Devoir maison n$^{\text{o}}$ 11 bis

À rendre le ? ? ?

L'usage de toute calculatrice est interdit.

Les problèmes sont indépendants et peuvent être traités dans un ordre quelconque. Le candidat composera sur des copies séparées pour chaque problème, et veillera à les identifier de façon claire.
Le correcteur sera particulièrement attentif à la clarté, à la rigueur et à la concision des raisonnement proposés. Tous les résultats demandés seront encadrés.

Définitions et notations

Dans tout le problème, nous utiliserons les notations suivantes.

— $\mathbb{N}$ représente l'ensemble des entiers naturels et on note $\mathbb{N}^* = \mathbb{N}\backslash\{0\}$, ensemble des entiers naturels non nuls ; $\mathbb{Z}$ est l'ensemble des entiers relatifs ; $\mathbb{R}$ est l'ensemble des nombres réels, $\mathbb{R}^+$l'ensemble des réels positifs ; $\mathbb{C}$ est l'ensemble des nombres complexes.

— Si p, q sont des entiers tels que $p \leqslant q$, on notera $[\![p; q]\!] = \{p, p+1, \ldots, q\}$ l'ensemble des entiers compris entre p et q inclus.

— Si X est une variable aléatoire, on notera $E(X)$ son espérance.

— On note $\mathscr{M}_n(\mathbb{R})$ (resp. $\mathscr{M}_n(\mathbb{C})$) l'ensemble des matrices carrées d'ordre n à coefficients dans $\mathbb{R}$ (resp. $\mathbb{C}$) et $\mathrm{GL}_n(\mathbb{R})$ (resp. $\mathrm{GL}_n(\mathbb{C})$) l'ensemble des matrices inversibles de $\mathscr{M}_n(\mathbb{R})$ (resp. de $\mathscr{M}_n(\mathbb{C})$). On note $\mathscr{M}_{n1}(\mathbb{R})$ l'ensemble des vecteurs colonne de taille n. Un vecteur colonne sera parfois noté $(z_1 z_2 \cdots z_n)^T$ pour des besoins de mise en page. Enfin, $\operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ représente la matrice diagonale dont les coefficients diagonaux sont $\lambda_1, \ldots, \lambda_n$.

Théorèmes utiles

— Si $\begin{array}{rccc} f: & [0;1]\times\mathbb{R} & \longrightarrow & \mathbb{R} \\ & (x,t) & \longmapsto & f(x,t) \end{array}$ est une fonction continue de deux variables, dérivable par rapport à t et si $\dfrac{\partial f}{\partial t}$ est une fonction continue par rapport à x et t, alors la fonction $t \longmapsto \displaystyle\int_0^1 f(x,t)\mathrm{d}x$ est de classe $\mathscr{C}^1$ et, pour tout $t \in \mathbb{R}$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_0^1 f(x,t)\mathrm{d}x = \int_0^1 \frac{\partial f}{\partial t}(x,t)\mathrm{d}x.$$

— si f est de classe $\mathscr{C}^k$ sur $[a;b]$, alors pour tout $\varepsilon \in [0; b-a]$, il existe $c \in]0; \varepsilon[$ tel que

$$f(a+\varepsilon) = f(a) + \sum_{i=1}^{k-1} \frac{\varepsilon^i}{i!} f^{(i)}(a) + \frac{\varepsilon^k}{k!} f^{(k)}(c)$$

— Soient $\mathrm{A}(t) = \displaystyle\sum_{n=0}^{+\infty} a_n t^n$ et $\mathrm{B}(t) = \displaystyle\sum_{n=0}^{+\infty} b_n t^n$ deux sommes de séries qui convergent pour tout réel de $]-R; R[$ où R est un réel supérieur ou égal à 1. On définit la suite $(c_n)_{n\in\mathbb{N}}$ par

$$\forall n \in \mathbb{N} \quad c_n = \sum_{i=0}^{n} a_i b_{n-i}$$

Alors la série $\sum c_n t^n$ converge pour tout réel $t \in]-R; R[$ et, de plus,

$$\forall t \in]-R; R[\quad \mathrm{C}(t) = \sum_{n=0}^{+\infty} c_n t^n = \mathrm{A}(t)\mathrm{B}(t).$$

