

# MIPS: Introduction à la statistique

Pr. Lacroix

SEATECH  
Université de Toulon

Promo 2020

## 1 Estimation

- Modèle paramétrique
- Estimateurs
- Gauss en dim  $> 1$
- Convergence
- Maximum de vraisemblance
- Information de Fisher

## 2 Intervalles de confiance

- Intervalle de confiance
- Fractile / Quantile
- Student,  $\chi^2$ , Fisher
- IC moyenne Gaussien
- IC variance Gaussien
- IC Asymptotique

## 3 Tests

- Test d'hypothèse
- Biais
- $p$ -valeur
- Convergence
- Test UPP
- Rapport de vraisemblance

## 4 Paramétriques

- Moyenne Gaussien 1
- Moyenne Gaussien 2
- Variance Gaussien
- Proportion

## 5 Chi-deux

- Distance
- $\chi^2$  d'adéquation
- Symétrie
- Indépendance
- Adéquation 1

## 6 Moyennes empiriques

- Glivenko-Cantelli
- Kolmogorov-Smirnov
- Adéquation 2
- KS-comparaison

## 7 Divers

- Rapport de cotes
- Comparaison 2
- Comparaison 3
- ANOVA

## 8 Régression linéaire

- Estimation
- Utilité des régresseurs
- Coefficient de détermination

- On suppose que l'on observe un échantillon  $X = (X_1, \dots, X_n)$  iid, et que la loi commune appartient à une famille de probabilités  $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$  avec  $\Theta \subset \mathbb{R}^d$ . C'est ce que l'on appelle un **modèle paramétrique** de la loi du caractère aléatoire.
- Lorsque l'on notera  $\mathbb{P}_\theta(X \in A)$  et  $\mathbb{E}_\theta(f(X))$ , on supposera implicitement que la loi commune est  $\mathbb{P}_\theta$ .
- Si la loi commune est discrète ( $X_i \in F$  fini ou dénombrable) alors on notera  $p(f, \theta) = \mathbb{P}_\theta(X_i = f)$ . Alors si  $f \in F^n$ ,

$$\mathbb{P}_\theta(X = f) = p_n(f, \theta) = \prod_{i=1}^n p(f_i, \theta).$$

- Si la loi commune a une densité notée  $p(s, \theta)$ , alors  $X$  en a une sous  $\mathbb{P}_\theta$ , notée  $p_n(x, \theta)$ , et

$$p_n(x, \theta) = \prod_{i=1}^n p(x_i, \theta).$$

## Définition (Vraisemblance et statistique)

- Pour  $x = (x_1, \dots, x_n)$  fixé, on appelle **vraisemblance de la réalisation  $x$**  l'application  $\theta \in \Theta \mapsto p_n(x, \theta)$ .
- Une **statistique** est une variable aléatoire  $S = f(X_1, \dots, X_n)$  qui ne dépend pas de  $\theta$ .

La démarche va consister à estimer  $g(\theta)$  à partir de l'observation  $x$  de  $X$  où  $g : \Theta \rightarrow \mathbb{R}^q$  avec  $q < d$ .... Cela nécessite quelques mises en place.

### Définition (Estimateur)

On appelle **estimateur de  $g(\theta)$**  toute statistique à valeurs dans  $g(\Theta)$ .

### Exemple

Pour le modèle de Bernoulli  $\mathcal{P} = \{\mathcal{B}(p) : p \in [0, 1]\}$ ,  $\bar{X}_n$  est un estimateur de  $p$ .

### Définition (Biais)

Un estimateur  $Z$  est dit **sans biais** si il est intégrable et si  $\mathbb{E}_\theta(Z) = g(\theta)$ ,  $\forall \theta \in \Theta$  (égalité prise au sens vectoriel si nécessaire).

On suppose dorénavant que  $\mathbb{E}_\theta(Z^2) < +\infty$  (ou que  $\forall i, Z_i \in \mathcal{L}^2(\mathbb{P}_\theta)$ ).

### Définition (Risque quadratique)

On suppose  $q = 1$ .

- Le **risque quadratique de l'estimateur**  $Z$  est la fonction

$$\theta \in \Theta \mapsto R(Z, \theta) = \mathbb{E}_\theta((Z - g(\theta))^2).$$

- L'estimateur  $Z_1$  est dit **préférable** à l'estimateur  $Z_2$  si

$$\forall \theta \in \Theta, R(Z_1, \theta) \leq R(Z_2, \theta).$$

## Proposition

*Le risque quadratique se décompose par l'équation*

$$R(Z, \theta) = (g(\theta) - \mathbb{E}_\theta(Z))^2 + \text{Var}_\theta(Z),$$

*en la somme d'un premier terme lié au biais, et d'un second terme de variance (dispersion de l'estimateur).*

## Définition

Un VA  $X : \Omega \rightarrow \mathbb{R}^d$  est dit *Gaussien* si toute combinaison linéaire de ses composantes suit une loi Gaussienne ou constante sur  $\mathbb{R}$ .

## Proposition

Un VA  $X$  est Gaussien ssi pour tout  $u \in \mathbb{R}^d$ ,

$\Phi_X(u) = e^{iu \cdot \mu - \frac{1}{2} u \cdot \Gamma u}$ , où  $\mu \in \mathbb{R}^d$  et  $\Gamma \in \mathbb{R}^{d \times d}$  est une matrice symétrique positive ( $\forall u \in \mathbb{R}^d, u \cdot \Gamma u \geq 0$ ).

Alors  $\mu = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_d))$  et  $\Gamma_{i,j} = \text{Cov}(X_i, X_j)$ .

## Définition

$X$ ,  $\mu$  et  $\Gamma$  tels que ci-dessus font de  $X$  un vecteur Gaussien de dimension  $d$ , d'espérance  $\mu$  et de matrice de covariance  $\Gamma$ . On note cette loi  $\mathcal{N}_d(\mu, \Gamma)$ .



Si  $\Gamma$  est symétrique positive alors il existe  $n$  et  $A \in \mathbb{R}^{d \times n}$  telle que  $A^t A = \Gamma$ . Alors si  $(G_1, \dots, G_n)$  est iid de loi commune  $\mathcal{N}(0, 1)$ ,  $\mu + A \cdot G$  suit une loi  $\mathcal{N}_d(\mu, \Gamma)$ .

Pour simuler un vecteur Gaussien, il suffit donc

- de calculer  $A$  (par exemple en diagonalisant  $\Gamma$ ,  $\Gamma = P D^t P$  pour une matrice orthogonale  $P$ , puis en prenant  $A = P \sqrt{D}$ ),
- tirer un vecteur iid  $G = (G_1, \dots, G_d)$  iid de loi  $\mathcal{N}(0, 1)$ ,
- renvoyer  $\mu + A \cdot G$ .

## Définition (Convergence d'estimateurs)

Soit  $(Z_n)_{n \geq 1}$  une suite d'estimateurs de  $((X_1, \dots, X_n))_{n \geq 1}$ .

