MATHÉMATIQUES E
(ÉPREUVE N° 289)
ANNÉE 2018
ÉPREUVE CONÇUE PAR HEC PARIS
VOIE ÉCONOMIQUE ET COMMERCIALE

1 – Le sujet

Le sujet de cette année était composé comme à l'accoutumée d'un exercice et d'un problème.

L'exercice, d'une longueur inhabituelle, portait sur l'algèbre linéaire (rang d'une matrice, polynômes annulateurs, etc.), les cardinaux de couples d'entiers ayant certaines propriétés et des questions de Scilab liées à ces cardinaux.

Le problème qui comportait trois parties, avait pour objet l'étude de sommes de variables aléatoires suivant une loi de Bernoulli de même paramètre mais non nécessairement indépendantes.

La partie I était consacrée à l'étude des valeurs possibles du coefficient de corrélation linéaire de deux variables aléatoires dans divers schémas de Bernoulli. Elle permettait de déterminer notamment une formule reliant la variance d'une somme de variables aléatoires, le paramètre commun aux deux variables aléatoires et leur coefficient de corrélation linéaire.

D'autres questions avaient pour objet de préciser un minorant du coefficient de corrélation.

La partie II, très progressive, étudiait les lois de probabilité bêta-binomiales et faisait la part belle aux convergences d'intégrales, à l'intégration par parties et à la formule du binôme.

Enfin, la partie III se concentrait sur le cas particulier où le modèle ne comporte que deux variables aléatoires en faisant appel à la loi d'un couple ainsi qu'à ses lois marginales et conditionnelles. Une dernière question proposait de compléter un script Scilab qui effectue une simulation de deux variables aléatoires.

2 – Barème

L'exercice et le problème comptaient respectivement pour 40% et 60% des points de barème. Plus précisément, la partie I du problème représentait 24% des points de barème tandis que la pondération des parties II et III était identique (18% de points de barème).
Le poids des questions de Scilab représentait 12% des points de barème et les questions les plus cotées étaient :

- dans l'exercice, les questions 2.c) et 3.b) ;
- dans le problème, les questions 1.a), 3.a), 8.a), 8.b) et 9.b).

3 – Remarques de correction

<u>Exercice.</u>

1.a) Trop de candidats ne savent pas que les valeurs propres d'une matrice triangulaire se lisent sur la diagonale principale.

On lit très souvent que la non inversibilité de M entraîne la non diagonalisabilité de M.

b) Pour le calcul du rang, il y a trop d'affirmations non justifiées.

c) Rares sont les candidats qui ont prouvé que les seuls polynômes annulateurs de degré 3 sont de la forme aX 3 ($a \neq 0$). On trouve également qu'il y a trois polynômes annulateurs : X 3, X 2.X et X.X.X !!!

2.a) Cette question a été bien traitée en général.

b) et c) Les réponses à cette question sont rarement correctes.

3.a) Question très peu abordée et très rarement réussie.

b) Les questions i) et ii) étaient tout à fait abordables mais nombre de candidats n'ont pas vu que les xk étaient des entiers positifs et ont trouvé des xk fractionnaires ou négatifs !!!

Les questions suivantes ont été très peu abordées et traitées ; en particulier, les définitions de P(k) et de Q(l,k) n'ont pas été comprises.

Problème.

1.a) Beaucoup de candidats ne connaissent pas la variance d'une variable de Bernoulli.

Trop peu de candidats ont trouvé la loi dans le cas où les variables sont égales.

De même, certains pensent que le calcul de la variance permet de trouver la loi : ainsi, puisque V ($\sum$Yk) = n2 p(1-p), la loi de la somme est binomiale de paramètres n2 et p.

b) Question rarement abordée et encore plus rarement bien résolue.

2. Les candidats qui raisonnent correctement sont peu nombreux ; on trouve souvent l'erreur suivante : P([X1=0] $\cap$ [X2=0]) = 1- P([X1=1] $\cap$ [X2=1]).

3. Question très peu abordée.

4. La majorité des candidats ne savent pas étudier la convergence d'une intégrale et quand ils pensent à prendre un équivalent, ils oublient de dire que la fonction est continue et positive sur]0,1/2].

Il y a encore trop de candidats qui écrivent que l'intégrale d'un produit est égal au produit des intégrales !!!

5.a) Plus de 90% des candidats effectuent directement l'intégration par parties au lieu de la faire sur un intervalle [a,b] et faire tendre ensuite a vers 0 et b vers 1.

Les questions suivantes furent très rarement abordées et correctement traitées. De plus, la notation (z)[m] s'est souvent transformée en zm et bien que les résultats soient exacts, les raisonnements qui ont permis de les trouver sont incorrects.

