

CHAPITRE 2. STATISTIQUES EMPIRIQUES

1. Moyenne, variance et moments empiriques

1.1. Échantillon d'une loi. — On considère une loi \mathcal{L} (mesure de probabilité) sur \mathbb{R} .

En modélisation, une situation courante est que l'on dispose de n observations X_1, \dots, X_n issues d'une expérience non déterministe (pensez aux jets de "pile ou face", aux pointages de nombres de personnes faisant la queue dans un bureau de poste le samedi matin ; ou encore les cours de change euro/dollar des trentes dernières journées).

En première approximation, ou après transformation et simplification, on est amené à considérer que le n -uplet (X_1, \dots, X_n) est constitué de n copies i.i.d. d'une loi \mathcal{L} . C'est ce que nous allons supposer dans la première partie du cours.

Nous appelons le n -uplet (X_1, \dots, X_n) un *échantillon de taille n* de la loi \mathcal{L} .

Pour simplifier quelquefois la présentation, on note souvent X une v.a. générique de loi \mathcal{L} . Nous verrons que dans une modélisation dite markovienne (par une chaîne de Markov ou un processus de Markov), les observations ne sont alors plus indépendantes.

1.2. Moyenne, variance et moments empiriques. — L'adjectif *empirique* réfère dans ce cours à une valeur ou quantité fonction d'un échantillon (X_1, \dots, X_n) . Mentionnons la définition du *Petit Robert* : « Empirique - Qui reste au niveau de l'expérience spontanée ou commune, n'a rien de rationnel ni de systématique ».

En modélisation, le calcul des probabilités, ou encore appelé *Statistique asymptotique*, vise à ce que les valeurs empiriques convergent vers les valeurs théoriques correspondantes ; et le travail de modélisation consiste à bâtir un modèle mathématique où on peut définir des valeurs théoriques vers lesquelles convergeront ces valeurs empiriques à l'issue d'un grand nombre d'expériences.

Exemples de valeurs théoriques. —

– Moyenne et variance de la loi \mathcal{L} : $m = \mathbb{E}[X]$, $\sigma^2 = \text{Var}(X)$.

– Moments : $\alpha_k = \mathbb{E}[X^k]$, $k \geq 1$.

– Moments centrés : $\mu_k = \mathbb{E}[(X - m)^k]$, $k \geq 1$.

On a : $m = \alpha_1$, $\sigma^2 = \alpha_2 - \alpha_1^2$.

Exemple de valeurs empiriques. —

– Moyenne et variance empirique :

$$\bar{X} = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad S^2 = S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 .$$

– Moments empiriques :

$$\hat{\alpha}_k = \frac{1}{n} \sum_{i=1}^n X_i^k .$$

On a : $S^2 = \hat{\alpha}_2 - \hat{\alpha}_1^2$.

Terminologie. On dit que X , ou encore sa loi \mathcal{L} a un moment d'ordre k si $\mathbb{E}|X|^k < \infty$.

Proposition 1.1. — *Supposons que la loi \mathcal{L} a un moment d'ordre $p \geq 1$. Alors*

1. *Loi des grands nombres : pour tout $1 \leq k \leq p$, $\hat{\alpha}_k \xrightarrow{p.s.} \alpha_k$.*

2. *TLC : soit $s = [p/2]$. Le vecteur*

$$\sqrt{n}(\hat{\alpha}_1 - \alpha_1, \dots, \hat{\alpha}_s - \alpha_s) \Rightarrow \mathcal{N}(0, \Gamma) ,$$

où la covariance Γ a pour éléments $\Gamma_{k\ell} = \text{Cov}(X^k, X^\ell) = \alpha_{k+\ell} - \alpha_k \alpha_\ell$.

1.3. Le Couple (\bar{X}, S^2) . —

Proposition 1.2. — Supposons que \mathcal{L} a un moment d'ordre 4. Alors

$$\sqrt{n} \left\{ \begin{pmatrix} \bar{X} \\ S^2 \end{pmatrix} - \begin{pmatrix} \alpha_1 \\ \mu_2 \end{pmatrix} \right\} \Rightarrow \mathcal{N}(0, \Sigma), \quad \Sigma = \begin{pmatrix} \mu_2 & \mu_3 \\ \mu_3 & \mu_4 - \mu_2^2 \end{pmatrix}.$$

