

Exemples d'exercices de PC pour le cours MAP433

Olivier Cappé

Janvier 2008

Table des matières

1 Petite classe 1	3
1.1 Loi de Bernoulli et lois limites	3
1.2 Théorème de la limite centrale	3
1.3 Loi conditionnelle	4
1.4 Modèle probit	4
2 Petite classe 2	5
2.1 Information de Fisher	5
2.2 Modèle à deux échantillons	5
2.3 Données censurées	6
2.4 Modèle de mélange	6
3 Petite classe 3	7
3.1 Distribution du χ^2	7
3.2 Modèle linéaire gaussien	7
4 Petite classe 4	8
4.1 Divergence de Kullback-Leibler	8
4.2 Maximum de vraisemblance pour la loi exponentielle translatée	8
5 Petite classe 5	9
5.1 Test de dispersion	9
5.2 Intervalle de confiance pour la loi uniforme	9
5.3 Intervalle de confiance	9
5.4 Modèle linéaire gaussien : modèle électoral	9
6 Petite classe 6	11
6.1 « Méthode delta »	11
6.2 Estimateur de la variance empirique	12
6.3 Test unilatère	12
7 Petite classe 7	14
7.1 Maximum de vraisemblance et moindres carrés non linéaires dans le modèle de choix binaire	14
7.2 Analyse de probabilité	14

7.3	Test de fiabilité	15
7.4	Intervalle de confiance pour le modèle de Bernoulli	15
7.5	Test de Wald	15
7.6	Test d'hypothèse sur le paramètre d'une loi de Poisson	16
7.7	Test de significativité dans le modèle de régression linéaire	16
8	Petite classe 8	17
8.1	Test d'ajustement	17
8.2	Test de Wald dans une loi multinomiale	17
8.3	Test de significativité dans le modèle de régression logistique	18
9	Petite classe 9	19
9.1	Test d'adéquation du khi-deux	19
9.2	Test d'adéquation à une loi discrète	19
9.3	Test d'endogamie sociale	20
9.4	Test d'homogénéité de deux populations	20

1 Petite classe 1 : Modélisation statistique

AU PROGRAMME :

- RAPPELS DE PROBABILITÉS CONCERNANT LA CONVERGENCE PRESQUE SÛRE, LA CONVERGENCE EN PROBABILITÉ ET LA CONVERGENCE EN LOI, EXEMPLE DE THÉORÈME DE LA LIMITE CENTRALE MULTIVARIÉ BIEN VENU.
- RAPPELS DE PROBABILITÉS CONCERNANT LE CALCUL DE LOIS CONDITIONNELLES, L'ESPÉRANCE CONDITIONNELLE « À $X=x$ » ÉTANT DÉFINIE COMME L'ESPÉRANCE SOUS LA LOI CONDITIONNELLE ET $E(\cdot|X)$ COMME CETTE FONCTION APPLIQUÉE À LA VARIABLE X ; EXERCICES MONTRANT QUE $E_\theta [h(Y)] = E \{E_\theta [h(Y)|X]\}$ ET

$$V_\theta [h(Y)] = V \{E_\theta [h(Y)|X]\} + E \{V_\theta [h(Y)|X]\}$$

BIEN VENUS

1.1 Loi de Bernoulli et lois limites

Soit $(X_n)_n$ une séquence de variables aléatoires de loi binomiale $B(n, \lambda_n)$. Ceci signifie que la probabilité que k succès soient observés parmi n tirages avec remise est donnée par

$$P(X_n = k) = \binom{n}{k} \lambda_n^k (1 - \lambda_n)^{n-k}$$

1. Calculez $E X_n$ et $V X_n$.
2. Donner des conditions qui garantissent que X_n converge en probabilité vers 0.
3. On considère maintenant que $\lambda_n = \frac{\lambda}{n}$; calculez la limite lorsque $n \rightarrow +\infty$ de $P(X_n \leq a)$. En déduire que X_n converge en loi (On rappelle la formule de Stirling : $n! \approx e^{-n} n^n \sqrt{2\pi n}$).

1.2 Théorème de la limite centrale

Soit $(X_n)_n$ une suite de variables aléatoires indépendantes centrées de variance commune σ^2 . Soit

$$S_n = \frac{1}{\sigma\sqrt{n}} \sum_{j=1}^n X_j$$

Par le théorème de la limite centrale, cette variable converge en loi vers la loi normale centrée réduite, c'est à dire que $\lim_{n \rightarrow +\infty} E e^{itS_n} = e^{-\frac{t^2}{2}}$. L'objet de cet exercice est de montrer que S_n ne converge pas en probabilité.

1. Décomposer la variable S_{2n} en fonction de S_n et d'une variable aléatoire indépendante de la précédente.
2. Calculer la fonction caractéristique de $S_{2n} - S_n$ et montrer que cette différence converge en loi.
3. En raisonnant par l'absurde, en déduire que S_n ne converge pas en probabilité.

1.3 Loi conditionnelle

Soit X une variable aléatoire qui suit une loi gamma $(2, \lambda)$ de densité

$$f(x) = \lambda^2 x e^{-\lambda x} \mathbb{1}_{[0, +\infty)}(x)$$

et soit Y une variable aléatoire dont la loi conditionnelle à $X = x$ est uniforme sur $[0, x]$.

- Donner la loi jointe de (X, Y) .
- Donner la loi marginale de Y et montrer que Y est indépendant de $X - Y$.

1.4 Modèle probit

Nous disposons d'une information relative au comportement de remboursement ou de non-remboursement d'emprunteurs :

$$Y_i = \begin{cases} 1 & \text{si l'emprunteur } i \text{ rembourse} \\ 0 & \text{si l'emprunteur } i \text{ est défaillant} \end{cases}$$

Afin de modéliser ce phénomène, on suppose l'existence d'une variable aléatoire Y_i^* normale, d'espérance m et de variance σ^2 , que l'on appellera « capacité de remboursement de l'individu i », telle que :

$$Y_i = \begin{cases} 1 & \text{si } Y_i^* > 0 \\ 0 & \text{si } Y_i^* \leq 0 \end{cases}$$

On note Φ la fonction de répartition de la normale $\mathcal{N}(0, 1)$.

