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

IS Math314 **Année 2008**

Examen, 26 juin 2008, durée 3 heures.

- Ce sujet comporte **4 pages**.
- 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, dictionnaire bilingue pour étudiants étrangers.
- Calculatrices autorisées.

Ex 1. *Échauffement gaussien (2 points)*

Pourquoi le théorème limite central est-il évident lorsqu'on l'applique au cas d'une suite de variables aléatoires X_i indépendantes et de même loi gaussienne $\mathfrak{N}(m, \sigma^2)$?

Ex 2. *Estimation du paramètre d'une loi de Poisson (8 points)*

1) La variable aléatoire X suit une loi de Poisson de paramètre $\alpha > 0$. Vérifiez par le calcul que $\mathbf{E}X = \alpha$ et $\operatorname{Var} X = \alpha$.

2) On se propose d'estimer la valeur inconnue de α au vu d'un échantillon de grande taille $X_1, \dots, X_n$ de variables aléatoires indépendantes et de même loi que X. On note $\overline{X}$ la moyenne empirique de l'échantillon (cette moyenne dépend de n). Que pouvez vous dire du comportement de cette moyenne lorsque n tend vers $= +\infty$?

3) Justifiez la convergence en loi suivante :

$$\sqrt{\frac{n}{\overline{X}}}\left(\overline{X} - \alpha\right) \xrightarrow[n\to+\infty]{\text{loi}} Z, \quad Z \sim \mathfrak{N}(0, 1). \tag{1}$$

4) En déduire un intervalle de confiance I au niveau 95% pour α.

5) Le tableau ci-dessous donne un échantillon observé de taille 100.

1	2	3	5	4	2	2	1	0	1	3	3	4	2	6	4	0	3	2	1
3	5	1	2	1	1	4	3	0	2	0	2	4	1	2	4	3	2	2	2
3	2	2	1	4	1	2	3	2	2	3	4	5	3	1	5	2	2	2	2
1	2	3	1	1	3	2	2	3	0	3	4	2	1	0	3	3	1	2	3
1	1	4	1	5	2	5	2	5	2	2	0	2	4	1	6	3	2	5	2

a) Reproduisez et complétez le tableau suivant. Utilisez le pour dessiner la fonction de répartition empirique associée à cet échantillon. Indiquez sur le dessin la hauteur de chaque saut.

Valeur	0	1	2	3	4	5	6
Fréquence	0,07						

b) Calculez la moyenne empirique et la variance empirique de l'échantillon en indiquant comme vous faites pour éviter de rentrer une à une les 100 valeurs dans votre calculatrice.

c) Donnez l'intervalle de confiance I calculé à partir de cet échantillon et l'intervalle de confiance J (toujours au niveau 95%) obtenu en appliquant le théorème limite central avec autonormalisation.

Ex 3. *Étendue d'un échantillon (3 points)*

Si $X_1, \dots, X_n$ est un échantillon, on appelle *étendue* de l'échantillon la statistique :

$$T_n := \max_{1 \leq i \leq n} X_i - \min_{1 \leq i \leq n} X_i.$$

Le but de cet exercie est de démontrer que si $\mathbf{E}|X_1| < +\infty$, $n^{-1}T_n$ converge vers 0 en probabilité. Autrement dit, l'étendue est un $o(n)$ en probabilité.

1) Dans cette question, on suppose les X_i de *même loi*, mais pas forcément indépendantes. Montrez que si $\mathbf{E}|X_1| < +\infty$,

$$\frac{1}{n} \max_{1 \leq i \leq n} |X_i| \xrightarrow[n \to +\infty]{\mathrm{Pr}} 0. \tag{2}$$

Indications : posez $t = n\varepsilon$, décomposez l'évènement $\{\max_{1 \leq i \leq n} |X_i| \geq t\}$ en une union d'évènements dont chacun ne fait intervenir qu'une seule des variables X_k, majorez sa probabilité et utilisez l'inégalité de Markov raffinée.

2) Proposez une majoration simple de l'étendue faisant intervenir $\max_{1 \leq i \leq n} |X_i|$ et concluez.

Ex 4. *Les ex-aequo d'un échantillon (7 points)*

1) On se propose de montrer que si X et Y sont deux variables aléatoires réelles *indépendantes* de fonctions de répartition respectives F et G et si F est *continue*, alors $P(X = Y) = 0$. Voici quelques indications. On note pour $M \in \mathbb{N}^*$,

$$\Delta_M := \{(x, y) \in [-M, M]^2;\ x = y\} \quad \text{et} \quad \Delta := \{(x, y) \in \mathbb{R}^2;\ x = y\}.$$

Il s'agit de prouver que $P_{(X,Y)}(\Delta) = 0$ (pourquoi ?). En remarquant [1] que pour tout $k \in \mathbb{N}^*$,

$$\Delta_M \subset \bigcup_{-Mk \leq j < Mk} \left[\frac{j}{k}, \frac{j+1}{k}\right]^2,$$

utiliser la continuité *uniforme* de F sur le *compact* $[-M, M]$ pour montrer que

$$\forall \varepsilon > 0, \quad P_{(X,Y)}(\Delta_M) \leq \varepsilon\bigl(G(M) - G(-M)\bigr).$$

En déduire que $P_{(X,Y)}(\Delta_M) = 0$, puis que $P_{(X,Y)}(\Delta) = 0$.

1. Dessin impératif !

2) Montrez que si $X_1, \dots, X_n$ est un échantillon de la loi de f.d.r. F continue, presque sûrement il n'y a pas d'*ex-aequo* dans l'échantillon (donc on observe n valeurs distinctes).

3) On suppose que la f.d.r. F est discontinue au point x_0 et on note

$$N_n(x_0) := \mathrm{card}\{i = 1, \dots, n;\ X_i = x_0\} = \sum_{i=1}^{n} \mathbf{1}_{\{X_i = x_0\}}.$$

Montrez que $N_n(x_0) \xrightarrow[n\to+\infty]{\text{p.s.}} +\infty$. Vous devriez pouvoir donner un résultat plus précis que cette affirmation.

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