Question de cours.

Exemple 1

Énoncer l'inégalité de Markov.

Exercice.

On considère d'une part deux urnes A et B et d'autre part 3 boules numérotées de 1 à 3.
On répartit initialement les boules entre les deux urnes, puis on effectue une série illimitée d'étapes selon le protocole suivant : à chaque étape, on tire au hasard un nombre entre 1 et 3 et on transfère la boule portant le numéro correspondant dans l'urne où elle n'était pas.
On note X_0 la variable aléatoire égale au nombre de boules présentes initialement dans l'urne A et pour tout entier naturel n non nul, on note X_n la variable aléatoire égale au nombre de boules présentes dans l'urne A après n étapes.
On suppose que X_0 suit la loi uniforme sur $[\![0,3]\!]$.

1. Écrire une fonction Python qui prend en argument une valeur de n, simule la réalisation de la variable aléatoire X_n et renvoie la valeur de X_n obtenue.
2. Soient $M = \begin{pmatrix} 0 & 1/3 & 0 & 0 \\ 1 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1 \\ 0 & 0 & 1/3 & 0 \end{pmatrix}$, et pour tout $n \in \mathbb{N}$, $U_n = \begin{pmatrix} P(X_n=0) \\ P(X_n=1) \\ P(X_n=2) \\ P(X_n=3) \end{pmatrix}$.
   Déterminer U_0 et démontrer que pour tout entier naturel n, $U_{n+1} = MU_n$.
3. Soit E l'espace vectoriel des fonctions polynomiales de dégré inférieur ou égal à 3 et $\mathcal{B} = (1, X, X^2, X^3)$ la base canonique de E .
   Soit φ l'application définie sur E par :
   $$\forall P \in E, \quad \varphi(P) = XP(X) + \frac{1}{3}(1-X^2)P'(X).$$
   (a) Montrer que φ est un endomorphisme de E, justifier que sa matrice représentative dans la base $\mathcal{B}$ est la matrice M.
   (b) Pour tout $k \in [\![0,3]\!]$, on pose : $P_k(X) = \dfrac{1}{8}(X-1)^k(X+1)^{3-k}$.
   Démontrer que pour $k \in [\![0,3]\!]$, $\varphi(P_k) = \left(1 - \dfrac{2}{3}k\right) P_k$.
   (c) En déduire que l'endomorphisme est diagonalisable et déterminer ses valeurs propres.
4. On pose : $Q = \dfrac{1}{4}(X^3 + X^2 + X + 1)$.
   (a) Expliciter les coordonnées de Q dans la base $\mathcal{B}$. Quel vecteur retrouve-t-on ?
   (b) Montrer que : $\forall n \in \mathbb{N}, \quad \varphi^n(Q) = P_0 + \left(-\dfrac{1}{3}\right)^n P_2$.
5. À l'aide des questions précédentes, calculer $\lim\limits_{n\to+\infty} P(X_n = 0)$, $\lim\limits_{n\to+\infty} P(X_n = 1)$, $\lim\limits_{n\to+\infty} P(X_n = 2)$ et $\lim\limits_{n\to+\infty} P(X_n = 3)$. De quelle loi pourrait-on approcher la loi de X_n pour une grande valeur de n ?
6. Vérifier le résultat de la question précédente à l'aide d'une simulation informatique.

Question de cours.

Exemple 2

Donner la définition de la fonction de répartition d'une variable aléatoire réelle.
Donner son tableau de variation.

Exercice.

Soit $(a_1, a_2, a_3) \in \mathbb{C}^3$. On considère la matrice de $\mathcal{M}_3(\mathbb{C})$ définie par : $A = \begin{pmatrix} a_1 & a_2 & a_3 \\ a_2 & a_3 & a_1 \\ a_3 & a_1 & a_2 \end{pmatrix}$.

On pose $j = -\dfrac{1}{2} + \mathrm{i}\dfrac{\sqrt{3}}{2}$.

1. Calculer j^2, j^3 et j^4.
2. (a) Soit r et s deux complexes non nuls.
   Montrer que la matrice $M = \begin{pmatrix} 0 & r^2 \\ s^2 & 0 \end{pmatrix}$ est diagonalisable.
   Donner une base de vecteurs propres de M.

   (b) La matrice $M = \begin{pmatrix} 0 & r^2 \\ s^2 & 0 \end{pmatrix}$ est-elle diagonalisable lorsque $r = 0$ ou $s = 0$?
3. Écrire une fonction `decalage(L)` qui renvoie, si $L = [a_1, \ldots, a_n]$, la liste $L_1 = [a_2, \ldots, a_n, a_1]$.
   Utiliser cette fonction pour écrire une fonction `matrice(a1,a2,a3)` qui renvoie la matrice A.
   On pourra par exemple compléter le script suivant :

```
def matrice(a1,a2,a3):
    A=....
    L=....
    for i in ...
        A.append(L[:]);
        ...
    return ...
```

4. Si a_1, a_2, a_3 sont réels, la matrice A est-elle diagonalisable ?
5. Montrer que $U = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ est un vecteur propre de A. Quelle est la valeur propre associée ?
6. On pose $X_1 = \begin{pmatrix} 1 \\ j \\ j^2 \end{pmatrix}$ et $X_2 = \begin{pmatrix} 1 \\ j^2 \\ j \end{pmatrix}$. On pourra utiliser sans justifier que la famille (U, X_1, X_2) est libre.

   (a) Calculer AX_1 et AX_2.
   En déduire qu'il existe des complexes r et s tels que $AX_1 = s^2 X_2$ et $AX_2 = r^2 X_1$.

