MA101 - ENSTA ParisTech 1ère Année

MA101
Deuxième partie

Bases de la statistique inférentielle

Christine Keribin
Université Paris Sud

9 janvier 2019

Chapitre 1

Introduction

La démarche statistique est utilisée dans tous les domaines : industriels, économiques, sciences du vivant, sciences de la nature, etc, pour traiter des données, et permettre de répondre à des questions aussi diverses que ces quelques exemples :

- quel est le résultat probable d'une élection ?
- le régime des pluies a-t-il changé en 100 ans ?
- quelle est la période de retour d'un événement extrême (pluie, crue, tempête) ?
- l'appareil de mesure récemment acheté est-il réellement plus précis que l'ancien ?
- un nouveau médicament est-il efficace ?
- la réussite des enfants dépend-elle de la csp de leurs parents ?
- quels sont les facteurs qui favorisent l'utilisation des moyens de transports modaux ?
- l'observation d'une averse permet-elle d'anticiper le débit d'une rivière en un point donné de l'aval ?
- Peut-on octroyer un crédit ?
- Retrouve-t-on un visage particulier dans une image ?

Les phénomènes générant les données sont en général complexes, partiellement connus ou observés. Ils peuvent faire preuve de variabilité : sous les mêmes conditions d'expérience, le résultat n'est pas forcément le même, d'où l'incertitude générée. Dans un milieu partiellement connu, il sera crucial de pouvoir calibrer le risque associé à la décision prise. Et quelles sont les limites ? : que peut-on faire dire (ou ne pas dire) à des jeux de données ?

Avec la place de plus en plus importante des "data" dans notre monde, et en particulier dans les processus de décision, il est important de définir un cadre d'étude des données. La statistique, en s'appuyant sur des outils mathématiques probabilistes et d'optimisation a développé ce cadre. Le mot statistique a d'ailleurs plusieurs sens :

- **ensemble de données** observées, voire résumées : on parlera de la statistique de l'emploi par exemple. Ces données pourront être observées sur la totalité de la population étudiée, ou sur une sous partie de cette population, appelé échantillon.
- **activité** qui consiste dans leur recueil, traitement et interprétation
- **discipline mathématique** qui fonde l'activité précédente

La statistique permet de définir un langage commun pour décrire les résultats d'expériences, et une méthodologie pour les traiter. C'est une branche des mathématiques appliquées qui a pris naissance au XIXème siècle, et qui se développe exponentiellement en ce moment grâce à l'essor de la puissance des ordinateurs et celui des moyens de stockage de données. Elle développe

des aspects théoriques (définitions, propriétés, théorèmes) qui fonde la démarche, des aspects méthodologiques pour mettre en œuvre les procédures et une part appliquée pour le traitement concret de jeux de donnés.

Dans ce cours sont posés les fondements de la statistique inférentielle : échantillonnage, modèle, estimateurs, intervalles de confiance et tests, dans un cadre essentiellement univarié et en général gaussien ou binomial. Il s'agit en quelque sorte d'apprendre les gammes qui permettront ensuite (en deuxième année) d'approfondir des questions théoriques comme l'optimalité des procédures utilisées, ou d'aborder des modélisations plus élaborées comme les modèles multivariés ou les séries d'observations irrégulières.

Objectifs pédagogiques A l'issue du cours, l'étudiant sera capable de
- Définir un modèle statistique paramétrique
- Proposer des estimateurs adaptés et en étudier les propriétés (biais, variance, consistance)
- Définir la loi d'une statistique (exacte ou asymptotique)
- Construire un intervalle de confiance d'un paramètre univarié
- Construire un test et savoir en interpréter les résultats.

Bien que l'utilisation de logiciel fasse partie intégrante de la science statistique, ce cours ne comprendra pas de partie pratique informatique. L'apprentissage de l'utilisation de logiciel (statistique) comme aide à la résolution se fera en deuxième année.

Prérequis Les pré-requis sont ceux de la première partie du cours MA101 (probabilités). En particulier, les notions de variables aléatoires discrètes et continues, convergence en probabilité et en loi, et les théorèmes limites ne doivent pas être inconnus, à défaut d'avoir été complètement acquis. Une connaissance en statistique descriptive est bienvenue, quelques rappels sont inclus dans un premier chapitre.

Bibliographie

- J.-F. Delmas. Introduction au calcul des probabilités et à la statistique. Les Presses de l'ENSTA, 2010.
- J-J. Daudin, S. Robin, C. Vuillet. Statistique inférentielle. Idées, démarches, exemples. Presses Universitaires de Rennes, Rennes, 2002.
- J. Pagès. Statistique générales pour utilisateurs. Presses Universitaires de Rennes, 2005.
- M. Lejeune. Statistique : la théorie et ses applications. Springer, 2011.

Chapitre 2

Statistique descriptive

On croit souvent que la statistique se résume à quelques éléments de **statistique descriptive** rencontrés au collège ou au lycée. Mais ce n'est qu'une toute petite partie de la discipline qu'il est cependant important de bien maîtriser pour aborder l'étude de données réelles. Elle permet comme son nom l'indique, de décrire des données, en général homogènes (faisant référence aux mêmes caractéristiques observées), soit de façon graphique, soit par l'intermédiaire d'indicateurs, résumés numériques de l'observation. Les outils diffèrent selon le type de variable observée qualitative, quantitative.

La statistique descriptive regroupe les études univariées (une variable) ou bi-variée (étude conjointe de deux variables). Quand le jeu de données comprend plus de trois variables, il n'est plus possible d'en faire une représentation graphique globale. L'analyse descriptive devient alors **analyse de données** et les **méthodes factorielles** prennent le relais pour décrire des plans d'observation permettant de capter la plus grande variabilité possible des observations et d'en dégager des facteurs sous sous-jacents. Cette question sera abordée dans la première partie du cours de deuxième année d'apprentissage statistique (MAE61).

Une étude de statistique descriptive est la première étape de la démarche statistique. Elle permet de se familiariser avec les données, faire le point sur les valeurs manquantes, avoir une première vision de la variabilité, proposer des indicateurs, structurer et résumer. Les données peuvent être présentées sous forme individuelle, une valeur de la variable observée par individu $(x_i)_{i=1,\dots,n}$: on enregistre la ville de résidence x_i de chaque individu i par exemple. Les données peuvent également être présentées sous forme groupée ou agrégée. La variable est alors le nombre d'individus (ou **effectif**) n_k pour chaque ville k : $(n_k)_{k=1,\dots,K}$ avec $\sum_{k=1}^{K} n_k = n$.

L'exemple suivant est un extrait du jeu de données étudié par P. Cortez et A. Morais [1] pour prédire le danger d'incendies

```
  X Y month day FFMC  DMC    DC  ISI temp RH wind rain area
1 7 5     3 fri 86.2 26.2  94.3  5.1  8.2 51  6.7  0.0    0
2 7 4    10 tue 90.6 35.4 669.1  6.7 18.0 33  0.9  0.0    0
3 7 4    10 sat 90.6 43.7 686.9  6.7 14.6 33  1.3  0.0    0
4 8 6     3 fri 91.7 33.3  77.5  9.0  8.3 97  4.0  0.2    0
5 8 6     3 sun 89.3 51.3 102.2  9.6 11.4 99  1.8  0.0    0
6 8 6     8 sun 92.3 85.3 488.0 14.7 22.2 29  5.4  0.0    0
```

1. **A Data Mining Approach to Predict Forest Fires using Meteorological Data**, Proceedings of the 13th EPIA 2007 pp. 512-523.

Ce sont des données individuelles. Chaque ligne est un incendie répertorié, pour lequel on a enregistré des variables

- **quantitatives** qui représentent des caractéristiques mesurables (température `temp`, pluviométrie `rain`, intensité du vent `wind`, par exemple).
- **qualitatives** qui représentent une propriété (qualité) du phénomène observé et prend ses valeurs dans un ensemble fixé discret (numérique ou non). On parle de variable **nominale** ou **catégorielle**. Les valeurs autorisées sont appelées **niveau**, ou **modalités** et la variable est un **facteur**. C'est le cas par exemple du jour `day` ou du mois `month`, même si ses niveaux sont codés numériquement.

Les variables `X` et `Y` sont numériques, elles représentent le numéro d'une zone en latitude et en longitude. La décision de les considérer comme quantitative ou qualitative pourra se faire au cours de l'étude : si on considère que la latitude et la longitude peuvent avoir une influence linéaire (par exemple) sur l'aire brûlée, on pourra avoir intérêt à les considérer comme quantitatives ; si ces zones correspondent à des régions, et qu'on souhaite les distinguer clairement sans lien les unes avec les autres, c'est l'option variable qualitative qui sera choisie. On le voit, décider du type de la variable n'est pas toujours si évident, et pourtant ce choix pourra avoir d'importantes conséquences sur l'optique prise pour traiter les données et donc sur les résultats de l'étude.

2.1 Variables qualitatives

Une variable qualitative ne peut prendre qu'un nombre fini de valeurs souvent codées de façon alphanumérique.

2.1.1 Indicateurs

Un résumé approprié est la **table de comptage**, qui répertorie le nombre d'individus par niveau du facteur. Par exemple,la variable `day` comporte 7 niveaux. Il y a 85 incendies qui ont eu lieu le vendredi `fri` :

```
mon tue wed thu fri sat sun
 74  64  54  61  85  84  95
```

Le comptage croisant les modalités de deux variables qualitatives s'appelle **table de contingence** : au mois de septembre (`9`), il y a eu 38 incendies répertoriés un vendredi :

```
      1  2  3  4  5  6  7  8  9 10 11 12
  mon 0  3 12  1  0  3  4 15 28  4  0  4
  tue 0  2  5  0  0  0  6 28 19  2  1  1
  wed 0  1  4  1  0  3  3 25 14  2  0  1
  thu 0  1  5  2  0  2  3 26 21  0  0  1
  fri 0  5 11  1  1  3  3 21 38  1  0  1
  sat 1  4 10  1  1  2  8 29 25  3  0  0
  sun 1  4  7  3  0  4  5 40 27  3  0  1
```

Les tables de comptage ou de contingence sont des données agrégées.

2.1.2 Graphiques

La visualisation des variables qualitatives se fait à partir des données agrégée, sous forme de diagrammes en barres pour lesquels la hauteur (ou longueur) de la barre est proportionnelle au nombre d'individus ayant ce niveau de facteur,

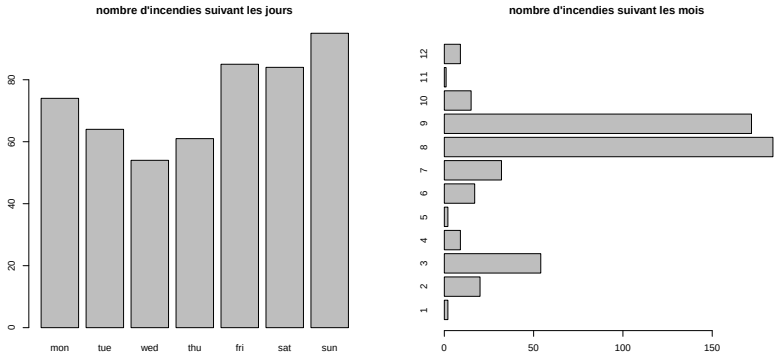


FIGURE 2.1 – Diagramme en barre sous forme verticale de la variable **day**, et sous forme horizontale de la variable **month**

ou sous forme de diagramme en barre des proportions (normalisation de la table de comptable par le nombre total d'individus du jeu de données)

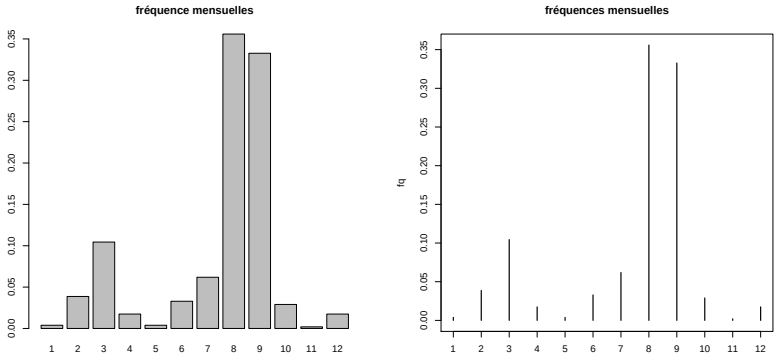


FIGURE 2.2 – Deux modes graphiques de diagrammes en barre des fréquences de la variable **month**

Le camembert est souvent utilisé, bien que cette représentation soit peu conseillée : en effet, l'œil conçoit plus difficilement le secteur angulaire que la longueur d'une part, et il est souvent difficile de distinguer des secteurs angulaires d'aires voisines d'autre part.

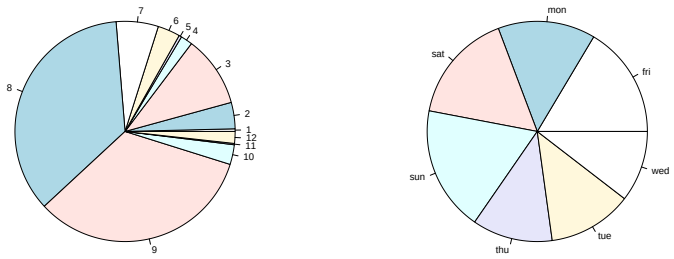


FIGURE 2.3 – Camemberts associés aux variables **day** et **month**

Deux variables qualitatives peuvent se représenter en empilant, pour chaque niveau de la première variable, les barres correspondants aux niveaux de la deuxième variable.

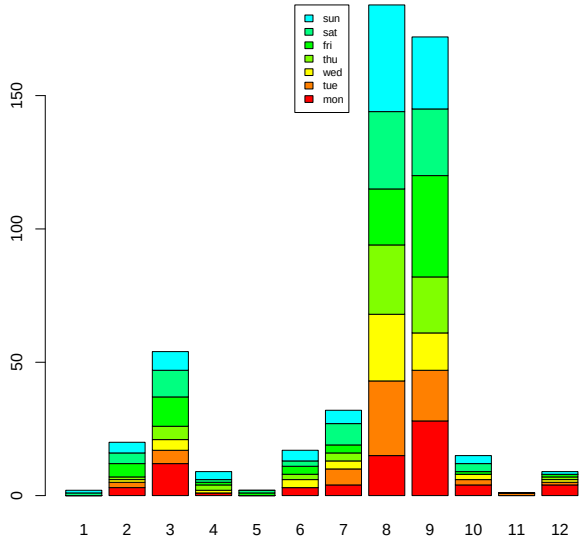


FIGURE 2.4 – Exemple de visualisation simultanée des variables **day** et **month**

2.2 Variables quantitatives

Une variable est quantitative quand elle est numérique et que son utilisation pour des calculs fait sens : une mesure physique, une valeur économique... On peut penser à une notion de grandeur continue, bien que de fait, on n'observe cette grandeur qu'avec un certain nombre de chiffres

significatifs. Elle peut parfois être discrète, ie prendre un (relativement) petit nombre de valeurs : nombre d'ouragans dans une année par exemple.

2.2.1 Indicateurs de tendance centrale

Certains indicateurs résument une notion de tendance centrale : quelle est la surface moyenne brûlée par incendie ? Quelle est la surface telle que la moitié des incendies ont eu une surface brûlée inférieure ?

Moyenne La **moyenne** (empirique) minimise la somme des carrés des distances à un point :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

On peut voir cet indicateur comme l'espérance de la loi qui charge chaque valeur x_i avec la probabilité $1/n$. Dans le cas où des groupes d'individus ont la même valeur, la **moyenne pondérée** sur les données groupées rend le calcul plus efficace :

$$\bar{x} = \frac{1}{n}\sum_{k=1}^{K} n_k x_k \text{ avec } \sum_k n_k = n$$

C'est l'espérance de la loi qui charge chaque valeur différente de la variable par la proportion d'individus qui l'ont.

Médiane La **médiane** minimise la somme des valeurs absolues des écarts à un point. Elle se calcule en commençant par trier les observations par valeur croissante $x_{(1)} \leq \ldots \leq x_{(n)}$. Si n est impair, la médiane est la valeur $x_{(\frac{n+1}{2})}$. Si n est pair, on définit par convention la médiane par la valeur $(x_{(\frac{n}{2})} + x_{(\frac{n+1}{2})})/2$.

Exemple : La médiane de `temp` vaut 19.3, presque égale à la moyenne 18.9, ce qui traduit une certaine symétrie des données pour cette variable. Mais la médiane peut sensiblement différer de la moyenne si la distribution n'est pas symétrique. La variable `area` a une moyenne de 12.84 et une médiane de 0.52.

La médiane est plus robuste à des valeurs extrêmes que la moyenne. Toutes les autres observations étant égales par ailleurs, si on change $x_{(n)} = 1$ par $x_{(n)} = 100$, on rajoute $99/n$ à la moyenne, mais la médiane ne change pas.

Mode Il s'agit de savoir si certaines zones concentre une grande part des observations. Quand la variable est discrète le mode est la valeur la plus fréquemment observée.

2.2.2 Indicateurs de dispersion

Il s'agit maintenant de comprendre comment les valeurs s'écartent de la tendance centrale.

Étendue C'est l'amplitude des valeurs, ie la distance $x_{(n)} - x_{(1)}$ entre la plus petite valeur $x_{(1)}$ et la plus grande $x_{(n)}$.

Variance C'est la moyenne de la somme des carrés des écarts à la moyenne :

$$V = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

On peut la voir comme la variance de la loi qui affecte à chaque valeur x_i la probabilité $1/n$. L'**écart-type** est la racine carrée de la variance $s = \sqrt{V}$. Plus la variance est grande, plus les valeurs sont dispersées autour de la moyenne.

Quantiles : le quantile q_α d'ordre α partage les observations triées par ordre croissant de telle sorte $n\alpha$ observations sont inférieures ou égales à q_α. En particulier, la médiane est le quantile d'ordre 0.5. D'autres quantiles remarquables sont :

- **premier quartile** Q_1 : valeur observable de la variable telle que 25% au moins des observations lui sont inférieures
- **troisième quartile** Q_3 : valeur observable de la variable telle que 75% au moins des observations lui sont inférieures

L'**intervalle inter quartiles** est la distance entre le premier et le troisième quartile : $IQ = Q_3 - Q_1$.

2.2.3 Graphiques

Boîte à moustaches Elle représente l'étendue des points allant de la plus petite valeur observée $x_{(1)}$ à la plus grande $x_{(n)}$. La boîte elle-même s'étend du premier quartile Q_1 au troisième quartile Q_3, et le trait central matérialise la médiane. Les moustaches peuvent atteindre les observations extrêmes $x_{(1)}$ ou $x_{(n)}$ si celles-ci ne sont pas trop éloignées des autres. Les observations de valeurs supérieures à $Q_3 + 1.5(Q_3 - Q_1)$ ou inférieures à $Q_1 - 1.5(Q_3 - Q_1)$ sont indiquées par des points isolés, marquant ainsi leur grand éloignement aux autres.

On remarque sur la figure 2.5 que la distribution de `aera` est très dissymétrique, avec la majeure partie des observations de faible surface brûlée, et des points extrêmes de valeur élevée. La transformation logarithme a permis d'atténuer cette dissymétrie. La distribution de `temp` est plus symétrique.

Histogramme : L'objectif de cette figure est de traduire la densité des observations suivant les plages de valeurs. Ainsi, la plage de valeurs étant partitionnée entre segments consécutif, un rectangle est tracé dont l'aire est proportionnelle au nombre des observations de la plage de valeurs considérée. Si les segments sont de longueur égale, la hauteur du rectangle est alors également proportionnelle au nombre des observations dans la plage observée. Dans le cas où les observations d'une variable quantitative sont suffisamment nombreuses, les segments sont contigus. C'est le même principe de représentation que celui du diagramme en barres pour les données qualitatives rencontré figure 2.2. Mais dans le cas de données discrètes, les segments sont non jointifs.

Le choix des coupures peut avoir une grande influence sur la représentation de l'histogramme. S'il n'y en a que deux, l'une au minimum, l'autre au maximum, il n'y a qu'un seul rectangle de hauteur la valeur moyenne. S'il y en a de telle sorte que toutes les observations sont séparées par une coupure, le graphique est une série de rectangles de même taille (quand il y a une observation) et de vide (quand il n'y en a pas).

Le nombre de coupures doit donc être choisi pour lisser suffisamment les irrégularités locales, sans trop occulter la tendance générale. C'est une représentation qui est sensible à la place des coupures.

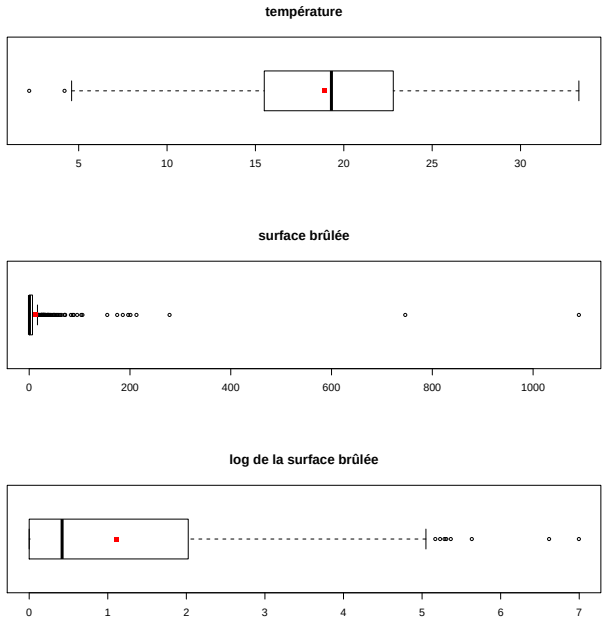


FIGURE 2.5 – Boîtes à moustaches de **temp**, **area** et **log(area)**. La valeur de la moyenne est indiquée par un carré rouge.

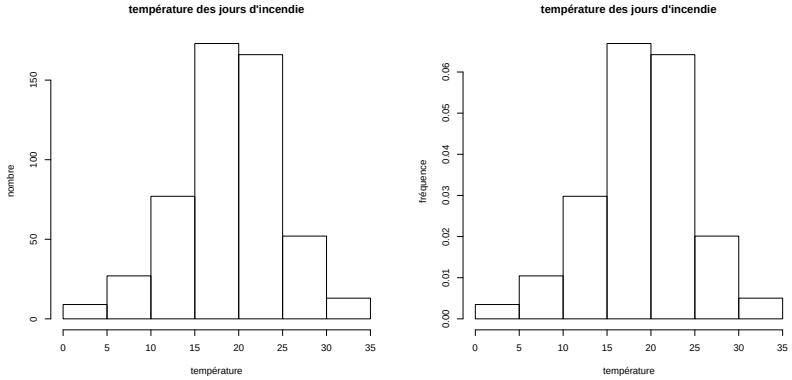


FIGURE 2.6 – Histogramme en comptage (gauche) et en proportion (droite) de la variable **temp**

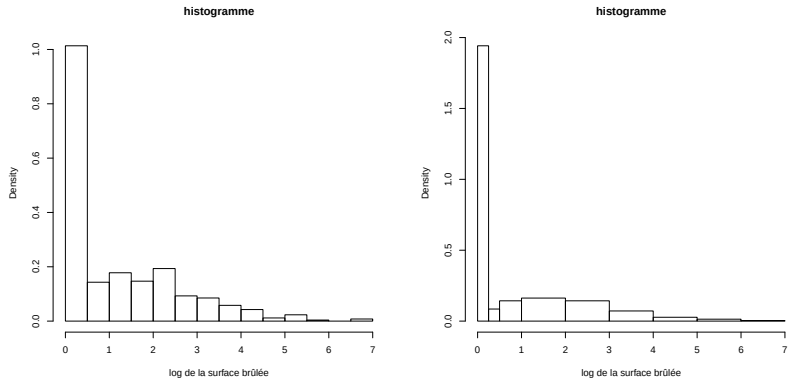


FIGURE 2.7 – Influence du nombre de coupures sur le tracé de l'histogramme de la surface brûlée.

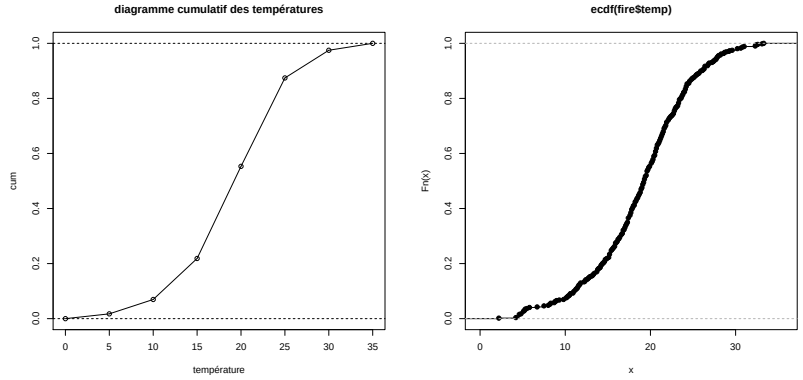


FIGURE 2.8 – Diagramme des fréquences cumulées pour 7 classes (gauche), pour une coupure à chaque observation (à droite)

Diagramme des fréquences cumulées A partir de l'effectif (ou de la proportion) de chaque classe, il est possible de calculer l'effectif (ou la proportion) cumulé(e). Par exemple, dans le cas de la température, en prenant 7 classes, on obtient les valeurs suivantes :

temp	0	5	10	15	20	25	30	35
fréquence		0.02	0.05	0.15	0.33	0.32	0.10	0.03
fréquence cumulée		0.02	0.07	0.22	0.55	0.87	0.97	1.00

ce qui permet de tracer le diagramme des fréquences cumulées. Si on considère des coupures à chaque observation (donc, à des intervalles non réguliers), le diagramme des fréquences cumulées représente la fonction de répartition empirique de la loi qui affecte à chaque observation une probabilité $1/n$.

2.3 Analyse bi-variée

Nous avons déjà vu comment traiter l'analyse conjointe de deux variables qualitatives. Quand les deux variables sont quantitatives, on trace le nuage de points 2D. Si le jeu de données a de

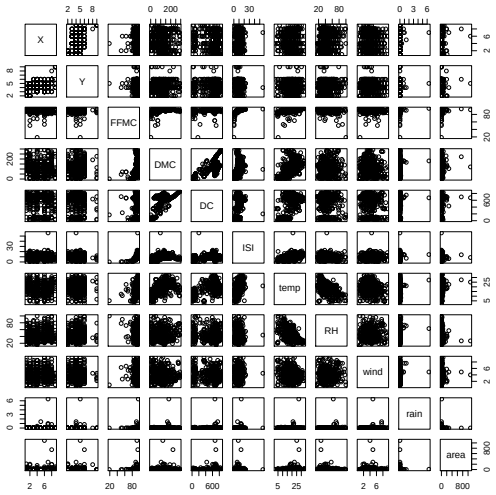


FIGURE 2.9 – scatter plot des variables quantitatives

nombreuses variables, il est possible de tracer un **scatter plot**, nuages de points des variables prises deux à deux (cf figure 2.9).

Les indicateurs utilisés avec deux variables quantitatives sont la **covariance** :

$$Cov(x,y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}$$

et la **corrélation**, covariance normalisée par l'écart-type de chaque variable :

$$Cor(x,y) = \frac{Cov(x,y)}{\sqrt{V(x)V(y)}}.$$

La corrélation mesure la force de la liaison **linéaire** entre deux variables. Si $Cor(x,y) = 1$, on a $x = ya + b$. Mais $Cor(x,y) = 0$ ne veut pas dire qu'il n'y pas de relation entre les deux variables, voir quelques exemples sur la figure 2.10.

