Université de Nantes – UFR des Sciences et Techniques
Département de Mathématiques

Master 1 Ingénierie mathématique **Année 2012-2013**

Exercices de Statistique

F. Lavancier, A. Philippe

Estimation

Ex 1. Montrer que les familles de lois suivantes appartiennent à la famille des lois exponentielles

- les lois de Poisson de paramètre λ.
  Rappel : X suit une loi de Poisson de paramètre λ si $\forall k \in \mathbb{N}$ par $P(X = k) = e^{-\lambda}\dfrac{\lambda^k}{k!}$.
- les lois gaussiennes de moyenne μ et de variance σ^2.

Ex 2. Convergence

1) Si $\hat{\theta}_n$ est un estimateur asymptotiquement sans biais de θ et si sa variance tend vers zéro,alors il est convergent (en quels sens?).

2) Si $\hat{\theta}_n$ est un estimateur de θ convergent en moyenne quadratique, alors il est asymptotiquement sans biais.

Ex 3. On dispose d'un n-échantillon $(X_1, \cdots, X_n)$ d'une loi de Poisson de paramètre $\lambda > 0$.

1) Montrer que la moyenne empirique $\bar{X}_n$ est un estimateur sans biais de λ

2) Montrer que $\bar{X}_n$ converge presque sûrement et dans L^2 vers θ, lorsque $n \to +\infty$.

Ex 4. On considère un n-échantillon $(X_1, \cdots, X_n)$ de la loi uniforme sur $[0, \theta]$. On pose $\hat{\theta}_n = \max\{X_1, \cdots, X_n\}$.

1) Montrer que $\hat{\theta}_n$ est asymptotiquement sans biais.

2) En déduire un estimateur sans biais de θ.

3) Montrer que $\hat{\theta}_n$ converge presque sûrement et dans L^2 vers θ, lorsque $n \to +\infty$.

Ex 5. On dispose d'un n-échantillon $X_1, \cdots, X_n$ d'une loi de Bernoulli de paramètre $p \in]0, 1[$ inconnu. Montrer qu'il n'existe pas d'estimateur sans biais de $\frac{1}{p}$.

Ex 6. Soit $X_1, \ldots, X_n$ une suite de variables aléatoires indépendantes et identiquement distribuées suivant la loi exponentielle de paramètre $\theta > 0$. Rappelons que X_1 admet alors pour densité par rapport à la mesure de Lebesgue

$$f(x) = \theta e^{-\theta x} \, \mathbb{I}_{\mathbb{R}^+}(x).$$

Les variables X_i ne sont pas observées, on observe uniquement les variables aléatoires Y_i définies par

$$Y_i = \mathbb{I}_{X_i > 2} \qquad i = 1, \ldots, n.$$

1) Donner la loi de Y_1, puis calculer $\mathbb{E}(Y_1)$.

On pose, pour tout entier $n \geq 1$,

$$\bar{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

et on estime le paramètre θ par

$$\hat{\theta}_n = \begin{cases} -\frac{1}{2} \log(\bar{Y}_n) & \text{si } \bar{Y}_n > 0 \\ 0 & \text{sinon} \end{cases}$$

2) Calculer la probabilité de l'événement $[\bar{Y}_n \neq 0]$. Montrer que presque sûrement, l'événement $[\bar{Y}_n \neq 0]$ est réalisé à partir d'un certain rang.

3) Montrer que $\hat{\theta}_n$ converge presque sûrement vers θ.

Ex 7. Soit un échantillon de variables aléatoires suivant la même loi paramétrée par un réel θ. On suppose disposer de deux estimateurs $\hat{\theta}_1$ et $\hat{\theta}_2$ de θ, calculés à partir de cet échantillon. On considère l'estimateur $\bar{\theta}$, défini comme la moyenne de θ_1 et θ_2, i.e.

$$\bar{\theta} = \frac{\hat{\theta}_1 + \hat{\theta}_2}{2}.$$

On note σ_1^2 (resp. σ_2^2) la variance de $\hat{\theta}_1$ (resp. de $\hat{\theta}_2$) que l'on supposera finie et ρ la corrélation entre $\hat{\theta}_1$ et $\hat{\theta}_2$.