La suite $c = (c_n)_{n\in\mathbb{N}}$ est appelée convolée (ou produit de convolution) de a et b, et on la note $c = a * b$.
On donne le tableau de quelques valeurs de la fonction $t \longmapsto \mathrm{e}^{-t}$:

t	1	2	3	4	5	6
e^{-t}	0,37	0,14	$5,0\cdot 10^{-2}$	$1,8\cdot 10^{-2}$	$6,7\cdot 10^{-3}$	$2,5\cdot 10^{-3}$

t	7	8	9	10	11	12
e^{-t}	$9,1\cdot 10^{-4}$	$3,4\cdot 10^{-4}$	$1,2\cdot 10^{-4}$	$4,5\cdot 10^{-5}$	$1,7\cdot 10^{-5}$	$6,1\cdot 10^{-6}$

Premier problème : équation de diffusion

Soit $g : [0;1] \longrightarrow \mathbb{R}^+$une fonction positive, continue, et telle que $\displaystyle\int_0^1 g(x)\mathrm{d}x = 1$.

On s'intéresse à une fonction $f : [0;1]\times\mathbb{R}^+ \to \mathbb{R}$, dépendant de deux variables x et t, vérifiant l'équation différentielle suivante, appelée équation de diffusion :

$$\forall (x,t) \in [0;1]\times\mathbb{R}^+ \quad \frac{\partial^2 f}{\partial x^2}(x,t) - \alpha\frac{\partial f}{\partial t}(x,t) = 0 \tag{E}$$

où $\alpha > 0$ est une constante, ainsi que la condition initiale

$$\forall x \in [0;1] \quad f(x,0) = g(x) \tag{CI}$$

et les conditions au bord

$$\frac{\partial f}{\partial x}(0,t) = \frac{\partial f}{\partial x}(1,t) = 0 \quad \forall t \geqslant 0. \tag{CB}$$

Cette équation permet notamment d'étudier la diffusion de la chaleur, ou bien la diffusion d'une substance dans un milieu : on suppose qu'à $t = 0$, la substance a une densité g sur $[0;1]$; la fonction $x \longmapsto f(x,t)$ représente alors la densité de cette substance à l'instant t.

1. Montrer que, pour tout $t \in \mathbb{R}^+$, $\displaystyle\int_0^1 f(x,t)\mathrm{d}t = 1$. Interpréter ce résultat.

2. On se propose de remplacer l'étude de l'équation (E) - qui fait intervenir des dérivées partielles - par l'étude d'un système différentiel - ne faisant intervenir que des équations différentielles ordinaires -, qui approche le problème précédent ; c'est ce qu'on appelle la discrétisation.
Pour cela, on choisit un entier naturel $n \geqslant 2$, et on pose

$$\Delta = \frac{1}{n+1} \qquad \text{et} \qquad x_k = k\Delta \qquad \text{pour } k \in \{0,\dots,n+1\}$$

On remplace l'équation (E) par le système

$$\forall k \in [\![1;n]\!] \quad \frac{f(x_{k+1},t) - 2f(x_k,t) + f(x_{k-1},t)}{\Delta^2} - \alpha\frac{\partial f}{\partial t}(x_k,t) = 0.$$

Pour des facilités d'écriture, on posera $\mathrm{Y}(t) = (y_1(t)\cdots y_n(t))^T$ et on étudiera le système d'équations différentielles

$$\forall k \in [\![1;n]\!] \quad \frac{y_{k+1}(t) - 2y_k(t) + y_{k-1}(t)}{\Delta^2} - \alpha y_k'(t) = 0. \tag{$\prime$}$$

De plus, la condition aux bords (CB) sera remplacée par

$$\forall t \in \mathbb{R}^+ \quad y_0(t) = y_1(t) \quad \text{et} \quad y_{n+1}(t) = y_n(t), \tag{CB'}$$

et on supposera que $\displaystyle\sum_{k=1}^n y_k(0) = 1$.

a) Montrer que le système (E′) assorti de sa condition aux bords (CB′) peut se mettre sous la forme

$$\mathrm{Y}'(t) = \frac{1}{\alpha\Delta^2}\mathrm{AY}(t)$$

Expliciter la matrice A.

b) Montrer que la quantité $\displaystyle\sum_{k=1}^{n} y_k(t)$ est constante par rapport au temps. Interpréter.