   (b) Déterminer le spectre de A.
   On pourra exprimer les valeurs propres à l'aide des complexes r et s introduits question 6 et utiliser la question 2.
7. Préciser dans les cas suivants si la matrice A est diagonalisable.

   (a) $A = \begin{pmatrix} 1 & j & j^2 \\ j & j^2 & 1 \\ j^2 & 1 & j \end{pmatrix}$ (b) $A = \begin{pmatrix} j & 1 & 0 \\ 1 & 0 & j \\ 0 & j & 1 \end{pmatrix}$

Question de cours.

Exemple 3

Définition de l'espérance d'une variable aléatoire X discrète à valeurs dans $\mathbb{N}$.

Exercice.

Dans tout l'exercice, on considère $\mathbb{R}^3$ muni du produit scalaire canonique. On rappelle que si $u = (u_1, u_2, u_3)$ et $v = (v_1, v_2, v_3)$ sont deux vecteurs de $\mathbb{R}^3$, alors leur produit scalaire est :

$$\langle u, v\rangle = u_1v_1 + u_2v_2 + u_3v_3.$$

Soit f un endomorphisme de $\mathbb{R}^3$ dont la matrice dans la base canonique de $\mathbb{R}^3$ est donnée par :

$$A = \frac{1}{6}\begin{pmatrix} 5 & -2 & 1 \\ -2 & 2 & 2 \\ 1 & 2 & 5 \end{pmatrix}$$

1. La matrice A est-elle diagonalisable ?
   Montrer que $\mathrm{Sp}(A) = \{0\,;1\}$ et déterminer les sous-espaces propres E_0 et E_1 de f.
2. La matrice A est-elle inversible ?
3. Écrire une fonction Python `proj` qui prend en argument une matrice carrée M (entrée par exemple sous la forme d'une liste de listes) et qui renvoie un booléen : `True` si $M^2 = M$, `False` sinon.
   Tester cette fonction sur la matrice A.
4. Montrer que $\mathrm{Im}\, f = E_1$
5. Écrire une fonction Python `ps` prenant en argument deux vecteurs de $\mathbb{R}^3$ codés sous forme de tableaux numpy ou de listes, et retournant leur produit scalaire.
6. (a) Montrer que les deux sous-espaces propres E_0 et E_1 sont orthogonaux, c'est-à-dire que :

   $$\forall u \in E_0, \quad \forall v \in E_1, \quad \langle u, v\rangle = 0.$$

   On pourra éventuellement utiliser la fonction `ps`.

   (b) En déduire que pour tout vecteur x de $\mathbb{R}^3$, on a :

   $$f(x) \in E_1 \quad \text{et} \quad \forall y \in E_1, \quad f(x) - x \;\text{ orthogonal à } y.$$

   On a donc montré que f était la projection orthogonale sur E_1.
7. Déterminer une base orthonormale de E_1. En déduire, en utilisant la fonction `ps`, une valeur approchée de la distance du vecteur $t = (1, 2, 1)$ à l'espace E_1 à 10^{-2} près.

Question de cours.

Exemple 4

À quelle(s) condition(s) sur sa fonction de répartition une variable aléatoire X admet-elle une densité de probabilité ? Comment détermine-t-on alors une densité de X ?

Exercice.

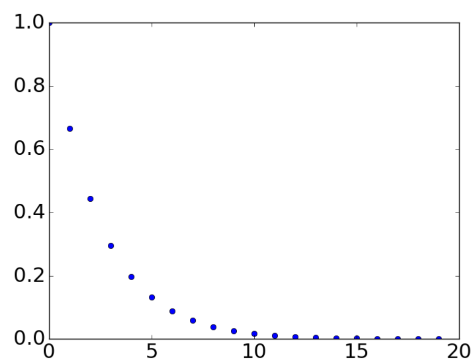
Rappel : algorithme de dichotomie. *On considère une fonction f continue sur un segment $[a,b]$. On suppose que f s'annule exactement une fois sur $[a,b]$, en un point que l'on note γ.*
On définit les suites $(a_k)_{k\geqslant 0}$ et $(b_k)_{k\geqslant 0}$ de la façon suivante:
• *$a_0 = a$ et $b_0 = b$.*
• *Pour tout entier naturel k, on note $c_k = \dfrac{a_k+b_k}{2}$ et :* $\left| \begin{array}{l} \text{si } f(a_k)f(c_k) \leqslant 0, \text{ alors } a_{k+1}=a_k \text{ et } b_{k+1}=c_k \\ \text{sinon, on pose } a_{k+1}=c_k \text{ et } b_{k+1}=b_k. \end{array}\right.$
On sait alors que les suites (a_k) et (b_k) convergent toutes les deux vers γ, en vérifiant:

$$\forall k \in \mathbb{N}, a_k \leqslant \gamma \leqslant b_k \text{ et } \forall k \in \mathbb{N}, b_k - a_k = \frac{b-a}{2^k}.$$

On peut montrer que si l'entier k est tel que $\dfrac{b-a}{2^k} \leqslant \varepsilon$, alors a_k et b_k sont des valeurs approchées de γ à ε-près.

