Probabilités discrètes et Statistiques descriptives

L2 Université Lyon 1 2018

Examen final du 22 mai 2018.
Durée : 2h
Les documents et calculatrices sont interdits. 2 pages
On prendra soin à JUSTIFIER les réponses aux exercices

- On note toujours $\mathbb{E}(Z)$ l'espérance d'une variable aléatoire Z et $\text{var}(Z)$ 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 $x \notin A$.
- Rappel : une variable aléatoire X suit la loi géométrique de paramètre $p \in]0,1]$ si X est à valeurs dans $\mathbb{N}^*$ et
$$\mathbb{P}(X = k) = p(1-p)^{k-1}, \quad \forall k \in \mathbb{N}^*.$$
- 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 $x \in \mathbb{R}$.

Questions de Cours (10 minutes maxi, 4 points) :

1. Énoncer le théorème de Moivre-Laplace.
2. Donner la définition de la variance empirique non biaisée d'un échantillon statistique.
3. Énoncer le théorème de sommation par paquet pour les familles sommables à termes positifs.
4. Énoncer l'inégalité de Markov.

Exercice 1. (4 points) Soit (X ; Y) un couple de v.a. discrètes à valeur $\{0,1\} \times \{-1,0,1\}$ dont la loi de probabilité est indiquée dans le tableau ci-dessous :

X/Y	-2	-1	0	1	2
0	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{18}$	$\frac{1}{6}$
1	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{9}$	$\frac{1}{9}$	0

1. Donner la loi de X et la loi de Y.
$$P(X=0) = \frac{1}{9} + \frac{1}{9} + \frac{2}{9} + \frac{1}{18} + \frac{1}{6} = 6/9 = 2/3.$$
$P(X = 1) = 1 - 2/3 = 1/3$
$$P(Y = -1) = 1/9 + 1/18 = 2/18 = 1/6 = P(Y = -2) = P(Y = 1) = P(Y = 2)$$
$$P(Y = 0) = 1 - 4/6 = 1/3.$$
2. Les variables X et Y sont elles indépendantes ? Non $P(Y = 2\ etX = 1) = 0 \neq P(Y = 2)P(X = 1) = (1/6).(1/3) = 1/18$.
3. Calculer les espérances de X et Y.
$E(X) = 1/3.1 + 0.2/3 = 1/3$
$$E(Y) = \sum_{i=-2}^{2} i.P(Y = i) = 1/6.(1 + 2 - 1 - 2) + 0.1/3 = 0$$

4. Calculer la variance de Y.
   $E(Y^2) = \sum_{i=-2}^{2} i^2.P(Y = i) = 1/6.(1 + 2^2 + 1 + 2^2) + 0.1/3 = 10/6 = 5/3$
   donc $var(Y) = E(Y^2) - E(Y)^2 = 5/3$.
5. Calculer la covariance $cov(X; Y)$. On calcule d'abord par transfert à 2 variables :

$$E(XY) = \sum_{i,j} i.j.P(Y = i\ etX = j) = -2.1\frac{1}{18} + -1.1\frac{1}{18} + 0\frac{1}{9} + 1.1.\frac{1}{9} + 0 = -\frac{1}{9} - \frac{1}{18} + \frac{1}{9} = -\frac{1}{18}$$

   donc :

$$Cov(X; Y) = E(XY) - E(X)E(Y) = -\frac{1}{18} - 0.$$

Exercice 2. (4 points) Soit $c > 0$. Soit f la fonction de $\mathbb{R}$ dans $\mathbb{R}$ donnée par

$$f(x) = cx^2\mathbf{1}_{[0,1]}(x).$$

1. Quel c doit on prendre pour que f soit une densité de probabilité ? Comme f est positive il suffit d'avoir $\int_{\mathbb{R}} f(x)dx = 1$ Soit $\int_0^2 cx^2 = [cx^3/3]_0^1 = c/3 = 1$ il faut donc $c = 3$.
2. On choisit le c de la question 1. Soit X une variable aléatoire de densité f. Calculer la moyenne $\mathbb{E}(X)$ par transfert :

