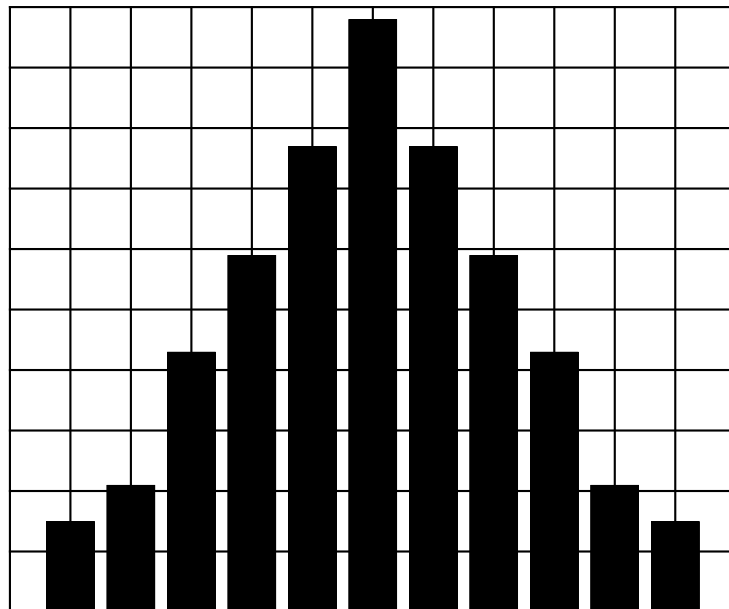


# (Petit) résumé de Probabilité et Statistique

Matteo DeGiacomi  
Samuel Robyr



23 juin 2005

Copyright © 2004-2005 Matteo DeGiacomi, Samuel Robyr.

Ce document est libre ; vous pouvez le redistribuer et/ou le modifier selon les termes de la Licence de Documentation Libre GNU (GNU Free Documentation License) publiée par la Free Software Foundation (version 1.2 ou toute autre version ultérieure choisie par vous).

Ce document est diffusé en espérant qu'il sera utile, mais SANS AUCUNE GARANTIE, ni explicite ni implicite. Reportez-vous à la licence FDL pour plus de détails.

## Préface

Ce document est un résumé du cours *Probabilité et Statistique* tenu par le professeur Thomas Mountford en 2003-2004 à une classe de malheureux informaticiens de deuxième année. Ce chef d'oeuvre résume presque tous ce qu'on a vu en classe (oui oui, on était là, nous !), les polycopiés du cours et des exercices, et le tout mélangé avec quelques petite notion tirée du livre *Introduction à la statistique* de Stephan Morgenthaler. Les seuls arguments qui n'ont pas été abordés dans ce document sont les régressions linéaires, ANOVA et R.

Naturellement, nous sommes entièrement responsables de tout erreur que l'on puisse trouver ici... Hélas, on est loin d'être des génies des stats ! On invite donc tous les chasseurs d'erreurs de soumettre à notre royale attention toute imprécision, pour qu'on puisse améliorer ce résumé !

Pour la préparation de ce document l'aide des étudiants a été sollicitée. De nombreuses personnes ont répondu et ont fourni des corrections, des suggestions et du texte pour améliorer ce document. On les remercie sincèrement. Un merci en particulier à : *Daniel Peppicelli*.

La version courante de ce document est disponible à l'adresse :

<http://epfl.neoch.net/>

Pour finir, nous ne somme en aucun cas responsables de votre note au propé;-). ... Mais on vous souhaite bonne chance à tous !

Matteo DeGiacomi <matteothomas.degiacom@epfl.ch>

Samuel Robyr <samuel.rob@epfl.ch>

*Faculté Informatique et Communications  
École Polytechnique Fédérale de Lausanne*

# Table des matières

<b>I</b>	<b>Probabilité</b>	<b>6</b>
<b>1</b>	<b>Analyse combinatoire</b>	<b>7</b>
1.1	Permutations . . . . .	7
1.2	Combinaisons . . . . .	7
1.2.1	Propriétés . . . . .	7
<b>2</b>	<b>Axiomes de probabilité</b>	<b>8</b>
2.1	Propriétés des opérations sur les événements . . . . .	8
2.2	Théorèmes fondamentales . . . . .	8
2.3	Calcul des probabilités : le comptage . . . . .	8
<b>3</b>	<b>Probabilité conditionnelle</b>	<b>9</b>
3.1	Formule de Bayes . . . . .	9
3.1.1	Formule des probabilités totales . . . . .	9
3.1.2	Formule de Bayes . . . . .	9
3.2	Indépendance . . . . .	9
<b>4</b>	<b>Variables aléatoires</b>	<b>10</b>
4.1	Fonctions de distribution (ou répartition) . . . . .	10
4.2	Fonctions de densité . . . . .	10
4.3	Distributions discrètes . . . . .	10
4.3.1	La distribution binomiale . . . . .	11
4.3.2	Les distributions géométriques et binomiales négatives . . . . .	11
4.3.3	La distribution hypergéométrique . . . . .	11
4.3.4	La distribution de Poisson . . . . .	12
4.4	Distributions continues . . . . .	12
4.4.1	La distribution uniforme . . . . .	12
4.4.2	La distribution exponentielle . . . . .	13
4.4.3	La distribution gamma . . . . .	13
4.4.4	La distribution gaussienne ou normale . . . . .	13
4.5	Variables aléatoires conjointes . . . . .	14
4.5.1	Fonction de distribution conjointe et marginale . . . . .	14
4.5.2	Fréquence conjointe . . . . .	14
4.5.3	Densité conjointe . . . . .	15
4.5.4	Distributions conditionnelles . . . . .	15
4.5.5	Fonctions de variables aléatoires conjointes . . . . .	15

<b>5</b>	<b>Espérance et Variance</b>	<b>17</b>
5.1	Espérance . . . . .	17
5.1.1	Quelque propriété . . . . .	17
5.2	Variance . . . . .	18
5.2.1	Quelque propriété . . . . .	18
5.2.2	Moyenne carrée d'erreur . . . . .	18
5.3	Covariance et corrélation . . . . .	18
5.3.1	Propriétés . . . . .	18
5.4	Espérance conditionnelle . . . . .	19
5.5	Prediction . . . . .	19
5.6	La fonction génératrice des moments . . . . .	19
<b>6</b>	<b>Distributions dérivées de la distribution normale</b>	<b>20</b>
<b>II</b>	<b>Statistique</b>	<b>21</b>
<b>7</b>	<b>Échantillonnage</b>	<b>22</b>
7.1	Échantillon simple . . . . .	22
7.2	Échantillon stratifié . . . . .	23
<b>8</b>	<b>Estimation de paramètres</b>	<b>25</b>
8.1	La méthode des moments . . . . .	25
8.2	La méthode du maximum de vraisemblance . . . . .	25
8.3	Information et intervalles de confiance . . . . .	26
8.3.1	Intervalles de confiances simples . . . . .	26
8.3.2	Intervalles de confiances quand la variance $\sigma^2$ est inconnue . . . . .	26
8.3.3	Information . . . . .	26
8.3.4	Intervalles de confiance avec $\mu$ estimé par la méthode du maximum de vraisemblance . . . . .	27
<b>9</b>	<b>Tests d'hypothèses</b>	<b>28</b>
9.1	Test optimal . . . . .	29
9.2	Test du rapport du maximum de vraisemblance . . . . .	29
9.3	z-test pour un échantillon . . . . .	29
9.4	t-test pour un échantillon . . . . .	30
9.5	z-test et t-test pour deux échantillons . . . . .	30
9.6	z-test et t-test pour deux échantillons appariés . . . . .	31
9.7	Tests khi-deux . . . . .	31
9.7.1	Test d'indépendance . . . . .	31
9.8	Tests non paramétriques (Test de Wilcoxon) . . . . .	32
9.8.1	Échantillons non appariés . . . . .	32
9.8.2	Échantillons appariés . . . . .	32
<b>III</b>	<b>Tables</b>	<b>33</b>

# Liste des tableaux

1	Distribution d'une variable normale centrée réduite $N(0, 1)$ . . . . .	34
2	Table des quantiles de la loi $\chi^2_v$ . . . . .	35
3	Propriétés des lois discrètes et continues les plus souvent utilisés . . . . .	36
4	Approximation de quelque loi . . . . .	37

# **Première partie**

## **Probabilité**

# Chapitre 1

## Analyse combinatoire

### 1.1 Permutations

Le nombre de rangements possibles de  $k$  objets choisis parmi  $n$  est

$$P_k^n = n \cdot (n-1) \cdot (n-2) \cdots (n-k+1)$$

En particulier le nombre d'arrangements ordonnés<sup>1</sup> possibles de  $n$  objets correspond à :

$$P_n^n = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1 = n!$$

### 1.2 Combinaisons

Le nombre de sous-ensembles de taille  $k$  que l'on peut former avec  $n$  objets distincts est<sup>2</sup>

$$C_k^n = \binom{n}{k} = \frac{P_k^n}{k!} = \frac{n \cdot (n-1) \cdots (n-k+1)}{k!} = \frac{n!}{(n-k)!k!}$$

#### 1.2.1 Propriétés

- $\binom{n}{k} = \binom{n}{n-k}$
- $\binom{n}{0} = 1$
- $\sum_{k=0}^n \binom{n}{k} = 2^n$ , nombre de sous-ensembles dans un ensemble de  $n$  objets.
- $\binom{n}{n_1 \dots n_r} = \frac{n!}{n_1! n_2! \dots n_r!}$ , regroupement de  $n$  objets en  $r$  classes de dimension  $n_i$ .

