Université des Sciences et Technologies de Lille
U.F.R. de Mathématiques Pures et Appliquées

IS Math314 **Année 2006**

Examen, 5 septembre 2006, durée 3 heures.

- Ce sujet comporte **4 pages**, incluant une table de la f.d.r. gaussienne standard.
- Le barème indiqué est là pour vous aider à gérer votre temps et n'a pas valeur contractuelle.
- Documents autorisés : polycopié du cours IPE, polycopié du cours d'IS.
- Calculatrices autorisées.

Exercice 1. *Échauffement gaussien (2 points)*

Soient X_1, X_2, X_3 trois variables aléatoires indépendantes et de même loi gaussienne standard $\mathfrak{N}(0,1)$. Quelle est la loi de la v.a. $Y := X_1 - 2X_2 + X_3$? Dans votre réponse, la partie calculatoire éventuelle ne devrait pas excéder 2 lignes.

Exercice 2. *Différence de moyennes empiriques (3 points)*

Soient $X_1, \dots, X_m$ un m-échantillon associé à une loi Q_1 et $Y_1, \dots, Y_n$ un n-échantillon associé à une loi Q_2. On suppose que ces échantillons sont indépendants l'un de l'autre, autrement dit que les vecteurs aléatoires $(X_1, \dots, X_m)$ et $(Y_1, \dots, Y_n)$ sont indépendants. On suppose que X_1 et Y_1 sont de carré intégrable et on note $\mu_1 := \mathbf{E}X_1$, $\mu_2 := \mathbf{E}Y_1$, $\sigma_1^2 := \operatorname{Var} X_1$, $\sigma_2^2 := \operatorname{Var} Y_1$. On note $\overline{X} := m^{-1}(X_1 + \dots + X_m)$ et $\overline{Y} := n^{-1}(Y_1 + \dots + Y_n)$ les moyennes empiriques de ces échantillons. Justifiez les formules

$$\mathbf{E}(\overline{X} - \overline{Y}) = \mu_1 - \mu_2, \quad \operatorname{Var}(\overline{X} - \overline{Y}) = \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}.$$

Exercice 3. *Tension de rupture (4 points)*

Le tableau 1 page 2 donne les tensions de rupture en kilogramme-force mesurées sur cent segments de fil de nylon.

Cette tension de rupture est une variable aléatoire Y de loi inconnue et on peut interpréter ce tableau comme les observations $Y_1(\omega), \dots, Y_{100}(\omega)$ d'un 100-échantillon de la loi de Y. On s'intéresse à la quantité $\theta := P(Y > 31)$.

1) Proposez une valeur numérique pour estimer θ en justifiant votre choix.

2) Proposez un intervalle de confiance au niveau 95% pour θ en utilisant la méthode avec majoration de la variance inconnue d'une certaine loi de Bernoulli qui apparaît naturellement dans ce problème.

TAB. 1 – Un 100-échantillon de tensions de rupture (en kgf) pour un fil de nylon

29,446	31,717	29,818	30,373	31,178	31,146	30,505	28,746	30,037	29,993
28,330	30,734	29,131	29,492	29,554	30,211	29,849	29,147	28,474	31,252
29,003	30,102	28,598	28,552	31,802	30,788	30,060	31,180	30,175	30,556
29,405	31,623	30,786	29,135	29,954	30,782	28,228	30,015	29,856	29,814
30,200	30,688	29,790	30,639	28,846	29,853	29,834	29,140	29,654	29,300
31,195	29,784	31,686	31,001	29,572	29,275	30,160	28,139	29,739	30,869
29,443	27,991	30,020	30,784	28,378	28,554	30,897	28,299	30,723	28,422
30,091	31,060	28,507	31,758	32,052	29,924	30,780	29,086	30,664	30,329
30,922	30,666	29,872	29,035	29,195	31,790	29,412	30,374	28,115	31,782
29,437	31,409	28,381	28,275	29,242	31,132	31,060	29,099	30,127	30,564