$$\mathbb{E}(X) = \int_0^1 xf(x)dx = \int_0^1 3x^3dx = 3[x^4/4]_0^1 = 3/4$$

   , ainsi que la fonction de répartition $F_X(t)$ $F_X(t) = \mathbb{P}(X \leq t) = \int_{-\infty}^{x} f(x)dx$
   si $t < 0$ $F_X(t) = 0$ vu l'intégration d'une fonction nulle.
   si $t \in [0, 1]$ $F_X(t) = \int_0^t 3x^2 = t^3$. Enfin par croissance $F_X(t) \geq F_X(1) = 1$ pour t>1 donc $F_X(t) = 1$
3. Soient U, V, W trois variables uniformes sur $[0, 1]$ indépendantes. Soit $Z = \max(U, V, W)$. Calculer la fonction de répartition $F_Z(t) = P(\max(U, V, W) \leq t) = P(U \leq t, V \leq t, W \leq t) = P(U \leq t)P(V \leq t)P(W \leq t)$ par indépendance.
   Or $P(U \leq t) = P(V \leq t) = P(W \leq t) = \max(0, \min(t, 1))$ (même calcul que question précédente) vu la loi uniforme, donc $F_Z(t) = F_X(t)$
4. La densité de la variable continue Z est f vu que X et Z ont même fonction de répartition donc même loi.

Exercice 3. (6 points)

Soit T une variable aléatoire à valeurs dans $\mathbb{N}^*$. Soit $p \in]0, 1[$. On suppose que

$$P(T = n) = p^2\ n(1 - p)^{n-1}.$$

1. Calculer $P(T \geq 2) = 1 - P(T = 1) = 1 - p^2$.

$$P(T\ impair\ et\ T \geq 2) = \sum_{k=1}^{\infty} P(T = 2k+1) = \sum_{k=1}^{\infty} p^2\ (2k+1)(1-p)^{2k} = p^2\frac{d}{dq}\sum_{k=0}^{\infty} q^{2k+1}]_{q=1-p} - p^2$$

   par dérivation d'une série entière sur sont disque de convergence (ici $[-1, 1]$ pour la série géométrique)
   donc

$$P(T\ impair\ et\ T \geq 2) = p^2\frac{d}{dq}\frac{q}{1 - q^2}]_{q=1-p}$$

   Or (décomposition en élément simple) $\frac{q}{1-q^2} = \frac{1}{2(1-q)} - \frac{1}{2(1+q)}$,

$$\frac{d}{dq}\frac{q}{1 - q^2} = \frac{1}{2(1 - q)^2} + \frac{1}{2(1 + q)^2}$$

donc

$$P(T\ impair\ et\ T \geq 2) = p^2(\frac{1}{2p^2} + \frac{1}{2(2-p)^2} - 1)$$

Finalement :

$$P(T\ impair \mid T \geq 2) = \frac{P(T\ impair\ et\ T \geq 2)}{P(T \geq 2)} = \frac{p^2}{2(1-p^2)}(\frac{1}{p^2} + \frac{1}{(2-p)^2} - 2)$$

2. Calculons la fonction génératrice de T est

$$G_T(s) = \mathbb{E}\left(s^T\right) = \sum_{n=1}^{\infty} \mathbb{P}(T = n)s^n,$$

$$G_T(s) = \sum_{n=1}^{\infty} p^2\, n(1-p)^{n-1}s^n = sp^2 \sum_{n=1}^{\infty} n[s(1-p)]^{n-1} = sp^2[\frac{d}{dq}\sum_{n=1}^{\infty} q^n]_{q=s(1-p)} = \frac{sp^2}{(1-s(1-p))^2}.$$

3. Soient R, S deux variables indépendantes de même loi géométrique de paramètre p.
   Calculons la fonction génératrice de $R + S - 1$.
   On a par indépendance $E(s^{R+S-1}) = \frac{E(s^R)E(s^S)}{s}$.
   Et $\mathbb{E}\left(s^R\right) = \sum_{n=1}^{\infty} p(1-p)^{n-1}s^n = \frac{sp}{1-s(1-p)}$ donc $G_{R+S-1}(s) = \frac{sp^2}{(1-s(1-p))^2}$
4. Calculer l'espérance et la variance de T. Comme T a même loi que $R + S - 1$ par l'égalité des fonctions génératrices et R, S i.i.d $E(T) = 2E(R) - 1$, $var(T) = 2var(R)$.
   Or $E(R) = g'_R(1) = [\frac{p}{1-s(1-p)} + \frac{sp(1-p)}{(1-s(1-p))^2}]_{s=1} = 1 + \frac{p(1-p)}{p^2} = \frac{1}{p}$
   donc $E(T) = 2/p - 1$.

