Université Claude Bernard - Lyon 1 — **Semestre automne 2024-2025**

L2 - Informatique — UE : Statistiques pour l'Informatique

Corrigé de l'examen final - 15/01/2025

durée 90 minutes

Les calculatrices sont permises en mode "examen". Les documents ne sont pas permis. Toute réponse doit être justifiée. Le barême est indicatif. L'examen sera notée sur 20 points.

Exercice 1. 6.5 pts. Dans une expérience visant à vérifier la loi de Hooke, la longueur l d'une corde de piano sur laquelle est suspendu un poids de masse m est mesurée. On obtient les valeurs suivantes :

masse m en kg	longueur l en cm
0	439,00
2	439,12
4	439,21
6	439,31
8	439,42
10	439,50

1. Calculez le coéfficient de corrélation $cor(m, l)$.
2. Déterminez les coefficients a et b de l'équation $l = am + b$ de la droite de régression.
3. Si on converti les masses m en grammes, comment cela affecte les grandeurs $cor(m, l)$, a et b ?

Corrigé : La moyenne de la masse est $\bar{m} = \frac{1}{6}\sum_{i=1}^{6} m_i = 5kg$.
La moyenne de la longeur est $\bar{l} = \frac{1}{6}\sum_{i=1}^{6} = l_i = 439,26cm$
La variance de la masse est

$$V(m) = \frac{1}{6}\sum_{i=1}^{6}(\bar{m} - m_i)^2 = \frac{25+9+1+1+9+25}{6}kg^2 = \frac{70}{6}kg^2 = 11,67kg^2$$

La variance de la longueur est

$$V(l) = \frac{1}{6}\sum_{i=1}^{6}(\bar{l} - l_i)^2 = \frac{0,0676+0,0196+0,0025+0,0025+0,0256+0,0576}{6}cm^2 = \frac{0,1754}{6}cm^2 = 0,0292cm^2$$

La covariance est

$$\begin{aligned} V(m,l) &= \frac{1}{6}\sum_{i=1}^{6}(\bar{m} - m_i)(\bar{l} - l_i) \\ &= \frac{-5\times -0,26 + -3\times -0,14 + -1\times -0,05 + 1\times 0,05 + 3\times 0,16 + 5\times 0,24}{6}kg\,cm \\ &= \frac{1,3+0,42+0,05+0,05+0,48+1,2}{6}kg\,cm = \frac{3,5}{6}kg\,cm \\ &= 0,583kg\,cm \end{aligned}$$

1. Le coefficient de corrélation est

$$cor(m,l) = \frac{V(m,l)}{\sqrt{V(m)}\sqrt{V(l)}} = \frac{3,5}{\sqrt{70\times 0,1754}} = 0,999$$

2. Les coefficients de la droite de régression sont

$$a = \frac{V(m,l)}{V(m)} = \frac{3,5}{70}cm\,kg^{-1} = 0,05cm\,kg^{-1}$$

$$b = \bar{l} - a\bar{m} = (439,26 - 0,05\times 5)cm = 439,01cm$$

3. Si on converti les masses m en grammes, alors les valuers de m sont à multiplier avec 1000. Donc $V(m)$ est à multiplier avec 1000^2 pendant que $V(m,l)$ est à multiplier avec 1000. D'où a est a diviser par 1000 pendant que le coefficient de corrélation et b ne changent pas.

<u>Exercice 2.</u> 5.5 pts. Soient X, Y deux variables aléatoires indépendantes de loi de binomiale $\mathcal{B}(n, \frac{1}{3})$.

1. Quelle est l'espérance de $X - Y$?
2. Quelle est la variance de $X - Y$?
3. Déterminez la loi du couple (X, Y) dans le cas $n = 2$.
4. Déterminez la loi de $X - Y$ dans le cas $n = 2$.
5. Quelle est la probabilité que $X > Y$ dans le cas $n = 2$?

