

# 1 Statistiques Descriptives

## 1.1 Variable Discrète

**Définition 1** Soit  $x$  une variable discrète, à valeurs dans  $V_x = \{\xi_1, \dots, \xi_K\}$  avec  $\xi_1 < \dots < \xi_K$ . Une distribution de  $n$  observations de  $x$  peut être représentée sous forme d'un tableau de fréquences où figure pour chaque modalité de  $\xi_K$  de  $x$ , le nombre  $n_k$  d'observations ayant la valeur  $\xi_K$ . La fréquence relative est  $f_k = \frac{n_k}{n}$ . La fréquence cumulée est  $F_k = \sum_{j=1}^k f_j$

## 1.2 Variable Continue

**Définition 2** Si  $x$  est continue, on partitionne le domaine de définition de  $x$  en  $K$  classes. On prendra la règle de Sturges pour calculer  $K$ .  $K = 1 + \frac{10}{3} \log_{10} n$ .

Pour une variable discrète, on va utiliser un diagramme en baton, alors que pour une variable continue, on va utiliser un histogramme, où l'aire de chaque rectangle est proportionnel à l'effectif.

## 1.3 Fonction de répartition empirique

**Définition 3** La fonction de répartition empirique est :

$$\hat{F} : \mathbb{R} \rightarrow [0, 1]$$

$$x \rightarrow \frac{1}{n} \#\{i : x_i \leq x\} = \frac{1}{n} \sum_{i=1}^n 1_{[x_i, +\infty[}(x)$$

## 1.4 Indicateur de tendance centrale

**Définition 4** Moyenne empirique :  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ . Dans le cas des classes, on a :  $\bar{x} = \frac{1}{n} \sum_{k=1}^K n_k c_k$

**Proposition 1** La somme des écarts à la moyenne empirique est nulle :  $\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - n\bar{x}$  La moyenne empirique est la valeur la plus proche au sens de la somme des carrés des écarts.

**Définition 5** Fractile d'ordre  $\alpha$

$$f_\alpha = (1 - \gamma)x_{(j)} + \gamma x_{(j+1)}$$

avec  $j = \lfloor \alpha(n+1) \rfloor$  et  $\gamma = \alpha(n+1) - j$

## 1.5 Indicateur de dispersion

**Définition 6** La variance empirique :  $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$

La variance empirique corrigée :  $s^{*2} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{n}{n-1} s^2$

## 1.6 Boite à Moustaches

**Définition 7** Il s'agit d'un graphique formé d'une boite délimitée par le premier et le troisième quartile, où figure aussi la médiane ( $f_{0.5}$ ) Des segments de droites s'étendent de part et d'autre de la boite jusqu'au point le plus extrême à une distance inférieure à 1.5H. Les autres points sont représentés individuellement. H est la hauteur de la boite.

# 2 Rappel de Probabilité

## 2.1 Probabilité

**Proposition 2** Propriétés des probabilités :

- $P(\emptyset) = 0$ ;
- $P(\bar{A}) = 1 - P(A)$ ;
- $A \subset B \Rightarrow P(A) \leq P(B)$ ;
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ ;
- $P(\cup A_i) \leq \sum P(A_i)$ ;

**Proposition 3** A et B sont indépendants si  $P(A \cap B) = P(A)P(B)$ . n événements  $A_1, \dots, A_n$  sont indépendants si pour toute partie I de  $[1, n]$  :  $P(\cap_{i \in I} A_i) = \prod_{i \in I} P(A_i)$ .

## 2.2 Variables Aléatoires

**Proposition 4** Propriétés de la densité

- $\forall x : f(x) \geq 0$
- $\int_{\mathbb{R}} f(x) dx = 1$
- $P(x < X < x + \Delta x) = f(x)\Delta x + o(\Delta x)$

## 2.3 Fonction de répartition

**Définition 8**  $F : \mathbb{R} \rightarrow [0, 1]$  définie par  $\forall x \in \mathbb{R} : F(x) = P(X \leq x)$ .

**Proposition 5** Propriété de la fonction de répartition

- F est croissante et continue à droite.
- $\lim_{x \rightarrow -\infty} F(x) = 0$
- $\lim_{x \rightarrow +\infty} F(x) = 1$
- $P(a < X \leq b) = F(b) - F(a)$
- Si X est continue, alors F est dérivable et on a  $F(x) = \int_{-\infty}^x f(t) dt$

## 2.4 Espérance mathématique

**Proposition 6** Propriétés de l'espérance

- Si  $X \geq 0$ , alors  $E(X) \geq 0$
- $E(a) = a$
- $E(aX + bY) = aE(X) + bE(Y)$

## 2.5 Variance

**Proposition 7** Propriétés de la variance

- $E((X - a)^2) = Var(X) + (E(X) - a)^2$
- $Var(X) = E(X^2) - (E(X))^2$
- $Var(aX + b) = a^2 Var(X)$
- $P(|X - E(X)| \geq k\sigma) \leq \frac{1}{k^2}$

## 2.6 Autres moments

**Définition 9**

- Moment centré d'ordre  $k$  :  $\mu_k = E([X - E(X)]^k)$
- Moment non centré d'ordre  $k$  :  $m_k = E(X^k)$

## 2.7 Convergence Stochastique

**Définition 10** La suite  $(X_n)$  converge en probabilité vers  $a$  si  $\forall \varepsilon > 0$  et  $\eta > 0$ ,  $\exists n_0$  tel que  $\forall n > n_0$

$$P(|X_n - a| > \varepsilon) < \eta$$

On note  $(X_n) \xrightarrow{P} a$ . La convergence en probabilité de  $X_n$  vers  $a$  revient à étudier la convergence de  $X_n - a$  vers 0. Condition : Si  $E(X_n) \rightarrow a$  et  $Var(X_n) \rightarrow 0$ , alors  $X_n \xrightarrow{P} a$ .

**Définition 11** La suite  $(X_n)$  converge presque sûrement vers  $a$  si  $\forall \varepsilon$  et  $\eta$ ,  $\exists n_0$  tel que  $n > n_0$  entraîne :

$$P((|X_n - a| > \varepsilon) \cup \dots \cup (|X_{n+i} - a| > \varepsilon) \cup \dots) < \eta$$

On note

- $X_n \xrightarrow{p.s.} a$ ;
- $X_n \xrightarrow{p.s.} X \Rightarrow (X_n - X) \xrightarrow{p.s.} 0$ .

