MATHEMATIQUES
Var discrètes et algèbre linéaire

Ce sujet se compose d'un problème, et d'un « complément » qu'on prendra soin de lire ***en entier****. On trouvera dans le complément des questions de cours qu'il est bon de savoir rédiger et qui sont apparues dans de nombreux sujets.*

Problème :

Un bus de nuit, sur son trajet pour rentrer au dépôt où il s'arrêtera définitivement, passe devant $n+1$ « arrêts » ou « stations » possibles, notées $S_0, S_1, \cdots, S_n$ (la station S_0 correspondant au dépôt de bus). On précise que les indices sont choisis de telle manière que les stations sont numérotées selon les distances croissantes, de la plus proche à la plus éloignée, du dépôt (à savoir $i < j \Rightarrow \text{dist}(S_0, S_i) < \text{dist}(S_0, S_j)$).

On suppose qu'à l'instant 0 le bus est à son arrêt le plus éloigné de S_0, à savoir S_n.
Pour tout entier naturel k et pour tout $r \in [\![0, n]\!]$, **si le bus est à l'instant k à la station S_r, il sera à l'instant $k+1$ arrêté à l'une des stations $S_0, S_1, \cdots, S_r$, avec équiprobabilité pour chacune de ces possibilités**. On suppose en effet que, lorsque sur une station il n'y a personne pour lui signifier l'arrêt, il poursuit sa route sans s'arrêter et que, s'il est arrêté et que quelqu'un arrive en courant, il peut lui arriver de patienter (c'est la nuit, il a du temps...).

Pour tout $k \in \mathbb{N}$, **on note Z_k la variable aléatoire égale au numéro de la station (pris dans $[\![0, n]\!]$) où se trouve le bus à l'instant k.**

Remarques de notations : Pour tout entier naturel n, I_n désigne l'ensemble des entiers naturels k tels que $0 \leq k \leq n$ et I_n^* l'ensemble des entiers naturels k tels que $1 \leq k \leq n$.
Pour une variable aléatoire X à valeurs dans $\mathbb{N}$, on pose, pour tout réel t pour lequel cela a un sens,

$$g_X(t) = \mathbb{E}(t^X) = \sum_{k=0}^{+\infty} \mathbb{P}(X = k) t^k$$

où $\mathbb{E}$ désigne l'espérance. La fonction g_X est appelée fonction génératrice de la variable aléatoire X.
Dans tout ce problème, on considérera que $0^0 = 1$, ce qui permet par exemple d'affirmer ici que $g_X(0) = \mathbb{P}(X = 0)$.
L'objet de la partie préliminaire est l'étude de quelques propriétés des fonctions génératrices afin de les exploiter par la suite dans notre modélisation.

Préliminaire : Fonction génératrice d'une variable à valeurs dans I_n

Soient n un entier naturel et X une variable aléatoire réelle à valeurs dans I_n. Pour tout $k \in I_n$, on note $a_k = \mathbb{P}(X = k)$ la probabilité que X prenne la valeur k.

① a) Montrer que g_X est une fonction polynôme à coefficients réels dont on précisera le degré maximal. Quelle est la valeur de $g_X(1)$?

b) On souhaite montrer que si g_X est donnée, alors la loi de X est entièrement connue :

i. Pour tout $k \in I_n$ fixé, montrer que pour tout entier naturel i, la dérivée i-ième du monôme X^k vaut :

$$(X^k)^{(i)} = \begin{cases} \dfrac{k!}{(k-i)!} X^{k-i} & \text{si } i \leq k \\ 0 & \text{sinon} \end{cases}$$

ii. Exprimer $g_X^{(i)}(t)$ sous forme d'une somme qui dépend des a_k, t et i. En déduire la valeur de $g_X^{(i)}(0)$.

iii. Conclure que $\forall k \in I_n$, $\mathbb{P}(X = k) = \dfrac{g_X^{(k)}(0)}{k!}$.

c) Justifier la dérivabilité première et seconde de g_X sur $\mathbb{R}$ et montrer que $g'_X(1) = \mathbb{E}(X)$, $g''_X(1) = \mathbb{E}(X(X-1))$. Que vaut $\mathbb{V}(X)$ en fonction de g_X ?