1. Exprimer la loi de Y_i en fonction de Φ .
2. Les paramètres m et σ^2 sont-ils identifiables ?

2 Petite classe 2 : Information et bornes d'estimation

AU PROGRAMME : PROPRIÉTÉS DE BASE DE L'INFORMATION DE FISHER (ADDITIONNÉ ETC.), AU MOINS UN EXEMPLE DE CALCUL DANS UN MODÈLE CONDITIONNEL ET DANS UN CAS OÙ LE PARAMÈTRE N'EST PAS SCALAIRE.

2.1 Information de Fisher

Écrire la vraisemblance pour un échantillon de n observations IID et calculer l'information de Fisher pour les lois suivantes :

1. Loi de poisson de paramètre λ ,

$$P(Y = k) = e^{-\lambda} \frac{\lambda^k}{k!}, k \in \mathbb{N}$$

la moyenne empirique est-elle un estimateur efficace de λ ?

2. Loi de Pareto de paramètres α et θ avec $\alpha > 1$ et $\theta > 0$ où θ est donné, de densité :

$$f(y) = \frac{\alpha - 1}{\theta} \left(\frac{\theta}{y}\right)^\alpha 1_{[\theta, +\infty[}(y)$$

Calculer l'espérance de $\log \frac{Y}{\theta}$. En déduire un estimateur sans biais de $\frac{1}{\alpha-1}$. Est-ce un estimateur efficace ?

3. Loi de Weibull de paramètres α et θ avec $\alpha > 0$ et $\theta > 0$ de densité où α est donné :

$$f(y) = \alpha \theta y^{\alpha-1} e^{-\theta y^\alpha} 1_{[0, +\infty[}(y)$$

Calculer l'espérance de la variable aléatoire Y^α ? En déduire un estimateur sans biais de $\frac{1}{\theta}$, est-il efficace ?

4. Loi normale de moyenne m et de variance σ^2 de densité :

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-m)^2}{2\sigma^2}}$$

La moyenne empirique est-elle un estimateur efficace de m ? L'estimateur standard de variance est-il efficace ?

2.2 Modèle à deux échantillons

Soit un échantillon de référence $\mathbf{X} = X_1, \dots, X_n$ où les X_i sont supposés indépendants et de loi $\mathcal{N}_{\mu, \sigma^2}$ et un échantillon de comparaison $\mathbf{Y} = Y_1, \dots, Y_m$ mesuré après une action dont on cherche à évaluer l'impact. On supposera que ce second échantillon peut être modélisé par une loi $\mathcal{N}_{\mu+\Delta, \sigma^2}$ (en d'autres termes, l'effet de l'action se traduit, en moyenne, par une translation des réponses mesurées).

1. Calculer la matrice d'information de Fisher jointe $I_{\mathbf{X}, \mathbf{Y}}(\mu, \Delta)$ et commenter l'expression obtenue (on aura intérêt à considérer plutôt $I_{\mathbf{X}, \mathbf{Y}}^{-1}(\mu, \Delta)$).
2. Quel estimateur simple de μ et Δ est *efficace* dans ce modèle ?

2.3 Données censurées

On observe un échantillon de taille n de durées $X_i \geq 0$ supposées indépendantes et distribuées selon une même loi exponentielle de paramètre λ .

1. Calculer l'information de Fisher associée aux observations pour le paramètre λ .
2. A partir du résultat de la question 1, calculer l'information de Fisher pour le paramètre $1/\lambda$.
3. Calculer le biais et la variance de l'estimateur de $1/\lambda$ donné par $1/n \sum_{i=1}^n X_i$. Qu'en conclure ?

On suppose maintenant que la durée totale fixée pour la collecte des observations est finie : au bout d'un temps t fixé, on observe la valeur des observations pour lesquelles $X_i \leq t$, les autres durées ne sont pas observées et on les prendra par convention égales à t . Les observations suivent donc le modèle $Y_i = \min\{X_i, t\}$ où les X_i suivent le modèle étudié précédemment.

4. Calculer la matrice d'information de Fisher pour le paramètre $1/\lambda$ dans le modèle de données censurées. Comment ce résultat se compare-t-il à celui obtenu à la question 1 ?
5. Déterminer l'estimateur du maximum de vraisemblance de $1/\lambda$ et commenter le résultat obtenu.
6. Utiliser la méthode delta pour étudier le comportement asymptotique de cet estimateur.

2.4 Modèle de mélange

1. (Preliminaire) Soit Z et X des variables aléatoires dont la loi (jointe) dépend d'un paramètre θ ; montrer que $I_{Z,X}(\theta) \geq I_X(\theta)$ en commentant le cas d'égalité.

Si $(f_i)_{1 \leq i \leq k}$ sont des densités de probabilités, on appelle « modèle de mélange à k composantes » la densité de probabilité

$$p(x) = \sum_{i=1}^k \pi_i f_i(x)$$

où les π_i sont des réels positifs dits « poids du mélange » qui vérifient $\sum_{i=1}^k \pi_i = 1$.

2. Montrer que le modèle de mélange peut s'interpréter comme un modèle partiellement observé dans lequel :

$$\begin{cases} Z & \sim \mathcal{M}_1(\pi_1, \dots, \pi_k) \\ X|Z=i & \sim f_i \end{cases}$$

On considère dans la suite le cas particulier d'un mélange à $k = 2$ composantes avec des densités f_1 et f_2 connues (et correspondant à des lois de probabilité distinctes). On note $\pi = \pi_1$ le paramètre inconnu de ce modèle.

3. Calculer en utilisant le résultat de la question 1 une borne supérieure de $I_X(\pi)$, et interpréter le cas où cette borne est atteinte.

3 Petite classe 3 : Modèle linéaire

AU PROGRAMME : MOINDRES CARRÉS ; RAPPELS SUR LA LOI GAUSSIENNE MULTIVARIÉE ; DISTRIBUTIONS DU KHI-DEUX ET DE STUDENT (DONT EXERCICES DE BASE SUR CES DEUX LOIS) ET MODÈLE LINÉAIRE.

3.1 Distribution du χ^2

Par définition, la loi du $\chi^2(k)$ « à k degrés de liberté » est celle de $\sum_{i=1}^k X_i^2$ où $(X_i)_{1 \leq i \leq k}$ sont des variables IID de loi $\mathcal{N}(0, 1)$.

