Math ECG 2. 2022-2023

Mathématiques Appliquées - F. Gaunard
http://frederic.gaunard.com
ENC Bessières, Paris 17e.

Devoir surveillé n°3 - sujet B

Samedi 3 Décembre
Durée : 4 heures

Les questions précédées de (∗) sont réservées aux khubes.

Exercice 1

(1) Pour tout $k \geq 2$, on a $\dfrac{\ln(k)}{k} \geq 0$. De plus, il est clair que

$$\frac{1}{k} = o\left(\frac{\ln(k)}{k}\right), \qquad k \to +\infty$$

Comme la série $\sum \dfrac{1}{k}$ diverge (série harmonique ou Riemann divergente), le critère de comparaison par négligeabilité pour les séries à termes positifs permet de conclure que la série $\displaystyle\sum_{k\geq 2} \frac{\ln(k)}{k}$ diverge.

(2) On introduit la fonction f définie sur $\mathbb{R}_+^*$ par $f(x) = \dfrac{\ln(x)}{x}$.

(a) Par algèbre des limites, on a $\displaystyle\lim_{x\to 0^+} \frac{\ln(x)}{x} = -\infty$ et f ne se prolonge pas par continuité en 0.

(b) Sur $]0, +\infty[$, f est quotient de fonctions usuelles dérivables (ln et un polynôme) dont le dénominateur ne s'annule pas, elle y est donc elle-même dérivable. De plus, pour tout $x > 0$, on a

$$f'(x) = \frac{(1/x)\cdot x - \ln(x)}{x^2} = \frac{1 - \ln(x)}{x^2}.$$

(c) On obtient sans mal le signe de $f'(x)$, puis les variations de f. Par croissance comparée, $f(x) \to 0,\ x \to +\infty$.

x	0		e		$-\infty$
$f'(x)$	‖	$+$	0	$-$	
f	‖ $-\infty$	↗	$1/e$	↘	0

(3) Dans toute la suite, on note, pour $n \geq 2$, $S_n = \displaystyle\sum_{k=2}^{n} \frac{\ln(k)}{k}$.

(a) Soit $k \geq 3$. Par décroissante de f sur $[3, +\infty[$, on a, pour tout $t \in [k, k+1]$,
$$f(k+1) \leq f(t) \leq f(k).$$
Par positivité de l'intégrale (les bornes étant dans le sens croissant), on a alors
$$f(k+1) = \int_k^{k+1} f(k+1)\mathrm{d}t \leq \int_k^{k+1} f(t)\mathrm{d}t \leq \int_k^{k+1} f(k)\mathrm{d}t = f(k),$$
ce qui est l'encadrement demandé.

(b) En observant que
$$S_n = \sum_{k=2}^{n} f(k) = \frac{\ln(2)}{2} + \sum_{k=3}^{n} f(k) = \frac{\ln(2)}{2} + \frac{\ln(3)}{3} + \sum_{k=3}^{n-1} f(k+1),$$
on somme l'encadrement précédent et avec la relation de Chasles, pour obtenir, d'une part
$$S_n \leq \frac{\ln(2)}{2} + \frac{\ln(3)}{3} + \sum_{k=3}^{n-1} \int_k^{k+1} f(t)\mathrm{d}t = \frac{\ln(2)}{2} + \frac{\ln(3)}{3} + \int_3^n \frac{\ln(t)}{t}\mathrm{d}t$$
et d'autre part
$$S_n \geq \frac{\ln(2)}{2} + \sum_{k=3}^{n} \int_k^{k+1} f(t)\mathrm{d}t = \frac{\ln(2)}{2} + \int_3^{n+1} \frac{\ln(t)}{t}\mathrm{d}t \geq \frac{\ln(2)}{2} + \int_3^n \frac{\ln(t)}{t}\mathrm{d}t,$$
ce qui donne bien l'encadrement attendu.