Si l'une des variables est qualitative et l'autre quantitative, on peut tracer pour chaque niveau de la variable qualitative mise en abscisse, les valeurs des observations correspondantes de la variable quantitative ou tracer les boîtes à moustache en prenant soin de les tracer dans une échelle commune, figure 2.11 :

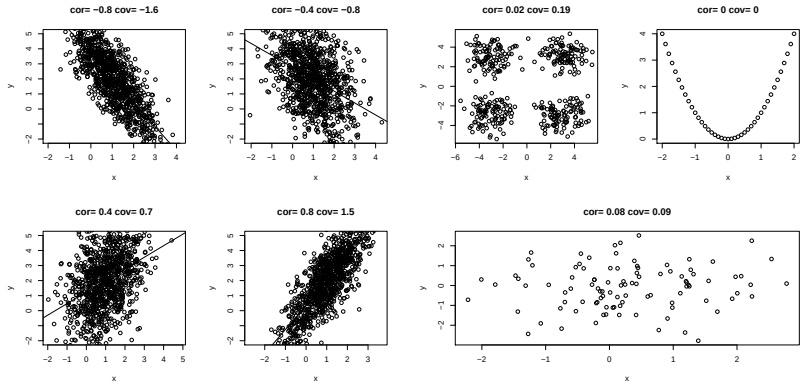


FIGURE 2.10 – Exemple de données corrélées (à gauche) et non corrélées (à droite)

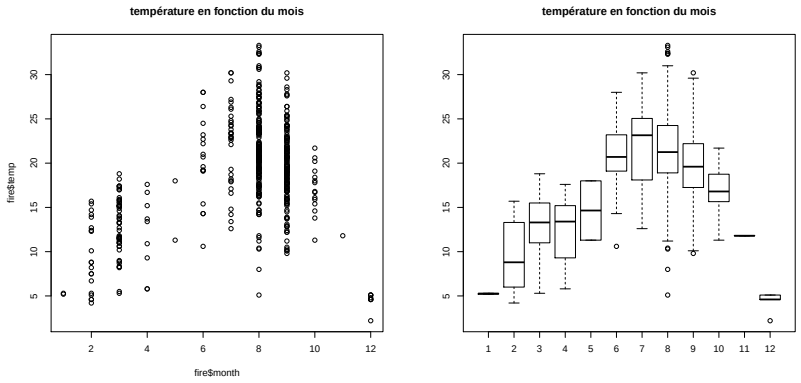


FIGURE 2.11 – Exemple de visualisation pour l'étude conjointe d'une variable qualitative et une variable quantitative.

Chapitre 3

Échantillonnage

On peut distinguer deux sources de variabilité dans les phénomènes observés : d'une part, l'aléatoire peut être provoqué par un mécanisme de tirage d'individus dans une population, la quantité observée sur l'individu tiré étant connue sans erreur ; d'autre part, l'aléatoire peut provenir du manque de précision de la mesure faite sur l'observation. Ces deux approches peuvent d'ailleurs se combiner. La procédure d'échantillonnage est une méthode courante d'acquisition d'information. Elle consiste à tirer des individus dans une population qui peut être finie ou infinie.

3.1 Population

Définition 3.1. *Une **population** $\mathcal{P}$ est un ensemble d'**individus** définis par des **caractéristiques**.*

L'individu doit être pris au sens général d'**unité statistique** : ce peut être une ville (sur laquelle on observe le taux d'ensoleillement annuel par exemple), un texte (dans lequel on compte le nombre de mots), une arête dans un graphe. Les caractéristiques des individus sont appelées **variables**. Nous avons vu qu'elles peuvent être

— **quantitative** (valeur numérique sur laquelle les calculs ont un sens) : discrète ou continue
— **qualitative** (attribut ou modalité) : nominale ou ordinale

Si la population est finie de **taille** N, on peut en faire une étude exhaustive ou **recensement**. Mais celui-ci peut s'avérer impossible, soit parce que la population est trop grande et le recensement demanderait trop de temps, soit parce que l'obtention de la caractéristique nécessite la destruction de l'objet : pour savoir si un lot de pièces mécaniques est acceptable, il est nécessaire de procéder à des études de fatigue qui amènent à la rupture de la pièce.

De même, il est évident qu'il n'est pas possible de faire le recensement d'une population infinie, qu'on étudiera par échantillonnage.

3.2 Échantillonnage dans une population finie

Définition 3.2. *On appelle **base de sondage** la liste des unités statistiques de la population. L'**échantillonnage** est le processus de tirage de n individus dans la population $\mathcal{P}$. Un **échantillon** est sous-ensemble d'unités de la population, résultat d'un processus d'échantillonnage.*

On notera n la taille de l'échantillon (son nombre d'individus). Quand la population est finie de taille N, le **taux de sondage** $f = n/N$ est la fraction des individus tirés. L'échantillon ne contient qu'une information partielle de la population, qui induit une variabilité d'un échantillon à l'autre, pourtant tiré à partir de la même population, et une différence entre les quantités calculées sur la population et sur l'échantillon. Ce sont ces différences et cette variabilité qu'il est important de savoir quantifier

Soit y une variable mesurable sur chaque unité u_ℓ, $\ell = 1, \ldots, N$ d'une population $\mathcal{P}$. La moyenne μ de la population totale vaut

$$\mu = \frac{1}{N} \sum_{\ell=1}^{N} y_\ell$$

où y_ℓ est la valeur de y sur l'unité u_ℓ. La moyenne de la population est l'espérance de la loi équiprobable sur $\{y_1, \ldots, y_N\}$. La moyenne observée sur un échantillon d'indices $(i_1, \ldots, i_n)$ est

$$\bar{Y} = \frac{1}{n} \sum_{j=1}^{n} y_{i_j}.$$

Comment choisir l'échantillon (ie, les indices $i, \ldots, n$) ? Il y a plusieurs techniques

3.2.1 Échantillonnage aléatoire simple

C'est une méthode assurant que chaque échantillon a la même probabilité d'être sélectionné. L'échantillon E est constitué de n individus tirés sans remise parmi les N unités de la population $\mathcal{P}$. Il y a donc $\binom{N}{n}$ échantillons et

$$\mathbb{P}(u_j \in E) = \frac{\begin{pmatrix} N-1 \\ n-1 \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} = \frac{n}{N}$$

La moyenne de l'échantillon $\bar{Y} = \sum_{j \in E} u_j/N$ est aléatoire parce que les indices j des unités de l'échantillon le sont. On a (cf TD) :

$$\mathbb{E}(\bar{Y}) = \mu$$

$$\mathrm{var}(\bar{Y}) = \frac{\sigma^{*2}}{n}(1-f) \text{ avec } \sigma^{*2} = \frac{1}{N-1} \sum_{\ell=1}^{N} (y_\ell - \mu)^2$$

Quand la taille de l'échantillon est multipliée par m, le coût de traitement est multiplié par m, mais l'erreur ou précision d'échantillonnage mesurée par l'écart type de $\bar{Y}$ n'est divisée que par un facteur $\sqrt{m}$.

La variance diminue avec n. Quand on a sélectionné toute la population sans remise ($n = N$), la variance est nulle, ce qu'on retrouve grâce au terme correctif impliquant le taux de sondage.

3.2.2 Échantillonnage avec remise

Si l'échantillonnage est effectué **avec remise**, il y a N^n échantillons possibles, certaines unités peuvent être présentes à plusieurs reprises dans un échantillon. Le tirage de chaque unité de l'échantillon suit la même loi, et se fait de façon indépendante au tirage de l'unité précédente. On a toujours $\mathbb{E}(\bar{Y}) = \mu$, mais $\mathrm{var}(\bar{Y}) = \frac{\sigma^2}{n}$. La variance diminue également avec n, mais ne

s'annule plus quand $N = n$ puisqu'on n'a plus le terme correctif. Il faudra une infinité de tirages pour arriver à une variance nulle. C'est le cas par exemple lorsqu'on tire avec remise des boules d'une urne.

3.2.3 Échantillonnage aléatoire stratifié

Si on sait que la population n'est pas homogène, et qu'on en connaît a priori une partition, on dit que la population est stratifiée. Les sous-groupes connus sont appelés des **strates**.

Le principe de l'échantillonnage stratifié est de calculer la moyenne d'un échantillon de chaque strate, puis de reconstituer la moyenne générale à partir de celles des strates.

Il y a plusieurs stratégies pour définir la taille des échantillons de chaque strate, voir le TD pour plus de précision.

3.3 Échantillonnage dans une population infinie

Quand la population est infinie, le taux de sondage est nul. Le tirage de chaque unité suit la même loi, et de façon indépendante au tirage de l'unité précédente. On se retrouve donc comme dans un cas analogue à celui du tirage avec remise dans une population finie et

$$\mathbb{E}(\bar{Y}) = \mu; \;\; \mathrm{var}(\bar{Y}) = \frac{\sigma^2}{n}.$$

Le cas de la population infinie peut s'appliquer par exemple à la mesure d'une caractéristique physique. Supposons qu'on souhaite déterminer le poids d'un objet. A cause de l'incertitude de mesure, l'appareil peut rendre toutes les valeurs possibles dans une plage donnée (potentiellement elle aussi infinie). La grandeur étant continue, c'est donc une infinité de valeurs possibles. En pratique, la mesure est faite avec un certain nombre de chiffres significatifs, donc le nombre de valeurs possible n'est pas infini, mais très grand. Étant très grand, on négligera le taux de sondage.

Chapitre 4

Estimation paramétrique ponctuelle

Etudier une population à partir d'un échantillon induit de la variabilité, que cette population soit finie ou infinie. Par ailleurs, l'incertitude associée à un phénomène qu'on ne sait pas complètement décrire induit aussi de la variabilité. La modélisation probabiliste permet de poser une loi de probabilité associée au phénomène étudié.

4.1 Statistique inférentielle

Considérons le cas d'une population finie constituée de 5 boules noires et 8 boules rouges. Le tirage au hasard d'une boule va donner la couleur rouge avec probabilité 8/13 et noire avec probabilité 5/13. L'événement "la boule est rouge" est modélisé par une variable aléatoire X de loi de Bernoulli $\mathcal{B}(1, \theta = 8/13)$. Supposons maintenant qu'on tire 5 boules avec un tirage avec remise et que les valeurs observées soient $1, 0, 1, 0, 0$.

En **probabilité**, on étudie les propriétés de cette modélisation, en supposant connue la proportion θ de boules rouges dans la population. On va par exemple déterminer la probabilité d'avoir obtenu le tirage indiqué en fonction de θ connu, ou, plus généralement, la loi de l'échantillon $X_1, \ldots, X_n$ et ses propriétés (moyenne, variance, etc...). On étudie donc les propriétés d'une loi dont la définition est connue.

Le point de vue est différent en **statistique**. Supposons que les boules soient dans une urne opaque et qu'on ne connait pas la proportion p de boules rouges. Peut-on déduire des couleurs des 5 boules une information sur la proportion θ qui nous est inconnue ? Une idée est de dire que la proportion θ inconnue ne doit pas être "loin" de la proportion observée $\widehat{\theta}$. C'est vrai sur cet exemple, la proportion observée sur l'échantillon est $\widehat{\theta} = 3/5$, très proche de 8/13. Mais nous l'avons vu, un échantillon n'est qu'une sous partie de la population. Il pourrait être possible d'avoir tiré un échantillon avec 5 boules noires (et la proportion observée est 0, mais de façon très rare), ou échantillon avec 5 boules rouges (et la proportion observée est 1, un peu moins rarement). Cela a-t-il un sens de prendre une décision en environnement incertain ? Et dans ce cas, peut-on quantifier le risque de se tromper, et d'ailleurs comment le définir ?

Si on s'intéresse à la précision de l'estimation, à n fixé, on peut utiliser l'inégalité de Tchebychev (cf A.4) et écrire que, pour tout $\varepsilon > 0$

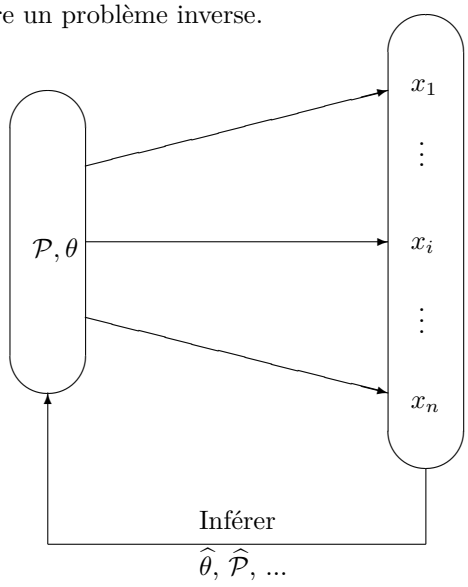
$$\mathbb{P}\left(|\widehat{\theta} - \theta| \geq \varepsilon\right) \leq \frac{\mathrm{var}(X_i)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$$

en utilisant le fait que la variance d'une loi de Bernoulli est $\theta(1-\theta)$, majorée uniformément par 4. Le choix de $\varepsilon = 1/(2\sqrt{n\alpha})$ conduit à

$$\mathbb{P}\left(|\widehat{\theta} - \theta| \leq \frac{1}{2\sqrt{n\alpha}}\right) \geq 1 - \alpha$$

On dit que $\left[\widehat{\theta} - \frac{1}{2\sqrt{n\alpha}}; \widehat{\theta} + \frac{1}{2\sqrt{n\alpha}}\right]$ est un intervalle de confiance pour θ de niveau de confiance $1-\alpha$. Pour $\alpha = 5\%$ et $n = 5$, la demi longueur de l'intervalle est égale à 1 ! En prenant $n = 100$ observations, celle-ci vaut encore 0.23. La majoration grossière obtenue ici n'est pas très précise. Il y a d'autres façons de contrôler $\mathbb{P}\left(|\widehat{\theta} - \theta| > \varepsilon\right)$.

La problématique précédente s'étend au tirage sans remise, et au cas d'une population infinie prenant des valeurs discrètes ou continues. La statistique propose un cadre théorique et méthodologique à ces questions : à partir d'un ensemble d'**observations** d'une loi inconnue, la statistique mathématique permet d'**inférer** les propriétés de cette loi, c'est à dire d'étendre les propriétés vues sur l'échantillon à celles de la loi. On parle ainsi de **statistique inférentielle**. Il s'agit en quelque sorte de résoudre un problème inverse.



L'**estimation** de la valeur des paramètres de la loi ayant généré les observations n'est pas le seul objectif de la statistique inférentielle. Son but est également de pouvoir **tester** des hypothèses (peut-on valider le fait que la proportion de boules rouges est 5/3 ? ou y a-t-il eu un truquage au remplissage de l'urne ?), comparer des échantillons, **prédire** la caractéristique d'une nouvelle unité non encore observée ou **classer** des individus.

Les applications sont très nombreuses, et dépassent largement l'exemple académique précédent. La statistique inférentielle a de larges applications dans un grand nombre de domaines très divers : sondage électoral, fiabilité, test d'efficacité d'un nouveau médicament, détection de

courriers indésirables, prévision de la consommation électrique, dimensionnement d'un standard téléphonique, classification automatique d'images... Dans ce cours, nous aborderons la compréhension de la problématique de la statistique inférentielle et de ses fondements. Il s'agit d'appréhender la définition du modèle probabiliste et d'une structure d'échantillonnage, le choix d'objets mathématiques permettant l'inférence (estimateurs, tests), et l'appréciation de la fiabilité de l'information obtenue.

4.1.1 Modèle statistique

La modélisation probabiliste est à la base de toute inférence statistique. Modéliser l'expérience, c'est proposer une loi théorique pour l'échantillon $X = (X_1, \dots, X_n)$.

Définition 4.1. *Un **modèle statistique** est la donnée d'un espace $\mathcal{X}^n$ mesuré par une tribu $\mathcal{A}^n$ et une famille de lois de probabilité $(\mathcal{P}_\theta^n)_{\theta\in\Theta}$. Le modèle associé est l'espace probabilisé noté*

$$\mathcal{M} = (\mathcal{X}^n, \mathcal{A}^n, \mathcal{P}_\theta^n, \theta \in \Theta)$$

*Quand il existe $d \in \mathbb{N}^*$ tel que $\Theta \subset \mathbb{R}^d$, le modèle est dit **paramétrique**. Sinon, il est **non paramétrique**.*

Dans le cas paramétrique, la forme de la loi n'est connue qu'à la valeur du paramètre θ près, à inférer avec l'observation de l'échantillon. Dans le cas non paramétrique, la loi est considérée comme un élément d'un espace de dimension infinie, et il s'agira d'estimer ses coordonnées dans cette base. Nous considérerons uniquement le cas paramétrique dans ce cours.

Définition 4.2. *Un **échantillon** de loi $\mathcal{P}^n$ est un ensemble de n variables aléatoires $X_1, \dots, X_n$ suivant la loi $\mathcal{P}^n$.*

Dans le cas d'un tirage indépendant dans une population infinie, ou le cas d'un échantillonnage avec remise dans une population finie, la loi de l'échantillon se factorise en produit des lois de chacune des composantes de l'échantillon.

Définition 4.3. *On appelle **n-échantillon i.i.d.** le modèle statistique de n composantes **i**ndépendantes et de même loi $\mathcal{P}_\theta$ (**i**dentiquement **d**istribuées)*

$$\mathcal{M} = (\mathcal{X}^n, \mathcal{A}^n, \mathcal{P}_\theta^{\otimes n}, \theta \in \Theta).$$

*La loi $\mathcal{P}_\theta$ est parfois appelée **loi mère** de l'échantillon. On dit que l'échantillon est de **taille** n.*

Dans ce cours, tous les échantillons seront considérés comme iid, et on les appellera simplement n-échantillons.

Définition 4.4. *Une **observation** est une composante de l'échantillon, variable aléatoire X_i de loi $\mathcal{P}_\theta$. Les **données** sont les réalisations (valeurs) $x_1, \dots, x_n$ prises par l'échantillon $X_1, \dots, X_n$.*

Par abus de langage, on appelle parfois observation le résultat $X_i(\omega) = x_i$ de cette variable aléatoire.

Exemples Le choix de la loi mère dépend du phénomène observé : si le phénomène est binaire, la loi de Bernoulli $\mathcal{B}(1, \theta)$, $\theta \in]0; 1[$ est un choix naturel, comme dans l'exemple introductif :

$$\mathbb{P}(X = 1) = \theta; \;\; \mathbb{P}(X = 0) = 1 - \theta.$$

Si on étudie une variable quantitative, la loi gaussienne $\mathcal{N}(\mu, \sigma^2)$, de densité

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}; \quad x \in \mathbb{R}$$

sera souvent utilisée. Le paramètre est $\theta = (\mu, \sigma^2)$.

La loi exponentielle d'espérance μ modélise les temps de survie

$$f(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}}; \quad x \in \mathbb{R}^+,$$

tandis que la loi de Poisson $\mathcal{P}(\mu)$

$$\mathbb{P}(X = k) = e^{-\mu} \frac{\mu^k}{k!}; \quad k \in \mathbb{N}$$

est utilisée dans les comptages associés à des événements rares, par exemple pour le nombre d'appels arrivant à un standard en une minute.

4.1.2 Démarche statistique

Des modèles plus développés seront abordés en deuxième année. Mais le principe de la démarche statistique sera le même. A partir des données d'un n-échantillon, déduire -ou inférer- certaines propriétés du modèle inconnu :

- acquérir et **préparer** les données
- définir un **modèle** adapté à la situation observée.
- **estimer** les paramètres du modèle grâce aux observations.
- vérifier l'**adéquation** de l'estimation aux observations.
- **proposer** d'autres modèles et **choisir** le plus adapté à un objectif donné (interprétation, prédiction)
- **utiliser** et **décider** !

tout en se rappelant de la maxime de G. Box [1] : *Tous les modèles sont faux, mais certains sont plus utiles que d'autres.*

Convention. Autant que faire ce peut, les paramètres seront notés avec des lettres grecques, les variables aléatoires par des lettres capitales et leur résultat par des lettres minuscules.

4.2 Estimateur

La statistique descriptive résume la réalisation d'une variable par quelques indicateurs.

Définition 4.5. *Une **statistique** T_n est variable aléatoire, fonction réelle ou vectorielle mesurable de l'échantillon $X = (X_1, \ldots, X_n)$. On peut la noter :*

$$T_n = t(X) = t(X_1, \ldots, X_n)$$

Une statistique est entièrement calculable à partir des données. Par exemple, si X est un n-échantillon de loi mère gaussienne $\mathcal{N}(\mu, \sigma^2)$, $\bar{X} = \sum_i X_i/n$ ou $S_n^2 = \frac{1}{n}\sum_i (X_i - \bar{X})^2$ sont des statistiques. En revanche, $\bar{X} - \mu$ n'en est pas une si μ est inconnue.

Une statistique ayant pour but d'inférer un paramètre est appelé estimateur.

1. George Box (1919-2013), statisticien britannique ayant apporté d'importantes contributions aux domaines du contrôle qualité, des séries temporelles et de l'inférence bayésienne

Définition 4.6. *Soit θ le paramètre d'une loi $\mathcal{P}_\theta$, $\theta \in \Theta$. Un **estimateur** de θ est une **statistique** à valeurs dans Θ. Elle est souvent notée $\widehat{\theta}$ ou $\widehat{\theta}_n$. Cette définition s'étend au cas d'une grandeur $\nu(\theta)$ calculée à partir de la loi $\mathcal{P}_\theta$: un estimateur $\widehat{\nu}_n$ de $\nu(\theta)$ est une statistique à valeurs dans $\nu(\Theta)$.*

La notation $\widehat{\theta}_n$ permet de rappeler que la statistique dépend de n, la taille de l'échantillon. Si on utilise la notation "chapeau", on ne fait pas en général de différence entre la notation de l'estimateur $\widehat{\theta} = t(X)$ ou $\widehat{\theta}_n = t(X)$ et celle de sa réalisation, notée à nouveau $\widehat{\theta} = t_n = t(x)$ ou $\widehat{\theta}_n = t_n = t(x)$ qui fournit l'estimation sur les données considérées. Mais l.e.a statisticien.ne saura bien distinguer ces deux cas !

Exemple Soit $X \sim \mathcal{P}_\theta$. Nous avons déjà rencontré la moyenne empirique d'un échantillon. C'est un estimateur de l'espérance $\mu = \mathbb{E}_\theta(X)$ de la loi $\mathbb{P}_\theta$ ayant généré les données :

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

La moyenne observée sur les données est également notée $\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$ est une estimation de μ. On peut proposer d'autres estimateurs :

$$T_n = 0; \quad \tilde{T}_n = X_1; \quad \tilde{\tilde{T}}_n = \frac{1}{[n/2]}\sum_{i=1}^{[n/2]} X_{2i} \tag{4.1}$$

Comment choisir ?

4.3 Biais, variance et risque

L'étude des propriétés d'un estimateur va apporter des éléments de réponse.

4.3.1 Biais

Définition 4.7. *Soit $\widehat{\nu}_n$ un estimateur de ν, défini à partir d'un n-échantillon de loi $\mathbb{P}_\theta$. Le **biais** de l'estimateur $\widehat{\nu}_n$ est défini par*

$$B_\theta(\widehat{\nu}_n, \nu) = \mathbb{E}_\theta(\widehat{\nu}_n) - \nu$$

*Si $B_\theta(\widehat{\nu}_n, \nu) = 0$, alors $\widehat{\nu}_n$ est dit **non biaisé** ou sans biais.*

Le biais représente l'erreur moyenne systématique due à la fluctuation de $\widehat{\nu}_n$ autour de son espérance $\mathbb{E}_\theta(\widehat{\nu}_n)$, plutôt qu'autour de la valeur à estimer ν. Il est donc a priori souhaitable d'utiliser des estimateurs sans biais.

Retour à l'exemple 4.1 L'estimateur constant $T_n = 0$ de l'espérance μ d'une loi est d'espérance nulle quelque soit la valeur de μ. Son biais vaut donc $-\mu$, nul uniquement si $\mu = 0$. T_n est donc biaisé, et même très fortement pour des grandes valeurs de μ. La linéarité de l'espérance et l'hypothèse de distribution identique des observations permettent de montrer que les autres estimateurs $\widehat{\mu}$, $\tilde{T}_n$ et $\tilde{\tilde{T}}_n$ de μ définis par 4.1 sont de leur côté non biaisés.

Remarque Si T_n est un estimateur sans biais de θ, alors $\nu(T_n)$ n'est pas forcément un estimateur sans biais de $\nu(\theta)$.

4.3.2 Variance

Définition 4.8. *La **variance** de l'estimateur $\widehat{\nu}_n$ est*

$$\mathrm{var}(\widehat{\nu}_n) = \mathbb{E}_\theta[(\widehat{\nu}_n - \mathbb{E}_\theta(\widehat{\nu}_n))^2]$$

La variance détermine donc la fluctuation aléatoire de $\widehat{\nu}_n$ autour de sa valeur moyenne (on devrait dire de son espérance). Il semble souhaitable d'utiliser des estimateurs de variance la plus faible possible.

Retour à l'exemple 4.1 Soit σ^2 la variance de la loi mère :

$$\mathrm{var}(T_n) = 0;\ \ \mathrm{var}(\tilde{T}_n) = \sigma^2;\ \ \mathrm{var}(\tilde{\tilde{T}}_n) = \frac{\sigma^2}{[n/2]};\ \ \mathrm{var}(\widehat{\mu}) = \frac{\sigma^2}{n}$$

Parmi les estimateurs sans biais, c'est T_n qui a la plus faible variance.

4.3.3 Risque quadratique

Mais la variance ne mesure que le carré de l'écart à l'espérance de l'estimateur, et ne prend pas en compte le biais. Le risque quadratique permet de comparer des estimateurs en prenant en compte le biais et la variance.

Définition 4.9. *Le **risque quadratique** ou **erreur quadratique moyenne** de l'estimateur $\widehat{\nu}_n$ pour l'estimation de ν est l'espérance de sa perte quadratique*

$$\nu \mapsto R_\theta(\widehat{\nu}_n, \nu) = \mathbb{E}_\theta[(\widehat{\nu}_n - \nu)^2],$$

Il mesure en moyenne (on devrait dire en espérance) le carré de la distance entre l'estimation et le paramètre à estimer. Le risque quadratique de l'estimateur empirique de l'espérance μ est

$$R_\theta(\bar{X}, \mu) = \mathbb{E}_\theta[(\bar{X} - \mu)^2] = \mathrm{var}_\theta(\bar{X}) = \frac{\sigma^2}{n}.$$

Propriété 4.10 (Décomposition biais-variance)**.** *Le risque quadratique s'écrit*

$$R_\theta(\widehat{\nu}_n, \nu) = \mathrm{var}_\theta(\widehat{\nu}_n) + (B_\theta(\widehat{\nu}_n))^2$$

Preuve. En effet, on a

$$\mathbb{E}_\theta\left[(\widehat{\nu}_n - \mathbb{E}_\theta(\widehat{\nu}_n))(\mathbb{E}_\theta(\widehat{\nu}_n) - \nu)\right] = \mathbb{E}_\theta\left[(\widehat{\nu}_n - \mathbb{E}_\theta(\widehat{\nu}_n))\right]\left[(\mathbb{E}_\theta(\widehat{\nu}_n) - \nu)\right] = 0$$

d'où

$$R_\theta(\widehat{\nu}_n, \nu) = \mathbb{E}_\theta[(\widehat{\nu}_n - \nu)^2] = \mathbb{E}_\theta[(\widehat{\nu}_n - \mathbb{E}(\widehat{\nu}_n))^2] + (\mathbb{E}(\widehat{\nu}_n) - \nu)^2 = \mathrm{var}_\theta(\widehat{\nu}_n) + (b_\theta(\widehat{\nu}_n))^2.$$

◊

Le risque quadratique permet de définir une relation d'ordre partiel sur les estimateurs

Définition 4.11. *Un estimateur δ_1 de $\nu(\theta)$ **domine** l'estimateur δ_2 si, pour tout $\theta \in \Theta$,*

$$R_\theta(\delta_1, \nu) \leq R_\theta(\delta_2, \nu)$$

*cette inégalité étant stricte pour au moins une valeur de ν. Un estimateur est **admissible** s'il n'existe aucun estimateur le dominant.*

Retour à l'exemple 4.1 $\tilde{T}_n$ et $\tilde{\tilde{T}}_n$ ne sont pas admissibles pour $n \geq 2$, car pour tout μ, $R(\bar{X}, \mu) < R(\tilde{T}_n, \mu) < R(\tilde{\tilde{T}}_n, \mu)$. On les écartera donc au profit de $\bar{X}$. Si $\mu = 0$, l'estimateur $T_n = 0$ est sans biais et de variance nulle, et donc de risque minimum. Il est donc admissible. Mais il n'est plus de risque minimum pour $\mu > \sigma/\sqrt{n}$: en particulier, il ne domine pas $\bar{X}$. Il n'est pas possible en général de minimiser le risque quadratique uniformément : on recherchera donc des estimateurs optimaux dans des sous classes, par exemple celle des estimateurs sans biais. Cette recherche d'estimateurs **Uniformément de Variance Minimale parmi les estimateurs sans Biais** sera abordée en deuxième année.