Condition :  $X_n \xrightarrow{p.s.} a \Rightarrow \max_i |X_{n+i} - a| \xrightarrow{P} 0$

**Définition 12** La suite  $(X_n)$  converge dans  $L^p$  vers  $X$  si  $\lim_{n \rightarrow +\infty} E(|X_n - X|^p) = 0$

**Définition 13** La suite  $(X_n)$  converge en loi vers  $X$  de fonction de répartition  $F$  si en tout point de continuité de  $F$  la suite  $(F_n)$  des fonctions de répartition de  $(X_n)$  converge vers  $F$ . On note  $(X_n) \xrightarrow{L} X$

**Proposition 8** On a  $(X_n) \xrightarrow{L} X$  et  $(Y_n) \xrightarrow{L} C$ .

- Cvg p.s.  $\Rightarrow$  Cvg en P  $\Rightarrow$  Cvg en Loi
- $X_n + Y_n \xrightarrow{L} X + C$
- $X_n Y_n \xrightarrow{L} CX$
- $\frac{X_n}{Y_n} \xrightarrow{L} \frac{X}{C}$  Si  $C \neq 0$

**Théorème 1 (Loi des grands nombres)** Si  $(X_n)$  est une suite de v.a. indépendantes de même loi et d'espérance  $\mu$ , alors  $\frac{X_1 + \dots + X_n}{n}$  converge p.s. vers  $\mu$ .

**Théorème 2** Théorème central limite :  $(X_n)$  suite de v.a. indépendantes de même loi, d'espérance  $\mu$  et de variance  $\sigma^2$  finie, alors

$$\frac{\frac{1}{n} \sum_{i=1}^n X_i - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{L} \mathcal{N}(0, 1)$$

quand  $n \rightarrow \infty$

Pour la loi Binomiale, on a :  $P(X = x) \simeq \Phi\left(\frac{x - np + 0.5}{\sqrt{np(1-p)}}\right)$

**Proposition 9** Autres approximations :

- Si  $X_\lambda \sim P(\lambda)$  alors  $\frac{X_\lambda - \lambda}{\sqrt{\lambda}} \xrightarrow{L} \mathcal{N}(0, 1)$  quand  $\lambda \rightarrow \infty$
- Si  $X_n \sim \chi_n^2$  alors  $\frac{X_n - n}{\sqrt{2n}} \xrightarrow{L} \mathcal{N}(0, 1)$  quand  $n \rightarrow \infty$

## 3 Echantillonnage

### 3.1 Notation

**Définition 14** On appellera échantillon le  $n$ -uplet de variables aléatoires  $(X_1, \dots, X_n)$ . Une réalisation de cet échantillon est alors le  $n$ -uplet  $(x_1, \dots, x_n)$ .

### 3.2 Statistique $\bar{X}$

**Définition 15** La statistique  $\bar{X}$  ou moyenne empirique est  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ . Si  $X$  a une espérance  $\mu$ , alors  $E(\bar{X}) = \mu$ . Si  $X$  a une variance  $\sigma^2$ , alors  $Var(\bar{X}) = \frac{\sigma^2}{n}$ . Si  $X$  a une espérance  $\mu$ , alors  $\bar{X} \xrightarrow{p.s.} \mu$ .

### 3.3 Statistiques $S^2$ et $S^{*2}$

**Définition 16** La statistique  $S^2$  ou variance empirique est

- $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ .  $E(S^2) = \frac{n-1}{n} \sigma^2 = E[X_i^2] - E[\bar{X}^2]$ ;
- $S^{*2} = \frac{n}{n-1} S^2$ ;
- $E[S^{*2}] = \sigma^2$ .

$S^{*2}$  est un estimateur sans biais de  $\sigma^2$ .

**Proposition 10** Si  $X$  a une variance  $\sigma^2$ , alors  $S^2 \xrightarrow{p.s.} \sigma^2$ .

### 3.4 Généralisation : moments empiriques

**Définition 17**

- Moment empirique centré d'ordre  $k$  :

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

- Moment empirique non centré d'ordre  $k$  :

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

Les moments empiriques d'un échantillon iid  $(X_1, \dots, X_n)$  convergent p.s. vers les moments théoriques correspondants de  $X$ .

### 3.5 Fonction de répartition empirique

**Définition 18** Fonction de répartition empirique :  $\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n 1_{(X_i \leq x)}$ . On a  $\forall x \in \mathbb{R} \hat{F}(x) \xrightarrow{ps} F(x)$ . De plus  $D_n = \sup_x |\hat{F}(x) - F(x)|$  est une v.a. qui ne dépend plus de  $X$ .

### 3.6 Echantillon gaussiens

On suppose dans ce paragraphe que  $X \sim \mathcal{N}(\mu, \sigma^2)$

**Définition 19**

- $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$
- $\frac{nS^2}{\sigma^2} \sim \chi_{n-1}^2$ .  $\bar{X}$  et  $S^2$  sont indépendantes.

### 3.7 loi jointe et vraisemblance

**Définition 20** Soit  $x_1, \dots, x_n$  une réalisation de  $X_1, \dots, X_n$  d'une v.a.  $X$ .  $X$  suit une probabilité dépendant d'un paramètre réel  $\theta$ .

- $X$  est continue :  $L(\theta; x_1, \dots, x_n) = f(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$
- $X$  est discrète :  $L(\theta; x_1, \dots, x_n) = P(X_1 = x_1, \dots, X_n = x_n; \theta) = \prod_{i=1}^n P(X_i = x_i; \theta)$
- L'application  $\theta \rightarrow L(\theta; x_1, \dots, x_n)$  est appelée vraisemblance de  $\theta$ . C'est une fonction de vraisemblance.
- $x_1, \dots, x_n \rightarrow L(\theta; x_1, \dots, x_n)$  est une probabilité ou une densité.

## 4 Estimation

### 4.1 formalisation et notation

**Définition 21** On appelle estimateur du paramètre de  $\theta$  toute fonction de l'échantillon  $X_1, \dots, X_n$ .

**Définition 22** Si  $T = \varphi(X_1, \dots, X_n)$  est un estimateur,  $t = \varphi(x_1, \dots, x_n)$ , réalisation de cet estimateur, est appelée estimation.