- Elle est dite **convergente** si  $\forall \theta \in \Theta$ ,  $Z_n$  converge en probabilité vers  $g(\theta)$ , sous  $\mathbb{P}_\theta$ .
- Elle est dite **fortement convergente** si

$$\forall \theta \in \Theta, Z_n \xrightarrow[\mathbb{P}_\theta - p.s.]{} g(\theta).$$

- Elle est dite **asymptotiquement normale** si il existe une matrice de covariance  $\Sigma(\theta) \in \mathbb{R}^{q \times q}$  telle que sous  $\mathbb{P}_\theta$ ,  $\sqrt{n}(Z_n - g(\theta))$  converge en loi vers  $\mathcal{N}_q(0, \Sigma(\theta))$ .

## Exemple

- Pour le modèle de Bernoulli, avec  $g(p) = p$ ,  $\bar{X}_n$  est un estimateur fortement convergent sans biais et asymptotiquement normal de matrice de covariance  $p(1 - p)$ .
- Idem pour  $\{\mathcal{P}(\lambda) : \lambda > 0\}$  ou  $\{\mathcal{E}(\alpha) : \alpha > 0\}$  (la variance change).
- Idem pour  $\{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$ .

## Définition (Estimateur du Maximum de Vraisemblance)

*On suppose que pour toute réalisation  $x = (x_1, \dots, x_n)$  de l'échantillon  $X = (X_1, \dots, X_n)$ , il existe une unique valeur  $\theta_n(x) \in \Theta$  telle que  $p_n(x, \theta_n(x)) = \max_{\theta \in \Theta} p_n(x, \theta)$ .*

*Alors la statistique  $\hat{\theta}_n = \theta_n(X)$  est appelée **Estimateur du Maximum de Vraisemblance** de  $\theta$ .*

Le **Principe du Maximum de Vraisemblance** consiste à considérer que l'observation réalisée de  $X$ ,  $x$ , l'a été avec un maximum de vraisemblance, c'est à dire consiste à inférer que  $x$  ayant été observé, la vraie valeur de  $\theta$  est  $\theta_n(x) = \hat{\theta}(x)$ .

Ce principe repose sur une considération ordinaire : **ce qui se produit est le plus probable !**

On cherche en statistiques à “deviner” la loi de  $X_i$  à partir d'observations de  $X = (X_1, \dots, X_n)$ . Pour cela on fait parfois des hypothèses qui conduisent à travailler avec un modèle paramétrique, et l'observation réalisée conduit alors à construire, ou tenter de construire,  $\hat{\theta}$  si il existe.

Maximiser  $p_n(x, \theta)$  revient à maximiser  $\ln p_n(x, \theta)$  qui est une somme. On cherchera  $\hat{\theta}$  tel que

$$\frac{\partial}{\partial \theta} \ln p_n(x, \theta) = 0.$$

On notera  $l_n(x, \theta) = \ln(p_n(x, \theta))$ .

## Exemple

- Pour le modèle  $\{\mathcal{B}(p) : 0 \leq p \leq 1\}$ ,  $\bar{X}_n$  est l'EMV de  $p$  ;
- Pour  $\{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}\}$ ,  $\bar{X}_n$  est l'EMV pour  $\mu$ , et  $s_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$  est l'EMV pour  $\sigma^2$  ;
- Pour un modèle  $\{\mathcal{U}(a, b) : -\infty < a < b < +\infty\}$ , on montre que  $(\min_{1 \leq i \leq n} X_i, \max_{1 \leq i \leq n} X_i)$  est l'EMV pour  $(a, b)$ .

## Théorème (Information de Fisher)

*Si le modèle présente de bonnes propriétés de régularité, alors l'Estimateur de Maximum de Vraisemblance (EMV)  $\hat{\theta}$  est fortement convergent et asymptotiquement normal de variance asymptotique  $I^{-1}(\theta)$  où*

$$I(\theta) = \mathbb{E}_{\theta}(\nabla_{\theta} l(X_1, \theta) \nabla_{\theta}^* l(X_1, \theta))$$

**s'appelle l'information de Fisher**

$$(\forall 1 \leq i, j \leq d, I_{i,j}(\theta) = \mathbb{E}_{\theta}(\frac{\partial}{\partial \theta_i} l(X_1, \theta) \frac{\partial}{\partial \theta_j} l(X_1, \theta))).$$

## Définition

Soit  $\alpha \in ]0, 1[$ . On dit qu'un intervalle  $I(X_1, \dots, X_n)$  est un **intervalle de confiance pour  $g(\theta)$  de niveau  $1 - \alpha$**  si

$$\forall \theta \in \Theta, \mathbb{P}_\theta(g(\theta) \in I(X_1, \dots, X_n)) = 1 - \alpha.$$

Si  $\forall \theta \in \Theta, \mathbb{P}_\theta(g(\theta) \in I(X_1, \dots, X_n)) \geq 1 - \alpha$ , on parle d'**intervalle de confiance de niveau  $1 - \alpha$  par excès**.

On choisit souvent les niveaux de confiance 90%, 95%, 99%. On cherche, pour un niveau de confiance donnée, à construire un intervalle de confiance de longueur minimale, afin de localiser au mieux la valeur  $g(\theta)$ .



## Définition

Soit  $Y$  une VAR de FR  $F_Y$ . Pour  $r \in ]0, 1[$ , on appelle **quantile** (ou **fractile**) d'ordre  $r$  de la loi de  $Y$  le nombre

$$q_r = \inf\{y \in \mathbb{R} : F(y) \geq r\}.$$

Si la loi de  $Y$  possède une densité strictement positive sur un intervalle et nulle à l'extérieur, alors  $F_Y$  est inversible et  $q_r = F_Y^{-1}(r)$ .

### Définition (Loi du Chi-deux)

*Si  $(X_1, \dots, X_n)$  sont iid de loi  $\mathcal{N}(0, 1)$ , alors  $X_1^2 + \dots + X_n^2$  suit une loi  $\chi^2(n)$ , dite loi du Chi-deux à  $n$  degrés de liberté.*

*Ces lois sont tabulées et de densité connues.*

### Définition (Student)

*Si  $Y$  suit une loi  $\mathcal{N}(0, 1)$  et si  $Z$  suit une loi  $\chi^2(n)$ , si  $Y$  et  $Z$  sont indépendantes, alors la loi de  $Y/Z$  est une loi de Student à  $n$  degrés de liberté, notée  $\mathcal{T}(n)$ . Sa densité est connue et elle est tabulée.*

### Définition (Fisher ou Fisher-Snedecor)

*La loi de Fisher à  $n$  et  $m$  degrés de liberté, est la loi de  $U/V$  où  $U$  et  $V$  de deux variables indépendantes de lois respectives  $\chi^2(n)$  et  $\chi^2(m)$ . On la note  $\mathcal{F}(n, m)$ . Sa densité est connue et elle est tabulée.*

Considérons le modèle Gaussien :  $\mathbb{P}_{(\mu, \sigma^2)} = \mathcal{N}(\mu, \sigma^2)$ . Alors  $\sqrt{n} \frac{\bar{X}_n - \mu}{S_n}$ , où  $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ , suit une loi  $\mathcal{T}(n-1)$ . Il s'ensuit que

$$\left[ \bar{X}_n - \frac{t_{n-1, 1-\frac{\alpha}{2}} S_n}{\sqrt{n}}, \bar{X}_n + \frac{t_{n-1, 1-\frac{\alpha}{2}} S_n}{\sqrt{n}} \right]$$

où  $t_{n-1, 1-\frac{\alpha}{2}}$  est le fractile d'ordre  $1 - \frac{\alpha}{2}$  de  $\mathcal{T}(n-1)$ , est un intervalle de confiance de niveau  $1 - \alpha$  pour  $\mu$ .