1. On considère pour tout entier $n \geqslant 2$, la fonction f_n définie sur $\mathbb{R}_+^*$ par : $f_n(x) = x^n - \ln x - n$.
   (a) Dresser le tableau de variations de f_n.
   (b) *On rappelle et on admet l'inégalité suivante :* $\forall x \in \mathbb{R}_+^*,, \ \ln(x) \leqslant x - 1 \quad (*)$.
   En déduire que pour tout entier $n \geqslant 2$, il existe un unique réel $u_n \in]0, n^{-\frac{1}{n}}[$ tel que $f_n(u_n) = 0$ et un unique réel $v_n \in]n^{-\frac{1}{n}}, +\infty[$ tel que $f_n(v_n) = 0$.
2. **Étude de la suite** (v_n)
   La suite $(v_n)_{n\geqslant 2}$ vérifie donc l'égalité : $\forall n \geqslant 2, v_n^n - \ln(v_n) - n = 0$.
   (a) Justifier en utilisant si besoin $(*)$ que $f_n((2n)^{\frac{1}{n}}) > 0$, puis en déduire que : $\forall n \geqslant 2,\ v_n \leqslant (2n)^{\frac{1}{n}}$.
   (b) En utilisant l'algorithme de dichotomie, déterminer des valeurs approchées à 10^{-3} près des termes v_n pour n allant de 2 à 30. Représenter la suite (v_n) graphiquement.
   On rappelle que la fonction **`plot`** *des modules* **`pylab`** *ou* **`matplotlib.pyplot`** *permet de faire des représentations graphiques comme le montre l'exemple suivant :*

```
import matplotlib.pyplot as plt
X=[n for n in range(20)]
Y=[(2/3)**k for k in X]
plt.plot(X,Y,'o')
plt.show()
```



   (c) Montrer que (v_n) converge et déterminer sa limite ℓ.
   (d) *On admet le résultat suivant : si $a_n \underset{n\to+\infty}{\sim} b_n$ avec $\lim\limits_{n\to+\infty} b_n = +\infty$, alors* $\ln(a_n) \underset{n\to+\infty}{\sim} \ln(b_n)$.
   On rappelle de plus que $\ln(x) \underset{x\to 1}{\sim} x - 1$.
   Déterminer un équivalent de $v_n - \ell$.
3. **Étude de la suite** (u_n)
   (a) Proposer une méthode permettant de déterminer des valeurs approchées à 10^{-3} près des termes u_n pour n allant de 2 à 8.
   (b) Calculer $f_n(u_{n+1})$. En déduire le sens de variation de (u_n).
   (c) Justifier que $\lim\limits_{n\to+\infty} u_n^n = 0$, puis montrer que u_n est équivalent à e^{-n}.

Question de cours.

Exemple 5

Rappeler la formule des accroissements finis.

Exercice.

Rappel : *la fonction définie ci-dessous permet de représenter graphiquement la loi d'une variable aléatoire à valeurs dans* $[\![0, n]\!]$ *(*$n \in \mathbb{N}$*),* `loi` *étant la liste* $[P(X_n = 0), P(X_n = 1), \ldots, P(X_n = n)]$.

```
from matplotlib.pyplot import *
def graphe(loi):
        lx = [i for i in range(len(loi))]
        bar(lx,loi)
        ylim(0,0.5)
        show()
```

Pour se rendre d'un endroit à un autre, les individus d'une fourmillère ont le choix entre deux trajets disjoints, que nous nommerons A et B. À chaque fois qu'une fourmi emprunte l'un des deux chemins, elle y dépose une certaine quantité de phéromone qui peut éventuellement dépendre de la quantité de phéromone déjà présente sur le chemin.

Notations : pour chaque $n \geqslant 1$, α_n (respectivement β_n) désigne la quantité de phéromone présente sur le chemin A (resp. B) après le n-ième trajet. A_n (resp. B_n) désigne l'événement « la n-ième fourmi choisit le trajet A (resp. B) ». Nous supposerons que chaque fourmi choisit de façon aléatoire le chemin qu'elle emprunte, en affectant à chacun une probabilité proportionnelle à la quantité de phéromone qui y est présente.

On a donc : $P_{[\alpha_n = a] \cap [\beta_n = b]}(A_{n+1}) = \dfrac{a}{a+b}$ et $P_{[\alpha_n = a] \cap [\beta_n = b]}(B_{n+1}) = \dfrac{b}{a+b}$

Enfin, X_n désignera le nombre de fourmis ayant choisi le trajet A lors des n premiers trajets.

Nous supposerons qu'initialement $\alpha_0 = \beta_0 = 1$ et qu'à chaque trajet une fourmi multiplie par un facteur $r > 1$ la quantité de phéromone déjà présente sur le chemin qu'elle emprunte.

1. Déterminer la loi des variables X_1, X_2 et X_3.
2. Rédiger une fonction `simulX` qui reçoit un entier n et un réel r, simule les déplacements de n fourmis suivant la règle énoncée et renvoie le nombre X_n de fois où le chemin A a été emprunté.
3. (a) Rédiger une fonction `loiX` qui reçoit un entier n et un réel r et renvoie, sous forme de liste, des valeurs approchées des probabilités $[P(X_n = 0), P(X_n = 1), \ldots, P(X_n = n)]$ obtenues en faisant 1000 simulations de la variable X_n.

