Probabilités discrètes
et Statistiques descriptives

L2 Université Lyon 1 2019

Examen final du 23 mai 2019
Durée : 2 h
Les documents et calculatrices sont interdits. Deux pages
On prendra soin à JUSTIFIER les réponses aux exercices

- On note toujours $\mathbb{E}(Z)$ l'espérance d'une variable aléatoire Z et $\mathrm{Var}(Z)$ sa variance.
- On notera $\mathbf{1}_A$ la fonction indicatrice d'un ensemble A, c'est-à-dire que $\mathbf{1}_A(x) = 1$ si $x \in A$ et $\mathbf{1}_A(x) = 0$ si $x \notin A$.
- On rappelle que la densité d'une variable de loi uniforme sur $[0,1]$ est donnée par $g(x) = \mathbf{1}_{[0,1]}(x)$ pour $x \in \mathbb{R}$.
- Tableau de valeurs de fractiles de la loi normale
  $P(N \leq z_\alpha) = 1 - \alpha$ pour N de loi normale centrée réduite $\mathcal{N}(0,1)$.

α	0.2	0.1	0.05	0.025	0.01
z_α	0.842	1.282	1.645	1.96	2.326

Questions de cours (10 minutes maxi, 4 points) (cf cours)

Exercice 1. (4 points)

Soit $(X;Y)$ un couple de variables aléatoires discrètes à valeurs dans $\{0,1,2\} \times \{-1,1\}$ et dont la loi de probabilité est indiquée dans le tableau ci-dessous, où a est un réel :

$X \setminus\!\setminus Y$	-1	1
0	$\frac{1}{12}$	$\frac{1}{12}$
1	$\frac{1}{6}$	$\frac{1}{6}$
2	$\frac{1}{6}$	a

1. Pour calculer le nombre réel a, il suffit d'utiliser la condition que la somme des probabilité vaut
$$1 = \frac{1}{12} + \frac{1}{12} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + a$$
donc : $a + 2/3 = 1$ donc $a = 1/3$.
2. Donnons la loi de X (c'est à dire la probabilité de chaque valeur possible).
$$P(X=0) = P(X=0, Y=1) + P(X=0, Y=-1) = \frac{1}{12} + \frac{1}{12} = \frac{1}{6}$$
$$P(X=1) = P(X=1, Y=1) + P(X=1, Y=-1) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$
$$P(X=2) = P(X=2, Y=1) + P(X=1, Y=-1) = \frac{1}{6} + a = \frac{1}{2}$$
Donnons la loi de Y :
$$P(Y=-1) = P(X=0, Y=-1) + P(X=1, Y=-1) + P(X=2, Y=-1) = \frac{1}{12} + \frac{1}{6} + \frac{1}{6} = \frac{5}{12}$$
donc $P(Y=1) = 1 - P(Y=0) = \dfrac{7}{12}$.

3. $P(X=0, Y=1) = \frac{1}{12} \neq P(X=0)P(Y=1) = \frac{1}{6}\frac{5}{12}$ vu $\frac{5}{6} \neq 1$
donc les variables X et Y ne sont PAS indépendantes.
4. Calculons les espérances de X et Y.

$$E(X) = \sum_{k=0}^{2} kP(X=k) = 0P(X=0) + 1P(X=1) + 2P(X=2) = 0\frac{1}{6} + 1\frac{1}{3} + 2\frac{1}{2} = \frac{4}{3}.$$

$$E(Y) = -1P(Y=-1) + 1P(Y=1) = -1\frac{5}{12} + 1\frac{7}{12} = \frac{1}{6}.$$

5. Calculons la variance de X : $Var(X) = E(X^2) - (E(X))^2$.
Or

$$E(X^2) = \sum_{k=0}^{2} k^2 P(X=k) = 0P(X=0) + 1P(X=1) + 2^2 P(X=2) = 0\frac{1}{6} + 1\frac{1}{3} + 4\frac{1}{2} = \frac{7}{3}.$$

Donc, on a :

$$Var(X) = \frac{7}{3} - \frac{4^2}{3^2} = \frac{21-16}{9} = \frac{5}{9}$$

6. Pour calculer la loi du maximum de X et Y, c'est-à-dire de la variable $Z = \max(X;Y)$, on trouve les valeurs possibles à savoir $\{0,1,2\}$ et on décompose les évènements : $\{Z=0\} = \{X=0, Y=-1\}$ car -1 est la seul valeur de Y en dessous de 0. Par contre $X=2$ est au dessus de toutes les valeurs de Y donc $\{Z=2\} = \{X=2\}$.
Donc $P(Z=0) = \frac{1}{12}$, $P(Z=2) = \frac{1}{2}$, on a :

$$P(Z=1) = 1 - P(Z=0) - P(Z=2) = 1 - \frac{1}{12} - \frac{1}{2} = \frac{5}{12}$$

7. Calculons l'espérance de la variable $Z = \max(X;Y)$:

$$E(Z) = 0P(Z=0) + 1P(Z=1) + 2P(Z=2) = \frac{5}{12} + 2\frac{1}{2} = \frac{17}{12}.$$