4 – Conseils aux futurs candidats

Pour ce qui concerne la forme, le jury conseille aux futurs candidats de lire attentivement le texte préliminaire qui précède toute épreuve écrite de mathématiques, dans lequel il est précisé notamment, que la lisibilité et la qualité de la rédaction entrent pour une part non négligeable dans l'appréciation des copies : un correcteur ne s'attarde pas à essayer de « décrypter » une copie illisible. Par contre, une copie propre et claire ne peut qu'avantager son auteur. Le jury rappelle également que les abréviations dans les copies doivent être proscrites et il conseille de bien numéroter les questions et d'encadrer les résultats.

De plus, les raisonnements doivent être clairs et précis, les affirmations étant étayées par une argumentation solide. Par exemple, le recours trop fréquent à des phrases du type « il est clair que… » doit être évité au profit d'une justification correcte fondée sur un apprentissage rigoureux et une très bonne maîtrise du cours.

Le jury recommande aux futurs candidats de prendre le temps de lire l'ensemble du sujet, non seulement pour s'en imprégner, mais aussi pour pointer les questions qui paraissent faciles à résoudre, lesquelles ne se situent pas nécessairement dans la première partie du sujet.

La recherche d'une solution à une question ne doit pas dépasser quatre à cinq minutes. Au-delà de ce délai, en cas d'échec, le candidat doit admettre le résultat de cette question (si la réponse figure dans l'énoncé), passer à la question suivante sans éprouver un sentiment de déstabilisation ou de découragement. Autrement dit, le jury recommande aux futurs candidats de faire preuve d'une grande ténacité.

5 – Statistiques

Sur les 2101 candidats ayant composé dans cette épreuve, la note moyenne est de 9,37 avec un écart-type de 4,75, ces statistiques étant très voisines de celles du concours 2017.

Environ 8%, soit 172 candidats, obtiennent une note supérieure à 16 et 20 candidats se voient attribuer la note maximale de 20. La note médiane est de 9,7 et les premier et troisième quartiles sont égaux à 5,5 et 12,7 respectivement.

Pour obtenir la note de 20, il fallait obtenir au moins 55% des points du barème.

Mathématiques HEC E 2018
Éléments de corrigé

EXERCICE

1. Soit $M = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$.

a) La matrice M étant triangulaire, son spectre est constitué de ses coefficients diagonaux :

$$\boxed{\mathrm{Sp}(M) = \{0\}}\,.$$

Si la matrice M était diagonalisable, la dimension du sous-espace propre associé à son unique valeur propre serait égale à 4, ce qui est impossible puisque M n'est pas la matrice nulle.
Par conséquent, M n'est pas diagonalisable.

b) Le rang de M est égal à 2 et celui de M^2 à 1, puisque $M^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$.

c) Comme la matrice M^3 est nulle, X^3 et plus généralement tous les polynômes $a\,X^3$ $(a \neq 0)$ sont des polynômes annulateurs de M dont le degré est égal à 3.

Pour montrer qu'il n'en existe pas d'autre, on considère un polynôme P de degré 3 annulateur de M.
Ce polynôme doit être de la forme $aX^3 + bX^2 + cX$ $(a \neq 0)$, puisque 0 en est une racine.
La matrice

$$aM^3 + bM^2 + cM = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & c & b \\ 0 & 0 & 0 & c \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

étant nulle, les deux réels b et c sont nécessairement nuls.

Finalement, les polynômes annulateurs de M dont le degré est 3 sont les polynômes de la forme $a\,X^3$ avec $a \neq 0$.

2.
$$\begin{cases} F_j = \mathrm{Im}(f^j) \\ r_j = \dim(F_j) \end{cases} \tag{1}$$

a) • $r_0 = \dim(F_0) = \dim\big(\mathrm{Im}(f^0)\big) = \dim\big(\mathrm{Im}(\mathrm{Id}_{\mathbb{R}^n})\big) = \dim(\mathbb{R}^n) = \boxed{n}\,.$

• $r_n = \dim(F_n) = \dim\big(\mathrm{Im}(f^n)\big) = \dim\big(\mathrm{Im}(0_{\mathcal{L}(\mathbb{R}^n)})\big) = \dim(\{0_{\mathbb{R}^n}\}) = \boxed{0}\,.$

b) Soit $j \in [\![0, n-1]\!]$.

i) L'image de g_j est l'image de F_j par f :

$$\mathrm{Im}(g_j) = f(F_j) = f\big(f^j(\mathbb{R}^n)\big) = f^{j+1}(\mathbb{R}^n) = F_{j+1}\ .$$

Il en résulte que le rang de g_j est égal à $\boxed{r_{j+1}}$.