1. Calculer les deux premiers moments de la loi $\chi^2(k)$ et en déduire une approximation gaussienne valable pour les grandes valeurs de k .
2. Montrer que la somme de variables indépendantes distribuées selon des lois du χ^2 suit également une loi du χ^2 .
3. Si $X \sim \mathcal{N}(\mu, \Sigma)$ est un vecteur gaussien multivarié de dimension p , montrer que
 - si Σ est définie positive, $(X - \mu)' \Sigma^{-1} (X - \mu)$ suit une loi $\chi^2(p)$,
 - si Σ est de rang $r < p$, $(X - \mu)' \Sigma^\# (X - \mu)$ où $\Sigma^\#$ est une « pseudo-inverse »* de Σ , suit une loi $\chi^2(r)$.
4. Si $X \sim \mathcal{N}(0, I_p)$ est un vecteur gaussien multivarié de dimension p , et P une matrice de projection orthogonale de \mathbb{R}^p sur un sous espace de dimension k (c'est-à-dire $P^2 = P$, $P' = P$ et $\text{rang}(P) = k$), alors $\|PX\|^2$ suit une loi du $\chi^2(k)$.

3.2 Modèle linéaire gaussien

On considère le modèle linéaire suivant

$$Y_i = \sum_{j=1}^k X_{ij} \beta_j + U_i \quad i = 1, \dots, n$$

où $(U_i)_{1 \leq i \leq n}$ sont des variables IID gaussienne d'espérance nulle et de variance σ^2 et où les (X_{ij}) sont des variables exogènes indépendantes, connues de l'utilisateur. On supposera que $E(X_i) = 0$, $E(|X_i|^2) < \infty$ et $\text{Cov}(X_i) = C$ où les vecteurs X_i sont définis par $X_i = (X_{i1}, \dots, X_{ik})'$.

1. A quelle condition les paramètres du modèle ($\beta = (\beta_1, \dots, \beta_k)'$ et σ^2) sont-ils identifiables ?
2. Déterminer l'estimateur du maximum de vraisemblance des paramètres (on montrera que l'estimateur du maximum de vraisemblance correspond à la solution d'un problème de projection orthogonale sur un sous-espace linéaire de \mathbb{R}^k).
3. Calculer la matrice d'information de Fisher du modèle linéaire. Les estimateurs obtenus sont-ils efficaces ?
4. Déterminer les lois (conditionnelles aux variables exogènes) des estimateurs des paramètres obtenus à la question 2.
5. Reprendre les questions 2 et 4 dans le cas d'un modèle de bruit plus général dans lequel on suppose que $\text{Cov}(U) = \Sigma$, supposée connue, où $U = (U_1, \dots, U_n)'$.

*C'est à dire une matrice symétrique positive telle que $\Sigma \Sigma^\# \Sigma = \Sigma$, ou en d'autres termes, une matrice qui se comporte comme l'inverse de Σ sur le sous-espace vectoriel $\text{Im}(\Sigma)$.

4 Petite classe 4 : Divergence de Kullback, maximum de vraisemblance, M-estimateurs

AU PROGRAMME : DIVERGENCE DE KULLBACK (DÉFINITION ET PROPRIÉTÉS), MAXIMUM DE VRAISEMBLANCE (EXEMPLES...), AU MOINS UN EXEMPLE DANS UN MODÈLE CONDITIONNEL.

4.1 Divergence de Kullback-Leibler

On note $I(p||q) = \int p(x) \log \frac{p(x)}{q(x)} dx$ la divergence de Kullback-Leibler (également appelée entropie relative) entre deux densités de probabilités p et q .

1. Rappeler pourquoi $I(p||q)$ est toujours bien défini et $I(p||q) \geq 0$ avec égalité si et seulement si p et q correspondent à la même loi de probabilité.
2. Calculer $I(\mathcal{N}_{\mu',v'} || \mathcal{N}_{\mu,v})$ où $\mathcal{N}_{\mu,v}$ désigne la densité gaussienne de moyenne μ et de variance v .
3. Soit p une loi telle que $E_p[X^2] < \infty$ et $E_p[|\log(p(X))|] < \infty$; on note E_p et V_p l'espérance et la variance sous la loi p . Montrer que

$$I(p || \mathcal{N}_{\mu,v}) = I(p || \mathcal{N}_{E_p, V_p}) + I(\mathcal{N}_{E_p, V_p} || \mathcal{N}_{\mu,v})$$

en déduire la meilleure approximation gaussienne de p au sens de la divergence de Kullback.

4.2 Maximum de vraisemblance pour la loi exponentielle translatée

Soit f la densité de la loi exponentielle de paramètre θ translatée de α :

$$f(y) = \frac{1}{\theta} \exp\left(-\frac{y-\alpha}{\theta}\right) \mathbb{1}_{[\alpha, +\infty[}(y)$$

1. Calculer l'estimateur du maximum de vraisemblance $\hat{\alpha}_n$ et $\hat{\theta}_n$ de α et θ .
2. Calculer la loi de $n(\hat{\alpha}_n - \alpha)$.
3. Déterminer la loi limite de $\sqrt{n}(\hat{\theta}_n - \theta)$.
4. Donner l'expression de la loi de l'ensemble ordonné (« statistiques d'ordre ») des n variables (Y_1, \dots, Y_n) noté $(Y_{(1)}, \dots, Y_{(j)}, \dots, Y_{(n)})$ et de la $i^{\text{ème}}$ variable $Y_{(i)}$. En déduire la loi du n -uplet

$$(nY_{(1)}, (n-1)(Y_{(2)} - Y_{(1)}), \dots, 2(Y_{(n-1)} - Y_{(n-2)}), Y_{(n)} - Y_{(n-1)})$$

En déduire l'indépendance de $\hat{\alpha}_n$ et de $\hat{\theta}_n$.

Pourquoi les résultats obtenus sont-ils différents du cas considéré dans le cours ?

5 Petite classe 5 : Tests et intervalles de confiance

AU PROGRAMME : TESTS ET INTERVALLES DE CONFIANCES NON ASYMPTOTIQUES, AU MOINS UN EXEMPLE DANS LE MODÈLE LINÉAIRE (PAR EXEMPLE ILLUSTRANT LA RÉGION DE CONFIANCE EN DIMENSION DEUX).

