ENS LYON — Probabilités, L3 Informatique Fondamentale
L3 Probabilités

Examen partiel
19 novembre 2025

On pourra utiliser les variantes suivantes des inégalités de Chernoff.

(I) Si X est la somme de n variables aléatoires i.i.d. de loi de Bernoulli de paramètre $1/2$, alors pour tout $a > 0$

$$\mathbf{P}(X \geqslant \frac{n}{2} + a) \leqslant \exp(-2a^2/n)$$

$$\mathbf{P}(X \leqslant \frac{n}{2} - a) \leqslant \exp(-2a^2/n)$$

(II) Si X est une somme de variables aléatoires indépendantes ayant chacune une loi de Bernoulli, alors en posant $\mu = \mathbf{E}[X]$, pour tout $\delta > 0$,

$$\mathbf{P}(X \leqslant (1-\delta)\mu) \leqslant \exp\left(-\delta^2\mu/2\right)$$

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

Exercice 1 Dans l'ascenseur

Un groupe de n personnes rentre dans un ascenseur situé au rez-de-chaussée d'un immeuble à k étages. Chaque personne appuie sur un des boutons 1 à k aléatoirement, indépendamment, selon la loi uniforme sur $\{1, \dots, k\}$. Quelle est l'espérance du nombre d'arrêts que fera l'ascenseur ?

Linéarité de l'espérance ! Soit A_i l'événement « une des personnes a appuyé sur le bouton i ». Par indépendance $\mathbf{P}(\overline{A_i}) = (1-1/k)^n$ et donc $\mathbf{P}(\overline{A_i}) = 1-(1-1/k)^n$. Le nombre X d'arrêts est $X = \sum_{i=1}^k \mathbf{1}_{A_i}$ a pour espérance

$$\mathbf{E}[X] = \sum_{i=1}^n \mathbf{P}(A_i) = n\left(1 - (1-1/k)^n\right)$$

Exercice 2 Records d'une permutation aléatoire

Pour un entier n, soit σ_n une permutation aléatoire de $\{1, \dots, n\}$, choisie uniformément parmi les $n!$ permutations possibles. On dit que $i \in \{1, \dots, n\}$ est un record de σ_n si on a $\sigma_n(j) < \sigma_n(i)$ pour tout $j < i$. On note A_i l'événement « i est un record de σ_n ».

1. Pour $i \in \{1, \dots, n\}$, calculer $\mathbf{P}(A_i)$.
2. Montrer que les événements $(A_i)_{1 \leqslant i \leqslant n}$ sont indépendants.
3. On note $R(\sigma_n)$ le nombre de records de σ_n. Montrer que

$$\mathbf{P}(R(\sigma_n) \geqslant 3\log n + 3) \leqslant \frac{1}{n}$$

On pourra utiliser l'inégalité $\log n \leqslant 1 + \frac{1}{2} + \cdots + \frac{1}{n} \leqslant \log n + 1$

Pour traiter les questions 1 et 2, on peut introduire $r_i \in \{1, \dots, i\}$ comme étant le rang de $\sigma_n(i)$ parmi l'ensemble $(\sigma_n(1), \dots, \sigma_n(i))$; ainsi $r_i = 1$ si et seulement si r_i est un record de σ_n. L'application $\sigma_n \mapsto (r_1, \dots, r_n)$ est une bijection de $\mathfrak{S}_n$ dans $E = \{1\} \times \{1,2\} \times \cdots \times \{1, \dots, n\}$. En particulier, la variable aléatoire $(r_1, \dots, r_n)$ suit la loi uniforme sur E. Cela revient à dire que les variables aléatoires r_i sont indépendantes et que r_i suit la loi uniforme sur $\{1, \dots, i\}$. Ainsi, les événements $A_i = \{r_i = 1\}$ sont indépendants et vérifient $\mathbf{P}(A_i) = 1/i$.