ii) En appliquant le théorème du rang à l'application linéaire g_j, on obtient

$$r_j = \dim(F_j) = \dim\big(\mathrm{Ker}(g_j)\big) + \dim\big(\mathrm{Im}(g_j)\big) = \dim\big(\mathrm{Ker}(f) \cap F_j\big) + r_{j+1}$$

d'après le résultat précédent et parce que

$$\mathrm{Ker}(g_j) = \{x \in F_j\ ;\ f(x) = 0_{\mathbb{R}^n}\} = \mathrm{Ker}(f) \cap F_j$$

d'où l'égalité :

$$\boxed{r_j - r_{j+1} = \dim\big(\mathrm{Ker}(f) \cap F_j\big)}\ .$$

c) Des inclusions $F_n \subset F_{n-1} \subset \cdots \subset F_1 \subset F_0$, on déduit les inclusions

$$\mathrm{Ker}(f) \cap F_n \subset \mathrm{Ker}(f) \cap F_{n-1} \subset \ldots \mathrm{Ker}(f) \cap F_1 \subset \mathrm{Ker}(f) \cap F_0$$

et, par passage aux dimensions, les inégalités

$$0 \leq r_{n-1} - r_n \leq \cdots \leq r_1 - r_2 \leq r_0 - r_1$$

d'où les inégalités

$$\boxed{n \geq r_0 - r_1 \geq r_1 - r_2 \geq \cdots \geq r_{n-1} - r_n \geq 0}$$

puisque $r_0 - r_1 \leq r_0 = n$.

3.

$$\forall i \in [\![1, n]\!],\ \ x_i = \mathrm{Card}\big(\{j \in [\![0, n-1]\!]\ ;\ r_j - r_{j+1} = i\}\big) \qquad (2)$$

a) $\displaystyle\sum_{i=1}^{n} i\, x_i = \sum_{j=0}^{n-1} (r_j - r_{j+1}) = r_0 - r_n = n$, ce qui prouve que $(x_1, x_2, \ldots, x_n)$ est un élément de $P(n)$.

b) i) Comme $M^0 = I_4$ et $M^3 = M^4 = 0_{\mathcal{M}_4(\mathbb{R})}$ de rangs respectifs 4 et 0, les rangs successifs de f^0, f^1, f^2, f^3, f_4 sont, grâce à 1.c,

$$r_0 = 4, r_1 = 2, r_2 = 1, r_3 = 0, r_4 = 0$$

d'où $r_0 - r_1 = 2, r_1 - r_2 = r_2 - r_3 = 1, r_3 - r_4 = 0$ et par conséquent :

$$\boxed{(x_1, x_2, x_3, x_4) = (2, 1, 0, 0)}\ .$$

ii) Les seules manières de décomposer l'entier 4 en somme de multiples de 1, 2, 3 et 4 sont :

$$4 = 0 + 0 + 0 + 4 = 1 + 0 + 3 + 0 = 0 + 2 \times 2 + 0 + 0 = 2 \times 1 + 2 + 0 + 0 = 4 \times 1 + 0 + 0 + 0\ .$$

Par conséquent : $\boxed{P(4) = \{(0, 0, 0, 1), (1, 0, 1, 0), (0, 2, 0, 0), (2, 1, 0, 0), (4, 0, 0, 0)\}}$ et $p(4) = 5$.

iii)

- Pour $f = 0_{\mathcal{L}(\mathbb{R}^4)}$, $r_0 = 4, r_1 = r_2 = r_3 = r_4 = 0$ et $(x_1, x_2, x_3, x_4) = (0, 0, 0, 1)$.

On fournit des endomorphismes f associés aux quatre autres éléments de $P(4)$ par leurs matrices N dans la base canonique de $\mathbb{R}^4$:

- Pour $N = M^2$, la matrice N^2 est nulle, d'où $r_0 = 4, r_1 = 1, r_2 = r_3 = r_4 = 0$ et $(x_1, x_2, x_3, x_4) = (1, 0, 1, 0)$.

- Pour $N = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$, $r_0 = 4, r_1 = 2, r_2 = r_3 = r_4 = 0$ et $(x_1, x_2, x_3, x_4) = (0, 2, 0, 0)$

.

- Pour $N = M$, $(x_1, x_2, x_3, x_4) = (2, 1, 0, 0)$ (voir a).

- Pour $N = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$, $r_0 = 4, r_1 = 3, r_2 = 2, r_3 = 1, = r_4 = 0$

et $(x_1, x_2, x_3, x_4) = (4, 0, 0, 0)$.

Pour les cinq éléments (x_1, x_2, x_3, x_4) de $P(4)$, on a donc bien trouvé un endomorphisme f de $\mathbb{R}^4$ tel que les relations (2) sont satisfaites pour les rangs définis par (1).

4. a) i) Les seuls k-uplets $(x_1, x_2, \dots, x_k)$ de nombres entiers naturels tels que

$$x_1 + x_2 + \cdots + x_k \leq 1$$

sont $(0, 0, \dots, 0)$ et les k vecteurs de la base canonique de $\mathbb{R}^k$.
Parmi ces k-uplets, le seul qui appartient à $P(k)$ est le dernier vecteur de la base canonique de $\mathbb{R}^k$.
Donc :

$$Q(1, k) = \{(x_1, x_2, \dots, x_k) \in \mathbb{N}^k \, ; \, x_1 + 2\,x_2 + \cdots + k\,x_k = k \text{ et } x_1 + x_2 + \cdots + x_k \leq 1\} = \boxed{\{(0, 0, \dots, 0, 1)\}}.$$

ii) Soit ℓ un nombre entier supérieur ou égal à k.
Pour établir l'égalité $Q(\ell, k) = P(k)$, on procède par double inclusion.