---

<sup>1</sup> $abc \neq cba$

<sup>2</sup>Le rangement des éléments dans l'ensemble ne compte pas :  $abc = acb$

# Chapitre 2

## Axiomes de probabilité

### 2.1 Propriétés des opérations sur les événements

1.  $A \cup A^c = \Omega$
2. Associativité de l'union :  $(E \cup F) \cup G = E \cup (F \cup G)$
3. Associativité de l'intersection :  $(E \cap F) \cap G = E \cap (F \cap G)$
4. Relations de De Morgan :  $(A \cup B)^c = A^c \cap B^c$  et  $(A \cap B)^c = A^c \cup B^c$   
En général :  $(E_1 \cup \dots \cup E_n)^c = E_1^c \cap \dots \cap E_n^c$  et  $(E_1 \cap \dots \cap E_n)^c = E_1^c \cup \dots \cup E_n^c$
5. Distributivité :  $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$  et  $(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$

### 2.2 Théorèmes fondamentales

1.  $P(A^c) = 1 - P(A)$ <sup>1</sup>
2.  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

### 2.3 Calcul des probabilités : le comptage

$$P(A) = \frac{|A|}{|\Omega|}$$

où  $\Omega$  est l'ensemble de tous les résultats possibles. Si une expérience se décompose en  $r$  étapes, chacune avec  $m_i$  résultats possibles, la totalité des résultats possibles est  $m_1 m_2 \dots m_r$ .

---

<sup>1</sup>En peut déduire :

- $P(A \cup B) = 1 - P(A^c \cap B^c)$
- $P(A^c \cap B^c \cap C^c) = 1 - P(A \cup B \cup C)$
- ...



# Chapitre 3

## Probabilité conditionnelle

La probabilité que A s'avère sachant que B est réalisé est dite *probabilité conditionnelle* et elle vaut :

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

ou plus généralement (avec  $B^1 \dots B^n$  disjoints) :

$$P(A) = \sum_{j=1}^n P(B_j) \cdot P(A|B_j)$$

Elle peut aussi se réécrire comme

$$P(A \cap B) = P(B) \cdot P(A|B) \quad \text{ou} \quad P(A \cap B) = P(A) \cdot P(B|A)$$

### 3.1 Formule de Bayes

#### 3.1.1 Formule des probabilités totales

$$P(A) = P(A \cap B) + P(A \cap B^c) = P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c)$$

#### 3.1.2 Formule de Bayes

$$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B)}$$

En utilisant la formule des probabilités totales on peut écrire la formule de Bayes sous la forme :

$$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c)}$$

### 3.2 Indépendance

Deux événements  $E$  et  $F$  sont *indépendants* si  $P(E) = P(E|F)$  et  $P(F) = P(F|E)$ . Cela équivaut à  $P(E \cap F) = P(E) \cdot P(F)$

# Chapitre 4

## Variables aléatoires

Une variable aléatoire est *discrète* si elle ne peut que avoir des valeurs définies, en cas contraire elle est *continue*.

### 4.1 Fonctions de distribution (ou répartition)

La *fonction de distribution* ou *fonction de répartition*, d'une variable aléatoire  $X$  est définie par

$$F_X(x) = P(X \leq x), \quad -\infty < x < \infty$$

C'est une fonction croissante qui satisfait

$$\lim_{b \rightarrow \infty} F_X(x) = 1 \quad \text{et} \quad \lim_{b \rightarrow -\infty} F_X(x) = 0$$

### 4.2 Fonctions de densité

La *fonction de densité* d'une variable aléatoire continue est continue par morceaux, **toujours plus grande ou égale à 0** et **l'aire au dessus d'elle est égale à 1**. Si la variable est discrète au lieu d'une intégrale on a une somme.

La probabilité que  $X$  se trouve dans l'intervalle  $[a, b]$  est

$$P(a \leq X \leq b) = P(a < X < b) = \int_a^b f(x) dx = F(b) - F(a)$$

La relation entre ces deux fonctions est :

$$F(X) = P(X \leq x) = \int_{-\infty}^x f(u) du, \quad F(X)' = f(x)$$

Ils existent plusieurs fonction de distributions caractéristiques, à dépendance du type de expérience aléatoire qu'on effectue.

### 4.3 Distributions discrètes

Une variable aléatoire ne pouvant prendre qu'une quantité dénombrable de valeurs est dite *discrète*. Pour une telle variable  $X$ , on définit sa *loi de probabilité*  $p$  par

$$p(a) = P(X = a)$$

On peut ainsi exprimer la fonction de répartition  $F$  :

$$F(a) = \sum_{x \leq a} p(x)$$

Si  $X$  est une variable aléatoire discrète avec une fonction de fréquence  $p(x)$ , l'espérance de  $X$  est

$$E[X] = \sum_{i=1}^n x_i p(x_i)$$

### 4.3.1 La distribution binomiale

Supposons qu'on effectue  $n$  épreuves identiques indépendantes et que chaque expérience ait comme résultat «succès» avec une probabilité  $p$  ou «échec» avec une probabilité  $(1 - p)$ . La variable aléatoire  $X$  qui compte le nombre de succès est dite variable aléatoire *binomiale* de paramètres  $(n, p)$  ; on note  $X \sim \text{Bin}(n, p)$ .

La loi de probabilité est donnée par

$$p(r) = P(X = r) = \binom{n}{r} p^r (1 - p)^{n-r}$$

Si  $X$  et  $Y$  sont deux variables aléatoires indépendantes,  $X + Y \sim \text{Bin}(n + m, p)$ .

Si  $n = 1$  on parle de distribution de Bernoulli.

### 4.3.2 Les distributions géométriques et binomiales négatives

On effectue des épreuves indépendantes. Chaque expérience a pour résultat «succès» avec une probabilité  $p$  ou «échec» avec une probabilité  $(1 - p)$ . Le nombre total d'essais nécessaires pour avoir un premier succès, noté  $X$ , est une variable aléatoire géométrique de paramètre  $p$ . La fonction de densité géométrique est

$$P(X = k) = (1 - p)^{k-1} p, \quad k = 1, 2, \dots$$

La distribution géométrique est un cas particulier ( $r = 1$ ) de la loi binomiale négative, définie ci-dessous.

On suppose maintenant qu'on effectue des épreuves indépendantes jusqu'à ce qu'on obtienne  $r$  succès. La fonction de densité est

$$P(X = k) = \binom{k-1}{r-1} p^r (1 - p)^{k-r}$$

On dit que  $X$  est une variable aléatoire *binomiale négative* de paramètres  $p$  et  $r$ .

### 4.3.3 La distribution hypergéométrique

Supposons qu'une urne contienne  $n$  boules, dont  $r$  sont noires et  $(n - r)$  blanches. On note  $X$  le nombre de balles noires dans un ensemble de  $m$  balles prises sans remise. On a :

$$P(X = k) = \frac{\binom{r}{k} \binom{n-r}{m-k}}{\binom{n}{m}}$$

### 4.3.4 La distribution de Poisson

La fonction de densité de *Poisson* de paramètre  $\lambda$  ( $\lambda > 0$ ) est

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots$$

On note  $X \sim P(\lambda)$ .

### Approximation poissonnienne de lois binomiales

On peut utiliser les variables aléatoire de Poisson pour approximer des variables aléatoires binomiales de paramètres  $(n, p)$  pour autant que  $n$  soit grand et  $p$  assez petit pour que  $\lambda = np$  soit d'ordre de grandeur moyes (cf table 4, page 37).

## 4.4 Distributions continues

La fonction de densité  $f$  d'une variable aléatoire continue est toujours positive, est continue par morceaux et l'aire au dessus d'elle est égale à 1.

La probabilité que  $X$  se trouve dans l'intervalle  $I = ]a, b[$  est donnée par

$$P(X \in I) = P(a \leq X \leq b) = P(a < X < b) = \int_a^b f(x) dx$$

La fonction  $f$  est appelée *densité de probabilité* (ou *fonction de densité*) de la variable aléatoire  $X$ .

$$f(x) = \begin{cases} X & \text{si } a \leq X \leq b \\ 0 & \text{sinon} \end{cases}$$

La *fonction de répartition* pour une variable aléatoire continue  $X$  est définie par

$$P(X < a) = P(X \leq a) = F_X(a) = \int_{-\infty}^a f(x) dx$$

de même

$$P(X > b) = P(X \geq b) = F_X(b) = \int_b^{\infty} f(x) dx$$

Si  $X$  est une variable aléatoire continue avec une fonction de densité  $f(x)$ , alors pour toute fonction réelle  $g$  on aura

$$E[g(x)] = \int_{-\infty}^{+\infty} g(x)f(x) dx$$

### 4.4.1 La distribution uniforme

La distribution *uniforme* (notée  $U[a, b]$ ), est définie sur un intervalle  $[a, b]$ , est nulle en dehors de cet intervalle :

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b, \\ 0 & \text{si } x > b \text{ ou } x < a \end{cases}$$

On peut alors facilement calculer la distribution  $F(x)$  :

$$F(x) = \int_{-\infty}^x f(u) du = \begin{cases} 0 & \text{si } x \leq a, \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b, \\ 1 & \text{si } x \geq b \end{cases}$$

### 4.4.2 La distribution exponentielle

La fonction de densité exponentielle est

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

La fonction de distribution est

$$F(a) = P(X \leq a) = \int_{-\infty}^a f(x) dx = \begin{cases} 1 - e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

### 4.4.3 La distribution gamma

La fonction de densité *gamma* dépend de deux paramètres  $t$  et  $\lambda$ , avec  $t, \lambda > 0$  :

$$f(x) = \begin{cases} \frac{\lambda^t e^{-\lambda x} x^{t-1}}{\Gamma(t)} & \text{si } t \geq 0 \\ 0 & \text{si } t < 0 \end{cases}$$

La fonction Gamma,  $\Gamma(t)$  est définie par

$$\Gamma(t) = \int_0^{+\infty} e^{-y} y^{t-1} dy$$

La loi  $\Gamma(1, \lambda)$  coïncide avec une exponentielle de paramètre  $\lambda$ .