Pour la question 3, soit $X_i = \mathbf{1}_{A_i}$. Alors $R(\sigma_n) = X_1 + \cdots + X_n$ est une somme de v.a. indépendantes ayant chacune une loi de Bernoulli. On a

$$\mathbf{E}[R(\sigma_n)] = \sum_{i=1}^n \mathbf{E}[X_i] = \sum_{i=1}^n \frac{1}{i} =: h_n$$

On a par Chernoff II

$$\mathbf{P}(R(\sigma_n) \geqslant (1+\delta)(1+\log(n))\mathbf{P}(R(\sigma_n) \geqslant (1+\delta)h_n) \leqslant \exp\left(-\frac{\delta^2}{2+\delta}h_n\right) \leqslant \exp\left(-\frac{\delta^2}{2+\delta}\log n\right)$$

et on choisit $\delta = 2$.

Exercice 3 Analyse plus fine de Randomized Quicksort

On étudie plus en détail l'algorithme Randomized Quicksort vu en cours pour montrer que le temps d'exécution pour trier un ensemble S de cardinal n est $O(n \log n)$ avec grande probabilité.

Quand un pivot est choisi aléatoirement au sein d'un ensemble $S' \subset S$, on dit que le choix est *chanceux* si le pivot p partage $S' \setminus \{p\}$ en deux sous-ensembles de cardinal $\leqslant \frac{2}{3}|S'|$, et *malchanceux* sinon.

1. Montrer que la probabilité qu'un choix de pivot soit chanceux est $\geqslant 2/3$. Cet énoncé est faux, comme vous êtes beaucoup à l'avoir remarqué. La probabilité qu'un choix soit chanceux est $\geqslant 1/3$ et cette valeur ne peut pas être amélioré.
   Si $\frac{1}{3}|Y| \leqslant \mathrm{rg}(p) \leqslant \frac{2}{3}|Y|$ alors le choix de p est chanceux, donc la probabilité qu'un choix de pivot soit chanceux est $\geqslant \frac{1}{3}$.
2. Dans une suite de N choix de pivots indépendants, soit X_N le nombre de choix chanceux. Montrer avec une inégalité de Chernoff que

$$\mathbf{P}(X_N \leqslant N/2) \leqslant \exp(-cN)$$

   pour une constante $c > 0$ à déterminer. Il faut changer l'énoncé en (par exemple) $\mathbf{P}(X_N \leqslant N/4) \leqslant \exp(-cN)$. Je vous présente mes excuses pour cette erreur et félicite les étudiants qui ont réussi à corriger l'énoncé pour traiter l'exercice.
   On utilise Chernoff II avec $X_i \sim \mathcal{B}(p_i)$, $X_N = \sum_i X_i$. Par la question 1, $p_i \geqslant \frac{1}{3}$ donc $\mathbb{E}[X_N] \geqslant \frac{N}{3}$. Aussi, $\mathbf{P}\left(X_N \leqslant \frac{N}{3}(1-\epsilon)\right) \leqslant \mathbf{P}(X_N \leqslant \mathbb{E}[X_N](1-\epsilon))$.

$$\mathbf{P}\left(X_N \leqslant (1-\epsilon)\mathbb{E}[X_N]\right) \leqslant \exp\left(-\frac{\epsilon^2}{2}\mathbb{E}[X_N]\right) \leqslant \exp\left(-\epsilon^2\frac{N}{6}\right) \tag{1}$$

   Pour retrouver $\mathbf{P}(X_N \leqslant \frac{N}{4})$, on pose $\epsilon = \frac{1}{4}$ et on trouve $c = \frac{1}{6\times 16} = \frac{1}{96}$.
3. Soit $x \in S$. Montrer qu'au cours de toute exécution de l'algorithme Quicksort (aléatoire ou déterministe), parmi les choix de pivots effectués dans un ensemble contenant x, au plus $C\log(n)$ sont chanceux (pour une constante C à déterminer). Soit $x \in X$. Soit Y_k^x l'ensemble contenant x juste après le $k$ème pivot chanceux. $Y_0^x = Y$.