4.4 Convergence

La variabilité de l'estimation (qui s'exprime dans la variance) est due au côté parcellaire de l'information apportée par l'échantillon. Si l'échantillon devient très grand, voire de taille infinie, quel est l'espoir de tomber juste ? Il est donc intéressant regarder les comportements quand n tend vers l'infini, c'est la **théorie asymptotique** de la convergence de variables aléatoires, part importante des probabilités et de la statistique mathématique.

4.4.1 Définitions

Nous rappelons ici des notions vues dans le cours de probabilité. Un estimateur étant une variable aléatoire, sa convergence est définie comme celle de la variable aléatoire

Définition 4.12. *Soit $\widehat{\nu}_n$ un estimateur de $\nu = \nu(\theta)$, défini à partir d'un n-échantillon de loi $\mathbb{P}_\theta$*

— *$\widehat{\nu}_n$ est (faiblement) **consistant** ou **convergent** ssi $\widehat{\nu}_n$ tend en probabilité vers ν quand $n \to \infty$:*

$$\forall \theta \in \Theta, \forall \varepsilon, \lim_{n\to\infty} \mathbb{P}_\theta(|\widehat{\nu}_n - \nu| > \varepsilon) = 0$$

On note : $\widehat{\nu}_n \xrightarrow{\mathcal{P}} \nu$.

— *$\widehat{\nu}_n$ est **fortement consistant** (convergence **presque sûre**) ssi $\forall \theta \in \Theta$,*

$$\mathbb{P}_\theta(\lim_{n\to\infty} |\widehat{\nu}_n - \nu| = 0\ ) = 1$$

On dit aussi $\widehat{\nu}_n \to \nu$ avec probabilité 1.

Ces définitions sont présentées sous une forme plus générale dans un cours de probabilité, où la limite de la suite d'estimateurs indexée par n peut être elle-même une variable aléatoire. Dans le cadre statistique, la limite recherchée ν est le paramètre à inférer, qui n'est pas aléatoire. En statistique, on regarde les phénomènes en moyenne, et c'est la consistance (faible) qui est en général utilisée. La consistance presque sûre nécessite la convergence de presque toutes les trajectoires, propriété évidemment plus forte : la convergence presque sûre entraîne la convergence en probabilité.

Une autre forme de convergence, bien utile aux statisticiens pour démontrer la convergence en probabilité, est la convergence en moyenne quadratique.

Définition 4.13. *Soit $\widehat{\nu}_n$ un estimateur de $\nu = \nu(\theta)$, défini à partir d'un n-échantillon de loi $\mathbb{P}_\theta$. $\widehat{\nu}_n$ converge en **moyenne quadratique** vers ν ssi son risque tend vers 0 quand la taille de l'échantillon tend vers l'infini :*

$$\forall \theta \in \Theta, \lim_{n\to\infty} R_\theta(\widehat{\nu}_n, \nu) = 0.$$

On note : $\widehat{\nu}_n \xrightarrow{L^2} \nu$.

Propriété 4.14. *La convergence en moyenne quadratique implique la convergence en loi.*

Preuve. En effet, pour tout ε,

$$\begin{aligned}\mathbb{P}(|\widehat{\nu}_n - \nu| > \varepsilon) &\leq \mathbb{P}(|\widehat{\nu}_n - \mathbb{E}(\widehat{\nu}_n)| > \varepsilon) + \mathbb{P}(|\mathbb{E}(\widehat{\nu}_n) - \nu)| > \varepsilon) \\ &\leq \frac{[\sigma(\widehat{\nu}_n)]^2}{\varepsilon^2} + \mathbb{1}_{|\mathbb{E}(\widehat{\nu}_n)-\nu)|>\varepsilon}\end{aligned}$$

où on a utilisé l'inégalité de Tchebychev A.4. Quand le risque de $\widehat{\nu}_n$ tend vers 0, son biais et sa variance aussi ce qui permet de conclure. ◊

Il faut bien sûr privilégier les estimateurs consistants, puisqu'ils assurent qu'en probabilité ils infèrent la vraie valeur quand la taille de l'échantillon est infini. Ils infèrent donc une valeur (très) proche quand l'échantillon est (très) grand.

Le lemme de l'application continue (A.3 en annexe A1) est très utile pour étudier les estimateurs : ainsi, si $\widehat{\theta}(X)$ est un estimateur (fortement) consistant de θ, et ν une fonction continue de θ, alors $\nu(\widehat{\theta})$ est un estimateur (fortement) consistant de $\nu(\theta)$.

4.4.2 LGN : premier théorème fondamental en statistique

La loi (faible) des grands nombres indique que la moyenne empirique $\bar{X}$ d'un échantillon est consistante quand la loi mère est de variance finie. Sous les mêmes conditions, on peut montrer que la moyenne empirique est fortement consistante, c'est la loi forte des grands nombres.

Théorème 4.15 (Loi des grands nombres (LGN))**.** *Soit $X_1, \ldots, X_n \sim \mathcal{P}_\theta$ un n-échantillon i.i.d. de loi mère d'espérance $\mathbb{E}(X_i) = \mu$ finie. La moyenne empirique $\bar{X}$ satisfait les propriétés suivantes :*

— *C'est un estimateur sans biais de μ :*

$$B_\theta(\bar{X}, \mu) = \mathbb{E}_\theta(\bar{X}) - \mu = 0$$

— *Si $\mathcal{P}_\theta$ a une variance $\sigma^2 = \mathrm{var}_\theta(X_i)$ finie, la variance de $\bar{X}$ est*

$$\mathrm{var}_\theta(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{var}_\theta(X_i) = \frac{\sigma^2}{n},$$

et la suite $\bar{X}$ est (fortement) consistante vers μ.

C'est un théorème dont l'utilisation est capitale en statistique, et obtenu avec des hypothèses peu contraignantes. La moyenne empirique d'un échantillon est proche de son espérance lorsque n tend vers l'infini, quelle que soit la loi mère de l'échantillon.

Il existe la variante suivante :

Théorème 4.16 (Loi des grands nombres de Kolmogorov)**.** *Soit $X_1, \ldots, X_n \sim \mathcal{P}_\theta$ un n-échantillon i.i.d. de loi $\mathcal{P}_\theta$, tel que $\mathbb{E}(|g(X_1)|)$ soit fini. Alors l'estimateur $\widehat{\nu}_n$*

$$\widehat{\nu}_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

est fortement consistant pour estimer $\mathbb{E}_\theta[g(X_1)]$.

La loi des grands nombres nous dit que $\bar{X}$ voit ses réalisations se concentrer autour de $\mu = \mathbb{E}(X)$, quelle que soit la loi de l'échantillon. C'est insuffisant pour donner une loi approchée de l'estimateur, c'est à dire, les probabilités que $\bar{X}$ que $\bar{X}$ appartienne à des ensembles donnés. Cette étude amènera au deuxième théorème fondamental de la statistique, le théorème central limite 5.11.

4.5 Méthodes de construction

4.5.1 Méthode des moments

Cette méthode permet de définir des estimateurs à partir des moments de la loi mère des observations. C'est une méthode simple, mais pouvant produire des estimateurs de qualité parfois médiocre. Le moment d'ordre k de $\mathcal{P}_\theta$ est fonction de θ :

$$\mathbb{E}(X_1^k) = m_k(\theta).$$

On définit le **moment empirique** par

$$\widehat{m}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k,$$

qui est fortement consistant si $\mathbb{E}(|X^k|)$ existe. L'estimateur de la méthode des moments est obtenu résolvant le système

$$\begin{aligned} m_1(\widehat{\theta}) &= \widehat{m}_1 \\ &\vdots \qquad\qquad\qquad\qquad \vdots \\ m_p(\widehat{\theta}) &= \widehat{m}_p. \end{aligned}$$

Exemple Soit $\theta = (\mu, \sigma^2)$ le paramètre formé de l'espérance et la variance de la loi mère, que l'on cherche à estimer par $\widehat{\theta} = (\widehat{\mu}, S^2)$ par la méthode des moments. On calcule

$$\begin{aligned} m_1(\theta) &= \mathbb{E}(X_1) = \mu \\ m_2(\theta) &= \mathbb{E}(X_1^2) = \mu^2 + \sigma^2 \end{aligned}$$

En remplaçant $m_1(\theta)$ et $m_2(\theta)$ par $\widehat{m}_1$ et $\widehat{m}_2$ et on obtient :

$$\widehat{\mu} = \widehat{m}_1, \quad S^2 = \widehat{m}_2 - [\widehat{m}_1]^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

qui sont les indicateurs de moyenne et variance rencontrés en analyse descriptive.

Extension si $\nu(\theta)$ s'écrit $\nu(\theta) = \phi(m_1(\theta), \ldots, m_k(\theta))$ où $m_k(\theta) = \mathbb{E}_\theta(g_k(X_1))$. Un estimateur *plug-in* de $\nu(\theta)$ est

$$\widehat{\nu}_n = \phi\left(\frac{1}{n}\sum_{i=1}^{n} g_1(X_i), \ldots, \frac{1}{n}\sum_{i=1}^{n} g_k(X_i)\right)$$

en remplaçant les espérances par leur version empirique. Si $\mathbb{E}_\theta(|g_k(X_1)|)$ est fini pour tout k, et ϕ est continue, alors la méthode est consistante.

4.5.2 Méthode du maximum de vraisemblance

C'est une des méthodes les plus connues pour construire un estimateur. Elle s'utilise dans le cas où le modèle paramétrique $\{\mathbb{P}_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$ est **dominé**. Un modèle statistique est dit dominé par une mesure positive ξ, s'il existe pour tout $\theta \in \Theta$ une densité de $\mathbb{P}_\theta$ par rapport à ξ. Dans ce cours, nous aurons les deux cas suivants :

- soit $\mathbb{P}_\theta$ est une loi continue de densité f_θ pour tout θ par rapport à la mesure de Lebesgue
- soit $\mathbb{P}_\theta$ est une loi discrète chargeant des valeurs $\{x_1, x_2, \ldots ...\}$ indépendantes de θ, telles que $\sum_j \mathbb{P}_\theta(x_j) = 1$ pour tout θ. On notera alors à nouveau $f_\theta(x) = \mathbb{P}_\theta(X = x)$ la fonction de probabilité, qui peut être vue comme une densité par rapport à la mesure de comptage.

Ainsi,

$$\forall A \in \mathcal{A}, \ \ \mathbb{P}_\theta(A) = \int_A f_\theta(x)\, d\xi(x)$$

Définition 4.17. *Dans un modèle paramétrique dominé, on appelle* ***vraisemblance*** *d'une réalisation $(x_1, \ldots, x_n)$ du n-échantillon, la fonction de θ :*

$$\theta \mapsto L(\theta; x_1, \ldots, x_n) = f_\theta(x_1, \ldots, x_n)$$

Pour un échantillon i.i.d. : $L(\theta; x_1, \ldots, x_n) = \prod_{i=1}^n f_\theta(x_i)$

Attention, la vraisemblance (fonction de θ) n'est pas la densité (fonction de x), même si pour θ et x fixés elles ont les mêmes expressions.

Exemples La loi gaussienne $\mathcal{N}(\mu, \sigma^2)$ d'espérance μ et de variance σ^2 a la densité suivante par rapport à la mesure de Lebesgue

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

La vraisemblance de l'observation n-échantillon i.i.d. de loi gaussienne $\mathcal{N}(\mu, \sigma^2)$ s'écrit donc

$$L(\mu, \sigma^2; x) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp - \left(\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2} \right)$$

La loi de Bernoulli $\mathcal{B}(1, \theta)$ d'espérance θ admet la densité $\theta^x(1-\theta)^{1-x}$ définie par rapport à la mesure discrète somme des Dirac en 0 et 1. La vraisemblance de l'observation n-échantillon i.i.d. de loi de Bernoulli $\mathcal{B}(1, \theta)$ s'écrit donc

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

La vraisemblance en un θ donné est la densité de l'observation quand le modèle est défini avec ce θ. L'idée de l'estimateur du maximum de vraisemblance est de proposer un $\widehat{\theta}$ qui rend la vraisemblance la plus grande possible, ie, qui rend la densité de l'observation la plus grande possible quand on la calcule avec ce $\widehat{\theta}$.

Définition 4.18 (EMV)**.** *On appelle* ***estimation*** *du maximum de vraisemblance, une valeur $\widehat{\theta}_n$ maximisant la vraisemblance*

$$\widehat{\theta}_n \in \arg\max_{\theta \in \Theta} L(\theta; x).$$

$\widehat{\theta}_n = t(x_1, \ldots, x_n)$ est une fonction des données, ce qui induit la statistique $t(X_1, \ldots, X_n)$ que l'on note traditionnellement avec la même notation : $\widehat{\theta}_n = t(X_1, \ldots, X_n)$ est appelé ***Estimateur du Maximum de Vraisemblance***

De même, on ne fait pas de différence de notation entre $L(\theta; X)$, vraisemblance de l'échantillon $X = (X_1, \ldots, X_n)$, vue comme une fonction de variables aléatoires et $L(\theta; x)$ vraisemblance définie comme fonction déterministe pour un résultat d'observation.

Dans le cas de modèles i.i.d., la vraisemblance de l'échantillon est le produit des vraisemblances individuelles de chacune des observations. Il est donc souvent pratique de travailler avec la **log-vraisemblance** pour remplacer le produit par une somme :

$$\log L(\theta; X) = \sum_{i=1}^{n} \log f_\theta(X_i)$$

Exemples L'estimateur du maximum de vraisemblance de θ dans le modèle i.i.d. de Bernoulli est $\widehat{\theta} = \bar{X}$.
L'estimateur du maximum de vraisemblance de $\theta = (\mu, \sigma^2)$ dans le modèle i.i.d. gaussien est

$$\widehat{\theta} = (\widehat{\mu}, \widehat{\sigma}^2); \;\; \widehat{\mu} = \bar{X}; \;\; \widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

L'EMV est en général biaisé à distance finie, mais sous des conditions de régularité des modèles, il est asymptotiquement sans biais, consistant et sa variance est optimale asymptotiquement. Ces propriétés seront détaillées en deuxième année.

4.6 Fonction de répartition empirique

Un échantillon $X_1, \ldots, X_n$ de la loi $\mathcal{P}_\theta$ permet également d'estimer la fonction de répartition F de cette loi.

Définition 4.19. *La fonction de* ***répartition empirique*** $\widehat{\mathrm{F}}_n$ *est la fonction de répartition de la loi de probabilité discrète uniforme de support* $\{X_1, \ldots, X_n\}$ *:*

$$\widehat{\mathrm{F}}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \leq x),$$

où $\mathbb{1}(X_i \leq x) = 1$ *si* $X_i \leq x$*, 0 sinon.*

Propriété 4.20. *Pour tout* x*,* $\widehat{\mathrm{F}}_n(x)$ *est un estimateur sans biais de* $\mathrm{F}(x)$*. Son risque est*

$$R_F(\widehat{\mathrm{F}}_n(x), \mathrm{F}(x)) = \frac{\mathrm{F}(x)(1 - \mathrm{F}(x))}{n}.$$

En effet, $Y = \mathbb{1}(X_i \leq x)$ suit une loi de Bernoulli $\mathcal{B}(0, \mathbb{P}(X_i \leq x) = F(x))$, d'espérance $F(x)$, $\widehat{F}(x)$ est donc sans biais, et de variance $\mathrm{F}(x)(1 - \mathrm{F}(x))$. $\widehat{\mathrm{F}}(x)$ est la moyenne d'un n-échantillon de Y, d'où le résultat. La figure 4.1 représente la fonction de répartition empirique de la loi gaussienne centrée réduite pour diverses tailles d'échantillon et superpose la fonction de répartition théorique. Quant la taille de l'échantillon augmente, la fonction de répartition empirique approche celle de la fonction de répartion théorique.

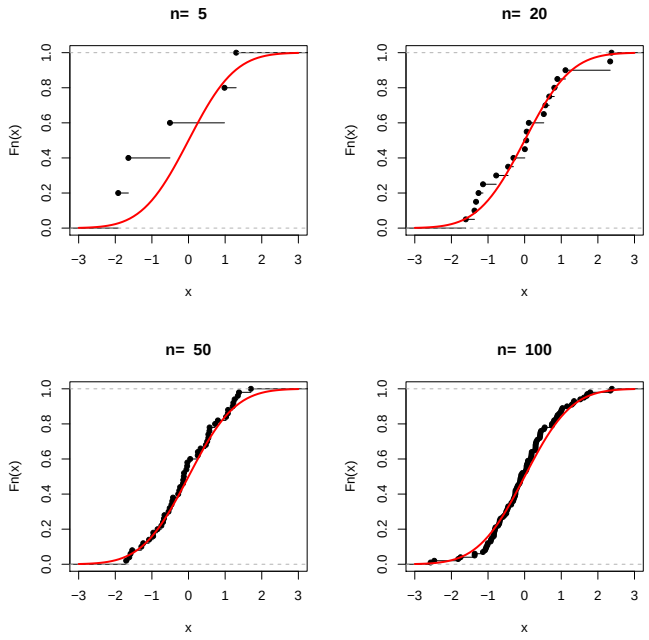


FIGURE 4.1 – La fonction de répartition empirique de la loi gaussienne centrée réduite pour diverses tailles d'échantillon.

Chapitre 5

Loi des estimateurs

Nous avons vu dans les chapitres précédents qu'il est possible de définir des estimateurs, variables aléatoires définies à partir d'un échantillon et permettant d'attribuer une valeur au paramètre inconnu. L'estimateur a des propriétés (biais, variance, risque) qui permettent d'évaluer l'erreur commise. La consistance assure d'estimer asymptotiquement la bonne valeur en probabilité. Mais l'estimation reste pour l'instant ponctuelle et ces informations souvent insuffisantes pour définir un cadre de décision.

Exemple Prenons par exemple le cas d'un constructeur automobile qui indique une consommation de $c_0 = 6.32\ell/100km$ pour les véhicules d'un type donné, dans des conditions précises de roulage, avec un écart-type de $\sigma_0 = 0.21\ell/100\ km$. Un organisme indépendant prend 30 véhicules au hasard, et les soumet aux conditions de roulage nominales. Il observe $\bar{x} = 6.43\ell/100\ km > c_0$, $\widehat{\sigma} = 0.25\ell/100\ km$. Les deux valeurs de consommation ne sont pas identiques. Mais la différence est-elle simplement due à la variabilité naturelle de l'expérience, c'est à dire au fait que l'échantillon tiré aléatoirement induit une imprécision de mesure sur la valeur nominale, et malencontreusement en légère défaveur du constructeur sur ces 30 voitures ? Où le constructeur a-t-il sous-estimé sciemment la consommation de ses véhicules ? Pour répondre à cette question, il faut pouvoir déterminer si le fait d'observer une moyenne plus grande que 6.42 est d'une probabilité forte ou pas sous les indications du constructeur, c'est à dire accéder à $\mathbb{P}(\bar{X} \geq 6,43)$ et donc, connaître la loi de $\bar{X}$.

Nous commencerons par étudier le cas gaussien pour lequel les calculs sont aisés, puis nous présenterons l'utilisation de l'approximation asymptotique quand le calcul à distance finie ne l'est pas. Le chapitre se termine par quelques éléments sur la loi empirique.

5.1 Cas gaussien

On suppose que la loi mère d'un n-échantillon $(X_1, \ldots, X_n)$ est gaussienne $\mathcal{N}(\mu, \sigma^2)$ d'espérance μ et de variance σ^2.

5.1.1 Loi de la moyenne empirique

Nous avons vu que l'estimateur empirique $\bar{X}$ de l'espérance est non biaisé, de variance σ^2/n où σ^2 est la variance de la loi mère. C'est donc également le cas pour une loi mère gaussienne.

Propriété 5.1. *L'estimateur empirique $\bar{X}$ de l'espérance d'une loi gaussienne $\mathcal{N}(\mu, \sigma^2)$, calculé à partir d'un échantillon i.i.d $(X_1, \ldots, X_n)$ de cette loi, est gaussien :*

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

Preuve. La propriété découle immédiatement des propriétés des variables gaussienne : en particulier, si $X \sim \mathcal{N}\left(\mu_x, \sigma_x^2\right)$ et $Y \sim \mathcal{N}\left(\mu_y, \sigma_y^2\right)$ sont deux variables gaussiennes indépendantes, et a et b deux réels, alors

$$aX + bY \sim \mathcal{N}\left(a\mu_x + b\mu_y, a^2\sigma_x^2 + b^2\sigma_y^2\right)$$

◊

Remarque On retrouve bien l'expression (déjà rencontrée) de la variance de $\bar{X}$, qui tend vers 0 quand n tend vers l'infini : la loi de l'estimateur se concentre autour de la valeur de l'espérance. Ceci est représenté sur la figure 5.1 : à gauche, la densité de la loi de $\bar{X}$ pour des échantillons de taille $n = 5$, 10, 30 et 100. Les moyennes calculées à partir d'un échantillon de taille 100 sont beaucoup moins dispersées que celles calculées à partir d'un échantillon de taille 5, l'estimation est donc bien plus précise quand n est grand. Les figures de droite représentent les histogrammes des moyennes générées à partir d'échantillons de 1000 valeurs de $\bar{X}$, chacune calculée à partir d'un n-échantillon. On retrouve la concentration autour de l'espérance de la loi.

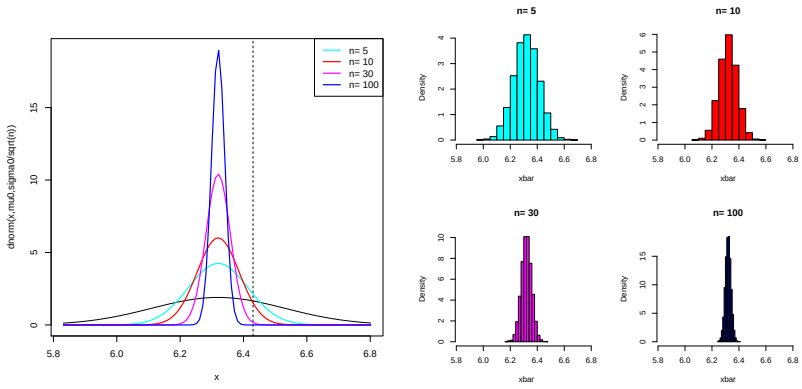


FIGURE 5.1 – La densité de $\bar{X}$ pour différentes taille d'échantillon (à gauche), son estimation (sous forme d'histogramme) d'échantillons de moyennes empiriques, elles-mêmes issues d'échantillons de tailles différentes (à droite)

Applications En revenant à l'exemple introductif, nous pouvons utiliser la connaissance de la loi de $\bar{X}$ de la façon suivante :

Cas 1 : Si on considère que la mesure de pollution d'un véhicule de marque donnée suit $X_1 \sim \mathcal{N}(c_0 = 6.32, \sigma_0^2 = 0.21^2)$, alors pour un échantillon de taille $n = 30$

$$\begin{aligned}\mathbb{P}(\bar{X} \geq 6.43) &= \mathbb{P}\left(\frac{\bar{X} - c_0}{\sigma_0/\sqrt{n}} \geq \frac{6.43 - c_0}{\sigma_0/\sqrt{n}}\right) \\ &= 1 - \mathrm{F}^*\left(\frac{6.43 - c_0}{\sigma_0/\sqrt{n}}\right) \simeq 0.002\end{aligned}$$

où F^* est la fonction de répartition de la loi gaussienne centrée réduite $\mathcal{N}(0,1)$. Ses valeurs sont tabulées, voir annexe B.1 (ou calculables via une calculette, un tableur ou un logiciel de statistique)

On peut donc remarquer que si la consommation suit la loi nominale, l'observation $\bar{x} = 6.43$ est rare, puisque la probabilité d'avoir une consommation encore plus élevée est environ de 2 pour 1000. Nous formaliserons cette approche dans le chapitre sur les tests.

Cas 2 : On peut aussi se demander quelle valeur q de la consommation moyenne sur 30 véhicules est dépassée avec une probabilité donnée (par ex, $\alpha = 5\%$)

$$\mathbb{P}(\bar{X} \geq q) = 1 - \mathrm{F}^*\left(\frac{q - c_0}{\sigma_0/\sqrt{n}}\right) = 0.05$$

Soit

$$q = c_0 + \underbrace{\mathrm{F}^{*-1}(0.95)}_{\text{quantile d'ordre 95\% de } \mathcal{N}(0,1)} \frac{\sigma_0}{\sqrt{n}} = 6.38$$

Certaines valeurs des quantiles de la gaussienne centrée réduite font partie des constantes que le-a statistien-ne connait sans logiciel ! Par exemple

$$\mathrm{F}^{*-1}(0.95) = 1.64; \quad \mathrm{F}^{*-1}(0.975) = 1.96; \quad \mathrm{F}^{*-1}(0.99) = 2.33; \quad \mathrm{F}^{*-1}(0.999) = 3.09$$

5.1.2 Loi de l'estimateur de la variance à espérance connue

Propriété 5.2. *Soit $V_n^* = \frac{1}{n}\sum_i (X_i - \mu)^2$, l'estimateur empirique de la variance d'un échantillon i.i.d. X de loi $\mathcal{N}(\mu, \sigma^2)$, μ connue. On a :*

$$n\frac{V_n^*}{\sigma^2} \sim \chi^2(n)$$

où $\chi^2(n)$ est la loi du Khi-deux à n degrés de liberté.

Ceci découle directement de la définition :

Définition 5.3 (loi du Khi-deux)**.** *Soit Z un vecteur gaussien **centré réduit** et de composantes **indépendantes** de dimension n. La loi de la somme du carré de ses composantes est la loi du **Khi-deux** (centré) à n degrés de liberté*

$$K_n = \sum_i Z_i^2 \sim \chi^2(n)$$

de densité sur $\mathbb{R}^+$

$$f_{K_n}(x) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} x^{n/2-1} e^{-x/2}$$

Sa fonction génératrice des moments est

$$\psi_{K_n}(t) = \mathbb{E}(e^{tK_n}) = \frac{1}{(1-2t)^{n/2}}$$

et on a : $\mathbb{E}(K_n) = n$; $\mathrm{var}(K_n) = 2n$

La figure 5.2 représente la densité de la loi du Khi-deux pour différents degrés de liberté. Dissymétrique pour de faibles valeurs de n, celle-ci devient symétrique pour des valeurs de n plus importantes.

