Statistiques
2010-2011

Feuille de TD n°1

Estimation, tests et régions de confiance

Exercice 1

Soit $(X_1, ..., X_n)$ un n-échantillon de la loi uniforme sur $[0, \theta]$, où $\theta > 0$ est inconnu.

1. (a) Calculer $\mathbb{E}_\theta(X_1)$ et en déduire un estimateur $\widehat{\theta}_n$ de θ par la méthode des moments.
   (b) Calculer l'estimateur du maximum de vraisemblance $\widetilde{\theta}_n$ de θ.
   (c) Les estimateurs $\widehat{\theta}_n$ et $\widetilde{\theta}_n$ sont-ils biaisés ?
   (d) Comparer les risques quadratiques de $\widehat{\theta}_n$ et $\widetilde{\theta}_n$.
   (e) Etudier la convergence en loi de $\sqrt{n}(\widehat{\theta}_n - \theta)$ et de $n(\widetilde{\theta}_n - \theta)$ lorsque n tend vers l'infini.
2. Soit $\theta_0 > 0$. On souhaite tester H_0 : "$\theta = \theta_0$" contre H_1 : "$\theta \neq \theta_0$".
   (a) Construire un test de niveau α en utilisant l'estimateur $\widetilde{\theta}_n$. Tracer la courbe de puissance de ce test.
   (b) Construire un test de niveau asymptotique α en utilisant l'estimateur $\widehat{\theta}_n$ et une approximation gaussienne.

Exercice 2

Soit $(X_1, ..., X_n)$ un n-échantillon de la loi uniforme sur $[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$, où θ est un réel inconnu. L'estimateur du maximum de vraisemblance de θ est-il bien défini ?

Exercice 3

Soit $X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ une variable aléatoire dont la loi appartient à une famille donnée $\{P_\theta,\ \theta \in \Theta\}$. La vraie valeur du paramètre de la loi de X, notée θ, est inconnue.

1. On dispose d'une région de confiance $\mathcal{I}(X)$ de coefficient de sécurité $1 - \alpha$ pour θ. Pour tout $\theta_0 \in \Theta$, construire un test de niveau α de $H_0(\theta_0)$: "$\theta = \theta_0$" contre $H_1(\theta_0)$: "$\theta \neq \theta_0$".
2. Pour tout $\theta_0 \in \Theta$, on dispose d'un test de niveau α de $H_0(\theta_0)$: "$\theta = \theta_0$" contre $H_1(\theta_0)$: "$\theta \neq \theta_0$". Construire une région de confiance $\mathcal{I}(X)$ de coefficient de sécurité $1 - \alpha$ pour θ.

Exercice 4

Soit $(X_1, ..., X_n)$ un n-échantillon de la loi uniforme sur $[0, \theta]$, où $\theta > 0$ est inconnu. En utilisant l'exercice 3 et la question 2a de l'exercice 1, construire un intervalle de confiance pour θ de coefficient de sécurité $1 - \alpha$.

Exercice 5

Quelques rappels de probabilités utiles en statistique asymptotique. On suppose dans toute la suite que $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires dans $\mathbb{R}^d$, convergeant en loi vers une v.a. X.

1. Soit g une fonction continue en tout point d'un sous-ensemble C de $\mathbb{R}^d$ tel que $\mathbb{P}(X \in C) = 1$. Montrer que $(g(X_n))_{n\in\mathbb{N}}$ converge en loi vers $g(X)$.
2. Dans cette question, on suppose que X est constante p.s, égale à 0. Soit g une fonction de $\mathbb{R}^d$ dans $\mathbb{R}^k$ telle que $g(h) = o(\|h\|^p)$ au voisinage de 0, pour un entier $p \geqslant 0$. Montrer que $g(X_n) = o_P(\|X_n\|^p)$, c'est à dire qu'il existe une variable aléatoire réelle Y_n dans $\mathbb{R}^k$, convergeant vers 0 en probabilité, telle que $g(X_n) = Y_n\|X_n\|^p$.
3. Soit $(Y_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires dans $\mathbb{R}$, convergeant en loi vers une constante c. Montrer que $X_n.Y_n$ converge en loi vers $X.c$.
4. Soit r_n une suite de réels tendant vers $+\infty$ et $\theta \in \mathbb{R}^d$ tel que $r_n(X_n - \theta)$ converge en loi vers une variable aléatoire T. Soit g de $\mathbb{R}^d$ dans $\mathbb{R}^k$, une fonction différentiable en θ. Montrer que $r_n(g(X_n) - g(\theta))$ converge en loi vers la variable aléatoire $Dg_\theta(T)$.

Exercice 6

Soit $(X_1, ..., X_n)$ un n-échantillon de la loi de Poisson de paramètre inconnu $\theta > 0$.

1. Estimer θ. On notera $\widehat{\theta}_n$ l'estimateur proposé.
2. Déterminer la loi limite de $\widehat{\theta}_n - \theta$ convenablment renormalisé (de sorte que la loi limite ne soit pas triviale), puis construire un intervalle de confiance asymptotique de coefficient de sécurité $1 - \alpha$ pour θ.
3. Proposer un estimateur asymptotiquement sans biais et consistant de $P_\theta(X_1 = 0)$.

Exercice 7

Soit $(X_1, ..., X_n)$ un n-échantillon de la loi Gaussienne $\mathcal{N}(\mu, \sigma^2)$, où $\mu \in \mathbb{R}$ et $\sigma > 0$. On note

$$\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^n X_i \quad \text{et} \quad \widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \widehat{\mu}_n)^2.$$

1. Préliminaires : Soient Y le vecteur aléatoire à valeurs dans $\mathbb{R}^n$ dont la $i$ème coordonnée est $(X_i - \mu)/\sigma$ et E_1 le sous-espace vectoriel de $\mathbb{R}^n$ engendré par le vecteur ${}^t(1, \ldots, 1)$.
   (a) Calculer les projections orthogonales de Y sur E_1 et sur ${}^\perp E_1$ et en déduire la loi de $(\widehat{\mu}_n, \widehat{\sigma}_n^2)$.
   (b) En déduire que si l'on note $\widehat{\sigma}_n$ la racine carrée positive de $\widehat{\sigma}_n^2$, alors $T \sim t(n-1)$, où
   $$T = \sqrt{n-1}\,\frac{\widehat{\mu}_n - \mu}{\widehat{\sigma}_n}$$
2. On suppose dans cette question que μ est inconnu et σ est connu.
   (a) Calculer l'estimateur du maximum de vraisembance de μ. Est-il biaisé ? consistant ?
   (b) On se donne $\alpha \in]0, 1[$ et $\mu_0 \in \mathbb{R}$. On souhaite construire le test du rapport des vraisemblances de $\mu = \mu_0$ contre $\mu \neq \mu_0$ au niveau α. Montrer que le rapport des vraisemblances Λ est
   $$\Lambda = \exp\left(\frac{n}{2\sigma^2}(\widehat{\mu}_n - \mu_0)^2\right),$$
   puis donner la région de rejet du test du rapport des vraisemblances.
3. On suppose dans cette question que σ est inconnu et μ est connu.
   (a) Estimer σ^2 et étudier les propriétés de l'estimateur proposé.