- L'inclusion $Q(\ell, k) \subset P(k)$ provient directement de la définition de $Q(\ell, k)$ comme un sous-ensemble de $P(k)$.

- Pour établir l'inclusion $P(k) \subset Q(\ell, k)$, il suffit d'observer que, si $(x_1, x_2, \dots, x_k)$ est un élément de $P(k)$, alors

$$x_1 + x_2 + \cdots + x_k \leq x_1 + 2\,x_2 + \cdots + k\,x_k = k \leq \ell$$

ce qui prouve que $(x_1, x_2, \dots, x_k)$ est un élément de $Q(\ell, k)$.

b) On suppose : $k > \ell \geq 2$.
Pour qu'un élément $(x_1, x_2, \dots, x_k)$ de $P(k)$ vérifie $x_1 + x_2 + \cdots + x_k = \ell$, il faut et il suffit que

$$x_2 + 2\,x_3 + \cdots + (k-1)x_k = k - \ell$$

c'est-à-dire que $x_{k-\ell+2} = \cdots = x_k = 0$, $x_2 + 2\,x_3 + \cdots + (k-\ell)x_{k-\ell+1} = k - \ell$ et $x_1 + x_2 + \cdots + x_{k-\ell+1} = \ell$ ou encore que

$$\begin{cases} x_{k-\ell+2} = \cdots = x_k = 0 \\ x_2 + 2\,x_3 + \cdots + (k-\ell)x_{k-\ell+1} = k - \ell \\ x_2 + \cdots + x_{k-\ell+1} \leq \ell \end{cases}$$

Il en résulte que le nombre d'éléments $(x_1, x_2, \dots, x_k)$ de $P(k)$ tels que $x_1 + x_2 + \cdots + x_k = \ell$ est égal au nombre de $(k-\ell)$-uplets $(x_2, x_3, \dots, x_{k-\ell+1})$ de nombres entiers naturels qui vérifient :

$$\begin{cases} x_2 + 2\,x_3 + \cdots + (k-\ell)x_{k-\ell+1} = k - \ell \\ x_2 + \cdots + x_{k-\ell+1} \leq \ell \end{cases}$$

c'est-à-dire $q(\ell, k - \ell)$.

On a donc démontré l'égalité :

$$\boxed{q(\ell, k-\ell) = \mathrm{Card}\big(\{(x_1, x_2, \ldots, x_k) \in P(k) \,;\, x_1 + x_2 + \cdots + x_k = \ell\}\big)}\,.$$

c) Soit un entier $\ell \geq 2$.

i) L'ensemble $Q(\ell, k)$ étant la réunion des deux ensembles disjoints $Q(\ell-1, k)$ et $\{(x_1, x_2, \ldots, x_k) \in P(k) \,;\, x_1 + x_2 + \cdots + x_k = \ell\}$, son cardinal $q(\ell, k)$ est la somme des cardinaux de ces deux ensembles.

On en déduit l'égalité demandée à l'aide du résultat démontré en b :

$$\boxed{q(\ell, k) = q(\ell-1, k) + q(\ell, k-\ell)}\,.$$

ii) Comme le seul ℓ-uplet $(x_1, x_2, \ldots, x_\ell)$ de $\mathbb{N}^\ell$ qui vérifie

$$x_1 + x_2 + \cdots + x_\ell = x_1 + 2x_2 + \cdots + \ell x_\ell = \ell$$

est $(\ell, 0, \ldots, 0)$, on a :

$$\boxed{q(\ell, \ell) - q(\ell-1, \ell) =} \mathrm{Card}\big(\{(x_1, x_2, \ldots, x_\ell) \in \mathbb{N}^\ell \,;\, x_1+x_2+\cdots+x_\ell = x_1+2x_2+\cdots+\ell x_\ell = \ell\} = \boxed{1}$$

5. a)

```
[5] if (K < L)  then q(L,K)=q(L-1,K);
[6] else   if (K==L) then q(L,K)=q(L-1,K)+1;
```

b) Les coefficients $p(1), p(2), \ldots, p(n)$ sont les coefficients diagonaux de la matrice « qmatrix(n) ».

```
--> n=input("valeur de n?")

valeur de n? 4  //exemple: n=4
 n  =
    4.

--> q=qmatrix(n);
--> q(n,n) //dernier coefficient diagonal de la matrice q
 ans  =
    5. // valeur de p(n) pour n=4
```

c) La lecture de la deuxième ligne de la matrice qmatrix(9) fournie par l'énoncé suggère la formule

$$q(2, k) = 1 + \lfloor \frac{k}{2} \rfloor$$

mais il ne s'agit encore que d'une conjecture, dont la démonstration suit.