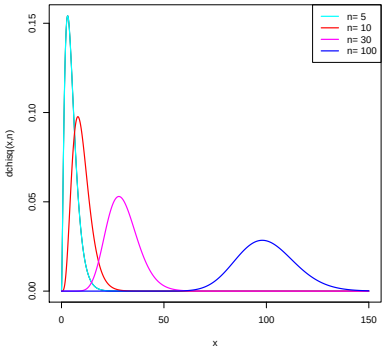


FIGURE 5.2 – Loi du $\chi^2(n)$ pour différents degrés de libertés

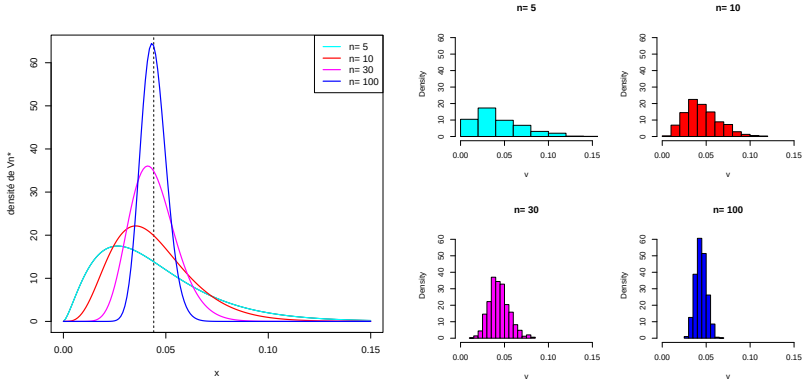


FIGURE 5.3 – Loi de V_n^* pour différents degrés de liberté : loi théorique (à gauche), empirique (à droite)

La figure 5.3 représente la densité de l'estimateur empirique de la variance à espérance connue. Comme pour la loi gaussienne, les lois du Khi-deux sont tabulées (ou calculables via une calculette, un tableur ou un logiciel de statistique).

Le cas de l'estimateur de la variance à espérance connue est cependant peu utilisé. On ne connait souvent ni l'espérance, ni la variance de la loi, d'où l'utilisation de l'estimateur empirique $S_n^2 = \frac{1}{n}\sum_i (X_i - \bar{X})^2$ étudié ci-après.

5.1.3 Loi de la variance empirique

Propriété 5.4. *Si X un n-échantillon gaussien de variance σ^2, alors*

$$\sum_i^n \frac{(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1)$$

On en déduit :

$$n\frac{S_n^2}{\sigma^2} \sim \chi^2(n-1); \quad (n-1)\frac{\widehat{\sigma}_n^2}{\sigma^2} \sim \chi^2(n-1)$$

où $S_n^2 = \frac{1}{n}\sum_i (X_i - \bar{X})^2$ *et* $\widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_i (X_i - \bar{X})^2$.

Ce théorème a été montré dans le cours probabilité en utilisant les fonctions génératrices, avec les principales étapes suivantes :

— La fonction génératrice des moments de X_i est $\psi_{X_i}(t) = \mathbb{E}(e^{tX_i}) = \exp\left(\mu t + \frac{\sigma^2}{2}t^2\right)$

— La fonction génératrice des moments du vecteur $(\bar{X}, X_1 - \bar{X}, \ldots, X_n - \bar{X})$ s'écrit

$$\psi(s, t_1, \ldots, t_n) = \psi_{X_1,\ldots,X_n}(u_1, \ldots, u_n) \text{ où } u_i = \frac{s}{n} + (t_i - \bar{t})$$

— Par indépendance des (X_i), elle se factorise $\psi(s, t_1, \ldots, t_n) = \psi_{\bar{X}}(s)\psi_T(t_1, \ldots, t_n)$ donc, $\bar{X}$ et $T = (X_1, \ldots, X_n)$ sont indépendantes, et par conséquent $\bar{X}$ et $\widehat{\sigma}^2$ (ou S_n) le sont aussi.

— La loi de $\widehat{\sigma}_n$ se déduit de la décomposition

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

La fonction génératrice des moments d'une loi $\chi^2(n)$ s'écrit, pour $t < \frac{1}{2}$, $\psi_{K_n}(t) = \left(\frac{1}{1-2t}\right)^{n/2}$. De plus, $\sum_i (X_i - \bar{X})^2$ et $\bar{X}$ sont indépendants. La fonction génératrice de la somme est donc le produit des fonctions génératrices, soit

$$\left(\frac{1}{1-2t}\right)^{n/2} = \psi_{nS_n^2/\sigma^2}(t)\left(\frac{1}{1-2t}\right)^{1/2}, \quad t < \frac{1}{2}$$

Il s'avère que la propriété 5.4 est une conséquence directe du théorème de Cochran :

Théorème 5.5 (Cochran)**.** *Si* $Y \sim \mathcal{N}_n(\mu, \sigma^2 I_n)$*, et si* $E_1 \oplus \ldots \oplus E_r = \mathbb{R}^n$ *est une décomposition de* $\mathbb{R}^n$ *en* r *sous-espaces orthogonaux, alors les projections orthogonales* $\Pi_1(Y), \ldots, \Pi_r(Y)$ *sur ces sous-espaces sont des vecteurs gaussiens indépendants tels que, pour tout* $j = 1, \ldots, r$

$$||\Pi_j(Y)||^2 \sim \sigma^2\chi^2(d_j = Dim(E_j), \mu_j = ||\Pi_j(\mu)||^2).$$

Preuve. Pour tout j, soit $(e_{j1}, \ldots, e_{jk})$ une base orthonormée de E_j où e_{jk} est le k-ième vecteur de la base de E_j. La décomposition sur cette base orthonormée du vecteur Y s'écrit

$$\Pi_j Y = \sum_{k=1}^{d_j} < e_{jk}, Y > e_{jk}$$

Soit U matrice de passage d'une base orthonormée de $\mathbb{R}^n$ dans base orthonormée de $\mathbb{R}^n$. Elle est telle que $UU' = Id_n$. On a $UY \sim \mathcal{N}_n(U\mu, Id_n)$. Les variables $e'_{jk}Y$ sont donc indépendantes quand j et k varient. Donc $\Pi_1(Y), \ldots, \Pi_r(Y)$ sont indépendantes.

Pour un sous-espace E_j, et pour $k = 1, \ldots, k_j$

$$e'_{jk}Y \sim \mathcal{N}(e'_{jk}\mu, e'_{jk}e_{jk} = 1)$$

d'où , avec $\mu_j = ||\Pi_j\mu||^2 = \sum_{k=1}^{d_j}(e'_{jk}\mu)^2$ on a

$$||\Pi_j(Y)||^2 = \sum_{k=1}^{d_j}||e'_{jk}Y||^2 \sim \chi^2(d_j, \mu_j),$$

$\Diamond$

En décomposant $\mathbb{R}^n = E_1 \oplus E_1^\perp$ où E_1 est la droite engendrée par le vecteur unitaire $(1, \dots 1)/\sqrt{n}$, on déduit du théorème de Cochran que pour un n-échantillon gaussien Y :
— $\bar{Y}$ et $\widehat{\sigma}^2$ sont indépendants
— $\sum_i (Y_i - \bar{Y})^2 \sim \sigma^2 \chi^2_{n-1}$

Application On peut ainsi calculer la probabilité qu'une observation de la variance calculée sur un échantillon de loi nominale soit supérieure à la variance observée sur l'échantillon des 30 véhicules testés :

$$\mathbb{P}(\widehat{\sigma}_n^2 > 0.25^2) = 1 - \mathrm{F}_{\chi^2(n-1)}(0.25^2/\sigma_0^2) \simeq 0.07$$

5.1.4 Conséquence : loi de Student

La variable $\bar{X}$ étant gaussienne, nous pouvons en dériver la variable centrée réduite ("normalisée") correspondante : $(\bar{X} - \mu)/(\sigma/\sqrt{n}) \sim \mathcal{N}(0, 1)$. Mais que devient cette loi, si la variance σ^2 n'étant pas connue, elle est remplacée par un estimateur ?

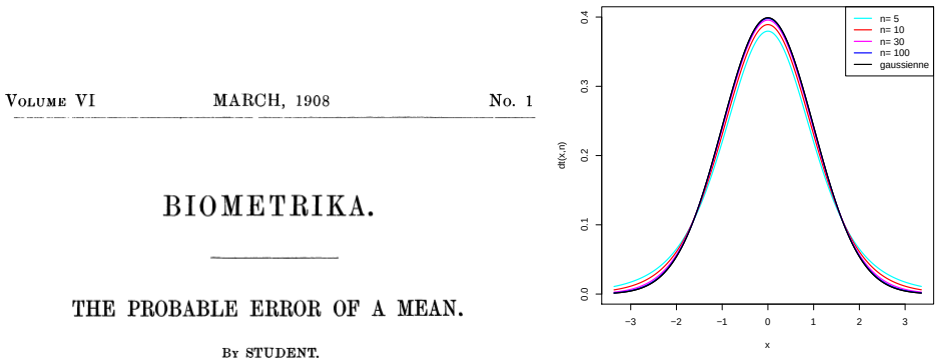
C'est William Gosset, élève de R. Fisher, qui a défini cette loi quand il travaillait sur l'estimation de l'erreur de la moyenne d'un échantillon gaussien. Ses articles ont été publiés sous le nom de Student, d'où la loi qui porte son nom :

Définition 5.6 (Loi de Student)**.** *Soit deux variables Z et K* ***indépendantes*** *telles que $Z \sim \mathcal{N}(0, 1)$ et $K \sim \chi^2(p)$. Alors, la v.a.*

$$T = \frac{Z}{\sqrt{\frac{K}{p}}} \sim \mathcal{T}(p)$$

suit une loi appelée loi de ***Student*** *à p degrés de liberté. T est appelée statistique de Student. Sa densité est déifnie sur $\mathbb{R}$:*

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$


VOLUME VI MARCH, 1908 No. 1

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

BY STUDENT.

FIGURE 5.4 – Quelques exemples de densité de la loi de Student pour différents degrés de liberté

La loi de Student est paire, d'espérance nulle si $n \geq 2$, de variance $\mathrm{var}(T) = \frac{n}{n-2}$ si $n \geq 3$. Pour x de valeur absolue suffisamment grande, la courbe de sa densité se situe au dessus de

celle de la gaussienne centrée réduite. On dit qu'elle est à queue de probabilité plus lourde que la gaussienne, prenant ainsi en compte l'incertitude sur l'estimation de la variance. La loi de Student tend vers la loi gaussienne quand le nombre de degrés de liberté tend vers l'infini.

Propriété 5.7. *Si $X_1, \ldots, X_n$ est un n-échantillon gaussien de loi $\mathcal{N}(\mu, \sigma^2)$,*

$$\sqrt{n}\frac{\bar{X} - \mu}{\widehat{\sigma}_n} \sim \mathcal{T}(n-1) \quad \textit{avec } \widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Cette propriété est une conséquence immédiate des lois de $\bar{X}$ et $\widehat{\sigma}_n^2$ et de leur indépendance dans le cas gaussien (cf section précédente).

Application : on peut reprendre le calcul de la probabilité de dépasser la valeur 6.43 observée sur l'échantillon, pour un échantillon de loi nominale.

$$\mathbb{P}(\bar{X} > 6,43) = 1 - \mathrm{F}_{\mathcal{T}}(\frac{6.43 - c_0}{0.25/\sqrt{n}}, n-1) \simeq 0.011$$

Cette probabilité est moins grande que dans le cas où σ est connue (11% à comparer avec 0.2%). L'incertitude sur le modèle (un paramètre supplémentaire inconnu) étant plus grande, on ne peut que moins s'étonner d'avoir une valeur éloignée de la valeur nomimale.

Notons que la statistique de Student utilisant S_n^2 s'écrit :

$$\sqrt{n}\frac{\bar{X} - \mu}{S_n\sqrt{\frac{n}{n-1}}} = \sqrt{n-1}\frac{\bar{X} - \mu}{S_n} \sim \mathcal{T}(n-1) \quad \text{avec } S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

5.1.5 Loi de Fisher

Définition 5.8 (Loi de Fisher)**.** *Soit deux variables K_1 et K_2 **indépendantes** telles que $K_1 \sim \chi^2(n_1)$ et $K_2 \sim \chi^2(n_2)$. Alors, la v.a.*

$$F = \frac{K_1/n_1}{K_2/n_2} \sim \mathcal{F}(n_1, n_2)$$

*suit une loi appelée loi de **Fisher** à (n_1, n_2) degrés de liberté. Sa densité est définie pour $x > 0$ par*

$$f(x) = \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}}\frac{x^{\frac{n_1-2}{2}}}{(1+\frac{n_1}{n_2}x)^{\frac{n_1+n_2}{2}}}$$

et vaut 0 sinon

Propriété 5.9. *$\mathbb{E}(F)$ existe pour $n_2 \geq 2$ et vaut $\mathbb{E}(F) = \frac{n_2}{n_2-2}$. $\mathrm{var}(F)$ existe pour $n_2 \geq 5$ et vaut $\mathrm{var}(F) = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}$*

La loi de Fisher est utile pour définir le rapport des estimateurs de deux variances.

Propriété 5.10 (Loi du rapport des estimateurs de variance)**.** *Soient deux échantillons gaussiens indépendants de taille n_1 et n_2, de **même variance** σ^2, et soient $\widehat{\sigma}_1^2$ et $\widehat{\sigma}_2^2$ les estimateurs **non biaisés** de la variance σ^2 dans chacun des deux échantillons. Alors, la v.a.*

$$\frac{\widehat{\sigma}_1^2}{\widehat{\sigma}_2^2} \sim \mathcal{F}(n_1 - 1, n_2 - 1)$$

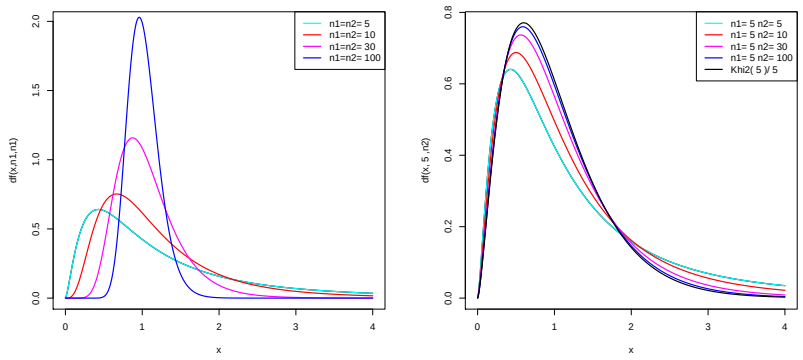


FIGURE 5.5 – Densités de lois de Fisher pour différents degrés de liberté

5.2 TLC et approximation gaussienne

La loi gaussienne n'est pas la seule qui permet de calculer la loi de la moyenne empirique à distance finie. Il est facile d'exprimer la loi de la moyenne empirique d'un échantillon de loi de Poisson, exponentielle, ou Gamma par exemple, voir TD. On pourrait, pour chaque type de loi mère, proposer des lois normalisées qui seraient tabulées.

En fait, si les échantillons sont (assez) grands, la loi de $\bar{X}$ a un comportement approximativement gaussien si la loi mère est de carré intégrable. Ainsi, même pour un échantillon non gaussien, on peut utiliser une **approximation gaussienne** de l'estimateur de l'espérance pour des échantillons **suffisamment grands**. Ceci est une conséquence du théorème central limite déjà rencontré dans cours de probabilité, et qui est le deuxième théorème d'importance que nous rencontrons en statistique.

Théorème 5.11 (central limite (TCL)). *Soit $\{X_n\}$ une suite de variables aléatoires i.i.d. admettant une espérance μ et une variance $\sigma^2 > 0$ finie. Alors, la suite des variables $\sqrt{n}(\bar{X}_n - \mu)$ converge en loi vers la v.a. $\mathcal{N}(0, \sigma^2)$ quand $n \to \infty$*

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

Preuve. Ce théorème a été vu dans le cours de probabilité. On rappelle ici les grandes lignes de la preuve. Soit $\varphi_X(t)$ la fonction caractéristique d'un X_i ; la fonction caractéristique de

$$W_n = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} = \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma\sqrt{n}} = \sum_i Z_i$$

est donc égale à $[\varphi_Z(t)]^n$ puisque les Z_i sont indépendants. Or, Z_i est une v.a. d'espérance nulle et de variance $1/n$.

Le développement en série de la fonction caractéristique de Z commence donc par $1 - \frac{t^2}{2n}$, les termes suivants sont des infiniment petits d'ordre $1/n^2$. En l'élevant à la puissance n, la fonction caractéristique de W_n est équivalente à $(1 - \frac{t^2}{2n})^n$ et tend vers $e^{-t^2/2}$ quand $n \to \infty$ selon un résultat classique. ◊

Remarque Ce théorème s'écrit encore

$$\mathbb{P}\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq x\right) \to F^*(x) = \int_{\infty}^{x} \frac{2}{\sqrt{2\pi}} e^{-t^2/2} dt$$

La loi de l'estimateur empirique de l'espérance d'une loi **quelconque** de moment d'ordre 2 fini peut être approximée par une loi gaussienne

$$\bar{X} \overset{appr}{\sim} \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

On peut calculer de façon approchée, pour n suffisamment grand, des probabilités d'événements relatifs à $\bar{X}_n$

$$\mathbb{P}\left(a \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq b\right) \simeq \mathbb{P}(a \leq Z \leq b) = F^*(b) - F^*(a)$$

où $Z \sim \mathcal{N}(0,1)$ et ceci, pour tout échantillon iid de loi mère de variance finie.

Exemple si on observe un n échantillon de loi de Bernoulli $\mathcal{B}(1,\theta)$, et si $n\theta > 5$ et $n(1-\theta) > 5$, alors l'approximation suivante est satisfaisante :

$$\bar{X} \overset{appr}{\sim} \mathcal{N}\left(\theta, \frac{\theta(1-\theta)}{n}\right), \text{ ou } n\bar{X} \overset{appr}{\sim} \mathcal{N}(n\theta, \theta(1-\theta))$$

Sous cette deuxième forme, $n\bar{X}$ suit exactement un loi binomiale $\mathcal{B}(n,\theta)$, et c'est simplement l'approximation de la loi binomiale par une loi gaussienne. La loi de $\bar{X}$ étant discrète, et celle de la gaussienne continue, il est possible de calculer une correction de continuité : soit $B \sim \mathcal{B}(n,\theta)$

$$\mathbb{P}(B = k) \simeq \mathbb{P}(k - \frac{1}{2} < U < k + \frac{1}{2}), \text{ où } U \sim \mathcal{N}(n\theta, n\theta(1-\theta))$$

$$\mathbb{P}(B \leq k) \simeq \mathbb{P}(U < k + \frac{1}{2})$$

En utilisant le lemme de Slutsky A.8, on montre que la statistique de Student converge en loi vers une gaussienne centrée réduite, puisque $\widehat{\sigma}$ tend en probabilité vers σ et $\sqrt{n}\frac{\bar{X}-\mu}{\sigma}$ est une gaussienne centrée réduite. C'est ce comportement qui a été représenté sur la figure 5.4, et qui est formalisé par la propriété suivante :

Propriété 5.12. *Si $X_1, \ldots, X_n$ est un n-échantillon de loi d'espérance μ et de variance σ^2 finie,*

$$\sqrt{n}\frac{\bar{X}-\mu}{\widehat{\sigma}_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \quad \textit{avec } \widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$\sqrt{n}\frac{\bar{X}-\mu}{S_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \quad \textit{avec } S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

5.2.1 Normalité asymptotique

De façon générale, tout estimateur correctement renormalisé convergeant en loi vers une gaussienne est dit asymptotiquement normal :

Définition 5.13. *Si un estimateur $\widehat{\nu}_n$ de $\nu \in \mathbb{R}^p$ de variance $\mathrm{var}(\widehat{\nu}_n) = V_n$ a un comportement asymptotique tel que*

$$V_n^{-1/2}(\widehat{\nu}_n - \nu) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, Id_p)$$

*on dit que l'estimateur est **asymptotiquement normal**. Si $nV_n \to V_0$ où $V_0 > 0$ est finie, on dit que la **vitesse** de l'estimateur est en $\sqrt{n}$*

Un estimateur est d'autant meilleur que sa vitesse de convergence est rapide et sa loi limite concentrée autour de 0.

5.2.2 Delta-méthode

La delta-méthode (de l'anglais delta-method) permet une linéarisation locale d'une fonction d'un paramètre asymptotiquement normal.

Théorème 5.14. *Si h est une fonction différentiable de $\nu \in \mathbb{R}^p$ et $\widehat{\nu}_n$ un estimateur asymptotiquement normal, alors $h(\widehat{\nu}_n)$ est un estimateur asymptotiquement normal de $h(\nu)$*

$$(D_\nu V_n(\nu) D'_\nu)^{-1/2}(h(\widehat{\nu}_n) - h(\nu)) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, Id_p)$$

avec $D_\nu = \begin{pmatrix} \partial h(\nu)/\partial \nu_1 & \dots & \partial h(\nu)/\partial \nu_p \end{pmatrix}$ et la notation D' désigne la matrice transposée de D. On peut alors également écrire

$$(h(\widehat{\nu}_n) - h(\nu))'(D_\nu V_n(\nu) D'_\nu)^{-1}(h(\widehat{\nu}_n) - h(\nu)) \xrightarrow{\mathcal{L}} \chi^2(p)$$

5.3 Loi empirique

Nous avons vu en section 4.6 un estimateur non-paramétrique de la fonction de répartition de la loi mère de l'échantillon que nous avons appelé fonction de répartition empirique. C'est la fonction de répartition d'une loi de probabilité dite empirique qui attribue la même masse à tous les points de l'échantillon $X_1, \dots, X_n \sim F$.

Définition 5.15. *La **loi de probabilité empirique** $\mathbb{P}_n$, ou loi empirique, définit la distribution uniforme sur l'ensemble fini des valeurs $(X_1, \dots, X_n)$ de l'échantillon :*

$$\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$

où δ_x est la masse de Dirac au point x.

Ainsi, la probabilité de tout intervalle I de $\mathbb{R}$ est égale à $\mathbb{P}_n(I) = \sum_{i=1}^n \mathbb{1}(X_i \in I)$.

Cette loi de probabilité admet comme fonction de répartition la fonction de répartition empirique que nous noterons ici sans chapeau

$$\mathrm{F}_n(x) = \int_{-\infty}^{x} d\mathbb{P}_n = \mathbb{P}_n(]-\infty; x]) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \leq x),$$

La loi empirique permet de définir des estimateurs dit empiriques : il sont définis en remplaçant $\mathbb{P}$ par $\mathbb{P}_n$. Le tableau suivant récapitule les correspondances entre les propriétés d'une loi et ses versions empiriques

Probabilité $X \sim F$ (connue)	**Statistique** $X_1, \dots, X_n \sim F$ (inconnue)
Loi de probabilité F	Mesure empirique $\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$
Fonction de répartition $\forall x \in \mathbb{R}, \mathrm{F}(x) = \mathbb{P}\{X \leq x\}$	Fonction de répartition empirique $\mathrm{F}_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}(X_i \leq x),$
Fonction quantile $\forall \alpha \in [0,1], q_F(\alpha) = \mathrm{F}^{-1}(\alpha)$	Quantile empirique $q_n(\alpha) = \mathrm{F}_n^{-1}(\alpha)$
Espérance $\mathbb{E}(X) = \int x dF(x)$	Espérance empirique $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$
Variance $\mathrm{var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]$	Variance empirique $S^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$
Moment $m_k(X) = \mathbb{E}(X^k)$	Moment empirique $\widehat{m}_k = \frac{1}{n}\sum_{i=1}^n X_i^k$

La loi empirique ne postule aucune connaissance a priori sur la loi mère, elle se construit de la même façon pour une loi discrète ou continue, une loi gaussienne, exponentielle ou de Bernoulli. Mais rappelons-nous que nous cherchons à définir la loi d'estimateurs, ie la loi de $\bar{X}$, S^2 par exemple. L'utilisation de la loi empirique devient peu aisé si on ne fait pas d'autres hypothèses de modélisation, alors qu'il y des cas de loi mère pour lesquels les lois des estimateurs s'expriment facilement.

Chapitre 6

Tests

Nous avons vu dans les chapitres précédents qu'un estimateur permet d'inférer une valeur d'un paramètre inconnu à partir de l'observation d'un n-échantillon. L'estimateur est une variable aléatoire. Sa réalisation sur un échantillon particulier est une valeur. Nous allons voir dans ce chapitre, que ces informations peuvent être utilisées dans un cadre **décisionnel**, afin de choisir entre deux hypothèses.

6.1 Introduction

Reprenons l'exemple d'un constructeur automobile annonçant une consommation moyenne $\mu_0 = 6.32\ell/100\ km$, avec un écart type $\sigma_0 = 0.21\ell/100\ km$, pour des véhicules d'un type donné. Un organisme indépendant suspecte une sous-estimation de cette consommation et indique que la consommation moyenne s'élèverait à $\mu_1 = 6.45\ell/100\ km$. Qui a raison ? La question est d'importance, parce qu'elle peut mener au paiement de pénalités si le constructeur a, même non intentionnellement, sous-estimé la consommation de ses véhicules. Il s'agit donc d'un problème de décision entre deux hypothèses :

- (H_0) une hypothèse communément admise : la consommation moyenne des véhicules d'un type donné est de $\mu_0 = 6.32\ell/100\ km$, avec un écart type $\sigma_0 = 0.21\ell/100\ km$.
- (H_1) contre une hypothèse en compétition : la consommation pourrait être plus importante, et valoir $\mu_1 = 6.45\ell/100\ km$

Il n'est pas envisageable de mesurer la consommation de tous les véhicules d'un type donné, et la décision va être prise à partir d'un échantillon de $n = 30$ véhicules. La consommation moyenne observée sur cet échantillon est $\bar{x} = 6.43\ell/100\ km$, supérieure à la valeur nominale μ_0 annoncée par le constructeur. Mais de part la taille finie de l'échantillon, cette différence est-elle simplement due à la fluctuation naturelle de l'échantillonnage ou le constructeur a-t-il sous-estimé la consommation de ses véhicules ? La **théorie des tests statistiques** offre un cadre mathématique à ce problème **décisionnel**.

Dans la théorie des tests de Neyman-Pearson que nous allons présenter, le choix entre les deux hypothèses (H_0) et (H_1) se fera en en assumant le **risque de première espèce** α (défini a priori à un faible niveau 5%, 10%,..), probabilité de choisir (H_1) alors que (H_0) est vraie. La construction d'un test suit les étapes suivantes :

1. **Modélisation** : la consommation X_i d'un véhicule suit une loi gaussienne $\mathcal{N}(\mu_0, \sigma_0^2)$. Les véhicules sont tirés de façon indépendantes dans le parc de l'ensemble des véhicules du type considéré.

2. On définit les hypothèses (H_0) et (H_1)
3. On considère la **statistique** $\bar{X}$, moyenne des consommations sur un n-échantillon. Sa **loi sous** (H_0) est gaussienne d'après les hypothèses :

$$\bar{X} \sim \mathcal{N}(\mu_0, \frac{\sigma^2}{n}) \text{ soit } T = \sqrt{n}\frac{\bar{X} - \mu_0}{\sigma_0} \sim \mathcal{N}(0,1)$$

4. On définit la procédure de décision en définissant une **région de rejet** $\mathcal{R}_\alpha$, plage de valeurs de probabilité α indiquant la décision de rejet du test. Ici, la région de rejet pour T est

$$\mathcal{R}_\alpha =]q^*_{1-\alpha}; +\infty[$$

avec $q^*_{1-\alpha}$ le quantile d'une loi $\mathcal{N}(0,1)$ d'ordre $1-\alpha$. En effet,

$$\mathbb{P}_{H_0}(T \in \mathcal{R}_\alpha) = \mathbb{P}_{H_0}(T > q^*_{1-\alpha}) = \alpha$$

Par abus de langage, on dit parfois que l'événement $\{T > q^*_{1-\alpha}\}$ est également la région de rejet du test. Si on ne remet pas en cause l'information sur la variance σ_0^2, on peut aussi définir sur cet exemple la région de rejet de risque α sur $\bar{X}$ par $]\mu_0 + q^*_{1-\alpha}\frac{\sigma_0}{\sqrt{n}}; +\infty[$, puisque

$$\mathbb{P}_{H_0}\Big(\bar{X} > \mu_0 + q^*_{1-\alpha}\frac{\sigma_0}{\sqrt{n}}\Big) = \alpha$$

Le risque (dit de **première espèce**) α est défini a priori. Il représente la probabilité de rejet de (H_0) à tort, ce sont des cas "extrêmes" qui peuvent arriver sous (H_0), mais avec une faible probabilité.