## 4.2 Estimateur convergent

**Définition 23** Un estimateur  $T$  du paramètre  $\theta$  est convergent si  $T$  converge en probabilité vers  $\theta$  quand  $n \rightarrow \infty$ , c-à-d si  $\forall \epsilon > 0, \lim_{n \rightarrow \infty} P(|T - \theta| > \epsilon) = 0$ .

## 4.3 Estimateur sans biais

**Définition 24**

- Un estimateur  $T$  est dit sans biais si quelle que soit la taille  $n$  de l'échantillon, et  $\theta$  la valeur du paramètre,  $E(T) = \theta$ .
- La quantité  $E(T) - \theta$  est appelée biais de l'estimateur  $T$ .
- Un estimateur est dit asymptotiquement sans biais si quel que soit  $\theta$ ,  $\lim_{n \rightarrow \infty} E(T) = \theta$ .
- Si  $T$  est un estimateur sans biais ou asymptotiquement sans biais et si  $\lim_{n \rightarrow \infty} Var(T) = 0$ , alors  $T$  est convergent.

## 4.4 Précision

**Définition 25** La précision d'un estimateur  $T$  est mesurée par  $R(T, \theta) = E[(T - \theta)^2] = Var(T) + (E(T) - \theta)^2$ .  $T_1$  est plus précis que  $T_2$  si  $R(T_1, \theta) < R(T_2, \theta)$ .

## 4.5 Exhaustivité

**Définition 26** Une statistique  $T$  est exhaustive si et seulement si il existe 2 fonctions  $g$  et  $h$  telles que :

- $L(\theta; x_1, \dots, x_n) = g(t, \theta)h(x_1, \dots, x_n)$ ;
- ou encore  $L(\theta; X_1, \dots, X_n) = g(T, \theta)h(X_1, \dots, X_n)$ .

## 4.6 Estimateur sans biais et de variance minimale

**Théorème 3**

- S'il existe un estimateur  $T$  de  $\theta$  sans biais de variance minimale, alors il est unique presque sûrement.
- S'il existe une statistique exhaustive  $U$  pour le paramètre de  $\theta$ , alors l'estimateur  $T$  de  $\theta$  sans biais de variance minimale ne dépend que de  $U$ .

## 4.7 Inégalité de Fréchet–Darmois–Cramer–Rao et estimateur efficace

**Théorème 4** Si les conditions suivantes de Cramer-Rao :

- Le support de  $X$  (ensemble pour lequel  $f(x, \theta) > 0$ ) est indépendant de  $\theta$
- $\frac{\partial L}{\partial \theta}(\theta; x_1, \dots, x_n)$  existe et est continue par rapport à  $\theta$
- $I_n(\theta) = E[(\frac{\partial \ln L}{\partial \theta}(\theta; X_1, \dots, X_n))^2]$  finie
- $\frac{\partial L}{\partial \theta}$  et  $\hat{u} \frac{\partial L}{\partial \theta}$  intégrables par rapport à l'échantillon

sont vérifiées, alors  $Var(\hat{u}) \geq \frac{(u'(\theta))^2}{I_n(\theta)}$  où  $u$  est une fonction quelconque. et l'on a  $I_n(\theta) = E[(\frac{\partial \ln L}{\partial \theta})^2] = -E(\frac{\partial^2 \ln L}{\partial \theta^2})$ .

**Proposition 11** Si les conditions de Cramer-Rao sont vérifiées, on a :

- $E[\frac{\partial \ln L}{\partial \theta}(\theta; X_1, \dots, X_n)] = 0$ ;
- $I_n(\theta) = E\left[\left(\frac{\partial \ln L}{\partial \theta}\right)^2\right] = -E\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right]$ .

**Définition 27** Les conditions de Cramer-Rao étant vérifiées, un estimateur sans biais dont la variance atteint la borne de Cramer-Rao est appelé estimateur efficace.

**Théorème 5** Les conditions de Cramer-Rao étant vérifiées, alors  $\hat{u}$  est un estimateur efficace de  $u(\theta) \iff \exists$  une fonction  $A$  indépendante de  $X_1, \dots, X_n$  telle que  $\frac{\partial \ln L}{\partial \theta}(\theta; X_1, \dots, X_n) = A(n, \theta)(\hat{u} - u(\theta))$ .

**Définition 28** Un estimateur sans biais est dit asymptotiquement efficace si les conditions de Cramer-Rao sont vérifiées et si la variance tend vers la borne de Cramer-Rao lorsque la taille de l'échantillon tends vers l'infini.

**Proposition 12** Si  $\hat{u}$  est un estimateur efficace de  $u(\theta)$ , alors  $Var(\hat{u}) = \frac{u'(\theta)}{A(n, \theta)}$ . De plus un estimateur efficace est exhaustif.

## 5 Méthodes d'estimation

### 5.1 Estimation de l'espérance et de la variance

Estimateur sans biais, convergents, a distribution asymptotique normale :

- de l'espérance :  $\hat{\theta} = \bar{X}$
- de la variance :  $\hat{\theta} = S^{*2}$

### 5.2 Maximum de vraisemblance

$$\begin{cases} \frac{\partial \ln L}{\partial \theta}(\theta; x_1, x_2, \dots, x_n) = 0 \\ \frac{\partial^2 \ln L}{\partial \theta^2}(\theta; x_1, x_2, \dots, x_n) < 0 \end{cases} \iff \theta = \hat{\theta}_{MV}$$

### 5.3 Propriétés

**Proposition 13** Soit  $\hat{\theta}$  un estimateur du maximum de vraisemblance de  $\theta$  et  $u$  une application de  $\mathbb{R} \mapsto \mathbb{R}$ . Alors, sous certaines conditions de régularité,  $u(\hat{\theta})$  est un estimateur du maximum de vraisemblance.

**Proposition 14** S'il existe un estimateur efficace de  $\theta$ , il est identique (p.s.) à l'unique solution de l'équation de vraisemblance.

**Proposition 15** S'il existe une statistique exhaustive  $T$ , alors l'estimateur du maximum de vraisemblance est fonction de  $T$ .