Exercice 2. (6 points)

Soit T une variable aléatoire à valeurs dans $\mathbb{N}$. On suppose que pour $n > 0$:

$$\mathbb{P}(T=n) = \frac{\mathrm{e}^{-1/2}}{2^{n+1}\, n!}.$$

1. Pour calculer $\mathbb{P}(T=0)$ on impose la condition $P(T=0) = 1 - P(T>0)$, or

$$P(T>0) = \sum_{k=1}^{\infty} P(T=k) = \sum_{k=1}^{\infty} \frac{\mathrm{e}^{-1/2}}{2^{n+1}\, n!} = \frac{\mathrm{e}^{-1/2}}{2}\left(\sum_{k=0}^{\infty} \frac{1}{2^n\, n!} - 1\right) = \frac{\mathrm{e}^{1/2}}{2}(\mathrm{e}^{-1/2} - 1) = \frac{1}{2}(1 - \mathrm{e}^{-1/2}).$$

Ainsi, on obtient :

$$\mathbb{P}(T=0) = 1 - \frac{1}{2}(1 - \mathrm{e}^{-1/2}) = \frac{1}{2}(1 + \mathrm{e}^{-1/2}).$$

2. Calculons la fonction génératrice de T (par la série de l'exponentielle on a convergence sur $\mathbb{R}$) :

$$g_T(s) = \sum_{n=0}^{\infty} s^n P(T=n) = \frac{1}{2}(1+\mathrm{e}^{-1/2}) + \sum_{n=1}^{\infty} \frac{\mathrm{e}^{-1/2} s^n}{2^{n+1}\, n!} = \frac{1}{2}(1+\mathrm{e}^{-1/2}) + \frac{\mathrm{e}^{-1/2}}{2}\left(\sum_{n=0}^{\infty} \frac{s^n}{2^n\, n!} - 1\right) = \frac{1}{2}(1+\mathrm{e}^{(s-1)/2}$$

3. Soient R et S deux variables indépendantes. On suppose que R suit une loi de Bernoulli $\mathscr{B}(1, \frac{1}{2})$ et S une loi de Poisson $\mathscr{P}(\frac{1}{2})$.
   Calculons la loi de la variable RS.
   Méthode 1 (directe)
   On rappelle que $P(R = k) = 1/2, k = 0, 1,\ P(S = l) = \frac{\mathrm{e}^{-1/2} s^l}{2^l\, l!}$.
   RS prend des valeurs entières et pour $k > 0$:
   $$P(RS = l) = P(R = 1, S = l) = \frac{1}{2}P(S = l) = \frac{\mathrm{e}^{-1/2} s^l}{2^{l+1}\, l!},$$
   donc $P(RS = l) = P(T = l)$ pour $l > 0$, donc aussi pour $l = 0$ seule valeur restante (puisque la somme des probabilités doit être 1). Donc RS et T ont même loi.
   Méthode 2 : (avec la fonction génératrice) On utilise la formule de transfert pour un couple :
   $$E(s^{RS}) = \sum_{k=0}^{1}\sum_{l=0}^{\infty} s^{kl} P(R = k, S = l)$$
   Or par indépendance : $P(R = k, S = l) = P(R = k)P(S = l) = \frac{1}{2}P(S = l)$ donc :
   $$E(s^{RS}) = \frac{1}{2} + \frac{1}{2}\sum_{s=0}^{\infty} s^l P(S = l) = \frac{1}{2} + \frac{1}{2}\mathrm{e}^{(s-1)/2}$$
   EN effet, on a
   $$g_S(s) = \sum_{s=0}^{\infty} s^l P(S = l) = \mathrm{e}^{-1/2}\sum_{n=0}^{\infty}\frac{s^n}{2^n\, n!} = \mathrm{e}^{(s-1)/2} \tag{1}$$
   En conclusion $g_{RS}(s) = g_T(s)$ par la question 1 donc la loi de RS est la même que celle de T :
4. Calculons l'espérance et la variance de T.
   Méthode 1 (directe) :Par le 3 puis indépendance, on a : $E(T) = E(RS) = E(R)E(S)$
   Donc $E(R) = 0P(R = 0 + 1P(R = 1) = \frac{1}{2}$
   et comme en TD,
   $$E(S) = \sum_{n=0}^{\infty}\frac{n}{2^n\, n!} = \mathrm{e}^{-1/2}\sum_{n=1}^{\infty}\frac{1}{2^n\,(n-1)!} = \frac{1}{2}$$
   donc $E(T) = E(RS) = \frac{1}{2}\frac{1}{2} = \frac{1}{4}$.
   De même, par le TD, on sait $Var(S) = \frac{1}{2}$, $E(S^2) = Var(S) + E(S)^2 = \frac{3}{4}$.
   $$E(T^2) = 0P(T = 0) + 1P(T = 1) = \frac{1}{2}.$$
   De même, par indépendance, on obtient :
   $$E(T^2) = E(R^2)E(S^2) = \frac{1}{2}\frac{3}{4}.$$
   $$Var(T) = E((RS)^2) - [E(RS)]^2 = \frac{6}{16} - \frac{1}{16} = \frac{5}{16}$$
   Méthode 2 (en utilisant g_T) :
   $g'_T(s) = \frac{1}{4}\mathrm{e}^{(s-1)/2}$
   et par le cours comme le rayon de convergence de g_T est $R = \infty > 1$, $E(T) = g'_T(1) = \frac{1}{4}$.
   on a $g''_T(s) = \frac{1}{8}\mathrm{e}^{(s-1)/2}$
   De même, par le cours vu $R > 1$, on a
   $$Var(T) = g''_T(1) + g'_T(1) - (g'_T(1))^2 = \frac{1}{8} + \frac{1}{4} - \frac{1}{16} = \frac{5}{16}.$$

