ENS LYON | Informatique

L3 Probabilités

Devoir maison numéro 1
Corection.

Exercice 1

Remarque. *Certains d'entre vous ont préféré snober l'indication et travailler avec la formule de Stirling, ce qui est plus fastidieux.*

Soit $k \in \mathbf{N}$. On a $\mathbf{P}(T_N > k+1 | T_N > k) = \mathbf{P}(X_{k+1} \notin \{X_1, \ldots, X_k\} | \operatorname{card}\{X_1, \ldots, X_k\} = k) = \frac{N-k}{N}$. On en déduit par récurrence sur k la formule $\mathbf{P}(T_N > k) = \prod_{i=1}^{k-1} \left(1 - \frac{i}{N}\right)$. En utilisant l'indication, on a donc si $k \leqslant N/2$,

$$\mathbf{P}(T_n > k) \leqslant \exp\left(-\sum_{i=1}^{k-1} \frac{i}{N}\right) = \exp\left(-\frac{k(k-1)}{2N}\right),$$

$$\mathbf{P}(T_n > k) \geqslant \exp\left(-\sum_{i=1}^{k-1} \frac{i}{N} - \frac{i^2}{N^2}\right) \geqslant \exp\left(-\frac{k(k-1)}{2N} - \frac{k^3}{N^2}\right).$$

Pour $k = \lfloor t\sqrt{N} \rfloor$ avec $t > 0$ fixé (qui vérifie $k \leqslant N/2$ pour N assez grand), les deux encadrements convergent vers $\exp(-t^2/2)$. On en déduit

$$\lim_{N\to\infty} \mathbf{P}\left(\frac{T_N}{\sqrt{N}} \leqslant t\right) = 1 - \lim_{N\to\infty} \mathbf{P}(T_N > \lfloor t\sqrt{N} \rfloor) = \begin{cases} 1 - \exp(-t^2/2) & \text{si } t > 0 \\ 0 & \text{sinon.} \end{cases}$$

On en déduit que le α cherché est $\sqrt{2\log(100/99)} \approx 0.14$.

Exercice 2

Remarque. *Une difficulté de l'exercice était de réussir à majorer astucieusement la borne obtenue par l'inégalité de Chernoff pour éviter des calculs inutilement lourds. Quelques uns d'entre vous ont proposé l'approche élégante présentée ici, qui n'utilise même pas l'hypothèse $\alpha \geqslant \frac{1}{100}$!*

On choisit les entiers $k_1, \cdots, k_n$ i.i.d. de loi uniforme dans $\{0, \ldots, n-1\}$. Alors la variable aléatoire $B_{i,j}$ suit une loi de Bernoulli de paramètre $c_i := \frac{1}{n}(A_{1,1} + \cdots + A_{1,n})$. Fixons un indice de colonne $j \in \{1, \ldots, n\}$. Le nombre de 1 dans la $j$ème colonne de B est $X_j := B_{1,j} + \cdots + B_{n,j}$. Pour tout j fixé, les variables aléatoires $(B_{1,j}, \ldots, B_{n,j})$ sont indépendantes et on peut donc appliquer l'inégalité de Chernoff (version II) pour conclure que pour tout $\varepsilon > 0$, en posant $\mu = \mathbf{E}[X_j]$

$$\mathbf{P}(X_j \geqslant (1+\varepsilon)\mu) \leqslant \exp\left(-\mu\frac{\varepsilon^2}{2+\varepsilon}\right).$$

On a $\mu = \alpha n$, et on choisit $\varepsilon = C\sqrt{n\log n}/\alpha n$ pour une constante C à déterminer. On a donc

$$\mathbf{P}(X_j \geqslant \alpha n + C\sqrt{n\log n}) \leqslant \exp\left(-\frac{\alpha n \cdot C^2 n \log n/(\alpha n)^2}{2+\varepsilon}\right) = \exp\left(-\frac{C^2 \log n}{\alpha(2+\varepsilon)}\right) < \exp\left(-\frac{C^2\log n}{2+C}\right),$$

où pour la dernière inégalité on a utilisé $\alpha \leqslant 1$ et $\sqrt{n\log n} < n$. Par la borne de l'union, on a donc

$$\mathbf{P}(\sup_{1\leqslant k\leqslant n} X_j \geqslant \alpha n + C\sqrt{n\log n}) < n\exp\left(-\frac{C^2\log n}{2+C}\right) = n^{1-\frac{C^2}{2+C}}.$$

Cette probabilité est strictement inférieure à 1 pour $C = 2$. Ainsi il existe une matrice $B \in \mathcal{A}$ dont chaque colonne contient au plus $\alpha n + 2\sqrt{n\log n}$ coefficients égaux à 1.