(b) Construire un intervalle de confiance pour σ^2.

4. On suppose dans cette question que σ et μ sont inconnus.

   (a) Calculer l'estimateur du maximum de vraisemblance de (μ, σ^2) et étudier ses propriétés.

   (b) On se donne $\alpha \in]0,1[$ et $\mu_0 \in \mathbb{R}$. On souhaite construire le test du rapport des vraisemblances de $\mu = \mu_0$ contre $\mu \neq \mu_0$ au niveau α. Montrer que le rapport des vraisemblances Λ est

$$\Lambda = \left(\frac{\widetilde{\sigma}_n^2}{\widehat{\sigma}_n^2}\right)^{n/2} = \left(1 + \frac{(\widehat{\mu}_n - \mu_0)^2}{\widehat{\sigma}_n^2}\right)^{n/2}$$

   où $\widetilde{\sigma}_n^2 = \sum_i (X_i - \mu_0)^2/n$ puis donner la région de rejet du test du rapport des vraisemblances.

   (c) Donner un intervalle de confiance pour μ de coefficient de sécurité 0.95 sur la base de l'observation suivante d'un échantillon gaussien :

-0.4326	-1.6656	0.1253	0.2877	-1.1465	1.1909	1.1892	-0.0376	0.3273
0.1746	-0.1867	0.7258	-0.5883	2.1832	-0.1364	0.1139	1.0668	0.0593
-0.0956	-0.8323	0.2944	-1.3362	0.7143	1.6236	-0.6918		

   (les réalisations de $\widehat{\mu}_n$ et de $\widehat{\sigma}_n$ sont ici 0.1171 et 0.8955).

Statistiques
2010-2011

FEUILLE DE TD N°2
TESTS NON PARAMÉTRIQUES

EXERCICE 8
Soit $(X_1, \ldots, X_n)$ un n-échantillon. On note $\mathbb{F}_n$ la fonction de répartition empirique associée :

$$\forall t \in \mathbb{R},\ \mathbb{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{X_i \leqslant t}.$$

Pour toute fonction de répartition F, on note

$$D_n(F) = \sup_{t \in \mathbb{R}} |F(t) - \mathbb{F}_n(t)|.$$

1. Soit $(U_1, ..., U_n)$ un n-échantillon de la loi uniforme sur $[0,1]$ et F une fonction de répartition. On note $\mathbb{U}_n$ la fonction de répartition empirique associée à $(U_1, ..., U_n)$, X une variable aléatoire de fonction de répartition F et F^{-1} l'inverse généralisé de F :

$$\forall u \in [0,1],\ F^{-1}(u) = \inf\{t \in \mathbb{R},\ F(t) \geqslant u\},$$

avec la convention $\inf \emptyset = +\infty$.

   (a) Montrer que pour tout réel t,

$$\{u,\ F^{-1}(u) \leqslant t\} = \{u,\ u \leqslant F(t)\}.$$

   (b) Montrer que $(F^{-1}(U_1), ..., F^{-1}(U_n))$ est un n-échantillon de la loi de X.

   (c) En déduire que si F est continue et si $(X_1, \ldots, X_n)$ est un n-échantillon de la loi de X, alors $D_n(F)$ est identique en loi à $\sup_{t \in [0,1]} |\mathbb{U}_n(t) - t|$, et la loi de $D_n(F)$ ne dépend pas de F (elle est tabulée).

2. On se donne une fonction de répartition continue F_0, et on note F la fonction de répartition commune des X_i. Construire un test de $F = F_0$ contre $F \neq F_0$ sur la base de l'observation de $(X_1, \ldots, X_n)$ et montrer que la puissance du test proposé tend vers 1 en tout point de l'alternative quand $n \to \infty$.

3. Tester si la variable aléatoire X suit une loi normale $\mathcal{N}(2,1)$, connaissant l'observation suivante d'un 21-échantillon : 0.3, 0.7, 0.9, 1.2, 1.4, 1.4, 1.5, 1.5, 1.6, 1.9, 2.0, 2.1, 2.1, 2.3, 2.5, 2.6, 2.7, 3.0, 3.8, 3.9, 4.0.

EXERCICE 9
Soient $(X_1, \ldots, X_n)$ un n-échantillon d'une loi continue μ et $(Y_1, \ldots, Y_m)$ un m-échantillon d'une loi ν, indépendant de $(X_1, ..., X_n)$. On note $\mathbb{F}_n$ la fonction de répartition empirique de $(X_1, ..., X_n)$, $\mathbb{G}_m$ la fonction de répartition empirique de $(Y_1, ..., Y_m)$ et on définit

$$D_{n,m} = \sup_{t \in \mathbb{R}} |\mathbb{F}_n(t) - \mathbb{G}_m(t)|.$$

On s'intéresse aux hypothèses H_0 : "$\mu = \nu$" et H_1 : "$\mu \neq \nu$".

1. Montrer que si les X_i et les Y_j ont la même loi, alors la loi de $D_{n,m}$ est libre de μ et ν.
2. En utilisant le fait que la loi de $D_{n,m}$ est connue sous H_0 (cette loi est tabulée), construire un test de H_0 contre H_1.
3. On souhaite comparer deux médicaments censés soulager la douleur post-opératoire. On a observé sur 16 patients dont 8 ont pris un médicament A habituel et les 8 autres un médicament B expérimental, les nombres suivants d'heures de soulagement. Y a-t-il une différence significative au niveau 5% entre A et B ?

A	6,8	3,1	5,8	4,5	3,3	4,7	4,2	4,9
B	4,4	2,5	2,8	2,1	6,6	0,0	4,8	2,3

Exercice 10

Soit X une variable aléatoire de fonction de répartition inconnue F, et soit $(X_1, ..., X_n)$ un n-échantillon de la loi de X. Pour tout $\theta > 0$, on note F_θ la fonction de répartition définie par $F_\theta(x) = (1 - \exp(-x/\theta))\ \mathbb{1}_{x>0}$.

