

Devoir surveillé n°10

Durée de l'épreuve : 4 heures
La calculatrice est interdite

Le devoir est composé d'un exercice et d'un problème de probabilité.
Lorsqu'une question est jugée, a priori, plus difficile, elle est précédée du symbole (*) voire (**).
La notation tiendra particulièrement compte de **la qualité de la rédaction, la précision des raisonnements et l'énoncé des formules utilisées.**

BON COURAGE

Exercice. ≈ 60 min.

Soient a et b des entiers naturels tels que $a \leq b$. On rappelle que $\llbracket a, b \rrbracket$ désigne l'ensemble des entiers naturels k tels que $a \leq k \leq b$.

Si S est un ensemble fini, on note $|S|$ son cardinal.

Si X est une variable à valeur dans une partie finie de $\mathbb{N}$, on note $\mathbf{E}(X)$ son espérance.

Soit n un entier naturel supérieur ou égal à 2 et soit ℓ un entier naturel non nul. Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes et de même loi uniforme sur l'ensemble $\llbracket 1, \ell \rrbracket$.

On note U_n le nombre de valeurs distinctes prises par les variables $X_1, \dots, X_n$: si $k_1, \dots, k_n$ sont les valeurs prises respectivement par $X_1, \dots, X_n$, alors U_n prend la valeur $|S|$ où $S = \{k_1, \dots, k_n\}$ pour tout $(k_1, \dots, k_n) \in \llbracket 1, \ell \rrbracket^n$.

Si S est une partie de $\llbracket 1, \ell \rrbracket$, on note $\{X_1, \dots, X_n\} = S$ la réunion des événements $(X_1, \dots, X_n) = (k_1, \dots, k_n)$ pour tout $(k_1, \dots, k_n) \in \llbracket 1, \ell \rrbracket^n$ tels que $S = \{k_1, \dots, k_n\}$.

- On suppose dans cette question seulement que $n = 2$ et $\ell \geq 2$.
 - Justifier que U_2 ne prend que les valeurs 1 et 2.
 - Calculer $\mathbf{P}(U_2 = 1)$ et $\mathbf{P}(U_2 = 2)$.
 - Calculer $\mathbf{E}(U_2)$.
- On se propose de simuler en Python la variable aléatoire U_n pour $n = 10$ dans le cas où $\ell = 25$.
 - Ecrire une fonction `simuU` qui renvoie une réalisation de U_{10} .
On pourra utiliser la fonction `random.randint`.
L'instruction `random.randint(1, 25)` fournit un nombre aléatoire dans $\llbracket 1, 25 \rrbracket$ uniformément.
 - Ecrire une fonction `espU` qui renvoie une approximation de l'espérance de U_{10} . Quel théorème utilisez-vous pour justifier que le résultat de cette fonction est une approximation de l'espérance de U_{10} ? Énoncez précisément ce théorème.
- Quel est l'ensemble des valeurs prises par U_n ?
- Soit i dans $\llbracket 1, n \rrbracket$. Soit S une partie de $\llbracket 1, \ell \rrbracket$. Quelle est la probabilité de l'événement $(X_i \in S)$ en fonction de $|S|$?
- Soit a dans $\llbracket 1, \ell \rrbracket$. Exprimer $\mathbf{P}(X_1 \neq a, \dots, X_{n-1} \neq a)$, la probabilité qu'aucune des variables $X_1, \dots, X_{n-1}$ ne prenne la valeur a , en fonction de n et ℓ .
- En déduire $\mathbf{P}(X_1 \neq X_n, \dots, X_{n-1} \neq X_n)$, la probabilité que la valeur prise par X_n soit différente de toutes les valeurs prises par les autres variables, en fonction de n et ℓ .
- Justifier

$$\mathbf{P}(X_1 \neq X_n, \dots, X_{n-1} \neq X_n) = \sum_{S \in \mathcal{P}_\ell} \mathbf{P}(\{X_1, \dots, X_{n-1}\} = S) \left(\frac{\ell - |S|}{\ell} \right)$$

où $\mathcal{P}_\ell$ désigne l'ensemble des parties non vides de $\llbracket 1, \ell \rrbracket$.