Soit $k \in \mathbb{N}^*$.

Si $(e_1, e_2, \ldots, e_k)$ désigne la base canonique de $\mathbb{R}^k$, on a :

- $\{(x_1, x_2, \ldots, x_k) \in P(k) \,;\, x_1 + x_2 + \cdots + x_\ell = 1\} = \{e_k\}$
- $\{(x_1, x_2, \ldots, x_k) \in P(k) \,;\, x_1+x_2+\cdots+x_\ell = 2\} = \begin{cases} \{e_1 + e_{k-1}, \ldots, e_i + e_{i+1}\} & \text{si } k = 2i+1 \\ \{e_1 + e_{k-1}, \ldots, e_{i-1} + e_{i+1}, 2\, e_i\} & \text{si } k = 2i \end{cases}$.

Il en résulte que

$$q(2, k) = \mathrm{Card}\big(Q(2, k)\big) = \begin{cases} i+1 & \text{si } k = 2i+1 \\ i+1 & \text{si } k = 2i \end{cases}$$

c'est-à-dire $\boxed{q(2, k) = 1 + \lfloor \frac{k}{2} \rfloor}$.

PROBLEME

Partie I.

Valeurs possibles du coefficient de corrélation linéaire dans divers schémas de Bernoulli

1. a) i) Si les variables aléatoires X_k sont mutuellement indépendantes, alors leurs covariances sont nulles et leur somme suit la loi binomiale $\mathcal{B}(n,p)$.

On a donc $r = 0$ et $V(\sum_{k=1}^{n} X_k) = np\,(1-p)$.

ii) Si les X_k sont égales, alors leurs coefficients de corrélation linéaire sont tous égaux à 1 et leur somme est égale à n fois leur valeur commune.

On a donc $r = 1$ et $V(\sum_{k=1}^{n} X_k) = V(n\,X_1) = n^2 p\,(1-p)$.

b)

On peut procéder par récurrence (finie) sur k.

• Initialisation

L'égalité est vraie pour $k = 1$ puisque : $V(X_1) = p\,(1-p)$.

• Hérédité

On suppose l'égalité vraie pour un entier $k \in [\![1, n-1]\!]$.

Dès lors :

$$V\big(\sum_{i=1}^{k+1} X_i\big) = V\big(\sum_{i=1}^{k} X_i\big) + V(X_{k+1}) + 2\,\mathrm{Cov}\big(\sum_{i=1}^{k} X_i, X_{k+1}\big)$$

$$= kp\,(1-p)\big(1+(k-1)r\big) + p\,(1-p) + 2\sum_{i=1}^{k} \mathrm{Cov}(X_i, X_{k+1})$$

$$= kp\,(1-p)\big(1+(k-1)r\big) + p\,(1-p) + 2krp\,(1-p) = (k+1)p\,(1-p)(1+kr)$$

ce qui achève la preuve, pour tout $k \in \mathbb{N}^*$, de l'égalité :

$$\boxed{V\big(\sum_{i=1}^{k} X_i\big) = kp\,(1-p)\big(1+(k-1)r\big)}\ .$$

c) Comme $V\big(\sum_{i=1}^{n} X_i\big) = np\,(1-p)\big(1+(n-1)r\big) \geq 0$ et $np\,(1-p) > 0$, on a bien :

$$\boxed{r \geq -\frac{1}{n-1}}\ .$$

2. a) La variable aléatoire X_1X_2 vaut 1 si X_1 et X_2 valent 1, et 0 sinon. Elle suit donc la loi de Bernoulli de paramètre $P([X_1 = 1] \cap [X_2 = 1])$.

Par conséquent :

$$\mathrm{Cov}(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2) = P([X_1 = 1] \cap [X_2 = 1]) - p^2 \cdot$$

Comme $r = \dfrac{\mathrm{Cov}(X_1, X_2)}{p\,(1-p)}$, on en déduit les équivalences :

$r=-1 \Longleftrightarrow P([X_1=1]\cap[X_2=1])-p^2=-p\,(1-p) \Longleftrightarrow P([X_1=1]\cap[X_2=1])=p\,(2p-1)\cdot$

b) Comme la probabilité $P([X_1=1])$ est égale à p, on a :

$$P([X_1=1]\cap[X_2=0])=P([X_1=1])-P([X_1=1]\cap[X_2=1])=p-p\,(2p-1)=2p\,(1-p)$$

d'où

$$P([X_1=0]\cap[X_2=0])=P([X_2=0])-P([X_1=1]\cap[X_2=0])=(1-p)-2p\,(1-p)=\boxed{(1-p)(1-2p)}\cdot$$

c) Comme les deux probabilités $P([X_1=1]\cap[X_2=1])$ et $P([X_1=0]\cap[X_2=0])$ sont nécessairement positives ou nulles, on déduit de a) et b) que r ne peut être égal à -1 que lorsque $p\,(2p-1)\geq 0$ et $(1-p)(1-2p)\geq 0$, c'est-à-dire lorsque $\boxed{p=\dfrac{1}{2}}\cdot$

Dans ce cas, on a :

$$P([X_1=1-X_2])=1-P([X_1=0]\cap[X_2=0])-P([X_1=1]\cap[X_2=1])=\boxed{1}\,.$$

Remarque
Le coefficient de corrélation linéaire r ne peut être égal à -1 que lorsque n est égal à 2, c'est pourquoi on a fait cette hypothèse dans cette question.
Le seul cas où $r=-1$ est celui où $n=2$ et où la somme X_1+X_2 est égale 1, presque sûrement.