1) Exprimer la variance de $\bar{\theta}$ en fonction de σ_1, σ_2 et ρ.

2) On suppose dans un premier temps que $\hat{\theta}_1$ et $\hat{\theta}_2$ sont sans biais et sans perte de généralité que $\sigma_2^2 = \alpha \sigma_1^2$ où $\alpha \geq 1$. Donner une condition sur ρ pour que $\bar{\theta}$ soit meilleur que $\hat{\theta}_1$ et $\hat{\theta}_2$ au sens du coût quadratique. Détailler cette condition dans les cas $\alpha = 1$ et $1 \leq \alpha \leq 2$.

3) Si $\hat{\theta}_1$ correspond au meilleur estimateur sans biais possible au sens du coût quadratique, que doit valoir $\hat{\theta}_2$ pour que $\bar{\theta}$ soit meilleur que $\hat{\theta}_1$ et $\hat{\theta}_2$?

4) On se place à présent dans le cas général où $\hat{\theta}_1$ et $\hat{\theta}_2$ peuvent être biaisés. On note B_1 (resp. B_2) le biais de $\hat{\theta}_1$ (resp. de $\hat{\theta}_2$). On note EQM_1 (resp. EQM_2) l'erreur quadratique moyenne de $\hat{\theta}_1$ (resp. de $\hat{\theta}_2$). On note enfin

$$\tilde{\rho} = \frac{E[(\hat{\theta}_1 - \theta)(\hat{\theta}_2 - \theta)]}{\sqrt{EQM_1 EQM_2}}.$$

Exprimer l'erreur quadratique moyenne de $\bar{\theta}$ en fonction des quantités introduites ci-dessus.

5) On suppose sans perte de généralité que $EQM_2 = \alpha EQM_1$ où $\alpha \geq 1$. Dans le cas où les biais de $\hat{\theta}_1$ et $\hat{\theta}_2$ sont de signes opposés, donner une condition simple sur α impliquant que $\bar{\theta}$ est meilleur que $\hat{\theta}_1$ et $\hat{\theta}_2$ au sens du coût quadratique.

6) Proposer d'autres conditions sous lesquelles $\bar{\theta}$ est meilleur que $\hat{\theta}_1$ et $\hat{\theta}_2$ au sens du coût quadratique.

Ex 8. Soit $X_1, \ldots, X_n$ une suite de variables aléatoires indépendantes et identiquement distribuées suivant la loi exponentielle de paramètre $\theta > 0$.

1) Montrer que ce modèle est dominé.
2) Écrire la vraisemblance du modèle.
3) Vérifier que le modèle est régulier.
4) Calculer l'information de Fisher apportée par le $n-$échantillon.

Ex 9. On considère un n-échantillon $(X_1, \cdots, X_n)$ d'une loi gaussienne $\mathcal{N}(\mu, \sigma^2)$.

1) Écrire la vraisemblance du modèle
2) Quelle est l'information de Fisher dans le cas suivants:

- σ^2 est connu, μ ne l'est pas.
- μ est connu, σ^2 ne l'est pas.
- Aucun des deux paramètres n'est connu.

Ex 10. On considère un échantillon $(X_1, \cdots, X_n)$ d'une loi gaussienne $\mathcal{N}(\mu, \sigma^2)$. On estime σ^2 par la variance empirique modifiée

$$S_n^2 = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \overline{X}_n)^2.$$

1) Quel est son biais, son risque quadratique? Est-il efficace?
2) Quel est parmi les estimateurs de la forme

$$c \sum_{j=1}^{n} (X_j - \overline{X}_n)^2$$

(où $c > 0$) celui qui minimise l'erreur quadratique? En déduire que S_n^2 n'est pas admissible.

Ex 11. On considère un n-échantillon $(X_1, \cdots, X_n)$ d'une loi gaussienne $\mathcal{N}(\mu, \sigma^2)$. Quel est l'estimateur du maximum de vraisemblance du couple (μ, σ^2)

Ex 12. On considère un n-échantillon $(X_1, \cdots, X_n)$ d'une loi logisitique de paramètre $\theta \in \mathbb{R}$, de densité

$$f(x) = \frac{e^{-(x-\theta)}}{\left(1 + e^{-(x-\theta)}\right)^2}, \quad x \in \mathbb{R}$$

1) Montrer que f est symétrique autour de θ. En déduire l'espérance de X_1.