   (b) Représenter graphiquement la loi de la variable X lorsque $n = 100$ et $r = 2$. Commenter.
4. Exprimer en fonction de n et de r la probabilité $P(X_n = n)$.
   On ne tentera pas de simplifier l'expression obtenue.
5. On pose : $\forall n \geqslant 1,\ p_n(r) = \dfrac{r}{r+1} \cdot \dfrac{r^2}{1+r^2} \cdots \dfrac{r^n}{1+r^n}$, et $\forall n \in \mathbb{N}, q_n(r) = \left(1 - \dfrac{1}{r}\right)\left(1 - \dfrac{1}{r^3}\right) \cdots \left(1 - \dfrac{1}{r^{2n+1}}\right)$.
   Démontrer que pour tout $r > 1$, les suites $(p_n(r))_{n \geqslant 1}$ et $(q_n(r))_{n \in \mathbb{N}}$ convergent.
   On notera $p(r)$ et $q(r)$ leurs limites respectives.
6. En remarquant que $\forall n \geqslant 1, p_n(r) = \dfrac{1}{1+r^{-1}} \cdot \dfrac{1}{1+r^{-2}} \cdots \dfrac{1}{1+r^{-n}}$, montrer que : $p(r) \geqslant \exp\left(-\dfrac{1}{r-1}\right)$.
   On pourra admettre sans le démontrer l'inégalité suivante : $\forall x \in \mathbb{R}, e^x \geqslant 1 + x$.
7. Démontrer que $\forall r > 1, q(r) \leqslant \exp\left(-\dfrac{r}{r^2-1}\right)$
8. On admet que $\forall r > 1, p(r) = q(r)$. Déduire des questions précédentes un encadrement de la limite de la probabilité $P(X_n = n)$ en fonction de r.

Conclusion : ce modèle vous semble-t-il approprié pour rendre compte du comportement des fourmis dans la réalité ?

Question de cours.

Exemple 6

Si f est la fonction définie sur $]0,1[$ par : $f(x)=\sqrt{1-x}$, déterminer l'expression de sa dérivée f'.

Exercice.

On souhaite estimer un paramètre $p \in]0,1[$. On note : $q=1-p$.
Soit un entier $n \geqslant 1$ fixé. On considère $X_1, X_2, \ldots, X_n$ des variables aléatoires, indépendantes, suivant une même loi de Bernoulli de paramètre p et définies sur un même espace probabilisé.
On note : $\overline{X}_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$.

1. (a) Justifier que $p(1-p) \leqslant \dfrac{1}{4}$.

   (b) En utilisant l'inégalité de Bienaymé-Tchebychev, montrer que l'intervalle $\left[\overline{X}_n - \sqrt{\dfrac{5}{n}}, \overline{X}_n + \sqrt{\dfrac{5}{n}}\right]$ est un intervalle de confiance de p au niveau de confiance 0,95.

2. Écrire une fonction Python `test(n,p,a,b)` qui prend en arguments un entier `n`, une probabilité `p`, deux flottants `a` et `b`, simule une réalisation de $\overline{X}_{\mathtt{n}}$ et retourne 1 si $\overline{X}_{\mathtt{n}}$ appartient à [`a`,`b`] et 0 sinon.

   On cherche par la suite un intervalle de confiance de p au niveau de confiance 0,95 d'amplitude plus petite.

3. On fixe un réel strictement positif t quelconque et ε un réel strictement positif quelconque.

   (a) Établir l'égalité : $P\Big(\overline{X_n} - p \geqslant \varepsilon\Big) = P\Big(e^{nt\overline{X}_n} \geqslant e^{nt(p+\varepsilon)}\Big)$.

   (b) En utilisant l'inégalité de Markov, établir l'inégalité suivante : $P\Big(\overline{X}_n - p \geqslant \varepsilon\Big) \leqslant e^{n\left(\ln(pe^t+q)-t(p+\varepsilon)\right)}$.

   (c) On admet l'inégalité : $\ln(pe^t+q)-tp \leqslant \dfrac{t^2}{8}$. Ainsi, on a l'inégalité suivante :

   $$\forall t \in \mathbb{R}_+, P\Big(\overline{X}_n - p \geqslant \varepsilon\Big) \leqslant e^{n(\frac{t^2}{8}-t\varepsilon)}.$$

   En déduire l'inégalité : $P\Big(\overline{X}_n - p \geqslant \varepsilon\Big) \leqslant e^{-2n\varepsilon^2}$.

4. On pose $\overline{Y}_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n}(1-X_k)$. Établir l'inégalité : $P\Big(\overline{Y}_n - q \geqslant \varepsilon\Big) \leqslant e^{-2n\varepsilon^2}$.

5. Déduire des questions 3(c) et 4 l'inégalité :

   $$\forall \varepsilon \in \mathbb{R}_+^*,\ P\Big(\left|\overline{X}_n - p\right| \geqslant \varepsilon\Big) \leqslant 2e^{-2n\varepsilon^2}$$

6. Comment choisir ε pour obtenir un intervalle de confiance de p au niveau de confiance 0,95 ? L'amplitude de l'intervalle de confiance est-elle plus réduite que celle obtenue à la question 1(b) ?

Question de cours.

Exemple 7

Condition nécessaire et suffisante pour qu'une application linéaire soit injective.

Exercice.

Soient $(X_k)_{k\geqslant 1}$ une suite de variables aléatoires mutuellement indépendantes, définies sur un même espace probabilisé, suivant chacune une même loi exponentielle de paramètre 1.

Pour tout $n \geqslant 2$, on note $Y_n = \max(X_1, X_2, \dots, X_n)$, et on note $S_n = \displaystyle\sum_{k=1}^{n} \frac{1}{k}$.