3. Montrer que 0 est valeur propre de A . En déduire toutes les solutions stationnaires du problème, c'est-à-dire les fonctions vectorielles $t \longmapsto \mathrm{Y}(t)$ constantes par rapport au temps.
4. On cherche à encadrer les valeurs propres de la matrice A .
   a) Rappeler pourquoi A est diagonalisable.
   b) On note Sp(A) l'ensemble des valeurs propres de A .
   Montrer que $\mathrm{Sp}(A) \subset [-4; 0]$.
   Indication : On pourra choisir un vecteur propre $\mathrm{Z} = (z_1 z_2 \cdots z_n)^T \in \mathscr{M}_{n1}(\mathbb{R})$ associé à une valeur propre λ.
5. a) On note $P \in GL_n(\mathbb{R})$ une matrice inversible telle que $P^{-1}AP$ est une matrice diagonale dont les coefficients diagonaux seront notés $\lambda_1, \ldots, \lambda_n$, classés dans l'ordre décroissant.
   Que vaut λ_1 ? Comparer λ_2 et 0 .
   b) On note $\mathrm{W}(t) = \mathrm{P}^{-1}\mathrm{Y}(t)$. Que vaut $\mathrm{W}'(t)$? En déduire $\lim\limits_{t\to+\infty} \mathrm{W}(t)$.
   c) En déduire que $\lim\limits_{t\to+\infty} \mathrm{Y}(t)$ existe et appartient à un certain sous-espace propre que l'on déterminera.
   En déduire la valeur de $\lim\limits_{t\to+\infty} \mathrm{Y}(t)$. Interpréter ce résultat.
6. **Exemple**. Dans cette question uniquement, on prend $n = 3$ et $\alpha = 4$.
   a) Écrire la matrice A . Trouver une matrice $\mathrm{P} \in \mathrm{GL}_3(\mathbb{R})$ telle que $\mathrm{P}^{-1}\mathrm{AP} = \mathrm{diag}(0, -1, -3)$.
   b) La condition initiale est donnée par $\mathrm{Y}(0) = \begin{pmatrix}1\\0\\0\end{pmatrix}$.
   Calculer W(0), puis W(t) et Y(t) pour tout $t \in \mathbb{R}^+$. Tracer l'allure des courbes $t \mapsto y_1(t), t \mapsto y_2(t)$ et $t \mapsto y_3(t)$.
   c) À partir de quelle valeur de t peut-on être sûr que la solution que nous venons de calculer atteint sa valeur limite avec une marge d'erreur de 10^{-5} ?
7. **Évaluation de l'erreur commise.** On suppose que, pour tout $t \in \mathbb{R}^+$, la fonction $x \longmapsto f(x,t)$ est de classe $\mathscr{C}^4$. Montrer qu'il existe une fonction $t \longmapsto \mathrm{M}(t)$, telle que

$$\forall k \in \{1, 2, \ldots, n\} \qquad \left| \frac{f(x_{k+1}, t) - 2f(x_k, t) + f(x_{k-1}, t)}{\Delta^2} - f''(x_k, t) \right| \leqslant \frac{\mathrm{M}(t)}{(n+1)^2}$$

   La fonction $t \longmapsto \mathrm{M}(t)$ dépend-elle de n ? Commenter.
8. Quels commentaires, remarques, critiques, pouvez-vous faire sur la méthode employée ?

Deuxième problème : détérioration d'une séquence génétique

On modélise une séquence génétique par une succession de sites, indicés par $\mathbb{N}$ ou par $\mathbb{Z}$. Pour chaque entier n, le site n pourra subir ou non une détérioration (mutation). On définit alors la variable aléatoire X_n en posant $\mathrm{X}_n = 1$ en cas de détérioration du site n, et $\mathrm{X}_n = 0$ dans le cas contraire.