2) En déduire un estimeur de θ.

3) En admettant que $Var(X_1) = \pi^2/3$, donner une approximation de la loi asymptotique de votre estimateur.

4) On admet que l'information de Fisher apportée par une observation vaut $I(\theta) = 1/3$. L'estimateur proposé précedement est-il asymptotiquement efficace?

5) Montrer que le maximum de vraisemblance existe. Peut-on en donner une formule explicite? Quelle est approximativement sa loi asymptotique?

6) Proposer un estimateur asymptotiquement efficace qui ne nécessite pas le calcul du maximum de vraisemblance.

Ex 13. On considère un n-échantillon $(X_1, \cdots, X_n)$ de la loi uniforme sur $[0, \theta]$.

1) Quel est l'estimateur du maximum de vraisemblance de θ. On le note $\hat{\theta}_n$.

2) Montrer que $n(\hat{\theta}_n - \theta)$ a une loi limite lorsque $n \to +\infty$. Quelle est cette loi limite?

3) Quels commentaires vous suggère ce résultat de convergence?

Ex 14. On dispose d'un n-échantillon $(X_1, \cdots, X_n)$ d'une loi de Poisson de paramètre $\lambda > 0$.

1) Quel est l'estimateur du maximum vraisemblance $\hat{\lambda}_n$ de λ? Montrer qu'il est sans biais et efficace.

2) Montrer que $\sqrt{n}(\hat{\lambda}_n - \lambda)$ converge en loi. Préciser la limite.

3) On désire estimer $e^{-k\lambda}$ où k est un entier positif fixé. Interprèter $e^{-k\lambda}$ comme la probabilité d'un événement.

4) En déduire un estimateur sans biais de $e^{-k\lambda}$. Est-il convergent? Efficace? Améliorez-le en utilisant la statistique exhaustive $\sum_{j=1}^{n} X_j$.

5) Comparez cet estimateur avec $e^{-k\hat{\lambda}_n}$.

Ex 15. On considère un n-échantillon $(X_1, \cdots, X_n)$ d'une loi gaussienne $\mathcal{N}(\mu, \sigma^2)$. Donner une statistique exhaustive.

Ex 16. Montrer que les statistiques suivantes sont exhaustives et totales

1) La somme pour le modèle de Poisson.

2) $\max\{X_1, \ldots, X_n\}$ pour un modèle uniforme sur $[0, \theta]$.

Ex 17. Soit $(X_1, \dots, X_n)$ une suite de longueur n, de variables aléatoires indépendantes et de même loi de Bernoulli de paramètre $p \in]0,1[$.

1) Écrire la vraisemblance.

2) Montrer que la statistique $S_n = \sum_{i=1}^{n} X_i$ est exhaustive.

3) Montrer que la statistique S_n est totale.

4) Calculer l'estimateur du maximum de vraisemblance de p.

5) Cet estimateur peut il être amélioré à l'aide de la statistique exhaustive S_n?

6) est-il un estimateur efficace du paramètre p ?

7) Proposer un estimateur sans biais de p^k où k ets un entier conuu.

8) En déduire un estimateur sans biais de variance minimale.

Ex 18. *Modèle de capture-recapture*
Dans un bassin, il y a un nombre inconnu N de poissons. On souhaite estimer N. Pour cela on capture n poissons et on les remet à l'eau après les avoir marqués. Ultérieurement on capture à nouveau K poissons ($K \geq n$). Soit X la variable aléatoire égale au nombre de poissons marqués parmi les K qui viennent d'être capturés.

1) Quelle est la loi de X?
2) En déduire l'estimateur du maximum de vraisemblance du paramètre N.
3) Que pensez vous de cet estimateur?
4) Une autre expérience destinée à estimer N consiste à capturer les poissons les uns après les autres (avec remise...) jusqu'à la capture du premier poisson marqué. On note Y la variable aléatoire égale au nombre de poissons capturés. Quelle est la loi de Y? Calculer son espérance et sa variance. En déduire un estimateur de N.