$$|Y_k^x| \leqslant \frac{2}{3}|Y_{k-1}^x| \tag{2}$$

   Si K est le nombre de choix de pivots chanceux, on a donc , $|Y_K^x| \leqslant \left(\frac{2}{3}\right)^K n$. Comme $|Y_K^x| \geqslant 1$, on a $K \leqslant C\log(n)$ pour $C = 1/\log(3/2)$.
4. Conclure que le temps d'exécution de l'algorithme Randomized Quicksort est $O(n \log n)$ avec grande probabilité. Pour $x \in S$, on note R_x la variable aléatoire donnant le nombre de comparaisons impliquant x dans une exécution de Randomized Quicksort. Le nombre total de comparaisons est majoré par $n \max_{x\in S} R_x$. Soit A_x l'événement «$R_x > kC\log n$» pour une constante $k > 4$ à déterminer. Si l'événement A_x est vrai, alors parmi les $kC\log n$ premiers choix de pivots dans un ensemble contenant x, au plus $C\log n$ étaient chanceux (par la question précédente, sinon l'algorithme se serait arrêté). On a donc pour $N = kC\log n$, avec les notations de la question 2

$$\mathbf{P}(A_x) \leqslant \mathbf{P}(X_N \leqslant C\log n) \leqslant \mathbf{P}(X_N \leqslant N/4) \leqslant \exp(-cN) = \exp(-kcC\log n)$$

   Si on choisit k tel que $kcC = 2$, alors $\mathbf{P}(A_x) \leqslant n^{-2}$. Par la borne de l'union, $\mathbf{P}(\max_x R_x \geqslant kC\log n) \leqslant 1/n$ et donc le temps d'exécution est majoré par $Ckn\log n$ avec probabilité $\geqslant 1 - 1/n$.

Exercice 4 MAX-CUT

Soit $G = (V, E)$ un graphe (non-orienté, sans boucle ni arête multiple). Pour $A \subset V$, on note $\mathrm{Cut}(A)$ la coupe selon A, c'est à dire l'ensemble des arêtes reliant un sommet de A à un sommet de $V \setminus A$. On considère la coupe maximale de G définie comme

$$\mathrm{MAXCUT}(G) = \max_{A \subset V} |\mathrm{Cut}(A)|$$

(on note $|X|$ le cardinal d'un ensemble X).

1. Montrer par la méthode probabiliste l'inégalité $\mathrm{MAXCUT}(G) \geqslant |E|/2$.

On considère l'algorithme glouton suivant pour approximer $\mathrm{MAXCUT}(G)$ (on suppose que l'ensemble V des sommets de G est muni d'un ordre total). On note $\mathrm{GreedyMAXCUT}(G)$ le cardinal de la coupe produite par cet algorithme. Il est évident que $\mathrm{GreedyMAXCUT}(G) \leqslant \mathrm{MAXCUT}(G)$.

$A \leftarrow \emptyset$, $B \leftarrow \emptyset$
Pour i **dans** V
  Soit a_i le nombre de voisins de i dans A et b_i le nombre de voisins de i dans B
  Si $a_i > b_i$ **alors** $B \leftarrow B \cup \{i\}$
  Sinon, $A \leftarrow A \cup \{i\}$
Renvoyer la coupe selon A.

---

2. A l'aide de cet algorithme, donner une autre preuve de l'inégalité $\mathrm{MAXCUT}(G) \geqslant |E|/2$.
3. On considère le graphe H_n de sommets $\{1, \dots, 2n\}$ et ayant pour arêtes

$$\{1,2\}, \{2,3\}, \dots, \{n-1, n\}\,;\ \text{ainsi que } \{i,j\} \text{ pour tous } i \in \{1, \dots, n\} \text{ et } j \in \{n+1, \dots, 2n\}.$$

Montrer que le rapport $\mathrm{MAXCUT}(H_n)/\mathrm{GreedyMAXCUT}(H_n)$ tend vers 2 quand n tend vers l'infini. L'algorithme glouton met tous les sommets de $n+1$ à $2n$ dans le même ensemble A ou B et produit donc une coupe de taille $(n-1) + \lceil n/2 \rceil \cdot n = \left(\frac{1}{2} + o(1)\right) n^2$. La coupe donnée car $A = \{1, \dots, n\}$ a cardinal n^2, d'où le résultat.