### 4.4.4 La distribution gaussienne ou normale

Elle est également appelée *distribution normale*.

La fonction de densité gaussienne de paramètres  $\mu$  et  $\sigma$ ,  $-\infty < \mu < +\infty$ ,  $\sigma > 0$  est :

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

On note  $X \sim N(\mu, \sigma^2)$ . Or  $X$  est une variable normalement distribuée de paramètres  $\mu$  et  $\sigma^2$ . La variable  $Z = (X - \mu)/\sigma$  est normalement distribuée de paramètres 0 et 1. Une variable normale ayant ces deux paramètres est dite *standard*. La fonction de répartition d'une *variable normale standard* est

$$P(X \leq a) = P\left(\frac{X - \mu}{\sigma} \leq \frac{a - \mu}{\sigma}\right) = P\left(N(0, 1) \leq \frac{a - \mu}{\sigma}\right) = \Phi\left(\frac{a - \mu}{\sigma}\right) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy$$

Les valeurs  $\Phi(x)$  pour des arguments  $x$  non négatifs sont données dans la table 1 (pag. 34). Pour les arguments  $x$  négatifs, on calcule  $\Phi(x)$  grâce à l'équation

$$\Phi(-x) = 1 - \Phi(x)$$

**Exemple 4.1.** Soit  $X \sim N(\mu, \sigma^2) = N(2, 9)$ . On a les probabilités suivantes :

$$\begin{aligned} P(X < 0) &= P\left(\frac{X - \mu}{\sigma} \leq \frac{a - \mu}{\sigma}\right) = P\left(\frac{X - 2}{3} < \frac{0 - 2}{3}\right) = P\left(\frac{X - 2}{3} < -\frac{2}{3}\right) = P\left(Z < -\frac{2}{3}\right) \\ &= \Phi\left(-\frac{2}{3}\right) = 1 - \Phi\left(\frac{2}{3}\right) \approx 1 - 0.745 = 0.255 \end{aligned}$$

$$\begin{aligned}
 P(X > 0) &= P\left(\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right) = P\left(\frac{X-2}{3} > \frac{0-2}{3}\right) = P\left(\frac{X-2}{3} > -\frac{2}{3}\right) = P\left(Z > -\frac{2}{3}\right) \\
 &= 1 - \Phi\left(-\frac{2}{3}\right) = 1 - \left(1 - \Phi\left(\frac{2}{3}\right)\right) = \Phi\left(\frac{2}{3}\right) \approx 0.745
 \end{aligned}$$

$$\begin{aligned}
 P(4 < X < 5) &= P\left(\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right) = P\left(\frac{4-2}{3} < \frac{X-2}{3} < \frac{5-2}{3}\right) = P\left(\frac{2}{3} < \frac{X-2}{3} < 1\right) \\
 &= \Phi(1) - \Phi\left(\frac{2}{3}\right) \approx 0.841 - 0.745 = 0.096
 \end{aligned}$$

■

## 4.5 Variables aléatoires conjointes

### 4.5.1 Fonction de distribution conjointe et marginale

Un couple de variables aléatoires  $(X, Y)$  peut être vue comme un point aléatoire de  $\mathbb{R}^2$ .

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv, \quad f(x, y) = \frac{\delta^2}{\delta x \delta y} F(x, y)$$

Pour ce couple la fonction de *distribution conjointe* ou *simultanée*  $F$  est définie par

$$F(a, b) = P(X \leq a, Y \leq b), \quad -\infty < a, b < \infty$$

En connaissant la fonction de densité conjointe  $f(x, y)$  on peut calculer la *distribution marginale* de  $X$  et  $Y$  :

$$\begin{aligned}
 f_X(x) &= F'_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy \\
 f_Y(y) &= F'_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx
 \end{aligned}$$

*Remarque 4.1.* La probabilité de tous les événements simultanément relatifs à  $X$  et  $Y$  peut être calculée grâce aux fonctions conjointes définies plus haut, ie :

$$\begin{aligned}
 P(X > a, Y > b) &= 1 - P((X > a, Y > b)^c) = \\
 &= 1 - P((X > a)^c \cup (Y > b)^c) = \\
 &= 1 - P((X \leq a) \cup (Y \leq b)) = \\
 &= 1 - [P(X \leq a) + P(Y \leq b) - P(X \leq a, Y \leq b)] = \\
 &= 1 - F_X(a) - F_Y(b) + F(a, b)
 \end{aligned}$$

### 4.5.2 Fréquence conjointe

Dans le cas où  $X$  et  $Y$  sont deux variables aléatoires discrètes, il est pratique de définir la fonction  $p$  suivante dite. La fonction de *fréquence conjointe* ou *loi de probabilité conjointe* de  $X$  et  $Y$  :

$$p(a, b) = P(X = a, Y = b)$$

On déduit la loi de probabilité marginale :

$$p_X(a) = P(X = a) = \sum_{b: p(a, b) > 0} p(a, b) \quad \text{et} \quad p_Y(b) = P(Y = b) = \sum_{a: p(a, b) > 0} p(a, b)$$

### 4.5.3 Densité conjointe

Les variables  $X$  et  $Y$  sont dites *conjointement continues* s'il existe une fonction  $f$  de deux arguments réels ayant pour tout sous-ensemble  $C$  du plan la propriété suivante :

$$P((X, Y) \in C) = \iint_{(x,y) \in C} f(x, y) dx dy$$

La fonction  $f$  est appelée *densité conjointe*<sup>1</sup> de  $X$  et  $Y$ . Soit  $C = \{(x, y) | x \in A \text{ et } y \in B\}$ . On obtient

$$P(X \in A, Y \in B) = \int_B \int_A f(x, y) dx dy$$

Or

$$F(a, b) = \int_{-\infty}^b \int_{-\infty}^a f(x, y) dx dy \quad \Rightarrow \quad f(a, b) = \frac{\partial^2}{\partial a \partial b} F(a, b)$$

On peut ainsi calculer la *densité marginale* de  $X$  et  $Y$  :

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy \quad \text{et} \quad f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx$$

### 4.5.4 Distributions conditionnelles

#### Cas discret

Si  $X$  et  $Y$  sont deux variables aléatoires discrètes conjointes, la probabilité conditionnelle de  $X$  sous la condition  $Y = y$  est

$$p_{X|Y}(x, y) = P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p(x, y)}{p_Y(y)} \quad \text{si } f_X(x) > 0 \text{ et 0 sinon}$$

C'est une *fonction de fréquence*. Si  $X$  et  $Y$  sont indépendants  $p_{X|Y}(x, y) = f_X(x)$ .

#### Cas continu

Soient  $X$  et  $Y$  des variables de densité conjointe  $f(x, y)$ . On définit la *densité conditionnelle* de  $X$  sous la condition  $Y = y$  et lorsque  $f_Y(y) > 0$  par la relation

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} \quad \text{si } f_Y(y) > 0 \text{ et 0 sinon}$$

Connaissant la densité marginale  $f_Y(y)$  et la densité conditionnelle  $f_{X|Y}(x|y)$ , on retrouve facilement la densité conjointe :

$$f(x, y) = f_Y(y) \cdot f_{X|Y}(x|y)$$

### 4.5.5 Fonctions de variables aléatoires conjointes

#### Somme

Soit  $Z = X + Y$ . Sa distribution est :

$$F_Z(z) = P(Z \leq z) = P(X + Y \leq z) = \int_{-\infty}^{+\infty} \int_{-\infty}^{z-u} f(u, v) dv du$$

---

<sup>1</sup>Rappel : la densité est toujours égale à 1 !

