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

IS Math314 **Année 2006**

Examen, 15 juin 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.

Ex 1. *Vitesse moyenne (3 points)*

On veut estimer par intervalle de confiance la vitesse moyenne des automobiles dans un certain virage d'une route à grand trafic. Pour cela on a enregistré à l'aide d'un radar les vitesses $X_1(\omega) = x_1, \dots, X_{400}(\omega) = x_{400}$ de 400 automobiles en une période de temps de 2 heures avec des conditions de circulation homogènes (météo, visibilité, densité de trafic,...). On a obtenu les statistiques suivantes :

$$\sum_{i=1}^{400} x_i = 35\ 200 \text{ km/h}, \quad \sum_{i=1}^{400} x_i^2 = 3\ 107\ 600 \text{ (km/h)}^2.$$

L'homogénéité des conditions de trafic permet de supposer que les variables aléatoires $X_1, \dots, X_{400}$ dont on a ainsi observé une réalisation sont indépendantes et de même loi. Proposez un intervalle de confiance au niveau 98% pour la vitesse moyenne $\mathbf{E}X_1$ en indiquant clairement quels résultats du cours légitiment les approximations faites. Les données numériques ci-dessus ont été « arrangées » pour vous permettre de faire facilement tous les calculs à la main si vous ne disposez pas d'une calculatrice.

Ex 2. *Estimation par maximum de vraisemblance (3 points)*

Soit $(\Omega, \mathcal{F}, (P_\theta)_{\theta\in\Theta})$ un modèle statistique où $\Theta =]0, +\infty[$. On note $X_1, \dots, X_n$ un échantillon associé à ce modèle, les X_i étant des v.a. à valeurs dans $]0, 1[$ ayant pour densité sous P_θ la fonction :

$$t \mapsto f(t, \theta) := \theta t^{\theta-1} \mathbf{1}_{]0,1[}(t).$$

On se propose d'estimer θ par maximum de vraisemblance.

1) Explicitez la fonction de vraisemblance $L(x_1, \dots, x_n, \theta)$ pour des x_i tous dans $]0, 1[$. Montrez qu'elle admet un maximum unique et en déduire l'estimateur T_n de θ par maximum de vraisemblance.

2) Calculez $\mathbf{E}_\theta(\ln X_1)$ après avoir justifié son existence.

3) Déduire de ce qui précède que T_n est un estimateur fortement consistant de θ.

Ex 3. *Convergence en loi (4 points)*

Soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathfrak{F}, P)$, indépendantes et de même loi telle que $\mathbf{E}X_1^2 = 1$ et $\mathbf{E}X_1 = 0$. On suppose de plus que pour tout i, $X_i(\omega)$ n'est jamais nul. On définit alors sur $(\Omega, \mathfrak{F}, P)$ les suites de variables aléatoires $(Y_n)_{n\geq 1}$ et $(Z_n)_{n\geq 1}$ par

$$Y_n := \sqrt{n}\frac{X_1 + \cdots + X_n}{X_1^2 + \cdots + X_n^2}, \quad Z_n := \frac{X_1 + \cdots + X_n}{(X_1^2 + \cdots + X_n^2)^{1/2}}.$$

Montrez que $(Y_n)_{n\geq 1}$ et $(Z_n)_{n\geq 1}$ convergent en loi quand n tend vers l'infini vers une v.a. gaussienne de loi $\mathfrak{N}(0, 1)$.

Ex 4. *En guise de problème de synthèse (12 points)*

Soit $(\Omega, \mathfrak{F}, (P_\theta)_{\theta\in\Theta})$ un modèle statistique où $\Theta =]0, +\infty[$. On note $X_1, \ldots, X_n$ un échantillon associé à ce modèle, les X_i étant des v.a. positives ayant pour densité sous P_θ la fonction :

$$t \mapsto f(t, \theta) := \frac{2t}{\theta^2}\mathbf{1}_{[0,\theta]}(t).$$

L'estimation de θ dans ce modèle sert de prétexte à une revue de l'ensemble du cours.

1) Calculez $\mathbf{E}_\theta X_1$ pour en déduire que $\frac{3}{2}\overline{X}$ est un estimateur sans biais et fortement consistant de θ. Calculez son EQM.

2) Proposez un intervalle de confiance au niveau 95% pour θ, basé sur l'estimateur de la question précédente. Appliquez cette méthode sur l'échantillon simulé (de taille 100) présenté à la table 1, page 3. Lisez les commentaires qui suivent cette table pour éviter de vous lancer dans des calculs longs et inutiles.

3) On pose

$$M_n := \max(X_1, \ldots, X_n).$$

Calculez la fonction de répartition $F_n(., \theta)$ de M_n : $F_n(x, \theta) := P_\theta(M_n \leq x)$.