5. Il reste à décider en fonction de la valeur de la statistique observée sur l'échantillon
   — si la statistique de test est dans la région de rejet, on **rejette** H_0 : les données sont significatives, ie suffisamment différentes des valeurs habituellement observées sous (H_0).
   — sinon, on **accepte** - on dit aussi on **conserve** - (H_0), faute de preuves suffisantes : les données ne sont pas significatives pour incriminer une consommation plus élevée.

Dans notre exemple $\bar{x} = 6.43\ell/100\ km$ soit

$$t_{obs} = \sqrt{30}(6.43 - 6.32)/0.21 = 2.86 > 1.64$$

Donc, en ayant défini un risque de première espèce de $\alpha = 5\%$, on a $q^*_{1-\alpha} = 1.64$, et on rejette (H_0). Les données sont significatives pour dire que le constructeur a sous-estimé la consommation, et on considère que ses véhicules consomment plus qu'annoncé.

Supposons maintenant que la même moyenne a été observée $\bar{x} = 6.435\ell/100\ km$ pour un échantillon de $n = 9$ véhicules de même consommation nominale de $\mu_0 = 6.32\ell/100\ km$. Dans ce cas,

$$t_{obs} = \sqrt{9}(6.43 - 6.32)/0.21 = 1.57 < 1.64$$

On accepte dans ce cas (H_0), le constructeur ne peut pas être mis en défaut, bien que la valeur ponctuelle de l'estimation soit supérieure à la valeur nominale. Il y a alors deux cas (qu'on ne sait pas départager) :

— le constructeur respecte effectivement la norme, et la décision a été prise à raison
— le constructeur ne respecte pas la norme mais il n'a pas été mis en défaut parce que l'échantillon ne donne pas une information suffisamment précise : la variabilité naturelle du phénomène masque le problème et il faudrait prendre un plus grand échantillon pour s'en rendre compte. La décision a été prise à tort, et la probabilité de conserver (H_0) alors que (H_1) est vraie s'appelle **risque de seconde espèce**.

6.2 Construction d'un test

Définition 6.1. *Un **test** est une procédure de décision qui permet de trancher, au vu des résultats d'un échantillon $X = (X_1, \ldots, X_n)$, entre deux hypothèses - l'**hypothèse nulle** (H_0) et une hypothèse **alternative** (H_1) - dont une seule est vraie.*
Un test peut s'écrire comme une fonction de l'échantillon qui ne peut prendre que deux valeurs : 0 pour accepter (H_0), et 1 pour rejeter (H_0) :

$$X = (X_1, \ldots, X_n) \mapsto \varphi(X) \in \{0, 1\}$$

*La **région critique** $\mathcal{R}$ (ou région de **rejet** du test) est l'ensemble des valeurs de la variable de décision $T(X)$ qui conduisent à écarter (H_0) au profit de (H_1).*

$$\varphi(X) = \mathbb{1}_{T(X) \in \mathcal{R}}$$

*La région d'**acceptation** du test est $\overline{\mathcal{R}}$.*

La construction d'un test suit les étapes que nous avons déroulées dans l'exemple introductif :

1. Définir le **modèle**
2. Définir les **hypothèses nulle** (H_0) et **alternative** (H_1)
3. Choisir une **statistique de test** $T(X)$, déterminer sa **loi sous** (H_0)
4. Définir la **règle de décision** en calibrant la région de rejet $\mathcal{R}$ suivant le risque α. Calculer le risque de seconde espèce.
5. Calculer la statistique de test observée et **décider** : rejet ou acceptation de (H_0).

La décision du test est prise à partir de la valeur observée t_{obs} de la statistique de test T. Cette décision dépend de l'échantillon, mais la fonction de test est déterministe : pour une valeur observée $\bar{x}$, le test prend toujours la même décision. Il est possible de construire des tests à fonction de test aléatoire, que nous ne développerons pas dans ce cours.

A l'issue du test, quatre situations sont donc possibles, suivant la véracité réelle de (H_0) - inconnue quand on fait le test - et la décision prise - acceptation ou rejet de (H_0). La décision est bonne si (H_0) est vraie et qu'on a accepté (H_0) ou si (H_0) et fausse et qu'elle a été rejetée. Elle est mauvaise si on rejette (H_0) alors qu'elle est vraie (erreur de première espèce) ou si on accepte (H_0) alors qu'elle est fausse (erreur de seconde espèce).

Définition 6.2. *Le risque de première espèce, noté α, est l'espérance de l'erreur de première espèce sous (H_0)*

$$\alpha = \mathbb{E}_{(H_0)}(\varphi(T(X)) = \mathbb{P}_{(H_0)}(T \in \mathcal{R})$$

Le risque de seconde espèce, noté β, est l'espérance de l'erreur de seconde espèce sous (H_1)

$$\beta = \mathbb{E}_{(H_1)}(\varphi(T(X)) = \mathbb{P}_{(H_1)}(T \in \overline{\mathcal{R}})$$

*On appelle **puissance** du test que l'on note π, la probabilité sous (H_1) de rejeter (H_0)*

$$\pi = 1 - \beta$$

La puissance est la capacité du test à détecter l'alternative. Les différentes situations sont récapitulées dans le tableau suivant :

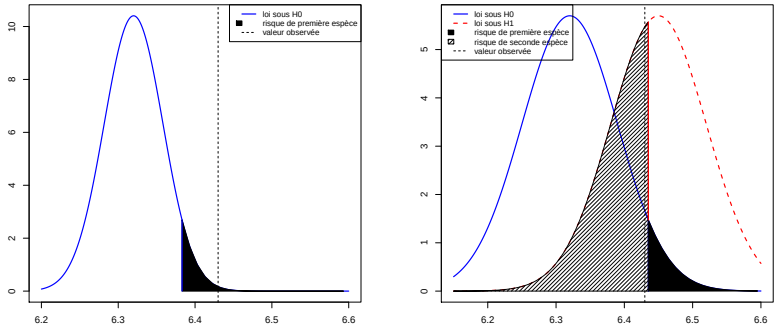


FIGURE 6.1 – Repésentation schématique du risque de première espèce dans le cas où $n = 30$ (à gauche) et des risques de première et seconde espèce dans le cas $n = 9$ (à droite)

	Choix (H_0)	Choix (H_1)
(H_0) vraie	$1-\alpha$ bonne décision	$\alpha = \mathbb{P}_{(H_0)}(T \in \mathcal{R})$ risque première espèce mauvaise décision
(H_1) vraie	$\beta = \mathbb{P}_{(H_1)}(T \notin \mathcal{R})$ risque seconde espèce mauvaise décision	$\pi = 1-\beta$ puissance bonne décision

La figure 6.1 représente sur la graphe de gauche le risque de l'erreur de première espèce (cas $n = 30$). Ce risque ne dépend pas de la valeur choisie pour μ_1 de l'hypothèse alternative. La valeur observée est dans la région de rejet de risque $\alpha = 5\%$, (H_0) est rejetée avec un risque de première espèce de 5%. Dans le cas $n = 9$, le test a accepté (H_0). Supposons que la variance sous (H_0) et (H_1) soient identiques et que la loi sous l'alternative est également gaussienne. Il est alors possible de calculer le risque de seconde espèce, matérialisé par les hachures sur le graphe de droite de la figure 6.1.

$$\begin{aligned}\beta = \mathbb{P}_{(H_1)}(T \in \overline{\mathcal{R}}) &= \mathbb{P}_{(H_1)}\left(\sqrt{n}\frac{\bar{X}-\mu_0}{\sigma_0} < q^*_{1-\alpha}\right)\\ &= \mathbb{P}_{(H_1)}\left(\sqrt{n}\frac{\bar{X}-\mu_1}{\sigma_0} < q^*_{1-\alpha} + \sqrt{n}\frac{\mu_0-\mu_1}{\sigma_0}\right)\\ &= F^*\left(q^*_{1-\alpha} + \sqrt{n}\frac{\mu_0-\mu_1}{\sigma_0}\right)\end{aligned}$$

où F^* est la fonction de répartition de $\mathcal{N}(0,1)$. L'application numérique donne ici $\beta = 0.41$. Le test est peu puissant à détecter (H_1), puisqu'il manque de rejeter (H_0) dans 40% des cas quand les données sont générées sous (H_1).

On voit que le risque de seconde espèce dépend de la valeur de μ_1 défini dans l'hypothèse alternative. Il y a dissymétrie de la situation de test envers les deux hypothèses :

— (H_0) et (H_1) ne sont pas interchangeables : le risque n'est contrôlé que pour (H_0) qui est l'hypothèse communément admise, (H_1) est celle dont on souhaite calibrer l'erreur d'un choix à tort. La véritable décision est celle qui rejette (H_0).
Exemple : pour tester la nocivité d'un nouveau médicament, il est préférable de choisir *dangereux* pour (H_0), et *inoffensif* pour (H_1), afin de ne pas choisir l'hypothèse d'inoffen-

sivité sans en calibrer le risque. Peut-être qu'on décidera qu'un médicament est dangereux avec un risque de seconde espèce important, et donc qu'on ne validera pas un médicament sans danger (et potentiellement efficace), mais c'est préférable à choisir l'inoffensité pour (H_0), qui pourrait amener à déclarer inoffensif un médicament avec un risque de seconde espèce non maîtrisé (à quel risque il pourrait être dangereux alors qu'il a étiqueté inoffensif).
Pour tester l'efficacité d'un médicament, il est pour les mêmes raisons préférable de choisir *inefficace* pour (H_0), et *efficace* pour (H_1), afin de ne mettre sur le marché que des médicaments qui ont significativement montré leur efficacité plutôt que d'autoriser des médicaments qui sont considérés par défaut comme efficaces.

— Il faut connaître la loi de la statistique de test sous (H_0) afin de déterminer le risque de première espèce qui calibre le test.
— Il faut que cette loi soit différente sous (H_1) pour que le test soit puissant à détecter l'alternative.
— Entre deux tests de même risque de première espèce α, il faut choisir le plus puissant. Par exemple, la région de rejet de la forme $\{T < q^*_{0.05}\}$ est aussi de risque 5%, mais elle n'a aucune puissance pour détecter le cas $\mu_1 > \mu_0$ et n'est donc pas intéressante pour tester une éventuelle sous-estimation de la consommation.

6.3 Hypothèses simples et composites

Les hypothèses précédentes sont **simples** : une valeur μ_0 pour (H_0) et une valeur μ_1 pour (H_1). En fait, il est souvent peu fréquent de souhaiter tester une valeur. Au contraire, on peut définir des hypothèses par des plages de valeurs. Par exemple dans le cas de la consommation des véhicules, on peut traduire la question de l'organisme indépendant par le test de l'hypothèse nulle (H_0) : *la consommation moyenne est inférieure au seuil* $\mu_0 = 6.32$ *indiqué par le constructeur*, contre l'hypothèse alternative (H_1) : *la consommation est supérieure à ce seuil* :

$$(H_0) : \mu \leq \mu_0 \text{ contre } (H_1) : \mu > \mu_0$$

Le hypothèses portant sur des ensembles de valeurs s'appellent des hypothèses **composites**.

6.3.1 Hypothèse alternative composite

Si l'on souhaite tester une hypothèse nulle simple contre une alternative composite

$$(H_0) : \theta = \theta_0 \text{ contre } (H_1) : \theta > \theta_0$$

on utilise la même région de rejet que pour le test d'hypothèse simple

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

En effet, l'erreur de première espèce $\alpha = \mathbb{P}_{(H_0)}(T \in \mathcal{R})$ est la même pout tout $\theta_1 > \theta$ mais la puissance devient une **fonction** de θ :

$$\text{Pour tout } \theta_1 \in \Theta_1 = \{\theta | \theta > \theta_0\}, \quad \pi(\theta_1) = \mathbb{P}_{\theta_1}(T \in \mathcal{R}) = 1 - \beta(\theta_1).$$

6.3.2 Hypothèse nulle composite

De même, pour tester

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

on utilise la même région de rejet que pour le test d'hypothèse simple

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

Le risque de première espèce devient une **fonction** de θ :

$$\text{Pour tout } \theta \in \Theta_0 = \{\theta | \theta \leq \theta_0\}, \quad \alpha(\theta) = \mathbb{P}_\theta(T \in \mathcal{R})$$

Dans notre exemple, pour $\theta \in \Theta_0$, $\alpha(\theta) \leq \alpha(\theta_0)$, et $\alpha(\theta_0) = \sup_{\alpha \in \Theta_0} \alpha(\theta)$ est la **taille** du test. Un test est de **niveau** α si sa taille $\leq \alpha$. On peut donc envisager les combinaisons d'hypothèses suivantes :

— Hypothèses simples

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

— Test **unilatéral** pour une hypothèse nulle composite

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

— Test **unilatéral** pour une hypothèse nulle composite

$$(H_0) : \theta \geq \theta_0 \text{ contre } (H_1) : \theta < \theta_0$$

— Test **bilatéral** pour une hypothèse nulle simple

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

De façon générale, on écrira les hypothèses sous la forme

$$(H_0) : \theta \in \Theta_0 \text{ contre } (H_1) : \theta \in \Theta_1 = \Theta \setminus \Theta_0$$

6.4 Propriétés des tests

Nous avons déjà rencontré la propriété de puissance, capacité du test à détecter (H_1). La puissance est une fonction du paramètre quand l'hypothèse alternative est composite. Entre deux tests de niveau fixé a priori, on choisit bien sûr le plus puissant.

Nous présentons ici deux autres propriétés, la biais et la convergence

Définition 6.3. *On considère un test statistique d'une hypothèse nulle contre une hypothèse alternative. Soit α le risque de première espèce, β le risque de seconde espèce et π la puissance.*

— *Le test est **sans biais** si $1 - \beta(\theta) = \pi(\theta) > \alpha$ pour tout $\theta \in \Theta_1$*

— *Le test est **consistant** (ou convergent) si la suite des puissance $\pi_n(\theta)$ tend vers 1 quand la taille n de l'échantillon tend vers l'infini*

La première propriété indique que la probabilité de détecter (H_1) à raison est supérieure à celle de rejeter (H_0) à tort ; la deuxième est une propriété asymptotique : c'est la capacité du test à détecter systématiquement l'alternative à raison pour un échantillon de taille l'infinie. On cherchera donc à construire des tests sans biais, consistant et de puissance maximale parmi les tests de même niveau.

6.5 Cadre de Neyman-Pearson

Le cadre de la théorie de Neyman-Pearson permet de construire des tests les plus puissants parmi les tests de niveau fixé

Définition 6.4. *Un test est* ***uniformément plus puissant*** *(UPP) si, quelle que soit la valeur de θ, sa puissance $\pi(\theta)$ est supérieure à la puissance de tout autre test de niveau α.*

Neyman et Pearson ont déterminé la forme de la région de rejet d'un tel test dans le cas du test de deux hypothèses simples.

Théorème 6.5 (Neyman-Pearson)**.** *Soit $L(\theta; x)$ la vraisemblance des observations. La région critique optimale (au sens UPP) du test de $\theta = \theta_0$ vs $\theta = \theta_1$ au niveau α est définie par*

$$\mathcal{R}_\alpha = \left\{ x \in \mathbb{R}^n; \frac{L(\theta_1; x)}{L(\theta_0; x)} > k_\alpha \right\}$$

Preuve. Nous ferons la preuve dans le cas d'une loi continue. On commence par noter que $RV(X) = L(\theta_1; x)/L(\theta_0; x)$ est bien une statistique parce que θ_1 et θ_0 sont donnés.

Supposons que k_α existe. Soit $\mathcal{R}_\alpha$ une autre région de rejet quelconque de niveau α. Alors

$$\mathbb{P}_{\theta_0}(X \in \mathcal{R}_\alpha) = \alpha$$

Il y a égalité des niveaux des deux régions :

$$\begin{aligned}
\mathbb{P}_{\theta_0}(X \in \mathcal{R}_\alpha) &= \mathbb{P}_{\theta_0}(X \in \mathcal{R}_\alpha \setminus \mathcal{R}_{opt}) + \mathbb{P}_{\theta_0}(X \in \mathcal{R}_\alpha \cap \mathcal{R}_{opt}) = \alpha \\
\mathbb{P}_{\theta_0}(X \in \mathcal{R}_{opt}) &= \mathbb{P}_{\theta_0}(X \in \mathcal{R}_{opt} \setminus \mathcal{R}_\alpha) + \mathbb{P}_{\theta_0}(X \in \mathcal{R}_\alpha \cap \mathcal{R}_{opt}) = \alpha
\end{aligned}$$

On en déduit $\mathbb{P}_{\theta_0}(X \in \mathcal{R}_\alpha \setminus \mathcal{R}_{opt}) = \mathbb{P}_{\theta_0}(X \in \mathcal{R}_\alpha \setminus \mathcal{R}_{opt})$, soit

$$\int_{x \in \mathcal{R}_\alpha \setminus \mathcal{R}_{opt}} L(\theta_0, x)\, dx = \int_{x \in \mathcal{R}_{opt} \setminus \mathcal{R}_\alpha} L(\theta_0, x)\, dx \tag{6.1}$$

On compare maintenant les puissances :

$$\begin{aligned}
\pi(\mathcal{R}_{opt}) = \mathbb{P}_{\theta_1}(X \in \mathcal{R}_{opt}) &= \int_{x \in \mathcal{R}_{opt} \setminus \mathcal{R}_\alpha} L(\theta_1, x)\, dx + \int_{x \in \mathcal{R}_{opt} \cap \mathcal{R}_\alpha} L(\theta_1, x)\, dx \\
\pi(\mathcal{R}_\alpha) = \mathbb{P}_{\theta_1}(X \in \mathcal{R}_\alpha) &= \int_{x \in \mathcal{R}_\alpha \setminus \mathcal{R}_{opt}} L(\theta_1, x)\, dx + \int_{x \in \mathcal{R}_{opt} \cap \mathcal{R}_\alpha} L(\theta_1, x)\, dx
\end{aligned}$$

Il reste à comparer le premier terme de chaque somme

$$\begin{aligned}
\int_{x \in \mathcal{R}_\alpha \setminus \mathcal{R}_{opt}} L(\theta_1, x)\, dx &\leq \int_{x \in \mathcal{R}_\alpha \setminus \mathcal{R}_{opt}} k_\alpha L(\theta_0, x)\, dx = \int_{x \in \mathcal{R}_{opt} \setminus \mathcal{R}_\alpha} k_\alpha L(\theta_0, x)\, dx \text{ en utilisant (6.1)} \\
&< \int_{x \in \mathcal{R}_{opt} \setminus \mathcal{R}_\alpha} L(\theta_1, x)\, dx
\end{aligned}$$

donc, $\pi(\mathcal{R}_\alpha) < \pi(\mathcal{R}_{opt})$. Soit $A(k) = \{x | L(\theta_1, x) > kL(\theta_0, x)\}$. L'application $k \to \mathbb{P}_{\theta_0}(A(k))$ est une fonction monotone, continue car la loi de X est continue. De plus $\mathbb{P}_{\theta_0}(A(0)) = 1$ et $\lim_{k \to +\infty} \mathbb{P}_{\theta_0}(A(k)) = 0$. Par le théorème des valeurs intermédiaires, il existe k_α tel que $\mathbb{P}_{\theta_0}(A(k_\alpha)) = \alpha$ ◊

Application Le test utilisé en introduction est UPP. On rappelle que le modèle est gaussien. Sous (H_0) les données suivent une loi $\mathcal{N}(\mu_0, \sigma_0^2)$ et sous (H_1) une loi $\mathcal{N}(\mu_1, \sigma_0^2)$ car on a supposé la même variance. La vraisemblance d'un échantillon gaussien $\mathcal{N}(\mu, \sigma^2)$ s'écrit

$$L(\mu, x) = \frac{1}{(2\pi\sigma)^{n/2}} \exp - \frac{\sum (x_i - \mu)^2}{2\sigma^2}$$

Il s'agit donc de trouver un constante k_α qui dépend du niveau α telle que

$$\begin{aligned}
\mathcal{R}_\alpha &= \left\{ x \in \mathbb{R}^n ; \frac{L(\theta_1; x)}{L(\theta_0; x)} > k_\alpha \right\} \\
&= \left\{ x \in \mathbb{R}^n ; \exp \frac{\mu_1 - \mu_0}{\sigma_0^2} \bar{x} > \frac{k_\alpha}{n} + \frac{\mu_1^2 - \mu_0^2}{2\sigma_0^2} \right\} \\
&= \left\{ x \in \mathbb{R}^n ; (\mu_1 - \mu_0)\bar{x} > \sigma_0^2 \log \left(\frac{k_\alpha}{n} + \frac{\mu_1^2 - \mu_0^2}{2\sigma_0^2} \right) \right\}
\end{aligned}$$

Le rapport de vraisemblance ne dépend donc des observations que par $\bar{x}$, la statistique de test concernée est $\bar{X}$ et le terme de droite de l'inégalité est une constante. De plus $\mu_1 > \mu_0$ sous l'alternative. On cherche donc une constante $\tilde{k}_\alpha$ telle que

$$\mathbb{P}_{\mu_0}(\bar{X} > \tilde{k}_\alpha) = \alpha$$

c'est à dire une constante $\tilde{\tilde{k}}_\alpha$ telle que

$$\mathbb{P}_{\mu_0} \left(\sqrt{n} \frac{\bar{X} - \mu_0}{\sigma_0} > \tilde{\tilde{k}}_\alpha \right) = \alpha$$

où l'on identifie immédiatement $\tilde{\tilde{k}}_\alpha = q^*_{1-\alpha}$. Le test que nous avons utilisé dans l'introduction est donc UPP. Ici, bien que la statistique du rapport de vraisemblance ne soit pas simple, le fait qu'elle soit une fonction **monotone** de $\bar{X}$ dont on connait la loi sous (H_0) suffit à construire le test. Pour tester $(H_0) : \mu = \mu_0$ contre $(H_1) : \mu = \mu_1 < \mu_0$, le théorème de Neyman-Pearson indique que la région de rejet est telle que

$$\mathbb{P}_{\mu_0} \left(\sqrt{n} \frac{\bar{X} - \mu_0}{\sigma_0} < q^*_\alpha \right) = \alpha$$

Il est immédiat de constater que l'extension de ce test à celui d'hypothèses composites $(H_0) : \mu \leq \mu_0$ contre $(H_1) : \mu > \mu_0$ est également UPP de niveau α, où α est le risque du test des hypothèses simples.

Si nous avons trouvé ici un test UPP, il n'en existe par toujours. Par exemple il n'existe pas de test UPP pour tester $(H_0) : \mu = \mu_0$ contre $(H_1) : \mu \neq \mu_0$. En effet, la région de rejet UPP pour l'alternative $(H_1) : \mu > \mu_0$ est de la forme $T > q^*_{1-\alpha}$ tandis que celle pour l'alternative $(H_1) : \mu < \mu_0$ est de la forme $T < q^*_\alpha$. La figure 6.2 superpose la puissance du test unilatéral de région de rejet à droite avec celle du test bilatéral de même niveau. On remarque que le test bilateral est de puissance inférieure au test unilatéral UPP pour $\mu > \mu_0$. Il a une forme symétrique par rapport à $\mu = \mu_0$, ce qui est conjoncturel à ce cas particulier.

6.6 p-value

La décision du test fait intervenir la valeur de la statistique de test observée t_{obs} et le quantile de la loi de la statistique de test sous (H_0). Il existe une méthodologie équivalente qui raisonne sur les probabilités plutôt que sur les quantiles.

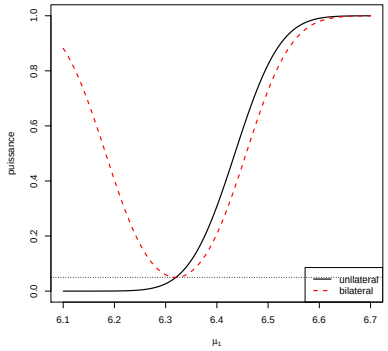


FIGURE 6.2 – Puissances d'un test unilatéral ou bilatéral de même niveau

Définition 6.6. *Soit la fonction test $\varphi(x;\alpha)$ associée à la région de rejet $\mathcal{R}_\alpha$ de niveau α. On appelle* ***probabilité critique ou p-value*** *le plus petit niveau qui fait rejeter (H_0) au vu des données :*

$$P_c(t_{obs}) = \inf\{\alpha \in [0,1]; \varphi(x_{obs};\alpha) = 1\}$$

La p-value est une variable aléatoire.

Par exemple dans le test de $\mu = \mu_0$ contre $\mu > \mu_0$ à variance connue, la région de rejet s'écrit

$$\mathcal{R} = \left\{ t; t = \sqrt{n}\frac{\bar{x} - \mu_0}{\sigma_0} > q^*(1-\alpha) \right\}.$$

Elle est de niveau : $\mathbb{P}_{H_0}(T(X) \in \mathcal{R}) = \alpha$. La valeur observée de la statistique de test : $t_{obs} = T(x_{obs})$ et la p-value est $P_c(t_{obs}) = \mathbb{P}_{H_0}(T(X) > t_{obs})$ Donc,

— si $P_c(t_{obs}) \leq \alpha$, c'est que t_{obs} est dans la région de rejet de (H_0)
— si $P_c(t_{obs}) > \alpha$, c'est que t_{obs} est dans la région d'acceptation de (H_0)

La règle de décision associée au test de niveau α est donc

$$\text{rejet de } (H_0) \Leftrightarrow \text{p-value} < \alpha$$

Dans un test de niveau α, (H_0) est rejetée si $\alpha >$ p-value, conservée si $\alpha <$ p-value. On peut utiliser les précisions suivantes :

— si $0.05 >$ p-value > 0.01, le test est significatif,
— si $0.01 >$ p-value > 0.001, le test est très significatif,
— si $0.001 >$ p-value, le test est hautement significatif.

Exemple Dans le cas de l'exemple introductif avec $n = 30$, l'application numérique pour n=30 est :

$$tobs = 2.86; \quad \text{p-value} = 1 - F^*(tobs) = 0.002$$

La pvalue vaut donc $0.002 < 5\%$, le test est très significatif et on rejette (H_0) au niveau 5%. Dans le cas $n = 9$,

$$tobs = 1.57; \quad \text{p-value} = 1 - F^*(tobs) = 0.058$$

la p-value est (légèrement) supérieure à 5%, il n'est pas possible de rejeter (H_0) avec un risque de première espèce maximum de 5%, et on conserve (accepte) (H_0) avec un risque de seconde espèce (inconnu si l'alternative est composite).

Remarque La p-value est la probabilité d'observer des valeurs "plus extrêmes" que celles de l'échantillon observé sachant que (H_0) est vraie. Ce n'est pas n'est pas la probabilité que (H_0) soit vraie sachant qu'on a observé l'échantillon. Il n'est pas possible de calculer cette dernière à partir de la seule connaissance de la p-value.