Démonstration. — — On considère d'abord le cas centré : $\alpha_1 = 0$. La proposition ci-dessus donne

$$\sqrt{n} \left\{ \begin{pmatrix} \bar{X} \\ \bar{X}^2 \end{pmatrix} - \begin{pmatrix} 0 \\ \alpha_2 \end{pmatrix} \right\} \Rightarrow \mathcal{N}(0, \Gamma), \quad \Gamma = \begin{pmatrix} \alpha_2 & \alpha_3 \\ \alpha_3 & \alpha_4 - \alpha_2^2 \end{pmatrix}.$$

Soit $\phi : (x, y) \mapsto (x, y - x^2)$. On a $(\bar{X}, S^2) = \phi(\bar{X}, \bar{X}^2)$ et $\phi'(0, \alpha_2) = I$. Par la méthode delta, on obtient $\sqrt{n} [(\bar{X}, S^2) - (0, \alpha_2)] \Rightarrow \mathcal{N}(0, \Gamma)$.

– Le cas général se ramène au cas précédent par le changement de variables $Y_i = X_i - \alpha_1$. □

Conséquences :—

- Pour la variance empirique, $\sqrt{n}(S^2 - \mu_2) \Rightarrow \mathcal{N}(0, \mu_4 - \mu_2^2)$.
- Les deux “estimateurs” \bar{X}, S^2 sont asymptotiquement indépendants si $\mu_3 = 0$: c'est le cas pour toute loi symétrique par rapport à sa moyenne, notamment les lois de Laplace ou gaussiennes.

2. Méthode des moments

2.1. Estimation d'un paramètre. — La loi \mathcal{L} est inconnue, mais on suppose qu'elle appartient à une famille de lois $\{P_\theta\}$ dépendant d'un paramètre $\theta \in \mathbb{R}^s$. Par exemple on pense que la loi \mathcal{L} est gaussienne mais on ne connaît ni sa moyenne ni sa variance. Cela amène à considérer la famille des lois à deux paramètres

$$P_\theta(dx) = g_{m, \sigma^2}(x)dx, \quad \theta = (m, \sigma^2) \in \mathbb{R}^2.$$

Un *estimateur* de θ est une fonction $T_n = \phi_n(X_1, \dots, X_n)$ de l'échantillon qui approxime θ . Chaque réalisation $T_n(\omega)$ fournit alors une *estimation* du paramètre inconnu θ .

L'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, traditionnel en calcul de probabilités est ici remplacé par une famille $(\Omega, \mathcal{A}, \{P_\theta, \theta \in \mathbb{R}^p\})$. De façon plus formelle, nous supposons que l'on dispose d'un espace mesurable (Ω, \mathcal{A}) d'événements (dits aléatoires). La famille de probabilités $\{P_\theta\}$ sont toutes définies sur cet espace. Dans les applications, on appelle un *modèle statistique paramétrique* le triplet $(\Omega, \mathcal{A}, \{P_\theta, \theta \in \mathbb{R}^s\})$, et sa *dimension paramétrique* l'entier s .

L'opérateur d'espérance mathématique associée à la famille $\{P_\theta\}$ est noté $\{E_\theta\}$.

2.2. Méthode des moments. —

Forme classique :— Pour chaque θ , soit le vecteur des s premiers moments

$$\psi(\theta) = [\mathbb{E}_\theta(X), \dots, \mathbb{E}_\theta(X^s)] = [\alpha_1(\theta), \dots, \alpha_s(\theta)].$$

Soient aussi les s premiers moments empiriques $[\hat{\alpha}_k]_{1 \leq k \leq s}$.

On estime θ par une solution du système de s *équations de moments*

$$\begin{cases} \hat{\alpha}_1 = \alpha_1(\theta), \\ \vdots \\ \hat{\alpha}_s = \alpha_s(\theta). \end{cases}$$

Exemple 2.1. — Supposons que les observations X_1, \dots, X_n suivent toutes une loi exponentielle de paramètre θ que l'on cherche à estimer. En d'autres termes, $\theta \in]0, \infty[$ et $\mathbb{P}_\theta(dx) = \theta e^{-\theta x} dx$ sur $]0, \infty[$.

La dimension paramétrique vaut $s = 1$. La méthode des moments utilise un seul des moments. En choisissant le premier d'entre eux, nous avons à résoudre l'équation

$$\bar{X} = \hat{\alpha}_1 = \mathbb{E}_\theta[X] = \frac{1}{\theta}.$$

La solution, $\hat{\theta}_n = 1/\bar{X}$ est un estimateur de moment du paramètre θ .