5.1 Test de dispersion

On considère (X_1, \dots, X_n) des variables gaussiennes indépendantes, de loi $\mathcal{N}(0, \sigma)$. Soient les hypothèses $H_0 : \sigma = \sigma_0$ et $H_1 : \sigma = \sigma_1$, avec $0 < \sigma_0 < \sigma_1$. Déterminez le test de Neyman-Pearson de niveau α pour H_0 contre H_1 .

5.2 Intervalle de confiance pour la loi uniforme

Soit X_1, \dots, X_n n v.a. distribuées suivant une loi uniforme sur l'intervalle $[0, \theta]$ et soit $X_{(n)} = \max(X_1, \dots, X_n)$.

1. Montrer que $X_{(n)}/\theta$ est une fonction pivotale pour θ .
2. En utilisant cette fonction pivotale, déterminer l'intervalle de confiance de probabilité de couverture $(1 - \alpha)$ de longueur minimale.
3. Discuter comment on peut utiliser cet intervalle de confiance pour construire un test de l'hypothèse $\theta = \theta_0$.

5.3 Intervalle de confiance

Soient X_1, \dots, X_n le nombre de minutes qu'un groupe d'utilisateurs-test d'internet passent connectés par semaine. Nous modélisons ces variables aléatoires par des lois exponentielles de paramètre λ . On cherche à construire un intervalle de confiance de niveau $1 - \alpha$ pour la fonction

$$q(\lambda) = P_\lambda[X \geq x] = \exp(-\lambda x)$$

la probabilité que les utilisateurs-tests passent plus de x heures connectés dans la semaine.

1. Montrer que si $X \sim \exp(\lambda)$, $2\lambda X_i$ suit une loi du χ^2 à deux degrés de liberté.
2. En déduire un intervalle de confiance de niveau $1 - \alpha$ pour λ puis pour $q(\lambda)$. La loi de $v(X, \lambda) = 2\lambda \sum_{i=1}^n X_i$ dépend-elle de λ ?
3. Comment faire si le nombre d'utilisateurs-tests est important?

5.4 Modèle linéaire gaussien : modèle électoral

Dans une étude sur le comportement électoral de la classe sociale ouvrière, un politologue recherche une relation entre le vote pour les candidats du parti P et la proportion d'ouvriers dans les circonscriptions. Il postule une relation de la forme

$$p_i = a + bx_i + u_i$$

où p_i est le pourcentage de voix que le candidat du parti P a obtenu dans la circonscription i , x_i est le pourcentage d'ouvriers, u_i est une perturbation normale centrée de variance σ^2 . Il dispose de $n = 30$ observations et calcule les quantités suivantes : $\bar{p} = 25,86$, $\bar{x} = 29,12$, $\sum_i (p_i - \bar{p})^2 = 13,64669$, $\sum_i (x_i - \bar{x})^2 = 17,64071$, $\sum_i (p_i - \bar{p})(x_i - \bar{x}) = 15,37778$.

1. Faites une analyse critique du modèle postulé par rapport à la nature du phénomène modélisé.
2. Calculer les estimateurs du maximum de vraisemblance de a et b . Quel lien y a-t-il avec les estimateurs des moindres carrés ordinaires ?
3. Calculer les estimateurs de σ^2 , de b ainsi que le coefficient de détermination R^2 du modèle (ratio de la variance expliquée par le modèle sur la variance de la variable à expliquer).
4. Comment tester l'hypothèse que l'augmentation de 1% du nombre d'ouvriers induit une augmentation de 1% des voix pour le candidat du parti P au seuil de 5%.
5. Donner une région de confiance (au niveau $1 - \alpha$) pour une prévision faite à l'aide de ce modèle.
6. Avant de nouvelles élections, on dispose de l'information que le pourcentage d'ouvriers dans la population est 22,2%. A quel pourcentage de voix pour le parti P peut-on s'attendre aux élections ? A l'issue du vote, on mesure que le score du candidat du parti P est 15,1%. Conclure quant à la validité du modèle.

6 Petite classe 6 : Eléments d'asymptotique

AU PROGRAMME : LEMME DE SLUTSKY ET MÉTHODES DELTA (**A moduler selon ce qui aura été fait en cours**), EXEMPLES D'APPLICATION À L'ÉTUDE DE PERFORMANCES ASYMPTOTIQUES D'ESTIMATEURS.

6.1 « Méthode delta »

On rappelle les définitions suivantes

Convergence en probabilité $X_n \xrightarrow{P} X$, si pour tout $\epsilon > 0$,

$$P(|X_n - X| > \epsilon) \longrightarrow 0$$

Convergence en loi $X_n \xrightarrow{\mathcal{L}} X$, si l'une de ces deux conditions équivalentes est vérifiée (Lemme de Pormanteau) :

1. $\forall B \in \mathcal{B}(\mathbb{R})$ tel que $P(X \in \delta B) = 0$, où $\delta B = \bar{B} - \overset{\circ}{B}$ désigne la frontière de B ,

$$P(X_n \in B) \longrightarrow P(X \in B)$$

On montre également qu'il suffit, par exemple, que cette condition soit vérifiée pour les intervalles $] -\infty, b]$ pour lequel $P(X = b) = 0$ (convergence de la fonction de répartition en ses points de continuité).

2. Pour toute fonction f continue bornée,

$$E[f(X_n)] \longrightarrow E[f(X)]$$

On montre de même qu'il suffit en fait que cette condition soit vérifiée, par exemple, pour les fonctions $f_t(x) = e^{itx}$, pour $t \in \mathbb{R}$ (la convergence en loi étant donc équivalente à la convergence ponctuelle des fonctions caractéristiques).

1. Montrer que si f est une fonction continue,

$$X_n \xrightarrow{P} X \implies f(X_n) \xrightarrow{P} f(X)$$

$$X_n \xrightarrow{\mathcal{L}} X \implies f(X_n) \xrightarrow{\mathcal{L}} f(X)$$

2. Montrer que $X_n \xrightarrow{\mathcal{L}} X$ et $(Y_n - X_n) \xrightarrow{P} 0$, implique $Y_n \xrightarrow{\mathcal{L}} X$.

