TD n°07

21 mars 2018

Exercice 1 - 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 2 - 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 $j > i$.

2.1 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).

2.2 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 ?

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

Exercice 3 - 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) $\mu = m/n$.

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

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

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

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

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

3.4 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.

3.5 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.

Exercice 4 - Numéros de Sécurité Sociale

Le numéro de sécurité sociale américain est composé de 9 chiffres. Les quatre derniers chiffres servent comme code secret. Supposons que les chiffres soient choisis indépendants et uniformément aléatoirement (et sans vérifier si un même numéro a déjà été donné).

4.1 On dénote par n le nombre d'individus. Événement : Au moins deux personnes ont le même code secret. Pour quelle valeur de n cet événement est plus probable d'arriver que de ne pas arriver ?

4.2 Même question avec l'événement : Au moins deux personnes ont le même numéro de sécurité sociale.

Exercice 5 - Uniformisation

On dispose d'un générateur de bits aléatoires indépendants suivant la même loi de Bernoulli de paramètre $0 < p < 1$. Mais on ne connait pas p. Proposez un algorithme qui utilise cette source d'aléa pour retourner un bit uniforme (i.e., suivant un Bernoulli de paramètre $1/2$) et analysez l'espérance du nombre d'appels au générateur.