PROBABILITÉS

Les trois parties de ce devoir sont indépendantes. Trois niveaux de difficulté/longueur :

— Piste verte : partie **1** et partie **2** questions **1)2)a)b)**.

— Piste bleue-rouge : parties **1** et **2**.

— Piste noire : parties **2** et **3** (éventuellement tout le devoir).

1 Pile ou face

Soit $p \in]0,1[$. On s'intéresse à une famille $(X_1,\ldots,X_n)$ de variables aléatoires indépendantes de loi de Bernoulli $\mathscr{B}(p)$.

1) On note T l'indice du premier 1 dans $(X_1,\ldots,X_n)$ s'il existe et on pose $T=+\infty$ sinon. Déterminer la loi de T.

2) Soient $r \in [\![1,n]\!]$ et $s \in [\![0,r]\!]$, puis $\varepsilon_0,\ldots,\varepsilon_{r-1}$ des entiers dont s valent 1 et $r-s$ valent 0. On note E l'événement « La séquence ε_0-ε_1-…-ε_{r-1} n'apparaît jamais dans cet ordre dans la famille $(X_1,\ldots,X_n)$ ».

a) Calculer pour tout $k \in [\![1,n-r+1]\!]$ la probabilité de l'événement « La séquence ε_0-ε_1-…-ε_{r-1} apparaît dans cet ordre dans $(X_1,\ldots,X_n)$ à partir de la position k ».

b) En déduire que $P(E) \leqslant \left(1-p^s(1-p)^{r-s}\right)^{\left\lfloor \frac{n}{r}\right\rfloor}$. Calculer la limite de $P(E)$ lorsque n tend vers $+\infty$ et interpréter le résultat.

2 Produit et somme de deux éléments de $\mathbb{U}_n$

Soit $n \in \mathbb{N}^*$. On se donne deux variables aléatoires X_n et Y_n indépendantes à valeurs dans $\mathbb{U}_n$.

1) Montrer que si X_n suit la loi uniforme sur $\mathbb{U}_n$, alors il en va de même de X_nY_n.

2) Dans cette question, X_n et Y_n suivent toutes les deux la loi uniforme sur $\mathbb{U}_n$. On note A_n et B_n les entiers uniques de $[\![0,n-1]\!]$ pour lesquels $X_n = \mathrm{e}^{\frac{2iA_n\pi}{n}}$ et $Y_n = \mathrm{e}^{\frac{2iB_n\pi}{n}}$.

a) Montrer que A_n et B_n sont indépendantes et déterminer leur loi.

b) Montrer que pour tout $k \in [\![1,n-1]\!]$: $\quad P\big(|A_n-B_n|=k\big) = \dfrac{2(n-k)}{n^2}$ et calculer $P\big(|A_n-B_n|=0\big)$.

c) En déduire que pour tout $\lambda \in [0,1]$: $\quad \lim\limits_{n\to+\infty} P\big(|A_n-B_n| \geqslant \lambda n\big) = \lim\limits_{n\to+\infty} P\big(|A_n-B_n| > \lambda n\big) = (1-\lambda)^2$.

d) Montrer que $|X_n+Y_n| = 2\left|\sin\left(\dfrac{\pi}{2} - \dfrac{|A_n-B_n|\,\pi}{n}\right)\right|$, puis soigneusement que pour tout $x \in [0,1]$:

$$P\big(|X_n+Y_n| \leqslant 2x\big) = P\left(-\frac{n}{\pi}\operatorname{Arcsin} x \leqslant \frac{n}{2} - |A_n-B_n| \leqslant \frac{n}{\pi}\operatorname{Arcsin} x\right).$$

e) En déduire que pour tout $x \in [0,1]$: $\quad \lim\limits_{n\to+\infty} P\big(|X_n+Y_n| \leqslant 2x\big) = \dfrac{2}{\pi}\operatorname{Arcsin} x$

3 Combien de tirages pour les atteindre tous ?

Soit $p \in \mathbb{N}^*$. On effectue une série de tirages successifs avec remise dans l'ensemble $[\![1,p]\!]$. Pour un nombre de tirages assez grand, il paraît vraisemblable que les éléments de $[\![1,p]\!]$ auront tous été tirés au moins une fois au bout d'un moment. Ce problème se propose de justifier cette intuition dans le cas où l'entier p est lui-même assez grand.