**Proposition 16** Soit  $\theta_0$  la vraie valeur du paramètre. Sous certaines conditions, il existe une suite  $(\hat{\theta})_{n \geq 1}$  de solutions de l'équation de vraisemblance qui converge en probabilité vers  $\theta_0$  :  $\hat{\theta}_n \xrightarrow{P} \theta_0$

**Proposition 17** Sous certaines conditions de régularité, pour toute suite  $(\hat{\theta}_n)_{n \geq 1}$  de solutions de l'équation de vraisemblance t.q.  $\hat{\theta}_n \xrightarrow{P} \theta_0$ , on a :

$$\frac{\hat{\theta}_n - \theta_0}{\sqrt{1/I_n(\theta_0)}} \xrightarrow{L} \mathcal{N}(0, 1)$$

### 5.4 Méthode des moments

#### 5.4.1 Principe

On écrit :  $(m_1, \dots, m_k) = g(\theta)$

Si  $g$  est inversible on en déduit :  $\theta = g^{-1}(m_1, \dots, m_k)$

La méthode des moments consiste à prendre :

$$\hat{\theta}_m = g^{-1}(m_1, \dots, m_k)$$

### 5.4.2 Propriétés

Les estimateurs  $\hat{\theta}_m$  obtenus par la méthode des moments sont :

- convergents :  $\hat{\theta}_m \xrightarrow{P} \theta$
- asymptotiquement gaussiens :

$$\sqrt{n}(\hat{\theta}_m - \theta) \xrightarrow{L} \mathcal{N}(0, \sigma_m^2)$$

## 6 Intervalles de confiance

### 6.1 Définition

$X_1, \dots, X_n$  étant un échantillon i.i.d dont la loi parente  $X$  a une distribution dépendante du paramètre réel  $\theta$ ,  $\alpha$  un réel quelconque fixé appartenant à  $]0, 1[$ , on appelle intervalle de confiance pour le paramètre  $\theta$  au niveau de confiance  $1 - \alpha$  tout intervalle  $[T_1, T_2]$  où  $T_1$  et  $T_2$  sont deux statistiques vérifiant  $P(T_1 \leq \theta \leq T_2) = 1 - \alpha$ .

### 6.2 Construction

- Choisir un estimateur  $T$  de  $\theta$  dont on connaît la loi de probabilité en fonction de  $\theta$ ;
- Déterminer  $f(T, \theta)$  dont la loi de probabilité ne dépend plus de  $\theta$  : la fonction *pivotal*;
- Si c'est possible, en déduire  $P(T_1 \leq \theta \leq T_2) = 1 - \alpha$  où  $T_1$  et  $T_2$  sont des statistiques fonctions de  $T$ .

### 6.3 Variable normale

Hypothèse :  $X_1, \dots, X_n, X_i \sim \mathcal{N}(\mu, \sigma^2)$ , niveau de confiance  $1 - \alpha$ , intervalle bilatéral.

#### 6.3.1 Moyenne

1. Si  $\sigma$  est connue, on a  $T = \bar{X}$  (estimateur) et  $\theta = \mu$ , on sait que  $f(T, \theta) = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0, 1)$ .

Comme  $P(u_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq u_{1 - \frac{\alpha}{2}}) = 1 - \alpha$ , on obtient comme I.C  $[\bar{X} - \frac{\sigma}{\sqrt{n}} u_{1 - \frac{\alpha}{2}}, \bar{X} + \frac{\sigma}{\sqrt{n}} u_{1 - \frac{\alpha}{2}}]$ . En unilatéral, on écrit  $P(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq u_{1 - \alpha}) = 1 - \alpha$  et on obtiendrait  $[\bar{X} - \frac{\sigma}{\sqrt{n}} u_{1 - \alpha}, +\infty[$ .

2. Sinon, on remplace  $\sigma$  par  $S^*$  pour remarquer que

$$\frac{\bar{X} - \mu}{\frac{S^*}{\sqrt{n}}} = \frac{\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{(n-1)S^{*2}}{\sigma^2}}} \sim \mathcal{N}(0, 1) \sim \tau_{n-1} \text{ (pleure pas).}$$

On obtient alors  $P(-t_{n-1; 1 - \frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\frac{S^*}{\sqrt{n}}} \leq t_{n-1; 1 - \frac{\alpha}{2}}) = 1 - \alpha$  d'où I.C =  $[\bar{X} - \frac{S^*}{\sqrt{n}} t_{n-1; 1 - \frac{\alpha}{2}}, \bar{X} + \frac{S^*}{\sqrt{n}} t_{n-1; 1 - \frac{\alpha}{2}}]$ .

#### 6.3.2 Variance

1. Si  $\mu$  est connue, l'estimateur  $T$  est l'estimateur de maximum de vraisemblance  $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$ , d'où  $f(T, \theta) = \frac{n\hat{\sigma}^2}{\sigma^2} = \sum_{i=1}^n \underbrace{\left(\frac{X_i - \mu}{\sigma}\right)^2}_{\sim \mathcal{N}(0, 1)} \sim \chi_n^2$  et on utilise la relation  $P\left(\frac{n\hat{\sigma}^2}{\chi_{n-1; 1 - \frac{\alpha}{2}}^2} < \sigma^2 < \frac{n\hat{\sigma}^2}{\chi_{n-1; \frac{\alpha}{2}}^2}\right)$ .

2. Sinon, l'estimateur est  $S^{*2}$  et  $f(T, \theta) = (n-1) \frac{S^{*2}}{\sigma^2} \sim \chi_{n-1}^2$ .

En utilisant

$$P(\chi_{n-1; \frac{\alpha}{2}}^2 \leq (n-1) \frac{S^{*2}}{\sigma^2} \leq \chi_{n-1; 1 - \frac{\alpha}{2}}^2) = 1 - \alpha$$

on obtient  $[(n-1) \frac{S^{*2}}{\chi_{n-1; 1 - \frac{\alpha}{2}}^2}, (n-1) \frac{S^{*2}}{\chi_{n-1; \frac{\alpha}{2}}^2}]$

## 6.4 Utilisation du maximum de vraisemblance

Soit  $\theta$  un paramètre,  $\hat{\theta}$  son estimateur de maximum de vraisemblance. Alors  $\frac{\hat{\theta} - \theta}{\sqrt{\frac{1}{I_n}}} \xrightarrow{L} \mathcal{N}(0, 1)$ , ce qui constitue une fonction asymptotiquement pivotale pour  $\theta$ .