Exercice 3

Remarque. *J'ai lu beaucoup de raisonnements incorrects dans la question 2. Notamment plusieurs d'entre vous ont affirmé que si k est le cardinal maximal d'un ensemble de vecteurs deux à deux à distance au moins $n/4$ dans H_n, alors $\mathbf{P}(d(X,Y) \geqslant n/4) \leqslant k/2^n$. C'est faux (considérer $n = 4$ si vous n'êtes pas convaincus).*

1. C'est une application facile de l'inégalité de Chernoff.
2. Si x, $y \in H_n$, les vecteurs $\frac{x}{\sqrt{n}}$ et $\frac{y}{\sqrt{n}}$ sont de norme 1. De plus $\left\| \frac{x}{\sqrt{n}} - \frac{y}{\sqrt{n}} \right\|^2 = \frac{4}{n} d(x,y)$. Il suffit donc de justifier l'énoncé suivant : pour tout n, il existe un sous-ensemble $A \subset H_n$ vérifiant $\operatorname{card} A \geqslant \exp(n/16)$ et $d(x,y) \geqslant n/4$ pour tous $x \neq y$ dans A.
   Soit $N = \lceil \exp(n/16) \rceil$, et $(X_1, \ldots, X_N)$ i.i.d. de loi uniforme sur H_n. Par la question précédente pour $\delta = n/4$ et la borne de l'union,
   $$\mathbf{P}(\exists\, 1 \leqslant i < j \leqslant N \,:\, d(X_i, X_j) < n/4) \leqslant \binom{N}{2} \exp(-n^2/8) \leqslant (N-1)^2 \exp(-n/8) < 1.$$
   Ainsi, avec probabilité strictement positive, l'ensemble $A = \{X_1, \ldots, X_N\}$ est formé de points (nécéessairement distincts) deux à deux à distance au moins $n/4$.

Exercice 4

Remarque. *Vous avez quasiment tous proposé la majoration facile $p \leqslant 1/2$ et la minoration $p \geqslant 1/3$ en considérant la stratégie qui consiste pour chaque joueur à choisir l'indice du premier 1 (ou, plus rarement, du premier 0 ; ça marche aussi !). Je suis un peu déçu que personne n'ait écrit d'algorithme pour chercher de meilleures stratégies. J'en ai appris plus sur ce problème sur Math Overflow, cf.* `https://mathoverflow.net/questions/326669`

On démontre les bornes $\frac{7}{20} \leqslant p \leqslant \frac{3}{8}$, soit $p \in [0.35, 0.375]$.

Borne inférieure. On considère la stratégie suivante. On définit une fonction $f : \{0,1\}^3 \to \{0,1,2\}$ par $f(110) = f(100) = 0$, $f(001) = f(101) = 1$, $f(010) = f(011) = 2$, et par une valeur quelconque pour $f(000)$ et $f(111)$ (peu importe le choix). Si Alice et Bob choisissent tous les deux cette stratégie en ne considérant que leurs 3 premiers bits, on vérifie en énumérant les 64 cas qu'ils gagnent le jeu avec probabilité $\frac{22}{64}$ — c'est déjà mieux que $\frac{1}{3}$!

On peut ensuite utiliser l'astuce suivante : soit k_A minimal telle que A_{3k_A}, A_{3k_A+1} et A_{3k_A+2} ne soient ni tous égaux à 0 ni tous égaux à 1. Alice choisit $a = 3k_A + f(A_{3k_A}, A_{3k_A+1}, A_{3k_A+2})$. Bob applique la même stratégie pour définir k_B et b. Conditionnellement à $k_A = k_B$ (événement de probabilité $\frac{3}{5}$, comme on le voit en sommant une série géométrique), ils gagnent dans 15 des 36 cas, donc avec probabilité $\frac{5}{12}$. Conditionnellement à $k_A < k_B$ ou $k_A > k_B$, ils gagnent avec probabilité $\frac{1}{4}$. La probabilité de victoire est donc $\frac{5}{12}\frac{3}{5} + \frac{1}{4}\frac{2}{5} = \frac{7}{20}$

Borne supérieure. Il est commode de supposer que les suites (A_n) et (B_n) sont à valeurs dans $\{-1,1\}$ au lieu de $\{0,1\}$, ce qui ne change rien au problème. Notons comme dans l'énoncé $a = f_A(A_n)$ et $b = f_B(B_n)$. Puisque les événements $\{A_b = 1\}$ et $\{B_a = 1\}$ ont probabilité $\frac{1}{2}$, on a

$$\mathbf{P}(A_b = B_a = 1) = \frac{1}{2} - \mathbf{P}(A_b = 1, B_a = -1) = \mathbf{P}(A_b = B_a = -1) = \frac{1}{2} - \mathbf{P}(A_b = -1, B_a = 1),$$

et donc

$$\mathbf{E}[A_b B_a] = \mathbf{P}(A_b = B_a) - \mathbf{P}(A_b = -B_a) = 4\mathbf{P}(A_b = B_a = 1) - 1.$$

Pour montrer $p \leqslant \frac{3}{8}$, il suffit donc de montrer que $\mathbf{E}[A_b B_a] \leqslant \frac{1}{2}$. L'astuce est de considérer $a' = f_A(-A_n)$, l'indice choisi par Alice lorsqu'elle voit la suite opposée à (A_n). Puisque (A_n, B_n) et $(-A_n, B_n)$ ont même loi, on a $\mathbf{E}[A_b B_a] = \mathbf{E}[-A_b B_{a'}]$. Ainsi

$$2\mathbf{E}[A_b B_a] = \mathbf{E}[A_b(B_a - B_{a'})] \leqslant \mathbf{E}|B_a - B_{a'}| \leqslant 1,$$

où la dernière inégalité vient du fait que conditionnellement à (A_n), $|B_a - B_{a'}|$ est soit nul (si $a = a'$) soit de même loi que $|B_1 - B_2|$ (qui est d'espérance 1).