1. (a) Soit U une variable aléatoire, de loi uniforme sur $]0,1]$. Vérifier que la variable $-\ln(U)$ suit une loi exponentielle de paramètre 1.
   (b) En déduire une fonction Python qui prend un entier n en entrée, et renvoie une simulation de la variable aléatoire Y_n.
   (c) En admettant que la variable aléatoire Y_n admet une espérance, à l'aide de la fonction Python précédente, conjecturer la valeur de $\displaystyle\lim_{n\to+\infty} \frac{E(Y_n)}{S_n}$.

Dans toute la suite de l'exercice, on fixe n un entier tel que $n \geqslant 2$.

2. On note F_n la fonction de répartition de Y_n.
   Montrer que :
   $$\forall x \in \mathbb{R},\ F_n(x) = \begin{cases} 0 & \text{si } x < 0, \\ (1-e^{-x})^n & \text{si } x \geqslant 0. \end{cases}$$
   En déduire que la variable Y_n est une variable à densité, et déterminer une densité f_n de Y_n.
3. (a) Montrer que pour tout réel u de $[0,1]$, on a :
   $$(1-u)^n \geqslant 1 - nu$$
   (b) En déduire que l'intégrale $\displaystyle\int_0^{+\infty} (1-F_n(x))\mathrm{d}x$ est convergente et que $\displaystyle\lim_{x\to+\infty} x(1-F_n(x)) = 0$.
4. (a) Pour tout $A > 0$, montrer à l'aide d'une intégration par parties que :
   $$\int_0^A x f_n(x)\mathrm{d}x = \int_0^A (1-F_n(x))\mathrm{d}x - A(1-F_n(A)),$$
   (b) En déduire que la variable Y_n admet une espérance, vérifiant :
   $$E(Y_n) = \int_0^{+\infty} (1-F_n(x))\mathrm{d}x$$
5. À l'aide du changement de variable $t = 1 - e^{-x}$, montrer que :
   $$E(Y_n) = \int_0^1 \frac{1-t^n}{1-t}\mathrm{d}t,$$
   et en déduire finalement que :
   $$E(Y_n) = S_n.$$

Question de cours.

Exemple 8

Définition d'une famille génératrice dans un espace vectoriel E.

Exercice.

Dans tout l'exercice, les variables aléatoires sont définies sur le même espace probabilisé $(\Omega, \mathcal{T}, P)$.
Si X *est une variable aléatoire, on notera* $E(X)$ *son espérance et* $V(X)$ *sa variance.*

Soient $a \in]0,1]$ et f la fonction définie par :

$$\forall x \in \mathbb{R}, \quad f(x) = \begin{cases} \dfrac{x}{a} \exp\left(-\dfrac{x^2}{2a}\right) & \text{si } x \geqslant 0 \\ 0 & \text{si } x < 0. \end{cases}$$

1. Montrer que f est une densité.
2. On considère dorénavant X une variable aléatoire de densité f.
   Déterminer la fonction de répartition F_X de X.
3. On considère la variable aléatoire Y donnée par :
   $$Y = \frac{X^2}{2a}.$$
   (a) Montrer que Y suit la loi exponentielle de paramètre 1.
   (b) On pose $U = 1 - e^{-Y}$ (de sorte que $Y = -\ln(1-U)$). Montrer que U suit la loi uniforme sur $]0,1[$.
   (c) En déduire une fonction Python `Y()` qui simule la variable Y.
   (d) Écrire une fonction Python `X(a)` qui prend en entrée un réel $a \in]0,1]$ et qui simule X.
4. (a) Donner une densité, que l'on notera g, d'une variable aléatoire suivant la loi normale d'espérance nulle et de **variance** a.
   (b) À l'aide d'un intégration par parties, en déduire que X possède une espérance et la calculer.
   (c) En utilisant la variable Y, montrer que X^2 possède une espérance et la calculer.
   (d) En déduire que $V(X) = \dfrac{(4-\pi)a}{2}$.
5. On considère désormais que la paramètre $a \in]0,1]$ est inconnu et on souhaite l'estimer.
   Soit $n \in \mathbb{N}^*$. On considère n variables aléatoires $X_1, \ldots, X_n$ indépendantes ayant toutes la même loi que X. On note :
   $$S_n = \frac{1}{2n} \sum_{k=1}^{n} X_k^2.$$
   (a) Montrer que S_n est un estimateur sans biais de a.
   (b) Montrer que X^2 admet une variance et montrer que $V(X^2) = 4a^2$.
   (c) Montrer que $V(S_n) \leqslant \dfrac{1}{n}$. Puis, à l'aide de l'inégalité de Bienaymé-Tchebychev, déterminer une valeur de n à partir de laquelle $\left]S_n - \dfrac{1}{10}, S_n + \dfrac{1}{10}\right[$ est un intervalle de confiance pour a avec un niveau de confiance au moins égal à 95%.

Question de cours.

Exemple 9

Si α est un réel quelconque, déterminer sur $]0, +\infty[$ l'expression d'une primitive de la fonction $x \mapsto \dfrac{1}{x^\alpha}$.

Exercice.

Une compagnie fait passer des entretiens d'embauche à n candidats. On suppose que la compétence de chaque candidat est quantifiée par une variable aléatoire X_i suivant une loi uniforme sur $[0, 1]$, d'autant plus élevée que le candidat est compétent. De plus, on suppose que les variables $(X_1, \ldots, X_n)$ sont mutuellement indépendantes. À la fin de chaque entretien, la compagnie doit immédiatement donner sa décision : soit elle embauche le candidat, soit elle passe au suivant, sans possibilité de revenir sur ses pas.
La compagnie cherche à élaborer une stratégie qui lui permettrait de maximiser l'espérance de la compétence du candidat qu'elle choisira. Pour ce faire, elle décide de fixer un seuil $s \in [0, 1]$. Si, parmi les $n - 1$ premiers candidats, aucun ne dépasse le seuil, la compagnie embauchera le dernier candidat. Sinon, elle choisira le premier candidat qui dépasse le seuil.

