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

M1-MEFM **Année 2010**

Examen, 5 janvier 2011, durée 3 heures.

- Ce sujet comporte **5 pages**, dont une table de la loi normale.
- 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 ainsi qu'*une feuille* manuscrite recto verso au format A4.
- Calculatrices autorisées.
- La qualité de la rédaction sera un élément important d'appréciation des copies.

Ex 1. *En guise de problème de synthèse (10 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. 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 2. Lisez les commentaires qui suivent cette table pour éviter de vous lancer dans des calculs longs et inutiles.

3) On pose

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

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

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 $\theta \mapsto L(x_1, \dots, x_n, \theta)$ et vérifiez qu'elle est nulle sur $]0, b[$ et monotone sur $[b, +\infty[$, où b est une fonction de $x_1, \dots, x_n$ 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'estimateurM_n. Exprimez $P_\theta(M_n \leqslant \theta < M_n + x)$ 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 $P_\theta(M_n \leqslant \theta < M_n + a_n\theta) = 1 - \varepsilon$. En déduire un intervalle de confiance pour θ au niveau exact $1 - \varepsilon$. Notez que cette méthode fonctionne pour tout $n \geqslant 1$ 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.

Table 1 – Un 100-échantillon simulé pour l'exercice 1

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 obtenir 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 $x_1, \dots, x_{100}$. 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.

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

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

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

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

Ex 2. *Inversions d'une permutation aléatoire (10 points)*

Pour a et b entiers, on note $[\![a, b]\!]$ l'ensemble des entiers k tels que $a \leqslant k \leqslant b$. On note Ω_n l'ensemble des permutations de $[\![1, n]\!]$, c'est à dire des bijections $\omega : [\![1, n]\!] \to [\![1, n]\!]$. Soient $k = \omega(i) \in [\![1, n]\!]$. On dit que k présente une inversion avec $\ell = \omega(j)$ si $j > i$ et $\ell < k$. Ainsi le nombre d'inversions dans ω de $k = \omega(i)$ est le nombre d'éléments inférieurs à $\omega(i)$ et situés à sa droite dans la liste $\omega(1), \dots, \omega(i), \dots, \omega(n)$. Le nombre d'inversions de k est ainsi toujours compris entre 0 et $k-1$, le nombre d'inversions de 1 est toujours nul.

Le but de cet exercice est d'établir une loi faible des grands nombres et un théorème limite central pour le nombre total d'inversions d'une permutation aléatoire. Notons pour cela $X_k(\omega)$ le nombre d'inversions de k dans ω, pour $k \in [\![1, n]\!]$ et $S_n(\omega) = \sum_{k=1}^{n} X_k(\omega)$.

1) Lorsque n n'est pas trop grand, on peut représenter ω en donnant la liste explicite des $\omega(i)$. Par commodité, nous écrirons chaque $k \in [\![1, n]\!]$ sur un jeton carré et chaque disposition sur une même ligne des n jetons représentera une permutation ω. Par exemple

$$\boxed{4}\,\boxed{3}\,\boxed{5}\,\boxed{1}\,\boxed{2}$$

représente la permutation ω définie par : $1 \mapsto 4$, $2 \mapsto 3$, $3 \mapsto 5$, $4 \mapsto 1$, $5 \mapsto 2$. L'élément $4 = \omega(1)$ a 3 inversions, puisqu'à sa droite on trouve 3, 1 et 2 qui lui sont inférieurs, donc $X_4(\omega) = 3$. De proche en proche, on voit ainsi que le vecteur d'inversions associé à ω est $(X_1(\omega), \dots, X_5(\omega)) = (0, 0, 2, 3, 2)$. Trouvez la (ou les) permutation(s) ayant pour vecteur d'inversion $(0, 1, 0, 3, 2)$. Même question pour le vecteur d'inversions $(0, 0, 1, 2, 0)$.

2) On note Δ_n le produit cartésien

$$\Delta_n = \{0\} \times [\![0, 1]\!] \times [\![0, 2]\!] \times \cdots \times [\![0, n-1]\!].$$

Tout vecteur d'inversions d'une permutation $\omega \in \Omega_n$ est dans Δ_n (pourquoi ?). Montrez que pour tout $x \in \Delta_n$, il existe *un unique* $\omega \in \Omega_n$ tel que $(X_1(\omega), \dots, X_n(\omega)) = x$. Autrement dit, l'application $\omega \mapsto (X_1(\omega), \dots, X_n(\omega))$ est une bijection de Ω_n sur Δ_n.