② Soient (m_1, m_2) deux entiers naturels et (Z_1, Z_2) deux variables aléatoires réelles à valeurs respectivement dans I_{m_1} et I_{m_2}. On suppose que Z_1 et Z_2 sont indépendantes.
En utilisant pour tout réel t l'expression $g_X(t) = \mathbb{E}(t^X)$, montrer que

$$g_{Z_1+Z_2}(t) = g_{Z_1}(t) \cdot g_{Z_2}(t)$$

③ **Application :**

a) Calculer la fonction génératrice d'une variable aléatoire X suivant la loi binomiale de paramètres n et p avec $p > 0$. Retrouver grâce à elle la valeur de son espérance.

b) Soit Y une variable aléatoire réelle suivant la loi binomiale de paramètres n' et p avec n' un entier naturel. On suppose que X et Y sont indépendantes. Montrer que $X + Y$ suit la loi binomiale de paramètres $n + n'$ et p.

Partie I : Simulation numérique

*On suppose avoir importé la fonction **`randint`** du module **`random`*** (on rappelle que pour tout couple d'entiers naturels (a, b) tels que $a \leq b$, `randint(a,b)` renvoie un entier aléatoire compris entre les entiers a et b, bornes incluses, y-compris lorsque $a = b$).

✐ Toute autre fonction est interdite dans cette partie, et cela inclus notamment `np.sum` et `np.mean`

① Écrire une fonction Python `simulZ(n,k)` d'argument l'entier n égale au nombre de stations distinctes du dépôt et un entier naturel k et qui renvoie une valeur simulée de la variable aléatoire Z_k.

② Écrire une fonction `freqZ(n,k,m)` d'argument l'entier n associé au nombre de stations, un entier naturel k et un entier naturel non nul m (supposé grand) et qui renvoie la liste de longueur $n + 1$ des fréquences de réalisation des événements $(Z_k = i)$ pour $i \in [\![0, n]\!]$ au cours de m simulations indépendantes de la variable aléatoire Z_k.

③ Écrire une fonction `moyEmpirZ(n,k,m)` d'argument l'entier n associé au nombre de stations, un entier naturel k et un entier naturel non nul m (supposé grand) et qui renvoie la moyenne de m simulations indépendantes de la variable aléatoire Z_k.

✐ *Dans les parties II et III qui suivent, et dans ces parties uniquement, on se place dans le cas particulier où $n = 2$. Le bus dessert donc les stations S_0, S_1 et S_2 et se trouve initialement en S_2.*

Partie II : Étude théorique.

① a) Déterminer les lois des variables aléatoires Z_0, Z_1 et Z_2.
b) Calculer l'espérance et la variance de chacune des variables aléatoires Z_0, Z_1 et Z_2.

② $\forall k \in \mathbb{N}$, on note C_k la colonne définie par $C_k = \begin{pmatrix} \mathbb{P}(Z_k = 0) \\ \mathbb{P}(Z_k = 1) \\ \mathbb{P}(Z_k = 2) \end{pmatrix}$ et B la matrice $\begin{pmatrix} 1 & 1/2 & 1/3 \\ 0 & 1/2 & 1/3 \\ 0 & 0 & 1/3 \end{pmatrix}$.

a) Démontrer que pour tout entier naturel k, on a : $C_{k+1} = B \cdot C_k$.
b) Déterminer les valeurs propres de B et les sous-espaces propres correspondants.
c) Montrer que la matrice B est diagonalisable et déterminer une matrice $P \in \mathcal{M}_3(\mathbb{R})$ inversible et dont tous les coefficients de la première ligne sont égaux à 1 et une matrice $D \in \mathcal{M}_3(\mathbb{R})$ diagonale dont les éléments diagonaux $(\lambda_1, \lambda_2, \lambda_3)$ vérifient $\lambda_1 > \lambda_2 > \lambda_3$ telles que : $B = PDP^{-1}$.
d) Justifier l'inversibilité de P et calculer P^{-1}.
e) Écrire pour tout entier naturel n la matrice B^k comme produit matriciel de matrices connues et déterminer la dernière colonne de B^k.
f) Préciser la valeur de C_0 et déduire de la question précédente, pour tout entier naturel k, la loi de la variable aléatoire Z_k.

✐ *Remarque :* On vérifiera notamment qu'on obtient : $\mathbb{P}(Z_k = 0) = 1 - 2\left(\frac{1}{2}\right)^k + \left(\frac{1}{3}\right)^k$.

g) Calculer l'espérance de Z_k pour tout entier naturel k.

Partie III : Étude du temps d'arrivée au dépôt.