- En déduire dans le cas où $n \geq 3$:

$$\mathbf{E}(U_{n-1}) = \ell(1 - \mathbf{P}(X_1 \neq X_n, \dots, X_{n-1} \neq X_n))$$

9. Exprimer $\mathbf{E}(U_n)$ en fonction de n et ℓ .
10. Déterminer la limite de $\mathbf{E}(U_n)$ lorsque ℓ est fixé et $n \rightarrow +\infty$. Interprétez votre résultat.
11. Déterminer la limite de $\mathbf{E}(U_n)$ lorsque n est fixé et $\ell \rightarrow +\infty$. Interprétez votre résultat.
12. On s'intéresse aux possibles partages de dates d'anniversaire dans un groupe de n personnes. On suppose que les années sont toutes de 365 jours et que les dates d'anniversaire sont uniformément réparties sur chaque jour de l'année. On fait aussi l'hypothèse que les dates d'anniversaire de n personnes choisies au hasard sont indépendantes mutuellement.
Soit D_n le nombre de dates d'anniversaire d'un groupe de n personnes choisies au hasard.
 - (a) Exprimer en fonction de n le nombre moyen de dates d'anniversaire d'un groupe de n personnes, c'est à dire $\mathbf{E}(D_n)$.
 - (b) Quelle est la limite de ce nombre moyen lorsque n tend vers $+\infty$.

Problème. Lemme de Fekete et théorème de Erdős-Szekeres

Le but de ce problème est d'étudier quelques applications probabilistes du lemme de sous-additivité de Fekete et du théorème de Erdős-Szekeres.

Dans tout le problème, $(\Omega, \mathcal{A}, \mathbf{P})$ désigne un espace probabilisé. On note $\mathbf{P}(A)$ la probabilité d'un événement A et on note $\mathbf{E}(X)$ l'espérance (si elle existe) d'une variable aléatoire réelle discrète X définie sur $(\Omega, \mathcal{A}, \mathbf{P})$.

A. Préliminaires

≈ 20 min.

Les deux questions de cette partie sont indépendantes.

Soit n un entier naturel non nul.

1. Montrer que pour toute variable aléatoire X réelle à valeurs dans $\{1, \dots, n\}$ et pour tout $m \in \{1, \dots, n\}$,

$$\mathbf{E}(X) \leq m - 1 + n\mathbf{P}(X \geq m)$$

2. A l'aide d'une comparaison entre une somme et une intégrale, montrer que

$$n \ln(n) - n + 1 \leq \sum_{k=1}^n \ln(k)$$

En déduire l'inégalité

$$\left(\frac{n}{e}\right)^n \leq n!$$

B. Le lemme de sous-additivité de Fekete

≈ 40 min.

Soit $u = (u_n)_{n \in \mathbb{N}^*}$ une suite réelle bornée. Pour tout $n \in \mathbb{N}^*$, on note $U_n = \{u_k ; k \geq n\}$. On définit les suites $\underline{u} = (\underline{u}_n)_{n \in \mathbb{N}^*}$ et $\bar{u} = (\bar{u}_n)_{n \in \mathbb{N}^*}$ par les formules

$$\underline{u}_n = \inf(U_n) \quad \text{et} \quad \bar{u}_n = \sup(U_n)$$

1. Justifier que $\underline{u}$ et $\bar{u}$ sont bien définies. Montrer qu'elles sont monotones puis qu'elles convergent.

Pour toutes suites réelles $v = (v_n)_{n \in \mathbb{N}^*}$ et $w = (w_n)_{n \in \mathbb{N}^*}$,

on dit que v est *plus petite* que w , noté $v \preceq w$, si pour tout $n \in \mathbb{N}^*$, on a $v_n \leq w_n$.

De façon équivalente, on dit aussi que w est *plus grande* que v .

2. Montrer que $\bar{u}$ est la plus petite suite (au sens de $\preceq$) qui est décroissante et plus grande que u . Montrer de même que $\underline{u}$ est la plus grande suite (au sens de $\preceq$) qui est croissante et plus petite que u .

Dans toute la suite du problème, on appelle limite inférieure $\underline{\lim}$ et limite supérieure $\overline{\lim}$ les limites suivantes :

$$\underline{\lim}_{n \rightarrow +\infty} u_n = \lim_{n \rightarrow +\infty} \underline{u}_n \quad \text{et} \quad \overline{\lim}_{n \rightarrow +\infty} u_n = \lim_{n \rightarrow +\infty} \bar{u}_n$$

3. Si $v = (v_n)_{n \in \mathbb{N}^*}$ est une suite réelle bornée plus grande que u , comparer les limites de $\bar{u}$ et $\bar{v}$.