1. On suppose que $F \in \mathcal{F}$, où $\mathcal{F} = \{F_\theta,\ \theta > 0\}$. Déterminer l'estimateur T du maximum de vraisemblance de θ, puis construire un test de $\theta = 100$ contre $\theta \neq 100$ au niveau 5%.
2. On note $\mathbb{F}_n$ la fonction de répartition empirique de $(X_1, ..., X_n)$, et on définit

$$\Delta_n = \sup_{t\in\mathbb{R}} |\mathbb{F}_n(t) - F_T(t)|.$$

Montrer que la loi de Δ_n est libre de θ lorsque $F \in \mathcal{F}$. En déduire un test de l'hypothèse $F \in \mathcal{F}$ contre $F \notin \mathcal{F}$. Si $n = 5$ et que l'on a observé les valeurs 133, 169, 8, 122 et 58, tester de deux manières l'hypothèse $F = F_{100}$ au niveau 10%.

Exercice 11

Partant de races pures, Bauer a croisé des mufliers ivoires avec des mufliers rouges. A la deuxième génération, après autofécondation des plantes de la première, il a obtenu 22 mufliers rouges, 52 pâles et 23 ivoires. Si la couleur des fleurs est gérée par un couple d'allèles, la probabilité théorique d'obtenir une fleur rouge (resp. pâle, resp. ivoire) est de 1/4 (resp. 1/2, resp. 1/4). Que conclure ?

Statistiques
2010-2011

Feuille de TD n°3
Modèles linéaires gaussiens

Remarque : les données se trouvent dans le dossier :
`/home/doc/rossigno/M1MFA/`

Exercice 12

Soient n_1 et n_2 des entiers positifs. On observe Y_{ij} $(i = 1, 2,\ j = 1, ..., n_i)$ où pour tout (i, j), Y_{ij} est de loi $\mathcal{N}(\mu_i, \sigma_i^2)$, μ_i et σ_i étant inconnus. Les variables Y_{ij} sont supposées indépendantes.

1. Tester l'égalité des variances σ_1^2 et σ_2^2.
2. On suppose ici les variances égales.
   (a) Ecrire le modèle sous forme matricielle.
   (b) Estimer les paramètres du modèle.
   (c) Construire le test de Fisher de l'hypothèse H_0 : "$\mu_1 = \mu_2$" contre H_1 : "$\mu_1 \neq \mu_2$".
3. On traite 12 parcelles de terrain identiques avec un engrais A et 12 autres avec un nouvel engrais B. On obtient avec A une production moyenne de 4,8 quintaux avec un écart-type empirique observé égal à 0,36. Les quantités correspondantes pour B valent 5,1 et 0,4. On modélise cette expérience comme ci-dessus. Tester l'égalité des variances et des espérances pour A et B, au niveau 10%.

Exercice 13

Soit p un entier positif, et pour tout $i = 1, ..., p$, soit n_i un entier positif. On suppose le modèle linéaire suivant régulier.

$$Y_{ij} = a_i + b_i x_{ij} + \varepsilon_{ij} \quad i = 1, ..., p, \quad j = 1, ..., n_i \tag{1}$$

Les x_{ij} sont des réels fixés et les ε_{ij} sont des variables aléatoires indépendantes entre elles, de loi $\mathcal{N}(0, \sigma^2)$. Les paramètres a_i, b_i $(i = 1, ..., p)$ et σ sont inconnus.

1. Estimer les paramètres du modèle.
2. Tester l'égalité des droites de régression (d'équations respectives $y = a_i + b_i x$).
3. On suppose les droites de régression identiques. Le modèle s'écrit alors $Y_{ij} = a + b x_{ij} + \varepsilon_{ij}$.
   (a) Déterminer les estimateurs du maximum de vraisemblance des paramètres.
   (b) Construire une région de confiance avec coefficient de sécurité $1 - \alpha$ $(\alpha \in]0, 1[)$ pour le paramètre (a, b).
   (c) Tester l'hypothèse "la droite de régression passe par le point (x_0, y_0)", où x_0 et y_0 sont des réels fixés.
4. Application :
   Des photographies aériennes de champs d'orge sont analysées au photomètre, qui mesure la brillance de chaque champ. On recherche la relation entre les rendements des parcelles et la brillance, en tenant compte de l'application d'un fongicide sur certaines parcelles pour combattre le mildiou. On considère le modèle linéaire décrit ci-dessus en (1), où x_{ij} désigne la brillance du $j$ème champ soumis au fongicide i, et Y_{ij} représente le rendement de ce champ. Les données sont dans le fichier `fongicide.dat`, sous la forme `(rendement, brillance, fongicide)`.

(a) Tracer sur un même graphe les rendements en fonction de la brillance, en distinguant avec deux couleurs différentes selon le groupe (i.e avec ou sans fongicide).

(b) Vérifier rapidement l'allure des résidus.

(c) Tester l'absence d'effet du fongicide sur le rendement.

(d) Estimer les paramètres du modèle (donner des intervalles de confiances).

Exercice 14

Soient p, q et r des entiers strictement supérieurs à 1. Pour tout $i \in \{1, ..., p\}$, $j \in \{1, ..., q\}$ et $k \in \{1, ..., r\}$ (k est un indice de répétition), on observe $Y_{i,j,k} = m_{i,j} + \varepsilon_{i,j,k}$, où les variables $\varepsilon_{i,j,k}$ sont indépendantes, de loi $\mathcal{N}(0, \sigma^2)$, σ^2 étant inconnu. On considère un modèle d'analyse de la variance à deux facteurs complet et équilibré, c'est-à-dire qu'on suppose qu'il existe μ, α_i, β_j et $\gamma_{i,j}$ (inconnus) tels que

$$Y_{i,j,k} = \mu + \alpha_i + \beta_j + \gamma_{i,j} + \varepsilon_{i,j,k}.$$

On suppose de plus la contrainte $\mathcal{C}$ suivante satisfaite :

$$\mathcal{C} \ : \quad \begin{cases} \sum_i \alpha_i = 0 \\ \sum_j \beta_j = 0 \\ \sum_j \gamma_{i,j} = 0, \quad \forall i \in \{1, ..., p\} \\ \sum_i \gamma_{i,j} = 0, \quad \forall j \in \{1, ..., q-1\} \end{cases}$$

1. Dans un modèle linéaire non régulier $Y = X\beta + \varepsilon$, on dit que $L\beta = 0$ est une contrainte d'identifiabilité si L est une application linéaire et si pour tout m de l'image de X, il existe un unique β tel que $X\beta = m$ et $L\beta = 0$. Vérifier que $\mathcal{C}$ est une contrainte d'identifiabilité.
2. Estimer les paramètres du modèle.
3. Tester les hypothèses suivantes :
   (a) absence d'effet du facteur j (i.e. $\beta_j = 0$ et $\gamma_{ij} = 0$ pour tout (i, j))
   (b) absence d'effet du facteur i (i.e. $\alpha_i = 0$ et $\gamma_{ij} = 0$ pour tout (i, j)).
   (c) absence d'interaction (i.e. $\gamma_{ij} = 0$ pout tout (i, j)).
4. Application :
   Une expérience est destinée à étudier l'adaptation de deux variétés de moutarde à la sécheresse. On a un dispositif à 4 traitements différents dérivant de 2 variétés (Clause et Gisilba) notées A et B et de 2 intensités lumineuses (29000 lux et 8000 lux) notées X et Y. On dispose de 4 plants pour chacun des traitements AX, AY, BX et BY. Un indicateur de l'adaptation à la sécheresse est l'apparition de racines courtes tubérisées. Le tableau suivant indique le nombre observé de ces racines.