(c) On commence par calculer l'intégrale ci-dessus
$$\int_3^n \frac{\ln(t)}{t}\mathrm{d}t = \left[\frac{1}{2}(\ln(t)^2\right]_3^n = \frac{(\ln(n))^2 - (\ln(3))^2}{2}.$$
En divisant l'encadrement obtenu à la question précédente par $(\ln(n))^2/2$ et en appliquant le théorème des gendarmes, on trouve alors
$$S_n \sim \frac{(\ln(n))^2}{2}, \quad n \to +\infty.$$

(4) Le but de cette question est de montrer qu'il existe un nombre réel κ tel que
$$S_n = \frac{(\ln(n))^2}{2} + \kappa + o(1), \qquad n \to +\infty.$$

(a) On rappelle que, au voisinage de 0, on a $\ln(1+u) = u - \dfrac{u^2}{2} + o(u^2)$. Il suit que

$$\begin{aligned}(\ln(n))^2 - (\ln(n-1))^2 &= (\ln(n) - \ln(n-1))(\ln(n) + \ln(n-1)) = -\ln\left(1 - \frac{1}{n}\right)\ln\left(n^2\left(1 - \frac{1}{n}\right)\right) \\ &= -2\ln(n)\ln\left(1 - \frac{1}{n}\right) - \left(\ln\left(1 - \frac{1}{n}\right)\right)^2\end{aligned}$$

Or,
$$\ln\left(1 - \frac{1}{n}\right) = -\frac{1}{n} - \frac{1}{2n^2} + o\left(\frac{1}{n^2}\right),$$
donc
$$\left(\ln\left(1 - \frac{1}{n}\right)\right)^2 = \frac{1}{n^2} + o\left(\frac{1}{n^2}\right)$$

Il suit que,

$$\begin{aligned}(\ln(n))^2-(\ln(n-1))^2 &= -2\ln(n)\ln\left(1-\frac{1}{n}\right)-\left(\ln\left(1-\frac{1}{n}\right)\right)^2\\ &= -2\ln(n)\left(-\frac{1}{n}-\frac{1}{2n^2}+o\left(\frac{1}{n^2}\right)\right)-\frac{1}{n^2}\\ &= \frac{2\ln(n)}{n}+\frac{\ln(n)}{n^2}+o\left(\frac{\ln(n)}{n^2}\right),\end{aligned}$$

ce qu'on voulait.

(b) On pose alors $u_n=\dfrac{\ln(n)}{n}-\dfrac{1}{2}\left((\ln(n))^2-(\ln(n-1))^2\right)$.

Par télescopage, on a

$$\begin{aligned}\sum_{k=2}^{n}u_k &= \sum_{k=2}^{n}\frac{\ln(k)}{k}-\frac{1}{2}\sum_{k=2}^{n}\left((\ln(k))^2-(\ln(k-1))^2\right)\\ &= S_n-\frac{1}{2}(\ln(n))^2.\end{aligned}$$

(c) On observe alors aussi, d'après ce qui précède, que

$$u_n=\frac{\ln(n)}{2n^2}+o\left(\frac{\ln(n)}{n^2}\right)=o\left(\frac{1}{n^{3/2}}\right)$$

Or, la série $\sum\dfrac{1}{n^{3/2}}$ converge (Riemann). Par critère de négligeabilité pour les séries à termes positifs, la série $\sum u_n$ converge également, et en notant κ sa somme on a

$$S_n-\frac{1}{2}(\ln(n))^2=\sum_{k=2}^{n}u_k\underset{n\to+\infty}{\longrightarrow}\kappa$$

ou encore

$$S_n=\frac{1}{2}(\ln(n))^2+\kappa+o(1).$$

Exercice 2

(1) La commande `donnees.head()` permet d'afficher les cinq premières lignes du tableau `donnees`. On ne sait pas, avec cet affichage combien de lignes comporte le tableau (il faudrait taper `donnees.shape` pour avoir cette information) mais on sait que tableau contient deux colonnes, la première comprenant les effectifs des populations (en dizaines de milliers d'habitants) et la seconde les profits (en dizaines de milliers d'euros).

(2) (a) Cette figure représente le nuage de points de la série statistique (x, y) où x représente la série statistique des populations des villes (en dizaines de milliers d'habitants) et y celle des profits (en dizaine de milliers d'euros).

(b) On observe un assez bon alignement des points du nuage, une approximation affine du type $y=ax+b$ semble donc raisonnable.