4. Montrer que $\bar{u}$ et $\underline{u}$ sont adjacentes si et seulement si u converge. En ce cas, que peut-on dire des limites des trois suites u , $\bar{u}$ et $\underline{u}$?

On dit qu'une suite réelle $u = (u_n)_{n \in \mathbb{N}^*}$ est *sous-additive* si pour tous i, j dans $\mathbb{N}^*$, on a $u_{i+j} \leq u_i + u_j$.

Dans le reste de cette partie, on ne suppose plus que la suite u est bornée, mais on suppose que u est positive et sous-additive.

5. Soient m et n deux entiers naturels non nuls tels que $m \geq 2n$. On note q le quotient et r le reste de la division euclidienne de m par n . Montrer que

$$u_m \leq (q-1)u_n + u_{n+r}$$

et en déduire l'inégalité

$$\frac{u_m}{m} \leq \frac{m-n-r}{m} \cdot \frac{u_n}{n} + \frac{\max\{u_n, u_{n+1}, \dots, u_{2n-1}\}}{m}$$

6. En déduire que la suite $\left(\frac{u_m}{m}\right)_{m \in \mathbb{N}^*}$ est bornée, puis que pour tout $n \in \mathbb{N}^*$,

$$\overline{\lim}_{m \rightarrow +\infty} \frac{u_m}{m} \leq \frac{u_n}{n}$$

7. En conclure que la suite $\left(\frac{u_n}{n}\right)_{n \in \mathbb{N}^*}$ converge.

C. Une application probabiliste

≈ 30 min.

Soit x un nombre réel et $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles mutuellement indépendantes et de même loi. Pour tout $n \in \mathbb{N}^*$, on note Y_n la variable aléatoire réelle définie par

$$Y_n = \frac{1}{n} \sum_{k=1}^n X_k$$

- Montrer que si $\mathbf{P}(X_1 < x) = 1$, alors pour tout $n \in \mathbb{N}^*$, $\mathbf{P}(Y_n < x) = 1$ et que si $\mathbf{P}(X_1 \geq x) > 0$ alors pour tout $n \in \mathbb{N}^*$, $\mathbf{P}(Y_n \geq x) > 0$.
- Soient m et n deux entiers naturels non nuls. Montrer l'inclusion d'événements suivante :

$$\left(\{Y_m \geq x\} \cap \left\{ \frac{1}{n} \sum_{k=m+1}^{m+n} X_k \geq x \right\} \right) \subset \{Y_{m+n} \geq x\}$$

et en déduire l'inégalité

$$\mathbf{P}(Y_{m+n} \geq x) \geq \mathbf{P}(Y_m \geq x) \mathbf{P}(Y_n \geq x)$$

- Démontrer la convergence de la suite $\left((\mathbf{P}(Y_n \geq x))^{\frac{1}{n}} \right)_{n \in \mathbb{N}^*}$

D. Le théorème de Erdős-Szekeres

≈ 30 min.

Si r est un entier naturel non nul, on note $\ell = (\ell_1, \dots, \ell_r)$ une liste de nombres réels de longueur r ; cette liste est *croissante* si $\ell_1 \leq \ell_2 \leq \dots \leq \ell_r$, *décroissante* si $\ell_1 \geq \ell_2 \geq \dots \geq \ell_r$. Une liste ℓ' de longueur $p \in \{1, \dots, r\}$ est *extraite* de ℓ s'il existe p indices strictement croissants $i_1 < i_2 < \dots < i_p$ dans $\{1, \dots, r\}$ tels que $\ell' = (\ell_{i_1}, \dots, \ell_{i_p})$.

Soit p et q deux entiers naturels non nuls et $a = (a_1, a_2, \dots, a_{pq+1})$ une liste de longueur $pq+1$ de nombres réels deux à deux distincts qui représentent les valeurs de $pq+1$ jetons numérotés $1, 2, \dots, pq+1$.

On range successivement les jetons en piles de gauche à droite par le procédé suivant :

- le jeton numéro 1 de valeur a_1 débute la première pile ;
- si $a_2 > a_1$, alors on pose le jeton numéro 2 de valeur a_2 sur le jeton numéro 1 ;
sinon on crée une nouvelle pile avec ce jeton numéro 2, située à droite de la première pile ;
- lors des étapes suivantes, disposant du jeton numéro k de valeur a_k , on le dépose sur la première pile en partant de la gauche telle que a_k est supérieur à la valeur du jeton au sommet de la pile,
si une telle pile existe ;
sinon on crée une nouvelle pile avec ce jeton, située à droite des précédentes.

