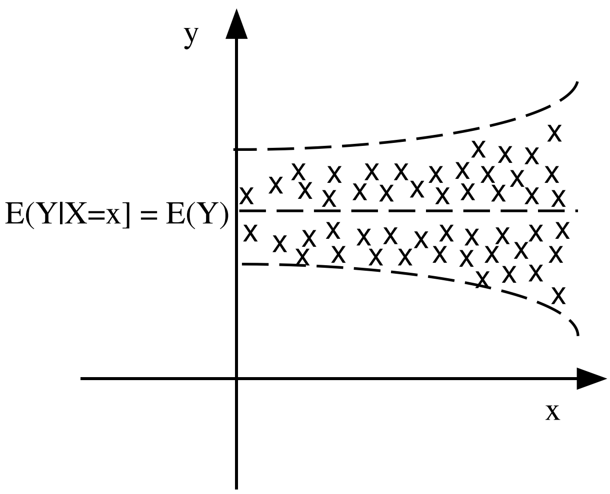
3. $\boldsymbol{\eta_{\mathbf{Y|X}} = 0} \Leftrightarrow VE(Y|X) = 0$

   $$\Leftrightarrow E(Y|X) = E(Y) = \text{constante}$$

   donc $Y \perp\!\!\!\perp X \Rightarrow \eta_{Y|X} = 0$

   Mais l'inverse est faux.

   Exemple graphique:



   Dans cet exemple :

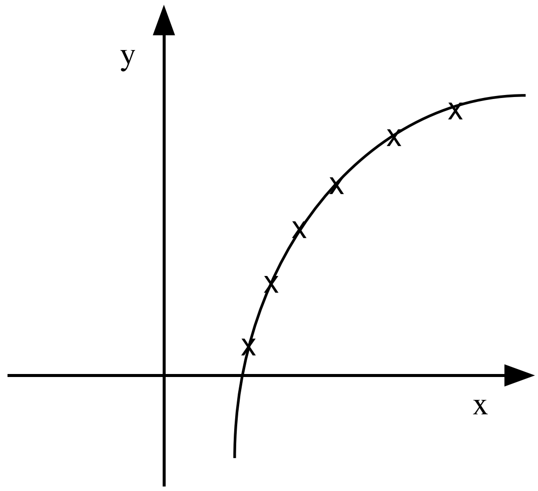
$$E[Y|X] = E[Y] \Rightarrow \begin{array}{l} V[E(Y|X)] = 0 \\ \mathrm{COV}(E(Y|X),X) = 0 \\ \mathrm{COV}(Y,X) = 0 \end{array}$$

MAIS $X \not\perp\!\!\!\perp Y$ car $V(Y|X)$ pas constant

4. $\eta_{\mathbf{Y|X}} = \mathbf{1} \Leftrightarrow E\,V(Y|X) = 0$

$$\Leftrightarrow V(Y|X = x) = 0 \qquad \forall : x$$

$$\Leftrightarrow \exists\, f : \quad Y = f(X)$$



5. **Cas linéaire**:

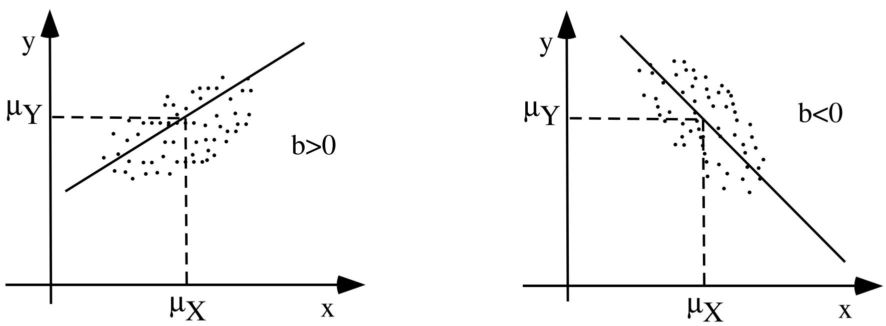
Si Fonction de régression linéaire: $E(Y|X) = a + bX$, alors:

$$\eta_{Y|X} = \eta_{X|Y} = [\rho(Y,X)]^2$$

$$b = \frac{C(Y,X)}{V(X)}$$

$$a = \mu_Y - b\mu_X \quad \Leftrightarrow \mu_Y = a + b\mu_X$$

Graphiquement:



Chapitre 6

Description de l'échantillonnage

6.1 Population et paramètres

6.1.1 Le point de vue de ce chapitre: échantillonnage en population finie

Population (finie)

$$\begin{aligned} \mathcal{N} &= \{1,2,\cdots N\} \\ &= \{k : 1 \leq k \leq N\} \end{aligned}$$

Remarque: "k" est une "étiquette" qui identifie chaque élément de la population

Caractéristique (ou: "paramètre") d'individus

$$\xi_k \qquad \eta_k \qquad \zeta_k \cdots$$

Remarque: ξ,η et ζ sont des lettres grecques lues : ksi, êta, zêta.

En résumé, on a ainsi défini :

$$\{(k,\xi_k,\eta_k,\zeta_k) : k \in \mathcal{N}\}$$

6.1.2 Différents types de variables statistiques(Rappel!)

- **variables qualitatives (ou catégorielles)**:
  - nominale
  - ordinale

  Exercice : ces variables induisent une partition finie de la population $\mathcal{N}$. Pourquoi ?

- **variables numériques (ou quantitatives)**
  - discrètes (dénombrables) "de comptage"
  - continues
    1. probabilité d'un poids "précis" $\equiv 0$
    2. probabilité d'un intervalle : seul intérêt.

6.2 Distribution de population

Considérons une population de caractéristiques: $\{\xi_1,\xi_2 \cdots \xi_N\}$

6.2.1 Effectif-fréquence

effectif (théorique)

ν_r: nombre d'individus dans la population qui ont la modalité r

Fréquence

$$\pi_r = \frac{\nu_r}{N}$$

Remarquons

$\nu_r \in I\!N = \{0,1,2,\cdots\}$ (entier naturel) $\qquad \sum_r \nu_r = N$

$\pi_r \in [0,1] \qquad$ et $\qquad \sum_r \pi_r = 1$

De plus, si toutes les caractéristiques ξ_k ont une valeur différente:

$$\nu_r = 1$$

$$\pi_r = \frac{1}{N}$$

6.2.2 Fonction de répartition ou de distribution

Soit ξ, une caractéristique numérique réelle

$\nu_\xi(x)$: nombre d'individus tels que la caractéristique $\xi \leq x$

Définition:

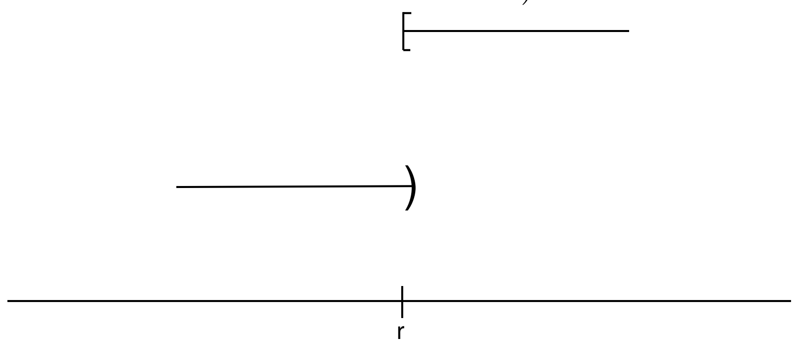
$$F_\xi(x) = \frac{\nu_\xi(x)}{N}$$

"Fonction de distribution (ou répartition)"de la population.

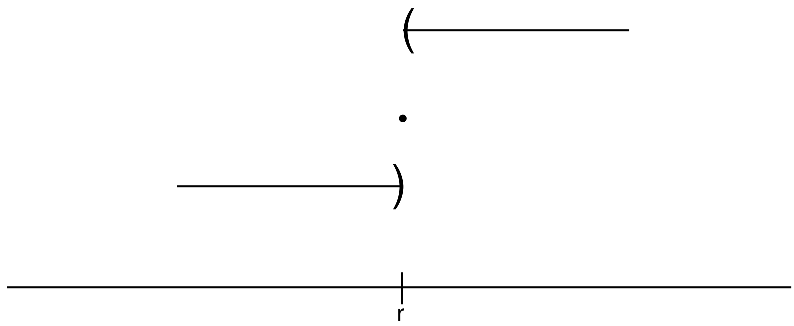
Propriétés

1. $\forall x \qquad 0 \leq F_\xi(x) \leq 1$
2. non décroissante: $x_1 < x_2 \quad \Rightarrow F_\xi(x_1) \leq F_\xi(x_2)$
3. $\lim\limits_{x \to -\infty} F_\xi(x) = 0$

   $\lim\limits_{x \to +\infty} F_\xi(x) = 1$
4. lorsqu'il y a un saut au point "r": la valeur de la fonction est la valeur supérieure (plus exactement: sa valeur immédiatement à la droite de r):



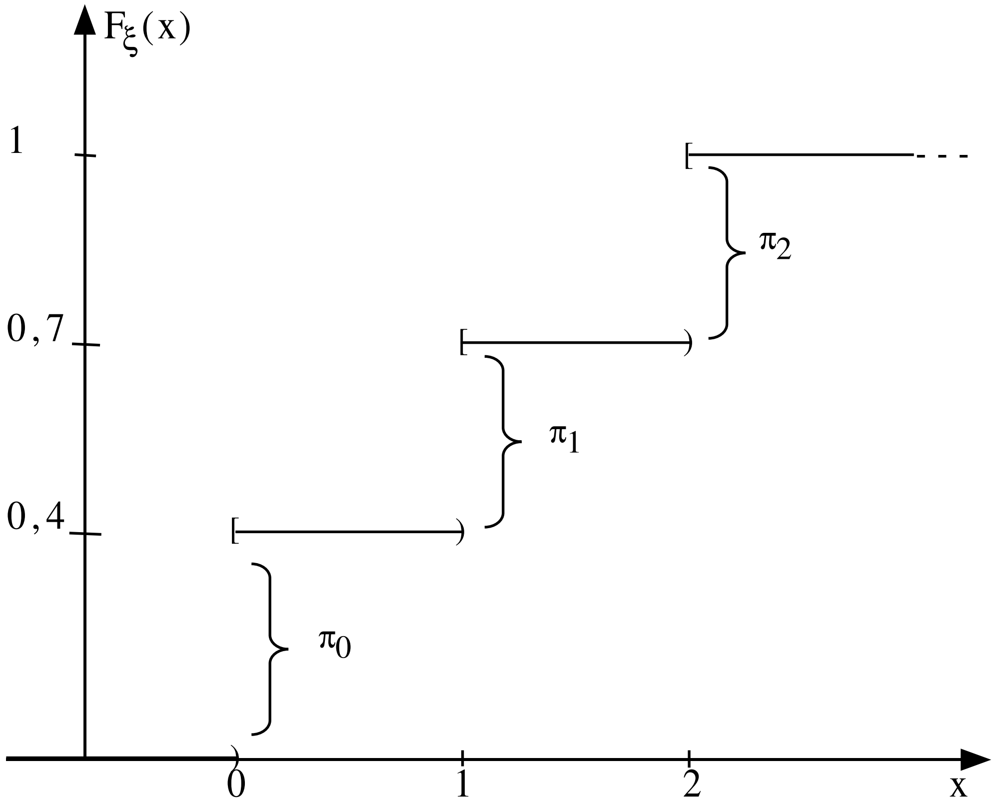
on exclut donc:



Remarque (pour les étudiants plus exigeants !): la fonction de distribution présente un "saut" aux points correspondants à des valeurs réalisées de la variable ξ. Le comportement spécifique de la fonction de distribution en ces points de discontinuité peut être exprimé de la façon suivante: $\forall x : \quad \lim\limits_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} \quad F_\xi(x + \varepsilon) = F_\xi(x)$

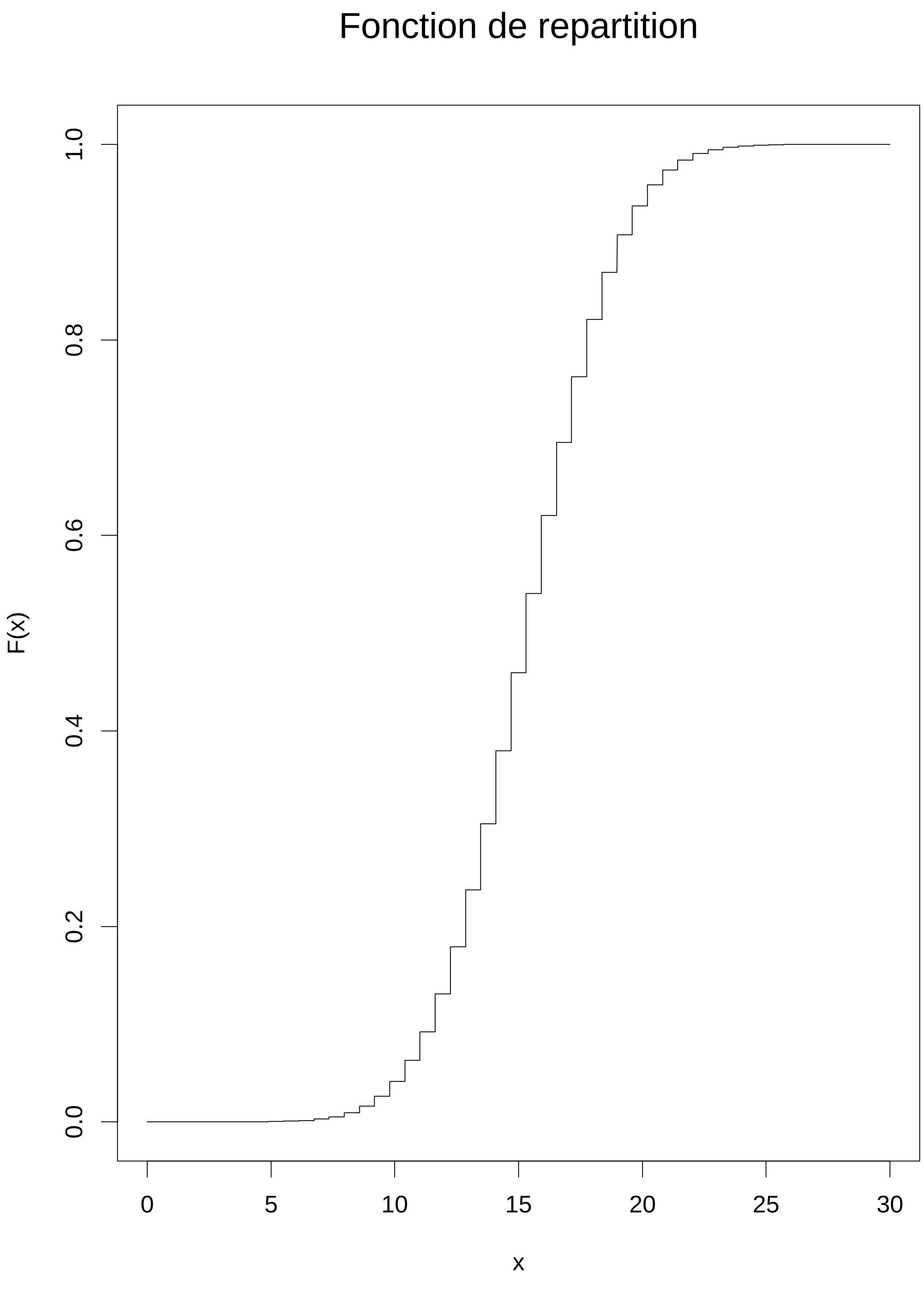
Exemple

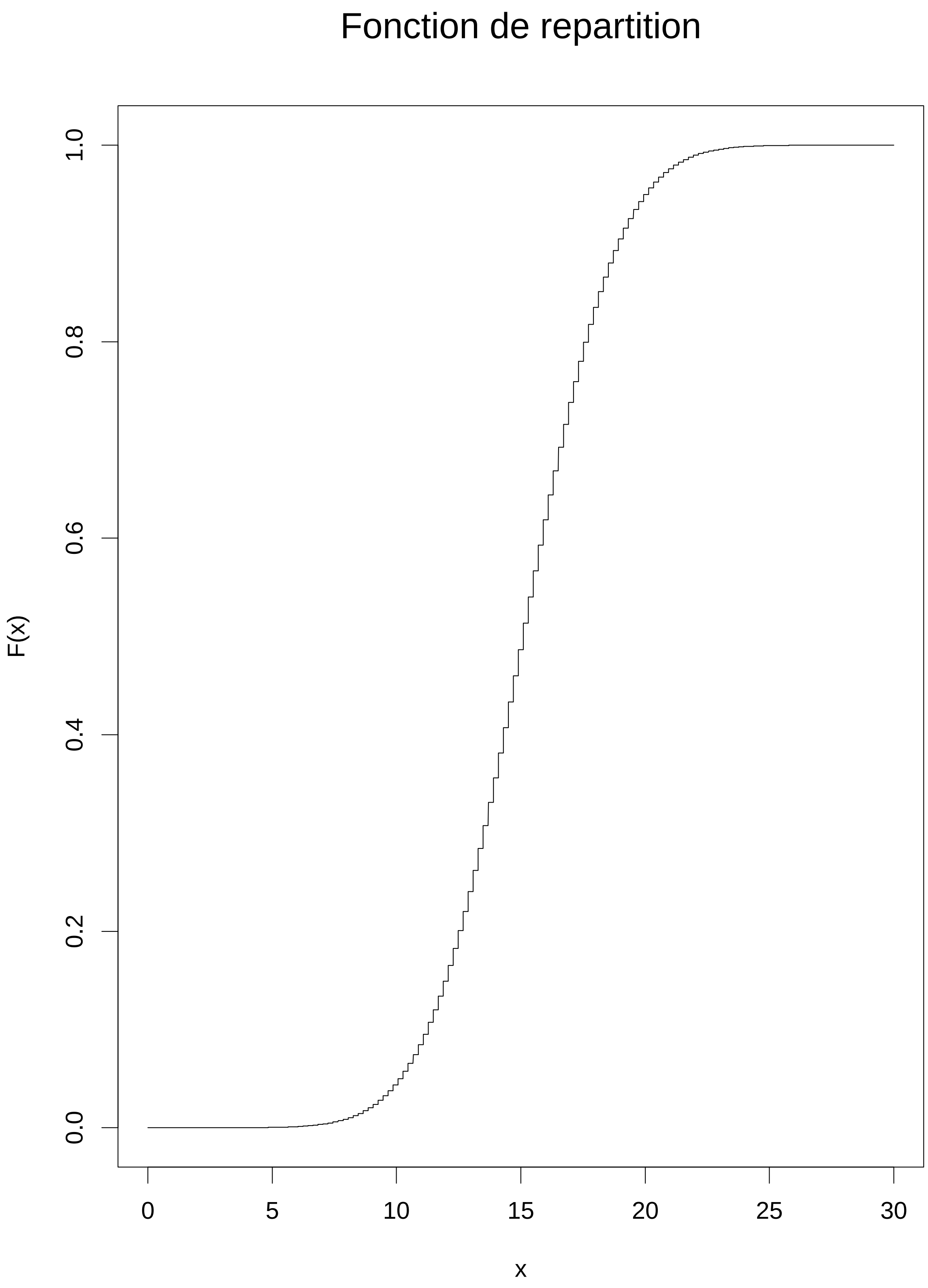
r	π_r	$F_\xi(r)$
0	0,4	0,4
1	0,3	0,4 +0,3 =0,7
2	0,3	0,7 + 0,3 =1



Question: Que se passe-t-il lorsque N augmente?

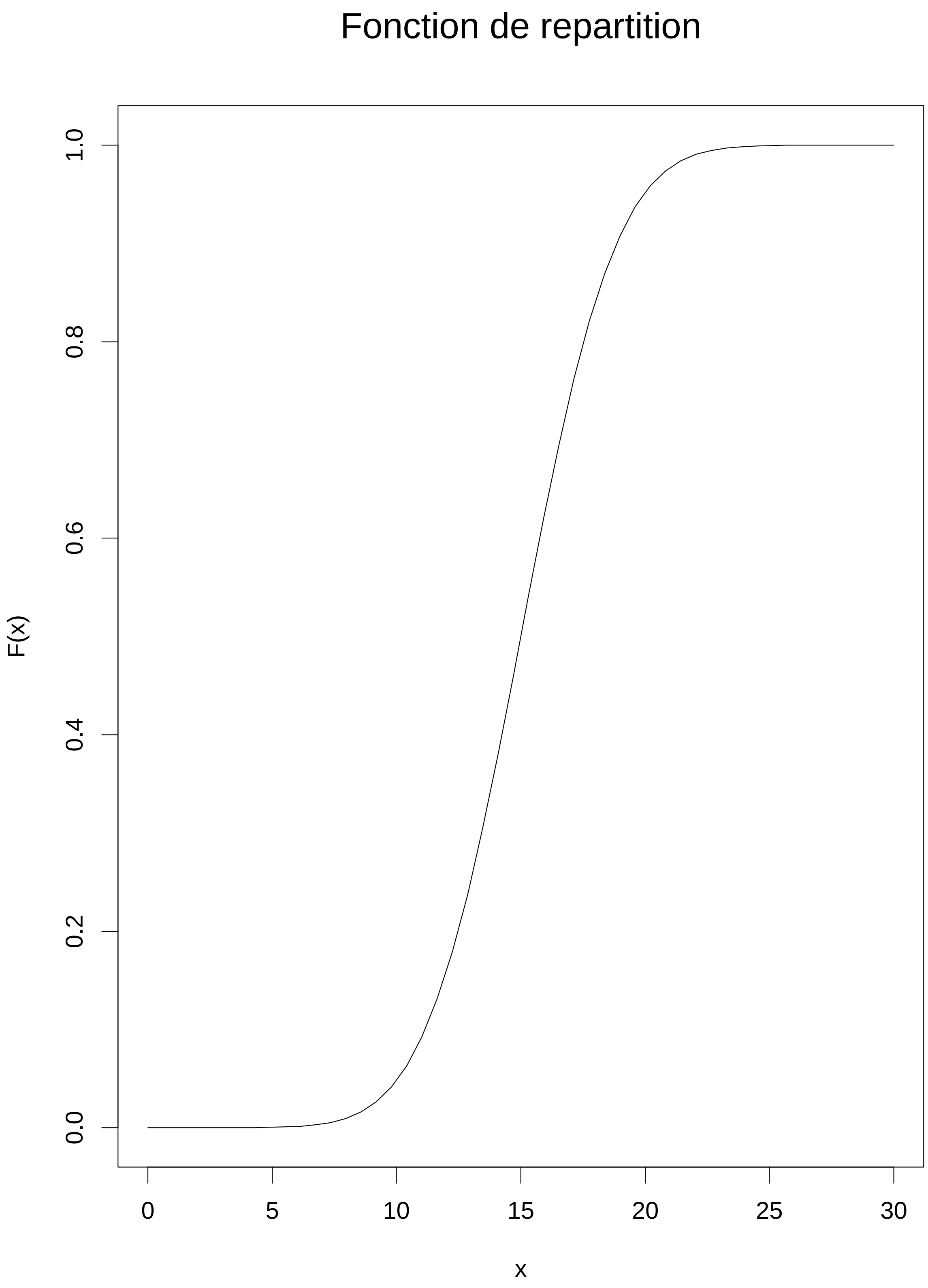
"Modèle théorique" c'est-à-dire une approximation continue de la "vraie" fonction de distribution.


Fonction de repartition
F(x)
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
x


Fonction de repartition
F(x)
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
x

Exercice : $F_\xi(x)$ continue $\Rightarrow$ [% d'individus tels que $\xi = x$]

$\equiv 0$



6.2.3 Caractéristiques (ou paramètres) de la distribution de population

Paramètre de localisation

$$\begin{aligned}
\mu_\xi &= m(\xi) = \overline{\xi} && \underset{\longleftarrow}{\text{notations}} \\
&= \frac{1}{N} \sum_{1 \leq k \leq N} \xi_k && \underset{\longleftarrow}{\text{définition}}
\end{aligned}$$

Paramètre de dispersion

$$\begin{aligned}
\sigma_\xi^2 &= v^2(\xi) \\
&= \frac{1}{N} \sum_{1 \leq k \leq N} (\xi_k - \overline{\xi})^2 \\
&= \frac{1}{N} \sum_k \xi_k^2 - \left(\frac{1}{N} \sum_k \xi_k \right)^2 \\
&= \overline{\xi^2} - (\overline{\xi})^2
\end{aligned}$$

6.3 Echantillon

6.3.1 Le problème

Examen exhaustif de la population:

$$\begin{cases} \text{impossible} \\ \text{inefficace} \end{cases}$$

6.3.2 Procédé

i) Choisir "aléatoirement" des individus

$$s = \{k_1, k_2, \cdots k_n\} \subset \mathcal{N} \qquad \text{“}\textit{échantillon ordonné}\text{”}$$

$$= \{k_i : 1 \leq i \leq n\}$$

k_i : "*étiquette*" *d'individus de la population* $\mathcal{N}$

ii) observer:

$$(\xi_{k_1}\ \xi_{k_2} \cdots \xi_{k_n})$$

noté:

$$x = (x_1\ x_2 \cdots x_n) \qquad \textit{observation}$$

données: "*observation complète*" $d = (s,x)$

par exemple

$$\begin{array}{lcc} \text{caractéristique de population:} & \xi & \rho \\ & \downarrow & \downarrow \\ \text{échantillon:} & x & r \end{array}$$

$$\text{donnée complète:} \quad d = \{(k_1,x_1,r_1),(k_2,x_2,r_2)\cdots(k_n,x_n,r_n)\}$$
$$= \{(k_i,x_i,r_i) : 1 \leq i \leq n\}$$

6.3.3 Considérons une caractéristique catégorielle (ou qualitative ou discrète)

Soit une valeur r de cette caractéristique

n_r	effectifs observés	(population : ν_r)
$f_r = \dfrac{n_r}{n}$	fréquence observée	(population : π_r)

6.3.4 Considérons une caractéristique numérique

Fonction de répartition "empirique"

$\hat{F}_\xi(x) = \%$ d'individus de l'échantillon tel que $\xi_i \leq x$.

6.4 Echantillonnage aléatoire simple

6.4.1 Choix "au hasard"

Convention d'écriture

Majuscule (par ex. K) pour la v.a. considérée comme fonction.

Minuscule (par ex. k) pour une valeur possible d'une réalisation.

