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

IS Math314 **Année 2007**

Examen deuxième session, 30 juin 2007, durée 3 heures.

- Ce sujet comporte **5 pages**, incluant 2 tables statistiques.
- 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. *Intervalle de confiance (2 points)*

Le tableau ci-dessous donne un 100-échantillon observé d'une loi de Bernoulli de paramètre inconnu p.

1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
0	0	0	1	0	0	0	1	1	1	0	0	0	0	1	0	0	0	0	0
1	0	0	1	1	0	0	0	0	0	1	0	0	0	1	1	1	0	1	1
0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0

1) Calculez simplement la moyenne empirique et la variance empirique de cet échantillon[1].

2) En déduire un intervalle de confiance au niveau 96 % pour p par la méthode avec *variance estimée*. Quel résultat du cours vous permet de justifier cette méthode ?

Ex 2. *Vous reprendrez bien un peu de vecteurs gaussiens ? (4 points)*

Soient X_1 et X_2 deux variables aléatoires définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$, indépendantes et de même loi gaussienne $\mathfrak{N}(0,1)$. Soient a et b des réels vérifiant $a^2 + b^2 = 1$. On définit les variables aléatoires Y_1 et Y_2 par l'égalité matricielle :

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}.$$

1) Expliquez *sans aucun calcul* pourquoi $V = (Y_1, Y_2)$ est un vecteur gaussien de $\mathbb{R}^2$.

2) Calculez le vecteur espérance et la matrice de covariance de (Y_1, Y_2). Comparez les lois des vecteurs (X_1, X_2) et (Y_1, Y_2).

3) Que peut-on dire du vecteur aléatoire image de (X_1, X_2) par une rotation de centre $(0,0)$ et d'angle 38° ?

[1] En évitant ici la méthode consistant à entrer les données une par une dans votre calculatrice !

Ex 3. *Un mélange de gaussiennes (6 points)*

On considère le modèle statistique $\big(\Omega, \mathcal{F}, (P_\theta)_{\theta\in[0,1]}\big)$ et une variable aléatoire X : $\Omega \to \mathbb{R}$, dont la loi sous P_θ a pour fonction de répartition F_θ définie par :

$$\forall\theta \in]0,1[,\ \forall x \in \mathbb{R}, \quad F_\theta(x) = (1-\theta)\Phi\Big(\frac{x-\mu}{\sigma}\Big) + \theta\Phi\Big(\frac{x-\mu}{5\sigma}\Big),$$

où Φ est la fonction de répartition de la loi $\mathfrak{N}(0,1)$. Dans tout l'exercice, les paramètres $\mu \in \mathbb{R}$ et $\sigma \in]0,+\infty[$ sont supposés connus.

1) Quelle est la loi de X sous P_0 ? Sous P_1 ? *Indication :* calculez $P_0\big((X-\mu)/\sigma \leq x\big)$.

2) Expliquez pourquoi la loi de X sous P_θ admet une densité f_θ et calculez la.

3) Calculez $\mathbf{E}_\theta X$.

4) Calculez $\operatorname{Var}_\theta X$.

5) Au vu du calcul précédent, proposez un estimateur T_n fortement consistant de θ, basé sur un échantillon $X_1, \dots, X_n$ de v.a. indépendantes de même loi que X sous P_θ.

6) Lors d'une séance de T.D. sur machine à laquelle vous avez échappé, l'enseignante a demandé à son groupe de proposer une méthode de simulation de la loi de X sous P_θ. L'étudiant G. Duflair a proposé de définir une variable aléatoire Y par

$$Y := (\sigma Z + \mu)\mathbf{1}_{\{U>\theta\}} + (5\sigma Z + \mu)\mathbf{1}_{\{U\leq\theta\}},$$

où U est une v.a. de loi uniforme sur $[0,1]$ et Z *indépendante* de U est gaussienne $\mathfrak{N}(0,1)$. Les variables U et Z sont fournies facilement par le générateur de nombre aléatoires utilisé. G. Duflair s'apprête à en tirer un programme Scilab, quand l'enseignante lui demande de justifier sa réponse. Pouvez vous l'aider ?