(c) Il faudrait calculer le coefficient de corrélation linéaire du couple statistique pour conforter l'idée précédente. Celui-ci peut se calculer, en utilisant les tableaux X et Y précédents, avec les commandes

```
cov = (X*Y).mean()-(X.mean())*(Y.mean())
rho = cov / (X.std()*Y.std())
print(rho)
```

(On observera qu'ici X et Y sont des tableaux et l'opération $X * Y$ est licite, ce qui ne serait pas le cas si X et Y étaient des listes.)

(d) D'après le cours, le coefficient de corrélation linéaire est une quantité entre -1 et 1. Si sa valeur absolue vaut 1, alors la relation affine est presque sûre. Dans le cas d'une valeur proche de ± 1, on estime que l'approximation affine est pertinente. Ici, on a une valeur proche de 1, ce qui est donc cohérent avec l'observation précédente.

(e) Il commencer par obtenir l'équation de la droite de regression qui permettra de faire la *prévision* demandée. D'après le cours

$$a = \frac{\operatorname{cov}(x,y)}{\hat{\sigma}_x^2}, \qquad b = \overline{y} - a\overline{x},$$

où $\overline{x}$ représenté la moyenne (empirique) de la série statistique x, $\overline{y}$ celle de y et $\hat{\sigma}_x^2$ la variance empirique. On applique ensuite la formule à 18.2354 qui est le nombre de dizaine de milliers d'habitants de *Legumeville*. Cela donne les instructions, en reprendant la valeur `cov` calculée ci-avant,

```
a=cov / (X.std()**2)
b=Y.mean() - a*X.mean()

print(a*18.2354 + b)
```

(3) La covariance est linéaire à gauche et à droite. Donc si on pose $Y' = 1.03Y$ on obtient

$$\operatorname{cov}(X, Y') = \operatorname{cov}(X, 1.03Y) = 1.03 \times \operatorname{cov}(X, Y).$$

Pour le coefficient de corrélation linéaire, on a

$$\rho_{X,Y'} = \frac{\operatorname{cov}(X,Y')}{\sqrt{V(X)V(Y')}} = \frac{(1.03)\operatorname{cov}(X,Y)}{\sqrt{V(X)(1.03)^2V(Y)}} = \frac{\operatorname{cov}(X,Y)}{\sqrt{V(X)2V(Y)}} = \rho_{X,Y}$$

et ce dernier est donc inchangé.

Exercice 3

On considère les fonctions e_1, e_2, e_3 et e_4 définies sur $\mathbb{R}_+^*$ par

$$e_1(x) = x, \quad e_2(x) = x^2, \quad e_3(x) = x\ln(x) \quad \text{et} \quad e_4(x) = x^2\ln(x)$$

et on note E l'espace vectoriel engendré par e_1, e_2, e_3 et e_4

$$E = \operatorname{Vect}(e_1, e_2, e_3, e_4).$$

Ainsi, E est un espace vectoriel dont les vecteurs sont des fonctions définies sur $\mathbb{R}_+^*$, son vecteur nul 0_E est la fonction identiquement nulle sur $\mathbb{R}_+^*$.

(1) Le but de cette question est de montrer que la famille (e_1, e_2, e_3, e_4) est libre. Pour cela, on considère quatre réels a, b, c et d tels que

$$ae_1 + be_2 + ce_3 + de_4 = 0_E,$$

c'est à dire que, **pour tout** $x \in \mathbb{R}$,

$$ax + bx^2 + cx\ln(x) + dx^2\ln(x) = 0 \qquad (\star)$$

(a) En choisissant $x = 1$ dans la relation $(\star)$, on fait disparaître les log et il reste $a + b = 0$.