Définition d'un échantillonnage aléatoire simple

$$P(K = k) = \frac{1}{N} \qquad \forall\, k \in \mathcal{N}$$

Par exemple

$$\begin{aligned} k &= 32 \\ &= 28 \end{aligned}$$

Conséquences

Soit

ρ_k caractéristique catégorielle (discrète)

R v.a. cette même caractéristique observée dans l'échantillon

alors:

$$\begin{aligned} P(R = r) &= \frac{\text{nombre d'individus de la population tels que } \rho_k = r}{N} \\ &= \pi_r \end{aligned}$$

Soit

ξ_k: caractéristique numérique

X v.a.: cette même caractéristique numérique observée dans l'échantillon.

alors, la *Probabilité* qu'un individu tiré au hasard soit tel que sa caractéristique $\xi \leq x$:

$$\begin{aligned} P(X \leq x) &= F_X(x) \quad \text{(par définition de } F_X) \\ &= \frac{\text{nombre d'individus de la population tels que } \xi \leq x}{N} \\ &\quad \text{(tirage au hasard)} \\ &= F_\xi(x) \quad \text{(par définition de } F_\xi) \end{aligned}$$

Remarque fondamentale:

$F_X(x)$ = fonction de distribution d'échantillonnage (Proba)

$F_\xi(x)$ = fonction de distribution de population (Descriptive)

6.4.2 Echantillonnage simple avec remise (ESAR)

Idée de base:

1. tirage séquentiel
2. chacun de ces tirages: mêmes conditions c.à.d.

$$\left[\begin{array}{l}\text{avec remise après chaque tirage}\\ \text{chaque tirage: aléatoire simple}\end{array}\right.$$

Elaboration

On choisit donc une suite d'individus: $K_1\ K_2 \cdots K_n$

Calculons-en la probabilité:

$$P\left[\{K_1 = k_1\} \cap \{K_2 = k_2\} \cdots \cap \{K_n = k_n\}\right]$$

$$= P(K_1 = k_1) \cdot P(K_2 = k_2) \cdot \cdots P(K_n = k_n)$$

(par indépendance)

$$= \frac{1}{N} \cdot \frac{1}{N} \cdots \frac{1}{N} \quad \text{(même distribution)}$$

$$= \frac{1}{N^n} \quad \text{(quels que soient } (k_1,k_2 \cdots k_n) \in \Omega\text{)}$$

On a donc dans ce cas:

$$\begin{aligned}\Omega\ &= \mathcal{N} \times \mathcal{N} \times \cdots \times \mathcal{N} = \mathcal{N}^{\mathrm{n}}\\ &= \{(\mathrm{k_1},\mathrm{k_2},\cdots,\mathrm{k_n}) | \mathrm{k_i} \in \mathcal{N}, 1 \leq \mathrm{i} \leq \mathrm{n}\}\end{aligned}$$

$$|\Omega| = N \times N \times \cdots N = N^n$$

6.4.3 Echantillonnage simple sans remise (E.S.S.R.)

Idée de base:

1. tirage séquentiel
2. chaque tirage
   - pas de remise de l'élément tiré antérieurement
   - aléatoire simple

Elaboration

On choisit donc une suite d'individus: $K_1\ K_2 \cdots K_n$. Calculons-en la probabilité:

$$P\left[\{K_1 = k_1\} \cap \{K_2 = k_2\} \cdots \cap \{K_n = k_n\}\right]$$

$$= P[K_1 = k_1] \cdot P[K_2 = k_2 | K_1 = k_1] \cdot$$

$$P[K_3 = k_3 | K_1 = k_1, K_2 = k_2] \cdots$$

$$P[K_n = k_n | K_1 = k_1, K_2 = k_2 \cdots K_{n-1} = k_{n-1}]$$

(probabilité composée)

$$= \frac{1}{N} \cdot \frac{1}{N-1} \cdot \frac{1}{N-2} \cdots \frac{1}{N-n+1}$$

SI $k_2 \neq k_1$ etc.
$k_3 \notin \{k_1,k_2\}$
$\ldots$
$k_n \notin \{k_1,k_2,\ldots,k_{n-1}\}$

SINON : 0

On a donc dans ce cas:

$$\begin{aligned}
\Omega &= \{(k_1,k_2\cdots k_n)|k_i \in \mathcal{N},\, i \neq j \Rightarrow k_i \neq k_j\} \\
|\Omega| &= N \cdot (N-1)(N-2)\cdots(N-n+1) \\
&= \frac{N!}{(N-n)!} = A_N^n \quad \text{(exercice)}
\end{aligned}$$

Dès lors :

$$P\left[\{K_1 = k_1\} \cap \{K_2 = k_2\} \cdots \cap \{K_n = k_n\}\right]$$
$$= \frac{1}{A_N^n} = \frac{(N-n)!}{N!}$$

quels que soient $(k_1,k_2,\cdots k_n) \in \Omega$.

Chapitre 7

Estimation ponctuelle

7.1 Introduction

7.1.1 Le problème général de "l'inférence statistique"

A revoir: section 1.4 de ce cours

"apprendre en observant"

↓

incertain

plusieurs façons d'**exprimer cette incertitude**

• estimation ponctuelle
•• test d'hypothèse
• • • intervalle de confiance

7.1.2 Procédé statistique:

Définition
Fonction définie sur l'espace d'observations et dont les valeurs possibles sont, selon les cas,

- estimation ponctuelle: espace du paramètre
- test d'hypothèse: {accepter, rejeter}
- intervalle de confiance: intervalle de l'espace paramétrique

De façon plus précise: $T : \mathcal{X} \to \mathcal{T}$ où $\mathcal{X}$ est un espace d'observation

$$\begin{aligned} \mathcal{T} &= \Theta && : \text{ estimation ponctuelle} \\ &= \{\text{Accepter, Rejet}\} && : \text{ test d'hypothèse} \\ &= \{\text{intervalle de } \Theta\} && : \text{ intervalle de confiance} \end{aligned}$$

Exercice : Commentez l'exemple suivant :

X_i poids de l'individu $i = 1, \ldots, n$
donc $X_i \in I\!R_+$
donc $\mathcal{X} = \underbrace{I\!R_+ \times I\!R_+ \ldots \times I\!R_+}_{n \text{ fois}} = I\!R_+^n$

Problème de base:
Comment définir:

- • "bonne" procédure?
- •• procédure "optimale"?

7.2 Echantillonnage aléatoire

7.2.1 Introduction: $n = 1$

Soit une population finie d'individus caractérisés par la variable ξ

$$\mathcal{N} = \{1, \cdots k \cdots N\}$$

$$\xi = (\xi_1 \cdots \xi_k \cdots \xi_N) \quad m(\xi) = \overline{\xi} = \frac{1}{N} \sum_{1 \le k \le N} \xi_k$$

$$v^2(\xi) = \sigma_\xi^2 = \frac{1}{N} \sum_{1 \le k \le N} (\xi_k - \overline{\xi})^2$$

Echantillon aléatoire:

De la population $\mathcal{N}$, on extrait "au hasard" un individu et on observera sa caractéristique ξ.

$$P(K = k) = \tfrac{1}{N} \qquad k = 1,2, \cdots N$$

$$X = \xi_K$$

Dans ce cas

$$F_X(x) = F_\xi(x)$$

Remarque importante:

$F_X(x)$: probabilité d'échantillonnage

$F_\xi(x)$: proportion de population

dès lors

$E(X) = \overline{\xi}$

$V(X) = \sigma_\xi^2$

Cas particulier: caractéristique de catégories.

7.2.2 Echantillonnage aléatoire

• Soit une **population finie** de N individus i:

$$\mathcal{N} = \{i_1, i_2, \cdots, i_N\}$$

•• **Echantillon**

On choisit aléatoirement n individus: $S = (i_1, i_2, \quad \cdots \quad, i_n)$

On observe: $\xi_{i_1} = x_1 \quad \xi_{i_2} = x_2 \quad \cdots \quad \xi_{i_n} = x_n$

On écrit : $x_j = \xi_{i_j}$

On obtient de la sorte : $x = (x_1, x_2, \cdots, x_n)$

• • • **Motivation**

Par exemple, on s'intéresse à:

$$m(\xi) = \overline{\xi} = \frac{1}{N}\sum_{k=1}^{N}\xi_k$$

Il s'agit de rechercher une fonction: $f(x_1, \cdots, x_n)$ pour estimer $\overline{\xi}$
Par exemple : la fonction "moyenne d'échantillon" :

$$(X_1, X_2, \cdots, X_n) \longrightarrow \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

• • •• **Plus formellement**, on définira les concepts suivants.

Définition: *échantillon aléatoire de taille n*

= suite ordonnée de n V.A.

$= (X_1, X_2, \cdots, X_n)$

$= (X_i \ : \ 1 \leq i \leq n)$

Définition: *statistique T*

= une fonction de l'échantillon

c'est-à-dire $T = f(X_1, X_2, \cdots, X_n)$

Définition: *estimateur*

= une statistique utilisée pour estimer (un paramètre inconnu)

Définition: *estimation*

= une valeur observée d'un estimateur

Remarque: attention de bien distinguer

$$\begin{cases} \text{estimateur} = \text{fonction} \\ \\ \text{estimation} = \text{valeur} \end{cases}$$

7.3 Le problème d'un "bon" estimateur

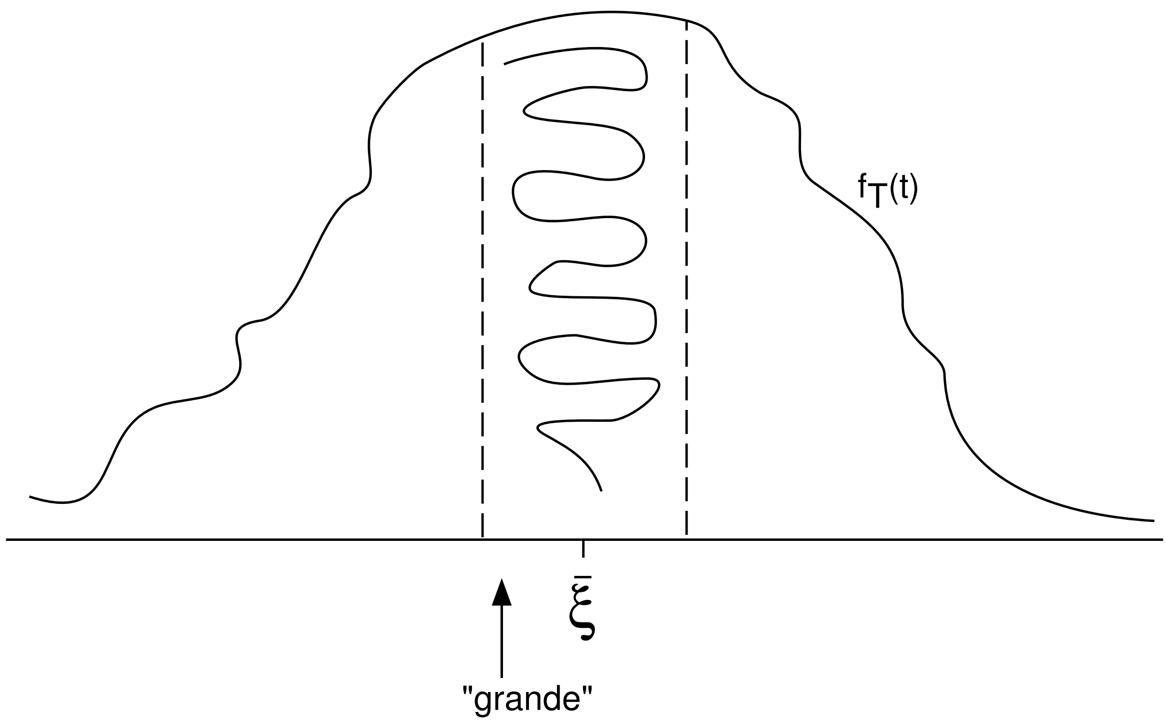
7.3.1 Distribution d'échantillonnage

Rappelons: $(X_1, \cdots, X_n)$ aléatoire $\Rightarrow T = f(X_1, \cdots, X_n)$ aussi une V.A.

La distribution de l'estimateur T s'appelle **distribution d'échantillonnage** et dépend:

- du plan d'échantillonnage: distribution de S
- •• de la distribution, dans la population, des caractéristiques $\xi_1, \xi_2, \cdots, \xi_N$

Question: la distribution de l'estimateur a-t-elle de "bonnes" propriétés?



Prob (T prenne une valeur proche de $\overline{\xi}$): "grande"?

Alternativement:

Soit
$T - \theta$ “l’erreur d’estimation”:

alors :
$T - \theta = 0$: pas d’erreur d’estimation de θ
$T - \theta > 0$: surestimation de θ
$T - \theta < 0$: sous-estimation de θ

Question: la distribution de l’erreur d’estimation est-elle concentrée autour de 0?

7.3.2 Estimation sans biais

Définitions:

biais de T (vis-à-vis de θ): $E(T) - \theta = b_T(\theta)$

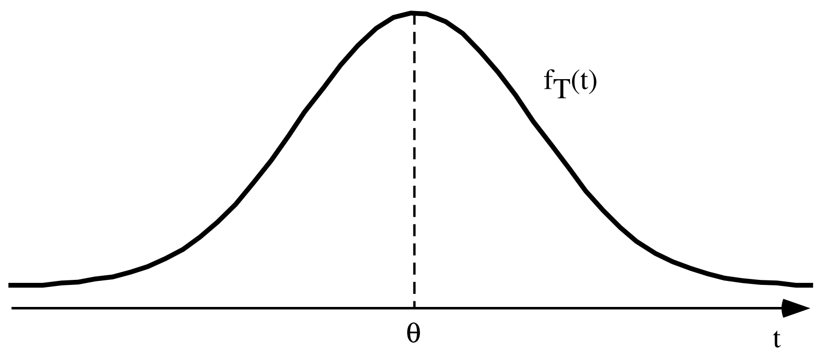
T estimateur *sans biais* pour estimer θ:
$E(T) = \theta \quad \forall\theta, \quad$ ou encore: $b_T(\theta) = 0 \qquad \forall\theta$

Alternativement:

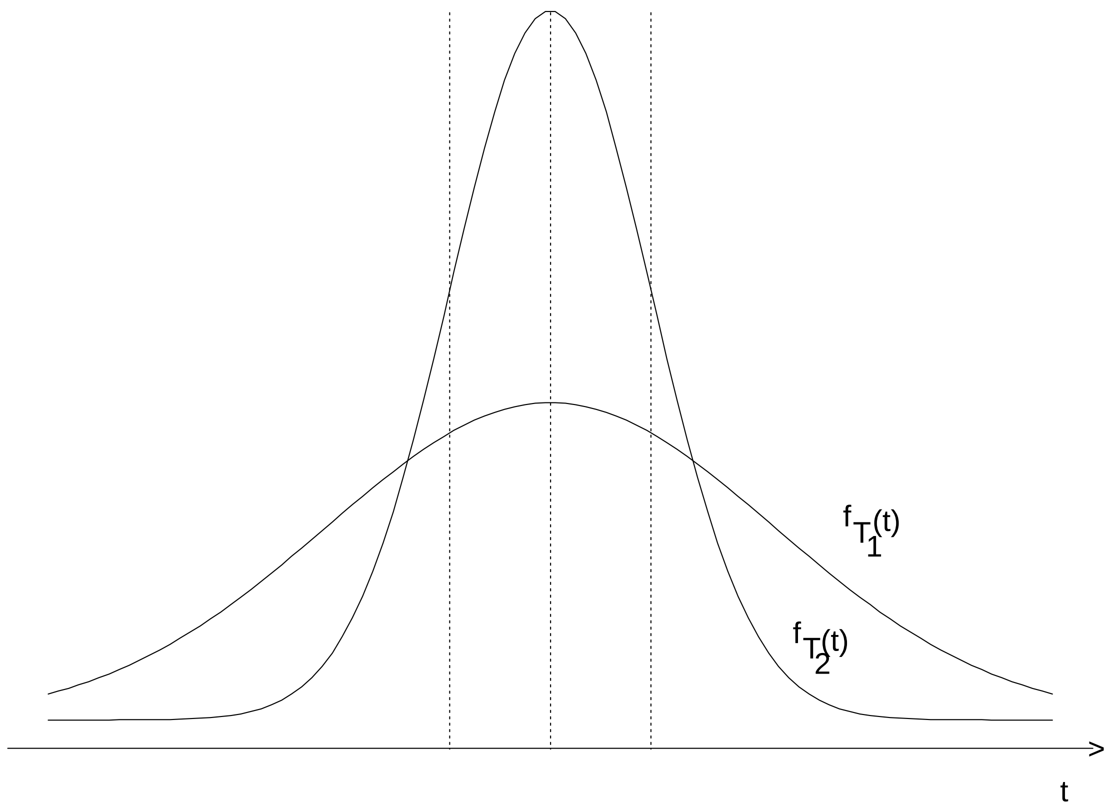
$$E(T - \theta) = 0 \qquad \forall\theta$$

la distribution de l'erreur d'estimation est centrée.

Graphiquement (dans le cas symétrique !)



Exemples



On remarque :

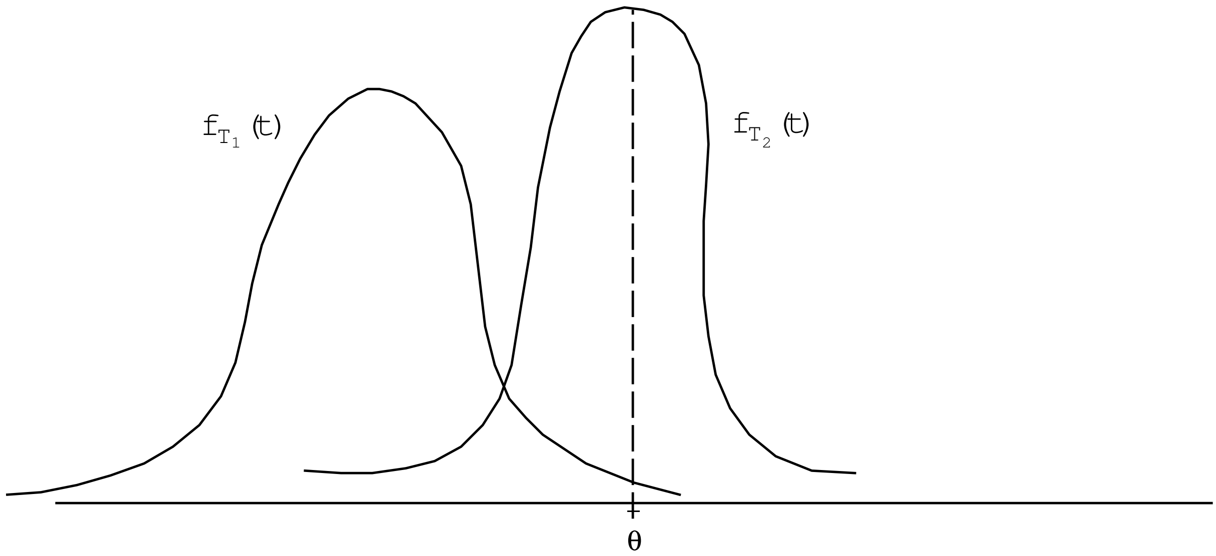
$E[T_1] = E[T_2] = \theta$

$P[|T_1 - \theta| < \varepsilon] < P[|T_2 - \theta| < \varepsilon]$ pour ε "petit"

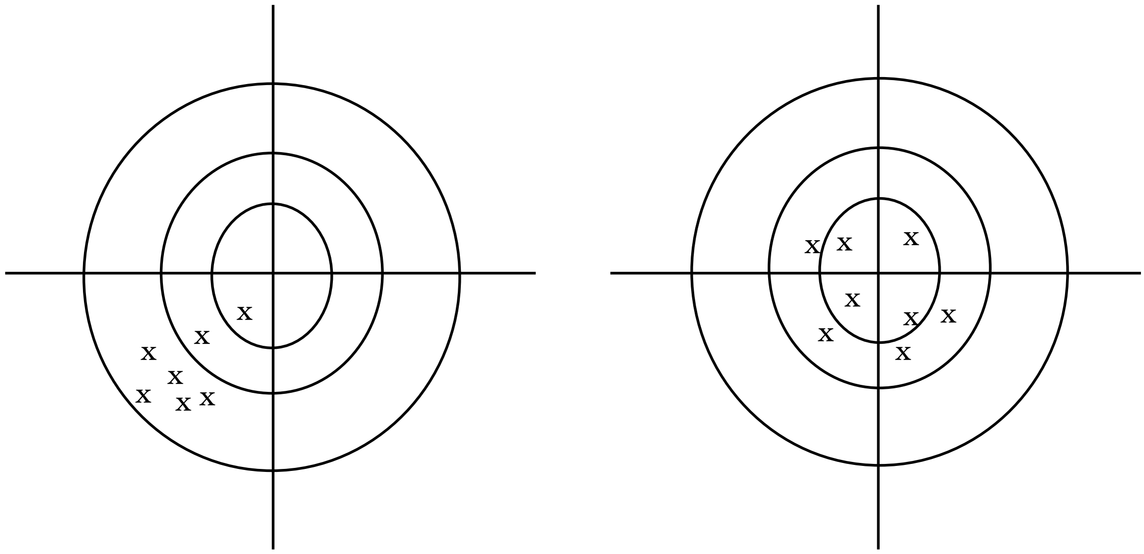
7.3.3 Erreur quadratique

Exemple 1

Considérons deux distributions d'échantillonnage relatives à des estimateurs T_1 et T_2:

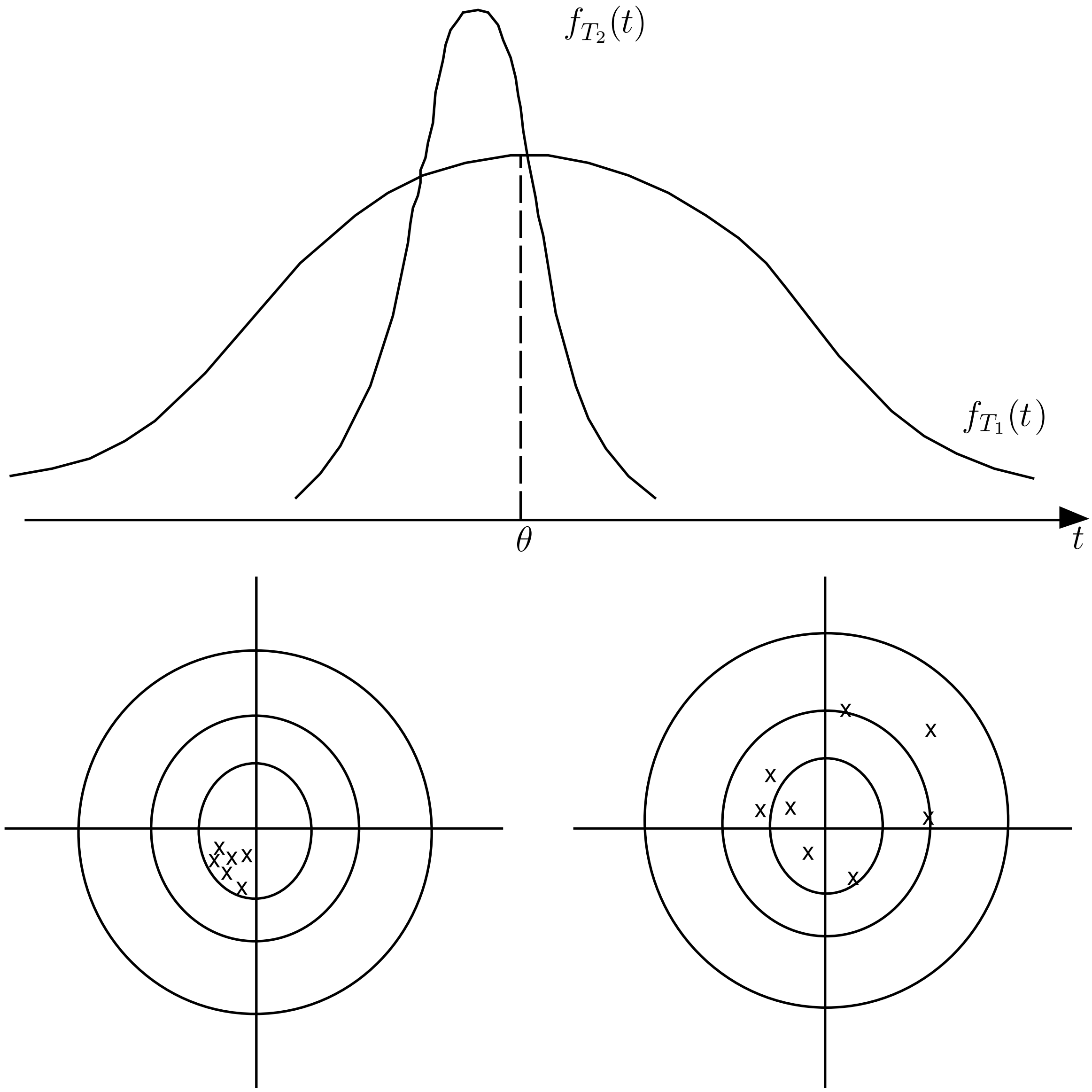


Analogie du tir sur une cible:



Exemple 2

Considérons les distributions d'échantillonnage de deux estimateurs T_1 et T_2.



Définitions

- Erreur quadratique de l'estimateur T

$$(T - \theta)^2$$

•• Erreur quadratique moyenne ou **risque** de l'estimateur T

$$E[(T - \theta)^2] = r_T(\theta)$$

Propriété essentielle:

En décomposant l'erreur d'estimation par rapport à l'espérance mathématique de T et au biais de T:

$$T - \theta = (T - E(T)) + (E(T) - \theta)$$

on peut démontrer (*Exercice :* Vérifier)

$$r_T(\theta) = V(T) + [b_T(\theta)]^2$$

c'est-à-dire:

Erreur quadratique moyenne = Variance + (biais)^2

Critère d'évaluation d'un estimateur:

Un estimateur est "bon" si $r_T(\theta)$ "petit pour tout θ"

7.3.4 Application : estimation d'une moyenne

Exemple $\overline{X}$

- $\overline{X} = \frac{1}{n} \sum X_i$
  $\mu_\xi = \dfrac{1}{N} \sum\limits_i \xi_i$

•• Ech. aléatoire simple
c'est-à-dire X_i mutuellement indépendant
même distribution
cette distribution
= distribution de population
donc "X_i est I.I.D."
p. ex. EASAR

dès lors $E(X_i) = \mu_\xi \quad \forall\, i$

$V(X_i) = \sigma_\xi^2 \quad \forall\, i$

Dès lors

$$\begin{aligned} E[\overline{X}] &= E[\tfrac{1}{n} \textstyle\sum X_i] = \tfrac{1}{n} \textstyle\sum E(X_i) \\ &= \tfrac{1}{n} \underbrace{\sum_{i=1}^{n} \mu_\xi}_{n \cdot \mu_\xi} \qquad\qquad E(X_i) = \overline{\xi} = \mu_\xi \\ &= \mu_\xi \end{aligned}$$

Exercice : Vérifiez.

$\therefore \overline{X}$ estimateur sans biais

$$
\begin{aligned}
V(\overline{X}) &= V\left[\tfrac{1}{n}\sum X_i\right] \\
&= \tfrac{1}{n^2} V\left(\sum X_i\right) \\
&= \tfrac{1}{n^2} \sum_i V(X_i) \qquad X_i \text{ indépendant} \\
&= \tfrac{1}{n^2} \underbrace{\left(\sum_i \sigma_\xi^2\right)}_{n\sigma_\xi^2} \qquad \sigma_\xi^2 = V(\xi) = \tfrac{1}{N}\textstyle\sum_i (\xi_i - \mu_\xi)^2 \\
&= \frac{\sigma_\xi^2}{n} \\
\text{EQM} &= r_{\overline{X}}(\mu,\sigma^2) = \frac{\sigma_\xi^2}{n}
\end{aligned}
$$

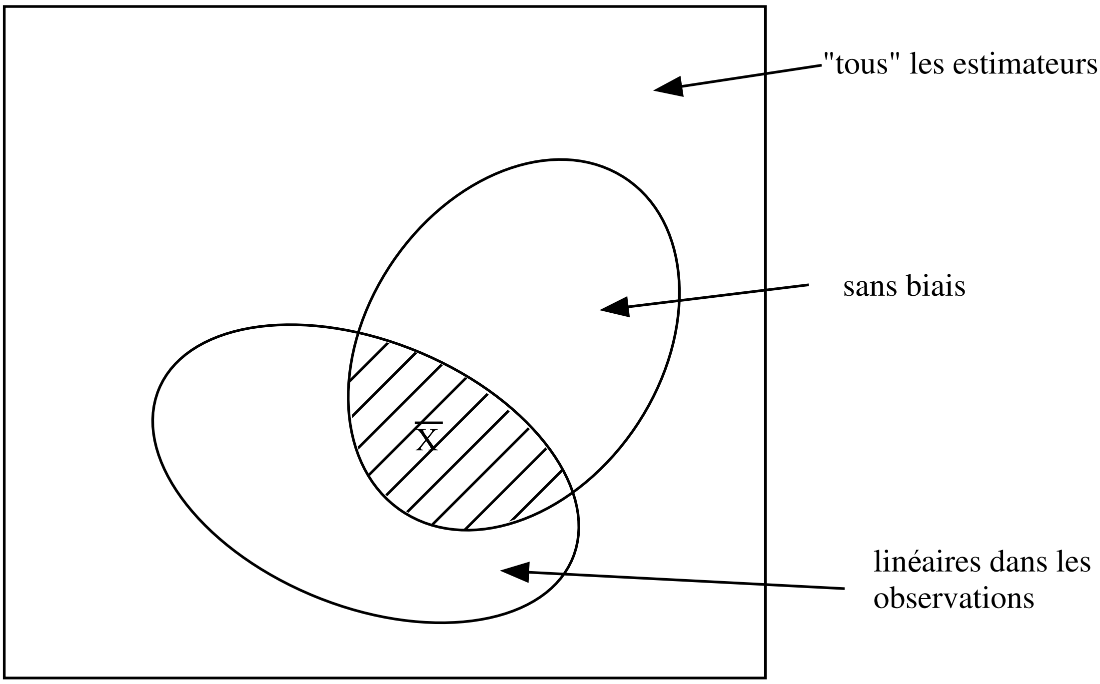
"optimal" au sens suivant

Théorème

Parmi tous les estimateurs linéaires et sans biais, $\overline{X}$ est celui de variance minimum.