4) Montrez que M_n est un estimateur fortement consistant de θ. *Indication :* étudiez pour $0 < \delta < \theta$, la convergence de la série $\sum_{n=1}^{+\infty} P_\theta(|M_n - \theta| > \delta)$.

5) Calculez $\mathbf{E}_\theta M_n$. En déduire que M_n est biaisé mais asymptotiquement sans biais. Corrigez ce défaut de M_n en le remplaçant par un estimateur $M'_n = c_n M_n$, pour que M'_n soit sans biais et fortement consistant.

6) Calculez l'EQM de M'_n et constatez que c'est un $O(n^{-2})$. Ce résultat vous paraît-il compatible avec l'inégalité de Cramér-Rao ?

7) Explicitez la fonction de vraisemblance $\theta \mapsto L(x_1, \ldots, x_n, \theta)$ et vérifiez qu'elle est nulle sur $]0, b[$ et monotone sur $[b, +\infty[$, où b est une fonction de $x_1, \ldots, x_n$ que vous préciserez. Quel est l'estimateur de θ par maximum de vraisemblance ?

8) On cherche un intervalle de confiance au niveau $1 - \varepsilon$ pour θ, construit à partir de l'estimateur M_n. Exprimez $P_\theta(M_n \leq \theta < M_n + x)$ pour $0 < x < \theta$ à l'aide de la f.d.r. de M_n calculée à la question 3. Déterminez ensuite $a_n \in]0, 1[$ en fonction de ε et de n pour que $P_\theta(M_n \leq \theta < M_n + a_n\theta) = 1 - \varepsilon$. En déduire un intervalle de confiance pour θ au niveau exact $1 - \varepsilon$. Notez que cette méthode fonctionne pour tout $n \geq 1$ et

tout $\varepsilon \in]0,1[$. Appliquez ceci avec l'échantillon de la table 1 ci-dessous et un niveau de confiance de 95%. Comparez avec l'intervalle obtenu à la question 2.

9) L'auteur de l'énoncé a hésité entre plusieurs méthodes pour simuler l'échantillon présenté dans la table 1. Quelle(s) méthode(s) pourriez vous suggérer ?

TAB. 1 – Un 100-échantillon simulé pour l'exercice 4

2,373 641	2,539 755	1,317 328	2,927 648	3,075 983	3,089 054	2,079 839	1,866 325
3,065 378	1,423 575	2,486 000	0,372 537	2,179 226	3,025 487	3,035 052	2,226 735
2,677 645	1,869 972	2,992 573	2,095 179	1,694 912	2,885 349	1,950 677	1,867 663
1,240 670	2,277 816	3,095 005	1,399 811	3,034 904	3,049 601	2,769 146	3,065 194
2,585 347	1,042 526	3,083 546	3,126 235	0,510 191	1,435 366	1,610 284	2,613 401
0,161 609	2,996 090	2,824 511	2,331 914	2,360 276	0,946 244	3,019 824	0,291 257
1,498 035	2,999 015	2,900 211	1,567 625	2,611 896	2,642 432	2,938 963	2,011 620
1,240 892	2,002 642	1,543 602	2,935 429	1,732 009	0,687 088	2,202 109	1,519 365
2,616 795	3,138 464	3,014 825	2,130 727	2,979 077	0,916 038	1,579 642	1,959 762
1,968 556	2,737 566	1,610 758	0,293 354	2,049 023	2,179 292	0,535 313	1,472 453
2,761 824	1,157 541	2,926 185	1,329 376	2,880 823	1,204 144	2,058 592	0,683 237
1,424 086	1,939 582	1,229 418	1,451 298	2,233 886	2,456 869	2,439 421	2,141 528
2,321 064	1,014 335	1,889 860	1,152 523				

Pour simuler cet échantillon, l'auteur a choisi une valeur θ_0 de θ et a utilisé l'un des algorithmes de simulation de v.a. de loi P_{θ_0}. La liste explicite des valeurs de l'échantillon ne vous sera pas vraiment utile. On a en effet calculé pour vous quelques statistiques associées à cet échantillon $x_1, \ldots, x_{100}$. Vous trouverez *parmi* elles toutes les valeurs utiles pour résoudre les questions 2 et 8, ce qui ne signifie pas que vous ayiez besoin de *toutes* ces valeurs.

$$\overline{x} = \frac{1}{100}\sum_{i=1}^{100} x_i = 2{,}068\,985.$$

$$\text{variance empirique } (s^2 = S^2(\omega)) \quad s^2 = 0{,}638\,195, \quad s = 0{,}798\,871.$$

$$\min(x_1, \ldots, x_{100}) = 0{,}161\,609, \quad \max(x_1, \ldots, x_{100}) = 3{,}138\,464.$$

$$\text{intervalle médian} \quad [2{,}130\,727;\ 2{,}141\,528].$$

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