Corrigé :

1. Par linéarité de l'espérance on a $E(X - Y) = E(X) - E(Y)$. X et Y étant de même loi, elles ont la même espérance, on a donc $E(Y) = E(X)$, et ainsi, $E(X - Y) = 0$.
2. Les variables aléatoires X et Y étant indépendantes, il en va de même pour les variables X et $-Y$, ce qui implique que $V(X - Y) = V(X) + V(-Y)$. Par la formule $V(\lambda Y) = \lambda^2 V(Y)$ appliquée en $\lambda = -1$, on obtient $V(-Y) = V(Y)$. X et Y suivant la loi $\mathcal{B}(n, \frac{1}{3})$ on en déduit que
$$V(X - Y) = V(X) + V(Y) = \frac{4}{9}n$$
3. Le couple est à valeur dans l'ensemble des tuples (k, l) avec $k, l \in \{0, 1, 2\}$. En outre par indépendance, et comme X et Y ont la même loi, on a $P(X = k, Y = l) = P(X = k)P(Y = l) = P(X = k)P(Y = l)$. Or $P(X = k) = \binom{2}{k}\left(\frac{1}{3}\right)^k\left(1 - \frac{1}{3}\right)^{2-k} = \binom{2}{k}\left(\frac{1}{3}\right)^k\left(\frac{2}{3}\right)^{2-k}$. On obtient le tableau

k	0	1	2
$P(X = k)$	$\frac{4}{9}$	$\frac{4}{9}$	$\frac{1}{9}$

On peut alors obtenir la table de la loi conjointe

$P(X = k, Y = l)$	0	1	2
0	$\frac{16}{81}$	$\frac{16}{81}$	$\frac{4}{81}$
1	$\frac{16}{81}$	$\frac{16}{81}$	$\frac{4}{81}$
2	$\frac{4}{81}$	$\frac{4}{81}$	$\frac{1}{81}$

4. La variable Z prend ses valeures dans l'ensemble $\{-2, -1, 0, 1, 2\}$. On calcul chaque valeurs à partir de la table de loi conjointe de (X, Y).
   — $P(Z = -2) = P(X = 0, Y = 2) = \frac{4}{81}$.
   — $P(Z = -1) = P(X = 0, Y = 1) + P(X = 1, Y = 2) = \frac{20}{81}$.
   — $P(Z = 0) = P(X = 0, Y = 0) + P(X = 1, Y = 1) + P(X = 2, Y = 2) = \frac{33}{81}$.
   — $P(Z = 1) = P(X = 1, Y = 0) + P(X = 2, Y = 1) = \frac{20}{81}$.
   — $P(Z = 2) = P(X = 2, Y = 0) = \frac{4}{81}$.

   On obtient donc la table

k	-2	-1	0	1	2
$P(Z = k)$	$\frac{4}{81}$	$\frac{20}{81}$	$\frac{33}{81}$	$\frac{20}{81}$	$\frac{4}{81}$

5. L'évènement $X > Y$ équivaut à $Z = X - Y > 0$, soit $Z = 2$ ou $Z = 1$. Ainsi
$$P(X > Y) = P(Z = 2) + P(Z = 1) = \frac{4}{81} + \frac{20}{81} = \frac{24}{81} = \frac{8}{27}$$

Exercice 3. 5.5 pts. Soit $N \geq 1$ un entier. Un sac contient 1 boule rouge et $N-1$ boules blanches. On étudie deux variantes d'un jeux : tirez une boule après l'autre du sac avec ou sans remise. Vous jouez jusqu'à tirer la boule rouge. Chaque droit de tirage vous coute 1 Euro. Si vous avez tiré la boule rouge, vous gagnez la somme de S Euro. Soit X la variable aléatoire qui designe la première fois que vous avez obtenue la boule rouge.

1. Vous tirez une boule après l'autre du sac *avec* remise.
   (a) Décrire la loi de X, si c'est une loi usuelle la citer avec son ou ses paramètres. Quelle est espérance et variance de X ?
   (b) A partir de quelle somme S ca vaut la peine de jouer ce jeu ?
2. Vous tirez une boule après l'autre du sac *sans* remise.
   (a) Décrire la loi de X, si c'est une loi usuelle la citer avec son ou ses paramètres. Quelle est espérance et variance de X ?
   Indication : Qu'est-ce qui se passe si $N = 2$ ou $N = 3$?
   (b) A partir de quelle somme S ca vaut la peine de jouer ce jeu ?