Toujours pour le modèle Gaussien,  $\frac{(n-1)S_n^2}{\sigma^2}$  suit une loi  $\chi^2(n-1)$ ,  
et donc

$$\left[ \frac{(n-1)S_n^2}{\chi_{n-1, 1-\frac{\alpha}{2}}^2}, \frac{(n-1)S_n^2}{\chi_{n-1, \frac{\alpha}{2}}^2} \right]$$

est un intervalle de confiance de niveau  $1 - \alpha$  pour la variance  $\sigma^2$ ,  
où  $\chi_{n-1, 1-\frac{\alpha}{2}}^2$  et  $\chi_{n-1, \frac{\alpha}{2}}^2$  désignent les fractiles appropriés de  
 $\chi^2(n-1)$ .

## Définition

*On appelle intervalle de confiance asymptotique pour  $g(\theta)$  de niveau  $1 - \alpha$  une suite d'intervalles  $(I_n(X_1, \dots, X_n))_n$  telle que*

$$\forall \theta \in \Theta, \lim_{n \rightarrow +\infty} \mathbb{P}_\theta(g(\theta) \in I_n(X_1, \dots, X_n)) = 1 - \alpha.$$

## Exemple

Dans le modèle de Bernoulli  $\{\mathcal{B}(p) : 0 < p < 1\}$ , si  $\phi_r$  désigne le quantile d'ordre  $r$  de  $\mathcal{N}(0, 1)$ ,

$$\left[ \bar{X}_n - \phi_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{X}_n(1 - \bar{X}_n)}{n}}, \bar{X}_n + \phi_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{X}_n(1 - \bar{X}_n)}{n}} \right]$$

est un intervalle de confiance asymptotique de  $p$  de niveau  $1 - \alpha$ . Comparer l'amplitude d'un intervalle de confiance de  $p$  de niveau 95% associé à un sondage donnant un cote de popularité de 37% à un politique, selon que le nombre de personnes interrogées est 100 ou 1000.

Nous noterons  $(H_0, H_1)$  une partition de l'espace des paramètres  $\Theta$  du modèle  $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$ . On suppose que l'échantillon statistique  $X = (X_1, \dots, X_n)$  parcourt un ensemble de la forme  $\mathbb{R}^{p \times n}$ .

### Définition

- On appelle **test d'hypothèse** une règle  $T : \mathbb{R}^{p \times n} \rightarrow \{0, 1\}$  de décision, binaire, qui étant donnée l'observation  $x$  de l'échantillon statistique  $X$  permet de décider si  $\theta \in H_0$ , l'**hypothèse nulle**, ou si  $\theta \in H_1$ , **hypothèse alternative**.
- Le test est déterminé par sa **région critique**  $W = T^{-1}(\{1\})$ , qui est l'ensemble des  $x$  conduisant selon  $T$  au rejet de  $H_0$ . Ainsi, si  $x \notin W$ , on accepte l'hypothèse  $H_0$ .

## Définition

- **L'erreur de première espèce** est le rejet de  $H_0$  alors qu'elle est vraie. Elle est quantifiée par le **risque de première espèce**  $\theta \in H_0 \mapsto \mathbb{P}_\theta(X \in W)$ .
- **L'erreur de seconde espèce** est le rejet de  $H_1$  alors qu'elle est vraie. Elle est quantifiée par le **risque de seconde espèce**  $\theta \in H_1 \mapsto \mathbb{P}_\theta(X \in W^c) = 1 - \rho_T(\theta)$ ,  $\rho_T(\theta)$  étant appelée la **puissance du test**.



On aura tendance à minimiser le risque de première espèce, qui représente les chances que l'on a de rejeter  $H_0$  à tort. On préfère donc accepter une hypothèse fausse que d'en rejeter une vraie, ce qui se conçoit dans la mesure où des tests complémentaires peuvent toujours venir compléter les premiers.

### Définition

On appelle **niveau du test** la quantité  $\alpha_T = \sup_{\theta \in H_0} \mathbb{P}_\theta(X \in W)$ .

Parmi les tests de niveau inférieur à un seuil  $\alpha$  fixé (en général 10, 5, 1%), on cherchera à minimiser le risque de seconde espèce, soit maximiser la puissance.

Si la puissance d'un test  $T$  devient moindre que son niveau, alors la probabilité de rejeter  $H_0$  à raison devient plus petite que la probabilité de la rejeter à tort. La règle de décision associée est alors incohérente.

Pour cette raison on préférera éviter cette situation.

### Définition (Biais)

*Un test de niveau  $\alpha$  est dit sans biais si sa puissance est supérieure à  $\alpha$ .*

En général, le test  $T$  sera construit à l'aide d'une statistique  $\tau_n = f(X)$ , la règle de décision sera alors de la forme

$\tau_n^{obs} \in \tilde{W}^c \Rightarrow H_0$  acceptée, où  $\tau_n^{obs} = f(x_1, \dots, x_n)$ . On supposera que l'on connaît la loi de  $\tau_n$  sous  $H_0$ .

Un test peut être

- **bilatéral** si  $\text{Rejet de } H_0 \Leftrightarrow \tau_n \notin [l, r]$  ;
- **unilatéral** si  $\text{Rejet de } H_0 \Leftrightarrow \tau_n^{obs} > r$  ou  
 $\text{Rejet de } H_0 \Leftrightarrow \tau_n^{obs} < l$  ;

## Définition

*Supposons que pour tout  $\alpha \in ]0, 1[$  on dispose d'un test  $T_\alpha$  de niveau  $\alpha$  de l'hypothèse nulle  $H_0$ . La **p-valeur** associée à cette famille de tests et à l'observation  $x$  est*

$$\hat{\alpha}(x) = \sup\{\alpha \in ]0, 1[: T_\alpha(x) = 0\}.$$

*Autrement dit,  $\hat{\alpha}(x)$  est le plus haut seuil autorisant l'acceptation de  $H_0$  compte tenu de l'observation  $x$ . Ainsi plus la p-valeur est faible, et moins les données confortent  $H_0$  (car plus l'acceptation de  $H_0$  au vu de ces données requiert une large région d'acceptation).*

La  $p$ -valeur indique donc un seuil limite et donne une indication sur la significativité du test à un seuil donné : plus faible sera la  $p$ -valeur, moins significatif sera le test.

Le cas typique est celui où la région de rejet du test  $T_\alpha$  est croissante en  $\alpha$ . Dans ce cas  $T_\alpha(x)$  accepte  $H_0$  tant que  $\alpha < \hat{\alpha}(x)$ .