3. En déduire $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{P} c$, où c est une constante déterministe, implique

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} X \\ c \end{pmatrix}$$

Montrer en particulier que (Lemme de Slutsky), $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{P} 0$, implique

$$- X_n + Y_n \xrightarrow{\mathcal{L}} X$$

$$- X_n Y_n \xrightarrow{\mathcal{L}} 0 \text{ (ce qui est équivalent à } X_n Y_n \xrightarrow{P} 0)$$

4. Soit un estimateur T_n (faiblement) consistant de θ (tel que $T_n \xrightarrow{P} \theta$), vérifiant

$$r_n(T_n - \theta) \xrightarrow{\mathcal{L}} T$$

où r_n est une suite positive croissante qui vérifie $r_n \uparrow +\infty$ (en général, pour un modèle régulier $r_n = \sqrt{n}$ et T est une loi Gaussienne).

Montrer que pour toute transformation ϕ différentiable en θ et telle que $\dot{\phi}(\theta) \neq 0$, on a $\phi(T_n) \xrightarrow{P} \phi(\theta)$ et

$$r_n(\phi(T_n) - \phi(\theta)) \xrightarrow{\mathcal{L}} \dot{\phi}(\theta)T \quad (\text{« méthode delta »})$$

Dans le cas où θ est un paramètre vectoriel et ϕ une transformation à valeur (éventuellement) vectorielle, on vérifiera que le résultat précédent reste correct à condition de remplacer $\dot{\phi}(\theta)$ par la matrice jacobienne $(J_{kl}) = \left(\frac{\partial \phi_k(\theta)}{\partial \theta_l} \right)$.

5. Application : Soit P_θ un modèle statistique et $\hat{\theta}_n$ un estimateur asymptotiquement efficace de θ . On considère une reparamétrisation $\lambda = \phi(\theta)$ où ϕ est une fonction inversible différentiable et d'inverse différentiable. Montrer que l'estimateur $\hat{\lambda}_n = \phi(\hat{\theta}_n)$ est un estimateur asymptotiquement efficace de λ .

6.2 Estimateur de la variance empirique

On suppose que $X_i \sim f(x - \theta)$, où f est une densité de probabilité sur \mathbb{R} symétrique dont on note $\mu_k = \mathbb{E}_f(X^k)$ les moments d'ordre $k = 2$ et $k = 4$.

1. Montrer que l'estimateur $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$, où $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$, vérifie un théorème central limite (indication : on montrera d'abord que l'on peut se ramener au cas où $\theta = 0$ puis on exprimera l'estimateur comme un transformation de $S_n = \frac{1}{n} \sum_{i=1}^n X_i^2$ et de \bar{X}_n).
2. En déduire que dans le cas où f est la densité Gaussienne, l'estimateur non biaisé $\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ de la variance σ^2 est asymptotiquement efficace.

6.3 Test unilatère

Soit X_1, \dots, X_n des observations IID de loi de médiane θ ($P(X \leq \theta) = P(X > \theta) = 1/2$) dont on notera $F_\theta(x) = F(x - \theta)$ la fonction de répartition. On souhaite tester l'hypothèse $H_0 : \theta = 0$ contre les alternatives de la forme $H_1 : \theta > 0$.

On considère tout d'abord un test très simple ne nécessitant aucune connaissance supplémentaire sur la forme de F : le test de signe basé sur la statistique

$$S_n = 1/n \sum_{i=1}^n \mathbb{1}\{X_i > 0\}$$

1. Montrer que sous P_θ ,

$$\sqrt{n}(S_n - \mu(\theta)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2(\theta))$$

où $\mu(\theta) = 1 - F(-\theta)$ et $\sigma^2(\theta) = (1 - F(-\theta))F(-\theta)$.

2. Comment construire un test de niveau asymptotique α à partir de S_n ?
3. Quelle est la puissance asymptotique de ce test sous P_θ (avec $\theta > 0$)[†] ?

Si la loi de fonction de répartition F admet une densité f symétrique (et telle que $\int x^2 f(x) dx < \infty$) de telle sorte que la médiane coïncide avec l'espérance, il est assez naturel de considérer un test (dit de Wald) basé sur la valeur de la moyenne empirique

$$\bar{X}_n = 1/n \sum_{i=1}^n X_i$$

4. Comment construire un test de niveau asymptotique α à partir de \bar{X}_n selon que la variance de la loi est ou non supposée connue ? Quelle est sa puissance asymptotique (on ne considérera ici que le cas où la variance est supposée connue, l'autre cas est plus complexe mais peut s'étudier à partir du résultat de l'exercice 6.2) ?
5. Dans le cas où f correspond à la loi $\mathcal{N}(0, \sigma^2)$, montrer que le test précédent est asymptotiquement plus puissant que le test de signe (pour les valeurs faibles de $\theta > 0$).

[†]On admettra le fait que dans le théorème central limite avec limite normale, la convergence de la fonction de répartition vers la fonction de répartition Φ normale se fait de façon uniforme, conséquence du second théorème de Dini : Si g_n est une suite de fonctions croissantes de $[a, b]$ dans \mathbb{R} convergeant vers une limite g continue, alors elle converge uniformément vers g .

7 Petite classe 7 : Asymptotique des M-estimateurs

AU PROGRAMME : COMPORTEMENT ASYMPTOTIQUE DES M-ESTIMATEURS, AU MOINS UN EXEMPLE AUTRE QUE SUR LE MAXIMUM DE VRAISEMBLANCE (AVEC $I(\theta_0)$ ET $J(\theta_0)$ DIFFÉRENTES) ET UN EXEMPLE DANS UN MODÈLE CONDITIONNEL ; APPLICATIONS AU TESTS ASYMPTOTIQUES D'AJUSTEMENT ET AUX INTERVALLES DE CONFIANCE ASYMPTOTIQUES.

7.1 Maximum de vraisemblance et moindres carrés non linéaires dans le modèle de choix binaire

On considère un modèle conditionnel dans lequel la variable de réponse Y_i peut prendre, conditionnellement à X_i , deux valeurs :

- 1 avec la probabilité $h_\theta(X_i)$,
- 0 avec la probabilité $1 - h_\theta(X_i)$.

où h_θ est une fonction connue dépendant de θ , X_i est la variable exogène (éventuellement vectorielle) et θ est un vecteur de paramètres inconnus. Cette classe de modèles contient en particulier le modèle Probit, pour $h_\theta(x) = \Phi(x'\theta)$ avec $\Phi(z) = \frac{1}{2\pi} \int_{-\infty}^z \exp(-t^2/2) dt$ et le modèle Logit (ou logistique), quand $h_\theta(x) = \frac{1}{1 + \exp(-x'\theta)}$. On considère l'estimateur du maximum de vraisemblance ainsi que l'estimateur des moindres carrés non-linéaires associés au modèle $P_\theta(Y_i = 1|X_i) = E_\theta[Y_i|X_i] = h_\theta(X_i)$.