Sa densité est

$$f_Z(z) = \int_{-\infty}^{+\infty} f(u, z-u) \, dv \, du$$

Si  $X$  et  $Y$  sont indépendantes, on a :

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(u) \cdot f_Y(z-u) \, du = \int_{-\infty}^{+\infty} f_X(z-v) \cdot f_Y(v) \, dv$$

### Quotient

Soit  $Z = X/Y$ . Sa distribution est

$$F_Z(z) = P(Z \leq z) = P(X/Y \leq z) = \int_0^{+\infty} \int_{-\infty}^{zx} f(x,y) \, dx \, dy + \int_{-\infty}^0 \int_{zx}^{-\infty} f(x,y) \, dx \, dy$$

Sa densité est

$$f_Z(z) = \int_{-\infty}^{+\infty} |x| f(x, xz) \, dx$$

Si  $X$  et  $Y$  sont indépendantes, on a :

$$f_Z(z) = \int_{-\infty}^{+\infty} |x| f_X(x) \cdot f_Y(xz) \, dx$$

### Cas général

Soient  $X$  et  $Y$  des variables aléatoires continues avec une fonction de distribution conjointe  $f_{X,Y}(x,y)$  et que  $X$  et  $Y$  soient envoyées sur  $U$  et  $V$  par transformation d'espace  $(x,y) \mapsto (u,v)$  donnée par  $u = h(x,y)$  et  $v = g(x,y)$ . La fonction de distribution conjointe de  $U$  et  $V$  au point  $(x_0, y_0)$  est :

$$f_{U,V}(h_0, g_0) = \frac{f_{X,Y}(x_0, y_0)}{|J(x_0, y_0)|}$$

où  $J(x_0, y_0)$  est la jacobienne de  $h$  et  $g$  au point  $(x_0, y_0)$  :

$$J(x_0, y_0) = \det \begin{bmatrix} \frac{\partial h}{\partial x}(x_0, y_0) & \frac{\partial h}{\partial y}(x_0, y_0) \\ \frac{\partial g}{\partial x}(x_0, y_0) & \frac{\partial g}{\partial y}(x_0, y_0) \end{bmatrix}$$

### Transformation de variables aléatoires

Soient  $X$  et  $Z$  des variables aléatoires, avec  $Z = g(X)$ . Soit  $f_X$  la densité de  $X$ . On a :

$$f_Z(z) = \begin{cases} \int f_X(g^{-1}(z)) \left| \frac{d}{dz} g^{-1}(z) \right| & \text{si } z = g(x) \text{ pour un certain } x \\ 0 & \text{si } z \neq g(x) \text{ pour tout } x \end{cases}$$



# Chapitre 5

## Espérance et Variance

### 5.1 Espérance

L'espérance (ou moyenne pondérée) d'une variable aléatoire est :

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

**On la note aussi  $\mu$ .** Si l'intégrale diverge il n'y a pas d'espérance.

*Remarque 5.1.* On a aussi :

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx$$

Si  $X_1 \dots X_i$  sont des variables aléatoires indépendantes :

$$E(X_1 \dots X_i) = E(X_1)E(X_2) \dots E(X_i)$$

Si  $X_1 \dots X_i$  sont des variables aléatoires conjointes

$$E(X_1 + \dots + X_i) = E(X_1) + E(X_2) + \dots + E(X_i)$$

Soient  $X$  et  $Y$  deux variables aléatoires avec  $Y = g(X)$  et  $f(x)$  la densité de  $x$ . L'espérance de  $Y$  est :

$$E(Y) = \int_{-\infty}^{+\infty} g(x)f(x)dx$$

L'espérance d'un grand nombre de variable aléatoires  $n$  identiquement distribuées (*loi des grands nombres*) est :

$$\frac{\sum_{i=1}^n X_i}{n} = \mu$$

#### 5.1.1 Quelques propriétés

Soit  $X$  une variable aléatoire et  $c \in \mathbb{R}$  une constante. On a :

- $E(c) = c$  ;
- $E(cX) = c \cdot E(X)$  ;
- $E(-X) = -E(X)$  ;

## 5.2 Variance

La *variance* d'une variable aléatoire avec espérance  $E(X)$  est :

$$\text{Var}(X) = E[(X - E(X))^2] = E(X^2) - E(X)^2$$

Elle est souvent notée  $\sigma^2$ . On appelle  $\sigma$  l'*écart type*, qui équivaut à  $\sqrt{\text{Var}(X)}$ . Si  $X$  est une variable aléatoire avec densité  $f(x)$ , sa variance est :

$$\text{Var}(X) = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx$$

### 5.2.1 Quelques propriétés

Soit  $X$  une variable aléatoire et  $c \in \mathbb{R}$  une constante. On a :

$$- \text{Var}(aX) = a^2 \text{Var}(X);$$

### 5.2.2 Moyenne carrée d'erreur

La valeur d'une v.a.  $X$  représentant la valeur d'un phénomène physique peut subir deux types d'erreur de mesure : une composante systématique appelée *biais*  $\beta$  et une composante aléatoire  $\epsilon$  avec  $\text{Var}(\epsilon) = \sigma^2$ . On a donc  $X = x_0 + \beta + \epsilon$ . L'écart entre la valeur de  $X$  et la valeur exacte recherchée  $x_0$  peut se mesurer avec la *moyenne carrée de l'erreur* :

$$\text{MCE} = E[(X - x_0)^2] = \beta^2 + \sigma^2 = \text{Var}(X) + (E(X) - x_0)^2$$

Si  $\beta = 0$  (non-biaisé), alors  $\text{MCE} = \text{Var}(\epsilon) = \sigma^2$ .

La table 1 (pag. 34) contient les valeurs de l'espérance et de la variance pour les principales distributions.

## 5.3 Covariance et corrélation

Pour mesurer le degré d'association de deux variables aléatoires conjointes  $X$  et  $Y$ , on peut en calculer la *covariance* et la *corrélation*.

La *covariance* de  $X$  et  $Y$  est :

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - E(X)E(Y)$$

Si  $X$  et  $Y$  sont indépendantes alors  $\text{Cov}(X, Y) = 0$ .

### 5.3.1 Propriétés

- $\text{Cov}(aX + b, cY + d) = ac \cdot \text{Cov}(X, Y)$
- $\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$

La *corrélation* entre deux variables est :

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}}$$

$-1 \leq \rho \leq 1$ , si elle existe.

## 5.4 Espérance conditionnelle

L'espérance conditionnelle de  $Y$  sachant  $X$  est :

$$E(Y|X = x) = \int y f_{Y|X}(y|x) dy$$

Si  $E(Y)$  existe la loi de l'espérance totale nous dit que :

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

## 5.5 Prediction

On observe une v.a.  $X$  et en se basant sur cette observation on essaye de *prédire* la valeur d'une deuxième v.a.  $Y$ . On note  $h(X)$  l'estimateur de  $Y$ .

- $Y$  cherche à estimer  $h(X)$ . La Moyenne carrée d'erreur ( $MCE$ ) entre ces deux valeurs est :

$$MCE = E[(Y - h(X))^2] = E[E[(Y - h(X))^2|X]]$$

La fonction qui minimise la  $MCE$  est  $h(X) = E(Y|X)$ . L'espérance conditionnelle de  $Y$  sachant  $X$  est donc le meilleur estimateur au sens de la moyenne carrée d'erreur.

- Si  $E(Y|X)$  n'est pas connue on cherche un *estimateur linéaire*  $h(x) = \alpha + \beta x$  pour que  $E[(Y - \alpha - \beta x)^2]$  soit minimal. Soit  $\hat{Y}$  le meilleur estimateur de  $Y$ . On a :

$$\frac{\hat{Y} - \mu_Y}{\sigma_Y} = \rho(X, Y) \left( \frac{X - \mu_X}{\sigma_X} \right)$$

On peut résoudre par rapport à  $\hat{Y}$  pour trouver la valeur du meilleur estimateur.

## 5.6 La fonction génératrice des moments

La fonction génératrice des moments d'une variable aléatoire  $X$  est :

$$M_X(t) = E(e^{tX})$$

Dans le cas continu :

$$M_X(t) = \int_{-\infty}^{+\infty} e^{tx} f(x) dx$$

- si  $M_X(t) = M_Y(t)$ ,  $X$  a la même loi de probabilité de  $Y$ .
- soit  $M_X(t)$  et  $Y = a + bX$ . La fonction génératrice des moments de  $Y$  est :

$$M_Y(t) = e^{at} M_X(bt)$$

- Soient  $X$  et  $Y$  deux variables aléatoire indépendantes, alors :

$$M_{X+Y}(t) = M_X(t) M_Y(t)$$

- Si  $X$  est la variable normale  $N(0, 1)$ ,  $M_X(t) = e^{t^2/2}$ .
- Si  $X$  est une variable Gamma  $\Gamma(\alpha, \lambda)$ ,  $M_X(t) = \left( \frac{\lambda}{\lambda - t} \right)^\alpha$
- Si  $X$  est une variable aléatoire de Poisson  $P(\lambda)$ ,  $M_X(t) = e^{\lambda(e^t - 1)}$

Le  $r^{\text{e}}$  moment d'une variable aléatoire  $X$  est  $E(X^r)$ , si elle existe. Si la fonction génératrice des moments existe, alors

$$E(X^r) = M_X^{(r)}(0)$$

# Chapitre 6

## Distributions dérivées de la distribution normale

On rappelle que si  $X \sim N(\mu, \sigma^2)$ ,  $\frac{X-\mu}{\sigma}$  suit une distribution normale standard.

- Soit  $X$  une distribution aléatoire normale standard. La distribution  $X^2$  est appelée distribution *chi-carré avec 1 degré de liberté*, notée  $\chi_1^2$ . Correspond à une distribution  $\Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$
- Soient  $U_1, U_2, \dots, U_n$  des variables aléatoires indépendantes  $\chi_1^2$ .  
 $U_1 + U_2 + \dots + U_n$  est une distribution *chi-carré avec  $n$  degrés de liberté*  $\chi_n^2$ , qui correspond à une distribution  $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$
- Soit  $Z \sim N(0, 1)$  et  $U \sim \chi_n^2$ .

On définit la distribution de Student  $t_n \sim \frac{Z}{\sqrt{U/n}}$ . Sa densité est :

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

- Soient  $U$  et  $V$  deux variables aléatoires indépendantes chi-carré.  
On définit la distribution de Fisher  $F_{m,n} \sim \frac{U/m}{V/n}$ . Sa densité est :

$$f(w) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{m/2} w^{m/2-1} \left(1 + \frac{m}{n}w\right)^{-(m+n)/2}$$

On observe que si  $X \sim t_n$ , alors  $X^2 \sim F_{1,n}$

# **Deuxième partie**

## **Statistique**

# Chapitre 7

## Échantillonnage

L'*échantillonnage* est utilisé pour obtenir des information sur une population en en examinant seulement une petite partie. Un échantillon doit être *représentatif* et *non biaisé*.