## Définition

Soit  $(T_n)_{n \geq 1}$  une suite de tests de régions critiques  $(W_n)_{n \geq 1}$ . On dit que

- la suite de tests est **convergente** si  
 $\forall \theta \in H_1, \lim_n \mathbb{P}_\theta(W_n) = 1$  (la puissance tend vers 1);
- la suite est **de niveau asymptotique**  $\alpha$  si

$$\lim_n \sup_{\theta \in H_0} \mathbb{P}_\theta(W_n) = \alpha.$$

### Définition (Test UPP)

Soient  $T$  et  $T'$  deux tests de niveau  $\alpha$  des hypothèses  $H_0$  et  $H_1$  pour une observation  $x$  de  $X$ . On dit que  $T$  est **uniformément plus puissant que**  $T'$  si

$$\forall \theta \in \Theta_1, \rho_T(\theta) \geq \rho_{T'}(\theta).$$

Un test est dit  $UPP(\alpha)$  s'il est uniformément plus puissant que tout test de niveau  $\alpha$ .

Un test  $UPP$  est nécessairement sans biais.

Il existe une méthode générale pour construire des tests  $UPP(\alpha)$ .  
La statistique utilisée pour sa construction est

$$h(x) = \frac{\sup_{\theta \in \Theta_1} p_n(x, \theta)}{\sup_{\theta \in \Theta_0} p_n(x, \theta)}.$$

En général le test s'écrit  $T(x) = \mathbf{1}_{h(x) > k_\alpha}$ .

### Théorème (de Neyman-Pearson)

*Soient  $\theta_0, \theta_1 \in \Theta$  et testons  $H_0 : \theta = \theta_0$  contre  $H_1 : \theta = \theta_1$ . Soit  $\alpha \in ]0, 1[$ . Soit  $h(x, \theta_0, \theta_1) = p_n(x, \theta_1)/p_n(x, \theta_0)$ . S'il existe  $k_\alpha$  tel que*

$$\mathbb{P}_{\theta_0}(\mathbf{1}_{h(x, \theta_0, \theta_1) > k_\alpha} = 1) = \alpha,$$

*alors le test  $T(x) = \mathbf{1}_{h(x, \theta_0, \theta_1) > k_\alpha}$  est  $UPP(\alpha)$ .*

### Proposition (Moyenne Gaussien $\sigma^2$ connue)

*Dans le modèle Gaussien  $\mathcal{P} = \{\mathcal{N}(\mu, \sigma_0^2) : \mu \in \mathbb{R}\}$  à  $\sigma_0^2$  connu, on souhaite tester  $H_0 : \mu = \mu_0$  contre  $H_1 : \mu \neq \mu_0$ . On choisit une région critique de la forme  $W = \{|\sqrt{n} \frac{\bar{x}_n - \mu_0}{\sigma_0}| > a_{1-\alpha/2}\}$  où  $a_{1-\alpha/2}$  est le fractile d'ordre  $1 - \alpha/2$  de  $\mathcal{N}(0, 1)$ .*

*La  $p$ -valeur est  $\mathbb{P}(|Z| > |\sqrt{n} \frac{\bar{x}_n - \mu_0}{\sigma_0}|)$  si  $\mathbb{P}_Z = \mathcal{N}(0, 1)$ .*

*Ce test est  $UPP(\alpha)$  sans biais convergent.*



### Proposition (Moyenne Gaussien $\sigma^2$ inconnue)

Dans le modèle Gaussien  $\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$ , on veut tester  $H_0 : \mu = \mu_0$  contre  $H_1 : \mu \neq \mu_0$  au niveau  $\alpha \in ]0, 1[$ . Alors si  $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ ,  $\zeta_n = \sqrt{n} \frac{\bar{X}_n - \mu_0}{S_n}$  suit une loi  $\mathcal{T}(n-1)$  sous  $H_0$ .

On choisit  $W_n = \{|\zeta_n| > t_{n-1, 1-\frac{\alpha}{2}}\}$ . Alors

$$\forall \sigma^2 > 0, \mathbb{P}_{(\mu_0, \sigma^2)}(W_n) = \alpha.$$

La  $p$ -valeur est  $\mathbb{P}(|T| > |\zeta_n^{obs}|)$  où  $\mathbb{P}_T = \mathcal{T}(n-1)$ .

Ce test est convergent UPP( $\alpha$ ) sans biais.

### Proposition ( $\sigma^2$ Gaussien)

Dans le modèle Gaussien  $\mathcal{P} = \{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$ , pour  $\sigma_0^2 > 0$ , on veut tester  $H_0 : \sigma^2 \geq \sigma_0^2$  contre  $H_1 : \sigma^2 < \sigma_0^2$  au niveau  $\alpha \in ]0, 1[$ . Alors si  $\tau_n = \frac{(n-1)S_n^2}{\sigma_0^2}$  suit une loi  $\chi^2(n-1)$  sous  $H_0$ .

On choisit  $W_n = \{\tau_n \leq \chi_{n-1, \alpha}^2\}$  où  $\chi_{n-1, \alpha}^2$  est le fractile d'ordre  $\alpha$  d'une loi  $\chi^2(n-1)$ .

La  $p$ -valeur est  $\mathbb{P}(C \leq \tau_n^{obs})$  où  $\mathbb{P}_C = \chi^2(n-1)$ .

Ce test est convergent  $UPP(\alpha)$  sans biais.

On suppose que  $\mathbb{P}_X = \mathcal{B}(n, p)$ .

### Proposition (Proportion)

*Sous l'hypothèse  $np_0$  et  $n(1 - p_0) \geq 10$ , on fait une approximation normale de  $\mathcal{B}(n, p)$  pour tester  $H_0 : p = p_0$  contre  $H_1 : p \neq p_0$ . Alors si  $a_{1-\frac{\alpha}{2}}$  désigne le fractile d'ordre  $1 - \frac{\alpha}{2}$  de  $\mathcal{N}(0, 1)$ , la région de rejet est définie par*

$$W = \left\{ x : \left| \frac{\sqrt{n}(\bar{x}_n - p_0)}{\sqrt{p_0(1 - p_0)}} \right| > a_{1-\frac{\alpha}{2}} \right\}.$$

### Définition ("Distance" du Chi-deux)

*Soient  $p$  et  $q$  deux vecteurs de probabilité de dimension  $d$ . Alors*

$$d_{\chi^2}(p, q) = \sum_{i=1}^d \frac{(p_i - q_i)^2}{q_i}.$$

*Remarquer que  $d_{\chi^2}(p, q) \neq d_{\chi^2}(q, p)$ .*

## Proposition ( $\chi^2$ d'adéquation cas discret)

On suppose la loi de  $X_i$  discrète, pouvant prendre  $d$  valeurs  $v_1, \dots, v_d$ . Le modèle paramétrique est

$\mathcal{P} = \{\mathbf{p} = (p_1, \dots, p_d) : 0 \leq p_i \leq 1; \sum p_i = 1\}$ . On veut tester  $H_0 : \mathbf{p} = \mathbf{p}^0$  contre  $H_1 : \mathbf{p} \neq \mathbf{p}^0$ . On note  $n_i = \text{Card}\{1 \leq j \leq n : x_j = v_i\}$ , et  $p_i^{\text{obs}} = \frac{n_i}{n}$ . On pose  $\mathbf{p}^{\text{obs}} = (p_i^{\text{obs}})_{1 \leq i \leq d}$ . Alors