Pour tout $k \in [\![1\,,\,n-1]\!]$, on note A_k l'évènement : « pour tout $i \in [\![1\,,\,k-1]\!]$, $X_i < s$, et $X_k \geqslant s$ », et B l'évènement : « pour tout $k \in [\![1\,,\,n-1]\!]$, $X_k < s$ ».
On définit par conséquent la variable aléatoire $Z_{n,s}$, compétence du candidat retenu, comme suit :

$$Z_{n,s} = \begin{cases} X_n & \text{si } B \text{ est réalisé,} \\ X_k & \text{si } A_k \text{ est réalisé } (1 \leqslant k \leqslant n-1). \end{cases}$$

1. (a) Écrire un programme Python qui prend en argument un réel $s \in [0, 1]$, un entier naturel non nul n, et retourne une réalisation de $Z_{n,s}$.
   (b) En déduire un programme qui retourne une valeur approchée de la compétence moyenne du candidat recruté via ce protocole.
2. Calculer $E(Z_{n,0})$ et $E(Z_{n,1})$.
3. Montrer que, pour tout $t \in [0, 1]$, on a : $P(B \cap [Z_{n,s} \leqslant t]) = s^{n-1}t$.

Dans toute la suite de l'exercice, on suppose que $0 < s < 1$.

4. Soit $t \in [0, 1]$. Pour tout $k \in [\![1\,,\,n-1]\!]$, montrer que :

$$P(A_k \cap [Z_{n,s} \leqslant t]) = \begin{cases} (t-s)s^{k-1} & \text{si } t \geqslant s, \\ 0 & \text{si } t < s. \end{cases}$$

5. En déduire la valeur de la probabilité $P(Z_{n,s} \leqslant t)$ en fonction de $t \in [0, 1]$.
6. Montrer que $Z_{n,s}$ est une variable à densité, et en donner une densité.
7. En déduire que : $E(Z_{n,s}) = \dfrac{1}{2}(1 + s - s^n)$.
8. Déterminer le seuil s_n^* qui maximise la compétence moyenne du candidat embauché en fonction de n.
9. À l'aide de la fonction programmée en 1.(b), tracer sur un graphique l'évolution de la valeur de $E[Z_{n,s}]$ en fonction de s pour $n = 5, 10, 50$, et vérifier les conclusions de la question précédente dans ces cas.

Rapport des oraux relatifs aux exemples publiés

Exemple 1

L'inégalité de Markov est en général bien connue des candidats. Cependant, le jury attend un rappel explicite des hypothèses requises lorsqu'il la demande en question de cours. Peu de candidats le font spontanément.

Cet exercice mêlait algèbre linéaire et probabilités discrètes, via une expérience classique de type chaîne de Markov. Sa particularité était d'étudier ici un endomorphisme de polynômes plutôt que la matrice de transition associée directement. Même si les définitions en algèbre sont comprises par les candidats, ces derniers manquent nettement de recul vis à vis des méthodes disponibles au programme et se retrouvent souvent démunis dans plusieurs questions.
La question informatique a en général été bien faite par une majorité des candidats. Rappelons que lorsqu'un sujet demande de simuler une variable aléatoire, il serait bon que les candidats exécutent spontanément plusieurs fois de suite leur programme, pour montrer en effet que leur programme simule bien une quantité aléatoire.
Pour la question 2, peu de candidats ont su expliquer correctement la relation $U_{n+1} = MU_n$. Le jury attendait ici un appel à la formule des probabilités totales, et se contente d'une explication de quelques probabilités conditionnelles. Beaucoup d'étudiants pensent devoir raisonner par récurrence.
La question 3 a mis en évidence de nombreuses lacunes concernant les méthodes fondamentales en algèbre linéaire. La question (a) et la question (c) devraient être un objectif raisonnable pour tout candidat admissible. La question (b) demandait de l'aisance en calcul, mais le résultat étant donné il était souvent plus intéressant d'admettre la question pour répondre à la suivante (c'est d'ailleurs ce qu'en général l'examinateur proposait au candidat). Il est dommage cependant que beaucoup de candidats ne voient pas le lien entre la question (b) et la question (c) (malgré le « en déduire ») et veulent revenir à l'étude de la matrice $M - \lambda I_4$.
Les questions 4 à 6 ont été moins abordées par les candidats, faute de temps, à part pour deux ou trois excellents étudiants qui ont su résoudre brillamment l'intégralité de la planche.
Les prestations ont donc été plutôt décevantes , surtout sur la question 2, 3(a) et 3(c) qui étaient les principales attentes du jury.

Exemple 2

La question de cours a été parfois mal comprise par les candidats. Certains candidats confondent les propriétés de la fonction de répartition (par exemple dans le cas où elle est associée à une variable à densité) avec sa définition qui elle est valable pour toutes les variables aléatoires. Le tracé du tableau de variations doit ici être rapide, on n'attend aucune justification.