3. a) Comme $E(X)=np$, on a nécessairement : $\boxed{p=\dfrac{1}{n}}\cdot$

Comme la variance de $\displaystyle\sum_{k=1}^{n}X_k$, variable aléatoire presque certaine, est nulle, on a d'après la formule démontrée en 1.b :

$$\boxed{r=-\frac{1}{n-1}}\cdot$$

b) Comme la somme des variables aléatoires de Bernoulli $X_1,X_2,\ldots,X_n$ ne prend que la valeur 1, une seule de ces variables peut prendre la valeur 1, les autres étant nulles.
Les valeurs possibles de $(X_1,X_2,\ldots,X_n)$ sont donc les vecteurs $e_1,e_2,\ldots,e_n$ de la base canonique de $\mathbb{R}^n$.
Pour tout $i\in[\![1,n]\!]$, on a :

$$\boxed{P([(X_1,X_2,\ldots,X_n)=e_i])=P([X_i=1])=p=\frac{1}{n}}\cdot$$

Partie II. Lois beta-binomiales

4. a) L'intégrande $t\longmapsto t^{x-1}(1-t)^{y-1}$ est une fonction continue sur $]0,1/2]$, positive, et équivalente quand t tend vers 0_+ à $1/t^{1-x}$.

L'intégrale de référence $\displaystyle\int_0^{1/2}\frac{1}{t^{1-x}}\,\mathrm{d}t$ est convergente si, et seulement si, l'exposant $1-x$ est strictement inférieur à 1.

Il en résulte, par comparaison, que l'intégrale $\displaystyle\int_0^{1/2}t^{x-1}(1-t)^{y-1}\,\mathrm{d}t$ est convergente sous la même condition, c'est-à-dire si, et seulement si, $x>0$.

b) Le changement de variable affine $u=1-t$ convient, puisque : $\begin{cases}\text{si } t=1/2,\ u=1/2\\ \text{si } t=1-\varepsilon,\ u=\varepsilon\\ \mathrm{d}t=-\mathrm{d}u\end{cases}$

d'où :

$$\int_{1/2}^{1-\varepsilon} t^{\,x-1}(1-t)^{\,y-1}\ \mathrm{d}t = \int_{1/2}^{\varepsilon} (1-u)^{\,x-1}u^{\,y-1}\ (-1)\mathrm{d}u = \int_{\varepsilon}^{1/2} u^{\,y-1}(1-u)^{\,x-1}\ \mathrm{d}u\ .$$

c) L'intégrale impropre $\displaystyle\int_0^1 t^{\,x-1}(1-t)^{\,y-1}\ \mathrm{d}t$ est convergente si, et seulement si, les deux intégrales $\displaystyle\int_0^{1/2} t^{\,x-1}(1-t)^{\,y-1}\ \mathrm{d}t$ et $\displaystyle\int_{1/2^1} t^{\,x-1}(1-t)^{\,y-1}\ \mathrm{d}t$ sont convergentes, ce qui est vrai si, et seulement si, x et y sont strictement positifs, d'après les résultats trouvés en a et b.

5. a) On procède par intégration par parties (sur une intégrale ordinaire) et passage à la limite.

$$\int_{\varepsilon}^{1-\varepsilon} t^{\,x}(1-t)^{\,y-1}\ \mathrm{d}t = \big[-\frac{t^x(1-t)^y}{y}\big]_{\varepsilon}^{1-\varepsilon} + \frac{x}{y}\int_{\varepsilon}^{1-\varepsilon} t^{\,x-1}(1-t)^{\,y-1}\ \mathrm{d}t$$

tend vers $\dfrac{x}{y}\,B(x,y)$ quand ε tend vers 0.

On a donc bien :

$$\boxed{B(x+1,y) = \frac{x}{y}\,B(x,y+1)}\ .$$

b)

$$B(x,y+1) = \int_0^1 t^{\,x-1}(1-t)(1-t)^{\,y-1}\ \mathrm{d}t = B(x,y) - B(x+1,y) = B(x,y) - \frac{x}{y}\,B(x,y+1)$$

d'où : $\boxed{B(x,y+1) = \dfrac{y}{x+y}\,B(x,y)}\ .$

6. Soit $(x,y) \in\ ]0,+\infty[^2$.
Une preuve « de proche en proche » de l'égalité peut être acceptable, mais une vraie récurrence est préférable.
Il s'agit d'une récurrence double : on démontre par récurrence sur ℓ la propriété

$$(\mathcal{P}_\ell)\quad \forall\, k \in [\![0,\ell]\!],\quad B(x+k,y+\ell-k) = \frac{(x)^{[k]}\,(y)^{[\ell-k]}}{(x+y)^{[\ell]}}\,B(x,y)$$

qui, pour un entier ℓ fixé, peut être démontrée par récurrence (finie) sur k.