Question. Si on avait utilisé le moment d'ordre deux à la place du moment d'ordre 1, que serait l'estimateur de moment de θ ?

Forme générale : — Plus généralement, on se donne une fonction vectorielle $f = (f_k)$ de \mathbb{R} dans \mathbb{R}^s et on considère le vecteur des moments théoriques ψ :

$$\theta \mapsto \psi(\theta) = \mathbb{E}_\theta f(X) = [\mathbb{E}_\theta f_k(X)].$$

Les moments empiriques sont

$$\hat{f}_k = \frac{1}{n} \sum_{i=1}^n f_k(X_i).$$

L'estimateur des moments de θ est une solution du système

$$\hat{f}_k = \mathbb{E}_\theta f_k(X), \quad 1 \leq k \leq s.$$

que l'on réécrit sous forme vectorielle

$$\hat{f} = \begin{pmatrix} \hat{f}_1 \\ \vdots \\ \hat{f}_s \end{pmatrix} = \begin{pmatrix} \mathbb{E}_\theta f_1(X) \\ \vdots \\ \mathbb{E}_\theta f_s(X) \end{pmatrix} = \psi(\theta).$$

Soit $\hat{\theta}_n$ une solution de ce système. Si la vraie loi est P_{θ_0} et que $\mathbb{E}_{\theta_0} \|f(X)\| < \infty$, par la loi forte des grands nombres, \hat{f} converge vers $\psi(\theta_0)$ presque-sûrement. Si de plus ψ a de bonne propriété d'inversion autour de θ_0 , on peut s'attendre à la convergence de $\hat{\theta}_n$ vers θ_0 .

Théorème 2.2. — Supposons que ψ est définie sur un ouvert Θ de \mathbb{R}^s , de classe \mathcal{C}^1 dans un voisinage de θ_0 , et que la matrice jacobienne $\psi'(\theta_0)$ est inversible. Supposons aussi

$$\mathbb{E}_{\theta_0} \|f(X)\|^2 < \infty.$$

Alors l'estimateur des moments $\hat{\theta}_n$ existe avec une probabilité tendant vers 1, et que

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \Rightarrow \mathcal{N}(0, \Gamma_0),$$

où

$$\Gamma_0 = \psi'(\theta_0)^{-1} \text{Cov}_{\theta_0}[f(X)f(X)^T] \psi'(\theta_0)^{-1T}.$$

Démonstration. — — Théorème de fonction implicite : ψ est un difféomorphisme $U \rightarrow V$, où U et V sont respectivement un voisinage de θ_0 et $\psi(\theta_0) = \mathbb{E}_{\theta_0} f(X)$.

$\hat{\theta}_n$ est bien défini dès que $\hat{f} \in V$; mais $\mathbb{P}_{\theta_0}(\hat{f} \in V)$ tend vers 1 par la LGN.

– TLC : $\sqrt{n}[\hat{f} - \psi(\theta_0)] \Rightarrow \mathcal{N}(0, \text{Cov}_{\theta_0}[f(X)f(X)^*])$.

– Par le lemme de Slutsky, $\mathbb{1}_{\hat{f} \in V} \sqrt{n}[\hat{f} - \psi(\theta_0)]$ converge en loi vers la même limite.

– Conclusion par la méthode delta appliquée à la fonction ψ^{-1} .

□

Exemple 2.3. — *Lois Beta.* La loi Beta(a, b), de paramètres $\theta = (a, b) > (0, 0)$, a pour densité

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} \mathbb{1}_{[0,1]}(x).$$

Avec $\theta = (a, b)$, les moments sont

$$\alpha_k(\theta) = \mathbb{E}_\theta(X^k) = \frac{a(a+1)\cdots(a+k-1)}{(a+b)(a+b+1)\cdots(a+b+k-1)}.$$

En particulier la variance vaut $ab(a+b)^{-2}(a+b+1)^{-1}$.

L'estimateur basé sur les deux premiers moments vaut

$$(\hat{a}_n, \hat{b}_n) = \frac{\bar{X} - \bar{X}^2}{S^2} (\bar{X}, 1 - \bar{X}).$$

Le théorème précédent s'applique ici donnant un TLC pour (\hat{a}_n, \hat{b}_n) .

