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

IS Math314 **Année 2009**

Examen 1^re session, 12 juin 2009, 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, dictionnaire bilingue pour étudiants étrangers.
- Calculatrices autorisées.

Ex 1. *Intervalles de confiance pour une probabilité inconnue (4 points)*

Dans cet exercice qui est une question de cours à peine déguisée, on dispose d'un 400-échantillon observé $(X_i(\omega))_{1\leq i\leq 400}$ d'une loi de Bernoulli. Le paramètre p de cette loi est inconnu, mais on sait avec certitude que $0 < p < 0{,}3$. Le nombre de valeurs 1 dans l'échantillon observé est 72.

1) Compte-tenu de ces informations, quel est le meilleur majorant que l'on peut proposer pour $\operatorname{Var} X_1$?

2) Donnez un intervalle de confiance I au niveau 95% pour p en utilisant la méthode avec variance *majorée* et en indiquant avec précision sur quel théorème du cours vous vous appuyez.

3) On note $\overline{X}$ et S^2 la moyenne et la variance empiriques de l'échantillon $(X_i)_{1\leq i\leq 400}$. Vérifiez que

$$S^2 = \overline{X}(1-\overline{X}).$$

Indication : comparez X_i et X_i^2 lorsque X_i est à valeurs dans $\{0,1\}$.

4) Donnez un intervalle de confiance J au niveau 95% pour p en utilisant la méthode avec variance *estimée* et en indiquant avec précision sur quel théorème du cours vous vous appuyez.

Ex 2. *Matrices de covariance et couples gaussiens (7 points)*

En dimension 1, il est facile de vérifier que tout nombre réel positif peut être la variance d'une certaine variable aléatoire (sauriez vous expliquer pourquoi en trois lignes maximum ?). On se propose d'examiner une question analogue en dimension 2 et d'en déduire une classification des lois de vecteurs gaussiens de $\mathbb{R}^2$.

1) Montrez que si M est la matrice de covariance d'un vecteur aléatoire (X,Y) de $\mathbb{R}^2$, elle est nécessairement de la forme :

$$M = \begin{pmatrix} a & c \\ c & b \end{pmatrix}, \qquad a,b\in\mathbb{R}^+,\ c\in\mathbb{R},\quad ab-c^2\geq 0. \tag{1}$$

Réciproquement, soit M une matrice 2×2 de la forme (1), on se propose de montrer que M est la matrice de covariance d'un certain couple aléatoire (X,Y) de $\mathbb{R}^2$.

2) Dans le cas particulier où $ab = 0$, vérifiez que M est la matrice de covariance d'un vecteur aléatoire ayant au moins une composante constante.

3) On suppose désormais que ab est non nul. On pose alors

$$\sigma_1 = \sqrt{a},\ \sigma_2 = \sqrt{b}, \quad \rho = \frac{c}{\sqrt{ab}} \quad (\text{donc } \rho \in [-1,1]),$$

on se donne un couple (U,V) de variables aléatoires réelles *indépendantes*, de carré intégrable et de variance 1 et on définit l'application linéaire g par :

$$g : \mathbb{R}^2 \to \mathbb{R}^2, \quad (u,v) \longmapsto (x,y),\ \text{avec} \begin{cases} x = \sigma_1 u, \\ y = \sigma_2 \rho u + \sigma_2 \sqrt{1-\rho^2}\, v. \end{cases} \tag{2}$$

Vérifiez que le vecteur aléatoire $(X,Y) = g(U,V)$ a pour matrice de covariance M.

4) Que peut-on dire de la loi de (X,Y) lorsque U et V ci-dessus sont gaussiennes $\mathfrak{N}(0,1)$?

5) Soit (X',Y'), un vecteur aléatoire gaussien d'espérance $(0,0)$ et de matrice de covariance K. Utilisez ce qui précède pour montrer que si le déterminant de K est non nul, alors la loi de (X',Y') a une densité qui se déduit de celle de (U,V) par un changement de variable linéaire. Donnez l'expression de cette densité à l'aide des variances σ_1^2 et σ_2^2 de X' et Y' et de leur coefficient de corrélation.

6) Montrez que si $\det K = 0$, soit le vecteur aléatoire (X',Y') est p.s. constant égal à $(0,0)$, soit sa loi est portée par une droite D, c'est-à-dire $P((X',Y') \in D) = 1$, dont vous donnerez l'équation.