Dans les parties A, B et C, nous considérerons des mutations ponctuelles, c'est-à-dire que $(X_n)_{n\in\mathbb{N}}$ sera une suite de variables indépendantes identiquement distribuées, suivant une loi de Bernoulli.

Dans la partie D , nous considérerons le cas des recombinaisons, où des segments entiers, de longueur aléatoire, sont détériorés.

Nous considérons maintenant un événement, noté $\mathscr{E}$, qui peut se produire à chaque site. On dira que cet événement est régénératif si et seulement si il vérifie la propriété suivante : les distances entre les occurrences successives de l'événement $\mathscr{E}$ sont des variables aléatoires mutuellement indépendantes et de même loi.

A Étude d'un événement régénératif : quelques relations fondamentales

Dans cette partie, la séquence génétique est indicée par $\mathbb{N}^* = \{1, 2, 3, \dots\}$.
Soit $p \in]0;1[$. Soit $(\mathrm{X}_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes et de même loi :

$$\mathrm{P}(\mathrm{X}_n = 1) = p \quad \text{et} \quad \mathrm{P}(\mathrm{X}_n = 0) = q = 1 - p$$

L'événement $\mathrm{X}_n = 1$ correspond à la détérioration du site n et l'événement $\mathrm{X}_n = 0$ correspond à la non-détérioration de ce site.

1. **Exemple** : On définit dans cette question l'événement $\mathscr{E}$ par : $\mathscr{E}$ a lieu au site n si et seulement si $\mathrm{X}_n = 1$. Montrer que $\mathscr{E}$ est régénératif.

Bien sûr, d'autres événements régénératifs peuvent être envisagés, comme par exemple la répétition d'un motif non recouvrant (cf. partie C).

Soit $\mathscr{E}$ un événement régénératif quelconque.
Pour tout $n \in \mathbb{N}^*$, notons a_n la probabilité que l'événement $\mathscr{E}$ ait lieu au n-ième site, et b_n la probabilité qu'il ait lieu pour la première fois au n-ième site :

$$\begin{cases} a_n = \mathrm{P}\{\mathscr{E} \text{ a lieu au site } n\} \\ b_n = \mathrm{P}\{\mathscr{E} \text{ a lieu pour la première fois au site } n\} \end{cases} \tag{I}$$

De plus, définissons par commodité $a_0 = 1$ et $b_0 = 0$. Ainsi, on supposera implicitement que $\mathscr{E}$ a lieu au site $n = 0$ (on ne cherchera pas de signification à cette hypothèse).

Enfin, posons

$$\mathrm{A}(x) = \sum_{n=0}^{+\infty} a_n x^n \quad \mathrm{B}(x) = \sum_{n=0}^{+\infty} b_n x^n \quad b = \sum_{n=0}^{+\infty} b_n = \mathrm{B}(1). \tag{2}$$

2. On note D la variable aléatoire exprimant la distance de la première occurrence de l'événement $\mathscr{E}$ (cette distance pouvant être $+\infty$ si l'événement $\mathscr{E}$ n'a pas lieu). On notera que, grâce à notre hypothèse sur a_0, D représente également la distance entre deux occurrences successives de l'événement $\mathscr{E}$.
   Expliciter la loi de D.
   Que représente le nombre $1 - b$ en terme d'occurrences de l'événement $\mathscr{E}$?
   Si $b = 1$, on dira que l'événement est récurrent.
3. On note $b_n^{(2)}$ la probabilité que $\mathscr{E}$ ait lieu pour la deuxième fois au n-ième site, et on définit

$$\mathrm{B}^{(2)}(x) = \sum_{n=0}^{+\infty} b_n^{(2)} x^n.$$

   Calculer, pour tout entier $n \in \mathbb{N}$, le coefficient $b_n^{(2)}$ en fonction de $b_1, b_2, \dots, b_{n-1}$.
   Écrire alors $b^{(2)} = \left(b_n^{(2)}\right)_{n\in\mathbb{N}}$ sous la forme d'un produit de convolution.
   En déduire $B^{(2)}$.