Remarque

aspect $\left.\begin{array}{l}\text{linéaire}\\ \text{sans biais}\end{array}\right\}$ simplifient l'évaluation des propriétés statistiques


"tous" les estimateurs
sans biais
$\overline{X}$
linéaires dans les observations

7.4 Propriétés asymptotiques

7.4.1 Nature du problème:

Jusqu'ici: n (taille d'échantillon) fixé

Considérons la limite $n \to \infty$.

Considérons un échantillonnage dont la taille croît indéfiniment:

$$
\begin{array}{ll}
X_1 & T_1 = f_1(X_1) \\
X_1,X_2 & T_2 = f_2(X_1, X_2) \\
X_1,X_2,X_3 & T_3 = f_3(X_1, X_2, X_3) \\
\vdots & \vdots \\
X_1, X_2, \cdots, X_n & T_n = f_n(X_1, X_2, \cdots, X_n) \\
\vdots & \vdots
\end{array}
$$

Par exemple: $T_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}_n$

7.4.2 Questions en jeu

Il convient de distinguer deux types de questions très différentes:

(i) $n \to \infty \Rightarrow$ "je connaîtrais parfaitement θ" ?
c'est-à-dire: question relative à la logique de l'inférence.

(ii) souvent, la distribution de T_n est "difficile"
Si n est "grand", est-il possible d'approximer cette distribution ?

7.4.3 Question relative a la logique d'inférence

Problème de "convergence" d'un estimateur

Convergence en probabilité "$T_n \xrightarrow{p} \theta$"

$$\lim_{n\to\infty} P[|T_n - \theta| > \varepsilon] = 0 \qquad \forall \varepsilon > 0$$
$$\Leftrightarrow \lim_{n\to\infty} P[|T_n - \theta| \leq \varepsilon] = 1 \qquad \forall \varepsilon > 0$$

Remarque : $|T_n - \theta| \leq \varepsilon \Leftrightarrow -\varepsilon \leq T_n - \theta \leq +\varepsilon$
$\Leftrightarrow \theta - \varepsilon \leq T_n \leq \theta + \varepsilon$

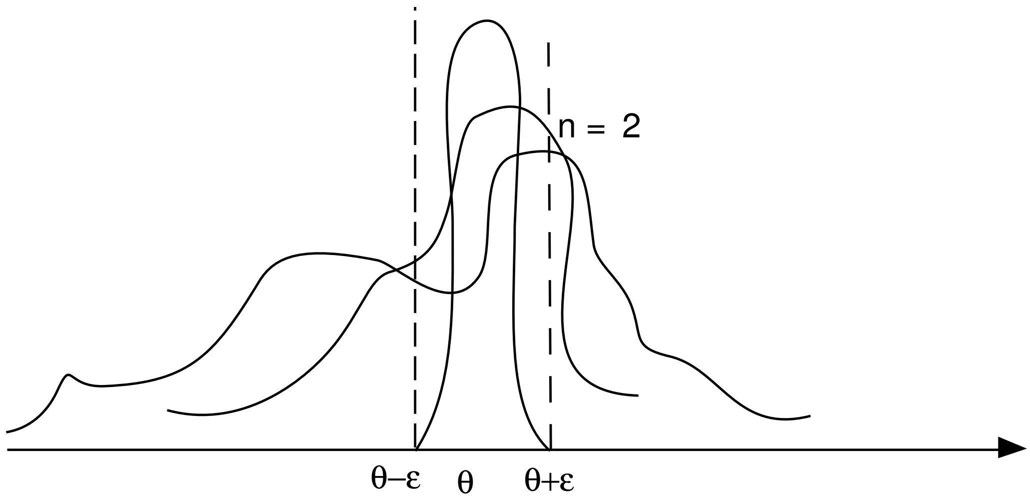
Théorème :

Si $E(T_n) \to \theta$
$V(T_n) \to 0$
alors $T_n \xrightarrow{p} \theta$

Remarque

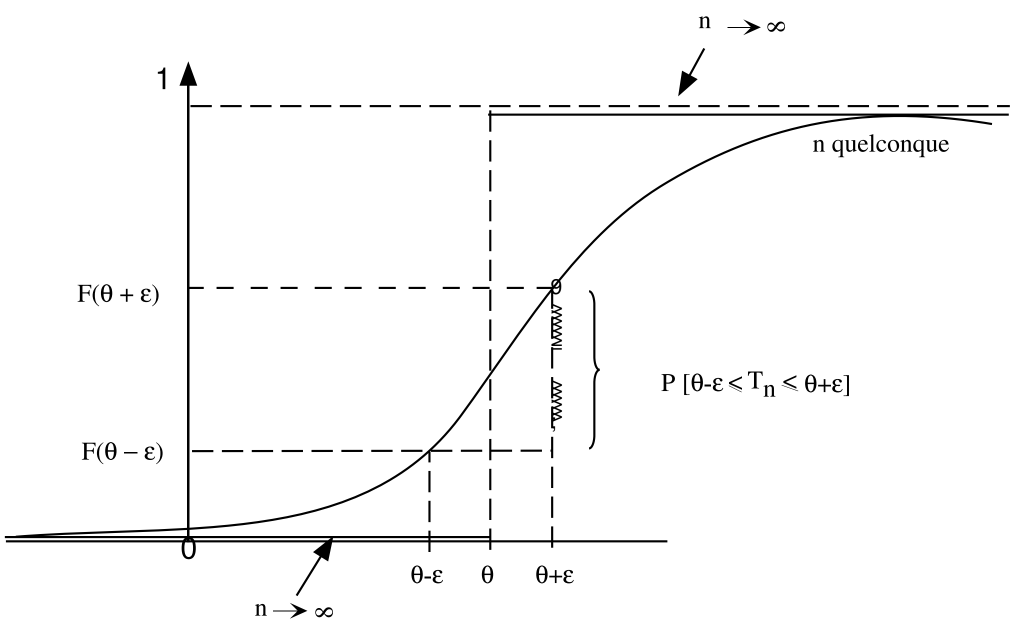
$[E(T_n) \to \theta$ et $V(T_n) \to 0] \Longleftrightarrow [r_\theta(T_n) \to 0] \Longrightarrow [T_n \xrightarrow{p} \theta]$

Au point de vue de la suite des fonctions de densité:



Au point de vue de la suite des fonctions de distribution

$$F_T(t) = P[T \leq t]$$



Rappel: $P[\theta\varepsilon \leq T_n \leq \theta + \varepsilon] = F_{T_n}(\theta + \varepsilon)F_{T_n}(\theta\varepsilon)$

Convergence presque sûre

Considérons la suite des estimateurs:

$T_1 \quad T_2 \quad T_3 \cdots$

Par exemple: pile ou face

$$X_i \qquad 1 \quad 0 \quad 0 \quad 1 \quad \cdots$$

$$T_i \qquad 1 \quad \frac{1}{2} \quad \frac{1}{3} \quad \frac{2}{4} \quad \cdots$$

Définition: convergence presque sûre de θ:

$$P\{T_n \to \theta\} = 1 \qquad \text{“}T_n \stackrel{p.s.}{\to} \theta\text{”}$$

Par exemple

$T_n \quad \cdots \quad .49$

$.495$

$.501$

$.5001$

Propriété importante:

convergence presque sûre $\Rightarrow$ convergence en probabilité.

MAIS, l'inverse est faux.

7.4.4 Approximation de la distribution de T_n

Soit T_n suite d'estimateurs

T_∞ une V.A.

Leurs fonctions de distribution $\begin{cases} F_n(x) = P[T_n \leq x] \\ \\ F_\infty(x) = P[T_\infty \leq x] \end{cases}$

Convergence en distribution "$T_n \xrightarrow{\mathcal{L}} T_\infty$" ou "$T_n \xrightarrow{d} T_\infty$"

$F_n(x) \to F_\infty(x) \qquad \forall\, x$ où F_∞ est continue.

ou encore

$P[T_n \leq x] \to P[T_\infty \leq x]$

Propriété importante :

Si $\quad X_n \xrightarrow{d} Y$
$\quad g(\cdot)$ est une fonction continue

Alors $\quad g(X_n) \xrightarrow{d} g(Y)$

7.5 Lois des grands nombres

Théorème

Si X_i indépendant, identiquement distribué
$E(X_i) = \mu < \infty$
$\overline{X}_n = \frac{1}{n} \sum\limits_i X_i$

Alors
(i) loi faible des grands nombres : $\overline{X}_n \xrightarrow{p} \mu$
(ii) loi forte des grands nombres : $\overline{X}_n \xrightarrow{p.s} \mu$

Application à s^2

$$
\begin{aligned}
s^2 &= \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 \\
&= \frac{n}{n-1} \left[\frac{1}{n} \sum X_i^2 - \left(\frac{1}{n} \sum X_i^2 \right) \right] \\
& \quad \downarrow \qquad\qquad \downarrow p.,ps \qquad\qquad \downarrow p,p.s \\
& \quad 1 \qquad\qquad \underbrace{[E(X^2) \quad - \quad \mu^2]}_{\equiv \sigma^2}
\end{aligned}
$$

Conclusion: $s^2 \xrightarrow{p,ps} \sigma^2$

7.6 Théorème de la limite centrée

7.6.1 Introduction

Si la moyenne d'échantillon, $\overline{X}_n$, est calculée à partir d'un échantillon aléatoire simple (c'est-à-direX_i: i.i.d.) et si la population des caractéristiques ξ est distribuée selon une loi normale, on peut démontrer le résultat suivant:

Proposition 1 *Si F_ξ est normale c'est-à-dire $\xi \sim N(\mu,\sigma^2)$*

$$\textit{alors } \overline{X}_n \sim N\left(\mu,\frac{\sigma^2}{n}\right) \qquad \forall n$$

7.6.2 Le théorème de la limite centrée

Remarque préliminaire: même si F_ξ n'est pas normale, on pourra cependant vérifier:

$$E(\overline{X}_n - \mu) = 0 \qquad \forall n$$

$$\sqrt{n}(\overline{X}_n - \mu) \qquad \begin{cases} E[\sqrt{n}(\overline{X}_n - \mu)] = 0 \\ V[\sqrt{n}(\overline{X}_n - \mu)] = \sigma^2 \qquad \forall n \end{cases}$$

en effet

$$\begin{aligned} V\left(\sqrt{n}(\overline{X}_n - \mu)\right) &= nV(\overline{X}_n - \mu) = nV(\overline{X}_n) \\ &= n \cdot \frac{\sigma^2}{n} = \sigma^2 \end{aligned}$$

Exercice : Justifiez ces égalités.

Heuristiquement:

Même si F_ξ n'est pas normale $\overline{X}_n$ est approximativement normale lorsque $n \to \infty$

Proposition 2 *(Théorème de la limite centrée)*

Soient X_i indépendantes identiquement distribuées (EASAR) telles que $E(X_i) = \mu$ et $V(X_i) = \sigma^2 < \infty \qquad 1 \leq i \leq n$

Alors

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} N(0,\sigma^2)$$

Alternativement:

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1)$$

Remarques:

1. L'intérêt pratique de ce théorème est qu'il est valable même si ξ n'est pas de distribution normale
2. on dira que σ^2 est la "variance asymptotique" de $\overline{X}_n$
3. $\sqrt{n}$ est appelée "la vitesse de convergence"

Utilisation pratique du T.L.C.

$$\frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma} \overset{\text{app.}}{\sim} N(0{,}1) \text{ lorsque } n \text{ est GRAND}$$

$$\overline{X}_n \overset{\text{app.}}{\sim} N(\mu, \frac{\sigma^2}{n})$$

Attention: le *piège* à éviter est d'écrire

$$\overline{X}_n \xrightarrow{d} N(\mu, \frac{\sigma^2}{n})$$

Exercice: deviner où est l'erreur !

7.7 Echantillonnage avec remise en population finie ou en population infinie

Objectif

Petite synthèse de ce chapitre.

Départ

par exemple $\overline{\xi} = \dfrac{1}{N} \sum\limits_i \xi_i = \mu =$ "$E(\xi)$"

de plus: σ^2 ou $\sigma^2_\xi = \dfrac{1}{N} \sum\limits_i (\xi_i - \overline{\xi})^2 =$ "$V(\xi)$"

observer: $x_1\ x_2\ \cdots\ x_n$

problème: estimer μ

Suggestion: estimateur $\overline{X}_n = \dfrac{1}{n} \sum X_i$

On peut vérifier:

$$E(\overline{X}_n) = \mu \qquad \forall n$$

$$V(\overline{X}_n) = \frac{\sigma^2}{n}$$

Remarques

1. $E(\ )$,$V(\ )$: fonction des paramètres de population (μ,σ^2)
2. $V(\overline{X}_n) = \dfrac{\sigma^2}{n} \rightarrow 0$ lorsque $n \rightarrow \infty$
3. $E(\overline{X}_n) = \mu$: estimateur sans biais pour μ.

donc :

- $r_{\overline{X}_n}(\mu,\sigma^2) = \dfrac{\sigma^2}{n} + 0^2 = \dfrac{\sigma^2}{n}$
- •• $\overline{X}_n \xrightarrow{p} \mu$ (L.f.G.N.)

4. De plus: $\overline{X}_n \stackrel{p.s.}{\rightarrow} \mu$ (L.F.G.N.)

5. ces résultats: obtenus sans connaître la distribution de ξ

6. MAIS la distribution de $\overline{X}_n$ dépend de la distribution de ξ et le TLC nous assurer que si n est suffisamment grand, la distribution de $\overline{X}_n$ sera *approximativement* normale.

Chapitre 8

Tests d'hypothèses : Principes généraux

8.1 Notions de base

Exemple: test relatif au paramètre d'une distribution binomiale.

Jeu de pile ou face: une pièce de monnaie

$$P(pile) = \pi$$

$$\begin{aligned} \hat{\pi}_n &= \text{proportion d'échantillon} \\ &= \frac{1}{n}\sum x_i \qquad x_i = \text{ indicatrice de pile} \end{aligned}$$

On sait (estimation ponctuelle)

$$E(\hat{\pi}_n) = \pi \qquad \text{donc: sans biais}$$

$$V(\hat{\pi}_n) = \frac{\pi(1-\pi)}{n} \qquad \to 0 \text{ lorsque } n \to \infty \qquad \Rightarrow \qquad \hat{\pi}_n \xrightarrow{p} \pi$$

de plus $\sqrt{n}(\hat{\pi}_n - \pi) \xrightarrow{d} N(0,\pi(1-\pi))$

Modèle statistique (de l'exemple)ou "Hypothèse maintenue" H_m

$X_i \sim \text{ind.} Be(\pi) \quad i = 1, \cdots ,n$

$\pi \in [0,1] = \Theta$

$X_i \in \{0,1\}$

On a donc :

$P[X_i = x] = \pi^x(1-\pi)^{1-x} \qquad x = 0,1$

Hypothèse "nulle": H_0

essentiellement: $H_0 \subset \Theta$

par exemple H_0: $\pi = \frac{1}{2}$

Hypothèse alternative: H_1

essentiellement: $H_1 \subset \Theta$ telle que $H_0 \cap H_1 = \emptyset$

par exemple, en considérant $H_o : \pi = \frac{1}{2}$,

$$H_1 : \ \pi \neq \frac{1}{2} \qquad\qquad H_m : \ 0 \leq \pi \leq 1$$

$$H_1 : \ \pi < \frac{1}{2} \qquad\qquad H_m : \ 0 \leq \pi \leq \tfrac{1}{2}$$

$$H_1 : \ \pi > \frac{1}{2} \qquad\qquad H_m : \ \tfrac{1}{2} \leq \pi \leq 1$$

$$H_1 : \ \pi = 0{,}75 \qquad\qquad H_m : \ \pi \in \{0{,}50 \quad 0{,}75\}$$

dans le 1° cas: $H_0 \cup H_1 = \Theta = [0{,}1]$
c'est-à-dire H_0, H_1: partition de Θ

Hypothèse simple

Par exemple $H_0 : \pi = \pi_0 \quad \left(\pi = \dfrac{1}{2}\right)$

ou encore $H_0 = \{\pi_0\}$

Hypothèse composée

Intervalle de plusieurs valeurs de π

Par exemple $H_1 : \quad \pi > \dfrac{1}{2}$

Test d'hypothèse

Idée de base:

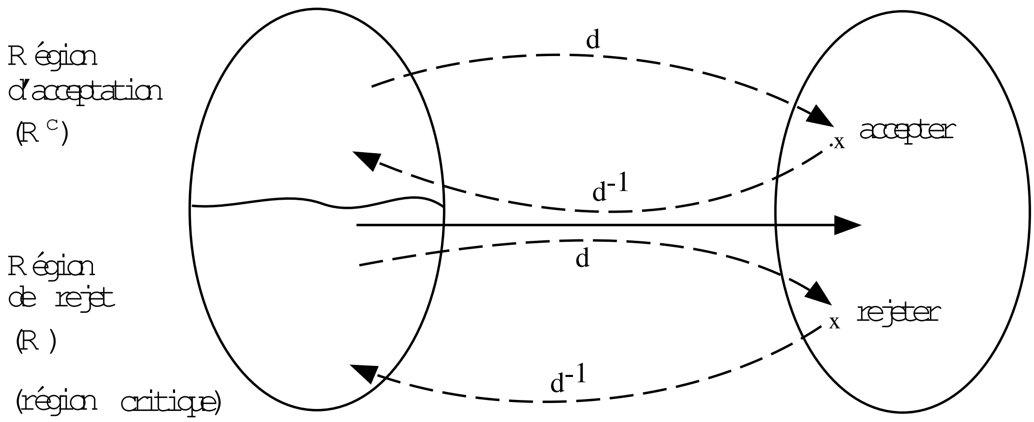
Observation $(x_1 \cdots x_n) \quad \rightarrow \quad \{$Accepter H_0, Rejeter $H_0\}$

Donc:

Test d'hypothèse c'est *une fonction* définie sur l'espace d'observation et à valeur dans un espace à deux points: {accepter, rejeter}

Graphiquement:

Espace d'observation



C'est aussi une partition de l'espace d'observations:

$\mathcal{X} = R \cup R^c \qquad$ et $\qquad R \cap R^c = \phi$

Remarque: (pour les étudiants plus exigeants).

$$R = d^{-1} \text{ (rejeter)} = \{x \in X | d(x) = \text{rejeter}\}$$

$$R^c = d^{-1} \text{ (accepter)} = \{x \in X | d(x) = \text{accepter }\}$$

Par exemple

1. $H_0 \qquad \pi = \frac{1}{2} \qquad H_1 : \pi \neq \frac{1}{2} \qquad n = 10$

   Un choix raisonnable serait :

   $R = \{0,1,2,3,7,8,9,10\}$

   $R^c = \{4,5,6\}$
2. $H_0 \qquad \pi = \frac{1}{2} \qquad H_1 : \pi > \frac{1}{2}$

   Un choix raisonnable serait :

   $R = \{7,8,9,10\}$

   $R^c = \{0,1,2,3,4,5,6\}$
3. $H_0 \qquad \pi \leq \frac{1}{2} \qquad H_1 : \pi > \frac{1}{2}$

   Un choix raisonnable serait :

   $R = \{7,8,\cdots,10\}$

   $R^c = \{0,1,\cdots,6\}$

8.2 Propriétés d'un test d'hypothèse

8.2.1 Types d'erreurs

"$\mathcal{E}_I$" **Erreur de type I**: rejeter H_0 lorsque H_0 est vraie

"$\mathcal{E}_{II}$" **Erreur de type II**: ne pas rejeter H_0 lorsque H_0 est fausse

8.2.2 Fonction de puissance

$$\begin{aligned}
P(\mathcal{E}_I) &= P[R|H_0] = P[R|\pi] \quad \pi \in H_0 \\
P(\mathcal{E}_{II}) &= P[R^c|H_1] = P[R^c|\pi] \quad \pi \in H_1 \\
&= 1 - P[R|\pi] \quad \pi \in H_1 \\
P(R|\pi) &= \Pi_R(\pi) \qquad \text{“fonction de puissance”}
\end{aligned}$$

On écrira aussi

$$\begin{aligned}
a(\pi) &= P(\mathcal{E}_I) = \Pi_R(\pi) \qquad \pi \in H_0 \\
b(\pi) &= P(\mathcal{E}_{II}) = 1 - \Pi_R(\pi) \quad \pi \in H_1
\end{aligned}$$

8.2.3 Exemple

$H_0 \qquad \pi \leq \frac{1}{2}$

$H_1 \qquad \pi > \frac{1}{2}$

$R = \{7,8,9,10\}$

Remarques

1. observation de base:

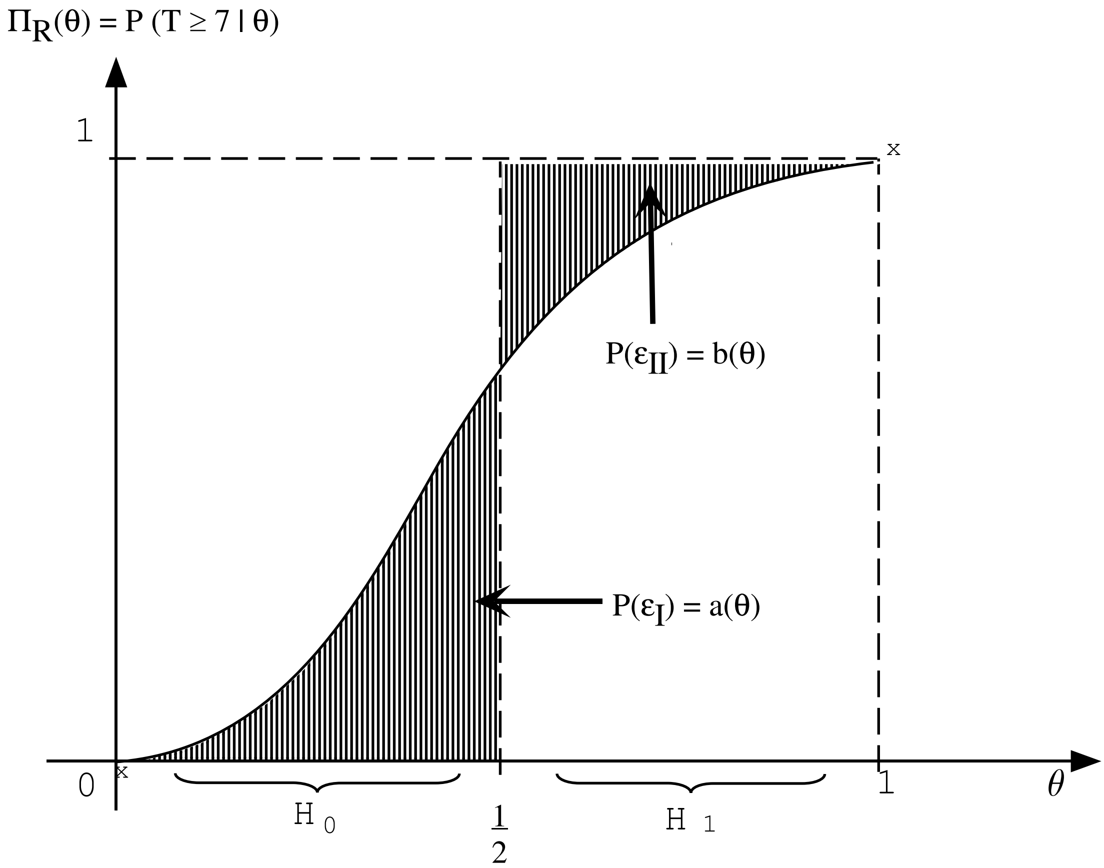
   $(x_1 \; x_2 \; \cdots \; x_n)$ par exemple: 1 0 0 1 ... 0

   $T = \sum X_i$ nombre de piles

   "*Statistique de test*"

2. Soit, par exemple, $R = \{t : t \geq 7\}$

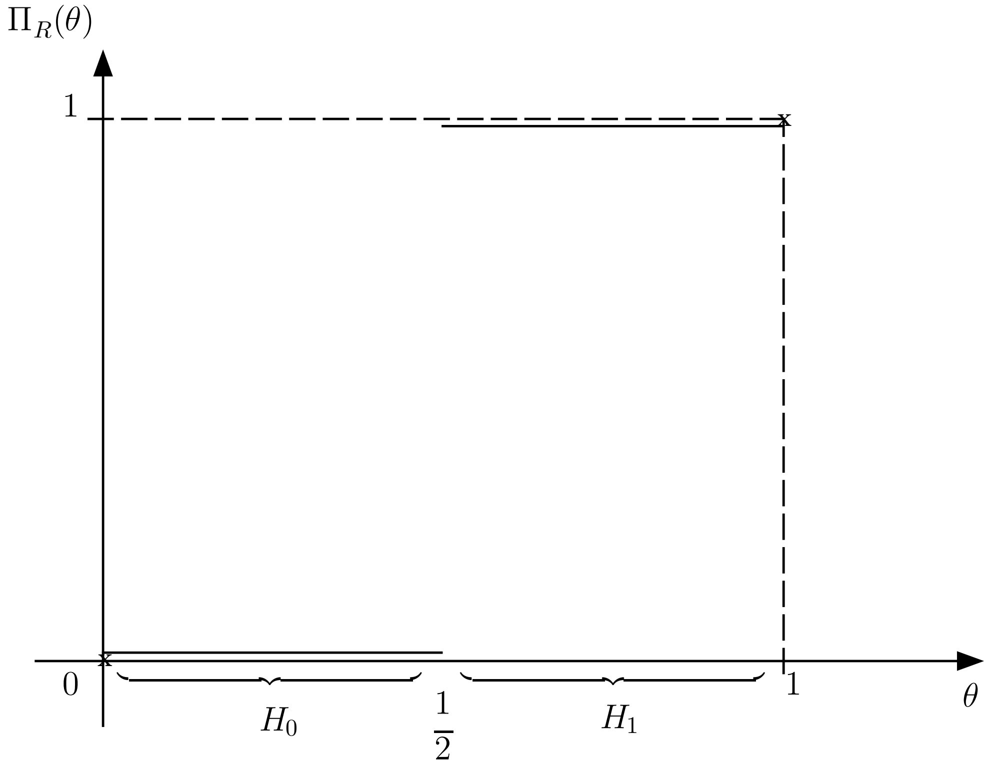
$$\begin{aligned} \Pi_R(\pi) &= P[T \geq 7|\pi] = \sum_{7 \leq t \leq 10} \binom{n}{t} \pi^t (1-\pi)^{n-t} \\ &= 1 - P[T < 7|\pi] \\ &= 1 - P[T \leq 6|\pi] \end{aligned}$$



Exercice: calculer les valeurs de cette fonction de puissance.