$$g''_R(1) = E(X(X-1)) = [\frac{2p(1-p)}{(1-s(1-p))^2} + 2\frac{sp(1-p)^2}{(1-s(1-p))^3}]_{s=1}$$

$$g''_R(1) = 2/p - 2 + 2(1-p)^2/p^2 = 2(1/p^2 - 2/p + 1 + 1/p - 1) = 2(1-p)/p^2$$

   donc $var(R) = g''_R(1) + E(R) - E(R)^2 = 2/p^2 - 2/p + 1/p - 1/p^2 = (1-p)/p^2$ d'où $var(T) = 2(1-p)/p^2$.
5. On utilise l'inégalité de Tchebychev $P(|T - E(T)| \geq \varepsilon) \leq Var(T)/\varepsilon^2$ ce qui se réécrit

$$P(T \notin]E(T) - \varepsilon, E(T) + \varepsilon[) \leq Var(T)/\varepsilon^2 = \alpha$$

   on prend donc $\varepsilon = \sqrt{Var(T)/\alpha}$ donc $a = 2/p - 1 - \sqrt{\frac{2(1-p)}{p^2 0.05}}$, $b = 2/p - 1 + \sqrt{\frac{2(1-p)}{p^2 0.05}}$ conviennent pour avoir

$$P(T \in [a, b]) \geq P(T \in]a, b[) \geq 1 - \alpha = 0,95.$$

Exercice 4. (3 points + Bonus : 3 points)

Une urne contient n boules rouges numérotées de 1 à n. Une deuxième urne contient n boules vertes numérotées de 1 à n. On tire 1 boule au hasard dans chaque urne. On appelle X la somme des valeurs obtenues.

1. Montrons que pour $k \in [\![2, n+1]\!]$

$$\mathbb{P}(X = k) = \frac{k-1}{n^2}.$$

   $\Omega = [\![1, n]\!]^2$ et $\{X = k\} = \cup_{i+j=k}\{\omega_1 = i, \omega_2 = j\}$ donc $Card(\{X = k\}) = k - 1$, d'où le résultat en utilisant la formule pour la probabilité uniforme :

$$\mathbb{P}(X = k) = \frac{Card(\{X = k\})}{Card(\Omega)}.$$

2. Calculer $P(X \leq n) = \sum_{k=2}^{n} \frac{k-1}{n^2} = \sum_{k=1}^{n-1} \frac{k}{n^2} = \frac{n(n-1)}{2n^2} = \frac{n-1}{2n}$.

3. Calculons la loi de X. Si on remplace le premier tirage T_1 par $n+1-T_1$ et de même le second tirage T_2 par $n+1-T_2$ on a même loi donc $2n+2-X$ a même loi que X Donc $P(X = 2n+2-k) = \mathbb{P}(X=k) = \frac{k-1}{n^2}$. soit pour $k \in [\![2n+2-n-1, 2n+2-2]\!] = [\![n+1, 2n]\!]$:

$$\mathbb{P}(X=k) = \frac{2n-k+1}{n^2}.$$

4. Calculons $\mathbb{E}(X)$.
Si T_i est le numéro de la ième urne, chaque T_i est uniforme dans $[\![1,n]\!]$ et les T_i sont indépendants, $X = T_1 + T_2$, $E(X) = 2E(T_1)$ $E(T_1) = \sum_{k=1}^{n} k/n = \frac{n+1}{2}$
donc $E(X) = n+1$.
Autre méthode plus calculatoire : $\mathbb{E}(X) = \sum_{k=2}^{2n} kP(X=k) = \sum_{k=2}^{n} k\frac{k-1}{n^2} + \frac{n(n+1)}{n^2} + \sum_{k=n+2}^{2n} k\frac{2n-k+1}{n^2}$ puis avec le changement de variable l=2n+2-k dans la dernière somme :

$$\mathbb{E}(X) = \sum_{k=2}^{n} k\frac{k-1}{n^2} + \frac{n(n+1)}{n^2} + \sum_{k=2}^{n}(2n+2-l)\frac{l-1}{n^2} = \sum_{k=2}^{n}(2n+2)\frac{k-1}{n^2} + \frac{n(n+1)}{n^2}$$

d'où

$$\mathbb{E}(X) = (2n+2)\frac{(n-1)}{2n} + \frac{(n+1)}{n} = n+1$$

5. **Bonus : 2 points** On considère de nouveau les 2 urnes et une troisième urne avec des boules bleus et on tire 1 boule au hasard dans chaque urne. Soit Z la somme des trois valeurs obtenues. Calculer la loi de Z.
$Z = T_1 + T_2 + T_3$ avec T_i 3 variables uniformes sur [1,n]