En suivant ce procédé avec tous les jetons, on obtient plusieurs piles de jetons, chaque pile ayant des valeurs rangées dans l'ordre croissant du bas vers le haut.

Par exemple, avec la liste

$$a = (1, 4, 2, 3, 7, 6, 5, 9, 10, 8)$$

dans cet ordre, on obtient de gauche à droite les trois piles suivantes

$$\begin{array}{ccc} & 10 & \\ & 9 & 8 \\ & 7 & 6 \\ & 4 & 3 \\ \hline & 1 & 2 & 5 \end{array}$$

1. A l'aide d'un raisonnement par récurrence sur le nombre s de piles, montrer qu'à l'issue de ce processus, pour tout jeton de valeur z de la dernière pile, il existe une liste $b = (b_1, \dots, b_s)$ de réels extraite de la liste a vérifiant :

- b est décroissante de longueur s ;
- pour tout $i \in \{1, \dots, s\}$, le jeton numéro i de valeur b_i est dans la i -ième pile en partant de la gauche ;
- $b_s = z$.

Par exemple, avec la liste $a = (1, 4, 2, 3, 7, 6, 5, 9, 10, 8)$ on a extrait une liste $b = (7, 6, 5)$.

2. En déduire que la liste a admet au moins une liste extraite croissante de longueur $p + 1$ ou une liste extraite décroissante de longueur $q + 1$.

E. Comportement asymptotique d'une suite aléatoire ≈ 60 min.

Soit n un entier naturel supérieur ou égal à 2. On note S_n l'ensemble des permutations de $\{1, 2, \dots, n\}$. Chaque élément σ de S_n est noté par la liste de ses n images $(\sigma(1), \sigma(2), \dots, \sigma(n))$.

Soit B une variable aléatoire à valeurs dans S_n de loi uniforme, c'est à dire que pour tout $\sigma \in S_n$, on a $\mathbf{P}(B = \sigma) = 1/\text{Card}(S_n)$. On définit la variable aléatoire A à valeurs dans S_n en posant pour tout $\omega \in \Omega$

$$A(\omega) = (B(\omega)(1), \dots, B(\omega)(n))$$

On note également, pour tout $k \in \{1, \dots, n\}$, $A_k(\omega) = B(\omega)(k)$. Enfin, on considère les variables aléatoires C_n et D_n définies par :

- C_n est la longueur de la plus longue liste croissante extraite de A ;
- D_n est la longueur de la plus longue liste décroissante extraite de A .

1. Les variables aléatoires réelles $A_1, A_2, \dots, A_n$ sont-elles mutuellement indépendantes ?
2. Soit $k \in \{1, \dots, n\}$ et $s = (s_1, \dots, s_k)$ une liste strictement croissante de longueur k d'éléments de $\{1, \dots, n\}$. On note A^s l'événement : « la liste $(A_{s_1}, \dots, A_{s_k})$ est croissante ». Montrer que $\mathbf{P}(A^s) = \frac{1}{k!}$.

3. Montrer que C_n et D_n ont le même loi.

Démontrer alors, à l'aide du résultat de la question D.2., que

$$\mathbf{E}(C_n) \geq \frac{\sqrt{n}}{2}$$

4. Démontrer que pour tout $k \in \{1, \dots, n\}$,

$$\mathbf{P}(C_n \geq k) \leq \frac{\binom{n}{k}}{k!}$$

5. Soit n un entier naturel non nul et α un réel strictement supérieur à 1. Justifier qu'il existe un entier naturel non nul k tel que $k - 1 < \alpha e \sqrt{n} \leq k$. A l'aide du résultat de la question A.2., déduire de la question précédente que

$$\mathbf{P}(C_n \geq \alpha e \sqrt{n}) \leq \left(\frac{1}{\alpha}\right)^{2\alpha e \sqrt{n}}$$

6. En déduire qu'il existe une suite $(\varepsilon_n)_{n \in \mathbb{N}^*}$ tendant vers 0 telle que, pour tout $n \in \mathbb{N}^*$,

$$\frac{\mathbf{E}(C_n)}{\sqrt{n}} \leq (1 + n^{-1/4})e + \varepsilon_n$$

En conclure que $\overline{\lim}_{n \rightarrow +\infty} \frac{\mathbf{E}(C_n)}{\sqrt{n}}$ existe et que $\overline{\lim}_{n \rightarrow +\infty} \frac{\mathbf{E}(C_n)}{\sqrt{n}} \leq e$.