Cet exercice d'algèbre linéaire permettait d'une part d'évaluer les connaissances des candidats sur les nombres complexes ainsi que sur la diagonalisation. Même si les candidats sont moins à l'aise avec les nombres complexes que les réels, le sujet était assez guidé ce qui permettait à tout étudiant sérieux de bien avancer dans les questions.
En informatique, les candidats ont parfois eu du mal à écrire la fonction `decalage`. Certains ont utilisé les fonctions `pop` et `append`, d'autres ont créé une nouvelle liste et l'ont rempli successivement. Lorsqu'une question propose de compléter un script comme ici pour guider les candidats, ce n'est qu'une simple aide, les candidats sont libres bien entendu d'utiliser d'autres démarches qui leur sembleraient plus naturelles.

Exemple 3

La question de cours a souvent été bien traitée par les candidats. La principale erreur était de se placer dans un cadre fini alors qu'ici on avait plus généralement $X(\Omega) \subset \mathbb{N}$, ou bien d'écrire la formule de l'espérance sans se soucier de son existence.

L'exercice avait pour but d'étudier un projecteur orthogonal sans l'indiquer au préalable. On étudiait un endomorphisme de $\mathbb{R}^3$ classique, dont la matrice était symétrique diagonalisable. Puis on établissait progressivement des propriétés pour parvenir au fait que f était un projecteur orthogonal, en revenant à la définition au programme (question 6(b)).
Les questions 1, 2, 4 sont un objectif minimal pour tout candidat admissible et malheureusement ces questions classiques ont été plutôt mal traitées lors des interrogations, ce qui a donné lieu à des notes faibles pour certains candidats.
La question 1 typiquement est mal lue par les candidats, l'intitulé de la question les encourageant à faire le moins de calculs possibles, mais ils se lancent néanmoins sur des calculs de rangs de $A - \lambda I_3$ dans le cas général, ce qui leur a posé problème notamment avec la gestion du facteur $1/6$.
La diagonalisabilité étant déjà établie au préalable, il suffisait aux candidats de résoudre les équations $AX = 0$ et $AX = X$ pour répondre à la question. La question mettait donc en évidence le manque de recul des candidats face aux différentes méthodes possibles, et quant à la mauvaise adéquation de leur méthode face à la question posée.
La question 3 a été parfois mal comprise par les candidats, ces derniers se contentant souvent d'écrire `dot(M,M)==M` ce qui ne fonctionnait pas lorsque `M` était donnée par une matrice numpy, ou qui pouvait donner des problèmes avec les nombres flottants. L'examinateur cherchait à voir si le candidat savait éventuellement parcourir les deux matrices M et M^2 par une double boucle `for`. La question 5 néanmoins a été souvent bien faite lorsqu'elle a été abordée.
La question 4 n'a pas souvent été bien faite. L'argument le plus rapide était d'utiliser le fait que $E_1 \subset \mathrm{Im}(f)$ par définition, puis de conclure avec les dimensions étudiées dans la question 1 ou avec le théorème du rang. Les examinateurs ont donc souvent dû guider les candidats dans ce sens, la méthode n'étant pas immédiate pour la plupart.
Les questions 6 et 7 n'ont pas souvent été abordées, mais quand il restait du temps, cela pouvait donner lieu à une demande au candidat des définitions du cours de « base orthonormale », ou de « distance de t à E_1 ».
Remarquons que dans la question 6(a), le sujet proposait d'utiliser la fonction `ps` notamment pour les candidats faibles qui ne parviendraient pas à montrer l'orthogonalité dans le cas général, et qui auraient pu étudier l'orthogonalité au moins pour les familles génératrices de E_1 et E_0.

Exemple 4

Dans la question de cours, les candidats confondent souvent les propriétés attendues. Une grande majorité citent la croissance ou les valeurs des limites pour justifier que X admet une densité.

L'exercice d'analyse se proposait d'étudier deux suites implicites, solutions de l'équation $x^n - \ln(x) - n = 0$ ($n \geqslant 2$). Le sujet était long, cela peut arriver, mais dans ce cas les exigences du jury peuvent être limitées aux premières questions, qui permettent en général assez bien d'évaluer le niveau du candidat interrogé. Il ne faut donc pas à tout prix essayer d'avancer dans les questions rapidement sans aboutir, mais plutôt traiter avec soin les premières questions pour ensuite essayer éventuellement de décrocher des points restants.
Il est assez regrettable que cette année encore, dans tous les sujets où l'algorithme de dichotomie était présent (et toujours rappelé dans le sujet), les candidats ne parviennent que très rarement à le programmer correctement. Ici encore, cela n'a pas manqué et beaucoup de candidats ont eu énormément de mal à rédiger leur programme.

Exemple 5

La question de cours a bien été traitée dans la majorité des candidats, les seules maladresses portant sur l'appel correct aux conditions nécessaires d'application de la formule donnée (égalité ou inégalité).

L'exercice proposait une modélisation mettant en jeu des variables aléatoires discrètes. Ce genre de sujet est en général un peu plus long à la lecture. L'évaluation en tient donc compte. Ici, les examinateurs attendaient des candidats qu'ils aient pu aborder environ la moitié du sujet (5 premières questions).
La question 1 et la question 4 ont en général été bien faites, les candidats ayant en général bien compris l'expérience aléatoire à l'aide des premiers exemples X_1, X_2 et X_3.
Pour les candidats s'étant arrêtés à la question 4, le jury a tenté d'aborder avec eux la convergence de la suite $(p_n(r))$ (question 5), et de guider pour obtenir la minoration de $p(r)$ (question 6).

Exemple 6

Cette question de cours était nouvelle, dans le sens où elle n'était pas présente dans la liste de celles données en 2018. On attendait simplement ici l'utilisation de la bonne formule de cours pour dériver une composée de fonctions dérivables.