## 7 Tests

### 7.1 Définitions

#### 7.1.1 Hypothèses

Soit  $X \sim P_\theta, \theta \in \Theta$ . Une hypothèse est une partie  $H$  de  $\theta : H \subset \Theta$ . Elle est simple si  $H$  est un singleton, composite sinon. Un problème de test consiste en un choix entre deux hypothèses  $H_0$  et  $H_1$ . Elles sont supposées complémentaires :  $H_0 \cap H_1 = \emptyset$  et  $H_0 \cup H_1 = \Theta$ .

#### 7.1.2 Règle de décision

Une règle de décision est une application  $\varphi : \mathbb{R}^n \mapsto D$  qui à toute réalisation de l'échantillon  $(x_1, \dots, x_n)$  associe une décision  $\varphi(x_1, \dots, x_n)$  (choix de  $H_0$  ou de  $H_1$ ). Pour résoudre le test, on partitionne l'espace en deux sous-ensembles :

- $W = \varphi^{-1}(d_1)$  des réalisations pour lesquelles on rejette  $H_0$  : c'est la région critique.
- $\bar{W} = \varphi^{-1}(d_0)$  des réalisations pour lesquelles on rejette  $H_1$  : la région d'acceptation.

**Remarque :** La détermination des hypothèses et de la région critique doit se faire avant de connaître le résultat de l'expérience.

#### 7.1.3 Risques et puissance

- Le *risque de première espèce* est la probabilité de rejeter à tort l'hypothèse nulle. Soit  $\forall \theta_0 \in H_0, \alpha_{\theta_0}(\varphi) = P_{\theta=\theta_0}((X_1, \dots, X_n) \in W)$ . Si  $H_0$  est une hypothèse simple, on notera  $\alpha(\varphi)$  ou  $\alpha$  s'il n'y a pas d'ambiguïté sur le test.
- La borne supérieure du risque de première espèce est le *niveau de signification* du test :  $\alpha^*(\varphi) = \sup_{\theta_0 \in H_0} \alpha_{\theta_0}(\varphi)$
- Le *risque de seconde espèce* est la probabilité d'accepter à tort l'hypothèse nulle.  $\forall \theta_1 \in H_1, \beta_{\theta_1}(\varphi) = P_{\theta=\theta_1}((X_1, \dots, X_n) \in \bar{W})$ .
- La quantité  $1 - \beta_{\theta_1}(\varphi)$  est appelée *puissance du test* (pour  $\theta = \theta_1$ ) : c'est la probabilité d'accepter à juste titre  $H_1$ . La fonction  $\theta_1 \mapsto 1 - \beta_{\theta_1}(\varphi)$  est la fonction de puissance.

## 7.2 Méthode de Neyman-Pearson

On contrôle le risque de première espèce en le rendant inférieur à une valeur fixée a priori à un  $\alpha^*$  donné. On se restreint donc à la classe  $\mathcal{C}(\alpha^*) = \{\varphi : \mathbb{R}^n \mapsto \{d_0, d_1\} | \alpha_{\theta_0}(\varphi) \leq \alpha^*, \forall \theta_0 \in H_0\}$ . On cherche alors  $\varphi^* \in \mathcal{C}(\alpha^*)$  telle que  $1 - \beta_{\theta_1}(\varphi^*) = \max_{\varphi \in \mathcal{C}(\alpha^*)} (1 - \beta_{\theta_1}(\varphi)) \forall \theta_1 \in H_1$ . On cherche donc le test le plus puissant pour un niveau de signification donné.

## 7.3 Cas particuliers

### 7.3.1 Hyp. simple vs Hyp. simple

On veut tester

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 (\neq \theta_0) \end{cases}$$

**Théorème 6 (Neyman et Pearson)** La région critique optimale vérifie une relation de la forme  $\frac{L(\theta_1; x_1, \dots, x_n)}{L(\theta_0; x_1, \dots, x_n)} > \lambda$ , où  $\lambda$  se calcule par  $P_{H_0}((X_1, \dots, X_n) \in W) = \alpha^*$

### 7.3.2 Hyp. simple vs Hyp. composite

1. On veut résoudre

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \in E(\theta_0 \notin E) \end{cases} \Rightarrow \begin{cases} H_0 : \theta = \theta_0 \\ H_i : \theta = \theta_i (\theta_i \in E) \end{cases}$$

Si la région critique  $W_i$  est indépendante de  $i$  et toujours égale à  $W$ , elle est optimale pour le test  $\theta = \theta_0$  contre  $\theta \in E$ . L'erreur de seconde espèce sera une fonction de  $\theta$ , mais sera toujours minimale pour  $\alpha^*$  donné. Le test est donc dit *UPP* (Uniformément Plus Puissant).

2. Cas non *UPP* : si  $W_i$  n'est pas constant, la stratégie de Neyman-Pearson ne peut s'appliquer. On conserve la contrainte  $\alpha = P_{H_0}((X_1, \dots, X_n) \in W) \leq \alpha^*$  fixé et on recherche une région critique en utilisant une fonction pivotale.

### 7.3.3 Hyp. composite vs Hyp. composite

1. Test *UPP*

**Définition 29** Une famille de lois  $L_\theta$  est une famille à rapport de vraisemblance monotone s'il existe une statistique  $T$  telle que  $\frac{L(\theta'; x_1, \dots, x_n)}{L(\theta; x_1, \dots, x_n)}$  soit une fonction croissante de  $t$  pour  $\theta' > \theta$ .

**Théorème 7 (Lehman)** Si la famille des lois  $L_\theta$  est à rapport de vraisemblance monotone, alors le problème de test

$$\begin{cases} H_0 : \theta \leq \theta_0 \\ H_1 : \theta > \theta_0 \end{cases}$$

admet un test *UPP* défini par une région critique de la forme  $T > A$  où  $A$  est défini par  $P_{\theta=\theta_0}(T > A) = \alpha^*$ .

2. Test non *UPP*

Forme générale :  $X$  suit une loi dépendant de plusieurs paramètres dont le paramètre réel  $\theta$  sur lequel porte le test :

$$\begin{cases} H_0 : \theta = \theta_0 & \text{autres paramètres inconnus} \\ H_1 : \theta >, <, \neq \theta_0 & \text{autres paramètres inconnus} \end{cases}$$

**Stratégie :**

- Trouver une fonction pivotale de  $\theta$
- Proposer une forme de région critique raisonnable
- Déterminer cette région critique en prenant  $\alpha = P_{H_0}(X \in W) = \alpha^*$ .