① Pour tout $k \in \mathbb{N}^*$, soit A_k l'événement : « le bus arrive (pour la première fois) au dépôt à l'instant k ». En remarquant qu'une fois que le bus arrive en S_0, il n'en repart plus, exprimer A_k en fonction des événements $(Z_k = 0)$ et $(Z_{k-1} = 0)$ puis démontrer l'égalité :

$$\forall k \in \mathbb{N}^*,\ \mathbb{P}(A_k) = \left(\frac{1}{2}\right)^{k-1} - \frac{2}{3}\left(\frac{1}{3}\right)^{k-1}$$

② Déterminer la probabilité de l'événement F : « le bus n'arrive jamais au dépôt ».

③ Soit T la variable aléatoire égale à l'instant où le bus arrive au dépôt. Donner la loi de T et calculer son espérance.

④ Écrire une fonction `simultT()` sans argument, qui renvoie une valeur simulée de la variable aléatoire T définie ci-dessus.

⑤ Proposer une méthode utilisant Python permettant de confirmer la valeur théorique de $\mathbb{E}(T)$ trouvée ci-dessus.

Partie IV : Retour au cas général.

Désormais le bus dessert les stations $S_0, \cdots, S_n$ où $n \in \mathbb{N}^*$ et il est à l'instant 0 à la station S_n.

IV/ A. Étude d'un endomorphisme

On note E_n l'espace vectoriel $\mathbb{R}_n[X]$ ds polynômes à coefficients réels de degré inférieur ou égal à n et $\mathcal{B} = (1, X, \cdots, X^n)$ la base canonique de E_n.
On considère f l'application qui à tout polynôme P de E_n associe le polynôme Q défini par :

$$Q = (X-1)P' + P$$

c'est-à-dire la fonction polynomiale Q définie par : $\forall x \in \mathbb{R}$, $Q(x) = (x-1)P'(x) + P(x)$.

① Montrer que l'application f ainsi définie est un endomorphisme de E_n.

② Déterminer la matrice A de f dans la base $\mathcal{B}$.

③ En déduire les valeurs propres de f. L'endomorphisme f est-il diagonalisable ? Que peut-on dire de la dimension des sous-espaces propres associés à ses valeurs propres ?

④ Soit λ une valeur propre de f. Soit P un vecteur propre de f associé à la valeur propre λ. Montrer que P est solution sur $]1, +\infty[$ d'une équation différentielle linéaire du premier ordre qu'on précisera.

⑤ Pour $k \in [\![0, n]\!]$, résoudre sur $]1, +\infty[$ l'équation différentielle $(E_k) : (x-1)y' - ky = 0$.

⑥ En déduire une base $\mathcal{B}' = (P_0, P_1, \cdots, P_n)$ de E_n dans laquelle la matrice de f est diagonale et telle que pour $0 \leq j \leq n$, le polynôme P_j soit de degré j et de coefficient dominant égale à 1.

⑦ Justifier que f est bijectif. On note alors $g = f^{-1}$.

⑧ Justifier que pour tout $k \in \mathbb{N}$ et tout $j \in [\![0, n]\!]$ l'égalité : $g^k(P_j) = \dfrac{1}{(j+1)^k} P_j$.

IV/ B. Étude de la variable aléatoire Z_k

Pour tout entier naturel k on note F_k la fonction génératrice de Z_k dont on rappelle qu'elle est définie par :

$$\forall x \in \mathbb{R},\ F_k(x) = \sum_{r=0}^{n} \mathbb{P}(Z_k = r)x^r = \mathbb{P}(Z_k = 0) + \mathbb{P}(Z_k = 1)x + \cdots + \mathbb{P}(Z_k = n)x^n$$

① Pour tout $k \in \mathbb{N}$, rappeler ce que valent $F_k(1)$ et $F_k'(1)$ pour la variable aléatoire Z_k.

② Préciser les polynômes F_0 et F_1 dans la base canonique.

③ A l'aide de la formule des probabilités totales, établir les égalités :

$$\forall k \in \mathbb{N},\ \forall r \in [\![0, n]\!],\ \mathbb{P}(Z_{k+1} = r) = \sum_{j=r}^{n} \frac{\mathbb{P}(Z_k = j)}{j+1}$$

④ En déduire les deux formules suivantes, valables pour tout entier $k \in \mathbb{N}$ et tout $r \in [\![0, n-1]\!]$:

$$\begin{cases} (\mathcal{R}_1) : & (n+1)\mathbb{P}(Z_{k+1} = n) = \mathbb{P}(Z_k = n) \\ (\mathcal{R}_2) : & (r+1)\mathbb{P}(Z_{k+1} = r) - (r+1)\mathbb{P}(Z_{k+1} = r+1) = \mathbb{P}(Z_k = r) \end{cases}$$