4. Montrer que, pour tout $n \in \mathbb{N}^*$, on a $a_n = (b * a)_n$, c'est-à-dire que $a_n = \displaystyle\sum_{k=0}^{n} b_k a_{n-k}$.
5. Montrer que $\mathrm{A}(x) = \dfrac{1}{1 - \mathrm{B}(x)}$ pour tout $x \in [0;1[$.

B Occurrence de sous-séquences entièrement détériorées

Soit K un entier naturel. On s'intéresse, dans la séquence des sites, aux occurrences de sous-séquences non recouvrantes de K sites détériorés. Plus précisément, on dira que l'événement $\mathscr{E}$ a lieu au site n si les sites $n-\mathrm{K}+1$ à n sont détériorés et si $\mathscr{E}$ n'a pas eu lieu aux sites $n-\mathrm{K}+1$ à n : ainsi, en représentant une détérioration par un « 1 » et un site sain par un « 0 », dans la séquence suivante

$$10111101111111010,$$

a-t-on des suites de 3 détériorations aux sites 5 , 10 et 13 (mais pas, par exemple, aux sites 6,11 ou 12).

6. Montrer que $\mathscr{E}$ est régénératif.
   On emploiera les notations introduites aux équations (1) et (2).
7. Soit $n \in \mathbb{N}$ tel que $n \geqslant \mathrm{K}$.
   a) Quelle est la probabilité d'avoir des sites détériorés aux rangs $(n-\mathrm{K}+1), \ldots, n-2, n-1, n$?
   b) En utilisant la formule des probabilités totales et en notant que, si les sites indexés $n-\mathrm{K}+1$ à n, sont détériorés, l'événement $\mathscr{E}$ a lieu à un et un seul de ces indices, trouver une relation entre $a_{n-\mathrm{K}+1}, \ldots, a_n$ et p.
   c) Que valent $a_1, \ldots, a_{K-1}$? On rappelle, pour la suite, que $a_0 = 1$ par définition.
   d) Montrer que
$$\sum_{n=\mathrm{K}}^{+\infty} p^{\mathrm{K}} x^n = \left(1 + xp + x^2p^2 + \cdots + x^{\mathrm{K}-1}p^{\mathrm{K}-1}\right) \sum_{n=\mathrm{K}}^{+\infty} a_n x^n$$
8. En déduire les expressions de $A(x)$ et de $B(x)$.
9. Calculer b et commenter.
10. Montrer que la distance moyenne à l'origine de la première occurrence de $\mathscr{E}$ (c'est-à-dire $E(D)$) vaut $\dfrac{1-p^{\mathrm{K}}}{qp^{\mathrm{K}}}$.
11. Quel est la distance moyenne de la première suite de 20 détériorations si $p = \dfrac{1}{2}$? Même question pour $p = \dfrac{1}{6}$.

C Occurrence de sous-séquences en partie détériorées

On s'intéresse aux occurrences non recouvrantes de sous-séquences de longueur 4, détériorées selon le motif suivant :

$$1101$$

On notera $\mathscr{E}^*$ l'événement correspondant, et $\mathrm{A}^*(x), a_n^*,\ \mathrm{B}(x), b_n^*, \mathrm{D}^*$ les quantités relatives à l'événement $\mathscr{E}^*$.
Exemple. Dans la séquence suivante :

$$01101101\underline{11010011001\cdots}$$

l'événement $\mathscr{E}$ a lieu aux sites 5 et 12 , mais pas au site 8 .

12. Montrer que $\mathscr{E}^*$ est régénératif.
13. Soit $n \in \mathbb{N}$ tel que $n \geqslant 4$.
    a) Quelle est la probabilité d'avoir le motif « 1101 » aux sites $n-3, n-2, n-1$ et n ?
    b) En remarquant que, si le motif « 1101 » a lieu aux sites $n-3, n-2, n-1$ et n, l'événement $\mathscr{E}$ a lieu à un et un seul de ces sites, montrer que $p^3q = a_n + a_{n-3}p^2q$ pour tout $n \geqslant 4$.
14. Déduire de ce qui précède l'expression de $A^*(x)$.
15. Calculer alors $\mathrm{B}^*(x)$. Montrer que $b^* = \mathrm{B}^*(1) = 1$.
16. Calculer $\mathrm{E}\left(\mathrm{D}^*\right)$.
17. Comparer les espérances de D et de D^* dans le cas $\mathrm{K} = 4, p = \dfrac{1}{10}$.
    Effectuer de même la comparaison dans le cas où $p = \dfrac{1}{2}$, et commenter.