1. Calculer les matrices de variance-covariance asymptotiques pour les deux estimateurs.
2. Montrer que si X et Y sont des vecteurs aléatoires d'espérance nulle pour lesquels on note $\Sigma_{XX} = E[XX']$, $\Sigma_{YY} = E[YY']$, $\Sigma_{XY} = E[XY']$ et on suppose que Σ_{YY} est inversible,

$$\min_m E[(X - mY)(X - mY)'] = \Sigma_{XX} - \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{XY}'$$

où le minimum est pris sur l'ensemble des matrices m déterministes de taille $\dim(X) \times \dim(Y)$; en déduire que $\Sigma_{XX} - \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{XY}'$ est une matrice positive[‡].

3. Utiliser le résultat précédent pour montrer que la matrice de variance-covariance asymptotique associée à l'estimateur du maximum de vraisemblance est inférieure à celle correspondant à l'estimateur des moindres carrés non-linéaires.

7.2 Analyse de probabilité

Un statisticien se propose d'analyser comment la probabilité de partir en vacance dépend du revenu. Il dispose pour cela d'un ensemble d'observations issues d'une enquête auprès de n personnes qui peuvent être regroupées selon leur niveau de revenu R en k classes. Il a calculé la proportion de personnes p_i qui partent en vacance pour un niveau de revenu donné R_i . Néanmoins pour prendre en compte le fait que la probabilité est une grandeur comprise entre 0 et 1, il retient le modèle suivant :

$$p_i = \frac{e^{a+bR_i}}{1 + e^{a+bR_i}} + \varepsilon_i \quad E(\varepsilon_i) = 0, V(\varepsilon_i) = s^2$$

[‡]Il s'agit de l'interprétation probabiliste d'un résultat connu sous le nom de *complément de Schur*.

1. Donner les équations définissant l'estimateur des moindres carrés non-linéaires des coefficients (a, b) . Donner leur matrice de variance-covariance asymptotique.

Un second statisticien considère les données initiales, pour chaque classe de revenu R_i il dispose de n_i observations et postule un modèle de la forme

$$\begin{cases} Y_i = \text{"part en vacance"} & y_i^* = \alpha + \beta R_i + v_i \geq 0 \\ Y_i = \text{"ne part pas en vacance"} & y_i^* < 0 \end{cases}$$

où v_i suit une loi logistique de densité

$$f(v) = \frac{e^{-v}}{(1 + e^{-v})^2}$$

2. Donner les estimateurs du maximum de vraisemblance des paramètres (α, β) de ce modèle. Donner leur matrice de variance covariance asymptotique. Quel est le lien entre ce modèle et le précédent ?
3. On dispose d'un grand nombre d'observations n , donner la prévision de la probabilité de partir en vacance pour une personne de revenu R_0 pour le dernier modèle ainsi que sa précision.

7.3 Test de fiabilité

Une entreprise reçoit n lots de k pièces avec la garantie que la proportion p de pièces défectueuses est la même dans chaque lot et inférieure à 5%. Pour vérifier que la garantie est satisfaite, un employé tire avec remise des pièces dans chaque lot, jusqu'à obtenir une pièce défectueuse par lot.

Soit Y_i la variable aléatoire désignant le nombre de tirages nécessaires dans le lot numéro i pour obtenir une pièce défectueuse.

1. Quelle est la loi de Y_i ? Calculer EY_i et VY_i .
2. Proposer un test de Wald de l'hypothèse $H_0 : p = 0,05$.

7.4 Intervalle de confiance pour le modèle de Bernoulli

On effectue un sondage sur un échantillon de 400 électeurs qui recueille 212 intentions de vote en faveur d'un candidat C .

1. On veut faire une estimation par intervalle de confiance de la proportion des intentions de vote en faveur du candidat C . Préciser et justifier la méthode choisie.
2. Donner un intervalle de confiance de probabilité de couverture asymptotique $1 - \alpha$ des intentions de vote en faveur du candidat C dans la population entière ; on prendra par exemple $\alpha = 0.05$.

A.N. : Intervalle confiance pour p : $[0.48, 0.58]$.

7.5 Test de Wald

On considère une hypothèse nulle $H_0 : g(\theta) = 0$, où g est à valeur dans \mathbb{R}^r et la statistique de Wald associée $\xi_n^W = ng'(\hat{\theta}_n)\hat{\Sigma}_W^{-1}g(\hat{\theta}_n)$ avec $\hat{\Sigma}_W = \left(\frac{\partial g}{\partial \theta'} J^{-1} I J^{-1} \frac{\partial g'}{\partial \theta}\right)_{\theta=\hat{\theta}_n}$. On considère également une hypothèse nulle $H_0(a)$ du type $a'g(\theta) = 0$ où a est un vecteur donné de \mathbb{R}^r .

1. Montrer que la statistique de Wald associée à $H_0(a)$ est :

$$\xi_n^W(a) = \frac{n[a'g(\hat{\theta}_n)]^2}{a'\Sigma_W a}$$

et que cette statistique suit asymptotiquement sous H_0 la loi $\chi^2(1)$.

2. Montrer que $\max_a \xi_n^W(a) = \xi_n^W$ et donc que $\max_a \xi_n^W(a)$ suit, sous H_0 la loi $\chi^2(r)$.

7.6 Test d'hypothèse sur le paramètre d'une loi de Poisson

On considère les réalisations de T variables aléatoires IID Y_1, Y_2, \dots, Y_T issues d'une loi de Poisson de paramètre λ inconnu. On s'intéresse au test de l'hypothèse : $H_0 : \lambda = \lambda_0$ contre l'alternative $H_1 : \lambda \neq \lambda_0$ où λ_0 est un réel (positif) connu.

