Examen du 30 Janvier 2007 (Durée 3h)
Calculatrice et cours autorisés

Statistiques - 2006-2007

Exercice 1
Un constructeur de piscine veut comparer 2 produits à dissoudre dans l'eau qui permettent de tuer les bactéries présentes. Les deux produits garantissent que 95 % des bactéries seront éliminées. Par contre il se peut que le pH soit plus basique et donc plus nocif pour les utilisateurs. Il mène une expérience dans son magasin où 2 piscines remplies sont exposées. Il met le produit A dans la piscine 1 et le produit B dans la piscine 2. Il a un pH-mètre vendu dans le commerce et il se doute que la mesure n'est pas fiable. Il prend donc 10 mesures dans chaque piscine.

Dans la piscine 1, il trouve $x_1, ..., x_{10}$ qui valent respectivement

$$7.33\ ;\ 6.17\ ;\ 7.46\ ;\ 8.13\ ;\ 6.68\ ;\ 6.76\ ;\ 7.97\ ;\ 6.76\ ;\ 6.81\ ;\ 8.40$$

La moyenne empirique vaut $\bar{x} = 7.247$ et la variance empirique (celui divisé par "$n-1$") vérifie $\sqrt{\widehat{\sigma_x^2}} = 0.73$.

Dans la piscine 2, il trouve $y_1, ..., y_{10}$ qui valent respectivement

$$10.40\ ;\ 7.27\ ;\ 8.99\ ;\ 7.28\ ;\ 9.18\ ;\ 9.10\ ;\ 7.96\ ;\ 7.71\ ;\ 9.59\ ;\ 9.61$$

La moyenne empirique vaut $\bar{y} = 8.709$ et la variance empirique (celui divisé par "$n-1$") vérifie $\sqrt{\widehat{\sigma_y^2}} = 1.08$.

Le constructeur n'a aucun parti pris entre les 2 produits et veut juste conseiller au mieux ses clients. Il voudrait pouvoir leur dire clairement "préférez A", "préférez B" ou "faites ce que vous voulez, les deux produits sont indiscernables vu mes mesures".

On sait que $pH = 7$ est neutre pour la peau tandis que $pH = 9$ est basique et donc nocif pour la peau.

Construire un test statistique qui permette au constructeur de décider au mieux.

Vous serez notés sur le choix et la justification du modèle statistique, des hypothèses, du test lui-même qui sera d'abord construit à un niveau α abstrait avant d'être calculé pratiquement. Vous devrez conclure au vu de vos calculs par l'une des 3 phrases "préférez A", "préférez B" ou "faites ce que vous voulez, les deux produits sont indiscernables vu les mesures". Pour vous aider voici un tableau de valeurs qui donne t tel que $P(U > t) = a$

	a=0.05	a=0.025	a=0.01	a=0.005	a=0.0025	a=1.10^{-4}
$U\sim \mathcal{N}(0,1)$	1.645	1.960	2.326	2.576	2.807	3.719
U∼T(9)	1.833	2.262	2.821	3.250	3.670	6.010
U∼T(10)	1.812	2.228	2,763	3.169	3.581	5.694
U∼T(18)	1.734	2.100	2.552	2,878	3.196	4.648
U∼T(20)	1.724	2.085	2.528	2.845	3.153	4.538

où T(k) est la loi du Student à k degrés de liberté.

Exercice 2 Soit n un entier fixé. Soient $X_1, ..., X_n$ n variables indépendantes et identiquement distribuées de densité f inconnue sur $[0,1]$ par rapport à la mesure de Lebesgue.

<u>*Modèle paramétrique*</u>

Soit D un entier fixé. Soit m la partition régulière de $[0,1]$ de pas $1/D$ c'est-à-dire que m est l'ensemble des intervalles I suivants

$$m = \left\{ \ [0, 1/D[\ ; \ [1/D, 2/D[\ ; ...; \ [(D-1)/D, 1] \ \right\}.$$

On suppose que la densité f inconnue est de la forme

$$\forall x \in [0,1], \quad f(x) = \sum_{I \in m} f_I \mathbf{1}_I(x),$$

où les f_I sont des réels inconnus.