Ex 4. *Podomètre (8 points)*

Un podomètre est un appareil qui, fixé à la ceinture d'un marcheur, compte ses pas. Il est aussi capable de calculer approximativement la distance parcourue par le marcheur. La distance parcourue à chaque pas par un marcheur est une variable aléatoire X d'espérance μ et d'écart type σ (ces quantités seront exprimées en mètres). Ces deux paramètres dépendent du marcheur et du type de terrain. On les supposera constants pour simplifier. Notons X_i la distance parcourue par le marcheur lors du i^{e} pas. Nous supposerons en outre que les X_i sont indépendantes et de même loi que X. Ce que le podomètre peut mesurer exactement est le nombre N de pas effectués. Pour estimer la distance parcourue en N pas, il affiche simplement la valeur du produit $N\mu$. En pratique, le podomètre est réglé en usine avec une certaine valeur par défaut μ_0 en mémoire et chaque utilisateur a la possibilité de la remplacer par sa valeur personnelle μ. La distance parcourue lors des k premiers pas du marcheur est notée :

$$S_k := \sum_{i=1}^{k} X_i \quad (k \in \mathbb{N}^*).$$

1) On suppose dans cette question que $\mu = 0{,}75$ m et $\sigma = 0{,}20$ m. Après une randonnée, le podomètre affiche $N = 12\,764$ pas et estime la distance parcourue à

9 573 m. En justifiant votre réponse, proposez un intervalle $[a, b]$ tel que la distance *réellement* parcourue S_N vérifie :

$$P(a \leq S_N \leq b) \simeq 0{,}99.$$

2) Le manuel de l'utilisateur du podomètre explique comment évaluer μ (marcher sur une distance connue, par exemple un kilomètre entre deux bornes sur une route, et diviser cette distance par le nombre de pas). Par contre, il ne dit rien de σ et on peut seulement lire que la distance calculée par le podomètre est fournie avec une *précision relative* de $\pm 10\%$.

a) Trouvez $\varepsilon(N, \mu, \sigma)$ tel que

$$P\Big(1 - \varepsilon(N, \mu, \sigma) \leq \frac{S_N}{N\mu} \leq 1 + \varepsilon(N, \mu, \sigma)\Big) \simeq 0{,}99.$$

b) En supposant μ connue et σ inconnu mais tel que $\sigma/\mu \leq 1/2$, critiquez l'affirmation du manuel sur la précision relative.

3) En fait l'utilisateur du podomètre a quelques notions de statistique. Il lui paraît légitime de supposer que les X_i sont gaussiennes de loi $\mathfrak{N}(\mu, \sigma)$ et il applique la procédure suivante pour estimer μ et σ.

– Il marche sur 10 km de borne à borne le long d'une route et note que son podomètre lui indique $n = 13\,158$ pas. Il en déduit une valeur estimée $\widehat{\mu_n}(\omega)$ de μ. Laquelle ?
– Il enduit ses semelles d'un liquide coloré et effectue 25 pas sur une route goudronnée. Grâce aux traces de ses semelles, il peut mesurer les valeurs observées des v.a. gaussiennes $Z_1, \dots, Z_{25}$, où Z_i est la distance parcourue lors du i^{e} pas. Il calcule alors la variance empirique de cet échantillon et trouve $0{,}008\,5$ m^2. Il détermine ensuite grâce au théorème de Student un intervalle de confiance de la forme $[0, \widehat{\sigma_{25}}(\omega)[$ pour σ, au niveau de confiance 99%. Expliquez comment et donnez la valeur de $\widehat{\sigma_{25}}(\omega)$.

Pourquoi n'utilise-t-il pas l'échantillon $Z_1, \dots, Z_{25}$ pour estimer μ ?

4) Notre utilisateur voudrait bien déduire de son travail une formule d'encadrement de S_N avec grande probabilité (ici N est quelconque). Il introduit alors les évènements *indépendants* :

$$A := \Big\{N\mu - 2{,}575\sigma\sqrt{N} \leq S_N \leq N\mu + 2{,}575\sigma\sqrt{N}\Big\} \tag{1}$$

$$B := \left\{\mu - \frac{2{,}575\sigma}{\sqrt{n}} \leq \widehat{\mu_n} \leq \mu + \frac{2{,}575\sigma}{\sqrt{n}}\right\} \quad (n = 13\,158) \tag{2}$$

$$C := \{\sigma < \widehat{\sigma_{25}}\}. \tag{3}$$

a) Justifiez brièvement les *égalités exactes*[2] $P(A) = P(B) = P(C) = 0{,}99$.

b) En déduire que $P(A \cap B \cap C) \geq 0{,}97$.

c) En déduire un encadrement de S_N vrai avec une probabilité d'au moins 97%, dont les bornes s'expriment en fonction de N et des valeurs estimées ci-dessus[3] pour μ et σ.

[2] En négligeant les erreurs d'approximation numérique dans le calcul de la f.d.r. Φ de la loi $\mathfrak{N}(0, 1)$.