- Si  $p_i > 0$ ,  $U_n = \sqrt{n} \left( \frac{p_1^{\text{obs}} - p_1^0}{\sqrt{p_1^0}}, \dots, \frac{p_d^{\text{obs}} - p_d^0}{\sqrt{p_d^0}} \right) \xrightarrow{w^*} \mathcal{N}_d \left( 0, I_d - {}^t \sqrt{\mathbf{p}^0} \sqrt{\mathbf{p}^0} \right)$ ;
- Sous  $H_0$ , si  $p_i^0 > 0$ ,  $nd_{\chi^2}(\mathbf{p}^{\text{obs}}, \mathbf{p}^0) \xrightarrow{w^*} \chi^2(d-1)$  ;
- Sous  $H_1$ ,  $nd_{\chi^2}(\mathbf{p}^{\text{obs}}, \mathbf{p}^0) \xrightarrow[\mathbb{P}]{} +\infty$ .

On obtient donc un test asymptotique de niveau  $\alpha$  :

Choisir  $H_0$  si  $nd_{\chi^2}(\mathbf{p}^{\text{obs}}, \mathbf{p}^0) \leq \chi_{d-1, 1-\alpha}^2$ .

Ce test est convergent sans biais.

On suppose que  $(X_k, Y_k)$  est un couple de VA à valeurs dans un ensemble fini  $I$ . On note  $p_{i,j} = \mathbb{P}((X_k, Y_k) = (i, j))$ ,  $(i, j) \in I^2$ . On échantillonne  $(X, Y) = ((X_k, Y_k))_{1 \leq k \leq n}$ . On note  $N_{i,j} = \text{Card}\{1 \leq k \leq n : (X_k, Y_k) = (i, j)\}$ . On veut tester l'hypothèse  $H_0 : \mathbb{P}_{(X_k, Y_k)} = \mathbb{P}_{(Y_k, X_k)}$  contre  $H_1 : \mathbb{P}_{(X_k, Y_k)} \neq \mathbb{P}_{(Y_k, X_k)}$

### Définition (Hypothèse composite)

*Une hypothèse est **composite** lorsque sous celle-ci, la loi de la statistique du test n'est pas entièrement connue. Sinon elle est dite **simple**.*

Dans notre exemple,  $H_0$  est composite, et on montre que l'EMV de  $p_{i,j}$  est  $\tilde{p}_{i,j} = (N_{i,j} + N_{j,i})/2n$ .  
Pour rendre  $H_0$  simple il faut estimer  $(\#I - 1)^2/2 + \#I - 1$  paramètres.

## Théorème (Test de symétrie)

On note  $p_{i,j}^{obs} = N_{i,j}/n$ . On note  $\#I = \text{Card}I$ .

- Sous  $H_0$ ,  $nd_{\chi^2}(\mathbf{p}^{obs}, \tilde{\mathbf{p}}) \xrightarrow{w^*} \chi^2\left(\frac{(\#I^2-1)}{2}\right)$ .
- Sous  $H_1$ ,  $nd_{\chi^2}(\mathbf{p}^{obs}, \tilde{\mathbf{p}}) \xrightarrow{\mathbb{P}} +\infty$ .
- Si  $\chi^2_{(\#I^2-1)/2, 1-\alpha}$  désigne le fractile d'ordre  $1 - \alpha$  de  $\chi^2((\#I^2 - 1)/2)$ , alors on en déduit un test qui accepte  $H_0$  ssi  $nd_{\chi^2}(\mathbf{p}^{obs}, \tilde{\mathbf{p}}) \leq \chi^2_{(\#I^2-1)/2, 1-\alpha}$ .

Dans la pratique, on demande  $np_{i,j}^{obs} \geq 5 \ \forall (i,j)$ , sinon on regroupe des classes.

Ce test est de niveau asymptotique  $\alpha$  et convergent.

On échantillonne deux caractères aléatoires  $((X_k, Y_k)_{1 \leq k \leq n})$  et on veut tester  $H_0 : X_k \perp\!\!\!\perp Y_k$  contre  $H_1 : X_k \not\perp\!\!\!\perp Y_k$ . On suppose que les valeurs de  $X_k$  et  $Y_k$  sont regroupées par classes,  $(C_i)_{1 \leq i \leq I}$  pour  $X_k$ , et  $(D_j)_{1 \leq j \leq J}$  pour  $Y_k$ .

Si  $p_{i,j} = \mathbb{P}_{(X_k, Y_k)}(C_i \times D_j)$ ,  $p_{i,\cdot} = \mathbb{P}_{X_k}(C_i)$  et  $p_{\cdot,j} = \mathbb{P}_{Y_k}(D_j)$ , alors sous  $H_0$ ,  $p_{i,j} = p_{i,\cdot} p_{\cdot,j}$ .

Par ailleurs, si  $N_{i,j} = \text{Card}\{k : (X_k, Y_k) \in C_i \times D_j\}$ , alors

- $N_{i,\cdot} = \text{Card}\{k : X_k \in C_i\} = \sum_j N_{i,j}$  ;
- $N_{\cdot,j} = \text{Card}\{k : Y_k \in D_j\} = \sum_i N_{i,j}$  ;
- $p_{i,j}^{obs} = N_{i,j}/n$  ;
- l'EMV de  $p_{i,j}$  est  $N_{i,\cdot} N_{\cdot,j} / n^2 = \tilde{p}_{i,j}$ .

L'hypothèse  $H_0$  est ici composite et pour la rendre simple il faut estimer  $I - 1 + J - 1$  paramètres.



## Théorème (Test d'indépendance)

Si  $p_{i,j} > 0 \forall (i,j)$ , alors

- Sous  $H_0$ ,  $nd_{\chi^2}(\mathbf{p}^{obs}, \tilde{\mathbf{p}}) \xrightarrow{w^*} \chi^2((I-1)(J-1))$ .
- Sous  $H_1$ ,  $nd_{\chi^2}(\mathbf{p}^{obs}, \tilde{\mathbf{p}}) \xrightarrow{\mathbb{P}} +\infty$ .
- Si  $\chi^2_{(I-1)(J-1), 1-\alpha}$  désigne le fractile d'ordre  $1-\alpha$  de  $\chi^2((I-1)(J-1))$ , on en déduit un test qui accepte  $H_0$  ssi  $nd_{\chi^2}(\mathbf{p}^{obs}, \tilde{\mathbf{p}}) \leq \chi^2_{(I-1)(J-1), 1-\alpha}$ .

Dans la pratique, on demande  $np_{i,j}^{obs} \geq 5 \forall (i,j)$ , sinon on regroupe des classes.

Ce test est de niveau asymptotique  $\alpha$  et convergent.