(b) Soit $x > 1$. En divisant dans la relation $(\star)$ par $x^2\ln(x)$ (qui est une quantité strictement positive), on a bien

$$\frac{a}{x\ln(x)} + \frac{b}{\ln(x)} + \frac{c}{x} + d = 0.$$

En faisant tendre $x \to +\infty$, un résultat de croissance comparée permet de voir que

$$d \underset{x\to+\infty}{\longleftarrow} \frac{a}{x\ln(x)} + \frac{b}{\ln(x)} + \frac{c}{x} + d = 0.$$

(c) La relation $(\star)$ est donc devenue

$$\forall x \in \mathbb{R}, \qquad ax + bx^2 + cx\ln(x)$$

en la divisant, pour $x > 1$, par x^2 (de sorte à isoler b), on obtient,

$$\forall x > 1, \qquad \frac{a}{x} + b + c\frac{\ln(x)}{x} = 0.$$

En passant à la limite lorsque $x \to +\infty$ et par croissance comparée

$$b \underset{x\to+\infty}{\longleftarrow} \frac{a}{x} + b + c\frac{\ln(x)}{x} = 0.$$

(d) Comme $a = -b = 0$ et $d = 0$. Il reste

$$\forall x \in \mathbb{R}, \qquad cx\ln(x) = 0.$$

En prenant par exemple $x = e$, on obient $c = 0$. Ainsi, on a bien montré que $a = b = c = d = 0$ et la famille (e_1, e_2, e_3, e_4) est libre.

(2) La famille (e_1, e_2, e_3, e_4) étant génératrice de E (par définition de E) et libre, elle en forme une base. On peut alors affirmer que

$$\dim(E) = 4.$$

(3) On note u l'application qui à toute fonction f de E associe la fonction $g = u(f)$ définie par

$$\forall x \in \mathbb{R}_+^*, \qquad g(x) = u(f)(x) = xf'(x).$$

(a) Soient f, g deux fonctions de E et λ, μ deux réels. Par linéarité de l'opération de dérivation, on a

$$\begin{aligned} u(\lambda f + \mu g)(x) &= x(\lambda f + \mu g)'(x) = x(\lambda f'(x) + \mu g'(x)) \\ &= \lambda x f'(x) + \mu x g'(x) \\ &= \lambda u(f)(x) + \mu u(g)(x) \end{aligned}$$

et u est bien une application linéaire.

(b) On calcule:

$$\begin{aligned} u(e_1)(x) &= xe_1'(x) = x \\ u(e_1) &= e_1 \end{aligned}$$

$$\begin{aligned} u(e_2)(x) &= xe_2'(x) = x(2x) = 2x^2 \\ u(e_2) &= 2e_2 \end{aligned}$$

$$\begin{aligned} u(e_3)(x) &= xe_3'(x) = x(\ln(x) + 1) = x + x\ln(x) \\ u(e_3) &= e_1 + e_3 \end{aligned}$$

$$\begin{aligned} u(e_4)(x) &= xe_4'(x) = x(2x\ln(x) + x) = x^2 + 2x^2\ln(x) \\ u(e_4) &= 2e_4 + e_2 \end{aligned}$$

(c) Les images par u des quatre vecteurs e_1, e_2, e_3, e_4 sont encore des éléments de E, ainsi, comme u est linéaire, il en sera de même pour l'image de toute combinaison de vecteurs de e_1, e_2, e_3 et e_4. Comme ces quatre vecteurs forment une base de E, on peut alors affirmer que l'image par u de tout vecteur de E sera encore un élément de E et u est bien un endomorphisme.

(4) (a) D'après les calculs précédents et la définition de la matrice de u dans la base (e_1, e_2, e_3, e_4), on peut écrire

$$A = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}.$$

(b) La matrice A est triangulaire supérieure sans zéro sur la diagonale; elle est donc inversible et l'endomorphisme qu'elle représente, u bijectif (donc un automorphisme).

(c) On résout. D'une part,

$$\begin{aligned} X = \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} \in \operatorname{Ker}(A - I) &\iff (A-I)X = 0 \\ &\iff \left\{ \begin{array}{rcl} z &=& 0 \\ y + t &=& 0 \\ t &=& 0 \end{array} \right. \\ &\iff X = \begin{pmatrix} x \\ 0 \\ 0 \\ 0 \end{pmatrix} = x \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \end{aligned}$$

Ainsi,

$$\operatorname{Ker}(A - I) = \operatorname{Vect}\left(\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right).$$