Indication: utiliser la loi binomiale.

8.2.4 Fonction de puissance d'un test "idéal"



8.2.5 Choix d'une région critique

Par exemple:

$$R_1 = \{7,8,9,10\}$$

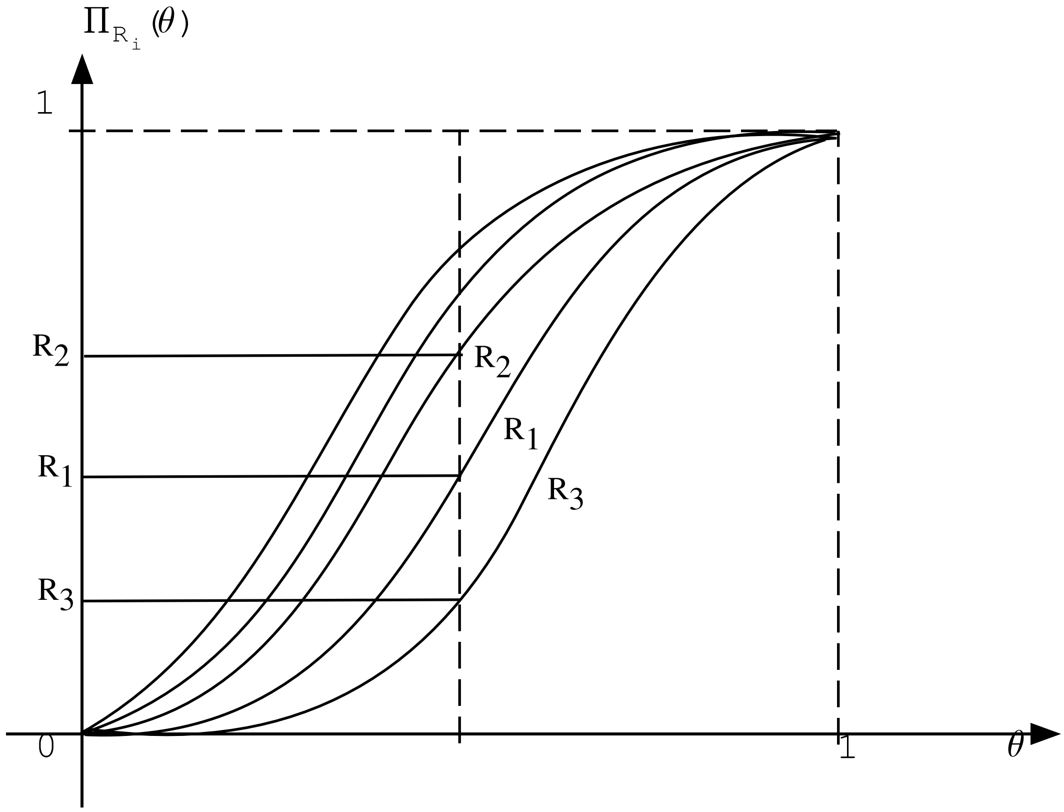
$$R_2 = \{6,7,8,9,10\}$$

$$R_3 = \{8,9,10\}$$

$$\vdots$$

$$R_3 \subset R_1 \subset R_2$$

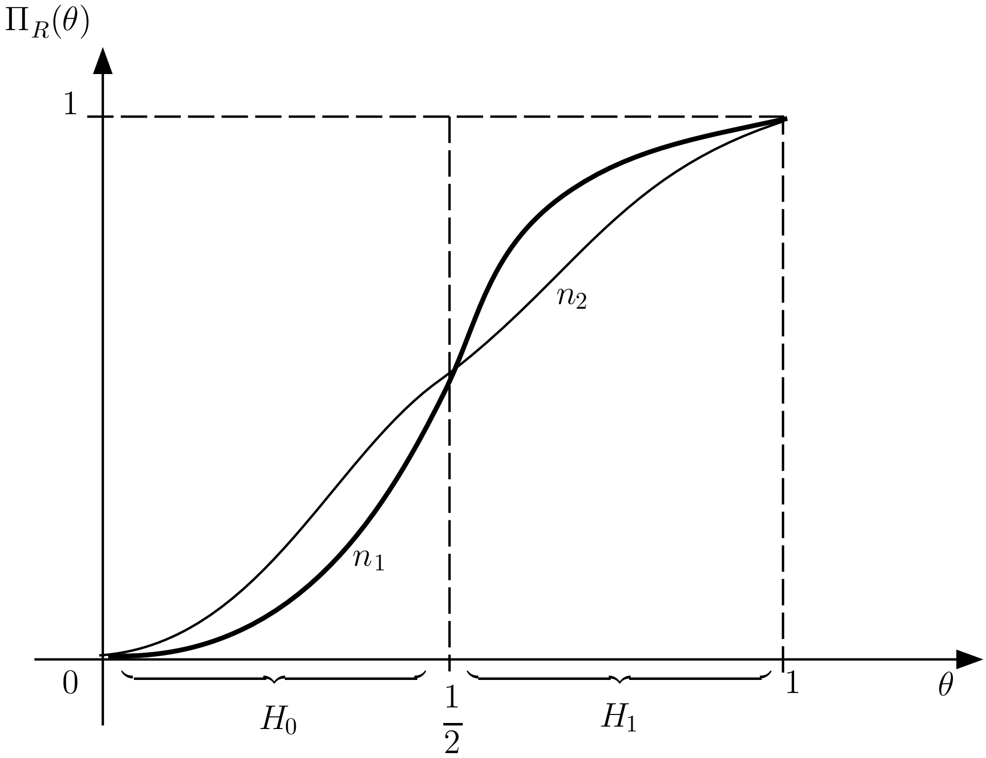
$$P(R_3|\theta) \leq P(R_1|\theta) \leq P(R_2|\theta) \qquad \forall\theta$$



8.2.6 Choix d'un test

Deux points de vue

Selon la taille d'échantillon



$$n_1 > n_2$$

Point de vue asymptotique: $n \to \infty$

Test "consistant"

$$P_n(\mathcal{E}_{II}) = b_n(\theta) \to 0 \qquad \forall\theta \in H_1$$

$$\Leftrightarrow \Pi_{R_n}(\theta) \to 1 \quad \forall\theta \in H_1$$

Le point de vue de Neyman

Une des deux erreurs est (beaucoup) plus grave . Par convention:

$\mathcal{E}_I$, c'est-à-dire: rejeter H_0 à tort, toujours la plus grave.

Exemple du lot avec pièces défectueuses

	H_0 $\theta \leq 2\%$	H_1 $\theta > 2\%$		H_1 $\theta \leq 2\%$	H_0 $\theta > 2\%$
accepter le lot		$\mathcal{E}_{II}$	accepter le lot		$\mathcal{E}_I$
rejeter le lot	$\mathcal{E}_I$		rejeter le lot	$\mathcal{E}_{II}$	
le plus grave: rejeter le lot alors qu'il est bon			le plus grave: accepter le lot alors qu'il est défectueux		

Il arrive souvent que les deux types d'erreur ne correspondent pas à des conséquences évaluables en termes économiques ou financiers, parce que les deux hypothèses se réfèrent à des réalités cognitives plutôt qu'économiques.
Par exemple: peut-on accepter l'hypothèse que le salaire des femmes est sous-évalué par rapport à leur compétence? que les étrangers commettent plus de délits que les nationaux? que le cancer atteint plus les habitants des villes que ceux des campagnes? etc.

L'enjeu n'est donc pas une décision économique à prendre dans l'immédiat mais plutôt la connaissance empirique d'un phénomène.
Dans ces cas, les hypothèses nulle et alternative seront spécifiées en tenant compte des aspects suivants:

– Parfois, l'hypothèse nulle représente une simplification d'un modèle plus complexe. Ce sera souvent le cas d'une hypothèse simple.

Par exemple: le remplissage moyen, μ, d'une bouteille de bière est égal à 25 cl. (alternative: différent de 25 cl., ou plus grand que, ou plus petit que)

– dans d'autres cas, l'hypothèse alternative représente une "nouvelle théorie" alors que l'hypothèse nulle représente une théorie davantage établie.

Par exemple: du temps de Galilée, pour "l'establishment de l'époque", l'hypothèse nulle était "le soleil tourne autour de la terre" et l'alternative était " la terre tourne autour du soleil".

Dans ces deux cas, on ne souhaite rejeter l'hypothèse nulle que si les observations lui sont "trop défavorables". C'est aussi une manière de dire que l'erreur de type I est plus grave que l'erreur de type II.

Exprimer les conclusions d'un test d'hypothèses sous la forme "On peut dire que ..." ou "On ne peut pas dire que ..." se réfère donc implicitement à l'hypothèse alternative car celle-ci est l'hypothèse qui n'est pas rejetée à tort (= erreur de type I) avec seulement beaucoup de précautions (c'est-à-dire au niveau α). Nous éviterons cependant une telle expression, sans pour autant la bannir, parce que les raisonnements de base des tests d'hypothèses sont faits explicitement sous l'hypothèse nulle (qui est de la sorte privilégiée dans les raisonnements) et parce que dans de nombreux cas l'hypothèse alternative est soit trop complexe soit non précisée de manière explicite. Nous préférerons donc l'expression "Au niveau α, l'hypothèse (nulle) est, ou n'est pas, rejetée"; si l'hypothèse nulle n'est pas rejetée, on dira aussi "qu'elle est acceptable".

8.2.7 Niveau d'un test:

Idées de base:
(i) imposer une limite au maximum possible de $P^\theta(\mathcal{E}_I) = a(\theta) \quad \theta \in H_0$
(ii) Comme $a(\theta)$ n'est défini que relativement à une région de rejet R, une écriture plus explicite sera $a_R(\theta)$.

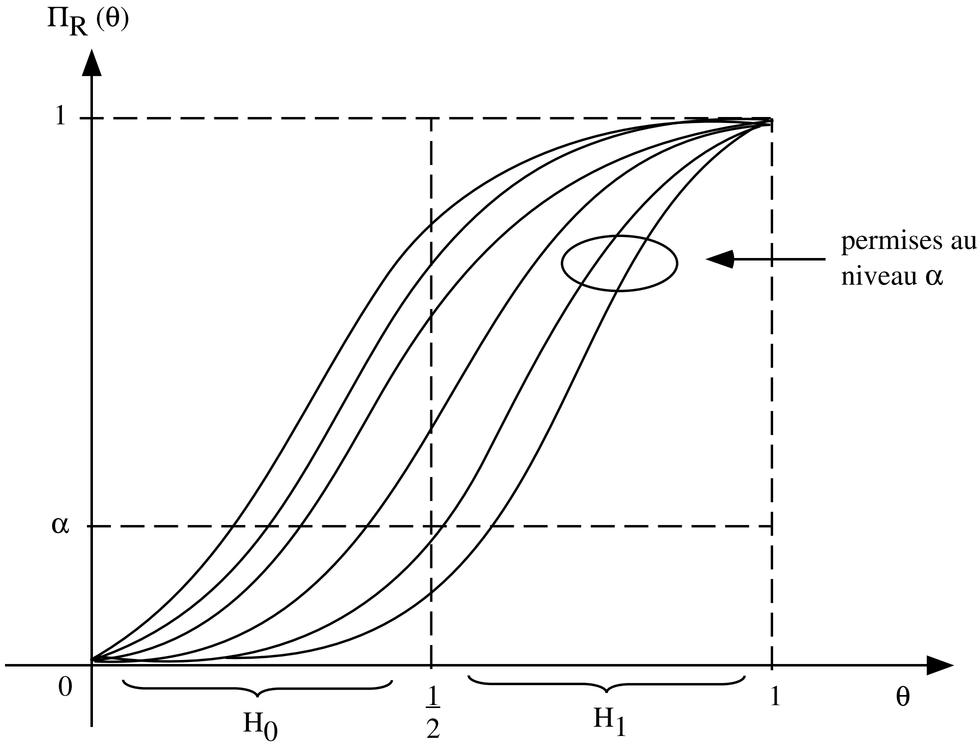
Définition: soit R, une région critique

$\alpha_R = \max\{a_R(\theta) : \theta \in H_0\}$ "niveau de la région critique R"

Utilisation de α_R: ne considérer que les régions critiques R telles que $\alpha_R \leq \alpha$ (α donné).

Souvent:

$$\begin{aligned} \alpha &= 0.05 \\ &= 0.01 \end{aligned}$$



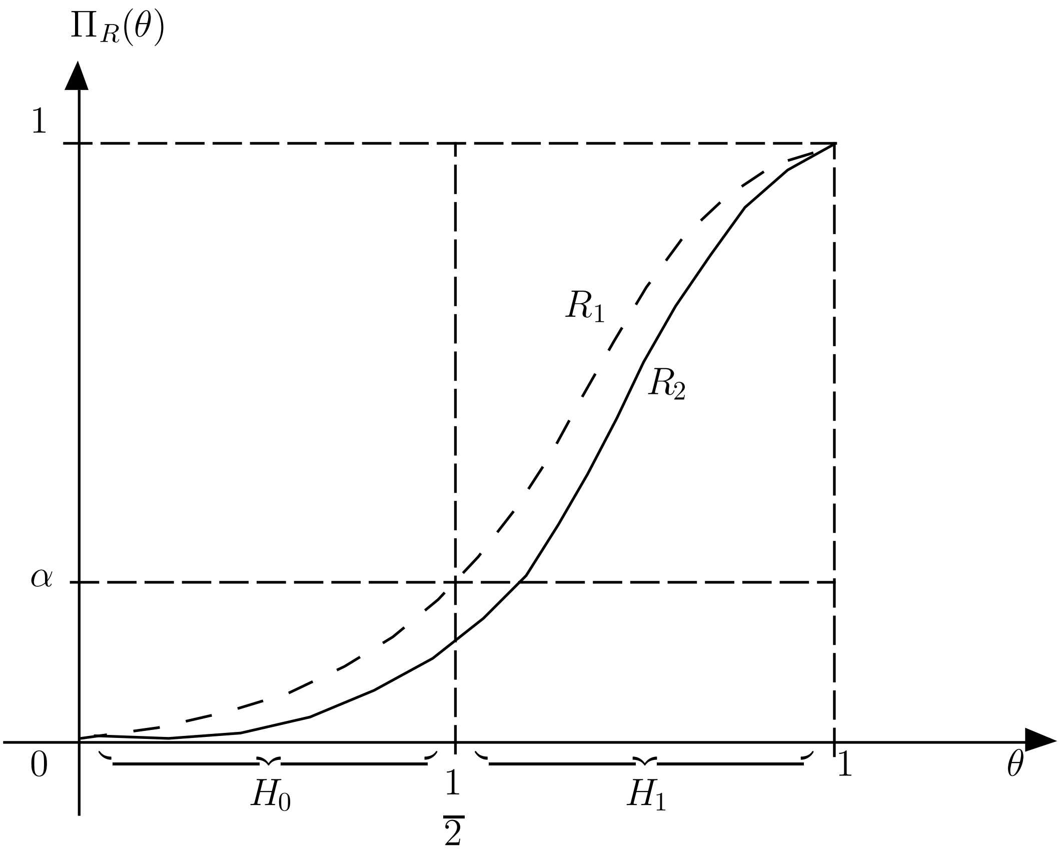
"permise au niveau α" : c'est-à-dire $\Pi_R(\theta) \leq \alpha \qquad \forall\, \theta \in H_0$.

8.2.8 Test le plus puissant

Idée de base

"Le test le plus puissant parmi les tests de niveau α": parmi les tests permis au niveau α, choisir celui dont $b_R(\theta) = P^\theta(\mathcal{E}_{II})$ est la plus petite, équivalent: $\Pi_R(\theta)$ est la plus grande dans la région $\theta \in H_1$.

Cas simple:

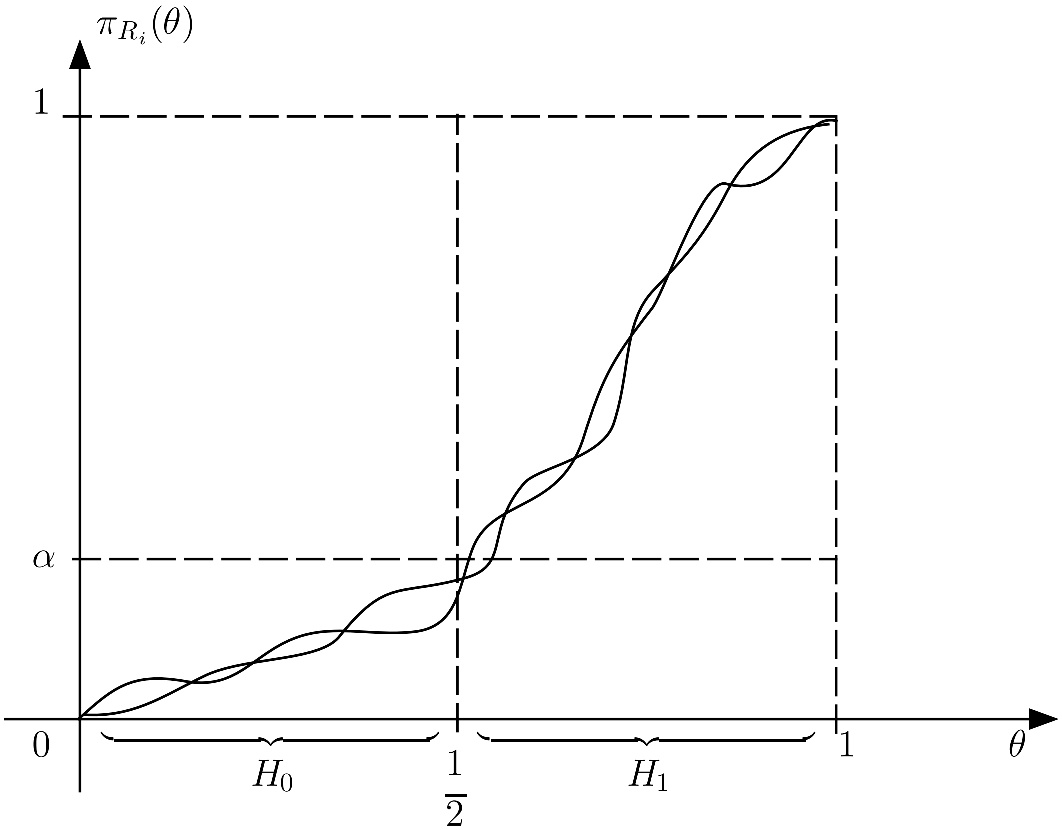


On remarque que $\forall\ \theta \in H_1$:

$$b_{R_1}(\theta) < b_{R_2}(\theta)$$
$$\Pi_{R_1}(\theta) > \Pi_{R_2}(\theta)$$

Dans ce cas: test $- - -$ est UNIFORMEMENT PLUS PUISSANT que le test —.

Cas difficile



8.3 Méthode de l'alpha critique

8.3.1 Cas unidirectionnel

Considérons plus particulièrement le cas unidirectionnel.

Par exemple: $H_0 \qquad \theta \leq \frac{1}{2} \qquad H_1 \qquad \theta > \frac{1}{2}$

Soit t une valeur observée de la statistique de test T ($T: \mathcal{X} \rightarrow \mathcal{T}$).

R_t ensemble des valeurs possibles de T plus défavorables à H_0 que t.

$$R_t = \{u \in \mathcal{T} | f_T^\theta(u) \leq f_T^\theta(t) \qquad \forall \theta \in H_0\}$$

Définition: La p-valeur, ou α-critique, ou niveau de signification de l'observation t:

$$\alpha_c(t) = \max\{P[R_t|\theta] : \ \theta \in H_0\}$$

Interprétations:

(i) probabilité d'observations au moins aussi défavorables à H_0 que t

(ii) niveau du test qui utiliserait R_t comme région critique

(iii) mesure de la "discordance" entre une observation et une hypothèse nulle

Utilisation pratique

1. choisir au départ un niveau α
2. calculer $\alpha_c(t)$
3. règle

$$\alpha_c(t) \leq \alpha \quad : \text{rejeter } H_0$$

$$\alpha_c(t) > \alpha \quad : \text{ne pas rejeter } H_0$$

Lorsque l'hypothèse nulle H_0 est rejetée au niveau α, on pourra éventuellement "affirmer, au seuil de signification α que $\theta \in H_1$"

Remarque

observation discrète:

en général: il n'existe pas de test de niveau *exactement* égal à α mais on peut calculer $\alpha_c(t)$

par exemple, T = nombre de piles

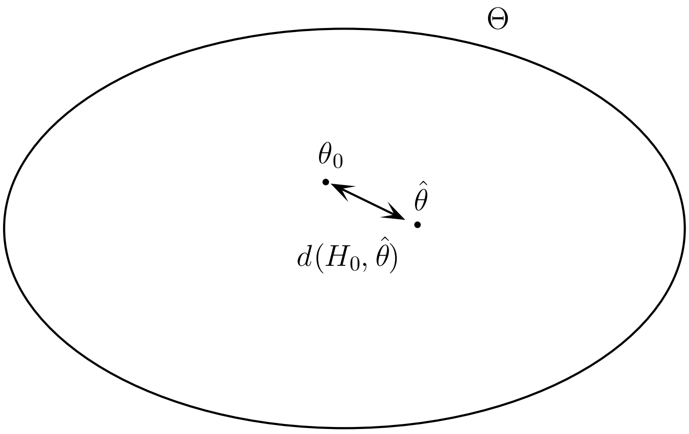
8.3.2 Cas général

Etendons la méthode antérieure à des cas plus complexes que le cas unidirectionnel.

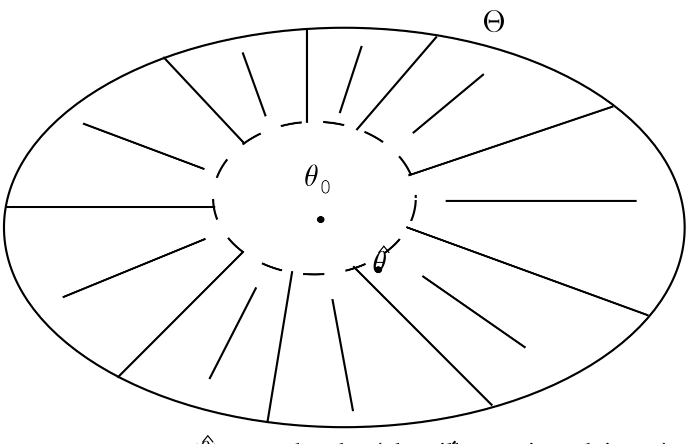
Idée de base: chercher une statistique de test qui "mesure" une distance entre H_0 et $\hat{\theta}$, l'estimation de θ, c'est-à-dire $T = d(H_0,\hat{\theta})$.

Graphiquement

Lorsque H_0 est simple: $H_0 : \theta = \theta_0$.



$R_d = \{d(H_0,\hat{\theta}) > d\}$
$R_d^c = \{d(H_0,\hat{\theta}) \leq d\}$ (où R_d^c est l'ensemble complémentaire R_d)



$\alpha_c(\hat{\theta})$ = proba. des échantillons qui conduisent à une estimation dans la région ////

$\alpha_c(t)$: proba. d'observations plus défavorables à H_o que t
α : maximum permis pour la proba. de $\mathcal{E}_I$

8.4 Intervalle de confiance

Remarque sur le non rejet de H_0

Considérons

$$\bullet \begin{cases} H_0 \\ \\ H_1 \end{cases}$$

$$\bullet\bullet \begin{cases} \text{rejeter } H_0 \quad (\text{ ou encore: accepter } H_1) \\ \\ \text{ne pas rejeter } H_0 \quad (\text{ ou encore: ne pas accepter } H_1) \end{cases}$$

Exemple

X_i contenu d'une bouteille de bière

H_0 $E(X_i) = \mu$ =25 cl

H_1 $\mu \neq 25$ cl

$\overline{X} = 24{,}958 \rightarrow H_0$ acceptable

De même pour:

H_0^* $\mu = 24{,}9$

$\mu = 24{,}8$

$\mu = 25{,}01$

$\vdots$

Intervalle de confiance:

Soit t valeur observée d'une statistique de test.

Définition d'un intervalle de confiance

$$I_\theta^\alpha(t) = \{\theta \in \Theta | P(R_t|\theta) \leq \alpha\} \subset \Theta$$

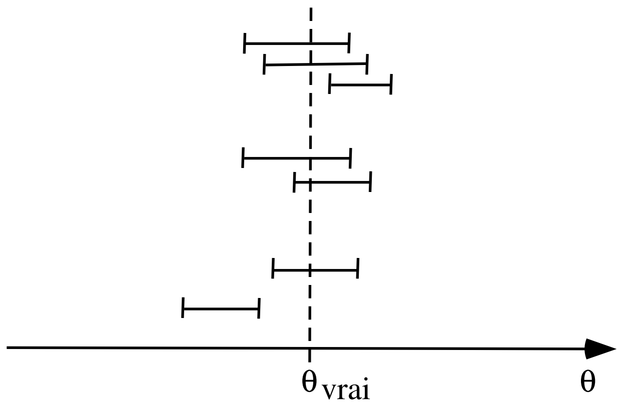
Interprétation: ensemble de valeurs de θ qui seraient une hypothèse nulle à ne pas rejeter au niveau α, après avoir observé t.

Construction: Lorsque le paramètre θ est à une dimension, l'intervalle de confiance prendra la forme :

$$I_\theta^\alpha(E) = [a(x) \qquad b(x)]$$

où $a(x)$ et $b(x)$ sont les valeurs de deux statistiques choisies de telle sorte que:

$$P([a(X) \qquad b(X)] \ni \theta) \geq 1 - \alpha \qquad \forall\, \theta \in H_0$$



Chapitre 9

Tests d'hypothèses sur des paramètres de localisation

9.1 Introduction

- paramètre de localisation (rappel 1ère Candi.)

  par exemple: espérance mathématique (μ), médiane ou mode

•• H_m: essentiellement distribution normale $N(\mu,\sigma^2)$

9.2 Variance connue

9.2.1 Exemple

Contexte: machine à remplir des bouteilles de bière.

"bon" remplissage: 75 cl

Supposons:

cl par bouteille: distribution normale, $N(\mu,\sigma^2)$

$\sigma^2 = \sigma_0^2 = 1\ cl^2$ (connu).

Echantillon:

$$n = 25$$

Observer:

$$\begin{aligned} x &= (x_1,x_2 \cdots x_{25}) \\ \overline{x} &= 74{,}5 \text{ cl} \end{aligned}$$

9.2.2 Hypothèse maintenue H_m

Idée de base: l'ensemble des hypothèses qui ne seront pas remises en question, quel que soit le résultat de l'observation.

Dans l'exemple:

X_i quantité observée

$X_i \sim \text{ind.}N(\mu,\sigma^2) \qquad i = 1,2,\cdots n$

$n = 25$

$\sigma^2 = \sigma_0^2 = 1 \text{ cl}^2$

Dès lors :

$H_m : X_i \sim \text{ ind. } N(\mu,1\ cl^2) \quad i = 1,2,\ldots,25$

9.2.3 Hypothèses à éprouver

Hypothèse statistique: hypothèse portant sur un ou plusieurs paramètres d'un modèle statistique.

Dans l'exemple:
fournisseur "honnête" $\mu \geq 75$
fournisseur pas "honnête" $\mu < 75$

9.2.4 Décisions à envisager

Prenons le point de vue du Ministère des Affaires Economiques : deux décisions possibles

$$\left[\begin{array}{l} \text{amende} \\ \\ \text{pas amende} \end{array}\right.$$

9.2.5 Hypothèses nulle et alternative H_0 et H_1

La question: fournisseur honnête ?

$$H_0 : \mu \geq 75$$

$$H_1 : \mu < 75$$

9.2.6 Erreurs en jeu

	H_0 $\mu \geq 75$	H_1 $\mu < 75$
amende	gros problème $\mathcal{E}_I$	OK
pas amende	OK	"encourager la fraude" $\mathcal{E}_{II}$

9.2.7 Statistique de test: $\overline{X}$

Idées de base:

- Statistique (en général): fonction des observations, $T = f(X_1 \cdots X_n)$. C'est aussi un "résumé" des observations.