Problème 4. *(13 points)*

Soit $(\Omega, \mathcal{F}, (P_\theta)_{\theta\in]0,1[})$ un modèle statistique et $X_1, \dots, X_n$ un échantillon associé à ce modèle, tel que pour tout $\theta \in]0,1[$, la loi sous P_θ des X_i soit la loi discrète sur $\{a,b,c\}$ donnée par

$$P_\theta(X_i = a) = \frac{1-\theta}{2}, \quad P_\theta(X_i = b) = \frac{1}{2}, \quad P_\theta(X_i = c) = \frac{\theta}{2}.$$

Les valeurs des réels a, b, c n'interviennent pas dans ce problème, la seule chose qui importe est qu'il s'agisse de 3 valeurs distinctes. On définit les statistiques

$$S_{n,a} := \sum_{i=1}^n \mathbf{1}_{\{X_i=a\}}, \quad S_{n,b} := \sum_{i=1}^n \mathbf{1}_{\{X_i=b\}}, \quad S_{n,c} := \sum_{i=1}^n \mathbf{1}_{\{X_i=c\}}.$$

1) Vérifiez que $\frac{2}{n} S_{n,c}$ est un estimateur sans biais et fortement consistant de θ.

2) Que peut-on dire de $S_{n,a} + S_{n,b} + S_{n,c}$? Quelle est la loi du vecteur aléatoire $(S_{n,a}, S_{n,b}, S_{n,c})$?

3) Montrez que

$$\sum_{n=1}^{+\infty} P_\theta(S_{n,b} = n) < +\infty.$$

4) On pose

$$\Omega' := \{\omega \in \Omega;\ \exists N(\omega) \in \mathbb{N},\ \forall n \geq N(\omega),\ S_{n,a}(\omega) + S_{n,c}(\omega) > 0\}.$$

Expliquez pourquoi $P_\theta(\Omega') = 1$ pour tout $\theta \in]0,1[$.

5) On définit la suite de variables aléatoires $(T_n)_{n\geq 1}$ par

$$T_n(\omega) := \begin{cases} \dfrac{S_{n,c}(\omega)}{S_{n,a}(\omega) + S_{n,c}(\omega)} & \text{si } S_{n,a}(\omega) + S_{n,c}(\omega) > 0, \\ 0 & \text{sinon.} \end{cases}$$

Montrez soigneusement que T_n est un estimateur fortement consistant de θ.

6) Montrez que T_n est un estimateur asymptotiquement sans biais de θ. Indication : notez que pour tout $\omega \in \Omega$, $T_n(\omega) \in [0,1]$, donc $(T_n)_{n\geq 1}$ est une suite de v.a. bornées par une même constante.

7) Dans l'échantillon observé $x_1 = X_1(\omega), \dots, x_n = X_n(\omega)$, on note $n_a := S_{n,a}(\omega)$, le nombre d'occurences de la valeur a et de même $n_b := S_{n,b}(\omega)$, $n_c := S_{n,c}(\omega)$. Exprimez à l'aide de ces notations la vraisemblance $L(x_1, \dots, x_n, \theta)$ de l'échantillon.

8) Montrez que pour $(x_1, \dots, x_n)$ tel que $n_c > 0$, la fonction $\theta \mapsto L(x_1, \dots, x_n, \theta)$ a un maximum unique atteint en un point $\widehat{\theta} \in]0,1[$ qui s'exprime simplement en fonction de n_a et n_c. En déduire un estimateur de θ par maximum de vraisemblance.

TAB. 2 – Table des valeurs de Φ, f.d.r. de la loi normale standard $\mathfrak{N}(0,1)$

$$\Phi(x) = P(Z \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(\frac{-t^2}{2}\right) \mathrm{d}t, \quad Z \sim \mathfrak{N}(0,1).$$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5754
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6627	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7122	0.7156	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7356	0.7389	0.7421	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7793	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8079	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8414	0.8438	0.8461	0,8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8622
1.1	0.8643	0.8665	0.8687	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0,8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9083	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9193	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9485	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9648	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9895	0.9898	0.9901	0.9903	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9926	0.9928	0.9930	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9944	0.9946	0.9948	0.9949	0,9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9958	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986