UNIVERSITÉ DE NANTES

Université de Nantes – UFR des Sciences et Techniques
Département de Mathématiques

Master 1 Ingénierie mathématique **Année 2012-2013**

Exercices de Statistique

F. Lavancier, A. Philippe

Tests d'hypothèses

Ex 19. Une machine automatique est censé fabriquer des comprimés de poids moyen 500 mg. Afin de vérifier si la machine ne se dérègle pas, on prélève régulièrement des échantillons de 40 comprimés et on en contrôle le poids. Lors d'un de ces contrôles, la moyenne du poids des comprimés est évaluée à 503 mg avec un écart-type de 11,8 mg. En supposant que les mesures sont distribuées suivant une loi normale, peut-on dire que la machine est déréglée?

Ex 20. Le taux normal de glycémie est de 1,0 g/L. On dose la glycémie chez 17 sujets diabétiques à jeun depuis 4 heures. La moyenne estimée est de 1,2 g/L avec un écart-type de 0,10 g/L.

En supposant que le taux de glycémie est distribué selon une loi normale, peut-on dire, au risque de 5%, que ces sujets sont hyperglycémiques ?

Ex 21. L'écart type de la teneur d'un composant dans un médicament est de 8 milligrammes. Un nouveau procédé de fabrication vise à diminuer cet écart type. Pour 10 mesures de teneur sur des unités fabriquées par le nouveau procédé on obtient (en mg):

725	722	727	718	723	731	719	724	726	726

On suppose que les mesures sont des v.a gaussiennes, identiquement distribuées et indépendantes. Faire un test permettant de voir si le but recherché est atteint.

Ex 22. Une usine fabrique une fibre synthétique dans deux ateliers. Un échantillon de 25 spécimens de fibres provenant de l'atelier 1 et de 30 provenant de l'atelier 2 donnent une résistance moyenne à la rupture égale à $64kg/cm^2$ pour le premier et à $61kg/cm^2$ pour le second. On suppose que la résistance à la rupture est une variable aléatoire gaussienne de variance 7 pour le premier atelier et 8 pour le second, et que les résistances des fibres sont indépendantes entre elles. Peut-on conclure au niveau 0.05 qu'il n'y a pas de différence concernant la résistance à la rupture des fibres produites par les deux ateliers?

Ex 23. On désire tester si un médicament a une influence sur le comportement psychomoteur. On choisit au hasard 20 sujets qu'on répartit au hasard en deux groupes: le groupe témoin et le groupe expérimental. On leur fait subir la même expérience psychomotrice. On a administré auparavant le médicament aux sujets du groupe expérimental et un placebo au groupe témoin. Les résultats sont les suivants:

Groupe témoin	166	167	169	170	174	173	172	170	166	173
Groupe expérimental	167	162	165	168	162	160	164	158	165	169

On suppose que dans chaque groupe les résultats sont distribués selon une loi gaussienne, que la variance est la même pour les deux groupes et que les performances des sujets sont indépendantes. Tester au niveau 0.05 l'hypothèse selon laquelle le médicament n'a aucun effet sur le comportement psychomoteur.

Ex 24. On considère n variables aléatoires, $(X_1, \ldots, X_n)$ indépendantes et identiquement distribuées suivant la loi qui admet pour densité

$$f_\theta(x) = \frac{1}{\theta} \exp\{\frac{-x}{\theta}\} \mathbb{I}_{[0,\infty[}(x)$$

Soit (θ_0, θ_1) deux valeurs de θ telles que $0 < \theta_0 < \theta_1$.

On souhaite maintenant tester l'hypothèse $\mathrm{H}_0 : \theta = \theta_0$ contre l'alternative $\mathrm{H}_1 : \theta = \theta_1$.

1) Donner la forme du test de Neyman-Pearson.

2) Rappeler les propriétés de ce test.

3) Exprimer la région critique de ce test au niveau α.

4) Proposer une approximation de la région critique quand $n \to \infty$.