3) On munit Ω_n de la tribu $\mathcal{F}_n = \mathcal{P}(\Omega_n)$ de tous ses sous-ensembles et de la probabilité P_n, loi uniforme sur Ω_n. On a donc $P_n(\{\omega\}) = \frac{1}{n!}$ pour tout $\omega \in \Omega_n$. Que vaut $P_n((X_1, \dots, X_n) = x)$ pour $x \in \Delta_n$? Quelle est la loi du vecteur aléatoire $(X_1, \dots, X_n)$?

4) Soit un espace probabilisé $(\Omega', \mathcal{F}', P')$ sur lequel on puisse définir des variables aléatoires discrètes $U_1, \dots, U_n$ indépendantes et telles que pour tout $k \in [\![1, n]\!]$, U_k suive la loi uniforme sur $[\![0, k-1]\!]$ (on ne vous demande pas de justifier l'existence de cet espace). Vérifiez que $(U_1, \dots, U_n)$ suit la loi uniforme sur Δ_n. En déduire que les X_k sont indépendantes et de même loi que les U_k.

5) Justifiez brièvement les formules sommatoires

$$\sum_{j=1}^{n} j = \frac{n(n+1)}{2} \tag{1}$$

et

$$\sum_{j=1}^{n} j^2 = \frac{n(n+1)(2n+1)}{6}. \tag{2}$$

6) On rappelle que $S_n = X_1 + \cdots + X_n$. Vérifiez que pour $n \geqslant 2$,

$$\mathbf{E}S_n = \frac{n(n-1)}{4} \sim \frac{n^2}{4}, \quad \operatorname{Var} S_n = \frac{n(n-1)(2n+5)}{72} \sim \frac{n^3}{36}.$$

7) Montrez que S_n vérifie la loi faible des grands nombres suivante :

$$\forall \varepsilon > 0, \quad P_n \left(\left| \frac{S_n}{n^2} - \frac{1}{4} \right| > \varepsilon) \right) \xrightarrow[n \to +\infty]{} 0.$$

8) On rappelle le théorème limite central de Lindeberg que l'on énonce ici sous une forme adaptée à notre problème.

Théorème 1 (Lindeberg)**.** *Considérons le « tableau triangulaire » de variables aléatoires de ligne numéro n ($n \in \mathbb{N}^*$) :*

$$X_{n,1}, \dots, X_{n,i}, \dots, X_{n,n},$$

où ces n variables sont définies sur le même $(\Omega_n, \mathcal{F}_n, P_n)$, indépendantes, de carré intégrable. On note

$$S_n := \sum_{i=1}^{n} X_{n,i}, \qquad \sigma_{n,i}^2 := \operatorname{Var} X_{n,i}, \qquad s_n^2 := \operatorname{Var} S_n = \sum_{i=1}^{n} \sigma_{n,i}^2.$$

On suppose $s_n > 0$ pour tout n. Si de plus le tableau vérifie la condition de Lindeberg *:*

$$\forall \varepsilon > 0, \quad R_{n,\varepsilon} := \frac{1}{s_n^2} \sum_{i=1}^{n} \mathbf{E}\big((X_{n,i} - \mathbf{E}X_{n,i})^2 \mathbf{1}_{\{|X_{n,i} - \mathbf{E}X_{n,i}| > \varepsilon s_n\}} \big) \xrightarrow[n \to +\infty]{} 0, \tag{3}$$

alors $s_n^{-1}(S_n - \mathbf{E}S_n)$ converge en loi vers une v.a. Z de loi $\mathfrak{N}(0,1)$.

Dans notre problème, les espaces probabilisés $(\Omega_n, \mathcal{F}_n, P_n)_{n \geqslant 1}$ sont ceux définis à la question 3. La v.a. notée X_k, quand on travaillait à n fixé, est notée maintenant $X_{n,k}$ (elle dépend de n puisqu'elle s'applique à des permutations ω sur $[\![1, n]\!]$). En utilisant le résultat de la question 6, montrez qu'il existe un rang $n_0 = n_0(\varepsilon)$ à partir duquel toutes les indicatrices dans $R_{n,\varepsilon}$ sont nulles (sur Ω_n tout entier) et qu'ainsi la condition de Lindeberg (3) est trivialement vérifiée. Utilisez ceci pour prouver que :

$$\frac{6}{n^{3/2}} \left(S_n - \frac{n^2}{4} \right) \xrightarrow[n \to +\infty]{\text{loi}} Z,$$

où Z suit la loi $\mathfrak{N}(0,1)$.

9) Justifiez l'affirmation « pour n grand, environ la moitié des permutations sur $[\![1, n]\!]$ ont leur nombre total d'inversions compris entre les bornes $\dfrac{n^2}{4} \pm 0{,}113\, n^{3/2}$ ».

Table des valeurs de Φ, f.d.r. de la loi normale standard $\mathfrak{N}(0,1)$ sur l'intervalle $[1,3]$

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

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