•• Statistique de test: statistique qui sert de base pour un test d'hypothèse.

Dans l'exemple:

Observe: $\overline{x} = 74.5$ cl.

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{en général}$$

$$\sim N\left(\mu, \frac{1}{25}\right) \quad \text{dans ce cas-ci}$$

ou encore, quelque soit la valeur inconnue de μ :

$$\frac{\overline{X} - \mu}{1/5} \sim N(0,1)$$

Sous H_0: $\mu = 75$

$$\frac{\overline{X} - 75}{1/5}$$

- est une statistique de test
- sa distribution est connue: $N(0{,}1)$

9.2.8 Partition critique

Idées de base:

- Partitionner l'espace des observations en 2 régions:

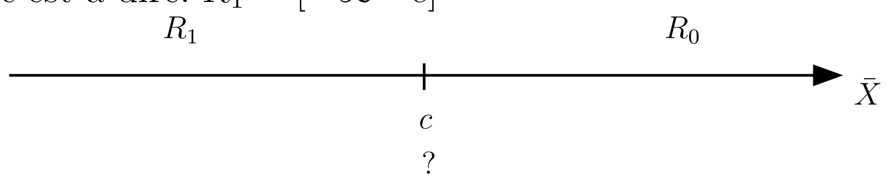
$$\left[\begin{array}{l}\text{Région d'acceptation}\\ \\ \text{Région de rejet - région critique}\end{array}\right.$$

•• on considère l'espace des observations de la statistique de test.

Dans l'exemple:

Intuitivement: Région critique: $\overline{X}$ "petit"

c'est-à-dire: $R_1 = [-\infty \quad c]$



A rechercher la valeur de c telle que

$$\begin{aligned} P(\mathcal{E}_I) &\leq \alpha && \text{en général} \\ &\leq 0{,}05 && \text{lorsqu'on prend } \alpha = 0{,}05 \end{aligned}$$

Dans ce cas-ci: c doit être tel que

$$P[\overline{X} \leq c|\mu] \leq \alpha \qquad\qquad \mu \geq 75$$

ou encore

$$P\left[\frac{\overline{X}-\mu}{1/5} \leq \frac{c-\mu}{1/5}\right] \leq \alpha, \quad \mu \geq 75$$

$$\downarrow$$

$$Z \sim N(0{,}1)$$

Dès lors:

$$P(\mathcal{E}_I) = \Phi\left(\frac{c-\mu}{1/5}\right) \qquad \mu \geq 75$$

Φ: fonction de distribution de $N(0{,}1)$, c'est-à-dire : $\Phi(a) = P(Z \leq a)$ où $Z \sim N(0{,}1)$

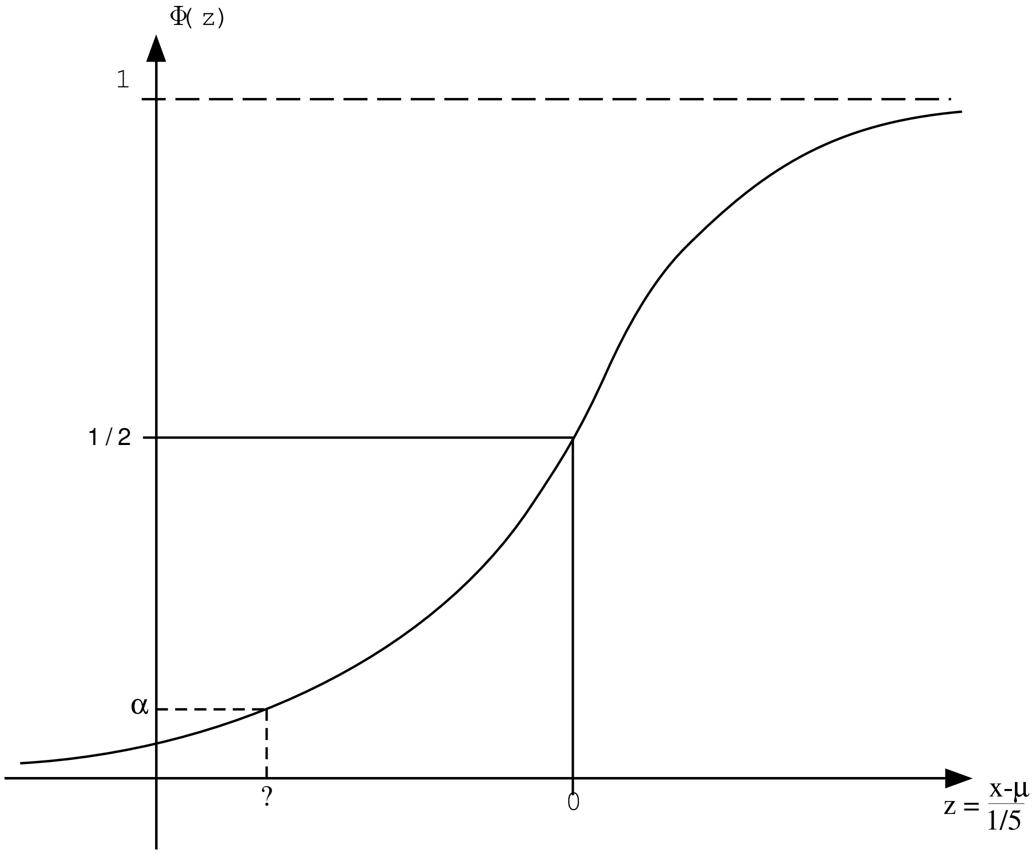
Remarquons:

$$P(\mathcal{E}_I) \leq 5\% \quad \forall\mu \in H_0 \Leftrightarrow \max\{P(\mathcal{E}_I) : \ \mu \in H_0\} \leq 5\%$$

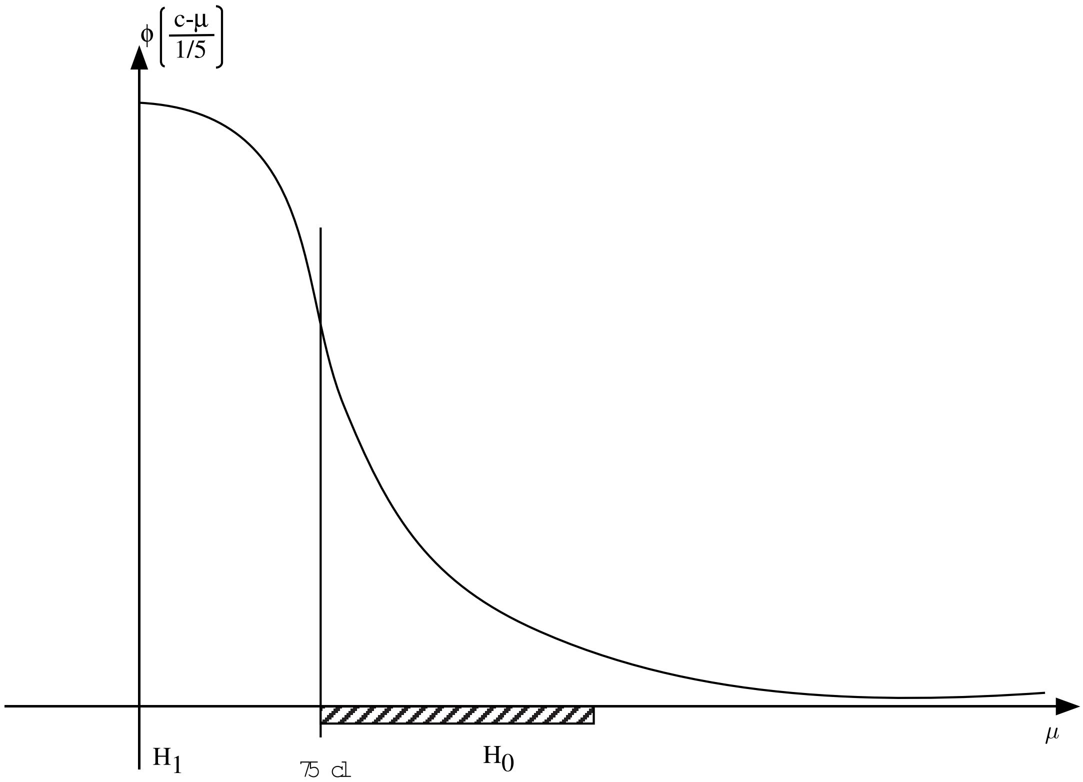
Il faut donc calculer c tel que:

$$\max\left\{\Phi\left(\frac{c-\mu}{1/5}\right) : \ \mu \geq 75\right\} \leq 5\%$$

Rappelons que la fonction de distribution d'une v.a. $N(0{,}1)$ est une fonction monotone croissante:


Φ(z)
1
1 / 2
α
?
0
z = (x-μ)/(1/5)

On peut en déduire que $\Phi\left(\frac{c-\mu}{1/5}\right)$ est une fonction monotone **décroissante** de μ:



Dès lors le maximum de $P(\mathcal{E}_I)$ est obtenu pour $\mu = 75$ c'est-à-dire à la frontière entre H_0 et H_1

Le problème devient donc:

$$\text{trouver } c \text{ tel que } \Phi\left(\frac{c-75}{1/5}\right) = 0{,}05$$

Table: $\Phi(-1.6449) = 0.05$

équivalent (selon les tables!): $\Phi(+1.6449) = 0.95$

Dès lors, la valeur de c est telle que:

$$\frac{c-75}{1/5} = 5(c-75) = -1.6449$$

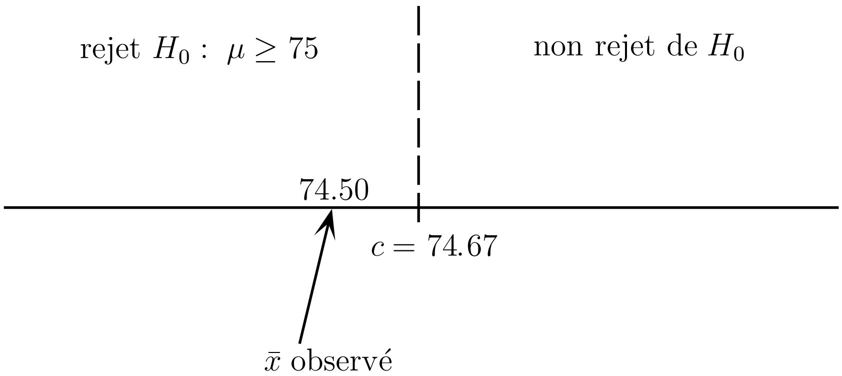
$$\rightarrow c - 75 = \frac{-1.6449}{5} = -0.33$$

$$c = 75 - 0.33 = 74.67$$

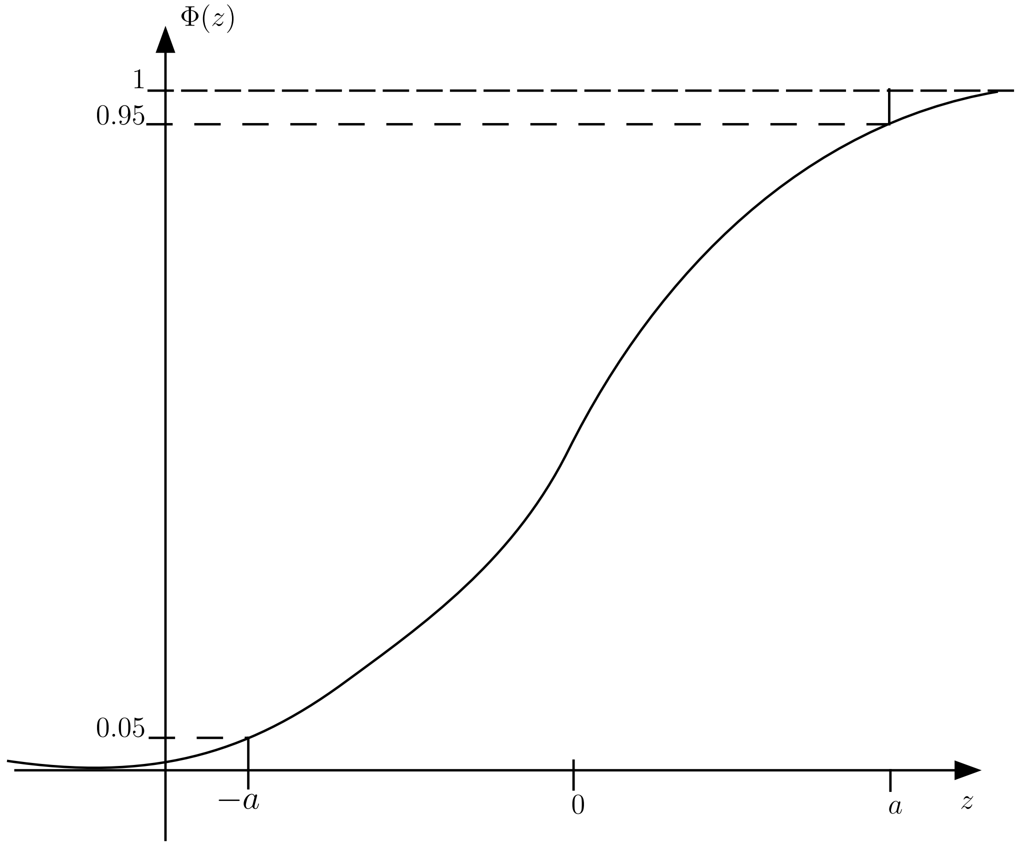
Conclusion

$c = 74{,}67$

$$\text{donc} \begin{cases} \text{rejet}: & \overline{X} \leq 74.67 \\ & \\ \text{non rejet}: & \overline{X} > 74.67 \\ \text{(acceptation)} & \end{cases}$$



Rappel: symétrie de la distribution normale



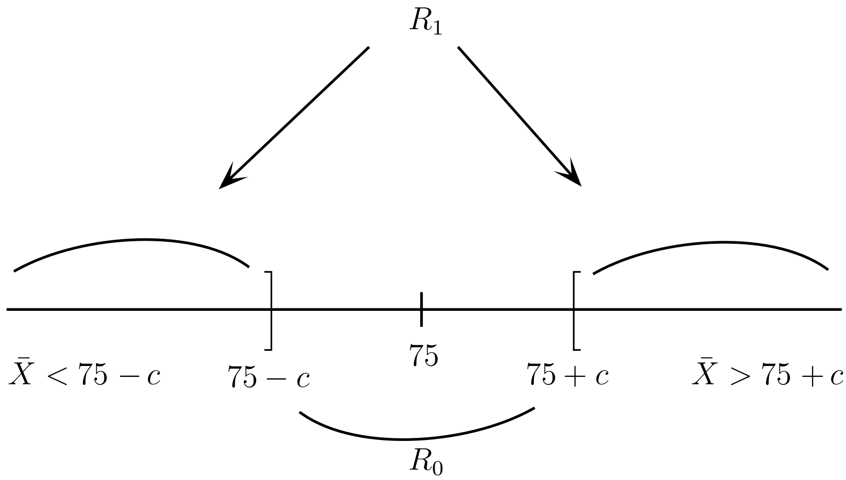
9.2.9 Partition critique pour différentes hypothèses alternatives

Considérons les 3 cas suivants:

(i)	H_0	$\mu = 75$	H_1	$\mu < 75$
(ii)	H_0	$\mu = 75$	H_1	$\mu > 75$
(iii)	H_0	$\mu = 75$	H_1	$\mu \neq 75$

Changement: forme de la région critique

1er cas ($\mu < 75$): déjà traité R_1 : $\overline{X} < 74.67$
2ème cas ($\mu > 75$): R_1 : $\overline{X} > 75.33$ (par symétrie de la distribution normale)
3ème cas ($\mu \neq 75$) : R_1 : $\overline{X}$ "loin de 75"



Le problème: chercher c tel que $P(\mathcal{E}_I) \leq .05 \qquad \forall \mu \in H_0$

Comme H_0 est une hypothèse simple: μ=75, cela devient:

$$P[75 + c < \overline{X} \text{ ou } \overline{X} < 75 - c] = 0.05 \quad \text{pour } H_0: \ \mu = 75$$

ou encore

$$P[75 - c < \overline{X} < 75 + c] = 0.95$$

$$\|$$

$$P\left[\frac{75 - c - \mu}{1/5} \leq \frac{\overline{X} - \mu}{1/5} \leq \frac{75 + c - \mu}{1/5}\right] = 0.95$$

Sous $H_0: \ \mu = 75$

$$P\left[-\frac{c}{1/5} \leq \frac{\overline{X} - 75}{1/5} \leq +\frac{c}{1/5}\right]$$

où encore avec $Z \sim N(0,1)$:

$$P[-5c \leq Z \leq +5c] = 0.95$$

Table:

$$\Phi(1.96) = .975$$

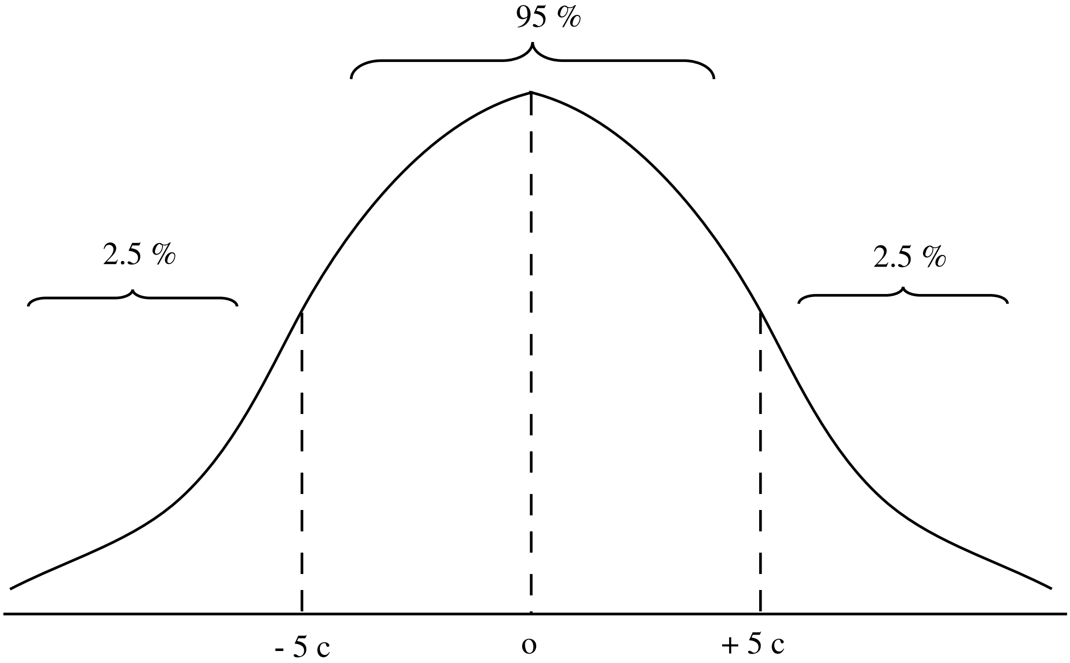
Donc

$$5c = 1.96 \quad c = \frac{1.96}{5} = .392$$

Dès lors:

$$\begin{aligned} R_0 &= (75 - .392; \quad 75 + .392) \\ &= (74.608; \quad 75.392) \\ R_1 &= [-\infty; \quad +74.608] \cup [75.392; \quad +\infty] \end{aligned}$$

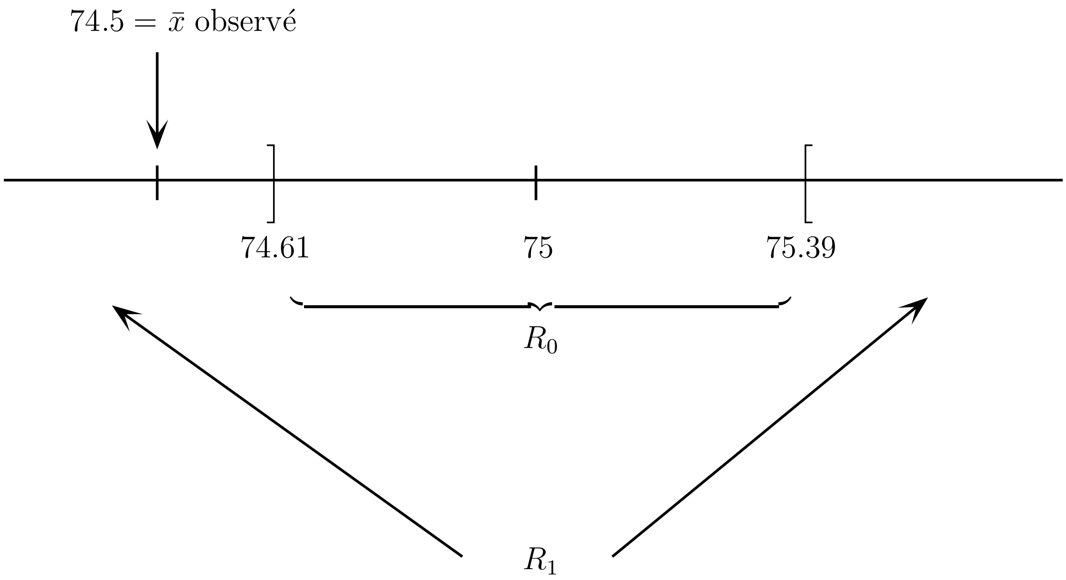
Graphiquement:



c'est-à-dire c tel que

$$\Phi\,(5c) = 0{,}975$$

Exemple:



Conclusion: rejet de $H_0: \ \mu = 75$ lorsque $H_1: \ \mu \neq 75$

9.2.10 Autre méthode: niveau de signification

Problème:

Est-ce que l'observation est "significativement éloignée de H_0"?

Exemple:

Comparons:

$$\begin{cases} H_0 \quad \mu = 75 \quad H_1 \; \mu < 75 \\ \\ H_0 \quad \mu \geq 75 \quad H_1 \; \mu < 75 \end{cases}$$

Observation: $\overline{x} = 74.5$

Question:

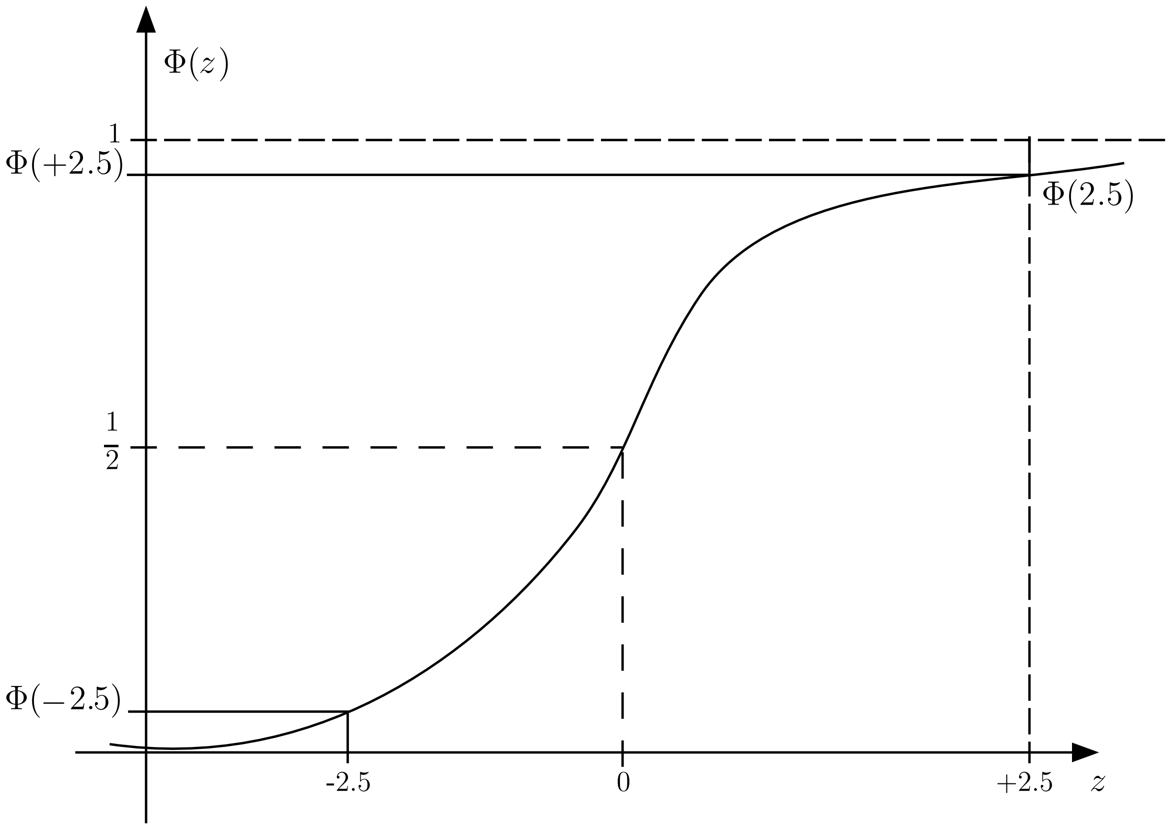
Quelle est la probabilité d'une observation *plus défavorable* à H_0 que $\overline{x} = 74.5$?

