TD n°06

14 mars 2018

Exercice 1 - Pièce biaisée

On a une pièce biaisée et on souhaite connaître le sens du biais, avec probabilité au moins $1-\delta$.

1.1 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.

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

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

1.3 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 $\mathbf{P}\{\forall j \in \{1,2,\ldots\},\ p \in I_j\}$.

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

Exercice 2 - 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 $\mathbf{P}\{\mathcal{B}(x) = \mathbf{1}_{x\in L}\} \geq 1 - 2^{-|x|}$. ($\mathbf{1}_{x\in L} = 1$ si $x \in L$ et 0 autrement, et $|x|$ représente la longueur du mot x.)

Exercice 3 - 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.)

Exercice 4 - Tri par seaux

On a un ensemble de $n = 2^m$ éléments à trier et chaque élément est un entier uniformément aléatoire dans $[0, 2^k - 1]$ et indépendant, où $k \geqslant m$ et k est connu. Dans ce cas, l'espérance du temps d'exécution du tri par seaux est $O(n)$.
Ce tri marche en deux phases. Dans la première phase (pré-tri), on place les éléments dans n seaux selon certaines règles. Dans la deuxième phase, on appelle un algorithme de tri simple, en temps quadratique, sur chaque seau. Enfin, on concatène les listes d'éléments obtenues. Pour que l'algorithme réponde correctement, il faut que le pré-tri place en seau i des éléments tous plus petits que les éléments en seau j avec $j > i$.

4.1 Donnez un moyen de faire le pré-tri qui satisfasse cette condition. On veut que pour chaque élément x, le choix du seau pour x se fasse en temps constant (en supposant qu'une opération arithmétique se fait en temps constant).

4.2 Soit X_i une variable aléatoire qui compte le nombre d'éléments dans le seau i (après le pré-tri). Quelle distribution suivent les X_i ?

4.3 Montrez que l'espérance de la complexité du tri par seau est $O(n)$.

Exercice 5 - Approximation de Poisson

On considère à nouveau le modèle *Boules et Bacs* (*Balls and Bins*) : on met aléatoirement m boules dans n bacs. Les nombres de boules dans les différents bacs ne sont pas indépendants. On veut approximer ce modèle par une loi de Poisson. $Y_1, \ldots, Y_n$ sont des variables aléatoires indépendantes suivant chacune une loi de Poisson avec paramètre (i.e., espérance) $\mu = m/n$.

5.1 Montrez que $Y = \sum_{i=1}^{n} Y_i$ suit une loi de Poisson et déterminez son paramètre.

5.2 Montrez que la distribution de $(Y_1, \ldots, Y_n)$ conditionné par $Y = m$ est la même que celle de $(X_1, \ldots X_n)$.

5.3 Soit f une fonction à n variables et à valeurs dans les réels positifs. Montrez que

$$\mathbf{E}\left\{f(X_1, \ldots, X_n)\right\} \leq e\sqrt{m}\mathbf{E}\left\{f(Y_1, \ldots, Y_n)\right\} \ .$$

Vous pourrez utiliser que $m! < e\sqrt{m}\left(\frac{m}{e}\right)^m$.

5.4 Appelons *cas de Poisson* les événements se produisant quand le nombre de boules dans les bacs sont prises comme variables de Poisson indépendantes d'espérance m/n, et *cas de Boules et Bacs* quand on jète m boules dans n bacs aléatoirement et indépendamment.
Quelle fonction f choisisseriez-vous dans le résultat précédent pour conclure que :
Chaque événement qui arrive avec probabilité p dans le cas de Poisson, arrive avec probabilité au plus $pe\sqrt{m}$ dans le cas Boules et Bacs.

5.5 Retrouvez la borne inférieure sur la charge maximum dans le cas où $m = n$, en utilisant l'approximation de Poisson.
Plus précisément, montrez que celle-ci est au moins $\ln n / \ln \ln n$ avec probabilité $1 - 1/n$ pour n suffisamment grand.