MPRI

Cours 2.15: Analyse d'algorithmes

— Partiel du 11 Décembre 2014, durée 3h—-

Documents autorisés : notes de cours. Chacun des 3 problèmes est noté sur 10 points, et la somme des points obtenus fait la note sur 20 (sic!).

Problème I (10 points)

Soit $\mathcal{A}$ un alphabet fini totalement ordonné. Étant donné un mot $w = w_1 w_2 \cdots w_n$, on appelle *record* (strict) toute lettre w_k qui est strictement plus grande que toutes les lettres qui la précèdent dans w :

$$(\forall j < k) \quad : \quad w_j < w_k.$$

Par convention la première lettre d'un mot est toujours un record et le mot vide n'a pas de record. Voici par exemple, avec $\mathcal{A} = \{\dashv, \lfloor, \rfloor, \ldots\}$ et l'ordre usuel $a < b < c < d \cdots$ sur les lettres, un mot dont les records sont encadrés:

$$\boxed{a}\ \boxed{l}\ g\ \boxed{o}\ \boxed{r}\ i\ \boxed{t}\ m\ i\ q\ \boxed{u}\ e.$$

On dira aussi que l'ensemble des lettres correspondantes est l'*ensemble des records*; sur l'exemple: $\{a, l, o, r, t, u\}$.

Question 1. On considère, dans un premier temps, l'alphabet à deux lettres $\mathcal{A} = \{\dashv, \lfloor\}$ (avec $a < b$). Donner les quatre spécifications (par constructeurs ou par expressions régulières) des mots sur $\mathcal{A}$ ayant pour ensemble de records:

$$\emptyset, \quad \{a\}, \quad \{b\}, \quad \{a, b\}.$$

Écrire les quatre fonctions génératrices ordinaires (OGF) associées.

Question 2. Soit $R(z, u)$ la fonction géneratrice (ordinaire, bivariée) où z marque la longueur des mots (sur l'alphabet à deux lettres $\mathcal{A}$) et u marque le nombre de records. Justifier la forme:

$$R(z, u) = \left(1 + u\frac{z}{1-z}\right)\left(1 + u\frac{z}{1-2z}\right).$$

On pose

$$K(z) = \left[\frac{\partial}{\partial u} R(z, u)\right]_{u=1}.$$

Calculer $K(z)$ et déterminer c tel que

$$K(z) \sim \frac{c}{1-2z} \qquad (z \to 1/2).$$

En déduire la valeur asymptotique du nombre moyen de records dans un mot de longueur n, lorsque $n \to \infty$.

Question 3. Soit $\{X_1, X_2, \ldots, X_r\}$ un ensemble d'indéterminées. Développer en une somme de monômes le produit suivant

$$\prod_{j=1}^{r} (1 + X_j)\,.$$

On considère désormais un alphabet à r lettres $\mathcal{A} = \{a_1, a_2 \ldots, a_r\}$, muni de l'ordre $a_1 < a_2 < \cdots$. Justifier la forme suivante de la fonction géneratrice $R(z,u)$, où z marque la longueur des mots sur $\mathcal{A}$ et u marque le nombre de records:

$$R(z,u) = \prod_{j=1}^{r} \left(1 + u\frac{z}{1-jz}\right).$$

(On pourra déterminer dans un premier temps la fonction géneratrice des mots pour lesquels l'ensemble des records est un certain sous-ensemble $\{a_{j_1}, \ldots, a_{j_m}\}$.)

Question 4. On pose de nouveau $K(z) = \left[\frac{\partial}{\partial u}R(z,u)\right]_{u=1}$. Calculer $K(z)$ (on pourra utiliser une derivation logarithmique et la forme réduite de $K(z)$ n'est pas demandée). Déterminer c tel que

$$K(z) \sim \frac{c}{1-rz} \qquad (z \to 1/r).$$

En déduire que le nombre moyen de records dans un mot de grande longueur est asymptotiquement (à r fixé) le nombre harmonique

$$H_r = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{r}.$$

Question 5. On considère maintenant la classe $\mathcal{P}$ des permutations, de fonction génératrice exponentielle $P(z) = 1/(1-z)$, vues aussi comme des mots particuliers écrits sur l'alphabet infini $\mathbb{Z}_{\geq 1}$ muni de l'ordre naturel. Donner une bijection de $\mathcal{P}_\backslash$ dans $\mathcal{P}_\backslash$ qui met en correspondance les records et les cycles (on pourra individualiser dans chaque cycle son "leader", c'est-à-dire, sa plus grande valeur). En déduire la fonction bivariée où u marque les records dans les permutations:

$$P(z,u) = \exp\left(u \log \frac{1}{1-z}\right).$$

Quel est le nombre moyen de records dans une permutation de taille n? Comment ce résultat se compare-t-il à celui de la question **Q4**?

