Fiche d'exercices n°1

Exercice 1. Une étude a été faite sur 320 familles de 5 enfants pour déterminer la répartition entre filles et garçons. Les résultats sont les suivants :

nombre de filles	0	1	2	3	4	5
nombre de familles	18	56	90	88	53	15

Ecrire le modèle statistique correspondant, en fonction des différentes hypothèses que l'on peut faire sur ces observations.

Exercice 2. On considère un échantillon de taille n de la loi de Poisson $\mathcal{P}(\lambda)$. Déterminer une statistique exhaustive de trois façons différentes.

Exercice 3. On considère un échantillon de taille n de la loi uniforme sur $[\theta, 2\theta]$, avec $\theta > 0$. Déterminer une statistique exhaustive et la statistique de maximum de vraisemblance.

Exercice 4. La RATP veut estimer la probabilité p qu'un passager ne valide pas son titre de transport sur le réseau du métro parisien aux heures de pointe de 7h30 à 9h30. Elle effectue pour cela une enquête dans $n = 40$ stations-types, un jour de semaine moyen : des contrôleurs sont mis en place à 7h30 et comptent le nombre de passagers validant leur ticket jusqu'à l'arrivée du premier fraudeur, celui-ci étant inclus. Les résultats sont les suivants :

11	15	35	60	53	48	3	58	26	61
83	68	39	28	68	44	73	36	43	84
16	93	12	81	2	97	1	41	98	29
37	84	38	98	82	73	96	1	11	9

1. Donner le modèle statistique correspondant à ces observations.
2. Estimer p par la méthode du maximum de vraisemblance. Cet estimateur est-il sans biais ?
3. Déterminer une statistique exhaustive et complète.
4. Déterminer l'estimateur sans biais et de variance minimale de $\varphi_k(p) = p(1-p)^k$, pour $k \in \mathbb{N}$. En déduire l'estimateur sans biais et de variance minimale de p.

Fiche d'exercices n°2

Exercice 1. On considère un modèle paramétrique continu $(\mathcal{X}, \mathcal{A}, \{P_\theta\,;\, \theta \in \Theta\ \subset \mathbb{R}^d\})$. On suppose que les hypothèses de régularité du chapitre 3 du cours sont vérifiées.

Montrer que le score est centré : $E\left[Z(\theta; X)\right] = 0$. Puis montrer que la matrice d'information de Fisher $\mathcal{I}(\theta)$ vérifie :

$$\forall (j,k) \in \{1, \ldots, d\}^2,\, \mathcal{I}_{jk}(\theta) = -E\left[\frac{\partial^2}{\partial\theta_j \partial\theta_k} \ln \mathcal{L}(\theta; X)\right]$$

Exercice 2. On considère le modèle d'échantillon de taille n de la loi normale $\mathcal{N}(m, \sigma^2)$, où σ^2 est supposée connue.

1. Calculer la quantité d'information de Fisher de ce modèle.
2. Vérifier de deux façons différentes que $\bar{X}_n$ est un estimateur efficace de m.

Exercice 3. On considère le modèle d'échantillon de taille n de la loi normale $\mathcal{N}(m, \sigma^2)$.

1. Calculer le vecteur des scores et la matrice d'information de Fisher.
2. Donner des bornes inférieures pour les variances d'estimateurs sans biais de m et σ^2.
3. Donner les lois asymptotiques des estimateurs de maximum de vraisemblance de m et σ^2 : $\bar{X}_n$ et S_n^2.
4. On suppose maintenant que $m \neq 0$ et on souhaite estimer le coefficient de variation σ/m.
   a. Donner la loi asymptotique de l'estimateur de maximum de vraisemblance $S_n/\bar{X}_n$.
   b. Donner un intervalle de confiance asymptotique de seuil α pour σ/m.

Exercice 4. On considère un échantillon $X_1, \ldots, X_n$ de variables aléatoires indépendantes et de même loi normale $\mathcal{N}(m, \sigma^2)$, où σ^2 est connue. On souhaite tester H_0 : "$m = m_0$" contre H_1 : "$m \neq m_0$ ".

1. Calculer la statistique du rapport des vraisemblances maximales :
$$v(x_1, \ldots, x_n) = \frac{\mathcal{L}(m_0; x_1, \ldots, x_n)}{\sup\limits_{m \in \mathbb{R}} \mathcal{L}(m; x_1, \ldots, x_n)}$$
2. Exprimer $-2 \ln v(X_1, \ldots, X_n)$ en fonction de $\bar{X}_n, m_0, \sigma^2$ et n.
3. Construire le test du rapport des vraisemblances maximales. Montrer que, dans ce cas, ce test est exact.
4. Application. On dispose d'un échantillon de taille 20 d'une loi $\mathcal{N}(m, 1)$. La moyenne empirique de cet échantillon vaut 0.25. Peut-on en conclure avec une confiance raisonnable que l'espérance de l'échantillon n'est pas nulle ?

Fiche d'exercices n°3

Exercice 1. Soit $(X_1, \ldots, X_n)$ un échantillon d'une loi continue et $(R_1, \ldots, R_n)$ la statistique de rang associée.