Pour modéliser les tirages en question, on se donne des variables aléatoires indépendantes $X_1,\ldots,X_{2^p}$ de loi uniforme sur $[\![1,p]\!]$. Comme on le comprendra peu à peu, le nombre 2^p de variables introduites est essentiellement arbitraire, on se donne juste « beaucoup » de variables.

On note T_p le plus petit entier $n \in [\![1,2^p]\!]$ pour lequel $\big\{X_1,\ldots,X_n\big\} = [\![1,p]\!]$ si jamais un tel entier existe, et sinon on pose $T_p = +\infty$. On définit ainsi une variable aléatoire T_p à valeurs dans $[\![1,2^p]\!] \cup \big\{+\infty\big\}$. Pour tous $t \geqslant 0$ et $i \in [\![1,p]\!]$, on pose enfin $E_{t,i} = \displaystyle\bigcap_{\substack{1\leqslant n\leqslant 2^p \\ n\leqslant t}} \big\{X_n \neq i\big\}$.

On observera en passant que chacune des variables $X_1,\ldots,X_{2^p}$ et chacun des événements $E_{t,i}$ dépend de la donnée initiale de l'entier p même si, par souci de légèreté, les notations n'en témoignent pas.

1) a) Exprimer pour tout $t>0$ l'événement $\left\{T_p>t\right\}$ en fonction des événements $E_{t,i}$, i décrivant $[\![1,p]\!]$.

b) Montrer que pour tous $t\in[0,2^p]$ et $i\in[\![1,p]\!]$: $\quad P(E_{t,i})\leqslant \mathrm{e}\times \mathrm{e}^{-\frac{t}{p}}$.

c) En déduire $\displaystyle\lim_{p\to+\infty} P\big(T_p>\alpha p\ln p\big)$ pour tout $\alpha>1$.

En résumé, T_p a peu de chances de dépasser nettement $p\ln p$ lorsque p est grand. La suite du problème donne un sens très précis à l'affirmation selon laquelle T_p est proche de $p\ln p$ avec grande probabilité pour p assez grand.

On rappelle que pour tout $x\in\mathbb{R}$: $\quad \displaystyle \mathrm{e}^x=\sum_{k=0}^{+\infty}\frac{x^k}{k!}$, $\quad$ i.e. que : $\quad \displaystyle \mathrm{e}^x-\sum_{k=0}^{n}\frac{x^k}{k!}\underset{n\to+\infty}{\longrightarrow}0$.

2) Déterminer pour tous $k\in\mathbb{N}$ et $x\in\mathbb{R}$ un équivalent simple lorsque p tend vers $+\infty$ de $\displaystyle\left(1-\frac{k}{p}\right)^{\lfloor p\ln p+px\rfloor}$

Soit $x\in\mathbb{R}$ fixé une fois pour toutes.

3) Dans cette question, on suppose p assez grand pour que l'entier $\varphi(p)=\lfloor p\ln p+px\rfloor$ appartienne à $[\![1,2^p]\!]$ et on pose pour tout $k\in[\![1,p]\!]$: $\quad \displaystyle\sigma_{p,k}=\sum_{1\leqslant i_1<\ldots<i_k\leqslant p} P\big(E_{\varphi(p),i_1}\cap\ldots\cap E_{\varphi(p),i_k}\big)$.

a) Exprimer $P\big(T_p>p\ln p+px\big)$ en fonction de $\sigma_{p,1},\ldots,\sigma_{p,p}$.

b) Montrer que pour tout $k\in[\![1,p]\!]$: $\quad \displaystyle\sigma_{p,k}\leqslant\frac{p^k}{k!}\,\mathrm{e}^{-\frac{k\varphi(p)}{p}}$.

c) En déduire que pour tout $r\in[\![1,p]\!]$: $\quad \displaystyle\left|\sum_{k=r+1}^{p}(-1)^{k+1}\sigma_{k,p}\right|\leqslant \mathrm{e}\sum_{k=r+1}^{+\infty}\frac{\mathrm{e}^{-kx}}{k!}$.

4) Que vaut $\displaystyle\lim_{p\to+\infty}\sigma_{p,k}$ pour tout $k\in\mathbb{N}$? En déduire proprement, en revenant à la définition de la limite, que :

$$\lim_{p\to+\infty}P\big(T_p>p\ln p+px\big)=1-\mathrm{e}^{-\mathrm{e}^{-x}}.$$