Soit  $x_1 \dots x_N$  une population de taille  $N$ . La *moyenne de la population* vaut :

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i$$

Le *total de la population*(ayant une certaine caractéristique) vaut :

$$\tau = \sum_{i=1}^N x_i = N\mu$$

La *variance de la population* :

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$$

### 7.1 Échantillon simple

Dans une population de taille  $N$  il y a  $\binom{N}{n}$  sous-ensembles de taille  $n$ . La *moyenne de l'échantillon* vaut :

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

Dans ce cas  $\bar{X}$  estime la moyenne de la population  $\mu$ .

$T = N\bar{X}$  estime le total de la population  $\tau$  ayant une certaine caractéristique.

**Un estimateur est *non-biaisé* si son espérance est égale à la quantité que l'on veut estimer :**  
 $E(\hat{\theta}) = \theta$ .  $\bar{X}$  et  $T$  sont non-biaisés.

La *variance de l'échantillon*(sans remise) vaut :

$$Var(\bar{X}) = \frac{\sigma^2}{n} \left( \frac{N-n}{N-1} \right)$$

Si  $n \ll N$ , ou dans le cas d'un échantillon avec remise,  $Var(\bar{X}) = \frac{\sigma^2}{n}$ . Dans ce cas on note  $\bar{X} \cong \mu \pm \frac{\sigma}{\sqrt{n}}$ .

La *variance de la la population* (ayant une certaine caractéristique) vaut :

$$Var(T) = N^2 \frac{\sigma^2}{n} \left( \frac{N-n}{N-1} \right)$$

On remarque que la variance de l'échantillon a la variance de la population  $\sigma^2$  comme paramètre. Si la population est très grande, cette valeur est inconnue, et doit être estimée<sup>1</sup>.

$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  est un estimateur biaisé pour  $\sigma^2$  (car  $E(\hat{\sigma}^2) \neq \sigma^2$ ).

Deux estimateurs non biaisés de  $\sigma^2$

$$s^2 = \hat{\sigma}^2 \frac{n(N-1)}{(n-1)N} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Estimateur non biaisé de  $Var(\bar{X})$

$$S_{\bar{X}}^2 = \frac{\hat{\sigma}^2}{n} \left( \frac{N-n}{N-1} \right) \left( \frac{n(N-1)}{(n-1)N} \right)$$

## 7.2 Échantillon stratifié

Utile dans le cas d'une population composée de plusieurs sous populations.

Soit une population de  $N$  individus composée de  $l$  sous populations de taille  $N_1 \dots N_l$ . On note  $x_{i,j}$  la valeur du  $j$ ème individu de la  $i$ ème sous population,  $W_i = \frac{N_i}{N}$  le poids de la sous population  $i$  et  $\mu_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{i,j}$  sa moyenne. La moyenne de la population est :

$$\mu = \sum_{i=1}^l W_i \mu_i$$

En considérant un échantillon de moyenne  $\bar{X}_i$  pour chaque sous population, on a pour  $\mu$  l'estimateur non-biaisé

$$\bar{X}_s = \sum_{i=1}^l W_i \bar{X}_i$$

La variance de l'échantillon choisi est :

$$Var(\bar{X}_i) = \sum_{i=1}^l \frac{W_i^2 \sigma_i^2}{n_i} \left( 1 - \frac{n_i - 1}{N_i - 1} \right)$$

$$Var(\bar{X}_i) \approx \sum_{i=1}^l \frac{W_i^2 \sigma_i^2}{n_i}, \text{ si } n_i \ll N_i \forall i$$

On s'intéresse à savoir de quelle dimension doit être l'échantillon de chaque population. Il y a deux possibles répartitions<sup>2</sup>.

– répartition optimale<sup>3</sup> : minimise la variance de  $\bar{X}_s$ . Pour un échantillon de taille<sup>4</sup>  $n = \sum_{i=1}^l n_i$  on a :

$$n_i = n \frac{W_i \sigma_i}{\sum_{k=1}^l W_k \sigma_k}$$

<sup>1</sup> notation : la valeur exacte est écrite en lettres grecques, l'estimateur en romaines !

<sup>2</sup> Si  $n$  n'as pas été fixé on peut écrire la répartition sous la forme  $\left( \frac{n_1}{n}, \dots, \frac{n_l}{n} \right)$

<sup>3</sup> La variance  $\sigma_i$  doit être connue !

<sup>4</sup>  $n$  est fixé !

– *répartition proportionnelle* : Dans ce cas, les  $n_i$  sont proportionnels aux  $W_i$  :

$$N_i = nW_i$$

On veut comparer les variances des différentes allocations ; soient  $Var(\bar{X})$  la variance de l'échantillon simple,  $Var(\bar{X}_{sp})$  celle de l'échantillon stratifié proportionnel et  $Var(\bar{X}_{so})$  celle de l'échantillon stratifié optimal.

$$Var(\bar{X}_{sp}) - Var(\bar{X}_{so}) = \frac{1}{n} \sum_{i=1}^l W_i (\sigma_i - \bar{\sigma})^2$$

$$\text{ou } \bar{\sigma} = \sum_{i=1}^l W_i \sigma_i$$

$$Var(\bar{X}) - Var(\bar{X}_{sp}) = \frac{1}{n} \sum_{i=1}^l W_i (\mu_i - \mu)^2$$



# Chapitre 8

## Estimation de paramètres

*Idee* : On a  $X_1 \dots X_i$  variables aléatoires indépendantes identiquement distribuées, on suppose qu'elles suivent une certaine distribution dont on ne connaît pas un ou plusieurs paramètres, que l'on veut estimer.

### 8.1 La méthode des moments

Soit  $X$  une variable aléatoire de même lois que  $X_i$ . Avec  $X_1 \dots X_n$  le  $k^e$  moment de l'échantillon.

$$\hat{\mu}_k = E(X^k) = \frac{1}{n} \sum_{i=1}^n X_i^k$$

$\hat{\mu}_k$  est un estimateur du  $k^e$  moment de  $X$ . De plus, **on estime  $\mu_1$  par  $\hat{\mu}_1 = \bar{X}$** .

*Remarque 8.1.* Rappel :

$$\sum_{i=1}^n X_i = n\bar{X} \quad \Leftrightarrow \quad \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}, \quad \sum_{i=1}^n (X_i - 1) = n(\bar{X} - 1) = n\bar{X} - n$$

Maintenant, pour estimer les  $n$  paramètres de la distribution, on pose :

$$\begin{cases} \mu_1 = E(X) = f_1(\theta_1, \dots, \theta_n), & \hat{\mu}_1 = \bar{X} \\ \mu_2 = E(X^2) = f_2(\theta_1, \dots, \theta_n) \\ \dots \\ \mu_n = E(X^n) = f_n(\theta_1, \dots, \theta_n) \end{cases}$$

On calcule les  $n$  espérances et on crée  $n$  equivalences du type  $\theta_i = f(\mu_1, \dots, \mu_n)$ . Ensuite on substitue aux  $\mu_i$  les estimateurs  $\hat{\mu}_i$ , on a ainsi des estimateurs pour les  $\theta_i$  !

### 8.2 La méthode du maximum de vraisemblance

Soient  $X_1 \dots X_n$  i.i.d. selon une loi de densité  $f_\theta(X_i)$ . On définit la *fonction de vraisemblance* :

$$V(\theta) = \prod_{i=1}^n f_\theta(X_i)$$

On a la fonction de vraisemblance logarithmique :

$$l(\theta) = \ln V(\theta) = \sum_{i=1}^n \ln f_\theta(X_i)$$

L'estimateur  $\hat{\theta}$  de  $\theta$  est celui qui maximise  $V(\theta)$ , et donc  $l(\theta)$ . Pour le trouver, il faut résoudre l'équation  $l'(\theta) = 0$ .

Le plus d'observation on réalise, le plus de précision on peut obtenir avec l'estimateur, en particulier :

$$|\hat{\theta} - \theta_0| \cong \frac{1}{\sqrt{n}}$$

### 8.3 Information et intervalles de confiance

Un intervalle de confiance est un ensemble de deux bornes tel que  $\mu$  se trouve avec une probabilité  $1 - \alpha$  à l'intérieur de ces deux bornes. Nous partons du fait que :

$$I_\alpha = [z_{\alpha/2}; z_{1-\alpha/2}] \text{ est un IC pour } y = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Car selon le théorème centrale limite  $y = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$  approche une loi normale  $N(0, 1)$

#### 8.3.1 Intervalles de confiance simples

But : trouver les bornes inférieur et supérieur  $b_i$  et  $b_s$  tel que :

$$P(b_i \leq \mu \leq b_s) = 1 - \alpha$$

$$\begin{aligned} P\left(z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{1-\alpha/2}\right) &\Rightarrow P\left(\frac{\sigma}{\sqrt{n}} z_{\alpha/2} \leq \bar{X} - \mu \leq \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}\right) \\ &\Rightarrow P\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \leq \mu \leq \bar{X} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right) \\ &\Rightarrow P\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \leq \mu \leq \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right) \end{aligned}$$

L'intervalle de confiance dans le cas d'une loi avec  $\sigma^2$  connu et  $\mu$  estimé grâce à la méthode des moments est :

$$P\left[\bar{X} - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}; \bar{X} + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}\right]$$