AX :	78	79	44	77
AY :	64	96	30	20
BX :	137	85	302	315
BY :	64	67	102	47

   Les données sont dans le fichier `moutarde.dat`, sous la forme :
   `(moutarde, lumière, nb de tubercules)`.
   (a) Vérifier rapidement l'allure des résidus.
   (b) Tester, au niveau 5%, l'absence d'interaction entre les deux facteurs.
   (c) Tester l'absence d'effet du facteur variété sur l'adaptation à la sécheresse.

EXERCICE 15

On veut comparer quatre variétés de céréales sur trois types de terrains. Les mesures Y_{ijk} , $1 \leqslant k \leqslant n_{ij}$ de rendement de la variété i de céréale sur le terrain j sont dans le fichier `cereales.dat`, sous la forme `(variété, terrain,rendement)`. On considère un modèle complet d'analyse de la variance à deux facteurs.

1. Vérifier rapidement l'allure des résidus.
2. Tester l'absence d'effet du facteur terrain.
3. Tester l'absence d'effet du facteur variété.

Statistiques
2010-2011

Feuille de TD n°4

Affinités et risque minimax

Exercice 16

Pour tout couple de réels (μ, μ'), on note $h(\mu, \mu')$ la distance de Hellinger entre la loi gaussienne d'espérance μ et de variance 1 et la loi gaussienne d'espérance μ' et de variance 1. On note $\rho(\mu, \mu')$ l'affinité de Hellinger entre ces deux lois.

1. Montrer que pour tout couple de réels (μ, μ'),

$$\rho(\mu, \mu') = \exp\left(-\frac{1}{8}(\mu - \mu')^2\right).$$

2. Soient $\mu \in \mathbb{R}$, $\lambda \in]0, 1[$ et $c_n \in \mathbb{R}$. On souhaite tester H_0 : "$\theta = \mu$" contre H_1 : "$\theta = \mu + c_n$" sur la base de l'observation d'un n-échantillon $(X_1, ..., X_n)$ de la loi $\mathcal{N}(\theta, 1)$. On rappelle qu'alors,

$$1 - \sqrt{2n}h(\mu, \mu + c_n) \leqslant \inf_{\varphi_n}\{E_\mu(\varphi_n) + E_{\mu+c_n}(1 - \varphi_n)\} \leqslant \exp(-nh^2(\mu, \mu + c_n)),$$

où l'inf est étendu à tous les tests φ_n.

   (a) Montrer que s'il existe un test φ_n tel que $E_\mu(\varphi_n) + E_{\mu+c_n}(1 - \varphi_n) \leqslant \lambda$ alors il existe $c_\lambda > 0$ tel que $|c_n| \geqslant c_\lambda n^{-1/2}$.

   (b) Réciproquement, soit $\lambda > 0$. On suppose qu'il existe $c > 0$ tel que $|c_n| \geqslant cn^{-1/2}$. Montrer que si c et n sont assez grands, alors il existe un test φ_n tel que $E_\mu(\varphi_n) + E_{\mu+c_n}(1 - \varphi_n) \leqslant \lambda$.

Exercice 17

Soient μ et μ' deux réels strictement positifs. On note $\rho(\mu, \mu')$ l'affinité de Hellinger entre la loi uniforme sur l'intervalle $[0, \mu]$ et la loi uniforme sur l'intervalle $[0, \mu']$.

1. Montrer que

$$\rho(\mu, \mu') = \left(1 - \frac{|\mu - \mu'|}{\mu \vee \mu'}\right)^{1/2}.$$

2. Soient $\mu > 0$, $\lambda \in]0, 1[$ et $c_n \in \mathbb{R}$. On souhaite tester H_0 : "$\theta = \mu$" contre H_1 : "$\theta = \mu + c_n$" sur la base de l'observation d'un n-échantillon $(X_1, ..., X_n)$ de la loi uniforme sur l'intervalle $[0, \theta]$. Donner une condition nécessaire puis une condition suffisante sur c_n pour qu'il existe un test φ_n tel que $E_\mu(\varphi_n) + E_{\mu+c_n}(1 - \varphi_n) \leqslant \lambda$.

Exercice 18

Soient w une application bijective croissante et concave de $[0, 1]$ sur $[0, 1]$ et $\mathcal{H}_w$ la famille des fonctions f définies sur $[-1, 1]$ telles que $|f(t) - f(0)| \leqslant w(|t|)$ pour tout $t \in [-1, 1]$. On considère le modèle suivant

$$Y_i = f(i/n) + \varepsilon_i, \quad i \in \mathbb{Z}(n) = \mathbb{Z} \cap [-n, n],$$

où f est une fonction inconnue définie sur $[-1,1]$ dont on sait qu'elle appartient à $\mathcal{H}_w$, et $\varepsilon_{-n},\dots,\varepsilon_n$ sont indépendantes et de même loi normale centrée et réduite. On souhaite étudier le risque quadratique minimax $R_n(\mathcal{H}_w)$ pour l'estimation de $f(0)$ sur la base des observations $Y_{-n},\dots,Y_n$:

$$R_n(\mathcal{H}_w) = \inf_T \sup_{f\in\mathcal{H}_w} E_f\left[(T-f(0))^2\right],$$

où l'inf est étendu à tous les estimateurs T.

1. On note P_f^n la loi du $(2n+1)$-uplet $(Y_{-n},\dots,Y_n)$ donné par le modèle ci-dessus. Montrer que :

$$R_n(\mathcal{H}_w) \geqslant \sup_{f,g\in\mathcal{H}_w} \left(\frac{f(0)-g(0)}{2}\right)^2 \left(1-\sqrt{\frac{1}{4}\sum_{i\in\mathbb{Z}(n)} (f(i/n)-g(i/n))^2}\right).$$

2. Pour tout entier positif $H \leqslant n$, on définit la fonction f_H sur $[-1,1]$ par

$$f_H(t) = \begin{cases} w(H/n) \text{ si } |nt| \leqslant H \\ 0 \text{ sinon} \end{cases}$$

On définit $H(n) = \sup\{H \leqslant n \ : \ (2H+1)w^2(H/n) \leqslant 1\}$. On suppose n assez grand pour que $w^2(1/n) \leqslant 1/3$.