Nous finissons cette présentation des p-values par l'exemple du calcul de la p-valeur d'un test bilatère de la moyenne de $\mu = \mu_0$ contre $\mu \neq \mu_0$ à variance connue. En prenant en compte la symétrie de la statistique de test $T(X) = \sqrt{n}\frac{\bar{X}-\mu_0}{\sigma_0}$, la région de rejet de niveau $\mathbb{P}_{(H_0)}(T(X) \in \mathcal{R}) = \alpha$ s'écrit

$$\mathcal{R} = \{x; |T(x)| > q^*(1-\alpha/2)\}\,.$$

La valeur observée de la statistique de test est $t_{obs} = T(x_{obs})$ et la p-value

$$\mathbb{P}_c(t_{obs}) = \mathbb{P}_{H_0}(|T(X)| > |t_{obs}|)$$

— si t_{obs} est supérieur à la médiane de T :

$$P_c(t_{obs}) = 2\ \mathbb{P}_{H_0}(T(X) > t_{obs})$$

— si t_{obs} est inférieur à la médiane de T :

$$P_c(t_{obs}) = 2\ \mathbb{P}_{H_0}(T(X) < t_{obs})$$

De façon générale, la p-value peut s'écrire

$$P_c(t_{obs}) = 2\ \mathbb{P}_{H_0}(T(X) < -|t_{obs}|)$$

Dans le cas de l'exemple ($t_{obs} = 1.57$, $n = 9$), p-value$= 2(1 - F^*(tobs)) = 0.116$)

6.7 Tests paramétriques usuels

6.7.1 Test de Student (dit de la moyenne)

Dans l'exemple introductif, nous avions supposé connaître la valeur de la variance. Si ce n'est pas le cas, il est possible de normaliser la statistique de test par l'écart type estimé de l'échantillon, ce qui amène au test (par exemple unilatéral) de l'espérance d'une loi $\mathcal{N}(\mu, \sigma^2)$ à variance inconnue $(H_0) : \mu = \mu_0$ contre $(H_1) : \mu > \mu_0$: la statistique de test est une statistique **pivotale** (c'est à dire dont la loi ne dépend d'aucun paramètre) : elle suit sous (H_0) une loi de Student à $n-1$ degrés de liberté :

$$T_n = \sqrt{n}\frac{\bar{X} - \mu_0}{\sqrt{\sum_i (X_i - \bar{X})^2/(n-1)}} \sim_{(H_0)} \mathcal{T}(n-1)$$

La région de rejet de niveau $\alpha = \mathbb{P}_{(H_0)}(\mathcal{R})$ est définie par $\mathcal{R} = \{T > qt(n-1, 1-\alpha)\}$ ou qt désigne le quantile d'une loi de Student.

Si on ne peut faire l'hypothèse gaussienne, mais que l'échantillon est suffisamment grand, il est possible d'utiliser une approximation asymptotique, les niveau et puissance calculés sont approximativement les niveau et puissance réels.

6.7.2 Test de la variance d'une loi gaussienne

Il s'agit de tester $(H_0) : \sigma = \sigma_0$ contre $(H_1) : \sigma \neq \sigma_0$ sur un échantillon de loi gaussienne d'espérance et variance inconnues. Sous (H_0), on a

$$C_n = \frac{\sum(X_i - \bar{X})^2}{\sigma_0^2} \sim \chi^2(n-1)$$

et C_n est une statistique de test pivotale qui suit sous (H_0) une loi du Khi-deux à $n-1$ degrés de liberté. On a

$$\mathbb{P}_{(H_0)}\left(qchisq(\alpha/2, n-1) < \frac{\sum(X_i - \bar{X})^2}{\sigma_0^2} < qchisq(1-\alpha/2, n-1)\right) = 1-\alpha$$

où $qchisq$ désigne le quantile d'une loi du Khi-deux. La région d'acceptation de niveau $1-\alpha$ est donc définie par $[qchisq(\alpha/2, n-1); qchisq(1-\alpha/2, n-1)]$. Son complémentaire est une région de rejet de niveau α.

Si on ne peut faire l'hypothèse gaussienne, mais que l'échantillon est suffisamment grand, il est possible d'utiliser une approximation asymptotique, les niveau et puissance calculés sont approximativement les niveau et puissance réels.

6.7.3 Test de comparaison des moyennes deux échantillons

C'est un autre test qui utilise une loi de Student quand les échantillons sont gaussiens.
voit TD et section 7.4.2

6.7.4 Test de comparaison des variances de deux échantillons

C'est un test qui utilise une loi de Fisher quand les échantillons sont gaussiens.
voit TD et section 7.4.3

6.8 Tests non paramétriques

On peut vouloir tester plus généralement la forme de la loi, et pas seulement la valeur d'un de ses paramètres. C'est typiquement le cas pour les tests dit d'adéquation, qui font souvent appel à des tests non-paramétriques.

La même méthodologie peut s'appliquer : définition du modèle, des hypothèses, de la loi sous (H_0), de la région de rejet. Sans forcément connaître le détail du test, il est possible d'en interpréter le résultat, en général en utilisant la p-value.

Des exemples de tests d'adéquation sont présentés au dernier chapitre, par exemple le test de Shapiro-Wilks (section8.3.3) pour tester le caractère gaussien d'une loi, ou le test de Kolmogorov-Smirnov (section 8.3.2) pour tester l'adéquation à une loi continue donnée.

Chapitre 7

Intervalle de confiance

Si le test nous permet de prendre un décision en étant conscient de son risque, il n'est pas aisé pour rendre compte de la précision de l'estimation. L'intervalle de confiance est une réponse à cette question, que nous avons déjà abordée en introduction 4.1 du chapitre sur l'estimation.

Nous verrons que la définition d'un test peut permettre celle d'un intervalle de confiance, et réciproquement. Il s'agit donc d'un autre angle de vue

Ce chapitre inclut des traces du logiciel R pour les applications numériques afin donner une idée de la façon dont les logiciels présentent les résultats. Vous pouvez sauter ces passages si vous souhaitez, et bien sûr faire les calculs à main (calculette), et vérifier les résultats.

7.1 Un autre angle de vue

Soit le test $(H_0) : \mu = \mu_0$ contre $(H_0) : \mu \neq \mu_0$ de l'espérance d'une loi gaussienne $\mathcal{N}(\mu, \sigma^2)$ à variance σ^2 connue. On accepte (H_0) au niveau α si et seulement si

$$|T(X)| = \left|\sqrt{n}\frac{\bar{X} - \mu_0}{\sigma}\right| \leq qnorm(1-\alpha/2) = q^*_{1-\alpha/2}$$

soit

$$-q^*_{1-\alpha/2} \leq \sqrt{n}\frac{\bar{X} - \mu_0}{\sigma} \leq q^*_{1-\alpha/2}$$

qu'on peut aussi écrire

$$\underbrace{\bar{X} - q^*_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\widehat{\mu}_{inf}} \leq \mu_0 \leq \underbrace{\bar{X} + q^*_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}}_{\widehat{\mu}_{sup}}$$

et

$$1 - \alpha = \mathbb{P}\left(|T(X)| \leq q_{1-\alpha/2}\right) = \mathbb{P}\left([\widehat{\mu}_{inf}; \widehat{\mu}_{sup}] \ni \mu_0\right)$$

Ainsi, pour qu'une valeur hypothétique de μ soit acceptée, il faut et il suffit qu'elle soit dans l'intervalle

$$IC(\mu) = [\widehat{\mu}_{inf}; \widehat{\mu}_{sup}]$$

Cet intervalle $IC(\mu)$ aux bornes aléatoires est appelé **intervalle de confiance** de niveau $1 - \alpha$ de l'espérance μ inconnue. Dans cet exemple, il y a équivalence pour μ entre prendre une valeur acceptée (H_0) dans le test de niveau α et le fait pour le μ_0 de l'hypothèse nulle d'être situé dans

l'intervalle de confiance de niveau (de confiance) $1-\alpha$. Rappelons que α est la probabilité de rejeter (H_0) à tort, et donc $1-\alpha$ est donc la probabilité d'accepter (H_0) à raison.

Le fait de fournir un **intervalle** (on dit aussi une **fourchette**, vocabulaire venant des sondages) permet de rendre en compte de la fluctuation d'échantillonnage plutôt que de donner une valeur ponctuelle $\widehat{\mu}$.

7.2 Définition et interprétation

Définition 7.1. *Soit $X=(X_1,\dots,X_n)$ un n-échantillon de loi $\mathbb{P}_\theta$, où $\theta\in\Theta\subset\mathbb{R}$ est inconnu. Un **intervalle de confiance** de **niveau** $1-\alpha$ pour θ est un intervalle $IC=[\widehat{\theta}_{inf}(X),\widehat{\theta}_{sup}(X)]$ dont les bornes sont **aléatoires**, telles que, pour tout $\theta\in\Theta$*

$$P_\theta(IC\ni\theta)\geq 1-\alpha.$$

où α est "petit".
Une réalisation $[\widehat{\theta}_{inf}(x),\widehat{\theta}_{sup}(x)]$ est obtenue à partir des données $x=(x_1,\dots,x_n)$.

Revenons sur l'exemple gaussien $X_i\sim\mathcal{N}(\mu,\sigma^2)$, de variance σ^2 connue pour déterminer s'il n'y a pas d'autres façons de déterminer les bornes aléatoires de l'intervalle de confiance de μ (Ici, μ joue le rôle du θ de la définition.

On cherche donc α_1 et α_2 tels que $0\leq\alpha_1,\alpha_2\leq\alpha$ et $\alpha_1+\alpha_2=\alpha$. Soit q^* la fonction quantile de $\mathcal{N}(0,1)$. Un intervalle de probabilité $1-\alpha$ de $T=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim\mathcal{N}(0,1)$ est $[q^*_{\alpha_1};q^*_{1-\alpha_2}]$

$$\begin{aligned}1-\alpha&=\mathbb{P}(q^*_{\alpha_1}<\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}<q^*_{1-\alpha_2})\\&=\mathbb{P}\left(\bar{X}-q^*_{1-\alpha_2}\frac{\sigma}{\sqrt{n}}<\mu<\bar{X}-q^*_{\alpha_1}\frac{\sigma}{\sqrt{n}}\right)\end{aligned}$$

d'où un intervalle de confiance de niveau $1-\alpha$ de μ de la forme

$$IC=\left[\bar{X}-q^*_{1-\alpha_2}\frac{\sigma}{\sqrt{n}};\bar{X}-q^*_{\alpha_1}\frac{\sigma}{\sqrt{n}}\right]$$

Notons que c'est le quantile d'ordre $1-\alpha_2$ qui intervient dans la minoration de l'intervalle, alors que le quantile d'ordre α_1 intervient dans sa majoration.

On voit donc qu'il y a une infinité d'intervalles de confiance possible de même niveau $1-\alpha$. Prenons un exemple numérique : $X_i\sim\mathcal{N}(\mu,\sigma^2)$, $\sigma^2=1$, $n=10$. La valeur observée sur un échantillon de taille $n=10$ est $\bar{x}=2.0686$ Le tableau suivant récapitule pour des couples de valeurs (α_1,α_2), les bornes inférieures et supérieures de l'intervalle de confiance.

```
    alpha1 alpha2   min   max longueur
IC1  0.000  0.050 1.904   Inf    Inf
IC2  0.015  0.035 1.887 2.286  0.398
IC3  0.025  0.025 1.873 2.265  0.392
IC4  0.050  0.000  -Inf 2.233    Inf
```

IC1 et IC4 sont appelés intervalles de confiance **unilatéraux**, alors que IC2 et IC3 sont **bilatéraux**. IC3 est l'intervalle de confiance bilatère symétrique (il équirépartit les probabilités à gauche et à droite) et est de plus faible étendue sur cet exemple. C'est l'intervalle que nous avions déduit du test de Student :

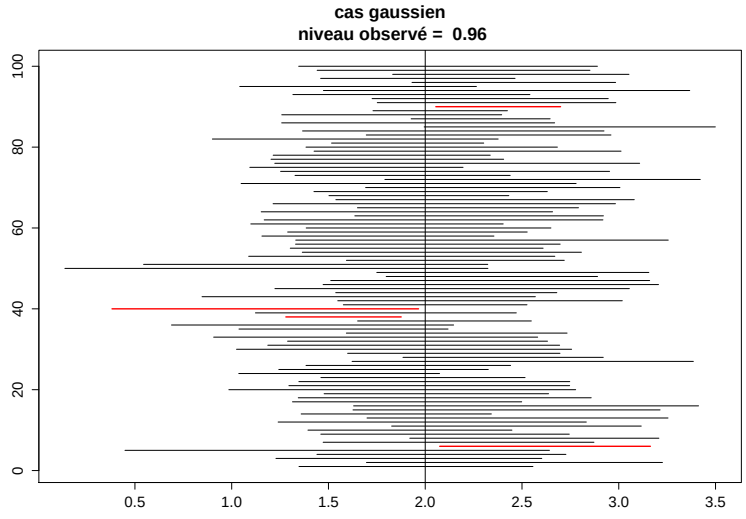


FIGURE 7.1 – Représentation de 100 intervalles de confiance construits à partir de 100 échantillons de taille 10 pour l'estimation de l'espérance d'une gaussienne à variance inconnue.

$$IC3(\mu) = \left[\bar{X} - q^*_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}; \bar{X} + q^*_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Il convient de bien interpréter la notion d'intervalle de confiance. Considérons $IC3$ l'intervalle de confiance symétrique de μ. Son observation sur le jeu de données vaut $[1.873; 2.265]$

Il est **incorrect** de dire : μ appartient à $[1.873; 2.265]$ avec probabilité $1-\alpha$. En effet (dans notre exemple) μ appartient à IC et $\mathbb{P}(\mu \in [1.873; 2.265]) = 1$ Mais il est **blue**correct de déclarer :

- La vraie valeur de μ (inconnue) **appartient ou (exclusif) n'appartient pas** à l'intervalle observé $[1.873; 2.265]$.
- Si on construit une centaine d'intervalles de confiance à partir d'une centaine de n-échantillons indépendants, **en moyenne** $100 \times \alpha$ IC observés **ne contiendront pas** μ. Mais on ne sait pas lesquels...

Ceci est représenté figure 7.1 : sur 100 ICs de niveau 95% calculés à partir de $B = 100$ échantillons de taille $n = 10$ d'une loi gaussienne $\mathcal{N}(2, 1)$, quatre ne contiennent pas la vraie valeur 2. Le niveau observé est de 96%, proche du niveau théorique de 95%. Quand B tend vers l'infini, le niveau observé tend vers le niveau réel. Ce type de simulation est un exemple de simulation de Monte-Carlo.

Remarques : de part la forme des bornes de l'intervalle de confiance, on voit que l'intervalle de confiance est d'autant plus large que α est petit. A l'extrême, l'IC de niveau de confiance $1 = 100\%$ contient toutes les valeurs possibles... mais il n'est plus informatif !

L'intervalle de confiance de l'espérance calculé précédemment est d'autant plus étroit que n est grand

Définition 7.2. *Une construction d'un intervalle de confiance est* ***convergente*** *si la différence des bornes de l'IC tend en proba vers 0 avec n*

A α et n fixés, l'intervalle de confiance est d'autant meilleur que sa longueur est faible (pour toute réalisation ou en moyenne)

7.3 Construction

Les intervalles de confiance peuvent être construit par la méthode **pivotale** : A partir d'un estimateur $\widehat{\theta}$ de θ, il s'agit de trouver une statistique pivotale $T_n(\widehat{\theta}, \theta)$ dont la loi ne dépend pas de θ, puis d'exprimer les bornes de l'intervalle de confiance en fonction de T_n et de ses quantiles. C'est la méthode que nous avons utilisée dans la section 7.2. Pour construire un intervalle de confiance de θ, on utilisera un estimateur de θ, dont on connaît la loi de probabilité pour tout θ, et le meilleur possible (de risque minimum ou sans biais et de variance minimum).

On peut également dériver un intervalle de confiance d'un paramètre à partir d'un test construit pour tester la valeur de ce paramètre, comme nous l'avons fait dans la section introductive de ce chapitre. Dans les exemples pris ici, les deux méthodes ont donné le même intervalle de confiance, mais ce n'est pas toujours le cas.

Comme pour les tests, il est possible de construire des intervalles de confiance **asymptotique** lorsque la loi de la statistique n'est pas connue à distance finie, mais tend asymptotiquement vers une loi pivotale. L'intervalle de confiance est constuit comme si la loi à distance finie était la loi limite. Le niveau de l'intervalle de confiance construit est **approximativement** $1-\alpha$ à distance finie, l'approximation s'améliorant avec n croissant.

Définition 7.3. *Une suite d'intervalle de confiance IC_n de $\theta \in \mathbb{R}$ construit avec un n-échantillon est de niveau asymptotique $1-\alpha$ si, pour tout $\theta \in \Theta$, on a*

$$\lim_{n\to\infty} \mathbb{P}(IC_n \ni \theta) = 1-\alpha.$$

Prenons par exemple l'intervalle de confiance de l'espérance θ d'une loi à variance inconnue (finie). L'estimateur (biaisé) de la variance est consistant

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \xrightarrow{\mathcal{P}} \sigma^2$$

Par ailleurs, la combinaison du Théorème de Limite Centrale et du lemme de Slutsky amène au comportement asymptotique suivant, quand l'échantillon est généré avec une loi d'espérance θ :

$$T_n = \frac{\bar{X}-\theta}{S_n/\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

d'où, avec q^* le quantile de $\mathcal{N}(0,1)$ d'ordre $1-\alpha/2$ l'intervalle de confiance du niveau $1-\alpha$

$$IC = \left[\bar{X} - q^*\frac{S_n}{\sqrt{n}}; \bar{X} + q^*\frac{S_n}{\sqrt{n}}\right] \text{ avec } \mathbb{P}(IC \ni \theta) \simeq 1-\alpha$$

Une solution alternative est d'utiliser l'estimateur non biaisé de la variance, qui est également consistant

$$\widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \xrightarrow{\mathcal{P}} \sigma^2$$

et la statistique ainsi renormalisée, tend, quand l'échantillon est généré avec une loi d'espérance θ, vers une loi gaussienne centrée réduite

$$\tilde{T}_n = \frac{\bar{X}-\theta}{\widehat{\sigma}_n/\sqrt{n}} \simeq \mathcal{T}(n-1) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

Par ailleurs, la loi de $\tilde{T}_n$ peut être approchée à distance finie par une loi de Student $\mathcal{T}(n-1)$, elle-même tendant vers une gaussienne quand n tend vers l'infini. D'où, avec $t = qt(1-\alpha/2, n-1)$ le quantile de $\mathcal{T}(n-1)$ d'ordre $1-\alpha/2$, l'intervalle de confiance du niveau $1-\alpha$

$$IC = \left[\bar{X} - t\frac{\widehat{\sigma}_n}{\sqrt{n}}; \bar{X} + t\frac{\widehat{\sigma}_n}{\sqrt{n}}\right] \text{ avec } \mathbb{P}(IC \ni \theta) \simeq 1-\alpha$$

Cet IC est **exact** si la loi de l'échantillon est **gaussienne**. Il est appelé intervalle de confiance de Student, pour mentionner la loi de la statistique de test utilisée.

Exemple Les logiciels statistiques permettent de calculer des intervalles de confiance dans des cas standards. Par exemple, pour l'intervalle de confiance de Student, on peut utiliser la fonction `t.test` de R : Dans le code suivant, un échantillon de taille $n = 9$ est généré à partir d'une loi gaussienne d'espérance $\mu = 10.5$ et de variance $\sigma^2 = 1$ (la fonction `set.seed` permet de fixer le germe du générateur aléatoire et de rendre la simulation reproductible). La fonction `t.test` calcule un intervalle de confiance de l'espérance quand la variance est inconnue. L'argument `alternative='greater'` indique qu'il s'agit d'un intervalle de confiance unilatéral à gauche, puis qu'il est déduit du test

$$(H_0) : \mu = \mu_0 = 10 \text{ contre } (H_1) : \mu > \mu_0 = 10$$

la valeur de μ_0 est définie par l'argument `mu=10`.

```
> mu=10.5; sigma=1; n=9
> set.seed(432); X=rnorm(n,mu,sigma)
> X       # les valeurs générées
9.842447 11.899532 11.101375 11.132721 10.974112 10.892170  9.827512
 10.016622 10.419826
> t.test(X,alternative="greater",mu=10)
One Sample t-test , data:  X
t = 2.9014, df = 8, p-value = 0.009925
alternative hypothesis: true mean is greater than 10
95 percent confidence interval: 10.24363      Inf
sample estimates: mean of x :  10.67848

> m=mean(X); t= sqrt(n)*(m-10)/sd(X)
> c(m,t,pt(t, n-1,lower.tail=FALSE))
[1] 10.678479644  2.901389730  0.009924681
> m - qt(0.95,n-1)*sd(X)/sqrt(n)
[1] 10.24363
```

On retrouve évidemment les sorties de la fonction par un calcul "à la main" les sorties de la fonction. Note : la valeur de l'argument `mu` n'a pas d'importance dans le calcul de l'intervalle de confiance fait par R, ce qui importe, c'est l'argument `alternative`, qui peut prendre les valeurs `greater`, `less` (intervalle de confiance unilatéral à droite) ou `two.sided` pour l'intervalle bilatéral symétrique. Par défaut, le niveau de l'intervalle est 0.95, et donc celui du test 0.05, mais on peut changer cette valeur en utilisant l'argument `conf.level`. L'information `df` désigne le nomber de degrés de liberté (degrees of freedom) de la statistique de Student : ici, $n - 1 = 8$.

Si on effectue sur le même exemple un test bilatéral $(H_0) : \mu = \mu_0 = 10$ contre $(H_1) : \mu \neq \mu_0 = 10$ la fonction `t.test` donne les résultats du test et de l'intervalle de confiance

```
> t.test(X,alternative="two.sided",mu=10)
    One Sample t-test , data:  X
t = 2.9014, df = 8, p-value = 0.01985
alternative hypothesis: true mean is not equal to 10
95 percent confidence interval:
 10.13923 11.21773
sample estimates: mean of x :  10.67848

> m +c(-1,1) * qt(0.975,n-1)*sd(X)/sqrt(n)
[1] 10.13923 11.21773
```

L'intervalle de confiance est bilatère, la statistique de test est la même que celle du test unilatéral, mais la p-value est différente.

7.4 Exemples

Nous avons vu le cas de l'intervalle de confiance de l'espérance d'une loi à variance inconnue (intervalle de confiance de Student). Nous présentons ici quelques autres exemples d'intervalles de confiance classiques, en mettant les tests en parallèle. Tous les ICs proposés sont bilatère, il suffit d'adapter dans le cas unilatéral.

7.4.1 IC et test de la variance d'une loi gaussienne

Soit $\widehat{\sigma}_n^2 = \sum_i (X_i - \bar{X})^2/(n-1)$ l'estimateur sans biais de la variance d'une loi $\mathcal{N}(\mu, \sigma^2)$ et on pose

$$T_n = \frac{(n-1)\widehat{\sigma}_n^2}{\sigma^2} \sim \chi^2_{n-1}$$

Soit $k_1 = qchisq(\alpha/2, n-1)$ et $k_2 = qchisq(1-\alpha/2, n-1)$ les quantiles d'ordre $\alpha/2$ et $1-\alpha/2$ d'une loi du Khi-deux à $n-1$ degrés de liberté. Alors,

$$\mathbb{P}(k_1 < \frac{(n-1)\widehat{\sigma}_n^2}{\sigma^2} < k_2) = 1 - \alpha$$

On en déduit un intervalle de confiance de niveau $1-\alpha$

$$IC = \left[\frac{(n-1)\widehat{\sigma}_n^2}{k_2}; \frac{(n-1)\widehat{\sigma}_n^2}{k_1}\right] \text{ avec } \mathbb{P}(IC \ni sigma^2) = 1 - \alpha$$

et le test de $(H_0) : \sigma^2 = \sigma_0^2$ contre $(H_1) : \sigma^2 \neq \sigma_0^2$ est de région d'acceptation $[k_1; k_2]$ pour T_n Attention, contrairement au test et intervalle de confiance de Student, ces constructions sont **peu robustes** vis à vis de l'hypothèse gaussienne.

7.4.2 Comparaison de l'espérance de deux lois

La comparaison des moyennes a été vue en TD4. On en rappelle ici les résultats. Soient deux échantillons gaussiens $X = (X_1, \ldots, X_n)$ et $Y = (Y_1, \ldots, X_m)$ indépendants de lois respectives $\mathcal{N}(\mu_1, \sigma^2)$ et $\mathcal{N}(\mu_2, \sigma^2)$. Les paramètres μ_1, μ_2 et σ^2 sont inconnus. La variance est inconnue, mais supposée identique dans les deux échantillons, est estimée par

$$\widehat{\sigma}^2 = \frac{1}{n+m-2}\left(\sum_{i=1}^{n}(X_i - \overline{X})^2 + \sum_{j=1}^{m}(Y_j - \overline{Y})^2\right)$$

Alors, la statistique T définie par

$$T = \frac{\overline{Y} - \overline{X}}{\widehat{\sigma}\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim \mathcal{T}(n + m - 2)$$

suit une loi de Student de paramètre $n + m - 2$. Soit q_t le quantile d'ordre $1 - \alpha/2$ d'une loi $\mathcal{T}(n + m - 2)$, l'intervalle de confiance bilatère de $\mu_2 - \mu_1$

$$IC(\mu_2 - \mu_1) = \left[\overline{Y} - \overline{X} \pm q_t\, \widehat{\sigma}\sqrt{\frac{1}{n} + \frac{1}{m}}\right]$$

est de niveau exact $1 - \alpha$ (ou approché si $n > 30$ et $m > 30$ et sans l'hypothèse gaussienne).

Le test bilatère $(H_0) : \mu_1 = \mu_2$ contre $(H_1) : \mu_1 \neq \mu_2$ est de région de rejet $\{|T| > q_t\}$ de niveau exact α (ou approché si $n > 30$ et $m > 30$ et sans l'hypothèse gaussienne). La fonction `t.test` permet de calculer les résultats du test, comme par exemple sur cet exemple où l'on compare les moyennes de deux petits échantillons d'espérance $\mu_0 = 1$ et $\mu_1 = 2$

```
> set.seed(15)
> mu0=1; mu1=2; sigma=2; n=9
> X0=rnorm(n,mu0,sigma); X0
 1.5176457  4.6622414  0.3207629  2.7943963  1.9760326 -1.5107716
  1.0455764  3.1815464  0.7357551
> X1=rnorm(n,mu1,sigma); X1
-0.1500026  3.7100215  1.2700397  2.3311086 -0.4855700  4.9185754
 1.9927745  1.9582337  2.0642120

> t.test(X0,X1,alternative = "two.sided", var.equal=TRUE)
Two Sample t-test
data:  data:  X0 and X1
t = -0.39, df = 16, p-value = 0.7017
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.063869  1.422489
sample estimates:
mean of x mean of y
 1.635909  1.956599
```

Sur cet exemple, la p-value du test vaut $0.7017 > 0.05$. Le test de niveau 5% accepte l'égalité des espérances, et commet une erreur de seconde espèce de risque $\mathbb{P}_{(H_1)}(\{|T| > q_t\})$. On peut remarquer que 0 appartient à l'intervalle de confiance, ce qui est cohérent avec le résultat du test.