#### 8.3.2 Intervalles de confiance quand la variance $\sigma^2$ est inconnue

On remplace simplement  $\sigma$  par  $\hat{\sigma}$  et la loi normale par la loi de student. L'intervalle de confiance devient alors :

$$P\left[\bar{X} - \frac{\hat{\sigma}}{\sqrt{n}} t_{n-1}(1 - \alpha/2); \bar{X} + \frac{\hat{\sigma}}{\sqrt{n}} t_{n-1}(1 - \alpha/2)\right]$$

#### 8.3.3 Information

On définit l'information  $I(\theta_0)$  par :

$$I(\theta_0) = E_{\theta_0} \left[ \frac{\partial}{\partial \theta} \ln f(X, \theta_0) \right]^2 = Var_{\theta_0} \left[ \frac{\partial}{\partial \theta} \ln f(X, \theta_0) \right]$$

Si  $f$  est assez régulière, alors<sup>1</sup> :

$$I(\theta_0) = -E_{\theta_0} \left[ \frac{\partial^2}{\partial \theta^2} \ln f(X, \theta_0) \right]$$

L'objectif d'un estimateur  $T$  d'un paramètre  $\theta$  est minimiser la moyenne carrée d'erreur. Il y a une limite de précision, cette borne inférieure est donnée par l'*inégalité de Cramer-Rao* :

$$\text{Var}(T) \geq \frac{1}{nI(\theta)} = v_n$$

où  $v_n$  est la *variance asymptotique*.

### 8.3.4 Intervalles de confiance avec $\mu$ estimé par la méthode du maximum de vraisemblance

Soit  $f$  une fonction suffisamment régulière par rapport à  $X$  et  $\theta$ . Alors

$$\sqrt{nI(\theta_0)}(\hat{\theta}_n - \theta_0) \rightarrow N(0, 1)$$

où  $\theta_0$  est la vraie valeur de  $\theta$ . En particulier on a

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \sim N\left(0, \frac{1}{I(\theta)}\right)$$

Cela implique

$$P\left(\sqrt{nI(\theta_0)}|\hat{\theta}_n - \theta_0| \leq z_{\alpha/2}\right) \approx 1 - \alpha$$

ou de façon similaire

$$P\left(\hat{\theta}_n - \frac{z_{\alpha/2}}{\sqrt{nI(\theta_0)}} \leq \theta_0 \leq \hat{\theta}_n + \frac{z_{\alpha/2}}{\sqrt{nI(\theta_0)}}\right) \approx 1 - \alpha$$

$\theta_0$  et donc  $I(\theta_0)$  sont inconnus. En remplaçant  $I(\theta_0)$  par  $I(\hat{\theta}_n)$  on a l'intervalle de confiance de niveau  $1 - \alpha$  suivant<sup>2</sup> :

$$\left[ \hat{\theta}_n - \frac{z_{\alpha/2}}{\sqrt{nI(\hat{\theta}_n)}}, \hat{\theta}_n + \frac{z_{\alpha/2}}{\sqrt{nI(\hat{\theta}_n)}} \right]$$

Où  $\hat{\theta}_n$  est un estimateur de  $\theta$ ,  $I(\hat{\theta}_n)$  se calcule en substituant à  $\theta$  dans l'information on estimateur  $\hat{\theta}_n$ , et  $z_{\alpha/2}$  est la valeur de la distribution normale pour  $\alpha/2$ .

<sup>1</sup> Attention : on dérive  $\ln f$  par rapport à  $\theta$  mais on calcule l'espérance  $E$  par rapport à  $X$

<sup>2</sup>  $(1 - \alpha)$  est la probabilité de trouver le paramètre  $\theta_0$  dans l'intervalle

# Chapitre 9

## Tests d'hypothèses

Soit  $X_1 \dots X_n$  un échantillon suivant une distribution  $F_\theta$ , avec  $\theta$  un paramètre inconnu. Soit  $H_0$  une hypothèse sur  $\theta$  dite nulle, soit  $H_A$  une hypothèse alternative. Soit on accepte  $H_0$  et on affirme qu'un croit cette hypothèse compatible avec les données, soit on la refuse en faveur de  $H_A$ , en affirmant ainsi qu'on ne croit pas  $H_0$  juste.

- On effectue un *erreur de type I* si  $H_0$  est juste mais on le rejette.
- On effectue un *erreur de type II* si  $H_0$  est fausse et on l'accepte.

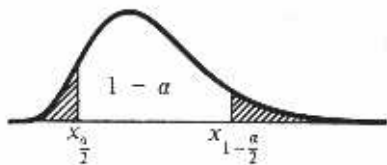
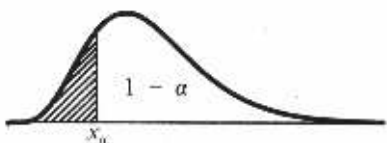
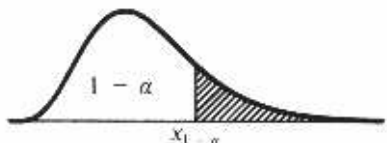
On a sous  $H_0$  une région d'acceptation A et une région de rejet R.

- Soit  $\alpha = P_{H_0}(\text{rejeter } H_0)$  le *niveau de signification*, ou *taille du test*.
- Soit  $1 - \beta = P_{H_A}(\text{rejeter } H_0)$  la *puissance du test*.

Chaque test essaye de minimiser la taille du test e maximiser la puissance, mais on ne peut pas obtenir les deux choses en même temps !

On définit la *p-valeur* la probabilité sous  $H_0$  que en exécutant une nouvelle mesure on obtienne un résultat égal ou plus extrême des données :  $p\text{-valeur} = P(T \geq T_{obs})$ . On rejette  $H_0$  si la *p-valeur* est inférieure à un limite  $\alpha$  donné.

**Types de tests.** On distingue trois types de tests :

Test	Probabilité valant $1 - \alpha$	Région critique (hachurée)
bilatéral	$P(x_{\alpha/2} < X < x_{1-\alpha/2})$	
unilatéral (à gauche)	$P(X > x_\alpha)$	
unilatéral (à droite)	$P(X < x_{1-\alpha})$	

## 9.1 Test optimal

Soient  $X_1 \dots X_n$  des variables aléatoires suivant une loi connue de paramètre inconnu. Soit  $H_0 : \theta = \theta_0$  et  $H_A : \theta = \theta_1 > \theta_0$ . Un test optimal est défini par le rapport des densités

$$\frac{f_0(\underline{X})}{f_A(\underline{X})}$$

La démarche à suivre ressemble ensuite à celle du rapport du maximum de vraisemblance.

## 9.2 Test du rapport du maximum de vraisemblance

À utiliser *si on ne connaît pas la statistique sous  $H_0$* .

Soit  $\Omega$  l'espace des paramètres possibles, et  $\omega_0$  un sous-espace, On veut tester  $H_0 : \theta \in \omega_0$  contre  $H_A : \theta \notin \omega_0$ . Soit  $V(\theta)$  la fonction de vraisemblance, la *fonction du rapport du maximum de vraisemblance* est :

$$\Lambda = \frac{\max_{\theta \in \omega_0} V(\theta)}{\max_{\theta \in \Omega} V(\theta)} = \frac{v(\theta_0)}{v(\theta_1)}$$

La région de réjet de  $H_0$  est celle qui inclut des valeurs de  $\bar{X}$  qui rendent petite  $\Lambda$ . Ce test est donc utile pour définir la région optimale de réjet, i.e. la région qui minimise  $\alpha$  tout en maximisant la puissance du test. Ex. : Si le rapport du maximum de vraisemblance est petit quand  $\bar{X}$  est grand, on aura une zone de rejet optimale, de la forme  $R_\alpha = (x_\alpha, \infty)$ . La valeur de  $x_\alpha$  se calcule avec la définition de  $\alpha$  :

$$\alpha := P_{H_0}(\text{rejeter } H_0) = P_{H_0}(\bar{X} > x_\alpha)$$

## 9.3 z-test pour un échantillon

Utilisé lorsqu'on a une grande quantité de données  $Y_1 \dots Y_n$  de **variance connue**<sup>1</sup>, suivant une loi normale  $N(\mu, \sigma^2)$ . Soit  $\bar{Y}$  la moyenne de l'échantillon et  $\mu_0$  la norme choisie. Soit  $z_{obs}$  un indicateur de la différence  $\bar{Y} - \mu_0$ . Le plus  $z_{obs}$  est grand le plus la probabilité que  $H_0$  soit fausse montent.  $z$  suit une distribution normale standard.

$$z_{obs} = \frac{\sqrt{n}(\bar{Y} - \mu_0)}{\sigma}$$

la p-valeur est la probabilité d'obtenir un résultat  $z$  plus extrême de  $z_{obs}$  :

- $p$ -valeur =  $1 - \Phi(z_{obs})$  pour un test unilatéral ;
- $p$ -valeur =  $2(1 - \Phi(z_{obs}))$  pour un test bilatéral ;

Pour rejeter  $H_0$ , ce valeur doit être plus petit du  $\alpha$  choisi.

On rejette  $H_A$  contre  $H_0(\mu = \mu_0)$

- $H_A(\mu \neq \mu_0)$  si  $|z_{obs}| \geq z_{\alpha/2}$
- $H_A(\mu > \mu_0)$  si  $z_{obs} \geq z_\alpha$
- $H_A(\mu < \mu_0)$  si  $z_{obs} \leq -z_\alpha$

<sup>1</sup>Si on connaît  $\bar{X}$  on connaît aussi la variance !