On suppose que  $\mathbb{P}_{X_k} \in \mathcal{P}$  un modèle paramétrique. On veut tester  $H_0 : \mathbb{P}_{X_k} = \mathbb{P}_0$  contre  $H_1 : \mathbb{P}_{X_k} \neq \mathbb{P}_0$ .

On partitionne l'image de  $X_k$  en classes  $C_1, \dots, C_c$ . On note  $N_t = \text{Card}\{k : X_k \in C_t\}$ ,  $1 \leq t \leq c$ , et  $p_t^{\text{obs}} = N_t/n$ . On note  $p_t = \mathbb{P}_0(C_t)$ .

L'hypothèse  $H_0$  est ici composite et pour la rendre simple il nous faut estimer  $d$  paramètres.

### Théorème (Adéquation en loi par le Chi-deux)

Si  $c - d - 1 > 0$  et  $p_t > 0$  pour tout  $t$ ,

- Sous  $H_0$ ,  $nd_{\chi^2}(\mathbf{p}^{\text{obs}}, \mathbf{p}) \xrightarrow{w^*} \chi^2(c - d - 1)$ .
- Sous  $H_1$ ,  $nd_{\chi^2}(\mathbf{p}^{\text{obs}}, \mathbf{p}) \xrightarrow{\mathbb{P}} +\infty$ .
- Si  $\chi_{c-d-1, 1-\alpha}^2$  désigne le fractile d'ordre  $1 - \alpha$  de  $\chi^2(c - d - 1)$ , on en déduit un test qui accepte  $H_0$  ssi  $nd_{\chi^2}(\mathbf{p}^{\text{obs}}, \mathbf{p}) \leq \chi_{c-d-1, 1-\alpha}^2$ .

À une suite iid  $(X_i)_{i \geq 1}$  de VARs, on peut associer les Diracs aléatoires  $(\delta_{X_i})_{i \geq 1}$  puis des mesures dites empiriques

$$\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \quad \mathbb{P}_n(\omega) = \frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)}.$$

### Théorème (de Glivenko-Cantelli)

Soit  $F = F_{X_k}$  et  $\forall \omega \in \Omega$ ,  $F_n(\omega, \cdot)$  la FR de  $\mathbb{P}_n(\omega)$ . Alors

$$\sup_{t \in \mathbb{R}} |F_n(\omega, t) - F(t)| \xrightarrow[\mathbb{P}\text{-p.s.}]{} 0.$$

## Théorème (de Kolmogorov-Smirnov)

Si  $F$  est continue sur  $\mathbb{R}$ , alors

$$K_n = \sqrt{n} \sup_{t \in \mathbb{R}} \| F_n(\cdot, t) - F(t) \|_{\infty} \xrightarrow{w^*} \mu_{KS}$$

où

- $\| f \|_{\infty} = \inf \{ M > 0 : \mathbb{P}(\{\omega : |f(\omega)| > M\}) = 0 \} ;$
- $\mu_{KS}$  est une loi de probabilités sur  $\mathbb{R}$  de FR  $F_{KS}$  telle que

$$F_{KS}(t) = \mathbf{1}_{[0, +\infty[}(t) \left( 1 + 2 \sum_{k \geq 1} (-1)^k e^{-k^2 t^2} \right).$$

On déduit du théorème précédent un test d'adéquation en loi : on va tester  $H_0 : \mathbb{P}_{X_k} = \mathbb{P}_0$  contre  $H_1 : \mathbb{P}_{X_k} \neq \mathbb{P}_0$ . On note  $F$  la FR de  $\mathbb{P}_0$  que l'on suppose continue. On notera  $k_{1-\alpha}$  le fractile d'ordre  $1 - \alpha$  de  $\mu_{KS}$ .

### Proposition (Test d'adéquation en loi de Kolmogorov-Smirnov)

*On note  $(X_{(i)})_{1 \leq i \leq n}$  l'échantillon réordonné ( $X_{(i)} \leq X_{(i+1)}$ ), commande sort en informatique). Alors*

$$K_n = \sqrt{n} \max_{1 \leq i \leq n} \left\{ \left| F(X_{(i)}) - \frac{i}{n} \right|; \left| F(X_{(i)}) - \frac{i-1}{n} \right| \right\}.$$

*On accepte  $H_0$  si  $K_n \leq k_{1-\alpha}$ , sinon on la rejette.  
ce test est de niveau asymptotique  $\alpha$  et convergent.*

On se donne deux échantillons  $X = (X_1, \dots, X_n)$  et  $Y = (Y_1, \dots, Y_m)$  indépendants, et on suppose que leurs lois respectives ont des FR respectives  $F$  et  $G$  continues. On veut tester  $H_0 : F = G$  contre  $H_1 : F \neq G$ . On note  $F_n$  et  $G_m$  les FR empiriques associées.

### Théorème (Comparaison de lois a prioriées)

*Sous  $H_0$ ,*

$$K_{n,m} := \sqrt{\frac{nm}{n+m}} \sup_{t \in \mathbb{R}} |F_n(\cdot, t) - G_m(\cdot, t)| \xrightarrow{w^*} \mu_{KS}.$$

*On rejette  $H_0$  si  $K_{n,m} > k_{1-\alpha}$ .*

*Il s'agit d'un test de niveau asymptotique  $\alpha$  et convergent.*

Le test du rapport de cotes est très utilisé pour tester si un symptôme est révélateur d'une maladie. Il se propose de tester d'une façon particulière l'indépendance de deux caractères binaires  $M$  et  $S$  :  $M$  pour malade,  $\bar{M}$  pour sain ;  $S$  pour symptôme,  $\bar{S}$  pour absence de symptôme. Le but ici n'est pas de conclure à l'indépendance mais plutôt à la dépendance orientée en ce que l'on espère que le symptôme diagnostique favorablement la maladie. On se donne un échantillon  $((M_i, S_i)_{1 \leq i \leq n}$ . On notera

- $n_{MS} = \text{Card}\{i : (M_i, S_i) = (M, S)\}$  ;
- $n_{\bar{M}S} = \text{Card}\{i : (M_i, S_i) = (\bar{M}, S)\}$  ;
- $n_{M\bar{S}} = \text{Card}\{i : (M_i, S_i) = (M, \bar{S})\}$  ;
- $n_{\bar{M}\bar{S}} = \text{Card}\{i : (M_i, S_i) = (\bar{M}, \bar{S})\}$ .

Dans une enquête visant à tester l'efficacité d'un symptôme pour diagnostiquer une maladie (mais ce peut être celle d'un médicament pour tester sa capacité à guérir), deux cas de figure se distinguent :

- étude de “*cohorte*” : le nombre de malades et non malades est aléatoire, on contrôle simplement le nombre d'individus présentant le symptôme et le nombre de ceux ne le présentant pas ;
- enquête “*cas témoins*” : ici on fixe le nombre de malades et de non malades et c'est le nombre de ceux présentant le symptôme qui est aléatoire.

Le **risque absolu** chez les symptomatiques est le quotient

$$R1 = \frac{n_{MS}}{n_S}, \text{ et chez les asymptomatiques il est de } R0 = \frac{n_{M\bar{S}}}{n_{\bar{S}}}.$$

Ils ne sont déterminés que dans une étude de cohorte.