Plus défavorable: voir forme R_1

C'est-à-dire observation $\overline{X} \leq 74.5$

$$\begin{aligned} \alpha_c(74.5) = P[\overline{X} \leq 74.5 | \mu = 75] \; &= P\left[\frac{\overline{X} - 75}{1/5} \leq \frac{74.5 - 75}{1/5}\right] \\ &= P[Z \leq (-0.5)5] = \Phi(-2.5) \end{aligned}$$

où $\dfrac{\overline{X} - 75}{1/5} = Z \sim N(0,1)$



$$
\begin{aligned}
& \Phi(+2.5) &&= 0.9938 \qquad \text{(table)} \\
\Rightarrow \quad & \Phi(-2.5) &&= 1 - .9938 \\
& &&= 0.0062 \qquad < 0.05
\end{aligned}
$$

Conclusion:

$$
\text{Rejet de } H_0 \text{ pour } \quad \alpha \begin{aligned} &= .05 \\ &= .01 \end{aligned}
$$

mais non rejet pour $\alpha = .001$

9.2.11 Puissance du test

Reprenons

$$H_0: \ \mu \geq 75 \qquad\qquad H_1: \ \mu < 75$$

$$\rightarrow \ R_1 = (-\infty \quad 74.67] \qquad \text{pour un test au niveau } \alpha = 0.5$$

Fonction de puissance

- En général: $\Pi(\theta) = P[R_1|\theta]$ = probabilité de rejeter H_0 comme fonction de θ

  Rappel

$$\begin{aligned} \Pi(\theta) &= P(\mathcal{E}_I) & \theta \in H_0 \\ &= 1 - P(\mathcal{E}_{II}) & \theta \in H_1 \end{aligned}$$

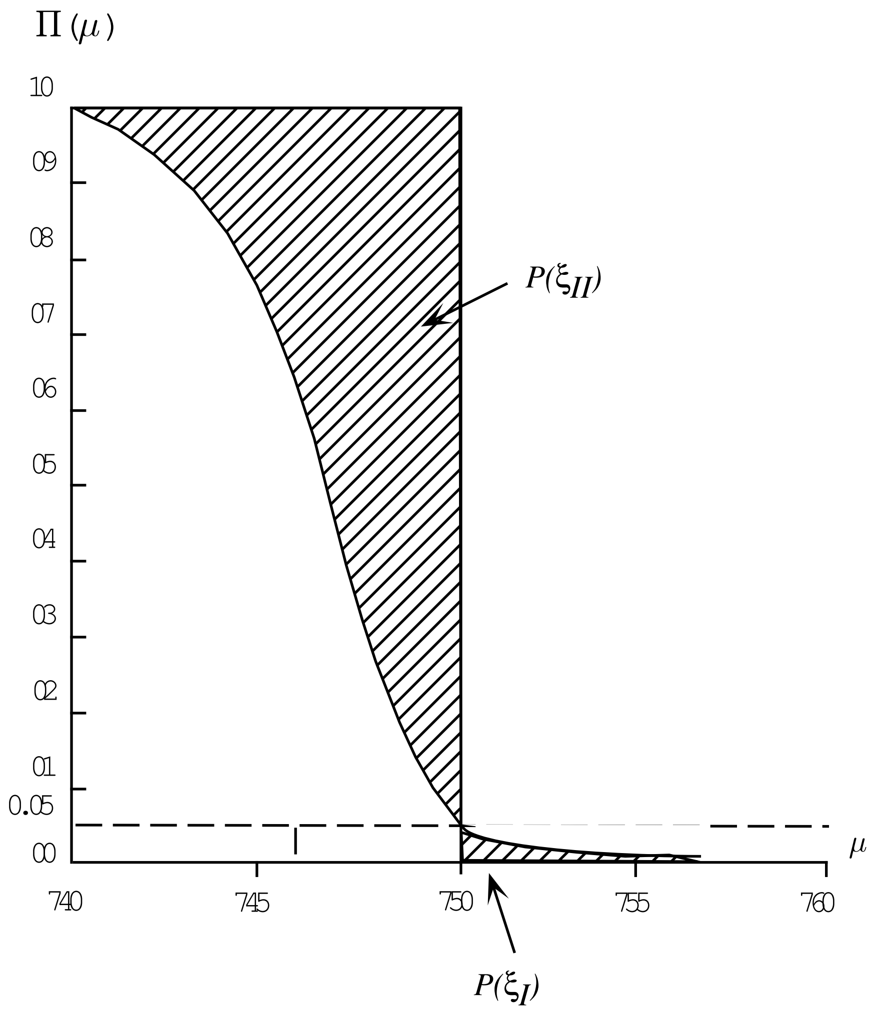
•• dans ce cas-ci:

$$\Pi(\theta)$$

$$\downarrow$$

$$\begin{aligned} \Pi(\mu) &= P(\overline{X} \leq 74.67|\mu) \\ &= P\left[\frac{\overline{X}-\mu}{1/5} \leq \frac{74.67-\mu}{1/5}\right] \\ &= \Phi[5(74.67-\mu)] \end{aligned}$$

Puissance du Test



1. à la frontière: $P(\mathcal{E}_I) + P(\mathcal{E}_{II}) = 1$
2. $P(\mathcal{E}_{II}) \searrow$ lorsque μ s'éloigne de la frontière

9.3 Test sur μ avec σ^2 aussi inconnue

9.3.1 Exemple: quotient intellectuel

Echantillon

$$x = (x_1 \cdots x_n)$$

$$n = 25$$

Hypothèse:

$$X_i \sim \text{ind.} N(\mu, \sigma^2)$$

Statistique pour "résumer" l'échantillon

$$\overline{x} = \frac{1}{n} \sum x_i \qquad = 105$$

$$s^2 = \frac{1}{n-1} \sum (x_i - \overline{x})^2 \quad = 81$$

Question:

$$\mu \quad \left\{ \begin{array}{c} > \\ = \\ < \end{array} \right\} \quad 100\,?$$

Plus précisément:

$$H_0 \qquad \mu \leq 100 \qquad H_1 \quad \mu > 100$$

Rappelons:

$\overline{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$

Dès lors:

$$\frac{\overline{X} - \mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$$

Remarquons

$\dfrac{\overline{X} - \mu}{\sqrt{\sigma^2/n}}$ **"quantité PIVOTALE"**, c'est-à-dire que sa distribution ne dépend pas de paramètre inconnu.

Dans le cas présent:

$$H_0 : \quad \begin{cases} \mu \leq 100 \\ \sigma^2 \geq 0 \end{cases} \qquad H_1 : \quad \begin{cases} \mu > 100 \\ \sigma^2 \geq 0 \end{cases}$$

$$H_m \quad X_i \sim \text{ind } N(\mu,\sigma^2) \quad -\infty < \mu < +\infty \quad \sigma^2 \geq 0$$

Donc, la région critique sera de la forme "grande valeur de $\overline{X}$". Précisons cette intuition.

9.3.2 Procédé de test

Intuition

Remplacer σ^2 par s^2.

Motivation: $E(s^2) = \sigma^2$ (sans biais)

On peut alors montrer:

$$\frac{\overline{X} - \mu}{\sqrt{s^2/n}} \sim t_{n-1} \text{ “Student à } n-1 \text{ degrés de liberté”}.$$

A remarquer

1. En général t_ν (Student à ν degrés de liberté). Distribution qui “ressemble ” à la distribution normale. Mais moins concentrée autour de 0, donc les queues de distribution de t_ν sont moins fines.
2. Lorsque $\nu \to \infty$, alors $t_\nu \to N(0,1)$ en pratique $\nu \geq 30$
3. ”Student” : pseudonyme utilisé par le statisticien anglais William S. Gosset (1876-1937)

Utilisation pratique:

Chaque fois que, dans la section 9.2, on a utilisé:

$\Phi(\cdot)$ fonction de distribution de $N(0,1)$

on la remplacera par:

$F_{t_\nu}(\cdot)$ fonction de distribution de t_ν

"Correction de petits échantillons"

Dans l'exemple

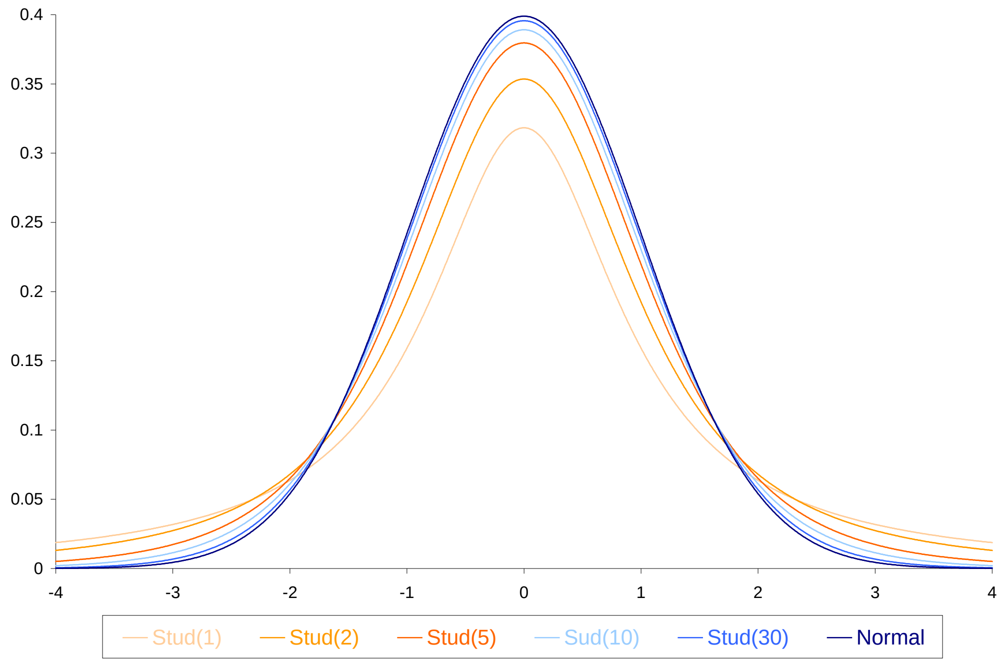
$$\begin{aligned} t_{obs} &= \frac{105-100}{\sqrt{\frac{81}{24}}} \\ &= 2{,}72. \end{aligned}$$

où $t \sim t_{(24)}$ lorsque $\mu = 100$

Dans les tables :

$0.005 < P[t_{(24)} \geq t_{obs}] < 0.01$

Donc rejet de H_0 lorsque $\alpha = 0.05$.



Graphique de quelques densités de Student

Test sur une Moyenne (logiciel SAS)

```
                     The UNIVARIATE Procedure
                        Variable:  saltest=salaire-45000

                              Moments

N                          504    Sum Weights                504
Mean                -903.97619    Sum Observations       -455604
Std Deviation        15610.273    Variance             243680625
Skewness            2.17976607    Kurtosis            13.2222816
Uncorrected SS      1.22983E11    Corrected SS        1.22571E11
Coeff Variation     -1726.8456    Std Error Mean      695.336827

                    Basic Statistical Measures

        Location                    Variability

    Mean      -903.98     Std Deviation            15610
    Median   -3000.00     Variance             243680625
    Mode         0.00     Range                   165000
                          Interquartile Range      14000

                    Tests for Location: Mu0=0

       Test           -Statistic-    -----p Value------

       Student's t    t  -1.30006    Pr > |t|    0.1942
       Sign           M       -47    Pr >= |M|   <.0001
       Signed Rank    S  -10437.5    Pr >= |S|   0.0003

                     Quantiles (Definition 5)

                     Quantile      Estimate

                      100% Max        130000
                      99%              45000
                      95%              24000
                      90%              15000
                      75% Q3            5000
                      50% Median       -3000
                      25% Q1           -9000
                      10%             -18000
                      5%              -22785
                      1%              -28000
                      0% Min          -35000

                        Extreme Observations

             -----Lowest----        -----Highest----

              Value      Obs          Value      Obs

             -35000      425          47000      305
             -31000       16          55000      152
             -30000      144          55000      340
             -30000       70         100000      283
             -30000       32         130000      469
```

9.4 Test d'égalité de deux moyennes

9.4.1 Exemple : quotient intellectuel (suite)

Echantillon

$n = 25$ dont 14 filles : $x = (x_1, \ldots, x_{14})$
11 garçons : $y = (y_1, \ldots, y_{11})$

Hypothèses

$$X_i \sim \text{ ind } N(\mu_X, \sigma_X^2) \quad i = 1, \ldots, n_X (= 14)$$

$$Y_i \sim \text{ ind } N(\mu_Y, \sigma_Y^2) \quad i = 1, \ldots, n_Y (= 11)$$

Statistique pour "résumer" les échantillons

$$\overline{x} = 109 \qquad \overline{y} = 100$$

$$s_X^2 = 85 \qquad s_Y^2 = 76$$

Question :

$$H_0 \quad \mu_X = \mu_Y$$

$$H_1 \quad \mu_X \neq \mu_Y$$

Difficulté nouvelle : $\sigma_X^2 = \sigma_Y^2 \quad ?$

9.4.2 Elaboration du test

1er cas : $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ (hypothèse maintenue)

- Estimation de σ^2

$$s_p^2 = \frac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_Y - 2}$$

Remarque : "p" pour "pooled", c'est-à-dire en combinant les deux échantillons.

Exercice : Expliquer pourquoi l'estimateur s_p^2 vous semble raisonnable.

Rappelons,

$$\overline{X} \sim N(\mu_X, \frac{\sigma^2}{n_X})$$

$$\overline{Y} \sim N(\mu_X, \frac{\sigma^2}{n_Y})$$

Dès lors,

$$\overline{X} - \overline{Y} \sim N\left(\mu_X - \mu_Y, \sigma^2\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)\right)$$

On peut alors montrer que,

$$t = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{s_p^2\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}}$$

est distribué selon $t_{(n_X+n_Y-2)}$

Dans l'exemple

$$s_p^2 = \frac{(13)(85) + (10)76}{23} = 81$$

$$\begin{aligned} t_{obs} &= \frac{109 - 100}{\sqrt{81\left(\frac{1}{14} + \frac{1}{11}\right)}} \\ &= 2{,}48 \end{aligned}$$

est une observation d'une v.a. $t_{(23)}$

Dans les tables

$$0.01 < P[t_{(23)} \geq 2{,}48] < 0.025$$

Dès lors, le niveau de signification - ou α critique - α_c sera

$$0.02 < \alpha_c < 0.05$$

Conclusion : rejet de H_0 lorsque $\alpha = 0.05$ mais non rejet si $\alpha = 0.01$

2ème cas : $\sigma_X^2 \neq \sigma_Y^2$

Si n_X et n_Y sont "grands", on négligera l'erreur d'estimation de σ_X^2 par s_X^2 (et de σ_Y^2 par s_Y^2).

Dès lors

$$\overline{X} \underset{app}{\sim} N\left(\mu_X, \frac{s_X^2}{n_X}\right)$$

$$\overline{Y} \underset{app}{\sim} N\left(\mu_X, \frac{s_Y^2}{n_Y}\right)$$

$$\overline{X} - \overline{Y} \underset{app}{\sim} N\left(\mu_X - \mu_Y, \frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}\right)$$

et donc

$$\frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}} \underset{app}{\sim} N(0,1)$$

Si n_X et n_Y ne sont pas "grands" : problème (théoriquement) délicat, plusieurs suggestions possibles, dont une est calculée dans SAS.

En pratique
On commence par tester l'égalité des variances et selon le résultat de ce test, on utilise le test d'égalité des moyennes en supposant soit l'égalité des variances soit leur différence (par exemple, dans ce cas, test de Satterthwaite).

Test sur une différence de deux moyennnes (logiciel SAS)

The TTEST Procedure

Statistics

Variable	sexe	N	Mean	Std Dev	Std Err
salaire	1	273	49273	16382	991.46
salaire	2	231	37977	12089	795.39

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
salaire	Pooled	Equal	502	8.67	<.0001
salaire	Satterthwaite	Unequal	493	8.89	<.0001

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
salaire	Folded F	272	230	1.84	<.0001

9.5 Test sur les proportions

9.5.1 Exemple : assuétude à la cigarette

Echantillon

$$X_i = \begin{cases} 1 & \text{si individu } i \text{ souffre d'assuétude} \\ 0 & \text{sinon} \end{cases}$$

Hypothèse

$$X_i \sim \text{ ind } Be(\pi) \qquad i = 1, \ldots, n$$

où $\pi = P(X_i = 1) = E[X_i]$

Statistique pour "résumer" l'échantillon

$$p = \overline{x} = \frac{\sum x_i}{n} = 0.15$$

Question

L'échantillon est pris dans la population des étudiants de l'U.C.L. On sait que dans cette tranche d'âge la proportion des fumeurs est de 0,20. La question posée est :

$$H_0 \quad \pi \geq 0{,}20 \qquad H_1 : \pi < 0{,}20$$

9.5.2 1er cas : n "petit"

Par exemple $n = 20$

Dans ce cas :(T="Total")

$$T = N_p = \sum_{i=1}^{n} X_i \sim B_i(n,\pi)$$

Région critique de la forme : $\{T \leq c\}$

On calcule donc (table de la loi binomiale) c tel que

$$P[T \leq c] \leq \alpha \qquad \text{où } T \sim B_i(n,\pi)$$

Par exemple : Pour $T \sim B_i(20{,}0.2)$, nous avons

$$\begin{aligned} P[T \leq 0] &= 0.0115 \\ P[T \leq 1] &= 0.0692 \\ P[T \leq 2] &= 0.2061 \qquad \text{etc} \end{aligned}$$

Donc pour $\alpha = 0{,}05$: $c = 0$
$T_{obs} = N_p = (20)(0.15) = 3$
Conclusion : non rejet de H_0

Exercice Observez attentivement, dans les tables de la loi binomiale, les conséquences du caractère discret de T. En particulier :

- • Quelle difficulté rencontre-t-on si on veut construire un test dont le niveau soit exactement $\alpha = 0.05$? (A relire: remarque à la fin de la section 8.3.1)
- •• Que se passe-t-il lorsque n augmente ?

9.5.3 2ème cas : n "grand"

Lorsque n est "grand", par exemple $n = 100$,

$$p = \frac{\sum X_i}{n} \underset{app}{\sim} N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$

Exercice Expliquer pourquoi

On comprend donc pourquoi les tables de la loi binomiale ne sont pas disponibles pour de grandes valeurs de n.

Dès lors, lorsque n est "grand" :

$$\frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \underset{app}{\sim} N(0,1)$$

Donc, pour éprouver l'hypothèse :

$$H_0 \quad \pi = 0{,}20$$

On aura :

$$\frac{p-0{,}20}{\sqrt{\frac{(0{,}2)(0{,}8)}{n}}} = \frac{\sqrt{n}(p-0{,}20)}{0{,}4} \underset{app}{\sim} N(0,1)$$

Dans l'exemple

$$\frac{\sqrt{100}(0{,}15-0{,}20)}{0{,}4} = -4{,}625$$

Conclusion : sans se fatiguer pour consulter les tables, on rejette H_0.

Exercice Expliquez pourquoi il n'est pas nécessaire de consulter les tables.

9.5.4 Comparaison de deux proportions

Modèle d'échantillonnage

$$\begin{array}{llllll}
\text{filles :} & X_i & \sim & \text{ind} & Be(\pi_X) & i = 1, \ldots, n_X \\
 & & \text{où } X_i & = 1 & \text{si assuétude} & \\
 & & & = 0 & \text{sinon} & \\
\text{garçons :} & Y_i & \sim & \text{ind} & Be(\pi_Y) & i = 1, \ldots, n_Y \\
 & & \text{où } Y_i & = 1 & \text{si assuétude} & \\
 & & & = 0 & \text{sinon} &
\end{array}$$

Question

$$H_0 \qquad \pi_X = \pi_Y$$

Données

$$p_X = 0{,}10 \quad n_X = 80$$

$$p_Y = 0{,}12 \quad n_Y = 100$$

Elaboration du test

$$p_X \underset{app}{\sim} N\left(\pi_X, \frac{\pi_X(1-\pi_X)}{n_X}\right)$$

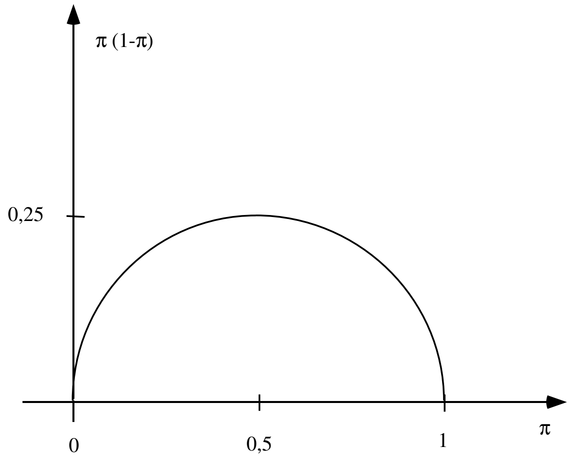
$$p_Y \underset{app}{\sim} N\left(\pi_Y, \frac{\pi_Y(1-\pi_Y)}{n_Y}\right)$$

$$p_X - p_Y \underset{app}{\sim} N\left(\pi_X - \pi_Y, \frac{\pi_X(1-\pi_X)}{n_X} + \frac{\pi_Y(1-\pi_Y)}{n_Y}\right)$$

Difficulté : sous $H_0 : \pi_X = \pi_Y$, on ne connaît pas la variance de la différence $p_X - p_Y$

Exercice En vous aidant de la figure ci-dessous, vérifiez que la variance la plus grande possible est :

$$0{,}25\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)$$



On a alors *deux solutions :*

a) solution "prudente" (si n_X et n_Y ne sont pas "très grands")

$$\frac{p_X - p_Y}{\sqrt{0{,}25\left(\frac{1}{n_X} + \frac{1}{n_Y}\right)}} \underset{app}{\sim} N(0{,}1)$$

b) solution moins pessimiste (si n_X et n_Y sont "vraiment grands")

$$\frac{p_X - p_Y}{\sqrt{\frac{p_X(1-p_X)}{n_X} + \frac{p_Y(1-p_Y)}{n_Y}}} \underset{app}{\sim} N(0{,}1)$$

Chapitre 10

Tables de contingence

10.1 Introduction

Soit

ρ et σ: deux caractéristiques catégorielles

R et S: les V.A. associées, dans un échantillon

Par exemple

ρ: fumeurs, non-fumeurs

σ: classe socio-professionnelle.

En général

$\rho \in \{1,2,\cdots,i,\cdots,r\}$

$\sigma \in \{1,2,\cdots,j,\cdots,s\}$

Table de la distribution conjointe

$\rho \downarrow \ \backslash \sigma \rightarrow$	1	2	$\cdots$	j	$\cdots$	s	Totaux
1	π_{11}	π_{12}		π_{1j}		π_{1s}	$\pi_{1\cdot}$
2	π_{21}	π_{22}		π_{2j}		π_{2s}	$\pi_{2\cdot}$
$\vdots$							
i	π_{i1}	π_{i2}		π_{ij}		π_{is}	$\pi_{i\cdot}$
$\vdots$							
r	π_{r1}	π_{r2}		π_{rj}		π_{rs}	$\pi_{r\cdot}$
Totaux	$\pi_{\cdot 1}$	$\pi_{\cdot 2}$		$\pi_{\cdot j}$		$\pi_{\cdot s}$	1

où π_{ij}= probabilité qu'un individu tiré au hasard appartienne à la catégorie i du critère ρ et j du critère σ.

10.2 Echantillonnage et hypothèses

Echantillon

$$s = (k_1 \cdots k_n) \subset \mathcal{N}$$

où (rappel !) k_j: "étiquette" (identification du jième individu de l'échantillon ($1 \leq j \leq n$)

Pour chaque individu, on observera alors sa catégorie pour chacun des 2 critères ρ et σ:

$k_1 \rightarrow R_1 = \rho_{k_1}$ et $S_1 = \sigma_{k_1}$

$k_2 \rightarrow R_2 = \rho_{k_2}$ et $S_2 = \sigma_{k_2}$

etc ...

Hypothèse maintenue

Un individu k tiré au hasard est donc classifié selon deux critères : R et S. L'observation est donc de la forme $X_k = (R_k, S_k)$.

Définissons

$$\pi_{r,s} = P[R_k = r, S_k = s] \qquad \pi_{r,s} \geq 0 \quad \sum_r \sum_s \pi_{r,s} = 1$$

Hypothèse maintenue: les X_k sont i.i.d.

Hypothèse nulle

$$H_0 : \rho \perp\!\!\!\perp \sigma$$

c'est-à-dire

$$\pi_{ij} = \pi_{i\cdot}\pi_{\cdot j}$$

ou encore

$$\pi_{i|j} = \pi_{i\cdot}$$

où

$$\pi_{i\cdot} = \sum_j \pi_{ij}$$

$$\pi_{\cdot j} = \sum_i \pi_{ij}$$

$$\pi_{i|j} = \frac{\pi_{ij}}{\pi_{\cdot j}}$$

10.3 Table de contingence

Soit: $N_{ij} = \#$ d'individus de l'échantillon tels que $\begin{cases} R = i \\ S = j \end{cases}$

$r \downarrow \backslash s \rightarrow$	1	2	$\cdots$	j	$\cdots$	s	totaux
1	N_{11}	N_{12}		N_{1j}		N_{1s}	$N_{1\cdot}$
2	N_{21}	N_{22}		N_{2j}		N_{2s}	$N_{2\cdot}$
$\vdots$							
i	N_{i1}	N_{i2}		N_{ij}		N_{is}	$N_{i\cdot}$
$\vdots$							
r	N_{r1}	N_{r2}		N_{rj}		N_{rs}	$N_{r\cdot}$
Totaux	$N_{\cdot 1}$	$N_{\cdot 2}$		$N_{\cdot j}$		$N_{\cdot s}$	n

N_{ij}: intersection $\left[\begin{array}{l} \text{ligne } i \\ \text{colonne } j \end{array} \right.$

$N_{\cdot j} = \sum_i N_{ij} = N_{1j} + N_{2j} + \cdots + N_{rj}$

$N_{i\cdot} = \sum_j N_{ij} = N_{i1} + N_{i2} + \cdots + N_{is}$

Remarque On peut aussi écrire:

$$N_{ij} = \sum_{i \leq k \leq n} \mathbb{1}_{\{R_k = i,\, S_k = j\}}$$

Exercice Vérifier que $N_{\cdot\cdot} = n$

Remarques:

1. N_{ij}: V.A.

2. échantillon aléatoire simple: $N_{ij} \sim Bi(n,\pi_{ij})$

   donc $E(N_{ij}) = n\pi_{ij}$
   $V(N_{ij}) = n\pi_{ij}(1-\pi_{ij})$

   Sous (H_0) : $\pi_{ij} = \pi_{i\cdot}\pi_{\cdot j}$
   $E(N_{ij}) = n\pi_{i\cdot}\pi_{\cdot j}$

Estimation de $E(N_{ij})$

en général $n \cdot \hat{\pi}_{ij} = n \, \dfrac{N_{ij}}{n} = N_{ij}$

sous H_0 $\quad n\,\hat{\pi}_{i\cdot} \cdot \hat{\pi}_{\cdot j} = n\,\dfrac{N_{i\cdot}}{n} \cdot \dfrac{N_{\cdot j}}{n} = \dfrac{N_{i\cdot}N_{\cdot j}}{n}$

10.4 Statistique de test: statistique de Pearson

$$U = \sum_{i,j} \sum \frac{\left(N_{ij} - \frac{N_{i.}N_{.j}}{n}\right)^2}{\frac{N_{i.}N_{.j}}{n}}$$

On peut aussi réécrire cette formule de la façon suivante:

$$U = n \left\{ \sum_i \sum_j \frac{N_{ij}^2}{N_{i.}N_{.j}} - 1 \right\}$$

c'est-à-dire une "**divergence**" entre 2 distributions ("distance")

$$\begin{cases} \dfrac{N_{ij}}{n} : \ \pi_{ij} \text{ estimé sous } H_1 \\[2ex] \dfrac{N_{i.}N_{.j}}{n^2} : \ \pi_{ij} \text{ estimé sous } H_0 \end{cases}$$

On écrira aussi, en posant $k = (i,j)$:

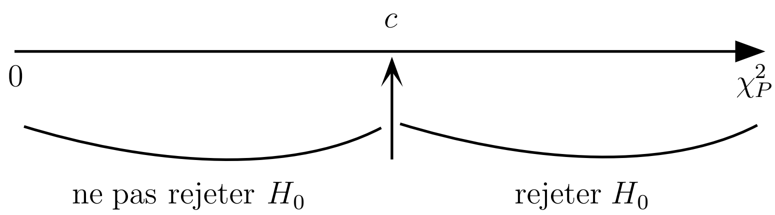
$$U = \sum_k \frac{(O_k - E_k)^2}{E_k}$$

où O_k est l'effectif observé dans la cellule k
E_k est l'effectif attendu dans la cellule k sous l'hypothèse d'indépendance

Partition critique:

Région critique: “U grand”

Région d’acceptation: “U petit”