1. On se propose d'utiliser un test de Wald pour tester cette hypothèse. Donner la forme du test.
2. Application numérique : $\lambda_0 = 1$, $\alpha = 0,05$ a) $T = 100$, $\bar{y} = 1,2$, b) $T = 200$, $\bar{y} = 1,1$.

7.7 Test de significativité dans le modèle de régression linéaire

Dans le modèle de régression linéaire, on cherche à s'assurer de la significativité d'un groupe de régresseurs en testant l'hypothèse $\theta_{\bullet} = \mathbf{0}$ où θ_{\bullet} est un sous-vecteur du vecteur θ des coefficients de régression (pour simplifier la présentation, on supposera que θ_{\bullet} correspond au p dernières composantes de θ). On suppose observer des couples de variables $(\mathbf{X}_i, Y_i)_{1 \leq i \leq n}$ IID, Y_i étant une réponse scalaire dépendant de la variable exogène vectorielle \mathbf{X}_i à travers le modèle de régression linéaire suivant

$$Y_i = \mu + \mathbf{X}_i' \theta + B_i$$

$(B_i)_{1 \leq i \leq n}$ étant une séquence IID de bruit indépendante des observations. On suppose que $\mathbb{E}(\mathbf{X}_i) = \mathbf{0}$ (on justifiera le fait que cette hypothèse n'est pas restrictive) et on note Σ la matrice de covariance de \mathbf{X}_i . La densité de probabilité du bruit B_i est supposée paire et on note respectivement σ^2 et μ_4 ses moments d'ordres 2 et 4.

2. Déterminer le test de Wald correspondant à l'hypothèse nulle $H_0 : \theta_{\bullet} = \mathbf{0}$. Montrer que sous H_0 , la loi limite du test ne dépend que de la distribution des variables exogènes.
3. Montrer qu'il est possible de tester indépendamment la significativité de chaque régresseur dans le cas où les variables exogènes sont décorréélées. Qu'en est-il lorsque la variance des variables exogènes est estimée à partir des données ?

8 Petite classe 8 : Tests d'Hausman du rapport de vraisemblance généralisé

AU PROGRAMME : TEST IMPLIQUANT UNE MAXIMISATION SOUS CONTRAINTE, EN PARTICULIER TEST D'HAUSMAN ET DU RAPPORT DE VRAISEMBLANCE GÉNÉRALISÉ.

8.1 Test d'ajustement

On recueille sur n jours, le nombre de véhicules arrivant à un péage. On modélise ces données comme la réalisation de n variables aléatoires IID de loi de Poisson. Soit $(X_i)_{1 \leq i \leq n}$ un échantillon IID de loi de Poisson $\mathcal{P}(\theta^*)$.

1. Etudier les tests de Wald et du rapport de vraisemblance pour l'hypothèse $H_0 : \{\theta; \theta = \theta_0\}$.
2. (Facultatif) Que peut-on dire des propriétés comparées du test de Wald et du rapport de vraisemblance à échantillon fini ?

On considère maintenant un échantillon IID de taille n de loi gaussienne multivariée $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. On suppose $\boldsymbol{\Sigma}$ connu et on veut tester l'hypothèse $\boldsymbol{\mu} = \boldsymbol{\mu}_0$.

3. Montrer que les stratégies de Wald et du rapport de vraisemblance conduisent au même test asymptotique.
4. Quelle est la forme de la région d'acceptation de ce test ?

8.2 Test de Wald dans une loi multinomiale

On considère une loi multinomiale à k modalités ($k \geq 3$) de probabilités p_1, p_2, \dots, p_k . On dispose de n observations indépendantes issues de cette loi. On notera N_i le nombre d'observations de la modalité i . On souhaite tester l'hypothèse :

$$H_0 : p_1 + p_2 = \frac{1}{2}$$

1. Donner les estimateurs du maximum de vraisemblance contraints et non contraints des $p_j, j = 1, \dots, n$. On les note \hat{p}_j et \hat{p}_j^c .
2. Calculer la statistique de Wald ξ_W associée au test de H_0 .

On suppose dorénavant que $k > 3$.

3. Montrer que

$$\lim_{n \rightarrow +\infty} V(\hat{p}_3 - \hat{p}_3^c) = \lim_{n \rightarrow +\infty} V(\hat{p}_3) - \lim_{n \rightarrow +\infty} V(\hat{p}_3^c) \quad (1)$$

4. Calculer $\lim_{n \rightarrow +\infty} V(\hat{p}_3)$ et $\lim_{n \rightarrow +\infty} V(\hat{p}_3^c)$ sous H_0 .
5. En déduire l'expression de la statistique du test de Hausman de H_0

$$\xi_H = n \frac{(\hat{p}_3 - \hat{p}_3^c)^2}{\hat{V}} \quad (2)$$

où \hat{V} est un estimateur consistant, sous H_0 , de $\lim_{n \rightarrow +\infty} V(\hat{p}_3 - \hat{p}_3^c)$.

6. Montrer que ξ_W et ξ_H sont asymptotiquement équivalentes sous H_0 .

8.3 Test de significativité dans le modèle de régression logistique

On veut étudier la proportion de ménages équipés de lecteur DVD. Pour cela, on tire de manière équiprobable avec remise un échantillon de n ménages et on observe pour chaque ménage $i = 1, 2, \dots, n$ la variable :

$$Y_i = \begin{cases} 1 & \text{si le ménage } i \text{ est équipé} \\ 0 & \text{sinon} \end{cases}$$

et son revenu $X_i = x_i$. On veut tester l'hypothèse selon laquelle la probabilité d'être équipé dépend du revenu du ménage. On définit pour cela un modèle statistique conditionnel aux X_i par

$$P_{a,b}(Y_i = 1 | X_i) = \frac{e^{a+bX_i}}{1 + e^{a+bX_i}}$$

où a et b sont deux paramètres réels inconnus. On s'intéresse au problème de test de l'hypothèse nulle

$$H_0 : b = 0$$

1. Calculer le score du modèle pour n observations, ainsi que la matrice d'information de Fisher pour une observation $I(a, b)$.
2. Expliciter le modèle contraint sous l'hypothèse nulle. Calculer les estimateurs du maximum de vraisemblance dans le modèle contraint (satisfaisant l'hypothèse nulle) et le score en ces valeurs.
3. Donner un estimateur \hat{I}_c de la matrice de Fisher en les estimateurs contraints. Donner la statistique du test du score de l'hypothèse H_0 . Quelle est sa loi asymptotique ?