(a) Vérifier que $1 \leqslant H(n) < n$.

(b) Vérifier que $f_H \in \mathcal{H}_w$ pour tout entier positif $H \leqslant n$. En déduire que

$$R_n(\mathcal{H}_w) \geqslant \frac{1}{8} w^2(H(n)/n).$$

(c) Soit $0 < \alpha \leqslant 1$ et $w(t) = t^\alpha$, $t \in [0,1]$. Démontrer que $\liminf_n \left[n^{\frac{2\alpha}{2\alpha+1}} R_n(\mathcal{H}_w)\right] > 0$.

3. Soient $0<\alpha\leqslant 1$ et $w(t) = t^\alpha$, $t\in[0,1]$. Soit $(h_n)_{n\in\mathbb{N}}$ une suite de nombres réels positifs tels que :

$$h_n \underset{n\infty}{\longrightarrow} 0 \quad \text{et} \quad nh_n \underset{n\infty}{\longrightarrow} \infty .$$

On pose $x_i = \frac{i}{n}$ et :

$$\widehat{f}(0) = \frac{1}{2\lfloor nh_n\rfloor + 1} \sum_{i=-n}^{n} Y_i \ \mathbb{1}_{|x_i|\leqslant h_n} .$$

Trouver $(h_n)_{n\in\mathbb{N}}$ de sorte que $\widehat{f}(0)$ soit un estimateur minimax, à une constante multiplicative près, de $f(0)$.

Statistiques
2010-2011

FEUILLE DE TD n^{o}5
ESTIMATEURS BAYESIENS

EXERCICE 19

Soit $(X_1, \ldots, X_n)$ un n-échantillon de la loi $\mathcal{N}(0, \sigma^2)$ et soit $\theta = \sigma^{-2}$ le paramètre à estimer. On considère la loi *a priori*

$$\nu(d\theta) = C_{a,h} e^{-a\theta}\, \theta^{h-1} \;\; \mathbb{1}_{\{\theta>0\}}\, d\theta$$

(ν est la loi $\Gamma(h, a)$).

1. Déterminer la loi *a posteriori* relative à ν.
2. Calculer l'estimateur bayesien T de θ relatif à ν et au risque quadratique.
3. L'estimateur T est-il admissible ?

EXERCICE 20

Soit $(X_1, ..., X_n)$ un n-échantillon de loi de Poisson de paramètre θ inconnu.

1. Calculer l'espérance et la fonction de risque quadratique de $\overline{X}$.
2. Calculer l'estimateur de Bayes T de θ relatif au risque quadratique et à la loi *a priori* $\Gamma(h, a)$.
3. Comparer les fonctions de risque de $\overline{X}$ et T.
4. Montrer que T est admissible.

EXERCICE 21

Soient $(X_1, ..., X_n)$ un n-échantillon de loi uniforme sur $\{1, 2, \ldots, \theta\}$ et $p \in]0, 1[$.

1. Calculer l'estimateur de Bayes T de θ relatif au risque quadratique et à la loi *a priori* définie par : $\forall \theta \in \mathbb{N}^*$, $\nu(\theta) = C_p \theta^n p^\theta$.
2. Cet estimateur est-il admissible ?

EXERCICE 22

Soit X une variable aléatoire réelle et $\alpha \in]0, 1[$. On pose, pour tout $a \in \mathbb{R}$:

$$F_\alpha(a) = \alpha \mathbb{E}[(X - a)_+] + (1 - \alpha)\mathbb{E}[(X - a)_-] \; .$$

1. Montrer que pour tout $a \in \mathbb{R}$,

$$F_\alpha(a) = \mathbb{E}[(X - a)_+] + (1 - \alpha)\mathbb{E}[a - X] \; ,$$

et

$$F_\alpha(a) = \mathbb{E}[(X - a)_-] + \alpha \mathbb{E}[X - a] \; .$$

2. Montrer que pour tous réels $a \leqslant b$,

$$F_\alpha(a) - F_\alpha(b) = \mathbb{E}[(X-a)\ \mathbb{1}_{X\in]a,b[}] + (b-a)(\mathbb{P}(X \geqslant b) - (1-\alpha))\ ,$$

et

$$F_\alpha(b) - F_\alpha(a) = \mathbb{E}[(b-X)\ \mathbb{1}_{X\in]a,b[}] + (b-a)(\mathbb{P}(X \leqslant a) - \alpha)\ .$$

3. On dit que $q \in \mathbb{R}$ est un quantile d'ordre α de X si et seulement si :

$$\mathbb{P}(X \leqslant q) \geqslant \alpha \quad \text{et} \quad \mathbb{P}(X \geqslant q) \geqslant 1-\alpha\ .$$

Montrer que :

$$F_\alpha(q) = \min_{a\in\mathbb{R}} F_\alpha(a) \quad \Longleftrightarrow \quad q \text{ est un quantile d'ordre } \alpha \text{ de } X\ .$$

4. On dit que m est une médiane de X si c'est un quantile d'ordre $1/2$ de X. En déduire que les médianes de X sont les réels qui minimisent $a \mapsto \mathbb{E}[|X-a|]$.

Exercice 23

Soit $(X_1, ..., X_n)$ un n-échantillon de la loi $\mathcal{N}(\theta, 1)$, $\theta \in \mathbb{R}$. Calculer l'estimateur de Bayes T_σ de θ relatif à la loi *a priori* $\mathcal{N}(0, \sigma^2)$ et à la perte $l(\theta, \theta') = |\theta - \theta'|$.

Exercice 24

On observe X, dont la loi est de la forme P_θ pour un paramètre $\theta \in \Theta$ inconnu. On souhaite estimer $g(\theta)$ sur la base de l'observation X, où g est définie sur Θ et à valeurs réelles. On considère la fonction de perte l et la loi *a priori* ν sur Θ.

1. On suppose qu'il existe un estimateur bayesien de risque constant, relatif à ν et l. On note $T(X)$ cet estimateur.
   (a) Montrer que $T(X)$ est minimax relativement à la perte l.
   (b) Montrer que ν est la loi *a priori* "la moins favorable", c'est-à-dire que si μ désigne une autre loi *a priori*,

   $$\inf_U \mathcal{R}^\nu(U) \geqslant \inf_U \mathcal{R}^\mu(U),$$

   où l'inf est étendu à l'ensemble des estimateurs $U(X)$ et où pour toute loi *a priori* τ,

   $$\mathcal{R}^\tau(U) = \int_\Theta \mathbb{E}_\theta\left[l(g(\theta), U(X))\right] d\tau(\theta).$$

2. Dans cette question, $X = (X_1, \ldots, X_n)$ est un n-échantillon de la loi de Bernoulli de paramètre inconnu $\theta \in]0,1[$. On considère la fonction de perte définie par

$$l(\theta, a) = \frac{(\theta-a)^2}{\theta(1-\theta)}$$

et on note $\bar{X} = \sum_{i=1}^n X_i/n$. On souhaite estimer θ.
   (a) Montrer que $\bar{X}$ est bayesien relativement à l et à la loi *a priori* uniforme sur $]0,1[$.
   (b) Montrer que $\bar{X}$ est de risque constant.
   (c) En déduire que $\bar{X}$ est minimax et que la loi uniforme sur $]0,1[$ est la loi *a priori* la moins favorable.