L'exercice proposé abordait les probabilités discrètes et les estimations par intervalle de confiance dans un cadre plutôt théorique. Il permettait de tester au moins sur ses premières questions les connaissances des candidats sur le cours des estimateurs ainsi que sur les formules de probabilités, en particulier les inégalités de concentration.
Les candidats connaissent en général les formules du cours (Bienaymé-Tchebychev, Markov) mais manquent d'autonomie dans leur utilisation au sein d'un exercice. Le calcul de l'espérance et de la variance de $\overline{X_n}$ reste difficile pour de nombreux candidats.
Le programme d'informatique a lui été plutôt bien réalisé.

Exemple 7

Certains candidats ont mal lu la question de cours, et mentionnent le lien entre éléments de l'ensemble d'arrivée et nombre d'antécédants. Dans ce cas, le jury insistait sur le fait qu'ici l'application était linéaire pour laisser une deuxième opportunité au candidat. Chaque examinateur a également demandé ici d'écrire la condition au tableau pour vérifier la bonne écriture mathématique du candidat, ce qui n'a malheureusement pas empêché de lire « $\mathrm{Ker}(f) = 0$ » ou « $\mathrm{Ker}(f) = \emptyset$ », ...

L'exercice est supposé classique et facile sur ses premières questions. Le jury se montre alors plus exigeant sur la capacité des candidats à traiter correctement et rapidement le début du sujet. Un candidat ne sachant pas répondre à la question 1(a) et/ou à la question 2 ne peut pas espérer avoir une note dépassant la moyenne (sur la partie Mathématiques Pratiques du moins).
La question 1(a) reste insurmontable pour une quantité trop grande de candidats. Les élèves manquant de rigueur ne parviennent pas à distinguer les différents cas à traiter. Nous rappelons qu'une étude préalable du support de la variable $-\ln(U)$ pourrait toujours aider les candidats les plus faibles, car cela limite ensuite les différents cas à étudier pour calculer la fonction de répartition de la variable aléatoire indiquée.

Exemple 8

La question de cours a été bien traitée par une majorité de candidats. Une réponse claire et rapide à l'oral peut suffire ici, certains candidats ayant perdu beaucoup de temps à écrire une définition longue et rigoureuse au tableau.

Ici encore, et comme c'est souvent le cas lors d'un exercice utilisant des variables aléatoires à densité, les premières questions étaient classiques et abordables. Les candidats ne parvenant pas à résoudre les questions 1 à 3 sont alors lourdement pénalisés, tant aux yeux du jury cela leur semble être un minimum requis pour tout candidat admissible après deux années de CPGE, tant les autres candidats sérieux parviennent à résoudre ces questions de manière immédiate et efficace en quelques minutes, ce qui creuse l'écart de manière drastique.
Les questions de type 3(b) sur un transfert de loi sont toujours le point qui peut bloquer les candidats faibles, cela devrait être cependant un exercice classique fondamental que tout candidat admissible devrait travailler lors de ses révisions du concours.
Le sujet étant cependant long, le jury se contente en général d'une résolution correcte (avec ou sans intervention du jury) des questions 1 à 4, la question 5 servant à départager les meilleurs candidats ayant montré une large autonomie dans le début du sujet.

Exemple 9

Comme pour l'exemple 6, cette question de cours était nouvelle, et a été fort mal résolue. Preuve qu'il ne faut pas que les candidats se contentent uniquement de « bachoter » les questions de cours tombées l'année précédente, mais doivent apprendre le cours dans leur intégralité.
Très peu de candidats ont pensé à dissocier le cas $\alpha = 1$, et même dans le cas général, la primitive proposée était souvent fausse.

Cet exercice proposait une modélisation conduisant à une étude ensuite de variables aléatoires à densité. Les notations requéraient une lecture attentive de l'énoncé, et la plupart des candidats ont bien compris comment définir la variable $Z_{n,s}$ en fonction du résultat des entretiens d'embauche.
Le programme d'informatique a alors été plutôt bien fait, et le jury est revenu en particulier sur la question 3 (pour écrire correctement l'événement $B \cap [Z_{n,s} \leqslant t]$ avec les variables aléatoires X_i), la question 5 (pour s'assurer que la formule des probabilités totales était comprise), la question 6 (sur la définition d'une variable aléatoire à densité).
Pour les candidats plutôt faibles, ou ceux disposant d'un peu de temps, la question 8 a permis de tester une rapide étude de fonction, mais l'utilisation des puissances pose encore des problèmes pour beaucoup d'étudiants.

Sujet	**Moyenne sur 10**	**Écart-type**	$\mathbf{0 \leqslant x \leqslant 2,5}$	$\mathbf{3 \leqslant x \leqslant 5}$	$\mathbf{5,5 \leqslant x \leqslant 7,5}$	$\mathbf{8 \leqslant x \leqslant 10}$
Ex 1	4,27	1,82	13%	61%	19%	6%
Ex 2	5,38	1,77	3%	50%	40%	7%
Ex 3	4,69	1,93	22%	31%	41%	6%
Ex 4	5,03	1,89	10%	55%	29%	6%
Ex 5	5,36	2,02	11%	25%	54%	11%
Ex 6	5,45	2,57	19%	23%	39%	19%
Ex 7	5,13	2,31	17%	30%	39%	13%
Ex 8	6,55	2,13	7%	18%	36%	39%
Ex 9	5,21	2,13	11%	36%	43%	11%

Résumé statistique (en 2019 exclusivement) des notes obtenues sur les exemples publiés