- Initialisation : pour $\ell = 0$, $(\mathcal{P}_\ell)$ est vraie puisque la seule valeur de k à considérer est 0, pour laquelle l'égalité est évidente $B(x,y) = B(x,y)$.
- Hérédité : on suppose $(\mathcal{P}_\ell)$ vraie pour un entier $\ell \in \mathbb{N}$.

On démontre alors $(\mathcal{P}_{\ell+1})$ par récurrence.

- Initialisation : $k = 0$.

$$B(x,y+\ell+1) = \frac{y+\ell}{x+y+\ell}\,B(x,y) = \frac{y+\ell}{x+y+\ell}\ \frac{(x)^{[0]}\,(y)^{[\ell]}}{(x+y)^{[\ell]}}\,B(x,y) = \frac{(x)^{[0]}\,(y)^{[\ell+1]}}{(x+y)^{[\ell+1]}}\,B(x,y)\ .$$

- Hérédité : on suppose l'égalité vraie pour un entier $k < \ell+1$.

$$B(x+k+1,y+\ell+1-k-1) = \frac{x+k}{x+y+\ell-k}\,B(x+k,y+\ell+1-k)$$

$$= \frac{x+k}{x+y+\ell-k}\ \frac{(x)^{[k]}\,(y)^{[\ell+1-k]}}{(x+y)^{[\ell+1]}}\,B(x,y) = \frac{(x)^{[k+1]}\,(y)^{[\ell+1-(k+1)]}}{(x+y)^{[\ell+1]}}\,B(x,y)\ .$$

On a prouvé, pour tout couple (k,ℓ) de nombres entiers tels que $0 \le k \le \ell$, l'égalité :

$$\boxed{B(x+k,y+\ell-k) = \frac{(x)^{[k]}\,(y)^{[\ell-k]}}{(x+y)^{[\ell]}}\,B(x,y)}\ .$$

7. a)

$$\sum_{k=0}^{n} p_k = \sum_{k=0}^{n} \binom{n}{k} \frac{B(a+k, b+n-k)}{B(a,b)} = \frac{1}{B(a,b)} \sum_{k=0}^{n} \binom{n}{k} \int_0^1 t^{\,a+k-1}(1-t)^{\,b+n-k-1}\, \mathrm{d}t$$

$$= \frac{1}{B(a,b)} \int_0^1 \Big(\sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k} \Big) t^{\,a-1}(1-t)^{\,b-1}\, \mathrm{d}t = \frac{1}{B(a,b)} \int_0^1 t^{\,a-1}(1-t)^{\,b-1}\, \mathrm{d}t = 1\,.$$

On a bien : $\boxed{\sum_{k=0}^{n} p_k = 1}$.

b) Pour $a = b = 1$, on a :

$$\forall\, k \in [\![0, n]\!],\ P([S = k]) = \binom{n}{k} \frac{(1)^{[k]}\,(1)^{[n-k]}}{(2)^{[n]}} = \binom{n}{k} \frac{k!\,(n-k)!}{(n+1)!} = \frac{1}{n+1}$$

et $\mathcal{B}(n; 1, 1)$ est donc la loi uniforme sur $[\![0, n]\!]$.

c)

$$E(S) = \sum_{k=1}^{n} k \binom{n}{k} \frac{(a)^{[k]}\,(b)^{[n-k]}}{(a+b)^{[n]}} = \sum_{k=1}^{n} n \binom{n-1}{k-1} \frac{a}{a+b} \ \frac{(a+1)^{[k-1]}\,(b)^{[n-1-(k-1)]}}{(a+b+1)^{[n-1]}} = \frac{na}{a+b}$$

puisque $sum_{k=1}^{n} \binom{n-1}{k-1} \dfrac{(a+1)^{[k-1]}\,(b)^{[n-1-(k-1)]}}{(a+b+1)^{[n-1]}} = \displaystyle\sum_{\ell=0}^{n-1} \binom{n-1}{\ell} \dfrac{(a+1)^{[\ell]}\,(b)^{[n-1-(\ell)]}}{(a+b+1)^{[n-1]}} = 1$

(en tant que somme des probabilités ponctuelles associées à la loi bêta-binomiale $\mathcal{B}(n-1; a+1, b)$).