[3] On ne vous demande pas de justification rigoureuse du fait que l'on peut remplacer les v.a. $\widehat{\mu_n}$ et $\widehat{\sigma_{25}}$ par les valeurs numériques trouvées ci-dessus. En fait, c'est là que l'indépendance de A, B et C serait utile.

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

Table des valeurs du réel x_q tel que $P(X > x_q) = q$, pour X de loi $\chi^2(d)$

d \ q	0,999	0,995	0,990	0,975	0,95	0,90	0,50	0,10	0,05	0,025	0,01	0,005	0,001
1	0,000	0,000	0,000	0,001	0,004	0,016	0,455	2,706	3,841	5,024	6,635	7,879	10,828
2	0,002	0,010	0,020	0,051	0,103	0,211	1,386	4,605	5,991	7,378	9,210	10,597	13,816
3	0,024	0,072	0,115	0,216	0,352	0,584	2,366	6,251	7,815	9,348	11,345	12,838	16,266
4	0,091	0,207	0,297	0,484	0,711	1,064	3,357	7,779	9,488	11,143	13,277	14,860	18,467
5	0,210	0,412	0,554	0,831	1,145	1,610	4,351	9,236	11,070	12,833	15,086	16,750	20,515
6	0,381	0,676	0,872	1,237	1,635	2,204	5,348	10,645	12,592	14,449	16,812	18,548	22,458
7	0,598	0,989	1,239	1,690	2,167	2,833	6,346	12,017	14,067	16,013	18,475	20,278	24,322
8	0,857	1,344	1,646	2,180	2,733	3,490	7,344	13,362	15,507	17,535	20,090	21,955	26,124
9	1,152	1,735	2,088	2,700	3,325	4,168	8,343	14,684	16,919	19,023	21,666	23,589	27,877
10	1,479	2,156	2,558	3,247	3,940	4,865	9,342	15,987	18,307	20,483	23,209	25,188	29,588
11	1,834	2,603	3,053	3,816	4,575	5,578	10,341	17,275	19,675	21,920	24,725	26,757	31,264
12	2,214	3,074	3,571	4,404	5,226	6,304	11,340	18,549	21,026	23,337	26,217	28,300	32,909
13	2,617	3,565	4,107	5,009	5,892	7,042	12,340	19,812	22,362	24,736	27,688	29,819	34,528
14	3,041	4,075	4,660	5,629	6,571	7,790	13,339	21,064	23,685	26,119	29,141	31,319	36,123
15	3,483	4,601	5,229	6,262	7,261	8,547	14,339	22,307	24,996	27,488	30,578	32,801	37,697
16	3,942	5,142	5,812	6,908	7,962	9,312	15,338	23,542	26,296	28,845	32,000	34,267	39,252
17	4,416	5,697	6,408	7,564	8,672	10,085	16,338	24,769	27,587	30,191	33,409	35,718	40,790
18	4,905	6,265	7,015	8,231	9,390	10,865	17,338	25,989	28,869	31,526	34,805	37,156	42,312
19	5,407	6,844	7,633	8,907	10,117	11,651	18,338	27,204	30,144	32,852	36,191	38,582	43,820
20	5,921	7,434	8,260	9,591	10,851	12,443	19,337	28,412	31,410	34,170	37,566	39,997	45,315
21	6,447	8,034	8,897	10,283	11,591	13,240	20,337	29,615	32,671	35,479	38,932	41,401	46,797
22	6,983	8,643	9,542	10,982	12,338	14,041	21,337	30,813	33,924	36,781	40,289	42,796	48,268
23	7,529	9,260	10,196	11,689	13,091	14,848	22,337	32,007	35,172	38,076	41,638	44,181	49,728
24	8,085	9,886	10,856	12,401	13,848	15,659	23,337	33,196	36,415	39,364	42,980	45,559	51,179
25	8,649	10,520	11,524	13,120	14,611	16,473	24,337	34,382	37,652	40,646	44,314	46,928	52,620
26	9,222	11,160	12,198	13,844	15,379	17,292	25,336	35,563	38,885	41,923	45,642	48,290	54,052
27	9,803	11,808	12,879	14,573	16,151	18,114	26,336	36,741	40,113	43,195	46,963	49,645	55,476
28	10,391	12,461	13,565	15,308	16,928	18,939	27,336	37,916	41,337	44,461	48,278	50,993	56,892
29	10,986	13,121	14,256	16,047	17,708	19,768	28,336	39,087	42,557	45,722	49,588	52,336	58,301
30	11,588	13,787	14,953	16,791	18,493	20,599	29,336	40,256	43,773	46,979	50,892	53,672	59,703