9 Petite classe 9 : Test du khi-deux et extensions

AU PROGRAMME : TEST DU KHI-DEUX, DES EXERCICES DE RÉVISION PEUVENT ÉGALEMENT ÊTRE BIEN VENUS.

9.1 Test d'adéquation du khi-deux

Soit \mathbf{X}_n une observation vectorielle de distribution multinomiale $\mathcal{M}(n, \mathbf{p})$, où $\mathbf{p} = (p_1, \dots, p_k)'$ est un vecteur de probabilité dont on supposera les éléments non nuls.

1. Montrer que $\sqrt{n}((\frac{\mathbf{X}_{n,1}}{n}, \dots, \frac{\mathbf{X}_{n,k}}{n})' - \mathbf{p}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \text{diag}(\mathbf{p}) - \mathbf{p}\mathbf{p}')$.
2. Montrer que $\sqrt{n}((\frac{\mathbf{X}_{n,1}}{n\sqrt{p_1}}, \dots, \frac{\mathbf{X}_{n,k}}{n\sqrt{p_k}})' - \sqrt{\mathbf{p}}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I - \sqrt{\mathbf{p}}\sqrt{\mathbf{p}}')$. Montrer que la matrice de covariance limite correspond à une matrice de projection sur un sous-espace vectoriel de dimension $k - 1$. En déduire la distribution limite du test d'ajustement du χ^2 à paramètre connu donné par

$$n \sum_{i=1}^k \frac{(X_{n,i}/n - p_i)^2}{p_i}$$

9.2 Test d'adéquation à une loi discrète

On étudie les arrivées d'appels dans un central téléphonique. On note lors de 200 observations consécutives, la répartition du nombre d'appels observés par seconde. On obtient le tableau suivant :

Nombre d'appels par seconde	Effectifs observés
0	6
1	15
2	40
3	42
4	37
5	30
6	10
7	9
8	5
9	3
10	2
11	1

On modélise les appels par un processus de Poisson, ce qui implique en particulier que le nombre N_t d'appels observés sur l'intervalle de temps $[0, t]$ suit une loi de Poisson de paramètre λt et que $N_{t+s} - N_t$ est indépendant de N_t et suit une loi de Poisson de paramètre λs .

1. Tester l'adéquation de la loi empirique à la loi de Poisson.
2. Quel est le niveau limite permettant d'accepter l'hypothèse d'adéquation à la loi de Poisson ?
3. A une date antérieure le nombre d'appels sur un intervalle de temps $[0, t]$ suivait une loi de Poisson de paramètre $4t$. Tester la stabilité de la description entre les deux dates.

On donne ci-dessous les probabilités $p_k(\lambda) = P(N = k)$, où N suit une loi de Poisson de paramètre λ .

	$p_k(3,7)$	$p_k(4)$
0	0,0247	0,0183
1	0,0915	0,0733
2	0,1684	0,1465
3	0,2087	0,1957
4	0,1930	0,1954
5	0,1428	0,1563
6	0,0881	0,1042
7	0,0465	0,0595
8	0,0215	0,0298
9	0,0099	0,0132
10	0,0046	0,0053
11	0,0021	0,0019
12	0,0001	0,0006
13	< 0,0001	0,0002
14	< 0,0001	0,0001

9.3 Test d'endogamie sociale

L'objet de ce test est d'utiliser le test du khi-deux pour construire un test d'indépendance. On dispose d'un ensemble de n observations $(X_i, Y_i)_{1 \leq i \leq n}$ décrivant des couples mariés, en particulier leur catégorie socio-professionnelle respective classées en k classes. On se propose de tester que pour un couple, les catégories socio-professionnelles sont indépendantes.

Ecrire le test du χ^2 correspondant à l'hypothèse d'indépendance de X et de Y et donner sa loi limite.

9.4 Test d'homogénéité de deux populations

Soit $(R_i)_{1 \leq i \leq n}$ des variables catégorielles IID à valeur dans $\{1, \dots, k\}$. On note $\mathbf{p} = (p_1, \dots, p_k)'$ le vecteur des probabilités $p_i = P(X = i)$ que l'on supposera strictement positives.

1. Rappeler pourquoi ce modèle est équivalent, du point de vue statistique, à l'observation d'une unique variable vectorielle \mathbf{X}_n de loi multinomiale $\mathcal{M}(n, \mathbf{p})$.
2. Montrer que le vecteur $(\frac{X_{n,1}}{\sqrt{p_1}}, \dots, \frac{X_{n,k}}{\sqrt{p_k}})'$ vérifie un théorème central limite dont la matrice de covariance limite correspond à une matrice de projection sur un sous-espace vectoriel de dimension $k - 1$. En déduire la distribution limite du test d'ajustement du χ^2 à paramètre connu donné par

$$\sum_{i=1}^k \frac{(X_{n,i} - np_i)^2}{np_i}$$

On considère dorénavant deux observations \mathbf{X}_m et \mathbf{Y}_n indépendants de lois respectives $\mathcal{M}(m, \mathbf{p})$ et $\mathcal{M}(n, \mathbf{q})$. Le but est de tester l'homogénéité des deux populations, c'est à dire l'égalité de

\mathbf{p} et \mathbf{q} . A cet effet, on construit le test du χ^2 composite suivant

$$\sum_{i=1}^k \frac{(X_{m,i} - m\hat{r}_i)^2}{m\hat{r}_i} + \sum_{i=1}^k \frac{(Y_{n,i} - n\hat{r}_i)^2}{n\hat{r}_i}$$

où $\hat{\mathbf{r}}$ est l'estimateur du maximum de vraisemblance du paramètre sous l'hypothèse nulle (c'est à dire en supposant que $\mathbf{p} = \mathbf{q}$).

3. Montrer que lorsque $m, n \rightarrow \infty$, la distribution limite du test sous l'hypothèse nulle est une distribution du khi-deux (indication : on exprimera le test ci-dessus en fonction des tests du χ^2 « habituels » associés aux deux échantillons \mathbf{X}_m et \mathbf{Y}_n en distinguant différents cas selon le comportement limite du rapport n/m).