## 9.4 t-test pour un échantillon

Lorsque la **variance est inconnue**, et on a peu de valeurs (moins de  $\sim 120$ ), on utilise le t-test, ou test Student. On pose :

$$t_{obs} = \frac{\sqrt{n}(\bar{Y} - \mu_0)}{\sqrt{s_n^2}} \sim t_{(n-1)}$$

ou  $s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  est un estimateur de la variance. On veut calculer  $t_{(n-1)}(t_{obs})$  pour définir la p-valeur. Comme les tables de la loi de Student ne montrent que des valeurs critiques, on cherche dans la ligne  $t_{(n-1)}$  les valeurs parmi lesquels  $t_{obs}$  est inclu ; ça donne une idée de l'entité de la p-valeur, à comparer avec le  $\alpha$  choisi.

On rejette

- $H_A(\mu \neq \mu_0)$  si  $|t_{obs}| \geq t_{n-1}(\alpha/2)$
- $H_A(\mu > \mu_0)$  si  $t_{obs} \geq t_{n-1}(\alpha)$
- $H_A(\mu < \mu_0)$  si  $t_{obs} \leq -t_{n-1}(\alpha)$

## 9.5 z-test et t-test pour deux échantillons

Soient deux échantillons de **même variance**  $\sigma^2$  suivant une loi Normale. On considère la différence  $\bar{X} - \bar{Y}$ , qui suit une loi Normale  $N(0, \frac{\sigma^2}{n} + \frac{\sigma^2}{m})$ . Pour un z-test bilatéral, on rejette  $H_0$  si  $|\bar{X} - \bar{Y}|$  est trop grand :

$$|\bar{X} - \bar{Y}| \geq z(\alpha/2) \sqrt{\sigma^2 \left( \frac{1}{n} + \frac{1}{m} \right)}.$$

Si on ne connaît pas la variance, on l'estime et on l'utilise pour un test de Student.

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^m (Y_j - \bar{Y})^2}{n + m - 2}$$

Sous  $H_0$  on a donc :

$$\frac{(\bar{X} - \bar{Y})}{\hat{\sigma} \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$$

Pour un t-test, on rejette  $H_0$  si :

$$\frac{|\bar{X} - \bar{Y}|}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \geq t_{n+m-2}(\alpha/2), \text{ pour un test } \mu_x \neq \mu_y$$

$$\frac{\bar{X} - \bar{Y}}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \geq t_{n+m-2}(\alpha), \text{ pour un test } \mu_x > \mu_y$$

$$\frac{\bar{X} - \bar{Y}}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \leq -t_{n+m-2}(\alpha), \text{ pour un test } \mu_x < \mu_y$$

## 9.6 z-test et t-test pour deux échantillons appariés

Si on a deux échantillons  $X_1 \dots X_n$  et  $Y_1 \dots Y_n$ , on crée  $n$  des couples  $(X_i, Y_i)$  où les deux composantes du couple sont semblables. Soit  $Z_i = X_i - Y_i$ ; on pose  $H_0 : \theta = 0$  contre  $H_A : \theta \neq 0$ , ou  $\theta = E(Z_i)$ .

Sachant que  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (Z_i - \bar{Z})^2$ , on rejette si :

$$\left| \frac{\sqrt{n}\bar{Z}}{s} \right| \geq t_{n-1}(\alpha/2)$$

On utilise dans ce cas un t-test pour des échantillons appariés. Si on connaît la variance, une autre possibilité est le z-test apparié :

$$\left| \frac{\sqrt{n}\bar{Z}}{s} \right| \geq z_{\alpha/2}$$

## 9.7 Tests khi-deux

Le test Khi-deux de Pearson est utilisé pour *tester l'adéquation d'une distribution théorique à des données*.

On effectue  $r$  fois une expérience aléatoire, avec  $s$  paramètres. Soient  $E_i$  l'espérance d'observations de type  $i$  sous  $H_0$  et  $O_i$  le nombre d'observations de type  $i$ . On rejette  $H_0$  si :

$$\sum_1^r \frac{(O_i - E_i)^2}{E_i} \geq q_{\chi^2_{r-1-s}}(1 - \alpha)$$

### 9.7.1 Test d'indépendance

Soit un échantillon de taille  $n$  dans lequel pour chaque individu on mesure deux variables qui ont respectivement  $I$  et  $J$  valeurs possibles, on se demande si ces deux variables sont indépendantes. Soient  $E_{i,j} = np_{i,j} = np_i q_j$  (sous  $H_0$ ) et  $O_{i,j} = n_{i,j}$ . On crée le suivant tableau empirique :

	1	2	...	J	
1	$O_{1,1}$	$O_{1,2}$	$\dots$	$O_{1,J}$	$O_{1.}$
2	$O_{2,1}$	$O_{2,2}$	$\dots$	$O_{2,J}$	$O_{2.}$
$\vdots$			$O_{i,j}$		$\vdots$
I	$O_{I,1}$	$O_{I,2}$	$\dots$	$O_{I,J}$	$O_{I.}$
	$O_{.1}$	$O_{.2}$	$\dots$	$O_{.J}$	$n$

Soit  $\hat{p}_i = \frac{O_i}{n}$  et  $\hat{q}_j = \frac{O_j}{n}$ , on a donc  $E_{ij} = \hat{p}_i \hat{q}_j n$ .

On teste  $H_0 : p_{i,j} = p_i q_j$  (soit les deux variables sont dépendantes) contre  $H_A$  : les variables sont dépendantes. On compare le tableau empirique avec un tableau de même dimension contenant les espérances théoriques. Chaque case de ce deuxième tableau se calcule ainsi :  $E_{i,j} = np_i q_j$ .

On rejette  $H_0$  si

$$\sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \geq q_{\chi^2_{(I-1)(J-1)}}(1 - \alpha)$$

Ce test est valable si tout  $E_{i,j} > 5$  et si  $n > 10$ .

## 9.8 Tests non paramétriques (Test de Wilcoxon)

### 9.8.1 Échantillons non appariés

Si on n'a **pas d'informations précises** sur la loi, et donc **son paramètre**, suivie par l'échantillon, on utilise des *tests non paramétriques*. Soient  $X_1 \dots X_n$  et  $Y_1 \dots Y_m$  deux échantillon à comparer.

Le *test Mann-Whitney-Wilcoxon* utilise le rang : on trie la suite  $(X_1, X_2, \dots, X_n, Y_1, \dots, Y_m)$  en ordre croissant.  $R_i$  correspond à la position du  $i^{\text{ème}}$   $X$  dans la suite triée<sup>2</sup>.

Si  $n, m > 10$  on a :

$$W = \sum_{i=1}^n R_i \sim N\left(\frac{n(n+m+1)}{2}, \frac{nm(n+m+1)}{12}\right)$$

Soit  $H_0$  : les échantillons suivent la même loi. On rejette  $H_0$  si :

$$\frac{\left| \left( \sum_{i=1}^n R_i \right) - \frac{n(n+m+1)}{2} \right|}{\sqrt{\frac{nm(n+m+1)}{12}}} \geq z(\alpha/2), \text{ pour un test } \mu_x \neq \mu_y$$

$$\frac{\left( \sum_{i=1}^n R_i \right) - \frac{n(n+m+1)}{2}}{\sqrt{\frac{nm(n+m+1)}{12}}} \geq z(\alpha), \text{ pour un test } \mu_x > \mu_y$$

$$\frac{\left( \sum_{i=1}^n R_i \right) - \frac{n(n+m+1)}{2}}{\sqrt{\frac{nm(n+m+1)}{12}}} \leq -z(\alpha), \text{ pour un test } \mu_x < \mu_y$$

### 9.8.2 Échantillons appariés

Il existe un *Test Wilcoxon pour deux échantillons appariés*<sup>3</sup>. On divise les deux échantillons  $X$  et  $Y$  de même taille en  $n$  couples, dont on calcule la différence  $D_i = X_i - Y_i$ . On range les  $D_i$  en ordre croissant, et on pose :

$$W^+ = \sum_{i=1}^n R_i I_{(X_i > Y_i)}$$

ça signifie qu'on somme seulement les positions des couples dans lesquelles  $X_i > Y_i$ . Pour  $n > 10$ , on a :

$$W^+ \sim N\left(\frac{n(n+1)}{4}, \frac{n(n+1)(2n+1)}{24}\right)$$

<sup>2</sup>Si on a par exemple la suite triée suivante  $(X_2, Y_3, Y_1, X_1, X_3, Y_2)$  on a :  $R_1 = 1, R_2 = 4$  et  $R_3 = 5$ .