## 8 Tests d'hypothèse : comparaison de 2 échantillons

### 8.1 Comparaison de deux échantillons gaussiens

**Définition 30 (loi de Fisher)** Si  $X$  et  $Y$  sont 2 variables aléatoires indépendantes de loi de proba respectives  $\chi_{\nu_1}^2$  et  $\chi_{\nu_2}^2$ , alors la variable  $\frac{X/\nu_1}{Y/\nu_2}$  suit une loi de Fisher à  $\nu_1$  et  $\nu_2$  degrés de liberté.

$$F_{\nu_1, \nu_2, \alpha} = \frac{1}{F_{\nu_2, \nu_1, 1-\alpha}} \quad (\text{voir les tables})$$

#### 8.1.1 Variations : test de Fisher

$$\begin{cases} H_0 : \sigma_X^2 = \sigma_Y^2 & (HC) \\ H_1 : \sigma_X^2 \neq \sigma_Y^2 & (HC) \end{cases} \quad \text{ou} \quad \begin{cases} H_0 : \frac{\sigma_X^2}{\sigma_Y^2} = 1 & (HC) \\ H_1 : \frac{\sigma_X^2}{\sigma_Y^2} \neq 1 & (HC) \end{cases}$$

Régions critiques :  $S_X^{*2}/S_Y^{*2} < F_{n-1, m-1, \alpha}$  ou  $S_X^{*2}/S_Y^{*2} > F_{n-1, m-1, \alpha}$ .

$$\text{Test unilatéraux : } \begin{cases} H_0 : \sigma_X^2 = \sigma_Y^2 & (HC) \\ H_1 : \sigma_X^2 < (\text{resp } >) \sigma_Y^2 & (HC) \end{cases}$$

Régions critiques :  $S_X^{*2}/S_Y^{*2} < F_{n-1, m-1, \alpha}$  (resp  $S_X^{*2}/S_Y^{*2} > F_{n-1, m-1, \alpha}$ ).

#### 8.1.2 Moyennes : test de Student

Estimateur de la variance commune à 2 populations gaussiennes ( $N = m + n$ ) :  $S^{*2} = \frac{1}{N-2}((n-1)S_X^{*2} + (m-1)S_Y^{*2})$ .

**Remarque :**  $\frac{(N-2)S^{*2}}{\sigma^2} = \frac{(n-1)S_X^{*2}}{\sigma^2} + \frac{(m-1)S_Y^{*2}}{\sigma^2} \sim \chi_{N-2}^2$ .  
 $X \sim \mathcal{N}(\mu_X, \sigma^2)$  et  $Y \sim \mathcal{N}(\mu_Y, \sigma^2)$  :

$$\begin{cases} H_0 : \mu_X = \mu_Y & (HC) \\ H_1 : \mu_X \neq \mu_Y & (HC) \end{cases}$$

Région critique :

- $\frac{\bar{X} - \bar{Y}}{S^* \sqrt{1/n + 1/m}} < t_{N-2, \alpha/2}$
- $\frac{\bar{X} - \bar{Y}}{S^* \sqrt{1/n + 1/m}} > t_{N-2, 1-\alpha/2}$

## 8.2 Tests non paramétriques

### 8.2.1 Test de Wilcoxon

$$\begin{cases} H_0 : F_X = F_Y \\ H_1 : F_X > F_Y \end{cases}$$

Suite  $W_X$  : fusion des réalisations des deux échantillons et classement par ordre croissant.

Si les deux échantillons ont la même distribution ( $F_X = F_Y$ ) alors  $E(W_X) = \frac{n(n+m+1)}{2}$  et  $Var(W_X) = \frac{nm(n+m+1)}{12}$ .

Région critique :  $W_X < \frac{n(n+m+1)}{2} - u_{1-\alpha^*} \sqrt{\frac{nm(n+m+1)}{12}}$

### 8.3 Échantillons appariés

Pour l'étude de couples, en règle général, on étudie la différence :  $D_i = Y_i - X_i$ .

$$\begin{cases} H_0 : m_e = 0 \\ H_1 : m_e > \text{ ou } < \text{ ou } \neq 0 \end{cases}$$

### 8.3.1 Test du signe

On compte le nombre de  $D_i > 0$  (soit  $Y_i > X_i$ ). Nous avons donc une  $B(n, p)$ . Le test d'égalité est donc  $p = 1/2$ .

$$\begin{cases} H_0 : p = 1/2 (\iff \text{médiane } M = 0) \\ H_1 : p > 1/2 (\iff M > 0) \end{cases}$$

En cas de valeurs égales, on supprime simplement les couples correspondant à ces égalité.

### 8.3.2 Test de Wilcoxon signé

On ordonne les valeur  $D_i$  par ordre de valeur absolues croissantes (en supprimant les ex-aequo), puis à calculer la somme  $W_+$  des rangs associés aux valeurs positives.

$$\text{Sous } H_0 : E(W_+) = \frac{n(n+1)}{4} \text{ et } \text{Var}(W_+) = \frac{n(n+1)(2n+1)}{24}$$

Pour la région critique, on utilise les tables statistiques, avec  $W : W^+ > A$ . Pour  $n > 19$ , on utilise le fait que  $W_+$  est asymptotiquement gaussien.

## 9 Tests d'adéquation

$$\begin{cases} H_0 : X \sim L \\ H_1 : X \text{ ne suit pas } L \end{cases}$$

### 9.1 Test du $\chi^2$

#### 9.1.1 Loi multinomiale

**Définition 31** On considère une expérience aléatoire pouvant prendre  $K$  valeurs avec les probabilité  $p_1, \dots, p_K$ . Ayant réalisé  $n$  fois cette expérience, il est possible de définir le vecteur aléatoire  $(N_1, \dots, N_K)$  où  $N_K$  est le nombre de fois où la valeur  $k$  a été obtenue. Ce vecteur suit alors par définition une loi multinomiale de paramètre  $p_1, \dots, p_K, n$ . Sa loi de probabilité est définie par  $P(N_1 = n_1, \dots, N_K = n_K) = \frac{n!}{n_1! \dots n_K!} p_1^{n_1} \dots p_K^{n_K}$ .