D'autre part,

$$\begin{aligned} X = \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} \in \operatorname{Ker}(A - 2I) &\iff (A-2I)X = 0 \\ &\iff \left\{ \begin{array}{rcl} -x + z &=& 0 \\ t &=& 0 \\ -z &=& 0 \end{array} \right. \\ &\iff X = \begin{pmatrix} 0 \\ y \\ 0 \\ 0 \end{pmatrix} = y \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \end{aligned}$$

Ainsi,

$$\operatorname{Ker}(A - I) = \operatorname{Vect}\left(\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \right).$$

Dans les deux cas, on a bien

$$\dim(\operatorname{Ker}(A - I)) = \dim(\operatorname{Ker}(A - 2I)) = 1.$$

☞ **Remarque**. En reprenant cet exercice après avoir travaillé sur le Chapitre 9, on pourra conclure que, la somme des dimensions des sous-espaces propres n'étant pas égale à la dimension de tout l'espace, l'endomorphisme n'est pas diagonalisable. C'est l'unique intérêt de cette question.

(5) On remarque en effet que

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = D + N,$$

avec D diagonale et $N^2 = 0$. Comme, de plus D et N commutent, ce qu'on vérifie

$$DN = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = ND,$$

on peut utiliser la formule du binôme pour le calcul des puissances de $A = D + N$. Comme de plus, par une récurrence immédiate, on a $N^k = 0$, pour $k \geq 2$, on a

$$\begin{aligned} A^n &= (D+N)^n = \sum_{k=0}^{n} \binom{n}{k} N^k D^{n-k} = D^n + nND^{n-1} \\ &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2^n & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2^n \end{pmatrix} + n \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2^{n-1} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2^{n-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2^n & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2^n \end{pmatrix} + n \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2^{n-1} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

et on a bien

$$A^n = \begin{pmatrix} 1 & 0 & n & 0 \\ 0 & 2^n & 0 & n2^{n-1} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2^n \end{pmatrix}.$$

Problème

Une urne contient des boules blanches (en proportion p), des boules noires (en proportion q) et des boules rouges (en proportion r). On a donc $p + q + r = 1$ (on suppose que $p, q, r \in]0; 1[$). On effectue des tirages successifs **avec remise** dans cette urne. Pour $i \in \mathbb{N}^*$, on note T_i la variable qui vaut 1 si la $i-$ème boule tirée est blanche, -1 si elle est noire et 0 si elle est rouge.
Les tirages étant effectués avec remise, les variables (T_i) sont donc mutuellement indépendantes.

Partie A

On note ensuite X_1 la variable aléatoire égale au numéro du tirage qui amène pour la première fois une boule blanche et X_2 celui correspondant au tirage où sort pour la première fois la **deuxième** boule blanche.

Par exemple, si les premiers tirages donnent *Noire, Rouge, Noire, Blanche, Rouge, Blanche, ...*, on a $X_1 = 4$ et $X_2 = 6$.

(1) Par définition, $T_i(\Omega) = \{-1, 0, 1\}$. Plus précisément, on peut écrire la loi de T_i sous forme de tableau

a	-1	0	1
$P(T_i = a)$	q	r	p

Le tableau permet de calculer sans mal espérance et variance.

$$\begin{aligned} E(T_i) &= (-1)q + 0 \cdot r + p = p - q \\ \\ E(T_i^2) &= (-1)^2 q + 0^2 r + p \\ &= p + q \\ V(T_i) &= E(T_i^2) - E(T_i)^2 \\ &= p + q - (p-q)^2 \end{aligned}$$

(2) Identifiant l'obtention d'une boule blanche à un *succès*, on reconnait pour X_1 le temps d'attente du premier succès lors de la répétition d'épreuves indépendantes de Bernoulli. On reconnait donc une loi géométrique de paramètre p

$$X_1 \hookrightarrow \mathcal{G}(p).$$

D'après le cours, on peut immédiatement affirmer que

$$E(X_1) = \frac{1}{p}, \qquad V(X_1) = \frac{(1-p)}{p^2}.$$

(3) Les variables X_1 et X_2 qu'on cherche à simuler ne dépendent que de la proportion de boules blanches p.

```
import numpy as np
import numpy.random as rd

def simul_X1_X2(p):
    x1=1 # on initialise avec le premier lancer
    while rd.rand() > p # tant qu'on a pas de blanche
        x1=x1+1 # on relance
    x2=x1+1 # au moins un lancer de plus pour la deuxième blanche
    while rd.rand() > p
        x2=x2+1
    return [x1, x2]
```