<sup>3</sup>Il n'y a aucun lien entre  $X_i$  et  $Y_i$



## **Troisième partie**

### **Tables**

$x$	$\Phi(x)$	$x$	$\Phi(x)$	$x$	$\Phi(x)$	$x$	$\Phi(x)$	$x$	$\Phi(x)$
0,00	0,500	0,60	0,726	1,20	0,885	1,80	0,964	2,40	0,992
0,02	0,508	0,62	0,732	1,22	0,889	1,82	0,966	2,42	0,992
0,04	0,516	0,64	0,739	1,24	0,893	1,84	0,967	2,44	0,993
0,06	0,524	0,66	0,745	1,26	0,896	1,86	0,969	2,46	0,993
0,08	0,532	0,68	0,752	1,28	0,900	1,88	0,970	2,48	0,993
0,10	0,540	0,70	0,758	1,30	0,903	1,90	0,971	2,50	0,994
0,12	0,548	0,72	0,764	1,32	0,907	1,92	0,973	2,52	0,994
0,14	0,556	0,74	0,770	1,34	0,910	1,94	0,974	2,54	0,994
0,16	0,564	0,76	0,776	1,36	0,913	1,96	0,975	2,56	0,995
0,18	0,571	0,78	0,782	1,38	0,916	1,98	0,976	2,58	0,995
0,20	0,579	0,80	0,788	1,40	0,919	2,00	0,977	2,60	0,995
0,22	0,587	0,82	0,794	1,42	0,922	2,02	0,978	2,62	0,996
0,24	0,595	0,84	0,800	1,44	0,925	2,04	0,979	2,64	0,996
0,26	0,603	0,86	0,805	1,46	0,928	2,06	0,980	2,66	0,996
0,28	0,610	0,88	0,811	1,48	0,931	2,08	0,981	2,68	0,996
0,30	0,618	0,90	0,816	1,50	0,933	2,10	0,982	2,70	0,997
0,32	0,626	0,92	0,821	1,52	0,936	2,12	0,983	2,72	0,997
0,34	0,633	0,94	0,826	1,54	0,938	2,14	0,984	2,74	0,997
0,36	0,641	0,96	0,831	1,56	0,941	2,16	0,985	2,76	0,997
0,38	0,648	0,98	0,836	1,58	0,943	2,18	0,985	2,78	0,997
0,40	0,655	1,00	0,841	1,60	0,945	2,20	0,986	2,80	0,997
0,42	0,663	1,02	0,846	1,62	0,947	2,22	0,987	2,82	0,998
0,44	0,670	1,04	0,851	1,64	0,949	2,24	0,987	2,84	0,998
0,46	0,677	1,06	0,855	1,66	0,952	2,26	0,988	2,86	0,998
0,48	0,684	1,08	0,860	1,68	0,954	2,28	0,989	2,88	0,998
0,50	0,691	1,10	0,864	1,60	0,955	2,30	0,989	2,90	0,998
0,52	0,698	1,12	0,869	1,72	0,957	2,32	0,990	2,92	0,998
0,54	0,705	1,14	0,873	1,74	0,959	2,34	0,990	2,94	0,998
0,56	0,712	1,16	0,877	1,76	0,961	2,36	0,991	2,96	0,998
0,58	0,719	1,18	0,881	1,78	0,962	2,38	0,991	2,98	0,999

TAB. 1 – Distribution d'une variable normale centrée réduite  $N(0, 1)$

$\nu$	$q\chi^2_\nu(1\%)$	$q\chi^2_\nu(2.5\%)$	$q\chi^2_\nu(5\%)$	$q\chi^2_\nu(95\%)$	$q\chi^2_\nu(97.59\%)$	$q\chi^2_\nu(99\%)$
1	0,031571	0,039821	0,003932	3,841	5,024	6,635
2	0,02010	0,05064	0,1026	5,991	7,378	9,210
3	0,1148	0,2158	0,3518	7,815	9,348	11,34
4	0,2971	0,4844	0,7107	9,488	11,14	13,28
5	0,5543	0,8312	1,145	11,07	12,83	15,09
6	0,8721	1,237	1,635	12,59	14,45	16,81
7	1,239	1,690	2,167	14,07	16,01	18,48
8	1,646	2,180	2,733	15,51	17,53	20,09
9	2,088	2,700	3,325	16,92	19,02	21,67
10	2,558	3,247	3,940	18,31	20,48	23,21
11	3,053	3,816	4,575	19,68	21,92	24,72
12	3,571	4,404	5,226	21,03	23,34	26,22
13	4,107	5,009	5,892	22,36	24,74	27,69
14	4,660	5,629	6,571	23,68	26,12	29,14
15	5,229	6,262	7,261	25,00	27,49	30,58
16	5,812	6,908	7,962	26,30	28,85	32,00
17	6,408	7,564	8,672	27,59	30,19	33,41
18	7,015	8,231	9,390	28,87	31,53	34,81
19	7,633	8,907	10,12	30,14	32,85	36,19
20	8,260	9,591	10,85	31,41	34,17	37,57
21	8,897	10,28	11,59	32,67	35,48	38,93
22	9,542	10,98	12,34	33,92	36,78	40,29
23	10,20	11,69	13,09	35,17	38,08	41,64
24	10,86	12,40	13,85	36,42	39,36	42,98
25	11,52	13,12	14,61	37,65	40,65	44,31
26	12,20	13,84	15,38	38,89	41,92	45,64
27	12,88	14,57	16,15	40,11	43,19	46,96
28	13,56	15,31	16,93	41,34	44,46	48,28
29	14,26	16,05	17,71	42,56	45,72	49,59
30	14,95	16,79	18,49	43,77	46,98	50,89
32	16,36	18,29	20,07	46,19	49,48	53,49
34	17,79	19,81	21,66	48,60	51,97	56,06
36	19,23	21,34	23,27	51,00	54,44	58,62
38	20,69	22,88	24,88	53,38	56,90	61,16
40	22,16	24,43	26,51	55,76	59,34	63,69
50	29,71	32,36	34,76	67,50	71,42	76,15
60	37,48	40,48	43,19	79,08	83,30	88,38
70	45,44	48,76	51,74	90,53	95,02	100,4
80	53,54	57,15	60,39	101,9	106,6	112,3
90	61,75	65,65	69,13	113,1	118,1	124,1
100	70,06	74,22	77,93	124,3	129,6	135,8

TAB. 2 – Table des quantiles de la loi  $\chi^2_\nu$

Loi	Notation	Espérance	Variance	Fréquences $f_X(k)$
Binomiale	$Bin(n, p)$	$np$	$np(1 - p)$	$\binom{n}{k} p^k (1 - p)^{(n-k)}$
Géométrique	$G(p)$	$\frac{1}{p}$	$\frac{(1 - p)}{p^2}$	$p(1 - p)^{k-1}$
Binomiale négative	$BN(n, p)$	$\frac{n}{p} - n$	$\frac{n(1 - p)}{p^2}$	
Poisson	$P(\lambda)$	$\lambda$	$\lambda$	$\frac{\lambda^k}{k!} e^{-\lambda}$
Hypergéométrique	$HG(n, N, B)$	$n \frac{B}{N}$	$n \frac{B}{N} \left(1 - \frac{B}{N}\right) \left(1 - \frac{n-1}{N-1}\right)$	$\frac{\binom{B}{k} \binom{N-B}{n-k}}{\binom{N}{n}}$
Loi	Notation	Espérance	Variance	Densité $f_X(x)$
Normale	$N(\mu, \sigma^2)$	$\mu$	$\sigma^2$	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
Uniforme	$U(a, b)$	$\frac{a+b}{2}$	$\frac{(a-b)^2}{12}$	$\frac{1}{b-a}$
Exponentielle	$\epsilon(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\lambda e^{-\lambda x}$
Gamma	$\Gamma(\alpha, \lambda)$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$	$\frac{\lambda}{\Gamma(\alpha)} (\lambda x)^{\alpha-1} e^{-\lambda x}$

TAB. 3 – Propriétés des lois discrètes et continues les plus souvent utilisés

$Bin(n, p) \sim N(np, np(1 - p))$
$Bin(n, p) \sim P(np)$ si $n$ grand et $p$ petit
$P(\lambda) \sim N(\lambda, \lambda)$ pour $\lambda \gg 1$

TAB. 4 – Approximation de quelque loi

# Index

- Analyse combinatoire, 6
- Arrangements, 6
- Bayes, formule de, 8
- Bernoulli, loi de -, 10
- Binomiale
  - loi -, 10, 35
  - loi - négative, 10, 35
- Chi-carré, 19
- Combinaisons, 6
- Comptage, 7
- Conjointe
  - loi de probabilité, 13
- Corrélation, 17
- Covariance, 17
- Densité, 11, 35
  - conjointe, 14
- Densité (fonction de), 9
- Densité conditionnelle, 14
- Distribution
  - conditionnelle, 14
  - conjointe, 13
  - marginale, 13, 14
- Fonction de -, 9–11
- Distributions dérivées de la loi normale, 19
- Écart-type, 17
- Échantillonnage, 21
- Espérance, 10, 11, 16, 35
- Espérance conditionnelle, 18
- Exponentielle (loi), 12, 35
- Fisher, 19
- Fréquence conjointe, 13
- Fréquences, 35
- $\Gamma$ , voir Gamma
- Gamma (loi), 12, 35
- Gaussienne (loi), 12
- Grands nombres, loi des -, 16
- Géométrique (loi), 10, 35
- Hypergéométrique (loi), 10, 35
- Indépendance, 8
- Information, 25
- Khi-deux, test -, 30
- Maximum de vraisemblance, 24
- MCE, voir Moyenne carrée d'erreur
- Moments, 18, 24
- Moyenne, 21
  - carrée d'erreur, 17, 18
- $N(\mu, \sigma^2)$  (loi gaussienne ou normale), 12
- Normale
  - loi -, 12, 35
  - Table des distributions, 33
- $\Omega$ , 7
- p-valeur, 27
- Permutations, 6
- Poisson (loi), 11, 35
- Prediction, 18
- Probabilité conditionnelle, 8
- Répartition (fonction de), voir Distribution
- t-test, 29, 30
- Tests d'hypothèses, 27
- Uniforme (loi), 11, 35
- Variables aléatoires, 9
- Variance, 17, 21, 35
- Wilcoxon, test de -, 31
- z-test, 28–30