Indication : Pour la question 2(a), on pourra utiliser la propriété suivante. Si pour tout $a > 0$ et tout $b > 0$, on note $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\, dx$, alors

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

où pour tout $b > 0$, $\Gamma(b) = \int_0^\infty t^{b-1}e^{-t}\, dt$.

Statistiques
2010-2011

Feuille de TD n°6
Théorie de Neyman-Pearson

Exercice 25

Soit X une variable aléatoire de loi $\mathcal{B}(n, p)$.

1. Soient p_0 et p_1 des réels fixés dans $]0, 1[$ tels que $p_0 < p_1$. Bâtir un test de Neyman-Pearson de niveau α de $p = p_0$ contre $p = p_1$.
2. Soit $p_0 \in]0, 1[$. Bâtir un test uniformément le plus puissant parmi les tests de niveau α de $p \leqslant p_0$ contre $p > p_0$.

Exercice 26

Soient $\alpha \in]0, 1[$ et $(X_1, ..., X_n)$ un n-échantillon de la loi $\mathcal{N}(\theta, 1)$, $\theta \in \mathbb{R}$. Bâtir un test uniformément le plus puissant parmi les tests de niveau α de $\theta \geqslant \theta_0$ contre $\theta < \theta_0$, puis déterminer la puissance de ce test.

Exercice 27

Soient $\tau_1, \tau_2, \ldots, \tau_m, \ldots$ des variables aléatoires indépendantes et de même loi exponentielle de paramètre θ inconnu. Pour tout $n \in \mathbb{N}$, on note $T_n = \tau_1 + \cdots + \tau_n$. Bâtir un test uniformément le plus puissant de $\theta \leqslant \theta_0$ contre $\theta > \theta_0$ lorsqu'on observe

1. $N_t = \sum_{n \geqslant 1} \mathbb{1}_{(T_n \leqslant t)}$ pour un $t > 0$ fixé (on montrera que N_t suit une loi de Poisson de paramètre θt) ;
2. $(T_1, \ldots, T_m)$, où m est un entier fixé ;

Exercice 28

Soient $\tau_1, \ldots, \tau_n$ les variables aléatoires associées à la durée de vie de n ampoules. On suppose que $\tau_1, \ldots, \tau_n$ sont indépendantes et de même loi exponentielle dont le paramètre, noté θ, est inconnu. On observe seulement $\tau_{(1)}, \ldots, \tau_{(m)}$, où $m \leqslant n$ est un entier fixé et $\tau_{(1)} < \cdots < \tau_{(n)}$ sont les statistiques d'ordre associées à l'échantillon. On note $\tau_{(0)} = 0$ et pour tout $i = 1, \ldots, n$,

$$X_i = \tau_{(i)} - \tau_{(i-1)}.$$

Enfin, on note

$$S_m = \tau_{(1)} + \cdots + \tau_{(m-1)} + (n - m + 1)\tau_{(m)}.$$

1. (a) Vérifier que $S_m = \displaystyle\sum_{i=1}^{m} (n - i + 1) X_i$.

   (b) Déterminer la loi de $(X_1, \ldots, X_n)$ et en déduire que $2\theta S_m$ suit une loi $\chi^2(2m)$.
2. Construire un test uniformément le plus puissant parmi les tests de niveau α de $\theta \geqslant \theta_0$ contre $\theta < \theta_0$.

Exercice 29

Soient $X = (X_1, ..., X_n)$ un n-échantillon de la loi uniforme sur $[0, \theta]$ et $\alpha \in]0, 1[$.

1. Soit φ un test de $\theta = \theta_0$ contre $\theta > \theta_0$. Montrer que φ est uniformément le plus puissant parmi les tests de niveau α si et seulement si $\mathbb{E}_{\theta_0}\varphi(X) = \alpha$ et $\varphi(X) = 1$ lorsque $\max(X_1, \ldots, X_n) > \theta_0$.
2. Montrer qu'il existe un unique test uniformément le plus puissant de $\theta = \theta_0$ contre $\theta \neq \theta_0$ au niveau α, donné par $\varphi(X) = 1$ lorsque $\max(X_1, \ldots, X_n) > \theta_0$ ou $\max(X_1, \ldots, X_n) \leqslant \theta_0\alpha^{1/n}$ et $\varphi(X) = 0$ sinon.

Exercice 30

Soit $(X_1, ..., X_n)$ un n-échantillon de la loi admettant la densité $x \mapsto a \exp(-a(x - b))\ \mathbb{1}_{(x \geqslant b)}$ par rapport à la mesure de Lebesgue.

1. On suppose que a est connu.
   (a) Déterminer la loi de $Y_i = \exp(-aX_i)$.
   (b) En utilisant l'exercice 29, montrer qu'il existe un unique test uniformément le plus puissant de $b = b_0$ contre $b \neq b_0$; décrire ce test.
2. Montrer qu'il existe un test uniformément le plus puissant de $a = a_0$, $b = b_0$ contre $a > a_0$, $b < b_0$; décrire ce test.

Statistiques
2010-2011

Feuille de TD n°7
Exhaustivité, efficacité

Exercice 31

1. Soit $(X_1, \dots, X_n)$ un n-échantillon de la loi de Bernoulli de paramètre inconnu $\theta \in]0,1[$. Déterminer un estimateur uniformément de variance minimale parmi les estimateurs sans biais de θ.
2. Soit $(X_1, \dots, X_n)$ un n-échantillon de la loi $\mathcal{N}(m, \sigma^2)$. Déterminer un estimateur uniformément de variance minimale parmi les estimateurs sans biais de m (resp. de σ^2).

Exercice 32

Soit $(X_1, ..., X_n)$ un n-échantillon de la loi uniforme sur $[0, \theta]$, $\theta > 0$.

1. Donner une statistique exhaustive et complète pour θ puis construire un estimateur uniformément de variance minimale parmi les estimateurs sans biais de θ.
2. En déduire sans calcul la valeur de $\mathbb{E}_\theta(\bar{X} \,|\, \max_{1 \leqslant i \leqslant n} X_i)$.