$$P(Z=l) = \sum_{k=2}^{l} P(X=k)P(T_3 = l-k) = \sum_{k=2}^{n \wedge l} \frac{k-1}{n^2}\frac{1}{n} + \sum_{k=(n+1)\vee l}^{l \wedge 2n} \frac{2n+1-k}{n^2}\frac{1}{n}$$

si $l \leq n$:

$$P(Z=l) = \sum_{k=1}^{l-1} \frac{k}{n^3} = \frac{l(l-1)}{2n^3}$$

Si $n+1 \leq l \leq 2n$

$$P(Z=l) = \sum_{k=2}^{n} \frac{k-1}{n^3} + \sum_{k=1}^{l-n} \frac{n+1-k}{n^3} = \frac{n-1}{n^2} + \frac{l-n}{n^2} - \sum_{k=1}^{l-n} \frac{k-1}{n^3} = \frac{l-1}{n^2} + \frac{(l-n-1)(l-n)}{2n^3}$$

6. **Bonus : 2 points** On rassemble les 2 urnes en une seule urne et on tire 2 boules simultanément au hasard. Soit Y la somme des valeurs obtenues. L'espace de réalisation est l'ensemble des parties à 2 éléments de l'ensemble des boules $B = \{1R, ..., nR, 1V, ..., nV\}$, soit $\Omega = \mathcal{P}_2(B)$. $Card(\Omega) = \frac{2n(2n-1)}{2}$ Calculons la loi de Y. Il fait distinguer 4 cas. D'abord comme avant pour $k \leq n$ $P(Y=k) = \sum_{(i,j):i+j=k} P(tirage\{i,j\})$ avec Y_i le tirage de chaque boule.
Si $i \neq j$ On a 4 paires de boules i, j donnant le tirage : les 2 rouges les deux verts, iR, jV ou i V et j R. Dans ce cas

$$P(tirage\{i,j\}) = \frac{4}{Card(\Omega)} = \frac{4}{n(2n-1)}.$$

Si i=j, il n'y a qu'une paire possible iR,iV donc :

$$P(tirage\{i,i\}) = \frac{1}{Card(\Omega)} = \frac{1}{n(2n-1)}.$$

Ce cas i,i n'intervient que si k=2i est pair.
Donc si $k = 2i+1 \leq n+1$ impair on a toujours $(k-1)/2$ paires $\{i,j\}$ et donc

$$P(Y=k) = \frac{(k-1)4}{2n(2n-1)} = \frac{(2k-2)}{n(2n-1)}$$

Si $k = 2i \leq n+1$, on a $(k-1)/2$ paires distinctes et 1 paire identique soit :

$$P(Y=k) = \frac{(k-2)4}{2n(2n-1)} + \frac{1}{n(2n-1)} = \frac{(2k-3)}{n(2n-1)}$$

Enfin par symétrie des indices des boules Y a même loi que $2n+2-Y$ donc pour $k \geq n+1$ impair $P(Y=k) = P(Y=2n+2-k) = \frac{(2n-k+1)4}{2n(2n-1)} = \frac{(2k-2)}{n(2n-1)}$ et :

$$P(Y=k) = P(Y=2n+2-k) = \frac{(2n-k+1)2}{n(2n-1)}$$

et pour $k \geq n+1$ pair :

$$P(Y=k) = P(Y=2n+2-k) = \frac{(4n-2k+1)}{n(2n-1)}.$$

L'ensemble de ces valeurs donne la loi de Y.

7. **Bonus : 1 point** On considère de nouveau les 2 urnes de départ et une troisième urne avec des boules bleues et on tire 1 boule au hasard dans chaque urne. Soit Z la somme des trois valeurs obtenues. Calculons la fonction génératrice de Z.

   $Z = X_1 + X_2 + X_3$ est la somme de 3 variables uniformes indépendantes. Donc

$$g_Z(s) = E(s^{X_1+X_2+X_3}) = g_{X_1}(s)g_{X_2}(s)g_{X_3}(s) = g_{X_1}(s)^3$$

   la dernière égalité vient du fait que les trois variables on même loi donc même fonction génératrice :

$$g_{X_1}(s) = \sum_{i=1}^{n} s^i P(X_1 = i) = \frac{1}{n}\sum_{i=1}^{n} s^i = \frac{s - s^{n+1}}{n(1-s)}$$

   donc

$$g_Z(s) = \frac{s^3(1-s^n)^3}{n^3(1-s)^3}$$