(4) (a) Pour le calcul, on utilise la modélisation suivante, $[T_k = 1]$ représente l'obtention d'une boule blanche au $k-$ième lancer. Ainsi, pour $i \in \mathbb{N}^*$ et $j \geq i+1$, et par indépendance des lancers (donc des variables (T_k)),

$$\begin{aligned} P(X_1 = i \cap X_2 = j) &= P\left(\left(\bigcap_{k=1}^{i-1} [T_k \neq 1]\right) \cap [T_i = 1] \cap \left(\bigcap_{k=i+1}^{j-1} [T_k \neq 1]\right) \cap [T_j = 1]\right) \\ &= \left(\prod_{k=1}^{i-1} P(T_k \neq 1)\right) \times P(T_i = 1) \times \left(\prod_{k=i+1}^{j-1} P(T_k \neq 1)\right) \times P(T_j = 1) \\ &= (1-p)^{i-1} \times p \times (1-p)^{j-1-i} \times p \\ &= (1-p)^{j-2} p^2 \end{aligned}$$

Si $j \leq i$, alors $P(X_1 = i \cap X_2 = j) = 0$ (la deuxième boule blanche arrive nécessairement *après* la première). Au final, on peut écrire

$$P(X_1 = i \cap X_2 = j) = \begin{cases} (1-p)^{j-2} p^2, & \text{si } j \geq i+1 \\ 0, & \text{si } j \leq i \end{cases}$$

(b) Par la formule des probabilités totales appliquée au s.c.e $\{(X_1 = i) : \ i \in \mathbb{N}^*\}$, on a, pour $j \in \mathbb{N}, j \geq 2$,

$$\begin{aligned} P(X_2 = j) &= \sum_{i=1}^{+\infty} P(X_1 = i \cap X_2 = j) \\ &= \sum_{i=1}^{j-2} P(X_1 = i \cap X_2 = j) \qquad (\text{car } P(X_1 = i \cap X_2 = j) = 0 \text{ si } i \geq j) \\ &= \sum_{i=1}^{j-2} (1-p)^{j-1} p^2 \\ &= (j-1)(1-p)^{j-2} p^2. \end{aligned}$$

(c) D'après le cours

$$\begin{aligned} X_2 \text{ admet une espérance} &\iff \sum jP(X_2 = j) \text{ converge absolument} \\ &\iff \sum j(j-1)(1-p)^{j-2}p^2 \text{ converge} \end{aligned}$$

Or,

$$j(j-1)(1-p)^{j-1}p^2 = p^2 j(j-1)(1-p)^{j-2}$$

et on reconnait le multiple du terme général de la série géométrique (de raison $1-p$) dérivée deux fois et convergente. Donc X_2 admet une espérance et

$$E(X_2) = p^2 \sum_{j=2}^{+\infty} j(j-1)(1-p)^{j-2} = p^2 \times \frac{2}{(1-(1-p))^3} = \frac{2}{p}.$$

(5) On note $U_2 = X_2 - X_1$.

(a) Soit $j \in \mathbb{N}^*$. C'est encore la formule des probabilités totales appliquée au même s.c.e que précédemment.

$$\begin{aligned} P(U_2 = j) &= \sum_{i=1}^{+\infty} P(U_2 = j \cap X_1 = i) \\ &= \sum_{i=1}^{+\infty} P(X_2 = j + i \cap X_1 = i) \\ &= \sum_{i=1}^{+\infty} (1-p)^{i+j-2} p^2 \\ &= (1-p)^{j-1} p^2 \sum_{i=1}^{+\infty} (1-p)^{i-1} \\ &= (1-p)^{j-1} p^2 \times \frac{1}{1-(1-p)} \\ &= (1-p)^{j-1} p \\ &= P(X_1 = j) \end{aligned}$$