Exercice 33

Soit $(X_1, ..., X_n)$ un n-échantillon de la loi exponentielle de paramètre θ, $\theta > 0$. On veut estimer $\mathbb{P}_\theta(X_1 \leqslant t)$ pour $t > 0$ donné. Montrer que l'estimateur empirique $\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(X_i \leqslant t)}$ est sans biais. Déterminer un estimateur uniformément de variance minimale parmi les estimateurs sans biais.

Exercice 34

Soit $(X_1, \dots, X_n)$ un n-échantillon de la loi $\mathcal{N}(\theta, \gamma\theta^2)$, où $\gamma > 0$ est connu et le paramètre est $\theta \in \mathbb{R}$. Montrer que $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ est une statistique exhaustive non complète.

Exercice 35

Soient $\{P_\theta,\ \theta \in \Theta\}$ un modèle dominé, P une dominante privilégiée, Z une statistique et T une statistique exhaustive.

1. Montrer que si Z est indépendante de T pour P, alors c'est une statistique libre.
2. Soit $X_1, \dots, X_n$ un n-échantillon de la loi uniforme sur $[\theta - 1/2, \theta + 1/2]$, $\theta \in \mathbb{R}$.
   (a) Montrer que $X_{(n)} - X_{(1)}$ est une statistique libre.
   (b) Montrer que $(X_{(1)}, X_{(n)})$ est exhaustive.
   (c) Montrer que la réciproque de 1. est fausse.
3. Montrer que si Z est libre et T est exhaustive complète, alors Z et T sont indépendantes pour P et pour P_θ, quel que soit $\theta \in \Theta$.
4. Soit $X_1, \dots, X_n$ un n-échantillon de la loi uniforme sur $[\theta - 1/2, \theta + 1/2]$, $\theta \in \mathbb{R}$. Vérifier que $(X_{(1)}, X_{(n)})$ est une statistique exhaustive non complète.

Exercice 36

Soit $(X_1, \dots, X_n)$ un n-échantillon de la loi F_θ. Montrer que $\bar{X}$ est un estimateur efficace de l'espérance de F_θ lorsque

1. F_θ est la loi de Bernoulli de paramètre $\theta \in]0, 1[$,
2. F_θ est la loi gaussienne $\mathcal{N}(\theta, 1)$, $\theta \in \mathbb{R}$,
3. F_θ est la loi exponentielle de paramètre $\theta > 0$.

Exercice 37

Soit $(X_1, ..., X_n)$ un n-échantillon de la loi de Poisson de paramètre inconnu $\theta > 0$.

1. Ecrire l'inégalité de Cramér-Rao pour un estimateur sans biais de $g(\theta) = \mathbb{P}_\theta(X_1 = 0)$.
2. On note $T = \mathbb{1}_{X_1=0}$. Vérifier que T est un estimateur sans biais de $g(\theta)$. En déduire qu'il existe un estimateur uniformément de variance minimale parmi les estimateurs sans biais de $g(\theta)$, puis montrer qu'il n'existe pas d'estimateur efficace de $g(\theta)$.

Exercice 38

Soit $(X_1, ..., X_n)$ un n-échantillon de la loi $\mathcal{N}(m, \sigma^2)$, où $(m, \sigma^2) \in \mathbb{R} \times]0, \infty[$ est inconnu.

1. Calculer l'information de Fisher de cet échantillon en (m, σ^2).
2. Calculer le risque quadratique de $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ pour estimer σ^2. Existe-t-il un estimateur efficace de σ^2 ?

Statistiques
2010-2011

Feuille de TD n°8
Efficacité asymptotique

Exercice 39
Soit $(P_\theta)_{\theta\in\Theta}$ une famille de probabilités distinctes sur un espace $\mathcal{X}$, où Θ est un intervalle ouvert non vide de $\mathbb{R}$. On considère $\phi : \mathcal{X} \to \mathbb{R}$ telle que $\mathbb{E}_\theta(\phi^2) < \infty$ pour tout θ, et on suppose que la fonction

$$\theta \mapsto g(\theta) := \mathbb{E}_\theta(\phi)$$

est dérivable sur Θ, de dérivée strictement positive.

On observe des variables aléatoires indépendantes et de même loi $X_1, \dots, X_n$, dont la loi inconnue appartient à la famille $(P_\theta)_{\theta\in\Theta}$. On estime $g(\theta)$ par

$$T_n = \frac{1}{n}\sum_{i=1}^{n} \phi(X_i)\ .$$

1. Quelle méthode a-t-on employée pour bâtir T_n ?
2. Montrer que pour tout θ, $\lim_{n\to\infty} \mathbb{P}_\theta(T_n \in g(\Theta)) = 1$.

   Dans la suite, on suppose (pour simplifier) que $T_n \in g(\Theta)$ et on estime θ par $\widehat{\theta}_n = g^{-1}(T_n)$.
3. Montrer que $\widehat{\theta}_n$ est un estimateur consistant de θ.
4. Montrer que quand n tend vers l'infini,

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{Var_\theta(\phi)}{g'(\theta)^2}\right)\ .$$

Exercice 40
Soit $(X_1, ..., X_n)$ un n-échantillon de la loi $\mathcal{N}(m, \sigma^2)$, où $\theta = (m, \sigma^2) \in \mathbb{R}\times]0, \infty[$ est inconnu. On note $\widehat{\sigma}_n^2$ l'estimateur du maximum de vraisemblance de σ^2 et $R_\theta(\widehat{\sigma}_n^2)$ son risque quadratique en θ. Montrer que

$$\sqrt{n}(\widehat{\sigma}_n^2 - \sigma^2) \xrightarrow{\mathcal{L}_\theta} \mathcal{N}(0, 2\sigma^4) \quad \text{quand } n \to \infty.$$

et que de plus, $\lim_{n\to\infty} nR_\theta(\widehat{\sigma}_n^2) = 2\sigma^4$.

Exercice 41
Soit $(X_1, ..., X_n)$ un n-échantillon d'une loi de Poisson de paramètre $\theta > 0$ inconnu. On souhaite estimer $g(\theta) = \exp(-k\theta)$, où k est un entier positif fixé.

1. Construire un estimateur asymptotiquement efficace de $g(\theta)$.
2. Soit $T_n = \left(1 - \frac{k}{n}\right)^{S_n}$, où $S_n = \sum_{i=1}^{n} X_i$.
   (a) Montrer que T_n est uniformément de variance minimale parmi les estimateurs sans biais de $g(\theta)$.
   (b) Montrer que T_n est asymptotiquement efficace.

(c) T_n est-il efficace ?

Exercice 42

Soit f une densité de probabilité de classe $\mathcal{C}^2$ sur $\mathbb{R}$, telle que

- pour tout $x \in \mathbb{R}$, $f(x)$ est strictement positif,
- les fonctions f'' et $(f')^2/f$ sont intégrables par rapport à la mesure de Lebesgue.
- la fonction $f''/f - (f'/f)^2$ est uniformément continue sur $\mathbb{R}$.