3. Statistiques d'ordre, de rangs

Dans tout ce paragraphe, on supposera que la loi commune \mathcal{L} d'un échantillon (X_1, \dots, X_n) est sans atome, de sorte que l'événement $\{X_1 = X_2\}$ est négligeable.

Lemme 3.1. — *Soient deux variables i.i.d. de loi commune \mathcal{L} qui n'a pas d'atomes. Alors, $\mathbb{P}(X_1 = X_2) = 0$.*

La preuve de ce lemme est laissé en exercice pour le lecteur.

On réordonne les observations (X_1, \dots, X_n) dans le sens croissant et on note l'échantillon ordonné $(X_{(k)})$ avec

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}.$$

On appelle *les statistiques d'ordre* associées à l'échantillon le vecteur $X_{\leq} = (X_{(k)})_{1 \leq k \leq n}$.

Soit R_{ni} le rang de X_i dans l'échantillon ordonné, i.e. $X_i = X_{(R_{ni})}$. Les *statistiques de rang* de l'échantillon est $R_n = (R_{ni})$. Remarquons que l'hypothèse d'absence d'atomes faite sur la loi \mathcal{L} entraîne qu'avec probabilité 1 il n'y ait pas d'ex-aequo et que le vecteur R_n prenne ses valeurs dans le groupe \mathcal{T}_n des permutations de $\{1, \dots, n\}$. Notons $\mathcal{O}_n = \{x = (x_i) \in \mathbb{R}^n : x_1 \leq x_2 \leq \cdots \leq x_n\}$.

Théorème 3.2. — *On suppose que la loi \mathcal{L} est sans atomes.*

1. Pour tout borélien $A \subset \mathcal{O}_n$,

$$\mathbb{P}\{(X_{(1)}, \dots, X_{(n)}) \in A\} = n! \mathbb{P}\{(X_1, \dots, X_n) \in A\}.$$

2. La loi de R_n est uniforme sur \mathcal{T}_n , i.e. pour tout $\sigma \in \mathcal{T}_n$, $\mathbb{P}(R_n = \sigma) = 1/n!$.

3. Les statistiques d'ordre X_{\leq} et les statistiques de rang R_n sont (des vecteurs aléatoires) indépendants.

Démonstration. — Soit $\sigma \in \mathcal{T}_n$ et σ^{-1} son inverse. Les 3 conclusions viennent toutes de la relation importante suivante

$$\begin{aligned} \mathbb{P}\{R_n = \sigma, (X_{(1)}, \dots, X_{(n)}) \in A\} &= \mathbb{P}\{(X_{\sigma^{-1}(1)}, \dots, X_{\sigma^{-1}(n)}) \in A\} \\ &= \mathbb{P}\{(X_1, \dots, X_n) \in A\}. \end{aligned}$$

□

Corollaire 3.3. — *On suppose que la loi \mathcal{L} est sans atomes.*

1. Pour tout $1 \leq k \leq n$, la k^{e} statistique de rang R_{nk} a la loi uniforme sur $\{1, \dots, n\}$.

2. Si la loi \mathcal{L} a une densité f , X_{\leq} a pour densité

$$n! f(x_1) \cdots f(x_n) \mathbb{1}_{\mathcal{O}_n}(x_1, \dots, x_n).$$

La proposition suivante est souvent utile.

Proposition 3.4. — Si la loi \mathcal{L} a une densité f et on note F sa fonction de répartition. Alors, la k^e statistique d'ordre $X_{(k)}$ a pour densité

$$h_{(k)}(x) = \frac{n!}{(k-1)!(n-k)!} F(x)^{k-1} f(x) [1 - F(x)]^{n-k}, \quad x \in \mathbb{R}.$$

Démonstration. — On a

$$\begin{aligned} H_{(k)}(x) &:= \mathbb{P}(X_{(k)} \leq x) \\ &= \mathbb{P}(\text{au moins } k \text{ des } n \text{ valeurs } \leq x) \\ &= \sum_{\ell=k}^n \binom{n}{\ell} [F(x)]^\ell [1 - F(x)]^{n-\ell}. \end{aligned}$$

En dérivant cette expression, on obtient $h_{(k)}(x) = H'_{(k)}(x)$ comme une somme télescopique qui se simplifie. \square

Exemple 3.5. — *Statistiques d'ordre uniformes.* Si la loi \mathcal{L} est la loi uniforme (sur $[0,1]$), les statistiques d'ordre U_{\leq} a pour densité $n! \mathbb{1}_{D_n}(x)$ où on a noté le simplex $D_n = \{x \in \mathbb{R}^n : 0 \leq x_1 \leq \dots \leq x_n \leq 1\}$.