Ex 3. *Estimation de la moyenne d'une loi de Poisson mélange (10 points)*

Pour les assurances, le nombre N de sinistres causés en une année par un assuré donné est souvent modélisé par une loi de Poisson de paramètre α. On a alors $\mathbf{E}N = \alpha$, ce qui permet d'interpréter le paramètre α comme un nombre moyen de sinistres. Ce modèle ne colle pas bien à la réalité lorsque l'on étudie l'ensemble des assurés correspondant à une police donnée (par exemple les conducteurs de véhicule d'un certain type dans une certaine zone géographique). Pour prendre en compte l'hétérogénéité des assurés, on remplace alors α par $R\alpha$ où R est une variable aléatoire positive d'espérance 1, appelée niveau de risque relatif. Le modèle « bons risques – mauvais risques » en assurance automobile est le plus simple de ce type. Il revient à partager l'ensemble d'assurés étudié en « bons conducteurs » pour lesquels le nombre de sinistres suit la loi de Poisson de paramètre $r_1\alpha$ (avec $r_1 < 1$) et en « mauvais conducteurs » pour lesquels le nombre de sinistres suit la loi de Poisson de paramètre $r_2\alpha$ avec $r_2 > 1$. Le problème est qu'il est difficile de décider au vu de l'historique des accidents qui est bon ou mauvais conducteur. Parmi les assurés n'ayant pas causé de sinistre lors des années précédentes, il peut très bien y avoir des mauvais conducteurs « chanceux » et parmi les assurés ayant causé au moins un sinistre, il peut tout aussi bien y avoir des bons conducteurs « malchanceux ». On considère donc que le niveau de risque relatif d'un assuré choisi au hasard est une variable aléatoire R vérifiant :

$$P(R = r_1) = p_1 > 0,\ P(R = r_2) = p_2 > 0, \quad p_1 + p_2 = 1, \quad \mathbf{E}R = 1. \tag{3}$$

Les paramètres p_i, r_i peuvent être estimés par l'observation approfondie des comportements au volant d'un échantillon d'assurés. Dans toute la suite, nous les supposerons connus. Le seul paramètre inconnu de notre modèle sera donc α. On s'intéresse au nombre N de sinistres causés par un assuré choisi au hasard (on ne sait pas s'il est bon ou mauvais conducteur).

1) Vérifiez que la loi de N est donnée par

$$\forall k \in \mathbb{N}, \quad P(N = k) = p_1 \frac{\mathrm{e}^{-r_1\alpha}(r_1\alpha)^k}{k!} + p_2 \frac{\mathrm{e}^{-r_2\alpha}(r_2\alpha)^k}{k!}. \tag{4}$$

On dit que N suit la loi de *Poisson mélange* de moyenne α et de niveau de risque relatif R donné par (3).

2) Vérifiez que

$$\mathbf{E}N = \alpha, \tag{5}$$

$$\operatorname{Var} N = \alpha + \alpha^2 \operatorname{Var} R. \tag{6}$$

On remarque que la loi de N est plus dispersée autour de α qu'une loi de Poisson classique de paramètre α dont l'espérance et la variance valent α.

3) On considère désormais le modèle statistique $(\Omega, \mathcal{F}, P_\alpha)_{\alpha \in]0,+\infty[}$ et un n-échantillon $N_1, \dots, N_n$ associé. Autrement dit, pour toute valeur de α, les variables aléatoires N_i sont P_α-indépendantes et de même loi sous P_α donnée par (4). On note $\overline{N}$ la moyenne empirique de cet échantillon. Justifiez brièvement les deux convergences suivantes :

$$\forall \alpha > 0, \quad \overline{N} \xrightarrow[n\to+\infty]{P_\alpha-\text{p.s.}} \alpha; \tag{7}$$

$$\forall \alpha > 0, \quad \sqrt{\frac{n}{\operatorname{Var}_\alpha N_1}} \big(\overline{N} - \alpha\big) \xrightarrow[n\to+\infty]{P_\alpha-\text{loi}} Z, \quad Z \sim \mathfrak{N}(0,1). \tag{8}$$

4) Justifiez la convergence en loi suivante où l'on a posé $\rho^2 = \operatorname{Var} R$ (la loi de R est connue et ne dépend pas de α) :

$$\forall \alpha > 0, \quad \sqrt{\frac{n}{\frac{1}{n} + \overline{N}(1 + \rho^2 \overline{N})}} \big(\overline{N} - \alpha\big) \xrightarrow[n\to+\infty]{P_\alpha-\text{loi}} Z, \quad Z \sim \mathfrak{N}(0,1). \tag{9}$$

5) En déduire un intervalle de confiance I au niveau 95% pour α. Pourquoi ne pouvait-on utiliser (8) pour obtenir un tel intervalle ?

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

0	1	0	2	1	0	0	2	0	0	0	0	1	0	0	1	0	0	2	0
0	1	1	0	0	0	2	0	0	1	0	1	0	1	1	0	0	0	2	1
1	0	0	2	0	0	1	1	1	0	1	0	0	1	0	0	0	1	0	0
0	3	1	0	0	0	0	1	1	1	1	0	0	1	0	0	1	0	0	0
1	0	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	2	0	3

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. Unité verticale conseillée 10 cm.

Valeur	0	1	2	3
Fréquence				0,02

b) Calculez la moyenne empirique $\overline{N}$ de cet échantillon observé en indiquant comment vous faites pour éviter de rentrer une à une les 100 valeurs dans votre calculatrice.

c) On « sait » que $p_1 = 0{,}25$, $r_1 = 0{,}7$ et $r_2 = 1{,}1$. Calculez $\rho^2 = \operatorname{Var} R$.

d) Donnez l'intervalle de confiance I calculé à partir de cet échantillon.

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