5) On souhaite à présent tester $\mathrm{H}_0 : \theta \leq \theta_0$ contre l'alternative $\mathrm{H}_1 : \theta \geq \theta_0$. Donner la forme du test uniformément le plus puissant au niveau α.

Ex 25. Justifier la forme des régions critiques proposées dans l'exercice 16 de la feuille de TP (dans l'énoncé puis la généralisation après la question 6).

Ex 26. On dispose de 10 observations $(X_1, \ldots, X_{10})$ indépendantes et identiquement distribuées suivant une loi de Poisson de paramètre θ.

On souhaite tester l'hypothèse $\mathrm{H}_0 : \theta = 0.5$ contre l'alternative $\mathrm{H}_1 : \theta = 1$.

1) Donner la forme du test de Neyman-Pearson.

2) Exprimer la fonction de test pour un test de niveau $\alpha = 5\%$

3) On observe $\sum_{i=1}^{10} X_i = 9$, comment peut-on conclure?

4) Proposer un test UPP (uniformément le plus puissant) pour tester $\mathrm{H}_0 : \theta \leq 0.5$ contre $\mathrm{H}_1 : \theta > 0.5$ au niveau α. Que conclure pour l'observation précédente?

Ex 27. On considère n variables aléatoires, $(X_1, \dots, X_n)$ indépendantes et identiquement distribuées suivant la loi uniforme sur $[0, \theta]$, où $\theta > 0$.
Soit $\theta_0 > 0$. On souhaite tester $\mathrm{H}_0 : \theta \leq \theta_0$ contre $\mathrm{H}_1 : \theta > \theta_0$. On note $M_n = \max_{1 \leq i \leq n} X_i$.

1) On propose comme région critique $RC = \{M_n > \theta_0\}$. Quel est le niveau du test?

2) Calculer la fonction puissance du test précédent et en déduire la forme de sa courbe représentative.

3) On considère à présent le test de région critique $RC = \{M_n > c\}$, où $c < \theta_0$. Trouver c pour que ce test soit de niveau $\alpha \in]0, 1[$. Calculer la fonction puissance associée.

4) On considère le test d'hypothèses simples $\mathrm{H}_0 : \theta = \theta_0$ contre l'alternative $\mathrm{H}_1 : \theta = \theta_1$, où $\theta_1 > \theta_0$. Donner la forme du test de Neymann-Pearson associé.

5) Supposons qu'on adopte le test de Neymann-Pearson précédent pour tester $\mathrm{H}_0 : \theta \leq \theta_0$ contre $\mathrm{H}_1 : \theta \geq \theta_0$. Tracer l'allure de sa fonction puissance. Commenter l'intérêt de ce test.

6) En vous inspirant de la forme du test de Neymann-Pearson, modifier le test initial de la question 1), en le rendant aléatoire, pour qu'il soit de niveau α. Que devient alors sa fonction puissance?

7) Entre tous les tests considérés, quel est celui uniformément le plus puissant au niveau α pour tester $\mathrm{H}_0 : \theta \leq \theta_0$ contre $\mathrm{H}_1 : \theta \geq \theta_0$?

8) Soit un test Φ de niveau α pour tester $\mathrm{H}_0 : \theta \leq \theta_0$ contre $\mathrm{H}_1 : \theta \geq \theta_0$.

a) Montrer qu'une condition nécessaire pour que Φ soit uniformément le plus puissant parmi les tests de niveau α est que $\Phi = 1$ si $M_n > \theta_0$.

b) Montrer que pour tout $\theta_1 > \theta_0$, $\mathbb{E}_{\theta_1}(\Phi\, \mathbb{I}_{M_n \leq \theta_0}) = \alpha(\theta_0/\theta_1)^n$

c) En déduire la fonction puissance des tests Φ de niveau α vérifiant $\Phi = 1$ si $M_n > \theta_0$.

d) En déduire que tout test Φ de niveau α vérifiant $\Phi = 1$ si $M_n > \theta_0$ est uniformément le plus puissant parmi les tests de niveau α pour tester $\mathrm{H}_0 : \theta \leq \theta_0$ contre $\mathrm{H}_1 : \theta \geq \theta_0$.