La loi de $U_{(k)}$ a pour densité

$$h_{(k)}(x) = \frac{n!}{(k-1)(n-k)!} x^{k-1} (1-x)^{n-k}, \quad x \in [0, 1].$$

Autrement dit, $U_{(k)}$ suit une loi Beta($k, n - k + 1$). En particulier

$$\mathbb{E}U_{(k)} = \frac{k}{n+1}, \quad \text{Var}(U_{(k)}) = \frac{k(n-k+1)}{(n+1)^2(n+2)}.$$

3.1. Statistique linéaire des rangs. — On note F la f.d.r. de la loi \mathcal{L} suppose continue. Une *statistique linéaire des rangs* est de la forme

$$T_n = \sqrt{12} \sum_{i=1}^n c_{ni} \frac{R_{ni}}{n+1},$$

où $(c_{ni})_{1 \leq i \leq n}$ est une famille de coefficients réels.

Théorème 3.6. — Supposons que les coefficients vérifient les conditions

$$\sum_{i=1}^n c_{ni} = 0, \quad \sum_{i=1}^n c_{ni}^2 = 1, \quad \max_{1 \leq i \leq n} |c_{ni}| \rightarrow 0 \quad \text{quand } n \rightarrow \infty. \quad [\mathbf{C}]$$

Alors $T_n \Rightarrow \mathcal{N}(0, 1)$ quand $n \rightarrow \infty$.

Démonstration. — On introduit la statistique

$$S_n = \sqrt{12} \sum_{i=1}^n c_{ni} F(X_i) = \sqrt{12} \sum_{i=1}^n c_{ni} \left[F(X_i) - \frac{1}{2} \right].$$

Puisque F est continue, les $\{U_i = F(X_i)\}$ forment un échantillon de la loi uniforme (on reviendra en détails sur ce calcul).

Étape 1 : par le TLC pour une suite de v.a. indépendantes vu au chapitre 1, $S_n \Rightarrow \mathcal{N}(0, 1)$.

Étape 2 : Nous allons montrer que $\mathbb{E}(T_n - S_n)^2 \rightarrow 0$. La loi du vecteur

$$(V_1, \dots, V_n) = \left(\frac{R_{n1}}{n+1} - U_1, \dots, \frac{R_{nn}}{n+1} - U_n \right)$$

est échangeable. En effet, il existe une fonction Φ telle que $(V_1, \dots, V_n) = \Phi(X_1, \dots, X_n)$, et la loi de ces vecteurs est invariante par permutation de ses coordonnées. Il en résulte que les V_i ont toutes une même loi, de même que pour tous les produits $V_i V_j$.

Par ailleurs,

$$\begin{aligned}
\frac{1}{12} \mathbb{E}(T_n - S_n)^2 &= \mathbb{E}\left(\sum_i c_{ni} V_i\right)^2 \\
&= \sum_i c_{ni}^2 \mathbb{E}V_i^2 + \sum_{i \neq j} c_{ni} c_{nj} \mathbb{E}(V_i V_j) \\
&= \left(\sum_i c_{ni}^2\right) \mathbb{E}V_1^2 + \left(\sum_{i \neq j} c_{ni} c_{nj}\right) \mathbb{E}(V_1 V_2) \\
&= \left(\sum_i c_{ni}^2\right) \{\mathbb{E}V_1^2 - \mathbb{E}(V_1 V_2)\} \\
&\leq 2\mathbb{E}V_1^2.
\end{aligned}$$

D'autre part,

$$\begin{aligned}
\mathbb{E}V_1^2 &= \sum_{k=1}^n \mathbb{E} \left\{ \left(\frac{R_{n1}}{n+1} - U_1 \right)^2 \mathbb{1}_{R_{n1}=k} \right\} \\
&= \sum_{k=1}^n \mathbb{E} \left\{ \left(\frac{k}{n+1} - U_{(k)} \right)^2 \mathbb{1}_{R_{n1}=k} \right\} \\
&= \sum_{k=1}^n \mathbb{E} \left\{ \left(\frac{k}{n+1} - U_{(k)} \right)^2 \right\} \mathbb{E} \{ \mathbb{1}_{R_{n1}=k} \} \quad (\text{indépendance entre } R_n \text{ et } U_{\leq}) \\
&= \frac{1}{n} \sum_{k=1}^n \frac{k(n-k+1)}{(n+1)^2(n+2)} \rightarrow 0.
\end{aligned}$$