Corrigé :

1. (a) La probabilité de tirer une boule rouge est la même pour chaque tirage, car on remet la boule. Comme il y a N boules dont une seule rouge, la proba de tirer une boule rouge est $p = \frac{1}{N}$. Les tirages sont indépendantes. Comme il y a N boules dont une seule rouge, la proba de tirer une boule rouge est $p = \frac{1}{N}$. Donc X suit la loi géométrique (premier succès) de paramètre $p = \frac{1}{N}$. Son éspérence est
$$E(X) = \frac{1}{p} = N$$
sa variance
$$V(X) = \frac{1-p}{p^2} = (N-1)N$$
(b) L'espérence du gain est $\sum_{i=1}^{\infty} 1P(X = i) = E(X) = N$. Donc si $S > N$ ca vaut la peine de jouer le jeu.
2. (a) La probabilité de tirer une boule rouge au premier tire est $\frac{1}{N}$. Donc $P(X = 1) = \frac{1}{N}$. L'évènement de tirer la boule au deuxieme tire correspond à avoir tiré une boule blanche pour le premier tire puis tirer une boule rouge. Comme il n'y a pas de remise c'est $P(X = 2) = \frac{N-1}{N}\frac{1}{N-1} = \frac{1}{N}$, pourvu que $N \geq 2$. Le même argument donne $P(X = k) = \frac{N-1}{N}\frac{N-2}{N-1}\cdots\frac{1}{N-(k-1)} = \frac{1}{N}$, pourvu que $N \geq k$. Clairement, X ne peut pas prendre des valeurs $> N$. Donc X suit la loi uniforme de paramètre N. Son éspérence est
$$E(X) = \frac{N+1}{2}$$
sa variance
$$V(X) = \frac{N^2-1}{12}$$
(b) L'espérence du gain est $\sum_{i=1}^{\infty} 1P(X = i) = E(X) = \frac{N+1}{2}$. Donc si $S > \frac{N+1}{2}$ ca vaut la peine de jouer le jeu.

Exercice 4. 6.5 pts. Pour un jeu on utilise un dé spécial, qui a deux faces avec 1 trou, deux faces avec 2 trous et deux faces avec 3 trous. Soit X la variable aléatoire qui designe le nombre de trous d'un jet.

1. Le Maître du jeu vous dit que le dé n'est pas truqué, c.a.d. la probabilité d'obtenir une face spécifique est uniforme. Vous faites confiance au Maître pour le raisonnement suivant :
   (a) Quelle est la loi de X, son espérance et son écart type ?
   (b) Soit Y la variable aléatoire qui designe le nombre total des trous obtenus en 100 jets. Quelle est l'espérance et l'écart type de Y ?
   (c) Soit Z la variable aléatoire centrée réduite associé à Y. En approchant la loi de Z avec une loi normale, estimez la probablité que la somme des trous des 100 jets soit plus grand que 210.
2. Vous avez joué le jeu, avez jeté le dé 100 fois et obtenu 23 fois 1, 34 fois 2 et 43 fois 3 trous. À cause de ce résultat vous avez un doute que le Maître a dit la vérité.
   (a) Donnez un intervalle de confiance de niveau 95% pour la moyenne de X. Vous pouvez approcher la loi de Student par la loi normale $\mathcal{N}(0,1)$, mais l'écart type de la loi de X n'est pas connu.
   (b) Est-ce que cet intervalle est compatible avec l'affirmation du Maître ?