Pour tout $\theta \in \mathbb{R}$ et $x \in \mathbb{R}$, on note

$$f_\theta(x) = f(x-\theta) \quad \text{et} \quad l_\theta(x) = \log f_\theta(x).$$

Pour tout $x \in \mathbb{R}$, la fonction $\theta \mapsto l_\theta(x)$ est alors de classe $\mathcal{C}^2$ sur $\mathbb{R}$; on note $l'_\theta(x)$ et $l''_\theta(x)$ ses dérivées première et seconde au point θ. On note de même $f'_\theta(x)$ la dérivée première de $\theta \mapsto f_\theta(x)$ au point θ. Enfin, on note X une variable aléatoire admettant la densité f_{θ_0}, $\theta_0 \in \mathbb{R}$.

1. Montrer les propriétés suivantes :
   (a) $\int_{\mathbb{R}} f''_{\theta_0}(x)\,dx = 0$,
   (b) Il existe un voisinage $\mathcal{V}$ de θ_0 et une fonction mesurable réelle h dans $L^1(P_{\theta_0})$ telle que pour tout $\theta \in \mathcal{V}$ et tout $x \in \mathbb{R}$, $|l''_\theta(x)| \leqslant h(x)$,
   (c) L'information de Fisher de l'observation X est finie et strictement positive. Elle ne dépend que de f et est égale à
$$I(f) = \int \frac{(f'(x))^2}{f(x)}\,dx.$$
2. Pour tout entier positif n, on note $\widehat{\theta}_n$ un estimateur du maximum de vraisemblance de θ calculé sur la base d'un n-échantillon de la loi de X (on suppose l'existence d'un tel estimateur). On note g une fonction réelle (donnée) dérivable au voisinage de θ_0. Déduire de la question précédente que si $\widehat{\theta}_n$ converge en probabilité vers θ_0 lorsque n tend vers l'infini, alors $g(\widehat{\theta}_n)$ est un estimateur asymptotiquement efficace en θ_0 de $g(\theta_0)$.
3. On suppose ici que $\sigma^2 = \int x^2 f(x)\,dx$ est finie et que $\int x f(x)\,dx = 0$.
   (a) Vérifier que θ_0 est l'espérance de X.
   (b) Sur la base d'un échantillon de la loi de X, construire un estimateur asymptotiquement efficace de θ_0.

Exercice 43

Soit $(\xi_1, ..., \xi_n)$ un n-échantillon d'une loi admettant une densité strictement positive par rapport à la mesure de Lebesgue, de fonction de répartition $F(\;.\; - \theta)$, où $\theta \in \mathbb{R}$ est inconnu et F est connue et de classe $\mathcal{C}^2$. A chaque tirage, on observe seulement

$$X_i = \left(\, \mathbb{1}_{]-\infty,a]}(\xi_i),\ \mathbb{1}_{]a,b]}(\xi_i),\ \mathbb{1}_{]b,\infty[}(\xi_i)\right).$$

On souhaite construire un estimateur asymptotiquement efficace de θ sur la base des observations $(X_1, ..., X_n)$.

1. Calculer l'information de Fisher de l'échantillon $(X_1, ..., X_n)$.
2. Soit $S_a = \sum_{i=1}^n \mathbb{1}_{X_i=(1,0,0)}$. On considère l'estimateur $\widetilde{\theta}_n = a - F^{-1}(S_a/n)$. Cet estimateur est-il asymptotiquement efficace ?
3. Construire un estimateur asymptotiquement efficace de θ.

Statistiques
2010-2011

Feuille de TD n^{o}9

Tests du χ^2 asymptotiques

Exercice 44

Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi de Poisson de paramètre $\theta > 0$. Construire un test du rapport de vraisemblance de niveau asymptotique $\alpha \in]0,1[$ de "$\theta = \theta_0$" contre "$\theta \neq \theta_0$".

Exercice 45

On observe deux caractères X et Y, où X (resp. Y) est à valeurs dans $\{1, ..., r\}$ (resp. $\{1, ..., s\}$).
On souhaite tester l'indépendance des variables X et Y, sur la base d'un n-échantillon $((X_1, Y_1), ..., (X_n, Y_n))$ de la loi du couple (X, Y). Pour tout couple $(i, j) \in \{1, ..., r\} \times \{1, ..., s\}$, on note N_{ij} le nombre d'observations égales à (i, j), $N_{io} = \sum_{j=1}^{s} N_{ij}$ et $N_{oj} = \sum_{i=1}^{r} N_{ij}$.

1. Montrer que si X et Y sont indépendantes, alors
$$n \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(N_{ij} - N_{io}N_{oj}/n)^2}{N_{io}N_{oj}} \xrightarrow{\mathcal{L}} \chi^2((r-1)(s-1)) \quad \text{quand } n \to \infty,$$
puis construire un test de l'hypothèse H_0 : "X et Y sont indépendantes".
2. Application : On a classé 217 enfants d'après leurs performances dans des tests de langage (L) et d'équilibre physique (E). Tester l'hypothèse de l'indépendance des performances de langage et d'équilibre.

	L1	**L2**	**L3**
E1	45	26	12
E2	32	50	21
E3	4	10	17

Exercice 46

Soit (X, Y) un couple de variables aléatoires à valeurs dans $\{1, ..., k\}^2$. On souhaite tester la symétrie de la loi de (X, Y), sur la base d'un n-échantillon $((X_1, Y_1), ..., (X_n, Y_n))$ de la loi du couple (X, Y). Pour tout couple $(i, j) \in \{1, ..., k\}^2$, on note N_{ij} le nombre d'observations égales à (i, j).

1. Montrer que si la loi de (X, Y) est symétrique, alors
$$\sum_{1 \leqslant i < j \leqslant k} \frac{(N_{ij} - N_{ji})^2}{N_{ij} + N_{ji}} \xrightarrow{\mathcal{L}} \chi^2(k(k-1)/2) \quad \text{quand } n \to \infty,$$
puis construire un test de l'hypothèse H_0 :"la loi de (X, Y) est symétrique".
2. Application : Le degré de vision de chacun des deux yeux de 7477 femmes agées de 30 à 40 ans a été classé en quatre groupes désignés par 1, 2, 3 et 4 par ordre croissant du pire au meilleur. Tester la symétrie puis l'indépendance du degré de vision de l'oeil droit et de l'oeil gauche sur la base des observations suivantes (les abréviations G et D font référence respectivement aux yeux gauches et droits) :

	G1	**G2**	**G3**	**G4**
D1	1520	266	124	66
D2	234	1512	432	78
D3	117	362	1772	205
D4	36	82	179	492