D'où la conclusion. □

4. Introduction au test d'hypothèses

En statistique une hypothèse est représentée par une famille de mesures de probabilités qui pourraient être la loi des observations $X = (X_1, \dots, X_n)$. On considère deux hypothèses : *l'hypothèse nulle* H_0 représentée par une famille de mesures de probabilités $\{P\}$ et *l'hypothèse alternative* H_1 représentée par une autre famille de mesures de probabilités $\{Q\}$. Naturellement les deux familles sont disjointes. Parfois on parle simplement de *l'hypothèse* H_0 et de *l'alternative* H_1 .

Lorsque la famille $\{P\}$ ne comporte qu'un seul élément, l'hypothèse H_0 est dite *simple*; autrement elle est dite *multiple*. De la même façon, on appelle H_1 simple ou multiple suivant le nombre d'éléments de la famille $\{Q\}$. Si les hypothèses sont simples, il est commode de noter P_0 et P_1 pour les deux probabilités impliquées.

Bien que d'un point de vue mathématique les deux hypothèses soient semblables, leur nature est différente dans la tête d'un statisticien. L'hypothèse nulle exprime bien souvent une notion d'égalité, de symétrie, ou encore d'indépendance et elle correspond à une position de réserve. L'hypothèse alternative cherche à cerner la présence d'une différence, d'une asymétrie, ou encore d'une dépendance. Le chercheur espère à prouver, du moins à conforter l'hypothèse alternative à l'aide des expériences. Ainsi le type et le nombre d'hypothèses nulles sont plutôt limités, tandis qu'il existe une abondance d'hypothèses alternatives. *Thus the tendency to adhere to null hypotheses as long as it is reasonable is fully justified by the economy of thought*

and the limited capacity of the human brain (Hájek et Šidák). Par ailleurs, le calcul des probabilités est bien plus simple sous l'hypothèse nulle que sous l'alternative.

Un test de l'hypothèse H_0 contre l'alternative H_1 consiste à décider laquelle des deux hypothèses est, ou devrait être, vraie. La décision est basée sur la valeur observée x des variables X de la manière suivante. Le domaine \mathcal{X} des variables X est partitionné en deux ensembles complémentaires,

$$\mathcal{X} = W + W^c .$$

La région W s'appelle *la région de rejet*, ou encore *la région critique*, de l'hypothèse (nulle), car si $x \in W$, on décide de la rejeter, ou encore que l'alternative est vraie. Si par contre $x \notin W$, on accepte l'hypothèse nulle. Le choix de W doit être fixé avant de mener des expériences.

La décision issue du test peut être erronée. Soit elle conduit à rejeter l'hypothèse nulle lorsqu'elle est vraie (erreur de première espèce, ou encore rejeter H_0 à tort), soit à l'accepter alors qu'elle est fautive (erreur de seconde espèce, ou encore accepter H_0 à tort). L'erreur de première espèce est une fonction croissante de la région critique W , tandis que celle de seconde espèce est décroissante. Il est donc impossible de minimiser ces deux erreurs simultanément (le risque zéro n'existe pas).

Une procédure de test consiste d'abord à positionner les hypothèses de manière à ce que l'erreur de première espèce soit maintenue à un niveau bas. On fixe alors un pourcentage faible α , généralement égal à 1%, 5% ou encore 10%. Et on impose la condition

$$\sup_{P \in H_0} P(X \in W) = \alpha .$$

Le nombre α s'appelle *le niveau* (de la procédure) du test. Notons que si $H_0 = \{P_0\}$ est simple, la condition devient simplement $P_0(X \in W) = \alpha$.

La procédure proprement dite est construite à l'aide d'une statistique $T(x)$, appelée *statistique du test*. La correspondance entre la région critique W et $T(x)$ revêt de l'une des trois formes suivantes :

- (1). $W = \{x : T(x) \geq d\}$,
- (2). $W = \{x : T(x) \leq c\}$,
- (3). $W = \{x : T(x) \notin [c, d]\}$.

