UNIVERSITÉ DE NANTES

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

Master Pro 1 **Année 2006-2007**

Exercices : Statistique
Estimation

A. Philippe

Ex 1. 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 2. 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 3. 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 4. 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 5. 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 6. 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 7. 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 8. 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 9. 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 10. 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 11. 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. Interprêter $e^{-k\lambda}$ 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 $\displaystyle\sum_{j=1}^{n} X_j$.

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

Ex 12. On considère un n-échantillon $(X_1, \cdots, X_n)$ d'une loi gaussienne $\mathcal{N}(\mu, \sigma^2)$. Donner une statistique exhaustive.

Ex 13. 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 14. Soit $(X_1, \ldots, X_n)$ 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 15. *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 Pro 1 **Année 2006-2007**

Exercices : Statistique
Tests d'hypothèses

A. Philippe

Ex 16. 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 $n \to \infty$

Ex 17. 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?

Ex 18. 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 19. 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 20. 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 21. On a fait une numération globulaire à un groupe de 20 personnes à deux périodes différentes de l'année. Pour chaque sujet on note les résultats des deux numérations (à multiplier par 10^5):

Sujet	01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20
Janvier-février	46 38 42 47 48 40 40 43 42 49 45 51 47 52 50 48 47 47 47 45
Sept.-oct.	48 47 44 45 51 44 47 48 47 57 49 55 48 48 46 48 54 54 44 48

On suppose que les sujets sont mutuellement indépendants et suivent une loi gaussienne. Tester au niveau 0.05 l'hypothèse selon laquelle les résultats de la numération sont les mêmes aux deux périodes.

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

Master Pro 1 **Année 2006-2007**

Exercices : Statistique

Tests du χ^2

A. Philippe

Ex 22. 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 23. 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 $jime$ 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 24. 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?

cheveux / yeux	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

Ex 25. *Extrait – Juin 2003*

La différence de potentiel mesurée aux bornes d'un conducteur ohmique traversé par un courant d'intensité x_i ($i = 1, \dots, n$) est modélisée par une variable aléatoire

$$U_i = rx_i + \varepsilon_i \qquad \forall i = 1, \dots, n$$

où

- r est un paramètre inconnu qui représente la résistance du conducteur ohmique.
- $(\varepsilon_i)_{1 \leq i \leq n}$ est une suite de variables aléatoires indépendantes et de même loi gaussienne de moyenne zéro et de variance σ^2.
- les intensités $(x_i)_{1 \leq i \leq n}$ ne sont pas aléatoires

On cherche à estimer les deux paramètres (r, σ^2).
On considère n mesures indépendantes $U_1, \dots, U_n$ réalisées pour des intensités $x_1, \dots, x_n$.

1) Écrire la vraisemblance

2) Donner une statistique exhaustive.

3) Calculer les estimateurs du maximum de vraisemblance des paramètres r, σ^2. On les note $\hat{r}_n^{MV}$ et $\hat{\sigma}_n^{2MV}$.

4) Montrer que $\hat{r}_n^{MV}$ est un estimateur sans biais de r. Calculer sa variance.

5) Donner la loi de $\hat{r}_n^{MV}$.

6) Calculer l'estimateur du paramètre r par la méthode des moindres carrés. Commenter.

7) On réalise les mesures suivantes

x_i (en ampère)	0.5	1.5	3	5	8
U_i (en volt)	1.73	2.41	8.18	9.86	16.11

Donner une estimation de la résistance r du conducteur ohmique.

8) On suppose désormais que le paramètre σ est connu et vaut 1. On souhaite tester l'hypothèse $H_0 : r = r_0 = 1.9$ contre l'hypothèse alternative $H_1 : r > 1.9$.

9) On construit une procédure de test dont la région critique est de la forme

$$R_0 = \{\sum_{i=1}^{n} U_i x_i > C\}.$$

Justifier ce choix

10) Quelle valeur de C faut-il choisir pour que le test associé soit de niveau $\alpha = 5\%$

11) Au vu des observations, quelle est la conclusion de ce test au niveau 5%

Régions de confiance

Ex 26. 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 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

Régression linéaire

Ex 27. Soit un modèle à deux variables explicatives dont les matrices $X'X$ et $X'Y$ sont calculées à partir d'un n-échantillon d'observations de variables centrées.

$$X'_n X_n = \begin{pmatrix} 200 & 150 \\ 150 & 113 \end{pmatrix} \; ; \; X'_n Y_n = \begin{pmatrix} 350 \\ 263 \end{pmatrix}.$$

L'ajout d'une observation modifie les resultats de la manière suivante :

$$X'_{n+1} X_{n+1} = \begin{pmatrix} 199 & 149 \\ 149 & 112 \end{pmatrix} \; ; \; X'_{n+1} Y_{n+1} = \begin{pmatrix} 347.5 \\ 261.5 \end{pmatrix}.$$

1) Calculer les coefficients de régression du modèle dans les deux cas.

2) Calculer le coefficient de corrélation linéaire entre les deux variables explicatives.
3) Commenter.