5. Calculons la fonction génératrice de $R+S$. Par indépendance, on a : $g_{R+S}(s) = E(s^R s^S) = g_R(s)g_S(s)$ Or on a
$$g_R(s) = s^0 P(R=0) + sP(R=1) = \frac{1}{2}(1+s)$$
et comme vu au 3 à l'équation (1) :
$$g_S(s) = \mathrm{e}^{(s-1)/2}.$$
DOnc on obtient $g_{R+S}(s) = E(s^R s^S) = g_R(s)g_S(s) = \frac{1}{2}(1+s)\mathrm{e}^{(s-1)/2}$
6. Calculons la probabilité (en décomposant selon la partition $\{R=0\} \cup \{R=1\}$,
$$\mathbb{P}(R+S=2) = P(R+S=2, R=0) + P(R+S=2, R=1) = P(S=2, R=0) + P(S=1, R=1).$$
Donc on obtient par indépendance :
$$\mathbb{P}(R+S=2) = P(S=2)P(R=0) + P(S=1)P(R=1) = \frac{1}{2}\Big(\frac{1}{2}e^{-1/2} + \frac{1}{2!2^2}e^{-1/2}\Big) = \frac{5}{16}e^{-1/2}.$$

Exercice 3. (4 points)
Soit $c > 0$. Soit f la fonction de $\mathbb{R}$ dans $\mathbb{R}$ donnée par
$$f(x) = c\frac{1}{\sqrt{x}}\mathbf{1}_{[1,4]}(x).$$
1. Quel c doit-on prendre pour que f soit une densité de probabilité ?
On doit avoir $\int_{-\infty}^{\infty} f(x)dx = 1$
donc
$$1 = c\int_1^4 \frac{1}{\sqrt{x}}dx = c[2\sqrt{x}]_1^4 = c(4-2)$$
et $c = 1/2$
2. On choisit la valeur de c trouvée à la question 1. Soit X une variable aléatoire de densité f. Par la formule de transfert l'espérance est
$$\mathbb{E}(X) = \int_1^4 xf(x)dx = c\int_1^4 \sqrt{x}dx = [\frac{x^{3/2}}{2.3/2}]_1^4 = \frac{2^3-1}{3} = \frac{7}{3}.$$
. Pour la fonction de répartition $F_X(t) = P(X \leq t)$, on distingue 3 cas :
-si $t \leq 1$, $P(X \leq t) = \int_- \infty^t 0dt = 0$
-si $t \in [1,4]$ $P(X \leq t) = c\int_1^t \frac{1}{\sqrt{x}}dx = [\sqrt{t}]_1^4 = \sqrt{t} - 1$
enfin, par croissance pour $t \geq 4$ $F_X(t) \geq F_X(4) = 1$, donc $F_X(t) = 1$ comme c'est une probabilité.
3. Soit U une variable suivant une loi uniforme sur $[1,2]$. Soit $Z = U^2$. Calculons la fonction de répartition $F_Z(t) = P(Z \leq t) = P(U^2 \leq t)$.
Déjà pour $t < 0$ c'est forcément 0 vu qu'un carré est positif, puis : pour $t \geq 0$ vu $U > 0$, $\{U^2 \leq t\} = \{U \leq \sqrt{t}\}$ donc
$$F_Z(t) = P(Z \leq t) = P(U^2 \leq t) = F_U(\sqrt{t})$$
Or de même on a $F_U(t) = 0 si t \leq 1$,
$F_U(t) = \int_0^t dt = (t-1)$ si $t \in [1,2]$ et $F_U(t) = 1$ pour $t \geq 1$ (par croissance).
Donc finalement $F_Z = F_X$.
4. Comme Z et U ont même fonction de répartition donc même loi, la densité de la variable continue Z est f.