Il faut donc chercher c tel que

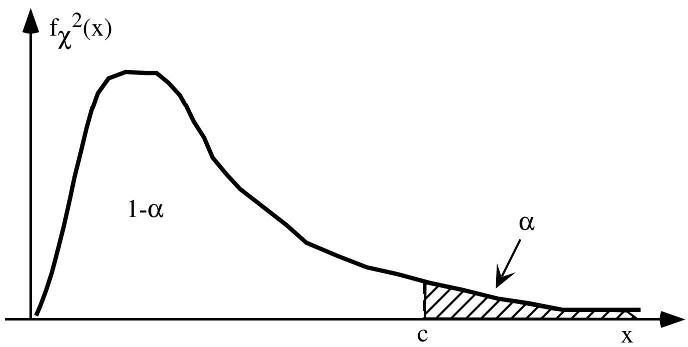
$P[U > c|H_0] \leq \alpha$

c’est-à-dire

$P(R_1|H_0) \leq \alpha$

ou encore

$P(\mathcal{E}_I) \leq \alpha$



Test d'indépendance (logiciel SAS)

```
                         The FREQ Procedure
                      Table of salcode by nsrev
salcode     nsrev

Frequency,
Percent  ,
Row Pct  ,
Col Pct  ,      1,      2,      3,      4,      5,      6,  Total
-------------------------------------------------------------
       1 ,      1 ,      3 ,      4 ,      3 ,      1 ,      2 ,      14
         , 0.20 , 0.60 , 0.79 , 0.60 , 0.20 , 0.40 ,   2.78
         , 7.14 ,21.43 ,28.57 ,21.43 , 7.14 ,14.29 ,
         , 3.13 , 8.82 , 4.26 , 1.97 , 0.76 , 3.33 ,
-------------------------------------------------------------
       2 ,     29 ,     27 ,     75 ,    104 ,     85 ,     32 ,   352
         , 5.75 , 5.36 ,14.88 ,20.63 ,16.87 , 6.35 ,69.84
         , 8.24 , 7.67 ,21.31 ,29.55 ,24.15 , 9.09 ,
         ,90.63 ,79.41 ,79.79 ,68.42 ,64.39 ,53.33 ,
-------------------------------------------------------------
       3 ,      2 ,      4 ,     13 ,     34 ,     39 ,     21 ,   113
         , 0.40 , 0.79 , 2.58 , 6.75 , 7.74 , 4.17 ,22.42
         , 1.77 , 3.54 ,11.50 ,30.09 ,34.51 ,18.58 ,
         , 6.25 ,11.76 ,13.83 ,22.37 ,29.55 ,35.00 ,
-------------------------------------------------------------
       4 ,      0 ,      0 ,      2 ,     11 ,      7 ,      5 ,    25
         , 0.00 , 0.00 , 0.40 , 2.18 , 1.39 , 0.99 , 4.96
         , 0.00 , 0.00 , 8.00 ,44.00 ,28.00 ,20.00 ,
         , 0.00 , 0.00 , 2.13 , 7.24 , 5.30 , 8.33 ,
-------------------------------------------------------------
Total          32      34      94     152     132      60     504
             6.35    6.75   18.65   30.16   26.19   11.90  100.00

              Statistics for Table of salcode by nsrev

        Statistic                      DF       Value      Prob
    ------------------------------------------------------------
        Chi-Square                     15     37.8504    0.0009
        Likelihood Ratio Chi-Square    15     41.6280    0.0003
        Mantel-Haenszel Chi-Square      1     27.1671    <.0001
        Phi Coefficient                        0.2740
        Contingency Coefficient                0.2643
        Cramer's V                             0.1582

         WARNING: 42% of the cells have expected counts less
                  than 5. Chi-Square may not be a valid test.
                           Sample Size = 504
```

10.5 Distribution de U

- compliquée
- •• approximation asymptotique: distribution limite “chi-carré à $(r-1)(s-1)$ degrés de liberté”

10.5.1 Par exemple

Soit

$$r = 2 \qquad s = 5 \qquad (r-1)(s-1) = 4$$

alors

$$U_n \xrightarrow{d} \chi^2_{(r-1)(s-1)} = \chi^2_4$$

10.5.2 Remarques

1. Conditions de l'approximation: règles ”fréquentes”:
   $N_{ij} \geq 2.5 \qquad \dfrac{N_{i.}N_{.j}}{n} \geq 2.5$
   $\frac{N_{i.}N_{.j}}{n} \geq 1 \quad \forall(i,j)$ et $\frac{N_{i.}N_{.j}}{n} \leq 5$ pour un maximum de 20% des cellules
2. Souvent: regroupement de catégories $\rightarrow$ A PRIORI
3. On peut représenter une variable aléatoire χ^2_l, à l degrés de liberté de la façon suivante:

   $$\chi^2_l = \sum_{i=1}^{l} Z_i^2 \quad \text{où } Z_i \sim \text{ ind. } N(0,1)$$

   On peut alors montrer :

   $$E[\chi^2_l] = l \;\; V(\chi^2_l) = 2l \;\text{ ou encore : }\; E\left[\frac{\chi^2_l}{l}\right] = 1;\; V\left(\frac{\chi^2_l}{l}\right) = \frac{2}{l}$$

Dès lors, en utilisant le théorème de la limite centrée, on obtient l'approximation suivante lorsque l est grand :

$$\frac{\chi_l^2}{l} \overset{app.}{\sim} N\left(1,\frac{2}{l}\right)$$

10.6 Mesures d'association

10.6.1 Le problème

- définir une mesure d'association entre deux variables catégorielles. Ces mesures sont basées sur la distribution conjointe des $[p_{ij}]$ ou, de façon équivalente dans la table de contingence $[N_{ij}]$, ou $p_{ij} = \dfrac{N_{ij}}{n}$
- déjà vu au chapitre 5 :
  - Covariance, corrélation (ρ)
  - indice de corrélation $\eta_{X|Y}$
- à remarquer (rappel !)
  $\rho(X,Y)$: seulement si X et Y sont quantitatives

  $$\eta_{X|Y} = \frac{VE(X|Y)}{V(X)} : \text{ seulement si } X \text{ est quantitative}$$
  $$Y : \text{ quantitative } \textit{ou} \text{ catégorique}$$

  Cette section : les *deux variables X et Y sont* catégoriques
- à remarquer : certaines mesures sont symétriques, entre X et Y, d'autres ne le sont pas.

10.6.2 Tables 2×2

Soit

	Critère S		Totaux
Critère R	p_{11}	p_{12}	$p_{1\cdot}$
	p_{21}	p_{22}	$p_{2\cdot}$
totaux	$p_{\cdot 1}$	$p_{\cdot 2}$	1

(i) **Coefficient lambda :** $\lambda = \dfrac{p_{11}\, p_{22}}{p_{12}\, p_{21}}$

Remarque Si $p_{12} = 0$ ou $p_{21} = 0$: permuter les lignes, ou les colonnes; dans ce cas : $\lambda = 0$

Propriétés :

- $\lambda \geq 0$
- $\lambda = 1 \Leftrightarrow R \perp\!\!\!\perp S$

(ii) **Association de Yule :** $Q = \dfrac{\lambda - 1}{\lambda + 1} = \dfrac{p_{11}\, p_{22} - p_{12}\, p_{21}}{p_{11}\, p_{22} + p_{12}\, p_{21}}$

(iii) **Collégation de Yule :** $Y = \dfrac{\sqrt{\lambda} - 1}{\sqrt{\lambda} + 1}$

Propriétés : $Y = Q = 0 \Leftrightarrow R \perp\!\!\!\perp S$

10.6.3 Table $r \times s$

C'est-à-dire : r lignes et s colonnes

(i)

$$\begin{aligned}\mathbf{\Phi^2} &= \sum_i \sum_j \frac{(p_{ij} - p_{i\cdot}p_{\cdot j})^2}{p_{i\cdot}p_{\cdot j}} \\ &= \frac{1}{n}\, U = \frac{1}{n} \sum_i \sum_j \frac{(N_{ij} - \frac{N_{i\cdot}N_{\cdot j}}{n})^2}{\frac{N_{i\cdot}N_{\cdot j}}{n}}\end{aligned}$$

Remarque SAS donne : Φ

Propriétés

- $\Phi^2 \geq 0$
- $\Phi^2 = 0 \Leftrightarrow R \perp\!\!\!\perp S$

Lorsque $r = s = 2$

$$\begin{array}{lll} \text{Soit } R_k & = 1 & \text{si l'individu } k \text{ est de catégorie } R \text{ égale à } 1 \\ & = 0 & \text{sinon} \\ S_k & = 1 & \text{si l'individu } k \text{ est de catégorie } S \text{ égale à } 1 \\ & = 0 & \text{sinon.} \end{array}$$

alors $\Phi^2 = [\text{corr}(R,S)]^2$

dès lors :

$$0 \leq \Phi^2 \leq 1$$

$$\Phi^2 = 1 \Leftrightarrow p_{12} = p_{21} = 0 : \text{association parfaite}$$

Lorsque $r = s > 2$

alors $\Phi^2 \leq r - 1$
$\Phi^2 = r - 1 \Leftrightarrow p_{ij} = 0 \quad i \neq j$ c'est-à-dire association parfaite.

Lorsque $r \neq s : \Phi^2 \leq \min\{r-1, s-1\}$

On a alors défini différentes manières de standardiser Φ^2.

(ii) **Coefficient de contingence de Pearson :**

$$C^2 = \frac{\Phi^2}{\Phi^2 + 1} = \frac{U}{U + n}$$

Remarque SAS : C

Propriétés

- $0 \leq C^2 < 1$
- $C^2 = 0 \leftrightarrow R \perp\!\!\!\perp S$
- lorsque $r = s : C^2 \leq \frac{r-1}{r}$
  par exemple si $r = s = 2 : C^2 \leq \frac{1}{2}$
- $C \leq 1 - \frac{1}{\min\{\rho, s\}}$

(iii) **Coefficient de Tsuprow** (pas dans SAS)

$$\begin{aligned} T^2 &= \frac{\Phi^2}{\sqrt{(r-1)(s-1)}} \\ &= 0 \Leftrightarrow R \perp\!\!\!\perp S \end{aligned}$$

(iv) **Coefficient de Cramer**

$$V^2 = \frac{\Phi^2}{\min\{r-1, s-1\}}$$

Propriétés

- $0 \leq V^2 \leq 1$
- $V^2 = 0 \Leftrightarrow R \perp\!\!\!\perp S$
- lorsque $r = s : V^2 = 1 \Leftrightarrow$ association parfaite
- SAS : V

Remarque : Lorsque $r = s = 2 : V^2 = T^2$

(v) **α-critique**

Soit $\alpha_c = P[\chi^2_{(r-1)(s-1)} \geq U]$ sous $H_0 : R \perp\!\!\!\perp S$

On peut interpréter :

α_c "grand" (par ex. 0,2 ou 0,8) : peu d'association

"petit" (par ex. 0.05 ou 0.001) : beaucoup d'association

donc $\gamma = 1 - \alpha_c$: mesure d'assocation dont la distribution (asymptotique) ne dépend pas de r ou de s. En effet, sous $H_0 : R \perp\!\!\!\perp S$ on a : $P[\gamma \geq g_{\text{obs}}] = 1 - g_{\text{obs}}$

Exercice. Vérifiez que l'alpha-critique est le même si on le base sur la statistique de Pearson, U, ou une des transformations définies plus haut (C^2,Φ^2,T^2,V^2).

10.6.4 Approche prédictive

Soit deux critères de classification :

$$R \in \{1,\dots,r\}$$

$$S \in \{1,\dots,s\}$$

de distribution conjointe p_{ij}

Problème : en prenant un individu au hasard, on veut prédire son niveau du critère S connaissant son niveau du critère R

Si on suppose $R \perp\!\!\!\perp S$, on prédit S_{j*} tel que $p_{\cdot j_*} \geq p_{\cdot j} \ \forall\ j$ dans ce cas, l'erreur de prédiction, notée q_1, vaut $q_1 = 1 - \max_j p_{\cdot j}$

Si on suppose R et S non indépendants :
Pour $R = i$ donné, on prédit $S = j_*(i)$ tel que :

$$\frac{p_{ij_*(i)}}{p_{i\cdot}} \geq \frac{p_{ij}}{p_{i\cdot}} \ \forall\ j$$

avec une erreur de prédiction $1 - \dfrac{p_{ij_*(i)}}{p_{i\cdot}}$
L'erreur moyenne de prédiction est donc :

$$q_2 = \sum_i p_{i\cdot}[1 - \max_j \frac{p_{ij}}{p_{i\cdot}}] = 1 - \sum_i \max_j p_{ij}$$

On peut montrer

- $q_2 \leq q_1$
- $q_1 = q_2 \Leftrightarrow j_*(i) = j_* \quad \forall\, i$
  c'est-à-dire: Un seul niveau j_* du critère S pour avoir des probabilités marginales et conditionnelles maximum

Coefficient λ (Goodman et Krushal) :

$$\lambda_{S|R} = \frac{q_1 - q_2}{q_1} = 1 - \frac{q_2}{q_1}$$

Propriétés

- $0 \leq \lambda_{S|R} \leq 1$
- $R \perp\!\!\!\perp S \Rightarrow \lambda_{S|R} = 0$
  Mais l'inverse est faux
- $\lambda_{S|R} = 1 \Leftrightarrow \ \forall\, i$ il existe un seul $j(i)$ tel que $q_{i,j(i)} > 0$
  c'est-à-dire prévisibilité parfaite
- en général : $\lambda_{S|R} \neq \lambda_{R|S}$.

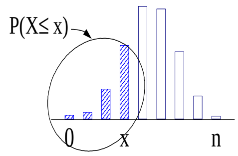
Annexe 1 : Alphabec grec

minuscule	majuscule	nom	équivalent français
α	A	*alpha*	*a*
β	B	*beta*	*b*
γ	Γ	*gamma*	*g*
δ	Δ	*delta*	*d*
ϵ	E	*epsilon*	*e*
ζ	Z	*zeta*	*z*
η	H	*eta*	*è*
θ	Θ	*theta*	*th*
ι	I	*iota*	*i*
κ	K	*kappa*	*k*
λ	Λ	*lambda*	*l*
μ	M	*mu*	*m*
ν	N	*nu*	*n*
ξ	Ξ	*xi*	*x*
o	O	*omicron*	*o*
π	Π	*pi*	*p*
ρ	P	*rho*	*r*
σ	Σ	*sigma*	*s*
τ	T	*tau*	*t*
υ	Υ	*upsilon*	*u*
ϕ	Φ	*phi*	*ph*
χ	X	*chi*	*kh*
ψ	Ψ	*psi*	*ps*
ω	Ω	*omega*	*o*

Annexe 2 : Tables statistiques

Table de la variable aléatoire Binomiale

Fournit la probabilité P(X≤x)
pour X ∼ Bi(n,p)



p		**0,1**	**0,2**	**0,25**	**0,3**	**0,4**	**0,5**	**0,6**	**0,7**	**0,75**	**0,8**	**0,9**
n=5	**x**											
	0	0,5905	0,3277	0,2373	0,1681	0,0778	0,0312	0,0102	0,0024	0,0010	0,0003	0,0000
	1	0,9185	0,7373	0,6328	0,5282	0,3370	0,1875	0,0870	0,0308	0,0156	0,0067	0,0005
	2	0,9914	0,9421	0,8965	0,8369	0,6826	0,5000	0,3174	0,1631	0,1035	0,0579	0,0086
	3	0,9995	0,9933	0,9844	0,9692	0,9130	0,8125	0,6630	0,4718	0,3672	0,2627	0,0815
	4	1,0000	0,9997	0,9990	0,9976	0,9898	0,9688	0,9222	0,8319	0,7627	0,6723	0,4095
	5	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
n=10	**x**											
	0	0,3487	0,1074	0,0563	0,0282	0,0060	0,0010	0,0001	0,0000	0,0000	0,0000	
	1	0,7361	0,3758	0,2440	0,1493	0,0464	0,0107	0,0017	0,0001	0,0000	0,0000	
	2	0,9298	0,6778	0,5256	0,3828	0,1673	0,0547	0,0123	0,0016	0,0004	0,0001	0,0000
	3	0,9872	0,8791	0,7759	0,6496	0,3823	0,1719	0,0548	0,0106	0,0035	0,0009	0,0000
	4	0,9984	0,9672	0,9219	0,8497	0,6331	0,3770	0,1662	0,0473	0,0197	0,0064	0,0001
	5	0,9999	0,9936	0,9803	0,9527	0,8338	0,6230	0,3669	0,1503	0,0781	0,0328	0,0016
	6	1,0000	0,9991	0,9965	0,9894	0,9452	0,8281	0,6177	0,3504	0,2241	0,1209	0,0128
	7	1,0000	0,9999	0,9996	0,9984	0,9877	0,9453	0,8327	0,6172	0,4744	0,3222	0,0702
	8		1,0000	1,0000	0,9999	0,9983	0,9893	0,9536	0,8507	0,7560	0,6242	0,2639
	9		1,0000	1,0000	1,0000	0,9999	0,9990	0,9940	0,9718	0,9437	0,8926	0,6513
	10				1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
n=15	**x**											
	0	0,2059	0,0352	0,0134	0,0047	0,0005	0,0000	0,0000				
	1	0,5490	0,1671	0,0802	0,0353	0,0052	0,0005	0,0000	0,0000			
	2	0,8159	0,3980	0,2361	0,1268	0,0271	0,0037	0,0003	0,0000	0,0000		
	3	0,9444	0,6482	0,4613	0,2969	0,0905	0,0176	0,0019	0,0001	0,0000	0,0000	
	4	0,9873	0,8358	0,6865	0,5155	0,2173	0,0592	0,0093	0,0007	0,0001	0,0000	
	5	0,9978	0,9389	0,8516	0,7216	0,4032	0,1509	0,0338	0,0037	0,0008	0,0001	
	6	0,9997	0,9819	0,9434	0,8689	0,6098	0,3036	0,0950	0,0152	0,0042	0,0008	0,0000
	7	1,0000	0,9958	0,9827	0,9500	0,7869	0,5000	0,2131	0,0500	0,0173	0,0042	0,0000
	8	1,0000	0,9992	0,9958	0,9848	0,9050	0,6964	0,3902	0,1311	0,0566	0,0181	0,0003
	9		0,9999	0,9992	0,9963	0,9662	0,8491	0,5968	0,2784	0,1484	0,0611	0,0022
	10		1,0000	0,9999	0,9993	0,9907	0,9408	0,7827	0,4845	0,3135	0,1642	0,0127
	11		1,0000	1,0000	0,9999	0,9981	0,9824	0,9095	0,7031	0,5387	0,3518	0,0556
	12			1,0000	1,0000	0,9997	0,9963	0,9729	0,8732	0,7639	0,6020	0,1841
	13				1,0000	1,0000	0,9995	0,9948	0,9647	0,9198	0,8329	0,4510
	14					1,0000	1,0000	0,9995	0,9953	0,9866	0,9648	0,7941
	15						1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

p		**0,1**	**0,2**	**0,25**	**0,3**	**0,4**	**0,5**	**0,6**	**0,7**	**0,75**	**0,8**	**0,9**
n=20	x											
	0	0,1216	0,0115	0,0032	0,0008	0,0000	0,0000					
	1	0,3917	0,0692	0,0243	0,0076	0,0005	0,0000					
	2	0,6769	0,2061	0,0913	0,0355	0,0036	0,0002	0,0000				
	3	0,8670	0,4114	0,2252	0,1071	0,0160	0,0013	0,0000				
	4	0,9568	0,6296	0,4148	0,2375	0,0510	0,0059	0,0003	0,0000			
	5	0,9887	0,8042	0,6172	0,4164	0,1256	0,0207	0,0016	0,0000	0,0000		
	6	0,9976	0,9133	0,7858	0,6080	0,2500	0,0577	0,0065	0,0003	0,0000	0,0000	
	7	0,9996	0,9679	0,8982	0,7723	0,4159	0,1316	0,0210	0,0013	0,0002	0,0000	
	8	0,9999	0,9900	0,9591	0,8867	0,5956	0,2517	0,0565	0,0051	0,0009	0,0001	
	9	1,0000	0,9974	0,9861	0,9520	0,7553	0,4119	0,1275	0,0171	0,0039	0,0006	0,0000
	10	1,0000	0,9994	0,9961	0,9829	0,8725	0,5881	0,2447	0,0480	0,0139	0,0026	0,0000
	11		0,9999	0,9991	0,9949	0,9435	0,7483	0,4044	0,1133	0,0409	0,0100	0,0001
	12		1,0000	0,9998	0,9987	0,9790	0,8684	0,5841	0,2277	0,1018	0,0321	0,0004
	13		1,0000	1,0000	0,9997	0,9935	0,9423	0,7500	0,3920	0,2142	0,0867	0,0024
	14			1,0000	1,0000	0,9984	0,9793	0,8744	0,5836	0,3828	0,1958	0,0113
	15				1,0000	0,9997	0,9941	0,9490	0,7625	0,5852	0,3704	0,0432
	16					1,0000	0,9987	0,9840	0,8929	0,7748	0,5886	0,1330
	17					1,0000	0,9998	0,9964	0,9645	0,9087	0,7939	0,3231
	18						1,0000	0,9995	0,9924	0,9757	0,9308	0,6083
	19						1,0000	1,0000	0,9992	0,9968	0,9885	0,8784
	20							1,0000	1,0000	1,0000	1,0000	1,0000
n=25	x											
	0	0,0718	0,0038	0,0008	0,0001	0,0000						
	1	0,2712	0,0274	0,0070	0,0016	0,0001	0,0000					
	2	0,5371	0,0982	0,0321	0,0090	0,0004	0,0000					
	3	0,7636	0,2340	0,0962	0,0332	0,0024	0,0001	0,0000				
	4	0,9020	0,4207	0,2137	0,0905	0,0095	0,0005	0,0000				
	5	0,9666	0,6167	0,3783	0,1935	0,0294	0,0020	0,0001				
	6	0,9905	0,7800	0,5611	0,3407	0,0736	0,0073	0,0003	0,0000			
	7	0,9977	0,8909	0,7265	0,5118	0,1536	0,0216	0,0012	0,0000			
	8	0,9995	0,9532	0,8506	0,6769	0,2735	0,0539	0,0043	0,0001	0,0000		
	9	0,9999	0,9827	0,9287	0,8106	0,4246	0,1148	0,0132	0,0005	0,0000	0,0000	
	10	1,0000	0,9944	0,9703	0,9022	0,5858	0,2122	0,0344	0,0018	0,0002	0,0000	
	11	1,0000	0,9985	0,9893	0,9558	0,7323	0,3450	0,0778	0,0060	0,0009	0,0001	
	12		0,9996	0,9966	0,9825	0,8462	0,5000	0,1538	0,0175	0,0034	0,0004	
	13		0,9999	0,9991	0,9940	0,9222	0,6550	0,2677	0,0442	0,0107	0,0015	0,0000
	14		1,0000	0,9998	0,9982	0,9656	0,7878	0,4142	0,0978	0,0297	0,0056	0,0000
	15		1,0000	1,0000	0,9995	0,9868	0,8852	0,5754	0,1894	0,0713	0,0173	0,0001
	16			1,0000	0,9999	0,9957	0,9461	0,7265	0,3231	0,1494	0,0468	0,0005
	17				1,0000	0,9988	0,9784	0,8464	0,4882	0,2735	0,1091	0,0023
	18				1,0000	0,9997	0,9927	0,9264	0,6593	0,4389	0,2200	0,0095
	19					0,9999	0,9980	0,9706	0,8065	0,6217	0,3833	0,0334
	20					1,0000	0,9995	0,9905	0,9095	0,7863	0,5793	0,0980
	21					1,0000	0,9999	0,9976	0,9668	0,9038	0,7660	0,2364
	22						1,0000	0,9996	0,9910	0,9679	0,9018	0,4629
	23						1,0000	0,9999	0,9984	0,9930	0,9726	0,7288
	24							1,0000	0,9999	0,9992	0,9962	0,9282
	25							1,0000	1,0000	1,0000	1,0000	1,0000

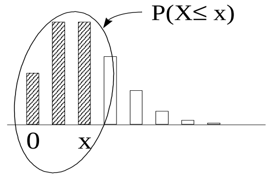
p		**0,1**	**0,2**	**0,25**	**0,3**	**0,4**	**0,5**	**0,6**	**0,7**	**0,75**	**0,8**	**0,9**
n=30	x											
	0	0,0424	0,0012	0,0002	0,0000							
	1	0,1837	0,0105	0,0020	0,0003	0,0000						
	2	0,4114	0,0442	0,0106	0,0021	0,0000						
	3	0,6474	0,1227	0,0374	0,0093	0,0003	0,0000					
	4	0,8245	0,2552	0,0979	0,0302	0,0015	0,0000					
	5	0,9268	0,4275	0,2026	0,0766	0,0057	0,0002					
	6	0,9742	0,6070	0,3481	0,1595	0,0172	0,0007	0,0000				
	7	0,9922	0,7608	0,5143	0,2814	0,0435	0,0026	0,0000				
	8	0,9980	0,8713	0,6736	0,4315	0,0940	0,0081	0,0002				
	9	0,9995	0,9389	0,8034	0,5888	0,1763	0,0214	0,0009	0,0000			
	10	0,9999	0,9744	0,8943	0,7304	0,2915	0,0494	0,0029	0,0000	0,0000		
	11	1,0000	0,9905	0,9493	0,8407	0,4311	0,1002	0,0083	0,0002	0,0000		
	12	1,0000	0,9969	0,9784	0,9155	0,5785	0,1808	0,0212	0,0006	0,0001	0,0000	
	13		0,9991	0,9918	0,9599	0,7145	0,2923	0,0481	0,0021	0,0002	0,0000	
	14		0,9998	0,9973	0,9831	0,8246	0,4278	0,0971	0,0064	0,0008	0,0001	
	15		0,9999	0,9992	0,9936	0,9029	0,5722	0,1754	0,0169	0,0027	0,0002	
	16		1,0000	0,9998	0,9979	0,9519	0,7077	0,2855	0,0401	0,0082	0,0009	
	17		1,0000	0,9999	0,9994	0,9788	0,8192	0,4215	0,0845	0,0216	0,0031	0,0000
	18			1,0000	0,9998	0,9917	0,8998	0,5689	0,1593	0,0507	0,0095	0,0000
	19			1,0000	1,0000	0,9971	0,9506	0,7085	0,2696	0,1057	0,0256	0,0001
	20				1,0000	0,9991	0,9786	0,8237	0,4112	0,1966	0,0611	0,0005
	21					0,9998	0,9919	0,9060	0,5685	0,3264	0,1287	0,0020
	22					1,0000	0,9974	0,9565	0,7186	0,4857	0,2392	0,0078
	23					1,0000	0,9993	0,9828	0,8405	0,6519	0,3930	0,0258
	24						0,9998	0,9943	0,9234	0,7974	0,5725	0,0732
	25						1,0000	0,9985	0,9698	0,9021	0,7448	0,1755
	26						1,0000	0,9997	0,9907	0,9626	0,8773	0,3526
	27							1,0000	0,9979	0,9894	0,9558	0,5886
	28							1,0000	0,9997	0,9980	0,9895	0,8163
	29								1,0000	0,9998	0,9988	0,9576
	30								1,0000	1,0000	1,0000	1,0000
n=40	x											
	0	0,0148	0,0001	0,0000	0,0000							
	1	0,0805	0,0015	0,0001	0,0000							
	2	0,2228	0,0079	0,0010	0,0001							
	3	0,4231	0,0285	0,0047	0,0006	0,0000						
	4	0,6290	0,0759	0,0160	0,0026	0,0000						
	5	0,7937	0,1613	0,0433	0,0086	0,0001						
	6	0,9005	0,2859	0,0962	0,0238	0,0006	0,0000					
	7	0,9581	0,4371	0,1820	0,0553	0,0021	0,0000					
	8	0,9845	0,5931	0,2998	0,1110	0,0061	0,0001					
	9	0,9949	0,7318	0,4395	0,1959	0,0156	0,0003					
	10	0,9985	0,8392	0,5839	0,3087	0,0352	0,0011	0,0000				
	11	0,9996	0,9125	0,7151	0,4406	0,0709	0,0032	0,0000				
	12	0,9999	0,9568	0,8209	0,5772	0,1285	0,0083	0,0001				
	13	1,0000	0,9806	0,8968	0,7032	0,2112	0,0192	0,0004				
	14	1,0000	0,9921	0,9456	0,8074	0,3174	0,0403	0,0012	0,0000			
	15		0,9971	0,9738	0,8849	0,4402	0,0769	0,0034	0,0000			