On peut calculer le **risque relatif**  $RR = \frac{R1}{R0}$  qui indique si le symptôme est favorable à la maladie ou pas ( $RR > 1$ ).



Si  $RR > 1$ , on peut faire un test du Chideux pour mesurer cette orientation en faisant un test d'indépendance de symptôme et maladie par un tableau de contingence.

Le **rapport de cotes** (*odds ratio*) est le rapport des cotes des risques absolus :  $RC = \frac{R1(1-R0)}{R0(1-R1)}$ . Si la maladie est rare, le

quotient  $(1 - R0)/(1 - R1)$  est proche de 1 et alors  $RC \sim RR$ .

Le **rapport de cote** (*odds ratio*) est alors

$$RC = \frac{\mathbb{P}(M|S)/\mathbb{P}(\bar{M}|S)}{\mathbb{P}(M|\bar{S})/\mathbb{P}(\bar{M}|\bar{S})} = \frac{\mathbb{P}(MS)\mathbb{P}(\bar{M}\bar{S})}{\mathbb{P}(M\bar{S})\mathbb{P}(\bar{M}S)}.$$

Noter que  $RC = 1$  si les deux caractères  $M$  et  $S$  sont indépendants.

On veut tester  $H_0 : M \amalg S$  contre  $H_1 : M \not\amalg S$ .

### Théorème (Test du rapport de cotes)

Pour une maladie rare ( $RC \sim RR$ ) le rapport de cotes est estimé par  $\hat{RC} = (n_{MS}n_{\bar{M}\bar{S}})/(n_{\bar{M}S}n_{M\bar{S}})$ . Si les effectifs observés sont nuls on leur ajoute à tous 0.5.

Si  $V = n_{MS}^{-1} + n_{\bar{M}\bar{S}}^{-1} + n_{\bar{M}S}^{-1} + n_{M\bar{S}}^{-1}$ , alors sous  $H_0$ ,

$$T = \frac{\ln(\hat{RC})}{\sqrt{V}} \xrightarrow{w^*} \mathcal{N}(0, 1).$$

Si  $a_{1-\frac{\alpha}{2}}$  désigne le fractile d'ordre  $1 - \frac{\alpha}{2}$  de  $\mathcal{N}(0, 1)$ , alors on rejette  $H_0$  si  $|T| \geq a_{1-\frac{\alpha}{2}}$ .

Ainsi un IC( $1 - \alpha$ ) asymptotique pour RR et RC est

$$\left[ \exp(\ln(\hat{RC}) \pm t_{1-\frac{\alpha}{2}} \sqrt{V}) \right].$$

On cherche à tester l'hypothèse  $H_0 : \mathbb{P}_{X_k} = \mathbb{P}_{Y_l}$  contre  $H_1 : \mathbb{P}_{X_k} \neq \mathbb{P}_{Y_l}$ . Ici on part de deux échantillons de VARs indépendants  $X = (X_k)_{1 \leq k \leq n}$  et  $Y = (Y_l)_{1 \leq l \leq m}$ .  
On réindexe les deux échantillons en ordre croissant :

$$Z_1 \leq \dots \leq Z_{n+m}.$$

On note  $W_x$  la somme des rangs des  $X_i$  dans ce réagencement.  
Le test de Wilcoxon repose sur le résultat de convergence suivant :

### Théorème (Test de Wilcoxon)

*Sous  $H_0$ ,*

$$\frac{W_x - n(n+m+1)/2}{\sqrt{nm(n+m+1)/12}} \xrightarrow[w^*]{n,m \rightarrow +\infty} \mathcal{N}(0,1).$$

*Le test qui en résulte est asymptotiquement de niveau  $\alpha$  et convergent.*

La loi de  $W_x$  est tabulée pour les petits échantillons et sinon le théorème précédent traite du cas des échantillons plus grands.

Le **test de Mann-Whitney** teste exactement le même chose que celui de Wilcoxon mais simplement repose sur l'observation que

$$U = \sum_{k=1}^n \sum_{l=1}^m \mathbf{1}_{X_k > Y_l} \Rightarrow W_x = U + \frac{n(n+1)}{2}.$$

Ici nous testons l'égalité des lois de deux échantillons Gaussiens indépendants de lois  $\mathcal{N}(\mu_x, \sigma_x^2)$  et  $\mathcal{N}(\mu_y, \sigma_y^2)$ . Les échantillons sont de tailles respectives  $n$  pour  $X$  et  $m$  pour  $Y$ .

Pour tester l'identité des lois on teste d'abord l'égalité des variances : tout repose sur l'observation de la page 20 :

$$\frac{(n-1)S_n^2}{\sigma^2} \text{ suit une loi } \chi^2(n-1) \implies$$

$$\frac{(n-1)S_{n,x}^2}{\sigma_x^2} \frac{\sigma_y^2}{(m-1)S_{m,y}^2} \text{ suit une loi } \mathcal{F}(n-1, m-1)$$

Donc sous  $H_0$ ,  $\frac{(n-1)S_{n,x}^2}{(m-1)S_{m,y}^2}$  suit une loi  $\mathcal{F}(n-1, m-1)$  d'où le test sur l'égalité des variances.

Si ce dernier est favorable alors on teste l'égalité des lois en terminant de tester l'égalité des moyennes grâce à l'observation suivante :

### Théorème (Égalité des moyennes Gaussien à variances égales)

$$T = \frac{\sqrt{n+m-2}(\bar{X}_n - \bar{Y}_m) - (\mu_x - \mu_y)}{\sqrt{\frac{1}{n} + \frac{1}{m}} \sqrt{(n-1)S_{n,x}^2 + (m-1)S_{m,y}^2}}$$

*suit une loi  $\mathcal{T}(n+m-2)$  si  $\sigma_x = \sigma_y$ .*

*Si  $n, m \rightarrow +\infty$ ,*

$$\frac{(\bar{X}_n - \bar{Y}_m) - (\mu_x - \mu_y)}{\sqrt{\frac{S_{n,x}^2}{m} + \frac{S_{m,y}^2}{n}}} \xrightarrow{w^*} \mathcal{N}(0, 1).$$

*Sous  $H_0$ ,  $\mu_x = \mu_y$  et on en déduit la fin du test.*

On suppose que l'on observe  $k$  groupes de tailles  $n_1, \dots, n_k$  d'individus sur lesquels on a échantillonné un caractère Gaussien  $X^{(h)}$  de loi  $\mathcal{N}(\mu_h, \sigma^2)$  pour le  $h^{\text{ième}}$  groupe. On suppose que  $\sigma^2$  ne dépend pas du groupe mais ceci pourrait faire l'objet d'un test préalable. On note  $n = \sum_i n_i$ .

### Définition

- On appelle **variance intra-classes** la moyenne des variances empiriques des  $k$  classes. On l'appelle aussi la **variance résiduelle**.
- On appelle **variance inter-classes** la variance des moyennes empiriques. On l'appelle aussi **variance expliquée**.