1. Pour $i \in \{1, \ldots, n\}$, déterminer la loi de probabilité de R_i, son espérance et sa variance.
2. Pour $i \neq j$, calculer la covariance $\text{Cov}(R_i, R_j)$ et le coefficient de corrélation linéaire $\rho(R_i, R_j)$.

Exercice 2. Soient X_1 et X_2 deux variables aléatoires réelles indépendantes et de même loi, de fonction de répartition F. Déterminer la fonction de répartition du couple (X_1^*, X_2^*).

Exercice 3. Soit $(X_1, \ldots, X_n)$ un échantillon d'une loi continue. Pour $i < j$, déterminer la densité du couple (X_i^*, X_j^*).

Exercice 4. Soient $X_1, \ldots, X_n$ n variables aléatoires réelles indépendantes et de même loi continue de fonction de répartition F et de densité f. On suppose que cette loi est symétrique par rapport à un réel μ, c'est-à-dire que :

$$\forall x \in \mathbb{R},\ F(\mu - x) = 1 - F(\mu + x) \qquad \text{ou bien} \qquad \forall x \in \mathbb{R},\ f(\mu - x) = f(\mu + x)$$

1. Montrer que la médiane et l'espérance de cette loi sont toutes les deux égales à μ.
2. Montrer que la loi de probabilité de la médiane empirique $\widetilde{X}_n$ est symétrique par rapport à μ. Pour simplifier, on supposera que n est impair.
3. En déduire que la moyenne empirique $\overline{X}_n$ et la médiane empirique $\widetilde{X}_n$ sont des estimateurs sans biais de μ.
4. Proposer un critère pour déterminer lequel des deux estimateurs $\overline{X}_n$ et $\widetilde{X}_n$ est asymptotiquement le meilleur.
5. Déterminer lequel des deux estimateurs $\overline{X}_n$ et $\widetilde{X}_n$ est asymptotiquement le meilleur dans les deux cas suivants :

   (a) loi normale $\mathcal{N}(\mu, \sigma^2)$: $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}$.

   (b) loi double exponentielle $De(\mu, \sigma)$: $f(x) = \dfrac{1}{2\sigma} e^{-\dfrac{|x-\mu|}{\sigma}}$.

ENSIMAG 2ème année **Statistique Inférentielle Avancée**

Fiche d'exercices n°4

Exercice 1. On considère la loi de Burr à deux paramètres $c > 0$ et $k > 0$ définie par sa fonction de répartition $F(x) = 1 - (1 + x^c)^{-k}$, $x \geq 0$.

1. Mettre F sous la forme $F(x) = 1 - x^{-1/\gamma} L(x)$ avec $\gamma > 0$, où L est une fonction à variations lentes que l'on précisera. En déduire le domaine d'attraction de la loi de Burr.

2. On pose pour $n > 0$, $a_n = F^{-1}(1 - 1/n)$. Soit X_n^* le maximum d'un échantillon de taille n de loi de Burr. On note Ψ la limite de la fonction de répartition de X_n^*/a_n quand n tend vers l'infini. Calculer $\Psi(x)$ pour tout $x \in \mathbb{R}$.

3. Montrer que le théorème de Gnedenko est vérifié, c'est-à-dire qu'il existe $\alpha > 0$ et $\beta \in \mathbb{R}$ tels que $\Psi(x) = H_\gamma(\alpha x + \beta)$ pour tout $x \in \mathbb{R}$.

Exercice 2. Soit $(X_1, \ldots, X_n)$ un échantillon d'une loi de probabilité admettant une densité f sur $\mathbb{R}$, que l'on suppose de classe C^∞. On se propose d'estimer f par la méthode du noyau :

$$\forall x \in \mathbb{R},\ \hat{f}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right)$$

où $\{h_n\}_{n\geq 1}$ est une suite de réels positifs convergent vers 0 et K est la densité d'une loi de probabilité symétrique par rapport à l'origine et de variance μ_2.

1. Calculer l'espérance et la variance de $\hat{f}(x)$ en fonction de K, h_n et f.

2. A l'aide d'un développement de Taylor de $f(x - h_n u)$, montrer les équivalents suivants, quand n tend vers l'infini :

   (a) $\left(E\left[\hat{f}(x)\right] - f(x)\right)^2 \sim \dfrac{h_n^4 \mu_2^2}{4} f''(x)^2$

   (b) $Var\left[\hat{f}(x)\right] \sim \dfrac{f(x)}{nh_n} \displaystyle\int_{-\infty}^{+\infty} K(u)^2 du$

3. On admet qu'un équivalent de l'erreur quadratique moyenne intégrée s'obtient en intégrant la somme des deux équivalents ci-dessus. Déterminer le paramètre de lissage optimal $h_{n,opt}$.

4. Calculer des équivalents de l'erreur quadratique moyenne intégrée minimale, du biais et de la variance de l'estimateur optimal.

5. Montrer que l'efficacité d'un noyau K est mesurée par $\mu_2 \left[\int_{-\infty}^{+\infty} K(u)^2 du\right]^2$. En déduire un classement des performances des noyaux rectangulaire, triangulaire, gaussien et d'Epanechnikov.

6. Calculer $h_{n,opt}$ quand on suppose que le noyau est gaussien et que la densité à estimer est celle d'une loi normale centrée.