

Structures et algorithmes aléatoires

TD7

20 novembre 2015

Exercice 1 Chemins disjoints

Soit $G = (V, E)$ un graphe non orienté et soient $(u_1, v_1), (u_2, v_2), \dots, (u_n, v_n)$ des couples de sommets. Pour chaque $1 \leq i \leq n$ soit F_i un ensemble de m chemins de u_i vers v_i .

Supposons que chaque chemin dans F_i ne partage des arêtes qu'avec au plus k chemins de chaque F_j avec $i \neq j$. Montrer que si $8nk/m \leq 1$, alors il existe $(P_1, \dots, P_n) \in F_1 \times \dots \times F_n$ tel que les chemins P_i ne partagent pas d'arêtes.

Exercice 2 Collectionneur de coupons

Soit X le nombre de boîtes ouvertes jusqu'à ce que le collectionneur ait obtenu tous les coupons. On veut montrer que

$$\lim_{n \rightarrow \infty} \mathbb{P}(X > n \log n + cn) = 1 - e^{-e^{-c}}$$

pour tout $c > 0$.

Pour ça, on interprète l'événement $X > n \log n + cn$ comme une urne étant vide après le jet de $m = \lfloor n \log n + cn \rfloor$ balles dans n urnes. On denote par $\mathcal{E}$ l'événement qu'aucune urne n'est vide si le nombre de balles dans chaque urne est une variable poissonnienne à espérance $\log n + c$ et par Y le nombre total de balles.

1. Montrer que $\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{E}) = e^{-e^{-c}}$.
2. Montrer que $\mathbb{P}(|Y - m| > \sqrt{2m \log m}) \rightarrow 0$ quand $n \rightarrow \infty$.
3. Montrer que $\mathbb{P}(\mathcal{E} \mid |Y - m| \leq \sqrt{2m \log m}) - \mathbb{P}(\mathcal{E} \mid Y = m) \rightarrow 0$ quand $n \rightarrow \infty$.
4. En déduire que $\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{E}) = \lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{E} \mid Y = m)$.
5. Conclure.

Exercice 3 Filtre de Bloom

Soit U un univers (ensemble fini de grand cardinal). On cherche à représenter un sous-ensemble S de U à m éléments. Pour ce faire, on utilise k fonctions de hachage parfaites dans un ensemble à n élément : on a des fonction $h_i : U \rightarrow [0, \dots, n-1]$ telles que pour tout $u \in U$ et $\ell \in [0, n-1]$, $\mathbb{P}(h_i(u) = \ell) = 1/n$, et les valeurs $h_i(u)$ sont mutuellement indépendantes.

On considère maintenant A un tableau à n cases, $A[0], \dots, A[n-1]$, initialisé à tout 0. Pour tout $i \in [1, k]$, $s \in S$, on pose $A[h_i(s)] = 1$.

Si l'on fait une requête sur un élément u , on retourne vrai si pour tout $i \in [1, k]$, $A[h_i(u)] = 1$. Si l'élément est bien dans S , la réponse est toujours correcte. Par contre, on peut avoir de faux positifs (un élément n'est pas dans S , mais on retourne vrai tout de même).

1. Quelle est la probabilité p que $A[\ell] = 0$? Quelle est la probabilité d'avoir un faux positif?

Quand n est grand, p est proche de $e^{-km/n}$. On fixe donc pour la suite $p = e^{-km/n}$.

2. Pour m et n fixés, donner alors le meilleur k possible (on pourra dans un premier temps s'affranchir de la contrainte « k est un entier »).

On reprend $p' = (1 - 1/n)^{km}$, la probabilité que $A[\ell] = 0$. Soit X le nombre de cases ℓ telles que $A[\ell] = 0$.

3. Montrer que $\mathbb{P}(X - np' > cn) \leq e\sqrt{n}e^{-nc^2/3p'}$.