Nous noterons :

- $\bar{X}^{(h)}$  la moyenne empirique de la classe  $h$  ;
- $V^{(h)}$  la variance empirique de la classe  $h$  ;
- $\bar{X}$  la moyenne empirique de l'échantillon global ;
- $V_{intra} = \sum_h \frac{n_h}{n} V^{(h)}$  la variance intra-classes ;
- $V_{inter} = \sum_h \frac{n_h}{n} (\bar{X}^{(h)} - \bar{X})^2$  la variance inter-classes ;
- $S^2$  la variance de l'échantillon global.

Alors

$$S^2 = V_{intra} + V_{inter} .$$



On veut tester l'hypothèse nulle  $H_0 : \mu_1 = \dots = \mu_k$ . Si  $H_0$  est fausse, on s'attend à ce que  $V_{inter}$  soit prépondérante par rapport à  $V_{intra}$ .

### Proposition

*Sous  $H_0$ ,*

- $nV_{intra}/\sigma^2$  suit une loi  $\chi^2(n - k)$  ;
- $nV_{inter}/\sigma^2$  suit une loi  $\chi^2(k - 1)$ .

*Il s'ensuit que*

$$\left( \frac{V_{inter}}{k - 1} \right) \left( \frac{n - k}{V_{intra}} \right) \text{ suit une loi } \mathcal{F}(k - 1, n - k) .$$

*On en déduit un test dit "ANOVA" pour  $H_0$ .*

On cherche à étudier l'effet de certains facteurs explicatifs (les **régresseurs**) sur un phénomène observé. Nous noterons  $X_i$  le  $i^{\text{ème}}$  caractère observé et  $(r_i^j)_{1 \leq j \leq p}$  les  $p$  variables explicatives associées. On suppose que

$$X_i = \beta_0 + \sum_{j=1}^p \beta_j r_i^j + \varepsilon_i$$

où les  $\varepsilon_i$  sont des résidus aléatoires (iid de loi  $\mathcal{N}_1(0, \sigma^2)$ ).

On suppose que  $p \ll n$ . Si  $X = (X_1, \dots, X_n)^t$  et  $\varepsilon = (\varepsilon_i)_{1 \leq i \leq n}^t$  et  $\gamma = (\beta_0, \beta_1, \dots, \beta_p)^t$ , si de plus

$$M = \begin{pmatrix} 1 & r_1^1 & \dots & r_1^p \\ \vdots & \vdots & \vdots & \vdots \\ 1 & r_n^1 & \dots & r_n^p \end{pmatrix},$$

alors le **modèle régressif** s'écrit

$$X = M\gamma + \varepsilon.$$

Dans le modèle de régression, le paramètre  $\theta = (\gamma, \sigma^2)$  doit être estimé. Nous supposons que  $M$  est de rang  $p + 1$ .

La densité de  $\varepsilon$  vaut  $\frac{1}{(2\pi\sigma^2)^{n/2}} \exp(-\frac{1}{2\sigma^2} \sum_{i=1}^n y_i^2)$ , on en déduit que

$$l_n(x, (\gamma, \sigma^2)) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{|x - M\gamma|^2}{2\sigma^2}.$$

En utilisant le fait que  $M$  soit de rang maximal, on obtient que l'EMV pour  $(\gamma, \sigma^2)$  est (on prend la distance euclidienne)

$$(\gamma, \sigma^2) = ((M^t M)^{-1} M^t X, \frac{|X - \overbrace{M(M^t M)^{-1} M^t X}^{X_E}|^2}{n}).$$

### Proposition

*Le vecteur aléatoire  $\hat{\gamma} = (M^t M)^{-1} M^t X$  est un estimateur sans biais de  $\gamma$ . On a  $\mathbb{P}_{\hat{\gamma}} = \mathcal{N}_{p+1}(\gamma, \sigma^2 (M^t M)^{-1})$ , et ce vecteur est indépendant de la variable  $\frac{|X - X_E|^2}{\sigma^2}$  qui suit une loi  $\chi^2(n - (p + 1))$ .*

On en déduit des intervalles de confiance pour les paramètres  $\beta_{k-1}$  et  $\sigma^2$ .

### Corollaire

Si  $t_{n-(p+1),r}$  désigne le fractile d'ordre  $r$  de  $\mathcal{T}(n - (p + 1))$ , alors

$$\left[ \hat{\gamma}_k \pm t_{n-(p+1), 1-\frac{\alpha}{2}} \sqrt{\frac{(M^t M)_{k,k}^{-1} |X - X_E|^2}{n - (p + 1)}} \right]$$

est un intervalle de confiance de niveau  $1 - \alpha$  pour  $\beta_{k-1}$ . Par ailleurs,

$$\left[ \frac{|X - X_E|^2}{\chi_{n-(p+1), 1-\frac{\alpha}{2}}^2}, \frac{|X - X_E|^2}{\chi_{n-(p+1), \frac{\alpha}{2}}^2} \right],$$

où  $\chi_{n-(p+1),r}^2$  désigne le fractile d'ordre  $r$  de  $\chi^2(n - (p + 1))$ , est un intervalle de confiance de niveau  $1 - \alpha$  pour  $\sigma^2$ .

Quitte à les permuter, on veut tester l'utilité des  $p - q$  derniers régresseurs, c'est à dire l'hypothèse nulle

$$H_0 : \{\beta_j = 0 : q + 1 \leq j \leq p\} \text{ contre } H_1 : \exists q + 1 \leq j \leq p, \beta_j \neq 0.$$

On note  $x_H = M_0(M_0^t M_0)^{-1} M_0^t x$  la projection orthogonal du vecteur  $M_0^t x$  sur  $H = \{M_0 u : u \in \mathbb{R}^{q+1}\}$ , avec

$$M_0 = \begin{pmatrix} 1 & r_1^1 & \dots & r_1^q \\ \vdots & \vdots & \vdots & \vdots \\ 1 & r_n^1 & \dots & r_n^q \end{pmatrix}.$$

Alors sous  $H_0$

$$F = \frac{|X_E - X_H|^2 / (p - q)}{|X - X_E|^2 / (n - (p + 1))}$$

suit une loi  $\mathcal{F}_{p-q, n-(p+1)}$ . Si l'on désigne ses fractiles par  $\mathcal{F}_{p-q, n-(p+1), r}$ , alors la région de rejet de  $H_0$  de niveau  $\alpha$  et la  $p$ -valeur du test sont

$$W = \{F \geq \mathcal{F}_{p-q, n-(p+1), 1-\alpha}, \quad p = \mathbb{P}(Z \geq F^{obs}) \text{ si } \mathbb{P}_Z = \mathcal{F}_{p-q, n-(p+1)}.$$

Si l'on note  $X_{1_n} = \bar{X}_n(1, \dots, 1)^t \in \mathbb{R}^n$ , on appelle **coefficient de détermination** le nombre

$$R^2 = \frac{|X_E - X_{1_n}|^2}{|X - X_{1_n}|^2}.$$

Il est lié au test d'utilité de l'ensemble des régresseurs ( $q = p$ ) par la formule

$$F = \frac{R^2}{1 - R^2} \frac{n - (p + 1)}{p}.$$

Il est fourni par plusieurs logiciels de statistique.