D Généralisation à la recombinaison

On modélise dans cette partie une séquence génétique par une succession de sites indicés par un entier relatif $n \in \mathbb{Z}$. On s'intéresse maintenant à la détérioration de cette séquence par un processus de recombinaison. À chaque site n, successivement, peut avoir lieu une recombinaison, qui détériore un segment, de taille aléatoire, dont l'extrémité droite est n. En d'autres termes, on a le processus suivant :

— à chaque site $n \in \mathbb{Z}$, une recombinaison a lieu avec la probabilité $1 - \alpha_0$;

— cette recombinaison résulte en la détérioration du site n seulement avec la probabilité α_1 ;

— pour tout $i \in \mathbb{N}^*$, elle résulte en la détérioration des sites $n-i+1, n-i+2, \cdots, n$ avec la probabilité α_i ; les réels α_i appartenant à $]0;1[$ et vérifiant $\displaystyle\sum_{i=0}^{+\infty} \alpha_i = 1$.

Un site sera sain s'il n'a subi aucune détérioration. Un site détérioré plusieurs fois reste détérioré.

18. **Préliminaires**

 a) Soit $(u_n)_{n\in\mathbb{N}}$ une suite de réels positifs tels que $0 \leqslant u_n < 1$ pour tout $n \in \mathbb{N}$. On pose, pour tout entier n, $\mathrm{P}_n = \displaystyle\prod_{k=0}^{n} (1-u_k)$. Montrer que la suite $(\mathrm{P}_n)_{n\in\mathbb{N}}$ converge et que sa limite ℓ appartient à $[0;1]$. On dira alors que le produit infini $\displaystyle\prod_{k=0}^{+\infty} (1-u_k)$ converge et on écrira

$$\prod_{k=0}^{+\infty} (1-u_k) = \ell$$

 b) Montrer que la série $\sum \ln(1-u_n)$ converge si et seulement si la série $\sum u_n$ converge.

 c) Donner une condition nécessaire et suffisante portant sur la série $\sum u_n$ pour que $\displaystyle\prod_{k=0}^{+\infty} (1-u_k) = 0$.

 d) Soit $\beta > 0$. En comparant $\displaystyle\sum_{k=1}^{n} \frac{1}{k^\beta}$ et la l'intégrale $\displaystyle\int_1^n \frac{1}{t^\beta}\mathrm{d}t$, donner une condition nécessaire et suffisante sur β pour que la série $\sum \dfrac{1}{n^\beta}$ converge.

19. Calculer, en fonction de $(\alpha_n)_{n\in\mathbb{N}}$, la probabilité qu'un site donné soit détérioré. Donner une condition nécessaire et suffisante pour que cette probabilité soit égale à 1 en fonction de $(\mathrm{R}_n)_{n\in\mathbb{N}}$, puis en fonction de $(\alpha_n)_{n\in\mathbb{N}}$.
 On pourra utiliser les notations suivantes :

$$S_n = \sum_{k=0}^{n} \alpha_k \quad \mathrm{R}_n = 1 - S_n = \sum_{k=n+1}^{+\infty} \alpha_k$$

20. Cette probabilité vaut-elle 1 ou non dans les cas suivants ?

 a) la longueur de la séquence détériorée à chaque recombinaison suit une loi de Poisson de paramètre $\lambda \in \mathbb{R}^+$, c'est-à-dire que $\alpha_k = \dfrac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$;

 b) elle suit la loi donnée par $\alpha_k = (1-\gamma)\gamma^k$, où $\gamma \in]0;1[$;

 c) elle suit la loi donnée par $\alpha_k = \dfrac{1}{(k+1)(k+2)}$ (on vérifiera au préalable que la suite $(\alpha_n)_{n\in\mathbb{N}}$ définit bien une loi de probabilité).