Ainsi U_2 et X_1 suivent toutes deux une loi géométrique de paramètre p. On en déduit que U_2 admet espérance et variance et que

$$E(U_2) = \frac{1}{p}, \qquad V(U_2) = \frac{(1-p)}{p^2}.$$

(b) On vérifie la condition définissant l'indépendance. Soient $i, j \in \mathbb{N}^* times \mathbb{N}^*$

$$\begin{aligned} P(X_1 = i \cap U_2 = j) &= P(X_1 = i \cap X_2 - X_1 = j) = P(X_1 = i \cap X_2 = i + j) \\ &= (1-p)^{i+j-2} p^2 \\ &= (1-p)^{i-1} p \times (1-p)^{j-1} p \\ &= P(X_1 = i) P(U_2 = j) \end{aligned}$$

et les variables X_1 et U_2 sont bien indépendantes.

(c) Le programme simule 10000 fois les variables X_1 et X_2 puis calcule la *covariance empirique* de X_1 et U_2. Les deux variables en question étant indépendantes, cette covariance est nulle. Le programme devrait afficher 0.

(d) Il est clair que $X_2 = X_1 + U_2$. Comme X_1 et U_2 admettent toutes deux une variance et qu'elles sont indépendantes, on peut alors écrire que

$$V(X_2) = V(X_1) + V(U_2) = 2V(X_1) = \frac{2(1-p)}{p^2}.$$

(e) On développe $V(U_2)$.

$$\begin{aligned} V(U_2) &= V(X_2 - X_1) \\ &= V(X_2) + V(X_1) - 2\text{cov}(X_1, X_2) \end{aligned}$$

Ainsi,

$$\begin{aligned} \text{cov}(X_1, X_2) &= \frac{1}{2}\left(V(X_2) + V(X_1) - V(U_2)\right) \\ &= \frac{1-p}{p^2} \end{aligned}$$

Cette covariance étant non nulle, on en conclut que les variables X_1 et X_2 ne sont pas indépendantes.

Partie B

On note W la variable correspondant au nombre de boules rouges obtenues avant l'obtention de la première boule blanche.

(6) Soit $i \in \mathbb{N}^*$. Sachant $[X_1 = i]$, la première boule blanche arrive au $n-$ième tirage. Il y a donc $i - 1$ tirages avec une boule rouge ou noire qui précèdent. Pour chacun de ces $n - 1$ tirages, on obtient une rouge ou une blanche. On compte le nombre de rouges obtenues. Sachant qu'on obtient pas de blanche, la probabilité d'obtenir une rouge à chaque tentative est

$$\frac{r}{q+r}.$$

On peut donc affirmer que, sachant $[X_1 = i]$, $\quad W \hookrightarrow \mathcal{B}\left(i-1, \frac{r}{q+r}\right)$, ou encore

$$P_{[X_1=i]}(W = k) = \begin{cases} \binom{i-1}{k} \left(\frac{r}{q+r}\right)^k \left(\frac{q}{q+r}\right)^{i-1-k}, & \text{si } 0 \leq k \leq i-1 \\ 0, & \text{si } k \geq i. \end{cases}$$

(7) Par la formule des probabilités totales, appliquée au s.c.e $\{(X_1 = i) : \ i \in \mathbb{N}^*\}$,

$$\begin{aligned}
P(W=k) &= \sum_{i=1}^{+\infty} P(X_1=i)P_{[X_1=i]}(W=k) \\
&= \sum_{i=k+1}^{+\infty} P(X_1=i)P_{[X_1=i]}(W=k) \\
&= \sum_{i=k+1}^{+\infty} \binom{i-1}{k}\left(\frac{r}{q+r}\right)^k\left(\frac{q}{q+r}\right)^{i-1-k}(1-p)^{i-1}p \\
&= p\left(\frac{r}{q+r}\right)^k \sum_{i=k+1}^{+\infty}\binom{i-1}{k}\left(\frac{q}{q+r}\right)^{i-1-k}(1-p)^{i-1}
\end{aligned}$$

(8) C'est un calcul facile que l'on a fait plusieurs fois.