Dans la suite de l'exercice, on suppose que G_n est un graphe aléatoire de loi $\mathsf{G}(n, 1/2)$, c'est-à-dire qu'il a pour ensemble de sommets $\{1, \dots, n\}$ et que chaque arête est présente indépendamment avec probabilité $1/2$.

4. Pour $A \subset \{1, \dots, n\}$, quelle est la loi de $|\mathrm{Cut}(A)|$? C'est la loi binomiale $\mathsf{B}(k(n-k), \frac{1}{2})$ où $k = |A|$
5. Montrer que pour une constante C à déterminer,

$$\lim_{n \to \infty} \mathbf{P}\left(\mathrm{MAXCUT}(G_n) \leqslant \frac{n^2}{8} + Cn^{3/2}\right) = 1$$

Indication. *On pourra combiner une inégalité de* Chernoff *avec la borne de l'union.* Par la borne de l'union,

$$\mathbf{P}(\mathrm{MAXCUT}(G_n) \leqslant \frac{n^2}{8} + Cn^{3/2}) \leqslant \sum_{A \in V} \mathbf{P}(|\mathrm{Cut}(A)| \geqslant \frac{n^2}{8} + Cn^{3/2})$$

Pour $1 \leqslant k \leqslant n$, on a $k(n-k) \leqslant n^2/4$ (en supposant n pair) et donc

$$\mathbf{P}(|\mathrm{Cut}(A)| \geqslant \frac{n^2}{8} + Cn^{3/2}) \leqslant \mathbf{P}(\mathsf{B}(n^2/4, 1/2) \geqslant \frac{n^2}{8} + Cn^{3/2}) \leqslant \exp(-2C^2n^3/(n^2/4)) = \exp(-8nC^2)$$

d'après Chernoff I. On a donc

$$\mathbf{P}(|\mathrm{Cut}(A)| \geqslant \frac{n^2}{8} + Cn^{3/2}) \leqslant 2^n \exp(-8nC^2)$$

qui tend vers 0 lorsque n tend vers l'infini.

6. Soit k un entier. Soient X et Y deux variables i.i.d. de loi $\mathsf{B}(k, 1/2)$.

(a) Calculer $\mathbf{E}(X-Y)^2$ et $\mathbf{E}(X-Y)^4$.

(b) En déduire que $\mathbf{E}[|X-Y|] \geqslant c_1\sqrt{k}$ pour une constante $c_1 > 0$.

Indication. *On pourra démontrer et utiliser l'inégalité* $\mathbf{P}\left(Z \geqslant \frac{1}{2}\mathbf{E}[Z]\right) \geqslant \frac{1}{4}\frac{(\mathbf{E}[Z])^2]}{\mathbf{E}[Z^2]}$ *valable pour une v.a. Z à valeurs positives.*

On peut réaliser X et Y comme $X = X_1 + \cdots + X_k$ et $Y = Y_1 + \cdots + Y_k$ avec X_i, Y_i i.i.d. de loi $\mathsf{B}(1/2)$. Les v.a. $W_i = X_i - Y_i$ sont i.i.d. et vérifient $\mathbf{E}[W_i] = 0$, $\mathbf{E}[W_i^2] = \mathbf{E}[W_i^4] = 1/2$. En particulier, $\mathbf{Var}[W_i] = 1/2$. On a par additivité de la variance pour des sommes indépendantes $\mathbf{E}[(X-Y)^2] = \sum \mathbf{Var}[W_i] = k/2$. Par ailleurs, comme dans la preuve de la loi forte des grands nombres

$$\begin{aligned}
\mathbf{E}[(X-Y)^4] &= \sum_{i_1,i_2,i_3,i_4} \mathbf{E}[W_{i_1}W_{i_2}W_{i_3}W_{i_4}] \\
&= k\mathbf{E}[W_1^4] + 3k(k-1)\mathbf{E}[W_1^2W_2^2] \\
&= \frac{k}{2} + \frac{3k(k-1)}{4} \\
&= \frac{3k^2-k}{4}
\end{aligned}$$

L'inégalité appliquée à $Z = (X-Y)^2$ donne

$$\mathbf{P}((X-Y)^2 \geqslant k/4) \geqslant \frac{1}{4}\frac{k^2/4}{(3k^2-k)/4} = \frac{1}{4(3-1/k)} \geqslant 1/8$$

et donc $\mathbf{E}|X-Y| \geqslant c_1\sqrt{k}$ pour $c_1 = \frac{1}{16}$.