Si on ne suppose plus l'égalité des variances des deux lois, et si $n > 30$ et $m > 30$, on peut proposer la statistique asymptotiquement pivotale

$$\tilde{T} = \frac{\overline{Y} - \overline{X}}{\sqrt{\frac{\widehat{\sigma_1^2}}{n} + \frac{\widehat{\sigma_2^2}}{m}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

et en déduire intervalle de confiance et test. Il est également possible d'utiliser une autre loi, extension de la loi de Student avec des degrés de liberté non entier

```
> t.test(X0,X1, alternative = "two.sided")

Welch Two Sample t-test

data:  X0 and X1
t = -1.6606, df = 14.808, p-value = 0.1178
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.9444877  0.3672109
sample estimates:
mean of x mean of y
 1.540758  2.829396
```

Il s'agit du test (et de l'intervalle de confiance) de Welch, calculé par défaut par la fonction `t.test`, puisque l'argument `var.equal` a été omis. Sur cet exemple, les résultats vont dans le même sens que ceux du test de Student à variances égales.

7.4.3 Rapport des variances de deux lois gaussiennes

Soient deux échantillons gaussiens $X = (X_1, \ldots, X_n)$ et $Y = (Y_1, \ldots, X_m)$ indépendants de lois respectives $\mathcal{N}(\mu_1, \sigma_1^2)$ et $\mathcal{N}(\mu_2, \sigma_2^2)$. Les paramètres μ_1, μ_2 et σ^2 sont inconnus. On s'intéresse maintenant à la comparaison de σ_1^2 et σ_2^2, par l'intermédiaire de leur rapport σ_1^2/σ_2^2. Sous (H_0),

$$\frac{\widehat{\sigma}_X^2/\sigma_X^2}{\widehat{\sigma}_Y^2/\sigma_Y^2} \sim \mathcal{F}(n-1, m-1)$$

On note $qf_\alpha^{n-1,m-1}$ le quantile d'ordre α d'une loi de Fisher de paramètres $n-1$ et $m-1$. On en déduit

$$IC(\sigma_X^2/\sigma_Y^2) = \left[qf_{\alpha/2}^{m-1,n-1} \frac{\widehat{\sigma}_X^2}{\widehat{\sigma}_Y^2};\ qf_{1-\alpha/2}^{m-1,n-1} \frac{\widehat{\sigma}_X^2}{\widehat{\sigma}_Y^2} \right]$$

ou

$$IC(\sigma_Y^2/\sigma_X^2) = \left[qf_{\alpha/2}^{n-1,m-1} \frac{\widehat{\sigma}_Y^2}{\widehat{\sigma}_X^2};\ qf_{1-\alpha/2}^{n-1,m-1} \frac{\widehat{\sigma}_Y^2}{\widehat{\sigma}_X^2} \right]$$

puisque $qf_{1-\alpha}^{m-1,n-1} = 1/qf_\alpha^{n-1,m-1}$. En effet, soit $K \sim \mathcal{F}(n,m)$

$$\alpha = \mathbb{P}\left(K \leq qf_\alpha^{n,m}\right) = \mathbb{P}\left(\frac{1}{qf_\alpha^{n,m}} \leq \frac{1}{K}\right) = 1 - \mathbb{P}\left(\frac{1}{K} \leq \frac{1}{qf_\alpha^{n,m}}\right)$$

soit $1-\alpha = \mathbb{P}\left(1/K \leq 1/qf_\alpha^{n,m}\right)$.

Le test bilatère de fonction de test

$$\varphi_\alpha(X,Y) = 1 - \mathbb{1}_{qf_{\alpha/2} < F(X,Y) < qf_{1-\alpha/2}}$$

est de niveau α.

```
> set.seed(15)
> sigma1=1; sigma2=3; n=9
> X1=rnorm(n,mean=0,sd=sigma1); X1
0.2588229  1.8311207 -0.3396186  0.8971982  0.4880163 -1.2553858
  0.0227882  1.0907732 -0.1321224
> X2=rnorm(n,mean=0,sd=sigma2);X2
```

```
-3.22500384  2.56503225 -1.09494042  0.49666288 -3.72835497
 4.37786306 -0.01083831 -0.06264952  0.09631801
> var.test(X1,X2,alternative="two.sided")

F test to compare two variances

data:  X1 and X2
F = 0.12537, num df = 8, denom df = 8, p-value = 0.008158
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.02827909 0.55579138
sample estimates:
ratio of variances
         0.1253686
```

Dans cet exemple, la pvalue est inférieure à 5%, on rejette l'égalité des variances avec un risque de 5%. De même, 1, valeur du rapport des variances si elles sont égales, n'appartient pas à l'intervalle de confiance de niveau 95%.

Notons que ces constructions sont peu robustes pour des lois qui s'écartent de la loi gaussienne.

7.4.4 Intervalle de confiance d'une proportion

Dans le cas où le phénomène observé est binaire, il est modélisé par une loi de Bernoulli : $Z_i \sim_{i.i.d.} \mathcal{B}(1,\pi)$. Le paramètre de la loi de Bernoulli est égal à l'espérance de la loi : $\pi = \mathbb{E}(Z_i)$. La somme de variables de Bernoulli de même loi suit une loi Binomiale $\sum_i Z_i \sim \mathcal{B}(n,\pi)$. Par application du Théorème de Limite Centrale,

$$T_n = \sqrt{n}\frac{\bar{Z}-\pi}{\sqrt{\pi(1-\pi)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \text{ et } \bar{Z}(1-\bar{Z}) \xrightarrow{\mathcal{P}} \pi(1-\pi)$$

Ainsi, si n est suffisamment grand,

$$\mathbb{P}(-q^* < T_n < q^*) \simeq 1-\alpha$$

où q^* est la quantile d'une loi gaussienne centrée réduite d'ordre $1-\alpha/2$. En remplaçant T_n par son expression en π, il reste à résoudre une équation du second degré en π qui amène à l'expression suivante de l'intervalle de confiance :

$$IC = \left[\frac{\widehat{\pi}+\frac{(q^*)^2}{2n}}{1+\frac{(q^*)^2}{n}} \pm \frac{1}{1+\frac{(q^*)^2}{n}}\sqrt{\frac{(q^*)^2}{n}\widehat{\pi}(1-\widehat{\pi})+\frac{(q^*)^4}{4n^2}}\right]$$

On considère que l'approximation asymptotique est valide quand $n\widehat{\pi}(1-\widehat{\pi}) > 12$ (quand π est connu, on admet l'approximation asymptotique quand $n\pi > 5$ et $n(1-\pi) > 5$). Or $\widehat{\pi}(1-\widehat{\pi}) < 1/4$ et l'expression de l'IC se simplifie

$$\tilde{IC} = \left[\widehat{\pi} - q^*\sqrt{\frac{\widehat{\pi}(1-\widehat{\pi})}{n}}; \widehat{\pi} + q^*\sqrt{\frac{\widehat{\pi}(1-\widehat{\pi})}{n}}\right]$$

Cette expression aurait pu être obtenue en remarquant, par application du Théorème de Limite Centrale et du lemme de Slutsky, que

$$\sqrt{n}\frac{\widehat{\pi}-\pi}{\sqrt{\widehat{\pi}(1-\widehat{\pi})}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \text{ et } \widehat{\pi}(1-\widehat{\pi}) \xrightarrow{\mathcal{P}} \pi(1-\pi)$$

Les IC sont de niveau approché α :

$$\mathbb{P}(\tilde{IC} \ni \pi) \geq \mathbb{P}(IC \ni \pi) \simeq 1 - \alpha$$

Or, $\widehat{\pi}(1-\widehat{\pi}) < 1/4$, on peut majorer la **précision** de l'intervalle

$$\Delta\pi = q^* \sqrt{\frac{\widehat{\pi}(1-\widehat{\pi})}{n}} \leq \frac{q^*}{2} \frac{1}{\sqrt{n}}$$

Par exemple, $n_{max} = q^2/(4 \times 0.01^2) \simeq 6765$ pour garantir une précision de $\pm 1\%$ d'un intervalle de confiance à 90% ($n = 9600$ pour un intervalle de confiance à 95%).

Notons que la loi de $\widehat{\pi}$ est discrète. Il peut être intéressant de prendre en compte une correction de continuité, en écrivant

$$\mathbb{P}(n\widehat{\pi} = k) \simeq \mathbb{P}\left(k - \frac{1}{2} < N \leq k + \frac{1}{2}\right) \text{ avec } N \sim \mathcal{N}(n\pi, n\pi(1-\pi))$$

Pour tester $(H_0) : \pi = \pi_0$ contre $(H_1) : \pi \neq \pi_0$, on utilise la statistique

$$T_n = \sqrt{n} \frac{\widehat{\pi} - \pi_0}{\sqrt{\pi_0(1-\pi_0)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

qui est asymptotiquement gaussienne sous (H_0) d'après le Théorème de Limite Centrale. On en déduit la région de rejet $\mathcal{R} = \{|T_n| > q^*\}$ avec $\mathbb{P}_{\pi_0}(\mathcal{R}) \simeq \alpha$, de niveau asymptotique α.

Chapitre 8

Tests du Khi-deux et tests d'adéquation

Les chapitres précédents ont principalement étudié des modèles dont les statistiques étaient de loi continue (ou à défaut, asymptotiquement gaussienne).

Nous étudions dans ce chapitre les variables catégorielles (extension des variables binomiales à plus de deux catégories), qui mèneront à la construction de tests dits du Khi-deux d'indépendance (entre deux caractères qualitatifs) et du Khi-deux d'adéquation (pour tester l'ajustement d'un type de loi à un jeu d'observations). Nous terminerons par la présentation de quelques tests d'adéquation non paramétriques.

8.1 Loi Multinomiale

La loi multinomiale est l'extension de la situation binaire à une situation à K catégories exclusives de probabilité $\pi_1, \ldots, \pi_K$, avec $\sum_{k=1}^K \pi_k = 1$.

$$P(Z_i = k) = \pi_k; \;\; Z_{ik} = \mathbb{1}_{\{Z_i = k\}}$$

On s'intéresse aux comptages observés dans chaque catégorie au cours de n observations répétées indépendantes.

$$N_n = (O_1, \ldots, O_K) \sim \mathcal{M}(n, \pi) \;\text{ avec } O_k = \sum_{i=1}^n Z_{ik}$$

C'est le cas par exemple dans un sondage d'opinion où la réponse peut prendre plusieurs niveaux : préférez-vous (choix unique) passer vos vacances au bord de mer, à la montage, à la campagne ou en ville ? Ici $K = 4$ et n le nombre de personnes interrogées.

8.1.1 Propriétés

La loi multinomiale est donc multivariée (elle a K composantes) et discrète, définie pour $(n_1, \ldots, n_K) \in \{0, 1, \ldots, n\}^K$ tels que $\sum_{k=1}^K n_k = n$ par

$$\mathbb{P}(N_n = (n_1, \ldots, n_K)) = n! \prod_{k=1}^K \frac{\pi_k^{n_k}}{n_k!}.$$

Ses lois marginales (loi d'une composante) sont des binomiales : $O_k \sim \mathcal{B}(n, \pi_k)$. Ainsi, le vecteur des espérances $\mathbb{E}(N_n) = (n\pi_1, \dots, n\pi_K)'$ est le vecteur des espérances des marginales, et la matrice de variance vaut

$$\mathrm{cov}(N_n) = \begin{pmatrix} n\pi_1(1-\pi_1) & -n\pi_1\pi_2 & \dots & -n\pi_1\pi_K \\ -n\pi_1\pi_2 & \pi_2(1-\pi_2) & \dots & -n\pi_2\pi_K \\ \vdots & \vdots & \ddots & \vdots \\ -n\pi_1\pi_K & -n\pi_2\pi_K & \dots & n\pi_K(1-\pi_K) \end{pmatrix} = \Sigma_n$$

Le paramètre à estimer est π, de dimension K, mais le modèle n'admet que $K-1$ paramètres indépendants puisque $\sum_k \pi_k = 1$.

L'estimateur des moments du paramètre π est le vecteur des proportions observées de chaque catégorie $\widehat{\pi}_k = O_k/n$.

8.1.2 Asymptotique Loi Multinomiale

Chaque composante suit une loi binomiale, à laquelle on peut appliquer un théorème de limite centrale

$$\sqrt{n}\left(\frac{O_k}{n} - \pi_k\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \pi_k(1-\pi_k))$$

La loi multinomiale suit de plus un théorème de limite centrale vectoriel

Théorème 8.1 (TLC vectoriel)**.** *Si X_i est un échantillon i.i.d. de vecteurs aléatoires de $\mathbb{R}^K$ de carré intégrable, alors*

$$Y_n = \sqrt{n}(\bar{X} - \mathbb{E}(X_1)) \xrightarrow{\mathcal{L}} Y \sim \mathcal{N}(0, \Sigma)$$

où la convergence en loi est définie si toute forme linéaire $a(Y_n)$ converge en loi vers $a(Y)$.

On en déduit

$$\frac{N_n - \mathbb{E}(N_n)}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathrm{cov}(N_1) = \Sigma_1)$$

Remarque : le déterminant de la matrice limite Σ_1 est nul, parce que les composantes de N_n (et donc de sa limite renormalisée) ne sont pas indépendantes.

Définition 8.2 (Statistique (du Khi-deux) de Pearson)**.** *La statistique K_n de Pearson est définie par*

$$K_n = \sum_{k=1}^{K} \frac{(O_k - n\pi_k)^2}{n\pi_k}$$

C'est la somme des carrés des écarts des comptages observés aux comptages théoriques renormalisés. Chaque composante de K_n suit une loi du Khi-deux à un degré de liberté, mais les composantes ne sont pas indépendantes, puisque les O_k ne le sont pas. Le théorème de Cochran permet de définir le comportement asymptotique de K_n :

Théorème 8.3. *La statistique K_n de Pearson suit asymptotiquement une loi du Khi-deux à $K-1$ degrés de liberté si K est le nombre de catégories :*

$$K_n = \sum_{k=1}^{K} \frac{(O_k - n\pi_k)^2}{n\pi_k} \xrightarrow{\mathcal{L}} \chi^2_{(K-1)}$$

Preuve. La preuve utilise un changement de repère judicieusement choisi. Soit $\Delta = Diag(1/\sqrt{\pi_1}, \ldots, 1/\sqrt{\pi_K})$ et $\sqrt{\pi} = (\sqrt{\pi_1}, \ldots, \sqrt{\pi_K})'$.

$$\Delta Y_n \xrightarrow{\mathcal{L}} \Delta Y \sim \mathcal{N}(0, \Delta\Sigma\Delta' = Id_K - \sqrt{\pi}(\sqrt{\pi})')$$

Soit Q une matrice orthogonale telle que $Q\sqrt{\pi} = (1\ 0\ \ldots\ 0)'$. On a $||Q\Delta Y_n||^2 = ||\Delta Y_n||^2 = \sum_{k=1}^K \frac{(O_k - n\pi_k)^2}{n\pi_k}$ et, d'après le théorème de Cochran,

$$||Q\Delta Y_n||^2 \xrightarrow{\mathcal{L}} \chi^2_{(K-1)}$$

$\Diamond$

8.1.3 Test du paramètre d'une multinomiale

La statistique de Pearson est utilisée dans le test du paramètre d'une multinomiale.

Posons les hypothèses $(H_0) : \pi = \pi_0$ contre $(H_1) : \pi \neq \pi_0$. Nous avons vu que la statistique de Pearson $K_n = \sum_{k=1}^K \frac{(O_k - n\pi_{0k})^2}{n\pi_{0k}}$ est asymptotiquement libre sous (H_0) et y suit une loi du Khi-deux à $K-1$ degrés de liberté. Si les données ne sont pas générées sous (H_0), le carré des écarts à l'effectif théorique sera plus important qu'attendu. La région de rejet de risque α est donc unilatérale à droite pour un test sans biais :

$$\mathbb{P}_{H_0}(K_n > \chi^2_{(K-1,1-\alpha)}) \to \alpha$$

où $\chi^2_{(K-1,1-\alpha)}$ est le quantile d'ordre $1-\alpha$ d'une loi du Khi-deux à $K-1$ degrés de liberté L'asymptotique est admise dès que : $n \min_k \pi_K > 5$. Le test est convergent. En effet, sous (H_1),

$$\begin{aligned}\frac{1}{n}K_n &= \frac{1}{n}\sum_{k=1}^K \frac{(O_k - n\pi_{0k})^2}{n\pi_{0k}} = \frac{n^2}{n^2}\sum_{k=1}^K \frac{(\widehat{\pi}_k - \pi_k + \pi_k - \pi_{0k})^2}{\pi_{0k}} \\ &\to \sum_{k=1}^K \frac{(\pi_k - \pi_{0k})^2}{\pi_{0k}} = \chi^2(\pi_0, \pi) > 0\end{aligned}$$

8.1.4 Cas particulier de la binomiale

La loi binomiale est un cas particulier de loi multinomiale, avec $K = 2$, $\pi_{01} = \pi_0$, $\pi_{02} = 1 - \pi_0$. En appliquant les résultats précédents, la statistique de Pearson s'écrit

$$\begin{aligned}K_n &= \frac{(O_1 - n\pi_0)^2}{n\pi_0} + \frac{(n - O_1 - n(1-\pi_0))^2}{n(1-\pi_0)} \\ &= \frac{(O_1 - n\pi_0)^2}{n\pi_0} + \frac{(O_1 - n\pi_0)^2}{n(1-\pi_0)} \\ &= \frac{(O_1 - n\pi_0)^2}{n\pi_0(1-\pi_0)} = \frac{(\widehat{\pi} - \pi_0)^2}{\frac{\pi_0(1-\pi_0)}{n}}\end{aligned}$$

La région d'acceptation du test d'égalité $\pi = \pi_0$ contre $\pi \neq \pi_0$

$$\{K_n < q_{\chi^2_{1,1-\alpha}}\} = \left\{ -\sqrt{q_{\chi^2_{1,1-\alpha}}} < \frac{\widehat{\pi} - \pi_0}{\frac{\pi_0(1-\pi_0)}{n}} < \sqrt{q_{\chi^2_{1,1-\alpha}}} \right\}$$

est identique à la région de rejet du test classique par approximation gaussienne pour une proportion, car $\sqrt{q_{\chi^2_{1,1-\alpha}}} = q^*_{1-\alpha/2}$.

8.1.5 Un autre test dans le modèle multinomial

Il est possible de définir un autre test, généralisation du test de Neyman-Pearson à des hypothèses composites : c'est le test du **rapport de vraisemblances maximales** (ou rapport de vraisemblance généralisé).

Dans un modèle i.i.d. de loi marginale $(f(x,\theta), \theta \in \Theta)$, $\Theta \subset \mathbb{R}^K$, la vraisemblance de l'échantillon $X = (X_1, \ldots, X_n)$ est le produit des vraisemblances de chacune des observations : $L(X;\theta) = \prod_i f(X_i, \theta)$. On teste $(H_0) : \theta \in \Theta_0$ contre $(H_1) : \theta \in \Theta_1 = \Theta \setminus \Theta_0$. La statistique de test du rapport de vraisemblance généralisé est définie par

$$\lambda_n(X) = \frac{\sup_{\theta\in\Theta_0} L(\theta; X)}{\sup_{\theta\in\Theta} L(\theta; X)}$$

La région de rejet est définie par $\mathcal{R} = \{\lambda_n(X) < q\}$ où q permet de vérifier $\mathbb{P}_{H_0}(\mathcal{R}) = \alpha$.

Théorème 8.4 (admis)**.** *Si le modèle est régulier et $\widehat{\theta}$ asymptotiquement normal, on a*

$$-2\log\lambda_n(X) \xrightarrow{\mathcal{L}} \chi^2(r)$$

où r désigne le nombre de composantes de θ spécifiées dans (H_0) par rapport à (H_1). Le test de rapport de vraisemblance (généralisé) de niveau approché α est défini par

$$\mathcal{R} = \{-2\log\lambda_n(X) > q_{\chi^2(r),1-\alpha}\};\ \ \mathbb{P}_{H_0}(\mathcal{R}) \simeq \alpha.$$

Définition 8.5. *On appelle **déviance** la statistique $-2\log\lambda_n(X)$*

Ceci peut être appliqué dans le cas de la loi multinomiale pour tester $(H_0) : \pi = \pi_0$ contre $(H_1) : \pi \neq \pi_0$. La vraisemblance admet son maximum pour θ égal à l'estimateur empirique déjà rencontré : $\widehat{\pi}_k = \frac{\sum_{i=1}^n Z_{ik}}{n} = \frac{O_k}{n}$. La statistique de test du rapport de vraisemblance généralisé

$$-2\log\lambda_n(Z) = 2\sum_{k=1}^K O_k \log\frac{O_k}{n\pi_{0k}} \xrightarrow{\mathcal{L}} \chi^2(K-1)$$

suit une loi du Khi-deux sous (H_0). Même si la loi limite est la même que celle du test de Pearson, les statistiques à distance finies sont différentes, et les tests sont donc différents. Mais on peut montrer leur équivalence asymptotique.

8.2 Test d'indépendance

Dans cette section, on considère deux caractères qualitatifs Y à J niveaux et Z à K niveaux observés sur chacun des n individus de l'échantillon. Les deux caractères (variables) sont-ils indépendants ? Par exemple, y a-t-il indépendance entre la couleur des yeux et la couleur des cheveux ?

Pour un individu i, $(Y,Z)_i$ suit une loi jointe discrète à $J \times K$ niveaux, définie par :

$$P(Y_i = j, Z_i = k) = \pi_{jk};\ \ \sum_{j=1}^J\sum_{k=1}^K \pi_{jk} = 1$$

soit $JK-1$ paramètres indépendants Le n-échantillon i.i.d. $((Y_1,Z_1)\ldots,(Y_n,Z_n))$ suit donc une multinomiale de paramètre π. Les effectifs observés o_{jk} sont les comptages des individus de même

caractéristique (j,k) :

$Y_1 \setminus Z_1$	1	...	K	total
1	o_{11}	...	o_{1K}	$o_{1\bullet}$
$\vdots$	$\vdots$	o_{jk}	$\vdots$	$o_{j\bullet}$
J	o_{J1}	...	o_{JK}	$o_{J\bullet}$
total	$o_{\bullet 1}$	$o_{\bullet k}$	$o_{\bullet K}$	n

dont on peut définir les marges

$$o_{j\bullet} = \sum_{k=1}^{K} o_{jk}; \quad o_{\bullet k} = \sum_{j=1}^{J} o_{jk}; \quad o_{\bullet\bullet} = \sum_{j=1}^{J} o_{j\bullet} = \sum_{k=1}^{K} o_{\bullet k} = n.$$

Le test d'indépendance teste l'hypothèse nulle (non-dépendance) des caractères contre l'alternative (dépendance). Soit $\mathbb{P}(Y_i = j) = \pi_{j\bullet}$ et $\mathbb{P}(Z_i = k) = \pi_{\bullet k}$. Sous (H_0),

$$P(Y_i = j, Z_i = k) = \pi_{j\bullet}\pi_{\bullet k}$$

soit $J - 1 + K - 1$ paramètres indépendants, tandis que sous (H_1), c'est une loi multinomiale quelconque à JK catégories, soit $JK - 1$ paramètres.

8.2.1 Test du Khi-deux d'indépendance de Pearson

Contrairement au test d'une valeur du paramètre de la section précédente, ici l'hypothèse nulle est composite : ce sont toutes les lois indépendantes qui sont testées. Pour exprimer l'hypothèse nulle, il faut donc estimer les valeurs des proportions pour Y et Z. Ainsi, à partir des effectifs attendus estimés sous (H_0)

$$n\widehat{\pi}^0_{j\bullet}\widehat{\pi}^0_{\bullet k} = n\frac{O_{j\bullet}}{n}\frac{O_{\bullet k}}{n} = \frac{O_{j\bullet}O_{\bullet k}}{n}$$

la statistique de Pearson s'écrit

$$K_n = \sum_{j=1}^{J}\sum_{k=1}^{K} \frac{\left(O_{jk} - \frac{O_{j\bullet}O_{\bullet k}}{n}\right)^2}{\frac{O_{j\bullet}O_{\bullet k}}{n}}$$

Théorème 8.6 (admis)**.** *Sous (H_0), la loi asymptotique de la statistique de Pearson K_n est celle d'un Khi-deux, de degré de liberté la différence entre la dimension du modèle sous (H_1) et (H_0)*

$$K_n \xrightarrow{\mathcal{L}} \chi^2((J-1)(K-1)).$$

La région de rejet s'en déduit $\mathcal{R} = \{K_n > q_{\chi^2_{(J-1)(K-1),1-\alpha}}\}$ avec $\mathbb{P}_{H_0}(\mathcal{R}) \simeq \alpha$.

La différence de dimension entre (H_0) et (H_1) est $(JK-1)-(J-1+K-1) = (J-1)(K-1)$

Exemple Agresti (2007) exploite la composition d'une assemblée pour déterminer si le taux de féminisation est indépendant du parti politique

	Democrat	Independent	Republican
F	762	327	468
M	484	239	477

Les données sont entrées pour créer le tableau de contingence des deux caractères sous forme d'un dataframe `M`. La fonction `chisq.test` indique le résultat du test

```
> M <- as.table(rbind(c(762, 327, 468), c(484, 239, 477)))
> dimnames(M) <- list(gender = c("F", "M"),
+                     party = c("Democrat","Independent", "Republican"))
>  chisq.test(M)
Pearson's Chi-squared test
data:  M
X-squared = 30.07, df = 2, p-value = 2.954e-07
> qchisq(0.95,2)
[1] 5.991465
```

La pvalue du test est inférieure à 5%, on rejette (H_0) et on décide donc que la féminisation dépend du parti, au risque de première espèce de 5%. Même décision si on compare la valeur observée 30.07 de la statistique de Pearson, qui est supérieure au quantile du Khi-deux à $6-4=2$ degrés de liberté et valant 5.99 et donc dans la région de rejet de niveau 5%.

8.2.2 Test d'indépendance du rapport de vraisemblance

La vraisemblance sous (H_0)

$$L_{H_0}(Y,Z;\pi) = \frac{n!}{\prod_{j,k} O_{jk}!} \prod_{j,k} (\pi_{j\bullet}\pi_{\bullet k})^{O_{jk}}$$

est maximisée par $\widehat{\pi}^0_{j\bullet} = O_{j\bullet}/n$, $\widehat{\pi}^0_{\bullet k} = O_{\bullet k}/n$ soit :

$$L(Y,Z;\widehat{\pi}^0) = \frac{n!}{\prod_{j,k} O_{jk}!} \prod_{j} \left(\frac{O_{j\bullet}}{n}\right)^{O_{j\bullet}} \prod_{k} \left(\frac{O_{\bullet k}}{n}\right)^{O_{\bullet k}}$$

Sous (H_1), la loi de (Y_i, Z_i) est discrète quelconque à $JK-1$ niveaux

$$L(Y,Z;\widehat{\pi}^0) = \frac{n!}{\prod_{j,k} O_{jk}!} \prod_{j,k} \left(\frac{O_{jk}}{n}\right)^{O_{jk}}$$

La statistique du rapport de vraisemblance généralisé, utilisée pour le test d'indépendance, s'écrit donc

$$\lambda_n(Y,Z) = \frac{\prod_j O_{j\bullet}^{O_{j\bullet}} \prod_k O_{\bullet k}^{O_{\bullet k}}}{n^n \prod_{jk} O_{jk}^{O_{jk}}}$$

Théorème 8.7 (admis)**.** *Sous (H_0), la loi asymptotique de la statistique du rapport de vraisemblance λ_n est celle d'un Khi-deux, de degré de liberté la différence entre la dimension du modèle sous (H_1) et (H_0)*

$$-2\log\lambda_n(Y,Z) \xrightarrow{\mathcal{L}} \chi^2((J-1)(K-1))$$

La région de rejet s'en déduit $\mathcal{R} = \{-2\log\lambda_n(Y,Z) > q_{\chi^2_{(J-1)(K-1),1-\alpha}}\}$ avec $\mathbb{P}_{H_0}(\mathcal{R}) \simeq \alpha$.