1. Expliquer pour quoi le modèle est paramétrique et donner d le plus petit nombre de paramètres possibles ainsi que l'ensemble Θ des paramètres ($\Theta \subset \mathbb{R}^d$).
2. Soit I un intervalle quelconque et N_I le nombre de variables X_i appartenant à l'intervalle I. Donner la loi de N_I.
3. Donner la loi de $(N_1, ..., N_D)$ où pour tout k, $N_k = N_{[(k-1)/D, k/D[}$.
4. Calculer $\hat{f}_D$ l'estimateur du maximum de vraisemblance dans ce modèle paramétrique. *On pourra remarquer que* $\log f = \sum_{I \in m} \log(f_I) \mathbf{1}_I$.
5. Pour toute fonction $g \in \mathbb{L}^2[0,1]$, on note $\|g\|^2 = \int_0^1 g^2(x)dx$ et

$$\gamma_n(g) = -\frac{2}{n} \sum_{i=1}^{n} g(X_i) + \|g\|^2$$

le contraste des moindres carrés. Montrer que $\gamma_n(\hat{f}_D) = \min_g \gamma_n(g)$ où le minimum en g est pris sur l'ensemble des fonctions g constantes par morceaux sur la partition m.
6. Soit f^o une densité sur $[0,1]$ constante par morceaux sur la partition m et qui ne soit jamais nulle.
On veut tester H_0 :"$f = f^o$ " contre H_1 : "$f \neq f^o$ ".
Donner la limite en loi quand n tend vers $+\infty$ de $n \left\| \frac{\hat{f}_D - f^o}{\sqrt{f^o}} \right\|^2$ sous H_0.
En déduire un test de H_0 contre H_1.

<u>*Modèle non paramétrique*</u>

On ne suppose plus dorénavant que f est constante par morceaux mais on suppose que f est C^1 et que pour tout x de $[0,1]$, $|f'(x)| \leq 1$. On envisage toujours $\hat{f}_D$ comme un estimateur de f même si il n'est plus dorénavant l'estimateur du maximum de vraisemblance. On pose

$$\forall x \in [0,1], f_D(x) = \sum_{I \in m} D \left[\int_I f(u)du \right] \mathbf{1}_I(x).$$

1. Montrer que $\forall x \in [0,1], \mathbb{E}_f(\hat{f}_D(x)) = f_D(x)$.

2. Montrer que f_D est la projection orthogonale de f sur l'ensemble des fonctions constantes par morceaux sur la partition m (au sens de $\|.\|$).
3. Montrer que $\mathbb{E}_f(\left\|\hat{f}_D - f_D\right\|^2) \leq D/n$.
4. Montrer que $\|f - f_D\|^2 = \int_0^1 \sum_{I\in m} \left[D \int_I (f(x) - f(u))du\right]^2 \mathbf{1}_I(x)dx$.
5. Montrer que $\|f - f_D\|^2 \leq 1/D^2$.
6. On sait qu'il existe $c > 0$ tel que

$$\inf_{\hat{f}} \sup_f \mathbb{E}_f(\left\|f - \hat{f}\right\|^2) \geq cn^{-2/3},$$

   où l'infimum est pris sur tous les estimateurs possibles et le supremum sur toutes les densités de $[0,1]$ C^1 et à derivée bornée par 1.

   Trouver D_n tel que $\hat{f}_{D_n}$ soit un estimateur minimax à constante près.
7. Soit f^o une densité de $[0,1]$ C^1 et à derivée bornée par 1. On suppose que f^o est minorée par r. On pose f^o_D la projection orthogonale de f^o sur l'ensemble des fonctions constantes par morceaux sur m.

   On veut tester H_0 :"$f = f^o$ " contre H_1 : "$f \neq f^o$ " dans ce modèle non paramétrique. On pose

$$T = n\left\|\frac{\hat{f}_{D_n} - f^o_{D_n}}{\sqrt{f^o_{D_n}}}\right\|^2$$

   Expliquer pourquoi on peut considérer que la loi de T est presque celle d'un khi-deux à $D_n - 1$ degrés de liberté du moment que $n^{2/3}r \geq 5$.

   Construire alors un test approximativement de taille α de H_0 :"$f = f^o$ " contre H_1 : "$f \neq f^o$ ".

   Montrer que sa puissance tend vers 1 quand n tend vers $+\infty$.