**Théorème 8** Si  $(N_1, \dots, N_K)$  est un vecteur aléatoire suivant une loi multinomiale de paramètres  $p_1, \dots, p_K, n$ , alors  $\sum_{k=1}^K \frac{(N_k - np_k)^2}{np_k}$  suit asymptotiquement une loi du  $\chi^2$  à  $K - 1$  degrés de liberté.

#### 9.1.2 Principe du test

La première étape de ce test consiste à définir  $K$  classes en répartissant les valeurs possibles de la v.a.  $X$  en  $K$  sous-ensemble. On peut alors associer à un échantillons de taille  $n$  le vecteur aléatoire  $(N_1, \dots, N_K)$ . Il est alors possible de calculer les probabilités  $p_k$  de chacune des classes à partir de la définition des classes et de la loi de probabilité  $L$ .

$$\text{Ce qui donne : } \begin{cases} H_0 : \text{les paramètres sont les } p_k \\ H_1 : \text{les paramètres ne sont pas les } p_k \end{cases}$$

$$\text{Fonction pivotale : } D^2 = \sum_{k=1}^K \frac{(N_k - np_k)^2}{np_k} \sim \chi_{K-1}^2$$

$$\text{Région critique : } D^2 = \sum_{k=1}^K \frac{(N_k - np_k)^2}{np_k} > \chi_{K-1, 1-\alpha}^2$$

#### 9.1.3 Remarques

- Ceci est applicable uniquement si  $np_k > 5$ . Sinon, il faut faire des regroupement par classe de façon à ce qu'on aie  $np_k > 5$ .
- Si certains paramètres de la loi de l'hypothèse  $H_0$  ne sont pas connus, il est possible de les remplacer par leur estimateur du maximum de vraisemblance, mais alors

$$D^2 = \sum_{k=1}^K \frac{(N_k - np_k)^2}{np_k} \sim \chi_{K-1-r}^2 \text{ où } r \text{ est le nombre de paramètres réel estimés.}$$

- En pratique, on utilise souvent l'expression suivante :

$$D^2 = \sum_{k=1}^K \frac{N_k^2}{np_k} - n$$

### 9.1.4 Application aux tables de contingence

On considère 2 variables aléatoires  $X$  et  $Y$  ne pouvant prendre respectivement que  $r$  et  $s$  valeurs (variables quantitatives). La donnée d'un échantillon  $(X_1, Y_1), \dots, (X_n, Y_n)$  permet alors de définir une table de contingence  $(N_{ij}, i \in [1, r], j \in [1, s])$  où la modalité  $i$  a été prise simultanément avec la modalité  $j$ . On définit aussi les marges  $N_{i.} = \sum_{j \in [1, s]} N_{ij}$  et  $N_{.j} = \sum_{i \in [1, r]} N_{ij}$  :

$X^Y$	1	...	$j$	...	$s$	
1	$N_{11}$	...	$N_{1j}$	...	$N_{1s}$	$N_{1.}$
...	...	...	...	...	...	...
$i$	$N_{i1}$	...	$N_{ij}$	...	$N_{is}$	$N_{i.}$
...	...	...	...	...	...	...
$r$	$N_{r1}$	...	$N_{rj}$	...	$N_{rs}$	$N_{r.}$
	$N_{.1}$	...	$N_{.j}$	...	$N_{.s}$	

On cherche à vérifier si deux variables  $X$  et  $Y$  sont indépendantes. Pour ceci, on effectue le test :

$$\begin{cases} H_0 : X \text{ et } Y \text{ sont indépendantes} \\ H_1 : X \text{ et } Y \text{ ne sont pas indépendantes} \end{cases}$$

On a donc

$$D^2 = \sum_i \sum_j \frac{(N_{ij} - \frac{N_{i.} N_{.j}}{n})^2}{\frac{N_{i.} N_{.j}}{n}} \sim \chi_{(r-1)(s-1)}^2$$

## 9.2 Test de Kolmogorov

Soit  $F$  la fonction de répartition empirique dont est issu l'échantillon et  $F_0$  la fonction de répartition de  $X$  sous  $H_0$ ,

$$\text{alors le test est : } \begin{cases} H_0 : F = F_0 \\ H_1 : F \neq F_0 \end{cases}$$

$D_n = \sup_x |\hat{F}(x) - F_0(x)|$  a une distribution qui, sous l'hypothèse  $H_0$ , ne dépend plus de la loi de  $X$ .

Région critique :  $D_n > d_{n, 1-\alpha^*}$  ( $d_{n, 1-\alpha^*}$  sur les tables statistiques)

$$D_n = \max_i \max \left( |\hat{F}(x_i) - F_0(x_i)|, |\hat{F}(x_i^-) - F_0(x_i)| \right)$$

Remarques :

- Le test n'est applicable que pour les v.a. continues.
- Si le test de K. est applicable, il est plus rigoureux que le test du  $\chi^2$  puisqu'il ne nécessite aucune approximation.
- La loi de référence doit être parfaitement connue. Lorsqu'on veut vérifier si une distribution est gaussienne sans préciser plus les paramètres de la loi, le test de K. n'est donc pas applicable; toutefois, un test de normalité s'appuyant sur la statistique  $D_n$  a pu être défini de la façon suivante : on estime les paramètres de la distribution par  $\bar{X}$  et  $S^{*2}$  et on prend comme région critique :
  - si  $\alpha = 0.05$   $(\sqrt{n} + \frac{0.85}{\sqrt{n}} - 0.01) D_n > 0.895$ ;
  - si  $\alpha = 0.01$   $(\sqrt{n} + \frac{0.85}{\sqrt{n}} - 0.01) D_n > 1.035$ .

$$\text{On a alors } D_n^* = \sup_x |\hat{F}(x) - \Phi_{\frac{x-\bar{x}}{S^*}}|$$

Ce test de normalité est appelé test de Stephens (ou de Lilliefors). De la même façon, on peut faire un test d'exponentialité : on détermine  $D_n$  en prenant  $\bar{X}$  comme estimateur du paramètre de la loi exponentielle et on prend comme région critique :

- si  $\alpha = 0.05$   $(\sqrt{n} + \frac{0.5}{\sqrt{n}} - 0.26)D_n > 1.094$ ;
- si  $\alpha = 0.01$   $(\sqrt{n} + \frac{0.5}{\sqrt{n}} - 0.26)D_n > 1.308$ .