Exercice 4. (4 points)

Dans une grande station de montagne, il y a 5 000 skieurs par semaine et ils restent exactement une semaine chacun. On sait qu'en moyenne 1 % des skieurs ont un accident pendant leur semaine de ski, et on va supposer que ce risque est le même pour chaque skieur indépendamment.

1. Soit X le nombre d'accidents survenus à la station pendant 2 semaines de vacances d'hiver.

   En deux semaines 10 000 skieurs visitent la station chacun une semaine. Soit X_i la variable à valeur dans $\{0,1\}$, qui vaut 1 si le i-ème skieur a un accident. Par hypothèse, c'est une variable de Bernoulli $\mathscr{B}(1, 0.01)$ (puisque la moyenne d'une variable de Bernoulli est son paramètre). Le nombre d'accident est $S_{10000} = \sum_{k=1}^{10\,000} X_k$ est la somme de 10 000 variables de Bernoulli indépendantes donc X est de loi binomiale $\mathscr{B}(10000, 0.01)$.

2. Par le théorème de Moivre Laplace $\frac{S_n - np}{\sqrt{np(1-p)}}$ tend vers une loi normale $\mathscr{N}(0,1)$ donc S_n est approché par une loi $\mathscr{N}(np, np(1-p))$.

   Pour $n = 10000; p = 0,01$ donc S_n est presque de loi $Y_n \sim \mathscr{N}(100, 99)$.

   quelle loi normale peut-on approcher X ?

   On utilise l'approximation par loi normale dans la suite de l'exercice.

3. Quelle est la probabilité qu'il y ait moins que 100 accidents au cours des deux semaines ?

   La probabilité $P(S_{10000} \leq 100) \simeq P(Y_{10000} \leq 100) = P(Y_{10000} - 100 \leq 0) = P(Y \leq 0)$ pour $Y = (Y_{10000} - 100)/\sqrt{99} \sim \mathscr{N}(0,1)$

   Or on a

   $$P(Y \leq 0) = \int_{-\infty}^{0} e^{-t^2/2} \frac{dt}{\sqrt{2\pi}} = \frac{1}{2}.$$

4. L'hôpital de la station a remarqué qu'un skieur blessé peut rester hospitalisé jusqu'à 2 semaines. Pour être sûr à 95 % d'accueillir toutes les victimes d'accidents sur les 2 semaines, combien de places l'hôpital de la station doit-il préparer ?

   On cherche donc x avec $P(Y_{10000} \leq x) = 0,95$

   soit $P((Y_{10000} - 100)/\sqrt{99} \leq (x-100)/\sqrt{99}) = P(Y \leq (x-100)/\sqrt{99}) = 0,95$. Comme Y normale standard on doit prendre $(x-100)/\sqrt{99} = z_{0,05}$, $x = 100 + \sqrt{99}.(1,645) \simeq 116,45$ Donc l'hôpital doit prévoir 117 lits.

5. Durant les deux semaines de vacances de l'année 2019, on observe 111 accidents. À partir de cette observation, donnez un intervalle de confiance à 95 % pour la probabilité p qu'un skieur ait eu un accident durant sa semaine de vacances.

   On rappelle la preuve du cours (ce n'était pas demandé), $P(|\frac{S_{10000} - 10000p}{\sqrt{10000p(1-p)}}| \leq z_{0,025}) = 0.95$

   donc $P(|\frac{S_{10000}}{10000} - p| \leq z_{0,025} \frac{\sqrt{p(1-p)}}{100}) = 0,95$

   Donc on obtient l'intervalle de confiance voulu pour la moyenne empirique $\overline{x} = \frac{S_{10000}}{10000}$:$[\overline{x} - z_{0,025}\frac{1}{200}, \overline{x} + z_{0,025}\frac{1}{200}] = [\overline{x}$ est un intervalle de confiance pour p à 95%. La valeur empirique étant $\frac{S_{10000}}{10000} = \frac{111}{10000} = 1,11\%$ et $z_{0,025} = 1.96$ donc

   $[1.11 - 1.96/2\%, 1.11 + 1.96/2\%] = [1.11 - 0,98, 1.11 + 0,98] = [0.13\%, 2.09\%]$ est l'intervalle de confiance cherché.

   On remarque que comme $0,01\%$ est dans l'intervalle, on ne peut pas rejeter l'hypothèse que ce soit la vraie valeur.