TD n°05

7 mars 2018

Exercice 1 - Borne de Chernoff pour des variables aléatoires continues

Soit X une variable aléatoire à valeurs dans $[0,1]$ avec $\mathbf{E}\{X\} = p$. Considérez la variable aléatoire $Y \in \{0,1\}$ avec $\mathbf{P}\{Y = 1\} = p$.

1.1 Montrez que pour tout $\lambda > 0$, $\mathbf{E}\left\{e^{\lambda X}\right\} \leqslant \mathbf{E}\left\{e^{\lambda Y}\right\}$.

1.2 En déduire que la borne de Chernoff montrée en classe est toujours valide si l'on remplace la condition $X_i \in \{0,1\}$ par $X_i \in [0,1]$.

Exercice 2 - Échantillonnage dans un rectangle

Soit $P \subseteq \mathbb{Z}^2$ de taille n. On veut pouvoir répondre rapidement à des requêtes du type : *quelle est la fraction des points de P dans le rectangle $r = [a_1, b_1] \times [a_2, b_2]$?*. Pour tout ensemble S, on définit la densité $r[S] = \frac{|S \cap r|}{|S|}$. On souhaite donc approximer efficacement $r[P]$ pour toute requête r. On dit que S (qu'on souhaite de taille bien inférieure à P) est une *ε-approximation de P* si pour tout r, on a $|r[P] - r[S]| \leqslant \varepsilon$.

2.1 Proposez une structure de données (ensemble) de taille m candidat pour être une ε-approximation de P. Comment répondez-vous à une requête r ?

2.2 Quelle valeur doit-on prendre pour m pour obtenir une ε-approximation avec probabilité $1 - \delta$?

Exercice 3 - Pièce biaisée

On a une pièce biaisée et on souhaite connaître à quel point, avec probabilité au moins $1 - \delta$.

3.1 Dans un premier temps, on vous assure que le biais est d'au moins ε. Combien de lancers faites-vous ?

Maintenant, vous n'avez plus d'hypothèse sur le biais. On souhaite trouver le sens du biais avec probabilité au moins $1 - \delta$.
Soit p la probabilité que la pièce tombe sur pile et $\hat{p}_n$ l'estimation de p obtenue après n lancers.

3.2 Trouvez une valeur de t (fonction de n et δ) pour laquelle on ait :

$$\mathbf{P}\{p \in [\hat{p}_n - t, \hat{p}_n + t]\} \geqslant 1 - \delta \ .$$

3.3 L'intervalle considéré ci-dessus est donc valable pour un n fixé. Modifions légèrement la procédure en la décomposant en plusieurs étapes successives. Nous allons faire n_1 lancers, puis n_2 lancers et ainsi de suite (on choisira la suite $(n_j)_{j \in \mathbb{N}}$ à la question suivante). On pose

$$I_j = \left[\hat{p}_{n_j} - \sqrt{\frac{\log\left(\frac{2}{\delta/2^j}\right)}{2n_j}}, \hat{p}_{n_j} + \sqrt{\frac{\log\left(\frac{2}{\delta/2^j}\right)}{2n_j}}\ \right] \ .$$

Donnez une borne inférieure de $\mathbf{P}\{p \in I_j\}$ puis de $\mathbf{P}\{\forall j \in \{1, 2, \ldots\},\ p \in I_j\}$.

3.4 Expliquez comment choisir la séquence $(n_j)_{j\in\mathbb{N}}$ et concluez.

Exercice 4 - Booster un algorithme randomisé générique

Vous connaissez un algorithme randomisé polynomial $\mathcal{A}$ pour décider si $x \in \{0,1\}^*$ is in the language L or not. $\mathcal{A}$ a la propriété suivante. Si $x \in L$, alors $\mathbf{P}\{\mathcal{A}(x) = 0\} \leq 1/4$ et si $x \notin L$, alors $\mathbf{P}\{\mathcal{A}(x) = 1\} \leq 1/3$. (Observez que la probabilité est prise sur toutes les graines aléatoires de l'algorithme et *non* sur les inputs x.)

Proposez un algorithme polynomial $\mathcal{B}$ tel que $\forall x \in \{0,1\}^*$, on ait $\mathbf{P}\{\mathcal{B}(x) = \mathbf{1}_{x\in L}\} \geq 1 - 2^{-|x|}$. ($\mathbf{1}_{x\in L} = 1$ if $x \in L$ and 0 otherwise, and $|x|$ denotes the length of the bitstring x.)

Exercice 5 - Collectionneur de vignettes

Soit X le nombre de paquets de céréales achetés avant d'avoir *toute la collec'*.
Montrez que
$$\mathbf{P}\{X \geq n \ln n + cn\} \leqslant e^{-c}.$$

(En classe, vous avez montré un résultat plus faible en utilisant l'inégalité de Chebyshev.)