Le test est *unilatère* dans les deux premiers cas, et *bilatère* dans le dernier cas.

Exemple : Test sur une moyenne. — La performance moyenne d'un médicament de référence est m_0 . Un laboratoire mène une expérience sur un nouveau médicament. Soient X_1, \dots, X_n ses performances relevées sur n patients.

En première approximation, on suppose que $X = (X_1, \dots, X_n)$ est un n -échantillon gaussien de moyenne m et de variance σ^2 . Il s'agit de tester l'hypothèse $H_0 : m = m_0$ contre l'alternative $H_1 : m > m_0$.

L'hypothèse nulle exprime le fait que le nouveau médicament a la même performance moyenne que le médicament de référence. Elle est multiple, associée à la famille des probabilités gaussiennes

$$\{P\} = \{\mathcal{N}(m_0, \sigma^2)^{\otimes n}, \sigma^2 > 0\} .$$

L'alternative décrit au contraire que le nouveau médicament est supérieur à celui de référence. Elle est également multiple, associée à la famille des probabilités gaussiennes

$$\{Q\} = \{\mathcal{N}(m, \sigma^2)^{\otimes n}, m \in]m_0, \infty[, \sigma^2 > 0\} .$$

On fixe $\alpha = 5\%$ et on utilise la statistique de Student comme statistique du test :

$$T_n = T_n(X) = \frac{\sqrt{n}}{S}(\bar{X} - m_0), \quad S = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 .$$

Pour toute probabilité $P \in H_0$, la loi de T_n est une loi de Student t_{n-1} à $n - 1$ degré de liberté. Soit q_α le quantile d'ordre $1 - \alpha$ de t_{n-1} . Alors

$$W = \{x \in \mathbb{R}^n : T_n(x) > q_\alpha\}$$

donne la procédure de test voulue.

5. Tests basés sur les rangs

5.1. Comparaison non-paramétrique de deux échantillons. — Soient deux échantillons X_1, \dots, X_{n_1} et Y_1, \dots, Y_{n_2} , indépendants entre eux et de loi respective \mathcal{L}_1 et \mathcal{L}_2 . On voudrait tester l'hypothèse $H_0 : \mathcal{L}_1 = \mathcal{L}_2$.

Si H_0 est vraie, l'échantillon combiné $X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}$ est un échantillon, de taille $n = n_1 + n_2$ de leur loi commune $\mathcal{L} = \mathcal{L}_1 = \mathcal{L}_2$. Notons $R_n = (R_{ni})$ le vecteur des rangs de l'échantillon combiné. Les statistiques

$$\sum_{i=1}^{n_1} R_{ni}, \quad \text{ou encore} \quad \sum_{i=1}^{n_1} \frac{R_{ni}}{n+1},$$

peuvent être utilisées pour tester la cohérence de l'hypothèse H_0 . En particulier le *test de Wilcoxon* s'appuie sur la statistique

$$W_n = \left(\frac{12n}{n_1 n_2} \right)^{1/2} \sum_{i=1}^{n_1} \left(\frac{R_{ni}}{n+1} - \frac{1}{2} \right).$$

Proposition 5.1. — (*Test de Wilcoxon*) Supposons $n_1 \wedge n_2 \rightarrow \infty$. Alors la statistique de Wilcoxon W_n converge en loi vers $\mathcal{N}(0, 1)$.

Démonstration. — On applique le théorème 3.6 avec

$$c_{ni} = \left(\frac{n_2}{nn_1} \right)^{1/2}, \quad 1 \leq i \leq n_1, \quad c_{ni} = - \left(\frac{n_1}{nn_2} \right)^{1/2}, \quad n_1 + 1 \leq i \leq n_1 + n_2. \quad \blacksquare$$

Détaillons la procédure du test. Pour les faibles effectifs n_1 et n_2 , la loi exacte de W_n est tabulée. On utilise en revanche l'approximation par la loi normale en cas des effectifs élevés. Si le niveau du test est fixé à α , la région critique du test est alors de la forme $A = \{|W_n| > q_\alpha\}$ où q_α est le quantile d'ordre $(1 - \alpha/2)$ de la loi de W_n . \square

5.2. Statistique des rangs pour indépendance. — On considère un couple de v.a. (X, Y) de loi \mathcal{L} sur \mathbb{R}^2 et soit $(X_1, Y_1), \dots, (X_n, Y_n)$ un échantillon de cette loi. On suppose que les deux lois marginales de \mathcal{L} sont continues.