## 10 Analyse de la variance

### 10.1 Estimation des paramètres du modèle

Soient  $k$  échantillons de taille  $n_k$  et de loi :

$$X_k^i \sim \mathcal{N}(\mu_k^i, \sigma^2), \forall i \in [0, \dots, n_k]$$

La méthode du maximum de vraisemblance donne :

$$\begin{aligned} - \mu_{k_{MV}} &= \bar{X}_k \\ - \hat{\sigma}_{MV}^2 &= \frac{1}{N} \sum_{k=1}^K \sum_{i=1}^{n_k} (X_k^i - \bar{X}_k)^2 = \frac{1}{N} \sum_{k=1}^K (n_k - 1) S_k^{*2} \end{aligned}$$

Mais il est préférable de retenir comme estimateur sans biais de l'espérance :  $\frac{1}{N-K} \sum_{k=1}^K (n_k - 1) S_k^{*2}$ , appelé  $MSW$

On en déduit également que :  $\bar{X} = \frac{1}{N} \sum_{k=1}^K n_k \bar{X}_k$

### 10.2 Formule de l'analyse de la variance

$$\underbrace{\sum_k \sum_i (X_k^i - \bar{X})^2}_{SST} = \underbrace{\sum_k \sum_i (X_k^i - \bar{X}_k)^2}_{SSW} + \underbrace{\sum_k (\bar{X} - \bar{X}_k)^2}_{SSB}$$

Dispersion totale = Dispersion intra-classe + Dispersion inter-classe

### 10.3 Moyenne des dispersions

- variance totale :  $MST = \frac{SST}{N-1}$ ;
- variance intra-modalités :  $MSW = \frac{SSW}{N-K}$ ;
- variance inter-modalités :  $MSB = \frac{SSB}{K-1}$ .

### 10.4 vérification des hypothèses du modèle

- test de normalité : test de Kolmogorov ou du  $\chi^2$
- test d'égalité des variances (Bartlett) :

$$\begin{cases} H_0 : \sigma_1 = \sigma_2 = \dots = \sigma_K \\ H_1 : \exists k, l \quad \sigma_k \neq \sigma_l, \end{cases}$$

On rejette  $H_0$  si  $B \geq \chi_{K-1, 1-\alpha}^2$ , avec :

$$B = (N - K) \ln(MSW) - \sum_{k=1}^K (n_k - 1) \ln S_k^{*2}$$

### 10.5 Test de l'analyse de la variance

Pour tester si les  $K$  population ont la même distribution, il suffit de faire le test :  $\begin{cases} H_0 : \sigma_1 = \sigma_2 = \dots = \sigma_K \\ H_1 : \exists k, l \quad \sigma_k \neq \sigma_l, \end{cases}$

et la région critique est de la forme :  $\frac{MSB}{MSW} > F_{K-1, N-K, 1-\alpha}$

On présente souvent les calculs intermédiaires dans un «tableau d'analyse de la variance» :

Variation	Inter-modalité	Intra-modalité	total
° de liberté	$K - 1$	$N - K$	$N - 1$
$\sum$ des $^2$	$SSB$	$SSW$	$SST$
Moy. des $^2$	$MSB = \frac{SSB}{K-1}$	$MSW = \frac{SSW}{N-K}$	
Rapport	$F = \frac{MSB}{MSW}$		

### 10.6 Effet du facteur : procédure LSD

- si on accepte  $H_0$ , le facteur n'a pas d'effet, et la procédure s'arrête.
- sinon, il est intéressant de tester l'égalité des moyennes 2 a 2. La région critique pour la comparaison des niveaux  $k$  et  $l$  est :

$$W_{kl} : \frac{|\bar{X}_k - \bar{X}_l|}{\sqrt{MSW (\frac{1}{n_k} + \frac{1}{n_l})}} > t_{N-K; 1-\frac{\alpha^*}{2}}$$

### 10.7 Le calcul du début à la fin

- Calculer les  $\bar{X}_k, S_k^{*2}$ ,  $k = 1, \dots, K$ . Rappel :  $\bar{X} = \frac{1}{N} \sum_k n_k \bar{X}_k$ .
- $SSW = \sum_k (n_k - 1) S_k^{*2}$ ,  $MSW = \frac{SSW}{N-K}$ .
- $SSB = \sum_k n_k (\bar{X}_k - \bar{X})^2 = \sum_k n_k \bar{X}_k^2 - N \bar{X}^2$ ,  $MSB = \frac{SSB}{K-1}$ .
- La région critique est  $W : F = \frac{MSB}{MSW} >$  (ou  $<$  selon le test)  $F_{K-1, N-K, 1-\alpha^*}$ .
- Éventuellement : tester la validité du modèle (test de normalité de chaque échantillon avec *Lillefors-Stephen*, test d'égalité des variances avec *Fisher* pour deux échantillons, *Bartlett* pour plus : voir **10.4**).

### 10.8 Le test de Kruskal-Wallis

C'est un équivalent non paramétrique de l'analyse de la variance, qui ne suppose pas la normalité des échantillons. Le test repose sur la statistique  $H = \frac{SSB}{MST}$  calculée sur les rangs : on mélange toutes les valeurs, on les range par ordre croissant et on leur attribue un rang.

Les valeurs de  $SSB$  et  $SST$  deviennent :

$$\begin{cases} SSB = \sum_{k=1}^K \frac{1}{n_k} (\sum_{i=1}^{n_k} R_k^i)^2 - \frac{N(N+1)^2}{4} \\ SST = \frac{N(N+1)(N-1)}{12} \end{cases}$$

et donc  $MST = \frac{SST}{N-1} = \frac{N(N+1)}{12}$  d'où :

$$H = \frac{SSB}{MST} = \frac{12}{N(N+1)} \sum_{k=1}^K \frac{1}{n_k} (\sum_{i=1}^{n_k} R_k^i)^2 - 3(N+1)$$

Si  $N$  n'est pas trop petit,  $H$  suit approximativement une loi  $\chi_{K-1}^2$ . La région critique du test est donc :

$$W : H > \chi_{K-1, 1-\alpha}^2$$