On peut montrer que les deux tests sont asymptotiquement équivalents.

8.3 Tests d'adéquation

L'**adéquation** est la qualité d'ajustement d'une loi à un échantillon : le modèle peut toujours être ajusté sur des données, mais ajuste-t-il bien ? Les tests du Khi-deux peuvent être utilisés pour tester l'adéquation à une loi discrète d'espace d'état fini. Mais elle peut aussi être utilisée pour tester l'adéquation d'une loi discrète à espace d'état infini (comme par exemple la loi de Poisson) ou d'une loi continue (gaussienne, exponentielle, ...)

8.3.1 Test du Khi-deux d'adéquation

Soit $(X_1, \ldots, X_n)$ un n-échantillon iid de loi F. On souhaite tester

$$(H_0) : F = F_0 \text{ contre } (H_1) : F \neq F_0$$

Il est alors possible de décomposer le support D de F en une partition de K intervalles A_k : $D = \cup_{k=1}^{K} A_k$ et $A_k \cap A_j = \varnothing$ pour $k \neq j$. On pose alors

$$Z_{ik} = \mathbb{1}_{X_i \in A_k} \text{ et } \pi_k = \mathbb{P}(X_i \in A_k), k = 1, \ldots, K$$

L'effectif attendu est $n\pi_k$ pour la classe k, l'effectif observé est $O_k = \sum_i Z_{ik}$. On considère que la loi est différente si elle est différente sur (au moins) un A_k : il s'agit donc du test du paramètre d'une loi multinomiale, et on utilise le test du Khi-deux : $\mathbb{P}_{H_0}(K_n > \chi^2_{(K-1,1-\alpha)}) \simeq \alpha$

Remarque Il est conseillé de diviser F_0 en K classes équiprobables ($np_k > 5$), ce qui fait des classes d'étendues différences (pense au cas de la loi de Poisson par exemple).

Extension Dans le cas plus général du type de la loi (gaussienne, exponentielle, Poisson) et pas seulement l'adéquation à une loi de paramètre complètement spécifié, (H_0) devient composite. On pourra utiliser la même méthodologie, mais en estimant le paramètre θ de la loi et en le remplaçant par $\widehat{\theta}$ dans la définition de la statistique de Pearson, qui suivra sous (H_0) asymptotiquement une loi du Khi-deux à $K - r - 1$ degrés de liberté, où r est le nombre de paramètres estimés sous (H_0) :

$$\mathbb{P}_{H_0}\left(\sum_{k=1}^{K} \frac{(O_k - n\pi(\widehat{\theta}))^2}{n\pi(\widehat{\theta})} > q_{\chi^2(K-r-1,1-\alpha)}\right) \simeq \alpha$$

8.3.2 Test de Kolmogorov-Smirnov (loi continue)

Il existe un autre test, basé sur le comportement de la fonction de répartition empirique. Soit $X_1, \ldots, X_n$ un échantillon i.i.d. de la loi de fonction de répartition F

Définition 8.8. *La fonction de* ***répartition empirique*** $\widehat{\mathrm{F}}_n$ *est la fonction de répartition de la loi de probabilité discrète uniforme de support* $\{X_1, \ldots, X_n\}$ *:*

$$\widehat{\mathrm{F}}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \leq x),$$

où $\mathbb{1}(X_i \leq x) = 1$ *si* $X_i \leq x$*, 0 sinon.*

Pour x donné, $\widehat{\mathrm{F}}_n(x)$ est un estimateur fortement consistant de $\mathrm{F}(x)$. Le théorème de Glivenko-Cantelli étend ces résultats pour montrer la convergence en tant que fonction.

Théorème 8.9 (Glivenko-Cantelli). *Soit $X_1, \ldots, X_n$ un échantillon i.i.d. de la loi de fonction de répartition* F*, et $\widehat{\mathrm{F}}_n$ sa fonction de répartition empirique. Alors, quand $n \to \infty$,*

$$D_n = \sup_{x \in \mathbb{R}} |\widehat{\mathrm{F}}_n(x) - \mathrm{F}(x)| \to 0 \ \ p.s.$$

Le théorème de Kolmogorov-Smirnov identifie une normalisation appropriée et une loi limite :

Théorème 8.10 (Kolmogorov-Smirnov). *Soit $X_1, \ldots, X_n$ un échantillon i.i.d. d'une loi continue. Pour $x > 0$,*

$$\lim_{n \to \infty} \mathbb{P}(\sqrt{n} D_n > x) = 2 \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2(kx)^2}$$

La loi limite de $\sqrt{n} D_n$ est libre, elle ne dépend pas de F, et cette propriété se conserve à n fini et elle est tabulée dans les logiciels.

On en déduit la région de rejet du test de Kolmogorov-Smirnov de

$$(H_0) : \mathrm{F} = \mathrm{F}_0 \text{ contre } (H_1) : \mathrm{F} \neq \mathrm{F}_0$$

pour une loi continue :

$$\mathcal{R} = \{\sqrt{n} \sup_{x \in \mathbb{R}} |\widehat{\mathrm{F}}_n(x) - \mathrm{F}_0(x)| > q_{\sqrt{n} D_n, 1-\alpha}\}, \ \ \mathbb{P}(\mathcal{R}) \simeq \alpha$$

Il n'y a pas de conclusion générale pour comparer la puissance du test du Khi-deux et celle du test de Kolmogorov-Smirnov, mais ce dernier est souvent plus puissant.

L'exemple suivant teste l'adéquation d'un échantillon gaussien $\mathcal{N}(2,1)$ à une loi gaussienne centrée réduite :

$$(H_0) : X_i \sim \mathcal{N}(2,1) \text{ contre } (H_1) : X_i \not\sim \mathcal{N}(0,1)$$

```
> set.seed(10)
> n=10
> x=rnorm(n,mean=2,sd=1);x
 2.0187462 1.8157475 0.6286695 1.4008323 2.2945451 2.3897943 0.7919238
 1.6363240 0.3733273 1.7435216
> ks.test(x,"pnorm", 0,1)

One-sample Kolmogorov-Smirnov test

data:  x
D = 0.64555, p-value = 0.0001364
alternative hypothesis: two-sided
```

Comme $0.0001364 < 5\%$, le test est significatif pour rejeter (H_0), au risque (de première espèce) de 5%. Quand on veut tester de façon plus générale

$$(H_0) : X_i \sim \mathcal{N} \text{ contre } (H_1) : X_i \not\sim \mathcal{N}$$

l'hypothèse (H_0) devient composite à cause des paramètres inconnus sous (H_0). On peut alors estimer le paramètre θ sous (H_0) par $\widehat{\theta}$, et le remplacer dans la définition de la statistique de test $\widehat{D}_n = \sup_{x \in \mathbb{R}} |\widehat{\mathrm{F}}_n(x) - \mathrm{F}_{\widehat{\theta}}(x)|$ dont la loi n'est plus libre. Il faut alors étudier chaque famille individuellement, mais on peut utiliser le seuil du test de l'hypothèse simple vu ci-dessus, qui conduit à un test conservateur parce que $\widehat{D}_n$ tend à sous-estimer l'écart réel. Le code ci-dessous opère un test de normalité de Kolmogorov-Smirnov : dans le premier cas, l'espérance et l'écart type sont estimés et entrés comme valeur de paramètre sous (H_0). Dans le deuxième, l'échantillon est centré et réduit, puis testé contre une loi normale $\mathcal{N}(0,1)$.

```
> ks.test(x,"pnorm",mean=mean(x),sd=sd(x))
D = 0.17198, p-value = 0.882
alternative hypothesis: two-sided

> ks.test((x-mean(x))/sd(x),"pnorm", 0,1)
D = 0.17198, p-value = 0.882
alternative hypothesis: two-sided
```

Ces deux méthodologies sont identiques ici, et donnent la même pvalue $0.882 > 5\%$. On accepte l'hypothèse de normalité, mais avec un risque de seconde espèce inconnue. Le test étant conservatif, sa puissance peut être faible, et il peut être difficile de détecter (H_1).

8.3.3 Un test d'adéquation à la loi gaussienne

Il existe des tests spécifiques à une loi donnée, par exemple pour la loi gaussienne le test de **Shapiro-Wilks**

$$(H_0) : X_i \sim \mathcal{N} \text{ contre } (H_1) : X_i \not\sim \mathcal{N}$$

La statistique de test est

$$W = \frac{\left(\sum_i a_i X_{[i]}\right)^2}{\sum_i (X_i - \bar{X})^2}$$

où $X_{[i]}$ est la i-ème statistique d'ordre (la ième observation quand on réordonne les observations d'un échantillon i.i.d. par valeur croissante), et a_i les coordonnées du vecteur

$$a = \frac{m' V^{-1}}{\sqrt{m' V^{-1} V^{-1} m}}$$

et $m = (m_1, \ldots, m_n)$ est le vecteur des espérances des statistiques d'ordre d'un échantillon i.i.d. d'une loi gaussienne centrée réduite. La région de rejet est unilatérale à gauche (les petites valeurs de la statistique de test font rejeter le test). Les logiciels calculent la pvalue, comme par exemple la sortie de la fonction `shapiro.test` appliquée à l'échantillon de la section précédente. Les données ne sont pas significatives pour rejeter (H_0) (pvalue$> 5\%$) qu'on accepte avec un risque de seconde espèce inconnu.

```
> shapiro.test(x)
Shapiro-Wilk normality test
data:  x
W = 0.9284, p-value = 0.4323
```

Annexe A

Rappels de convergence

A.1 Définitions

Définition A.1. *Soit X_n une suite de v.a. et X une v.a.. On étudie le comportement de X_n quand n tend vers l'infini.*

— *La suite X_n* ***converge en loi*** *vers la v.a. X si*

$$\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x)$$

pour tout x où la fonction de répartition de X est continue. On parle aussi de convergence en distribution ou de convergence faible. On note $X_n \xrightarrow{\mathcal{L}} X$.

— *La suite X_n* ***converge en probabilité*** *vers la v.a. X si*

$$\forall \varepsilon, \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$$

On note : $X_n \xrightarrow{\mathcal{P}} X$.

— *La suite X_n converge en* ***presque sûrement*** *(convergence forte) vers la v.a. X si*

$$\mathbb{P}(\lim_{n\to\infty} |X_n - X| = 0\ ) = 1$$

On dit aussi $X_n - X \to 0$ avec probabilité 1 et on note $X_n \xrightarrow{p.s.} X$.

— *La suite X_n converge en* ***moyenne quadratique*** *vers la v.a. X si*

$$\mathbb{E}[(X_n - X)^2] \to 0$$

On dit aussi que X_n converge vers X dans L^2 et on note : $X_n \xrightarrow{L^2} X$.

Définitions équivalentes de la convergence en loi ou lemme de porte-manteau

— $X_n \xrightarrow{\mathcal{L}} X$ ssi $\mathbb{E}[h(X_n)] \to \mathbb{E}[h(X)]$ pour toute fonction continue et bornée h

— $X_n \xrightarrow{\mathcal{L}} X$ ssi $\mathbb{E}[h(X_n)] \to \mathbb{E}[h(X)]$ pour toute fonction lipschitzienne et bornée h
On dit qu'une fonction $H : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ est lipschitzienne de constante $K > 0$ ssi $||h(x) - h(y)|| \leq K||x - y||$. En particulier, toute fonction continûment dérivable sur un intervalle fermé borné est lipschitzienne.

A.2 Relations entre les convergences

Théorème A.2. *On a les relations suivantes :*

1. $X_n \xrightarrow{p.s.} X$ *implique* $X_n \xrightarrow{\mathcal{P}} X$
2. $X_n \xrightarrow{L^2} X$ *implique* $X_n \xrightarrow{\mathcal{P}} X$
3. $X_n \xrightarrow{\mathcal{P}} X$ *implique* $X_n \xrightarrow{\mathcal{L}} X$
4. *Soit c est une constante réelle.* $X_n \xrightarrow{\mathcal{P}} c$ *ssi* $X_n \xrightarrow{\mathcal{L}} c$

Remarque En général, les implications réciproques de 2. et 3. sont fausses.

— La convergence en loi n'implique pas celle en probabilité. Contre-exemple : Soit $X \sim \mathcal{N}(0,1)$ et $X_n = -X$ pour tout n. Alors, X_n a même loi que X pour tout n, donc $\lim_n \mathbb{P}(X_n \leq x) = \mathbb{P}(X \leq x)$ pour tout x, d'où $X_n \xrightarrow{\mathcal{L}} X$. Mais $\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|2X| > \varepsilon) \neq 0$, donc X_n ne converge pas vers 0 en probabilité

— La convergence en probabilité n'implique pas celle en moyenne quadratique, en particulier parce que la convergence en probabilité ne nécessite pas d'avoir un moment d'ordre deux fini. Contre-exemple : Soit $U_{[0;1]}$ une variable uniforme sur $[0;1]$ et $X_n = \sqrt{n}\mathbb{1}_{U \leq 1/n}$.
Alors, $\mathbb{P}(|X_n| > \varepsilon) = \mathbb{P}(U \leq 1/n) = 1/n \to 0$ donc $X_n \xrightarrow{\mathcal{P}} 0$. Mais $\mathbb{E}(X_n^2) = \int_0^{1/n} \sqrt{n}^2 dt = 1$ pour tout n

Le résultat suivant est très utile pour étudier les estimateurs :

Lemme A.3 (de l'application continue)**.** *Soit $X_1, \ldots, X_n \sim \mathcal{P}_\theta$. Soit g une fonction continue (au moins en tout point x d'un ensemble A tel que $X \in A) = 1$), et soit X_n une suite de variables aléatoires :*

— *Si* $X_n \xrightarrow{p.s.} X$, *alors* $g(X_n) \xrightarrow{p.s.} g(X)$
— *Si* $X_n \xrightarrow{\mathcal{P}} X$, *alors* $g(X_n) \xrightarrow{\mathcal{P}} g(X)$
— *Si* $X_n \xrightarrow{\mathcal{L}} X$, *alors* $g(X_n) \xrightarrow{\mathcal{L}} g(X)$

Un outil bien pratique pour montrer la convergence probabilité est l'inégalité de Tchébychev

Propriété A.4 (Inégalité de Bienaymé-Tchebychev)**.** *Soit T une v.a. telle que $\mathbb{E}(T^2) < +\infty$. Alors,*

$$\forall t > 0, \quad \mathbb{P}(\{|T - \mathbb{E}(T)| > t\}) \leq \frac{\mathrm{var}(T)}{t^2}$$

A.3 Convergence de couples de variables aléatoires

On peut généraliser la définition des convergences en probabilité à des couples de variables aléatoires.

Définition A.5. *Le couple de variables aléatoires (X_n, Y_n) converge en probabilité vers (X, Y) si, pour tout $\varepsilon > 0$,*

$$\mathbb{P}\left(d((X_n, Y_n), (X, Y)) > \varepsilon\right) \to 0$$

où d est la distance euclidienne de $\mathbb{R}^2$

$$d((x_1, y_1), (x_2, y_2)) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

On a le résultat suivant

Propriété A.6. *Si* $X_n \xrightarrow{\mathcal{P}} X$ *et* $Y_n \xrightarrow{\mathcal{P}} Y$*, alors* $(X_n, Y_n) \xrightarrow{\mathcal{P}} (X, Y)$.
Le lemme de l'application continue est encore vrai pour les vecteurs aléatoires.

On a donc équivalence entre la convergence en probabilité du couple et la convergence en probabilité de chacune des marginales

Attention Cette assertion est fausse pour la convergence en loi. La connaissance de chacune des marginales ne détermine pas en général la loi jointe du couple. Il y a des exceptions : si X_n et Y_n sont indépendantes, ou si l'une des deux marginales converge vers une constate

Propriété A.7. *Quelques propriétés de la convergence en loi de couples de v.a.*
- *Si* $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, Y)$*, alors* $X_n \xrightarrow{\mathcal{L}} X$ *et* $Y_n \xrightarrow{\mathcal{L}} Y$
  La réciproque est fausse en général
- *si* X_n *et* Y_n *sont indépendants, et* X *et* Y *indépendantes, alors*
  si $X_n \xrightarrow{\mathcal{L}} X$ *et* $Y_n \xrightarrow{\mathcal{L}} Y$*, alors* $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, Y)$

Lemme A.8 (Lemme de Slutsky)**.** *Si* $X_n \xrightarrow{\mathcal{L}} X$ *et* $Y_n \xrightarrow{\mathcal{L}} c$ *où* c *est une constante, alors* $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, c)$.
En appliquant cette convergence jointe à une fonction continue, on a en particulier
- $X_n + Y_n \xrightarrow{\mathcal{L}} X + c$
- $X_n Y_n \xrightarrow{\mathcal{L}} cX$
- $X_n / Y_n \xrightarrow{\mathcal{L}} X/c$

Preuve. On utilise la caractérisation de la convergence en loi du lemme porte-manteau. Soit h une fonction lipschitzienne de constante K bornée par M.

$$|\mathbb{E}[h(X_n, Y_n)] - \mathbb{E}[h(X, c)]| \leq |\mathbb{E}[h(X_n, Y_n)] - \mathbb{E}[h(X_n, c)]| + |\mathbb{E}[h(X_n, c)] - \mathbb{E}[h(X, c)]|$$

Le second terme tend vers 0 car $X_n \xrightarrow{\mathcal{L}} X$ et en appliquant le lemme de l'application continue. On majore maintenant le premier terme

$$\begin{aligned}
|\mathbb{E}[h(X_n, Y_n)] - \mathbb{E}[h(X_n, c)]| \leq & |\left(\mathbb{E}[h(X_n, Y_n)] - \mathbb{E}[h(X_n, c)]\right) \mathbb{1}_{||Y_n - c|| > \varepsilon}| \\
& + |\left(\mathbb{E}[h(X_n, Y_n)] - \mathbb{E}[h(X_n, c)]\right) \mathbb{1}_{||Y_n - c|| \leq \varepsilon}| \\
& \leq 2 \sup_{x,y} ||h(x, y)||\ \mathbb{P}(||Y_n - c|| > \varepsilon) + K\mathbb{E}[||Y_n - c|| \mathbb{1}_{||Y_n - c|| \leq \varepsilon}] \\
& \leq 2M\ \mathbb{P}(||Y_n - c|| > \varepsilon) + K\varepsilon \mathbb{P}(||Y_n - c|| \leq \varepsilon)
\end{aligned}$$

Comme $Y_n \xrightarrow{\mathcal{L}} c$ le premier terme tend vers 0, et on majore le second terme par $K\varepsilon$ pour tout $\varepsilon > 0$. ◊

Annexe B

Tables

B.1 Table de probabilité de la loi gaussienne

t	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,5	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,10	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,20	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,30	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,40	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,50	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,60	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,70	0,758	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,80	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,90	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,00	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,10	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,20	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,30	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,40	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,50	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,60	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,70	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,80	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,90	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,00	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,10	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,20	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,30	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,40	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,50	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,60	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,70	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,80	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,90	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

t	3	3,1	3,2	3,3	3,4	3,5	3,6	3,8	4	4,5
P	0,99865	0,99903	0,99931	0,99952	0,99966	0,99977	0,99984	0,99993	0,99997	1

t	1,282	1,645	1,96	2,326	2,576	3,09
P	0,9	0,95	0,975	0,99	0,995	0,999

t	-3,719	-4,2649	-4,7534	-5,1993	-5,612	-5,9978
P	10^ -4	10^ -5	10^ -6	10^ -7	10^ -8	10^ -9

B.2 Quantiles de la loi de Student

n \ α	0,75	0,900	0,950	0,975	0,990	0,995
1	1,000	3,078	6,314	12,706	31,821	63,657
2	0,817	1,886	2,920	4,303	6,965	9,925
3	0,765	1,638	2,353	3,182	4,541	5,841
4	0,741	1,533	2,132	2,776	3,747	4,604
5	0,727	1,476	2,015	2,571	3,365	4,032
6	0,718	1,440	1,943	2,447	3,143	3,707
7	0,711	1,415	1,895	2,365	2,998	3,499
8	0,706	1,397	1,860	2,306	2,896	3,355
9	0,703	1,383	1,833	2,262	2,821	3,250
10	0,700	1,372	1,812	2,228	2,764	3,169
11	0,697	1,363	1,796	2,201	2,718	3,106
12	0,696	1,356	1,782	2,179	2,681	3,055
13	0,694	1,350	1,771	2,160	2,650	3,012
14	0,692	1,345	1,761	2,145	2,624	2,977
15	0,691	1,341	1,753	2,131	2,602	2,947
16	0,690	1,337	1,746	2,120	2,583	2,921
17	0,689	1,333	1,740	2,110	2,567	2,898
18	0,688	1,330	1,734	2,101	2,552	2,878
19	0,688	1,328	1,729	2,093	2,539	2,861
20	0,687	1,325	1,725	2,086	2,528	2,845
21	0,686	1,323	1,721	2,080	2,518	2,831
22	0,686	1,321	1,717	2,074	2,508	2,819
23	0,685	1,319	1,714	2,069	2,500	2,807
24	0,685	1,318	1,711	2,064	2,492	2,797
25	0,684	1,316	1,708	2,060	2,485	2,787
26	0,684	1,315	1,706	2,056	2,479	2,779
27	0,684	1,314	1,703	2,052	2,473	2,771
28	0,683	1,313	1,701	2,048	2,467	2,763
29	0,683	1,311	1,699	2,045	2,462	2,756
30	0,683	1,310	1,697	2,042	2,457	2,750
40	0,681	1,303	1,684	2,021	2,423	2,704
60	0,679	1,296	1,671	2,000	2,390	2,660
120	0,677	1,289	1,658	1,980	2,358	2,617
	0,675	1,282	1,645	1,960	2,327	2,576

B.3 Quantiles de la loi du Khi-deux

n \	0,01	0,025	0,05	0,1	0,5	0,9	0,95	0,975	0,99	0,995
1	0,0002	0,001	0,004	0,02	0,46	2,71	3,84	5,02	6,63	7,88
2	0,02	0,05	0,10	0,21	1,39	4,61	5,99	7,38	9,21	10,6
3	0,12	0,22	0,35	0,58	2,37	6,25	7,81	9,35	11,34	12,84
4	0,30	0,48	0,71	1,06	3,36	7,78	9,49	11,14	13,28	14,86
5	0,55	0,83	1,15	1,61	4,35	9,24	11,07	12,83	15,09	16,75
6	0,87	1,24	1,64	2,20	5,35	10,64	12,59	14,45	16,81	18,55
7	1,24	1,69	2,17	2,83	6,35	12,02	14,07	16,01	18,48	20,28
8	1,65	2,18	2,73	3,49	7,34	13,36	15,51	17,53	20,09	21,95
9	2,09	2,70	3,33	4,17	8,34	14,68	16,92	19,02	21,67	23,59
10	2,56	3,25	3,94	4,87	9,34	15,99	18,31	20,48	23,21	25,19
11	3,05	3,82	4,57	5,58	10,34	17,28	19,68	21,92	24,72	26,76
12	3,57	4,40	5,23	6,30	11,34	18,55	21,03	23,34	26,22	28,3
13	4,11	5,01	5,89	7,04	12,34	19,81	22,36	24,74	27,69	29,82
14	4,66	5,63	6,57	7,79	13,34	21,06	23,68	26,12	29,14	31,32
15	5,23	6,26	7,26	8,55	14,34	22,31	25	27,49	30,58	32,8
16	5,81	6,91	7,96	9,31	15,34	23,54	26,3	28,85	32	34,27
17	6,41	7,56	8,67	10,09	16,34	24,77	27,59	30,19	33,41	35,72
18	7,01	8,23	9,39	10,86	17,34	25,99	28,87	31,53	34,81	37,16
19	7,63	8,91	10,12	11,65	18,34	27,2	30,14	32,85	36,19	38,58
20	8,26	9,59	10,85	12,44	19,34	28,41	31,41	34,17	37,57	40
21	8,90	10,30	11,59	13,24	20,34	29,62	32,67	35,48	38,93	41,4
22	9,54	11,00	12,34	14,04	21,34	30,81	33,92	36,78	40,29	42,8
23	10,20	11,70	13,09	14,85	22,34	32,01	35,17	38,08	41,64	44,18
24	10,90	12,40	13,85	15,66	23,34	33,2	36,42	39,36	42,98	45,56
25	11,50	13,10	14,61	16,47	24,34	34,38	37,65	40,65	44,31	46,93
26	12,20	13,80	15,38	17,29	25,34	35,56	38,89	41,92	45,64	48,29
27	12,90	14,60	16,15	18,11	26,34	36,74	40,11	43,19	46,96	49,64
28	13,60	15,30	16,93	18,94	27,34	37,92	41,34	44,46	48,28	50,99
29	14,30	16,00	17,71	19,77	28,34	39,09	42,56	45,72	49,59	52,34
30	15,00	16,80	18,49	20,60	29,34	40,26	43,77	46,98	50,89	53,67
40	22,20	24,40	26,51	29,05	39,34	51,81	55,76	59,34	63,69	66,77
50	29,70	32,40	34,76	37,69	49,34	63,17	67,5	71,42	76,15	79,49
60	37,50	40,50	43,19	46,46	59,34	74,4	79,08	83,3	88,38	91,95
70	45,40	48,80	51,74	55,33	69,33	85,53	90,53	95,02	100,43	104,21
80	53,50	57,20	60,39	64,28	79,33	96,58	101,88	106,63	112,33	116,32
90	61,80	65,60	69,13	73,29	89,33	107,57	113,15	118,14	124,12	128,3
100	70,10	74,20	77,93	82,36	99,33	118,5	124,34	129,56	135,81	140,17

B.4 Quantiles de la loi de Fisher

p = degré du dénominateur \ n = degrés du numérateur																			
a=0,99	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	inf
1	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40	99,42	99,43	99,45	99,46	99,47	99,47	99,48	99,49	99,50
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35	27,23	27,05	26,87	26,69	26,60	26,50	26,41	26,32	26,22	26,13
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,37	14,20	14,02	13,93	13,84	13,75	13,65	13,56	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11	9,02
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,97	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,47	6,31	6,16	6,07	5,99	5,91	5,82	5,74	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,95	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,40	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00	3,91
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,40	4,25	4,10	4,02	3,94	3,86	3,78	3,69	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,45	3,36
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25	3,17
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,80	3,66	3,51	3,43	3,35	3,27	3,18	3,09	3,01
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,96	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,55	3,41	3,26	3,18	3,10	3,02	2,93	2,84	2,75
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,46	3,31	3,16	3,08	3,00	2,92	2,83	2,75	2,65
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,37	3,23	3,08	3,00	2,92	2,84	2,75	2,66	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,30	3,15	3,00	2,92	2,84	2,76	2,67	2,58	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,52	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,17	3,03	2,88	2,80	2,72	2,64	2,55	2,46	2,36
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,12	2,98	2,83	2,75	2,67	2,58	2,50	2,40	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,07	2,93	2,78	2,70	2,62	2,54	2,45	2,35	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,03	2,89	2,74	2,66	2,58	2,49	2,40	2,31	2,21
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	2,99	2,85	2,70	2,62	2,54	2,45	2,36	2,27	2,17
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,84	2,70	2,55	2,47	2,39	2,30	2,21	2,11	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,66	2,52	2,37	2,29	2,20	2,11	2,02	1,92	1,81
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,73	1,60
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,34	2,19	2,03	1,95	1,86	1,76	1,66	1,53	1,38
inf	6,64	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,19	2,04	1,88	1,79	1,70	1,59	1,48	1,33	1,05

p = degré du dénominateur \ n = degrés du numérateur																			
a=0,95	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	inf
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,43	19,45	19,45	19,46	19,47	19,48	19,49	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	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	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,37
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,41
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35	1,26
inf	3,84	3,00	2,61	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,52	1,46	1,40	1,32	1,22	1,03

Table des matières