$$k\binom{i}{k} = k \times \frac{i!}{(i-k)!k!} = \frac{i(i-1)}{(k-1)!(i-1-(k-1))!} = i\binom{i-1}{k-1}.$$

(9) On admet qu'il est licite de permuter les sommes

$$\sum_{k=0}^{+\infty}\sum_{i=k+1}^{+\infty}\cdots = \sum_{i=1}^{+\infty}\sum_{k=0}^{i-1}\cdots\ ,$$

et que W admet une espérance. Il suit que

$$\begin{aligned}
E(W) &= \sum_{k=0}^{+\infty} kP(W=k) = \sum_{k=1}^{+\infty} kP(W=k) \\
&= \sum_{k=1}^{+\infty}\sum_{i=k+1}^{+\infty}\binom{i-1}{k}kp\left(\frac{r}{q+r}\right)^k\left(\frac{q}{q+r}\right)^{i-1-k}(1-p)^{i-1} \\
&= \sum_{i=1}^{+\infty}\sum_{k=0}^{i-1}\binom{i-1}{k}kp\left(\frac{r}{q+r}\right)^k\left(\frac{q}{q+r}\right)^{i-1-k}(1-p)^{i-1} \\
&= p\sum_{i=1}^{+\infty}(1-p)^{i-1}\sum_{k=1}^{i-1}k\binom{i-1}{k}\left(\frac{r}{q+r}\right)^k\left(\frac{q}{q+r}\right)^{i-1-k} \\
&= p\sum_{i=1}^{+\infty}(1-p)^{i-1}(i-1)\sum_{k=1}^{i-1}\binom{i-2}{k-1}\left(\frac{r}{q+r}\right)^k\left(\frac{q}{q+r}\right)^{i-1-k} \\
&= p\left(\frac{r}{q+r}\right)\sum_{i=1}^{+\infty}(i-1)(1-p)^{i-1}\sum_{\ell=0}^{i-2}\binom{i-2}{\ell}\left(\frac{r}{q+r}\right)^\ell\left(\frac{q}{q+r}\right)^{i-2-\ell} \\
&= p\left(\frac{r}{q+r}\right)\sum_{i=1}^{+\infty}(i-1)(1-p)^{i-1} \qquad \text{(par la formule du binôme)} \\
&= p(1-p)\left(\frac{r}{q+r}\right)\sum_{s=1}^{+\infty}s(1-p)^{s-1} \\
&= p(1-p)\left(\frac{r}{q+r}\right)\times\frac{1}{1-(1-p))^2} \\
&= \frac{(1-p)r}{(q+r)p} = \frac{r}{p}.
\end{aligned}$$

Partie C

On note, pour $n \in \mathbb{N}^*$, $S_n = T_1 + ... + T_n$.

(10) En observant que $S_1 = T_1$, la toute première question de la Partie A permet de voir que $E(S_1) = p - q$ et $V(S_1) = p + q - (p-q)^2$.

(11) Par linéarité de l'espérance, on a

$$E(S_n) = \sum_{i=1}^{n} E(T_i) = n(p-q).$$

Par **indépendance** des variables T_i, on a

$$V(S_n) = \sum_{i=1}^{n} V(T_i) = n\left(p + q - (p-q)^2\right).$$

(12) Soit $t > 0$. On pose $V_n = t^{S_n}$.

(a) Par définition $V_1 = t^{T_1}$. On peut donc écrire le tableau de la loi de T_1:

a	$1/t$	1	t
$P(V_1 = a)$	q	r	p

Il suit que

$$E(V_1) = \frac{1}{t} \times q + r + t \times p = \frac{q + rt + t^2}{t}.$$

(b) On observe que

$$\begin{aligned} E(V_n) &= E(t^{S_n}) = E\left(\prod_{i=1}^{n} t^{T_i}\right) \\ &= \prod_{i=1}^{n} E\left(t^{T_i}\right) \end{aligned}$$

par indépendance des variables t^{T_i} par le **lemme des coalitions**. Comme les variables t^{T_i} ont toutes la même espérance, il suit que

$$E(V_n) = E(V_1)^n = \left(\frac{q + rt + t^2}{t}\right)^n.$$