9) Vérifier que les tests des questions 3) et 6) sont des tests uniformément les plus puissants parmi tous les tests de niveau α.

UNIVERSITÉ DE NANTES

Université de Nantes – UFR des Sciences et Techniques
Département de Mathématiques

Master 1 Ingénierie mathématique **Année 2012-2013**

Exercices de Statistique

F. Lavancier, A. Philippe

Tests du χ^2

Ex 28. Deux cent jets d'un même dé ont donné les résultats suivants:

j: numéro obtenu	1	2	3	4	5	6
n_j: nombre d'apparitions de j	30	26	29	30	40	45

On se demande si le dé est pipé.

1) Quelle statistique de test utiliseriez vous pour répondre à cette question? (Contentez vous d'indiquer les calculs numériques que vous auriez à faire, mais ne les faites pas).

2) Tous calculs faits on trouve que la valeur prise dans cette experience par la statistique de test est 8.26. Dans quelle table lisez vous la conclusion de votre test? Pourquoi? Quelle est cette conclusion?

Ex 29. On effectue N prélèvements dans un cours d'eau et on compte le nombre de particules d'or en suspension dans chaque prélèvement.

1) On suppose que les prélèvements sont indépendants et que le nombre de particules d'or en suspension dans un prélèvement est distribué selon une loi de Poisson de paramètre λ inconnu. On note X_j le nombre de particules dans le j-ième prélèvement. Quel estimateur $\hat{\lambda}_N$ de λ proposez vous? Justifiez brièvement votre réponse.

2) On a effectué $N = 518$ prélèvements. Les résultats sont rassemblés dans le tableau ci-dessous:

k: Nombre de particules	0	1	2	3	4	5	6	7
n_k: Nombre de prélèvements	116	174	129	65	24	7	2	1

Quelle valeur prend l'estimateur $\hat{\lambda}_N$ que vous avez proposé?

3) On désire tester la validité de l'hyposthèse H_0: *la loi des X_j est une loi de Poisson.* Etant donné le faible effectif des trois dernières cases du tableau précédent on vous conseille de les regrouper et de les attribuer aux prélèvements qui présentent un nombre de particules supérieur ou égal à 5. Quelle statistique de test proposez vous?

Quelle table devez-vous utiliser? Donnez le résultat du test de niveau 0.05.

Ex 30. Le tableau suivant indique le classement de 124 personnes selon la couleur de leurs yeux et la couleur de leurs cheveux. Peut on conclure à l'indépendance de ces deux caractères?

yeux \ cheveux	Blonds	Bruns	Noirs	Roux	Total
Bleus	25	9	3	7	44
Gris ou verts	13	17	10	7	47
Marrons ou noirs	7	13	8	5	33
Total	45	39	21	19	124

Université de Nantes – UFR des Sciences et Techniques
Département de Mathématiques

Master 1 Ingénierie mathématique **Année 2012-2013**

Exercices de Statistique

F. Lavancier, A. Philippe

Régions de confiance

Ex 31. On considère un $n-$échantillon $(X_1, \cdots, X_n)$ iid suivant une loi exponentielle de paramètre $\lambda > 0$ inconnu. La densité de la loi exponentielle est

$$f_\lambda(x) = \lambda e^{-\lambda x}\, \mathbb{I}_{\mathbb{R}^+}(x)$$

1) Montrer que la moyenne empirique $\bar{X}_n = \dfrac{1}{n}\sum_{i=1}^n X_i$ est un estimateur sans biais de $\frac{1}{\lambda}$ qui converge au sens L^2 et presque sûrement.

2) Construire un intervalle de confiance pour $\frac{1}{\lambda}$ asymptotiquement de niveau $1 - \alpha$

3) On souhaite maintenant estimer le paramètre λ. Proposer un estimateur et préciser ses propriétés.

4) Donner un intervalle de confiance asymptotiquement de niveau $1 - \alpha$ pour le paramètre λ.

5) On estime maintenant la durée de survie $\tau = P(X_1 > 1)$. Proposer un estimateur pour ce paramètre.

6) Donner un intervalle de confiance asymptotiquement de niveau $1 - \alpha$ pour le paramètre τ.