⑤ Pour tout entier $k \in \mathbb{N}$ établir en utilisant les égalités ci-dessus la relation suivante :

$(S) : (X-1)F_{k+1}'(X) + F_{k+1}(X) = F_k(X)$ (ce résultat peut être admis pour la suite)

⑥ *Calcul de l'espérance de Z_k :*

a) En dérivant une première fois la relation (S), établir une relation de récurrence vérifiée par la suite $(F_k'(1))_{k \in \mathbb{N}}$.

b) En déduire la valeur de l'espérance $\mathbb{E}(Z_k)$ en fonction de k et de n pour tout $k \in \mathbb{N}$.
c) Dans le cas particulier où $n = 2$, retrouver la valeur de $\mathbb{E}(Z_k)$ calculée dans la partie II.

⑦ **Détermination de la loi de Z_k :**

a) En utilisant le résultat de la question IV/B.5 et la question IV/A.7, justifier que $F_{k+1} = g(F_k)$ où g est l'endomorphisme réciproque de f défini dans la partie IV/A. En déduire pour tout $k \in \mathbb{N}$ l'égalité : $F_k = g^k(X^n)$.

b) Pour tout $k \in \mathbb{N}$, justifier grâce au binôme de Newton l'égalité : $X^n = \displaystyle\sum_{j=0}^{n} \binom{n}{j} P_j$ où $P_0, P_1, \cdots, P_n$ sont les polynômes définis en IV/A.6.

c) En déduire que pour tout $k \in \mathbb{N}$ l'égalité : $F_k = \displaystyle\sum_{j=0}^{n} \binom{n}{j} \frac{1}{(j+1)^k} P_j$.

d) Établir enfin, pour tout $k \in \mathbb{N}$ et tout $r \in [\![0, n]\!]$, l'égalité :

$$\mathbb{P}(Z_k = r) = \sum_{j=r}^{n} (-1)^{(j-r)} \frac{\binom{n}{j}\binom{j}{r}}{(j+1)^k}$$

Complément. Fonction génératrice de variables à valeurs dans $\mathbb{N}$.

Dans la suite, nous admettrons la propriété suivante, généralisant le résultat des Préliminaires, A.1.b) : si X est une variable aléatoire à valeurs dans $\mathbb{N}$, alors la connaissance de $g_X(t)$ pour tout $t \in [-1, 1]$ entraîne la connaissance de la loi de X. Ceci permet donc de reconnaître la loi d'une variable aléatoire dont on connaît la fonction génératrice.
Par ailleurs, si g_X est dérivable en 1, alors $\mathbb{E}(X)$ existe et vaut $\mathbb{E}(X) = g'_X(1)$.
Soit maintenant X une variable aléatoire réelle à valeurs dans $\mathbb{N}$. Pour tout $n \in \mathbb{N}$, on pose :

$$a_n = \mathbb{P}(X = n) \text{ et donc } g_X(t) = \sum_{n=0}^{\infty} a_n t^n \text{ en tout } t \text{ réel où la série converge}$$

① Montrer que pour tout $t \in [-1, 1]$, la série $\sum a_n t^n$ est absolument convergente. En déduire que g_X est définie sur $[-1, 1]$ et donner la valeur de $g_X(1)$.

② Soit Y une variable aléatoire réelle à valeurs dans $\mathbb{N}$. Montrer que si X et Y sont indépendantes, alors pour tout $t \in [-1, 1]$, $g_{X+Y}(t) = g_X(t) g_Y(t)$

③ a) On suppose que X suit une loi géométrique de paramètre $p \in]0, 1[$, à valeurs dans $\mathbb{N}^*$. Calculer $g_X(t)$ pour $t \in [-1, 1]$ (on pourra poser $q = 1 - p$). En déduire l'existence et la valeur de $\mathbb{E}(X)$. Retrouver ce résultat par la méthode de votre choix.

b) On suppose ici que X suit une loi de Poisson de paramètre $\lambda > 0$.

i. Calculer $g_X(t)$ pour $t \in [-1, 1]$ et retrouver l'existence et la valeur de $\mathbb{E}(X)$.
ii. Si X et Y sont deux variables indépendantes qui suivent respectivement une loi de Poisson de paramètre λ et μ. Déterminer $g_{X+Y}(t)$ et en déduire que $X + Y$ suit une loi de Poisson dont on précisera le paramètre. Retrouver ce résultat par une autre méthode de votre choix.