Partie III. Étude du cas où $n = 2$

8. a) Les probabilités associées aux quatre valeurs possibles du couple (X_1, X_2) sont données par :

$$\begin{cases} P([X_1 = 0] \cap [X_2 = 0]) = \dfrac{B(a, b+2)}{B(a,b)} = \dfrac{b(b+1)}{(a+b)(a+b+1)} \\[2ex] P([X_1 = 0] \cap [X_2 = 1]) = \dfrac{B(a+1, b+1)}{B(a,b)} = \dfrac{ab}{(a+b)(a+b+1)} \\[2ex] P([X_1 = 1] \cap [X_2 = 0]) = \dfrac{B(a+1, b+1)}{B(a,b)} = \dfrac{ab}{(a+b)(a+b+1)} \\[2ex] P([X_1 = 1] \cap [X_2 = 1]) = \dfrac{B(a+2, b)}{B(a,b)} = \dfrac{a(a+1)}{(a+b)(a+b+1)} \end{cases}$$

- $P([X_1 = 1]) = P([X_1 = 1] \cap [X_2 = 0]) + P([X_1 = 1] \cap [X_2 = 1]) = \dfrac{a}{a+b}$.

La variable aléatoire X_1 suit la loi de Bernoulli de paramètre $\dfrac{a}{a+b}$.

- $P([X_2 = 1]) = P([X_1 = 0] \cap [X_2 = 1]) + P([X_1 = 1] \cap [X_2 = 1]) = \dfrac{a}{a+b}$.

La variable aléatoire X_2 suit la loi de Bernoulli de paramètre $\dfrac{a}{a+b}$.

b)
$$\begin{cases} P([X_1+X_2=0]) = \dfrac{b(b+1)}{(a+b)(a+b+1)} = \dbinom{2}{0}\dfrac{(a)^{[0]}\,(b)^{[2-0]}}{(a+b)^{[2]}} \\[2ex] P([X_1+X_2=1]) = \dfrac{2ab}{(a+b)(a+b+1)} = \dbinom{2}{1}\dfrac{(a)^{[1]}\,(b)^{[2-1]}}{(a+b)^{[2]}} \\[2ex] P([X_1+X_2=2]) = \dfrac{a(a+1)}{(a+b)(a+b+1)} = \dbinom{2}{2}\dfrac{(a)^{[2]}\,(b)^{[2-2]}}{(a+b)^{[2]}} \end{cases}$$

La variable aléatoire $X_1 + X_2$ suit donc la loi bêta-binomiale $\mathcal{B}(2; a, b)$.

c) $P_{[X_1=1]}([X_2=1]) = \dfrac{P([X_1=1]\cap[X_2=1])}{P([X_1=1])} = \boxed{\dfrac{a+1}{a+b+1}}$.

9. a) La ligne [3] affecte à la variable u une simulation de la loi uniforme sur $[0, a+b]$. De même la ligne [4] affecte à la variable v une simulation de la loi uniforme sur $[0, a+b+1]$.

b) Si deux variables aléatoires indépendantes U et V suivent respectivement la loi uniforme sur $[0, a+b]$ et la loi uniforme sur $[0, a+b+1]$, alors :

$$P([U < a]) = \frac{a}{a+b}\ .$$

et

$$\begin{cases} P_{[X_1=1]}([X_2=1]) = \dfrac{a+1}{a+b+1} = P([V < a+1]) \quad \text{(voir 8.c)} \\[2ex] P_{[X_1=0]}([X_2=1]) = \dfrac{P([X_1=0]\cap[X_2=1])}{P([X_1=0])} = \dfrac{a}{a+b+1} = P([V < a]) \end{cases}$$

On peut donc compléter les lignes [5] et [6] comme suit :

```
[5] if (u < a)  then x(1,1)= 1 ; // P([X1=1])=P([U<a])
    if (v < a+1) then x(1,2)=1;  // P([X2=1]/[X1=1])=P([V<a+1])
    end;
[6] else if (v < a) then x(1,2)=1;  // P([X2=1]/[X1=0])=P([V<a])
    end;
```

10. a)

$$\operatorname{Cov}(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2) = \frac{a(a+1)}{(a+b)(a+b+1)} - \Big(\frac{a}{a+b}\Big)^2 = \frac{ab}{(a+b)^2(a+b+1)}$$

$$\sigma(X_1)\,\sigma(X_2) = V(X_1) = \frac{ab}{(a+b)^2}$$

d'où :

$$\boxed{\rho(X_1, X_2) = \frac{1}{a+b+1}}\ .$$

b)

$$\begin{cases} \dfrac{a}{a+b} = p \\[2ex] \dfrac{1}{a+b+1} = r \end{cases} \iff \begin{cases} a = \dfrac{p\,(1-r)}{r} \\[2ex] b = \dfrac{(1-p)(1-r)}{r} \end{cases}$$

On peut donc simuler deux variables aléatoires de Bernoulli de paramètre p et de coefficient de correlation linéaire r à l'aide de l'instruction

```
randbetabin(p*(1-r)/r,(1-p)*(1-r)/r)
```