Corrigé

1. (a) La variable X prend ses valeurs dans l'ensemble $\{1,2,3\}$. Chaque nombre de trou correspondant à exactement deux face, on trouve que X suit la loi uniforme $\mathcal{U}(3)$. En particulier, on obtient $E(X) = \frac{3+1}{2} = 2$, $V(X) = \frac{3^2-1}{12} = \frac{8}{12} = \frac{2}{3}$ et $\sigma(X) = \sqrt{\frac{2}{3}} \simeq 0.81$.
   (b) On peut écrire $Y = \sum_{i=1}^{100} X_i$ avec X_i des copies indépendantes de X. Par linéarité de l'espérance, et comme les X_i ont tous la même loi on a
   $$E(Y) = \sum_{i=1}^{100} E(X_i) = \sum_{i=1}^{100} E(X) = 100 \times E(X) = 200$$
   De même, par propriété d'additivité de la variance sur les sommes de variables indépendantes on a
   $$V(Y) = \sum_{i=1}^{100} V(X_i) = 100 \times V(X) = \frac{200}{3}$$
   Donc $\sigma(Y) = \sqrt{\frac{200}{3}} = 10\sqrt{\frac{2}{3}} \simeq 8.1$.
   (c) La variable Z est donnée par $Z = \frac{Y-\mu}{\sigma}$ avec $\mu = E(Y) = 200$ et $\sigma = \sigma(Y) = 10\sqrt{23} \simeq 8.1$. Comme $Y \geq 210$ équivaut à $Z \geq \frac{10}{\sigma} = \sqrt{32} \simeq 1.22$, et en utilisant la loi des grands nombre, on a
   $$P(Y \geq 210) = P(Z \geq \sqrt{3/2}) = 1 - P(Z < \sqrt{3/2}) \simeq 1 - F(1.22) \simeq 1 - 0.8888 \simeq 0.1112$$
   où F est la fonction de répartition de la loi $\mathcal{N}(0,1)$.
2. (a) La formule de l'intervalle de confiance à 95% est donnée par
   $$IC_\alpha = [\mu - \frac{s}{\sqrt{n}} t(n)_{\alpha/2}, \mu + \frac{s}{\sqrt{n}} t(n)_{\alpha/2}]$$
   où la valeur de $t(n)_{\alpha/2}$ est donnée par la loi de student. On peut ici remplacer cette loi par la loi normale, i.e. remplacer $t(n)_{\alpha/2}$ par $z_{\alpha+2}$. On a $n = 100$, μ est la moyenne empirique, s^2 la variance empirique sans biais et $\alpha = 0.95$. On calcul
   $$\mu = \frac{23 \times 1 + 34 \times 2 + 43 \times 3}{100} = 2.2$$

$$s^2 = \frac{23 \times (1-2.2)^2 + 34 \times (2-2.2)^2 + 43 \times (3-2.2)^2}{99} \simeq 0.63 \quad s \simeq 0.79$$

En outre $z_{\alpha/2} = 1.96$. On en déduit que l'intervalle est donné par $[2.07, 2.38]$.

(b) Avec95% de chance, l'espérance théorique doit se trouver dans l'intervalle de confiance. Or le maitre énonce une telle espérance de 2, qui n'est pas dans l'intervalle. Il est donc raisonnable de supposer que le maitre a menti.

Les lois discrètes usuelles.

	notation	valeurs de X	$P(X=k)$	espérance	variance
loi de Bernoulli	$\mathcal{B}(p)$	$k \in \{0,1\}$	$p^k(1-p)^{1-k}$	p	$p(1-p)$
loi binomiale	$\mathcal{B}(n,p)$	$k \in \{0,1,\cdots,n\}$	$\binom{n}{k} p^k(1-p)^{n-k}$	np	$np(1-p)$
loi géométrique	$\mathcal{G}(p)$	$k \in \mathbb{N}^*$	$p(1-p)^{k-1}$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
loi de Poisson	$\mathcal{P}(\lambda)$	$k \in \mathbb{N}$	$\exp(-\lambda)\frac{\lambda^k}{k!}$	λ	λ
loi uniforme	$\mathcal{U}(n)$	$k \in \{1,2,\cdots,n\}$	$\frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$

Loi normale $\mathcal{N}(0,1)$: tableau de valeurs de $z_{\frac{\alpha}{2}}$ telles que $\displaystyle\int_{-z_{\frac{\alpha}{2}}}^{z_{\frac{\alpha}{2}}} \rho_{\mathcal{N}(0,1)}(x)dx = 1-\alpha$

$1-\alpha$	0.80	0.85	0.90	0.95	0.99
$z_{\frac{\alpha}{2}}$	1.28	1.44	1.645	1.96	2.58

Valeurs $\mathcal{F}(z)$ pour la fonction de répartition de la loi $\mathcal{N}(0,1)$.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
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	0.5753
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	0.6628	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	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	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	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	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	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	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	0.9484	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	0.9649	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	0.9772	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	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	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

Grandes valeurs de z :

z	3.0	3.1	3.2	3.3	3.4	3.5	4.0	4.5
$F(z)$	0.9987	0.99904	0.99931	0.99952	0.99966	0.99976	0.999968	0.999997