Soient $R_n = (R_{ni})$ le vecteur des rangs des (X_i) et $V_n = (V_{ni})$ celui des (Y_i) . Si X et Y sont indépendants, il en est de même entre R_n et V_n ; inversement si par exemple X et Y sont positivement corrélées, les deux vecteurs R_n et V_n ont tendance à devenir parallèles.

Pour tester l'hypothèse $H_0 : "X \text{ et } Y \text{ sont indépendants}"$, le test de Spearman propose la statistique qui suit, égale au coefficient de corrélation empirique entre R_n et V_n :

$$\begin{aligned} \rho_n &= \frac{\sum_i (R_{ni} - \bar{R}_n)(V_{ni} - \bar{V}_n)}{\{\sum_i (R_{ni} - \bar{R}_n)^2 \sum_i (V_{ni} - \bar{V}_n)^2\}^{1/2}} \\ &= \frac{12}{n(n-1)(n+1)} \left[\sum_{i=1}^n R_{ni} V_{ni} - \frac{1}{4} n(n+1)^2 \right]. \end{aligned}$$

La dernière relation vient du fait que la loi uniforme sur $\{1, \dots, n\}$ a pour moyenne $(n+1)/2$ et variance $(n^2 - 1)/12$.

Proposition 5.2. — Sous H_0 , $\sqrt{n}\rho_n$ converge en loi vers $\mathcal{N}(0, 1)$.

Démonstration. — Soit R_n^0 le vecteur rangs des (X_i) obtenu après avoir ordonné au préalable les couples (X_j, Y_j) selon l'ordre croissant des (Y_j) : $Y_1 \leq Y_2 \leq \dots \leq Y_n$. Autrement dit,

$$\begin{aligned} \begin{pmatrix} X_1, \dots, X_n \\ Y_1, \dots, Y_n \end{pmatrix} &\rightarrow \begin{pmatrix} \tilde{X}_1, \dots, \tilde{X}_n \\ Y_{(1)}, \dots, Y_{(n)} \end{pmatrix}, \\ \tilde{X}_i &= \tilde{X}_{(R_n^0(i))} = X_{(R_n^0(i))} \\ \tilde{X}_i &= X_{V_n^{-1}(i)}, \quad \text{par définition.} \end{aligned}$$

Il en résulte que $R_n^0 = R_n \circ V_n^{-1}$. Ainsi

$$\sum_{i=1}^n R_{ni} V_{ni} = \sum_{i=1}^n R_{nV_n^{-1}(i)} i = \sum_{i=1}^n R_{ni}^0 i.$$

On a alors

$$\rho_n = \frac{12}{n(n-1)} \sum_{i=1}^n \left(i - \frac{n+1}{2} \right) \frac{R_{ni}^0}{n+1},$$

et

$$\sqrt{n}\rho_n = \frac{(n^2-1)^{1/2}}{n-1} \times \frac{12}{\{n(n^2-1)\}^{1/2}} \sum_{i=1}^n \left(i - \frac{n+1}{2} \right) \frac{R_{ni}^0}{n+1}.$$

Admettons d'abord que sous l'hypothèse d'indépendance, R_n^0 a la même loi que R_n . Alors le deuxième facteur du membre de droite ci-dessus est une statistique de la forme de T_n du théorème 3.6 avec les coefficients

$$c_{ni} = (12)^{1/2} \{n(n^2-1)\}^{-1/2} \left(i - \frac{n+1}{2} \right).$$

On vérifie que ces coefficients satisfont aux conditions [C]; et la preuve est complète par application de ce théorème.

Preuve du point admis : pour toute fonction $\phi : \mathcal{T}_n \rightarrow \mathbb{R}$ nous avons

$$\begin{aligned} \mathbb{E}\phi(R_n^0) &= \mathbb{E}\phi(R_n \circ V_n^{-1}) = \sum_{\sigma} \mathbb{E} \left[\phi(R_n \circ \sigma) \mathbf{1}_{V_n^{-1}=\sigma} \right] \\ &= \sum_{\sigma} \mathbb{E} [\phi(R_n \circ \sigma)] \mathbb{E} \left[\mathbf{1}_{V_n^{-1}=\sigma} \right] = \mathbb{E}\phi(R_n). \end{aligned}$$

□