Enfin, démontrons l'inégalité suggérée en notant A l'événement «$Z \geqslant \frac{1}{2}\mathbf{E}[Z]$»

$$\mathbf{E}[Z] = \mathbf{E}[Z\mathbf{1}_A] + \mathbf{E}[Z\mathbf{1}_{\overline{A}}] \leqslant \left(\mathbf{E}[Z^2]\right)^{1/2}\mathbf{P}(A)^{1/2} + \frac{1}{2}\mathbf{E}[Z]$$

d'où le résultat en réarrangeant.

7. Montrer que si A et B sont des v.a. indépendantes de lois respectives $\mathsf{B}(a,1/2)$ et $\mathsf{B}(b,1/2)$, alors

$$\mathbf{E}[\max(A,B)] \geqslant \frac{a+b}{4} + c_2\sqrt{a+b}$$

pour une constante $c_2 > 0$.

Sans perte de généralité supposons $a \geqslant b$. Si $a \geqslant 2b$ on peut écrire

$$\mathbf{E}[\max(A,B)] \geqslant \mathbf{E}[A] = \frac{a}{2} \geqslant \frac{a+b}{3} \geqslant \frac{a+b}{4} + \frac{\sqrt{a+b}}{12}.$$

Si $b \leqslant a < 2b$, écrivons $A = A_1 + A_2$ avec $A_1 \sim \mathsf{B}(a-b,1/2)$ et $A_2 \sim \mathsf{B}(b,1/2)$. On a

$$\begin{aligned}
\mathbf{E}[\max(A,B)] &\geqslant \mathbf{E}[\frac{A+B+|A-B|}{2}] \\
&= \frac{a+b}{4} + \frac{1}{2}\mathbf{E}|A-B| \\
&\geqslant \frac{a+b}{4} + \frac{1}{2}\mathbf{E}(A-B)_+ \\
&\geqslant \frac{a+b}{4} + \frac{1}{2}\mathbf{E}(A_2-B)_+
\end{aligned}$$

Par symétrie, $\mathbf{E}(A_2-B)_+ = \mathbf{E}(B-A_2)_+$ et donc $\mathbf{E}(A_2-B)_+ = \frac{1}{2}\mathbf{E}|A_2-B| \geqslant \frac{c_1}{2}\sqrt{b}$. On a donc

$$\mathbf{E}\max(A,B) \geqslant \frac{a+b}{4} + \frac{c_1}{4}\sqrt{b} \geqslant \frac{a+b}{4} + \frac{c_1}{4\sqrt{3}}\sqrt{a+b}$$

8. En déduire que pour une constante $c_3 > 0$

$$\mathbf{E}[\mathrm{GreedyMAXCUT}(G_n)] \geqslant \frac{n^2}{8} + c_3 n^{3/2}.$$

On peut de manière équivalente générer le graphe aléatoire "en temps réel" au cours de l'exécution de l'algorithme glouton. Au $i$ème tour de boucle, on a $|A| + |B| = i - 1$; les variables aléatoires a_i et b_i sont indépendantes et de lois respectives $\mathsf{B}(|A|, 1/2)$ et $\mathsf{B}(|B|, 1/2)$. Le nombre d'arêtes ajoutées est $\max(a_i, b_i)$. D'après la question précédente,

$$\mathbf{E} \max(a_i, b_i) \geqslant \frac{i-1}{4} + c_3\sqrt{i-1}.$$

Par linéarité de l'espérance, la coupe produite a un cardinal supérieur à

$$\sum_{i=1}^{n} \left(\frac{i-1}{4} + c_3\sqrt{i-1} \right),$$

d'où le résultat