p		**0,1**	**0,2**	**0,25**	**0,3**	**0,4**	**0,5**	**0,6**	**0,7**	**0,75**	**0,8**	**0,9**
n=40	**x**											
	16		0,9990	0,9884	0,9367	0,5681	0,1341	0,0083	0,0001			
	17		0,9997	0,9953	0,9680	0,6885	0,2148	0,0189	0,0003	0,0000		
	18		0,9999	0,9983	0,9852	0,7911	0,3179	0,0392	0,0009	0,0000		
	19		1,0000	0,9994	0,9937	0,8702	0,4373	0,0744	0,0024	0,0002	0,0000	
	20		1,0000	0,9998	0,9976	0,9256	0,5627	0,1298	0,0063	0,0006	0,0000	
	21			1,0000	0,9991	0,9608	0,6821	0,2089	0,0148	0,0017	0,0001	
	22			1,0000	0,9997	0,9811	0,7852	0,3115	0,0320	0,0047	0,0003	
	23				0,9999	0,9917	0,8659	0,4319	0,0633	0,0116	0,0010	
	24				1,0000	0,9966	0,9231	0,5598	0,1151	0,0262	0,0029	
	25				1,0000	0,9988	0,9597	0,6826	0,1926	0,0544	0,0079	0,0000
	26					0,9996	0,9808	0,7888	0,2968	0,1032	0,0194	0,0000
	27					0,9999	0,9917	0,8715	0,4228	0,1791	0,0432	0,0001
	28					1,0000	0,9968	0,9291	0,5594	0,2849	0,0875	0,0004
	29					1,0000	0,9989	0,9648	0,6913	0,4161	0,1608	0,0015
	30						0,9997	0,9844	0,8041	0,5605	0,2682	0,0051
	31						0,9999	0,9939	0,8890	0,7002	0,4069	0,0155
	32						1,0000	0,9979	0,9447	0,8180	0,5629	0,0419
	33						1,0000	0,9994	0,9762	0,9038	0,7141	0,0995
	34							0,9999	0,9914	0,9567	0,8387	0,2063
	35							1,0000	0,9974	0,9840	0,9241	0,3710
	36							1,0000	0,9994	0,9953	0,9715	0,5769
	37								0,9999	0,9990	0,9921	0,7772
	38								1,0000	0,9999	0,9985	0,9195
	39								1,0000	1,0000	0,9999	0,9852
	40									1,0000	1,0000	1,0000
n=50	**x**											
	0	0,0052	0,0000	0,0000								
	1	0,0338	0,0002	0,0000								
	2	0,1117	0,0013	0,0001	0,0000							
	3	0,2503	0,0057	0,0005	0,0000							
	4	0,4312	0,0185	0,0021	0,0002							
	5	0,6161	0,0480	0,0070	0,0007	0,0000						
	6	0,7702	0,1034	0,0194	0,0025	0,0000						
	7	0,8779	0,1904	0,0453	0,0073	0,0001						
	8	0,9421	0,3073	0,0916	0,0183	0,0002						
	9	0,9755	0,4437	0,1637	0,0402	0,0008						
	10	0,9906	0,5836	0,2622	0,0789	0,0022	0,0000					
	11	0,9968	0,7107	0,3816	0,1390	0,0057	0,0000					
	12	0,9990	0,8139	0,5110	0,2229	0,0133	0,0002					
	13	0,9997	0,8894	0,6370	0,3279	0,0280	0,0005					
	14	0,9999	0,9393	0,7481	0,4468	0,0540	0,0013	0,0000				
	15	1,0000	0,9692	0,8369	0,5692	0,0955	0,0033	0,0000				
	16	1,0000	0,9856	0,9017	0,6839	0,1561	0,0077	0,0001				
	17		0,9937	0,9449	0,7822	0,2369	0,0164	0,0002				
	18		0,9975	0,9713	0,8594	0,3356	0,0325	0,0005				
	19		0,9991	0,9861	0,9152	0,4465	0,0595	0,0014				
	20		0,9997	0,9937	0,9522	0,5610	0,1013	0,0034	0,0000			

p		**0,1**	**0,2**	**0,25**	**0,3**	**0,4**	**0,5**	**0,6**	**0,7**	**0,75**	**0,8**	**0,9**
n=50	**x**											
	21		0,9999	0,9974	0,9749	0,6701	0,1611	0,0076	0,0000			
	22		1,0000	0,9990	0,9877	0,7660	0,2399	0,0160	0,0001			
	23		1,0000	0,9996	0,9944	0,8438	0,3359	0,0314	0,0003	0,0000		
	24			0,9999	0,9976	0,9022	0,4439	0,0573	0,0009	0,0000		
	25			1,0000	0,9991	0,9427	0,5561	0,0978	0,0024	0,0001		
	26			1,0000	0,9997	0,9686	0,6641	0,1562	0,0056	0,0004	0,0000	
	27				0,9999	0,9840	0,7601	0,2340	0,0123	0,0010	0,0000	
	28				1,0000	0,9924	0,8389	0,3299	0,0251	0,0026	0,0001	
	29				1,0000	0,9966	0,8987	0,4390	0,0478	0,0063	0,0003	
	30					0,9986	0,9405	0,5535	0,0848	0,0139	0,0009	
	31					0,9995	0,9675	0,6644	0,1406	0,0287	0,0025	
	32					0,9998	0,9836	0,7631	0,2178	0,0551	0,0063	
	33					0,9999	0,9923	0,8439	0,3161	0,0983	0,0144	0,0000
	34					1,0000	0,9967	0,9045	0,4308	0,1631	0,0308	0,0000
	35					1,0000	0,9987	0,9460	0,5532	0,2519	0,0607	0,0001
	36						0,9995	0,9720	0,6721	0,3630	0,1106	0,0003
	37						0,9998	0,9867	0,7771	0,4890	0,1861	0,0010
	38						1,0000	0,9943	0,8610	0,6184	0,2893	0,0032
	39						1,0000	0,9978	0,9211	0,7378	0,4164	0,0094
	40							0,9992	0,9598	0,8363	0,5563	0,0245
	41							0,9998	0,9817	0,9084	0,6927	0,0579
	42							0,9999	0,9927	0,9547	0,8096	0,1221
	43							1,0000	0,9975	0,9806	0,8966	0,2298
	44							1,0000	0,9993	0,9930	0,9520	0,3839
	45								0,9998	0,9979	0,9815	0,5688
	46								1,0000	0,9995	0,9943	0,7497
	47								1,0000	0,9999	0,9987	0,8883
	48									1,0000	0,9998	0,9662
	49									1,0000	1,0000	0,9948
	50										1,0000	1,0000

Table de la variable aléatoire de Poisson

Fournit la probabilité P(X≤x)
pour X ∼ Po(λ)

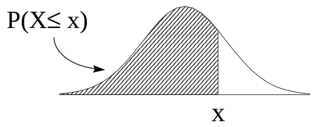


λ	**0,05**	**0,10**	**0,20**	**0,30**	**0,40**	**0,50**	**0,60**	**0,70**	**0,80**	**0,90**	**1,00**
x											
0	0,9512	0.9048	0.8187	0.7408	0.6703	0,6065	0.5488	0.4966	0.4493	0.4066	0,3679
1	0,9988	0.9953	0.9825	0.9631	0.9384	0,9098	0.8781	0.8442	0.8088	0.7725	0,7358
2	1,0000	0.9998	0.9989	0.9964	0.9921	0,9856	0.9769	0.9659	0.9526	0.9371	0,9197
3		1,0000	0.9999	0.9997	0.9992	0,9982	0.9966	0.9942	0.9909	0.9865	0,9810
4			1,0000	1,0000	0.9999	0,9998	0.9996	0.9992	0.9986	0.9977	0,9963
5					1,0000	1,0000	1,0000	0.9999	0.9998	0.9997	0,9994
6								1,0000	1,0000	1,0000	0,9999
7											1,0000

λ	**1,20**	**1,40**	**1,60**	**1,80**	**2**	**3**	**4**	**5**	**6**	**8**	**10**
x											
0	0.3012	0.2466	0.2019	0.1653	0,1353	0,0498	0,0183	0,0068	0,0025	0,0003	0,0000
1	0.6626	0.5918	0.5249	0.4628	0,4060	0,1992	0,0916	0,0404	0,0174	0,0030	0,0005
2	0.8795	0.8335	0.7834	0.7306	0,6767	0,4232	0,2381	0,1247	0,0620	0,0138	0,0028
3	0.9662	0.9463	0.9212	0.8913	0,8571	0,6472	0,4335	0,2650	0,1512	0,0424	0,0103
4	0.9923	0.9857	0.9763	0.9636	0,9473	0,8153	0,6288	0,4405	0,2851	0,0996	0,0293
5	0.9985	0.9968	0.9940	0.9896	0,9834	0,9161	0,7851	0,6160	0,4457	0,1912	0,0671
6	0.9997	0.9994	0.9987	0.9974	0,9955	0,9665	0,8893	0,7622	0,6063	0,3134	0,1301
7	1,0000	0.9999	0.9997	0.9994	0,9989	0,9881	0,9489	0,8666	0,7440	0,4530	0,2202
8	1,0000	1,0000	1,0000	0.9999	0,9998	0,9962	0,9786	0,9319	0,8472	0,5925	0,3328
9				1,0000	1,0000	0,9989	0,9919	0,9682	0,9161	0,7166	0,4579
10						0,9997	0,9972	0,9863	0,9574	0,8159	0,5830
11						0,9999	0,9991	0,9945	0,9799	0,8881	0,6968
12						1,0000	0,9997	0,9980	0,9912	0,9362	0,7916
13							0,9999	0,9993	0,9964	0,9658	0,8645
14							1,0000	0,9998	0,9986	0,9827	0,9165
15								0,9999	0,9995	0,9918	0,9513
16								1,0000	0,9998	0,9963	0,9730
17									0,9999	0,9984	0,9857
18									1,0000	0,9993	0,9928
19										0,9997	0,9965
20										0,9999	0,9984
21										1,0000	0,9993
22											0,9997
23											0,9999
24											1,0000

Table de la variable aléatoire Normale réduite

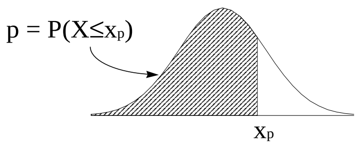
Fournit la probabilité P(X≤x)
pour X ∼ N(0,1)



x	**.00**	**.01**	**.02**	**.03**	**.04**	**.05**	**.06**	**.07**	**.08**	**.09**
0,0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0,1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0,2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0,3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0,4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0,5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0,6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0,7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0,8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0,9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Table des quantiles de la v. a. Normale réduite

Fournit les quantiles x_p tels que
$P(X \leq x_p) = p$
pour $X \sim N(0,1)$

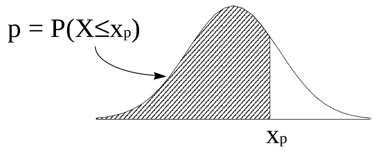


p	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00		-2,3263	-2,0537	-1,8808	-1,7507	-1,6449	-1,5548	-1,4758	-1,4051	-1,3408
0,10	-1,2816	-1,2265	-1,1750	-1,1264	-1,0803	-1,0364	-0,9945	-0,9542	-0,9154	-0,8779
0,20	-0,8416	-0,8064	-0,7722	-0,7388	-0,7063	-0,6745	-0,6433	-0,6128	-0,5828	-0,5534
0,30	-0,5244	-0,4959	-0,4677	-0,4399	-0,4125	-0,3853	-0,3585	-0,3319	-0,3055	-0,2793
0,40	-0,2533	-0,2275	-0,2019	-0,1764	-0,1510	-0,1257	-0,1004	-0,0753	-0,0502	-0,0251
0,50	0,0000	0.0251	0.0502	0.0752	0.1004	0.1257	0.1510	0.1764	0.2019	0.2275
0,60	0.2533	0.2793	0.3055	0.3319	0.3585	0.3853	0.4125	0.4399	0.4677	0.4959
0,70	0.5244	0.5534	0.5829	0.6128	0.6433	0.6745	0.7063	0.7388	0.7722	0.8064
0,80	0.8416	0.8779	0.9154	0.9542	0.9945	1,0364	1,0803	1,1264	1,1750	1,2265
0,90	1,2816	1,3408	1,4051	1,4758	1,5548	1,6449	1,7507	1,8808	2,0537	2,3263

p	0,9000	0,9500	0,9750	0,9900	0,9950	0,9975	0,9990	0,9995	0,9999
	1,2816	1,6449	1,9600	2,3263	2,5758	2,8070	3,0902	3,2905	3,7190

Table des quantiles de la v.a. de Student

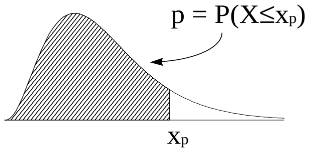
Fournit les quantiles x_p tels que
$P(X \leq x_p) = p$
pour $X \sim t_n$



n / p	**0,7500**	**0,9000**	**0,9500**	**0,9750**	**0,9900**	**0,9950**	**0,9975**	**0,9990**
n								
1	1,0000	3,0780	6,3140	12,7060	31,8210	63,6570	127,3213	318,3088
2	0,8160	1,8860	2,9200	4,3030	6,9650	9,9250	14,0891	22,3271
3	0,7650	1,6380	2,3530	3,1820	4,5410	5,8410	7,4533	10,2145
4	0,7410	1,5330	2,1320	2,7760	3,7470	4,6040	5,5976	7,1732
5	0,7270	1,4760	2,0150	2,5710	3,3650	4,0320	4,7733	5,8934
6	0,7180	1,4400	1,9430	2,4470	3,1430	3,7070	4,3168	5,2076
7	0,7110	1,4150	1,8950	2,3650	2,9980	3,4990	4,0293	4,7853
8	0,7060	1,3970	1,8600	2,3060	2,8960	3,3550	3,8325	4,5008
9	0,7030	1,3830	1,8330	2,2620	2,8210	3,2500	3,6897	4,2968
10	0,7000	1,3720	1,8120	2,2280	2,7640	3,1690	3,5814	4,1437
11	0,6970	1,3630	1,7960	2,2010	2,7180	3,1060	3,4966	4,0247
12	0,6950	1,3560	1,7820	2,1790	2,6810	3,0550	3,4284	3,9296
13	0,6940	1,3500	1,7710	2,1600	2,6500	3,0120	3,3725	3,8520
14	0,6920	1,3450	1,7610	2,1450	2,6240	2,9770	3,3257	3,7874
15	0,6910	1,3410	1,7530	2,1310	2,6020	2,9470	3,2860	3,7328
16	0,6900	1,3370	1,7460	2,1200	2,5830	2,9210	3,2520	3,6862
17	0,6890	1,3330	1,7400	2,1100	2,5670	2,8980	3,2225	3,6458
18	0,6880	1,3300	1,7340	2,1010	2,5520	2,8780	3,1966	3,6105
19	0,6880	1,3280	1,7290	2,0930	2,5390	2,8610	3,1737	3,5794
20	0,6870	1,3250	1,7250	2,0860	2,5280	2,8450	3,1534	3,5518
21	0,6860	1,3230	1,7210	2,0800	2,5180	2,8310	3,1352	3,5272
22	0,6860	1,3210	1,7170	2,0740	2,5080	2,8190	3,1188	3,5050
23	0,6850	1,3190	1,7140	2,0690	2,5000	2,8070	3,1040	3,4850
24	0,6850	1,3180	1,7110	2,0640	2,4920	2,7970	3,0905	3,4668
25	0,6840	1,3160	1,7080	2,0600	2,4850	2,7870	3,0782	3,4502
26	0,6840	1,3150	1,7060	2,0560	2,4790	2,7790	3,0669	3,4350
27	0,6840	1,3140	1,7030	2,0520	2,4730	2,7710	3,0565	3,4210
28	0,6830	1,3130	1,7010	2,0480	2,4670	2,7630	3,0469	3,4082
29	0,6830	1,3110	1,6990	2,0450	2,4620	2,7560	3,0380	3,3962
30	0,6830	1,3100	1,6970	2,0420	2,4570	2,7500	3,0298	3,3852
35	0,6820	1,3060	1,6900	2,0300	2,4380	2,7240	2,9960	3,3400
40	0,6810	1,3030	1,6840	2,0210	2,4230	2,7040	2,9712	3,3069
45	0,6800	1,3010	1,6790	2,0140	2,4120	2,6900	2,9521	3,2815
50	0,6790	1,2990	1,6760	2,0090	2,4030	2,6780	2,9370	3,2614
100	0,6770	1,2900	1,6600	1,9840	2,3640	2,6260	2,8713	3,1737
infinity	0,6745	1,2816	1,6449	1,9600	2,3263	2,5758	2,8070	3,0902

Table des quantiles de la v.a. Chi-Carré

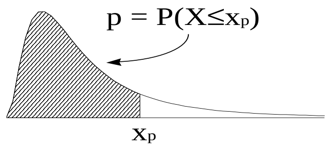
Fournit les quantiles x_p tels que
$P(X \leq x_p) = p$
pour $X \sim \chi^2_n$



n / p	**0,005**	**0,010**	**0,025**	**0,050**	**0,100**	**0,250**	**0,500**	**0,750**	**0,900**	**0,95**	**0,975**	**0,990**	**0,995**
n													
1	0,00	0,00	0,00	0,00	0,02	0,10	0,45	1,32	2,71	3,84	5,02	6,64	7,88
2	0,01	0,02	0,05	0,10	0,21	0,58	1,39	2,77	4,61	5,99	7,38	9,21	10,60
3	0,07	0,11	0,22	0,35	0,58	1,21	2,37	4,11	6,25	7,82	9,35	11,35	12,84
4	0,21	0,30	0,48	0,71	1,06	1,92	3,36	5,39	7,78	9,49	11,14	13,28	14,86
5	0,41	0,55	0,83	1,15	1,61	2,67	4,35	6,63	9,24	11,07	12,83	15,09	16,75
6	0,68	0,87	1,24	1,64	2,20	3,45	5,35	7,84	10,64	12,59	14,45	16,81	18,55
7	0,99	1,24	1,69	2,17	2,83	4,25	6,35	9,04	12,02	14,07	16,01	18,48	20,28
8	1,34	1,65	2,18	2,73	3,49	5,07	7,34	10,22	13,36	15,51	17,53	20,09	21,95
9	1,74	2,09	2,70	3,33	4,17	5,90	8,34	11,39	14,68	16,92	19,02	21,67	23,59
10	2,16	2,56	3,25	3,94	4,87	6,74	9,34	12,55	15,99	18,31	20,48	23,21	25,19
11	2,60	3,05	3,82	4,58	5,58	7,58	10,34	13,70	17,28	19,68	21,92	24,72	26,76
12	3,07	3,57	4,40	5,23	6,30	8,44	11,34	14,85	18,55	21,03	23,34	26,22	28,30
13	3,57	4,11	5,01	5,89	7,04	9,30	12,34	15,98	19,81	22,36	24,74	27,69	29,82
14	4,08	4,66	5,63	6,57	7,79	10,17	13,34	17,12	21,06	23,68	26,12	29,14	31,32
15	4,60	5,23	6,26	7,26	8,55	11,04	14,34	18,25	22,31	25,00	27,49	30,58	32,80
16	5,14	5,81	6,91	7,96	9,31	11,91	15,34	19,37	23,54	26,30	28,85	32,00	34,27
17	5,70	6,41	7,56	8,67	10,09	12,79	16,34	20,49	24,77	27,59	30,19	33,41	35,72
18	6,27	7,02	8,23	9,39	10,87	13,68	17,34	21,61	25,99	28,87	31,53	34,81	37,16
19	6,84	7,63	8,91	10,12	11,65	14,56	18,34	22,72	27,20	30,14	32,85	36,19	38,58
20	7,43	8,26	9,59	10,85	12,44	15,45	19,34	23,83	28,41	31,41	34,17	37,57	40,00
21	8,03	8,90	10,28	11,59	13,24	16,34	20,34	24,94	29,62	32,67	35,48	38,93	41,40
22	8,64	9,54	10,98	12,34	14,04	17,24	21,34	26,04	30,81	33,92	36,78	40,29	42,80
23	9,26	10,20	11,69	13,09	14,85	18,14	22,34	27,14	32,01	35,17	38,08	41,64	44,18
24	9,89	10,86	12,40	13,85	15,66	19,04	23,34	28,24	33,20	36,42	39,36	42,98	45,56
25	10,52	11,52	13,12	14,61	16,47	19,94	24,34	29,34	34,38	37,65	40,65	44,31	46,93
26	11,16	12,20	13,84	15,38	17,29	20,84	25,34	30,43	35,56	38,89	41,92	45,64	48,29
27	11,81	12,88	14,57	16,15	18,11	21,75	26,34	31,53	36,74	40,11	43,19	46,96	49,64
28	12,46	13,56	15,31	16,93	18,94	22,66	27,34	32,62	37,92	41,34	44,46	48,28	50,99
29	13,12	14,26	16,05	17,71	19,77	23,57	28,34	33,71	39,09	42,56	45,72	49,59	52,34
30	13,79	14,95	16,79	18,49	20,60	24,48	29,34	34,80	40,26	43,77	46,98	50,89	53,67
40	20,71	22,16	24,43	26,51	29,05	33,66	39,34	45,62	51,81	55,76	59,34	63,69	66,77
50	27,99	29,71	32,36	34,76	37,69	42,94	49,33	56,33	63,17	67,50	71,42	76,15	79,49
60	35,53	37,48	40,48	43,19	46,46	52,29	59,33	66,98	74,40	79,08	83,30	88,38	91,95
70	43,28	45,44	48,76	51,74	55,33	61,70	69,33	77,58	85,53	90,53	95,02	100,4	104,2
80	51,17	53,54	57,15	60,39	64,28	71,14	79,33	88,13	96,58	101,9	106,6	112,3	116,3
90	59,20	61,75	65,65	69,13	73,29	80,62	89,33	98,65	107,6	113,1	118,1	124,1	128,3
100	67,33	70,06	74,22	77,93	82,36	90,13	99,33	109,1	118,5	124,3	129,6	135,8	140,2

Table des quantiles de la v.a. de Fisher

Fournit les quantiles x_p tels que
$P(X \leq x_p) = p$
pour $X \sim F_{n1;n2}$



p=0.95

n1	1	2	3	4	5	6	7	8	9	10	15	20	30	50	100	inf
n2																
1	161	200	216	225	230	234	237	239	241	242	246	248	250	252	253	254
2	18,5	19,0	19,2	19,3	19,3	19,3	19,4	19,4	19,4	19,4	19,43	19,5	19,5	19,5	19,5	19,5
3	10,1	9,6	9,3	9,1	9,0	8,9	8,9	8,8	8,8	8,8	8,70	8,7	8,6	8,6	8,5	8,5
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,86	5,80	5,75	5,70	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,62	4,56	4,50	4,44	4,41	4,37
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	3,94	3,87	3,81	3,75	3,71	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,51	3,45	3,38	3,32	3,28	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,22	3,15	3,08	3,02	2,98	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,01	2,94	2,86	2,80	2,76	2,71
10	4,97	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,85	2,77	2,70	2,64	2,59	2,54
11	4,84	3,98	3,59	3,36	3,20	3,10	3,01	2,95	2,90	2,85	2,72	2,65	2,57	2,51	2,46	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,62	2,54	2,47	2,40	2,35	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,53	2,46	2,38	2,31	2,26	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,46	2,39	2,31	2,24	2,19	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,40	2,33	2,25	2,18	2,12	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,35	2,28	2,19	2,12	2,07	2,01
17	4,45	3,59	3,20	2,97	2,81	2,70	2,61	2,55	2,49	2,45	2,31	2,23	2,15	2,08	2,02	1,96
18	4,41	3,56	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,27	2,19	2,11	2,04	1,98	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,23	2,16	2,07	2,00	1,94	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,20	2,12	2,04	1,97	1,91	1,84
25	4,24	3,39	2,99	2,76	2,60	2,49	2,41	2,34	2,28	2,24	2,09	2,01	1,92	1,84	1,78	1,71
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,17	2,02	1,93	1,84	1,76	1,70	1,62
35	4,12	3,27	2,87	2,64	2,49	2,37	2,29	2,22	2,16	2,11	1,96	1,88	1,79	1,70	1,64	1,56
40	4,09	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	1,92	1,84	1,74	1,66	1,59	1,51
45	4,06	3,20	2,81	2,58	2,42	2,31	2,22	2,15	2,10	2,05	1,89	1,81	1,71	1,63	1,55	1,47
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,87	1,78	1,69	1,60	1,53	1,44
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,84	1,75	1,65	1,56	1,48	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97	1,81	1,72	1,62	1,53	1,45	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,79	1,70	1,60	1,51	1,43	1,32
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99	1,94	1,78	1,69	1,59	1,49	1,41	1,30
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,98	1,93	1,77	1,68	1,57	1,48	1,39	1,28
150	3,90	3,06	2,67	2,43	2,27	2,16	2,07	2,00	1,94	1,89	1,73	1,64	1,54	1,44	1,35	1,22
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,99	1,93	1,88	1,72	1,62	1,51	1,41	1,32	1,19
inf	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,67	1,57	1,46	1,35	1,24	1,00

p=0.99

n1 / n2	1	2	3	4	5	6	7	8	9	10	15	20	30	50	100	inf
1	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6157	6209	6261	6303	6334	6366
2	99	99	99	99	99	99	99	99	99	99	99,4	99,4	99,5	99,5	99,5	99,5
3	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,3	27,2	26,9	26,7	26,5	26,4	26,2	26,1
4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5	14,20	14,02	13,84	13,69	13,58	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,72	9,55	9,38	9,24	9,13	9,02
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,56	7,40	7,23	7,09	6,99	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,31	6,16	5,99	5,86	5,75	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,52	5,36	5,20	5,07	4,96	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	4,96	4,81	4,65	4,52	4,41	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,56	4,41	4,25	4,12	4,01	3,91
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,25	4,10	3,94	3,81	3,71	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,01	3,86	3,70	3,57	3,47	3,36
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	3,82	3,66	3,51	3,38	3,27	3,17
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,66	3,51	3,35	3,22	3,11	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,52	3,37	3,21	3,08	2,98	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,41	3,26	3,10	2,97	2,86	2,75
17	8,40	6,11	5,19	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,31	3,16	3,00	2,87	2,76	2,65
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,23	3,08	2,92	2,78	2,68	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,15	3,00	2,84	2,71	2,60	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,09	2,94	2,78	2,64	2,54	2,42
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	2,85	2,70	2,54	2,40	2,29	2,17
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,70	2,55	2,39	2,25	2,13	2,01
35	7,42	5,27	4,40	3,91	3,59	3,37	3,20	3,07	2,96	2,88	2,60	2,44	2,28	2,14	2,02	1,89
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,52	2,37	2,20	2,06	1,94	1,81
45	7,23	5,11	4,25	3,77	3,45	3,23	3,07	2,94	2,83	2,74	2,46	2,31	2,14	2,00	1,88	1,74
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,78	2,70	2,42	2,27	2,10	1,95	1,82	1,68
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,35	2,20	2,03	1,88	1,75	1,60
70	7,01	4,92	4,07	3,60	3,29	3,07	2,91	2,78	2,67	2,59	2,31	2,15	1,98	1,83	1,70	1,54
80	6,96	4,88	4,04	3,56	3,26	3,04	2,87	2,74	2,64	2,55	2,27	2,12	1,94	1,79	1,65	1,49
90	6,93	4,85	4,01	3,53	3,23	3,01	2,84	2,72	2,61	2,52	2,24	2,09	1,92	1,76	1,62	1,46
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50	2,22	2,07	1,89	1,74	1,60	1,43
150	6,81	4,75	3,91	3,45	3,14	2,92	2,76	2,63	2,53	2,44	2,16	2,00	1,83	1,66	1,52	1,33
200	6,76	4,71	3,88	3,41	3,11	2,89	2,73	2,60	2,50	2,41	2,13	1,97	1,79	1,63	1,48	1,28
inf	6,64	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,04	1,88	1,70	1,52	1,36	1,00