Question 6. On considère finalement un alphabet à r lettres $\mathcal{A} = \{a_1, a_2, \ldots, a_r\}$, muni de l'ordre $a_1 < a_2 < \cdots$ et où chaque lettre a une probabilité p_j d'apparaître. Par extension, la probabilité d'un mot de longueur n est alors le produit des probabilités des n lettres qui le composent.
Justifier la forme de la fonction géneratrice $R(z,u)$, où z marque la longueur des mots sur $\mathcal{A}$ et u marque le nombre de records:

$$R(z,u) = \prod_{j=1}^{r} \left(1 + u\frac{p_j z}{1 - \sum_{i=1}^{j} p_i z}\right).$$

Question 7. On pose de nouveau $K(z) = \left[\frac{\partial}{\partial u}R(z,u)\right]_{u=1}$. Calculer $K(z)$ et montrer que

$$K(z) \sim \frac{c}{1-z} \qquad (z \to 1/r),$$

où

$$c = \sum_{j=1}^{r} \frac{p_j}{p_j + p_{j+1} + \cdots + p_r}.$$

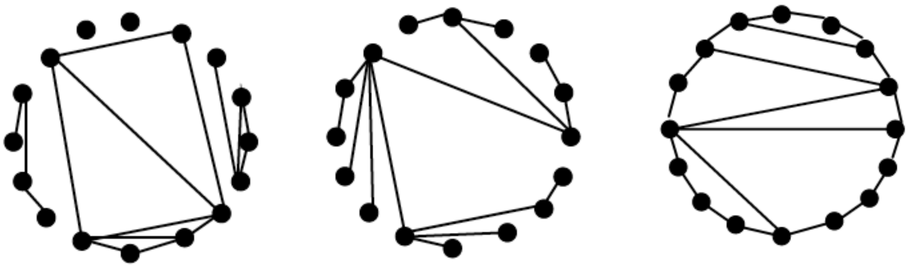
En déduire le nombre moyen asymptotique de records dans un mot de grande longueur sur un alphabet à r lettres dont les probabilités d'occurrences sont $(p_1, p_2, \ldots, p_r)$.

Problème II (10 points)

Soit $P_n = \{s_1, \ldots, s_n\}$ un ensemble de points, ordonnés en sens trigonométrique, constituant les sommets d'un polygone convexe régulier. On considère des configurations non croisées de P_n formées de l'union de certains côtés du polygone ainsi que de certaines de ses diagonales, qui ne peuvent s'intersecter qu'en leurs extrémités. La taille d'une configuration est son nombre n de sommets.

- un *graphe non croisé* est un graphe planaire dont les sommets sont $\{s_1, \ldots, s_n\}$ et les arêtes sont des côtés et des diagonales non croisées de P_n.
- un *arbre non croisé* est un graphe non croisé qui est un arbre; on considère que cet arbre est enraciné en s_1.
- une *dissection* est formée de l'ensemble des côtés de P_n et d'une collection de diagonales qui se coupent uniquement en leurs extrémités.

La figure suivante montre, de gauche à droite, un graphe, un arbre et une dissection de taille 16. De plus le graphe a 16 arêtes et la dissection comporte 6 régions.(On n'a pas fait figuré les étiquettes $s_1, \ldots s_{16}$.)



I. Arbres non croisés

Question 1. Soit $\mathcal{T}$ la classe des arbres non croisés enracinés, et $\mathcal{V}$ la classe des arbres non croisés dont on a supprimé la racine (et les arêtes qui en partent). Montrer que:

$$\mathcal{T} = \bullet \times \mathcal{V}, \quad \mathcal{V} = \mathrm{Seq}(\mathcal{V} \times \bullet \times \mathcal{V})$$

Question 2. En déduire que la série génératrice $T(z) = \sum T_n z^n$, où T_n est le nombre d'arbres non croisés enracinés sur P_n, vérifie:

$$T(z) = z + zy(z) \quad \text{avec} \quad y(z) = z(1 + y(z))^3$$

Question 3. Donner une expression exacte de T_n.

Question 4. Quelle est la valeur ρ du rayon de convergence de $T(z)$? Donner un développement de $T(z)$ autour de $z = \rho$ et en déduire la valeur asymptotique de T_n.

Question 5. Soit $T(z, u)$ la série bivariée des arbres non croisés, où u marque le nombre de feuilles (sommets de P_n avec 1 seul voisin). Montrer que

$$T(z, u) = zu + zy(z, u) \quad \text{et} \quad T'_u(z, u)|_{u=1} = z + z^3 \frac{y'(z)}{y(z)}.$$

En déduire que le nombre moyen μ_n de feuilles dans un arbre non croisé sur P_n est asymptotiquement linéaire : $\mu_n \sim cn$. Déterminer la constante c.

II. Dissections et graphes non croisés

Question 6. Soit Δ est une dissection de P_n, et soit λ la région de Δ qui contient le côté s_1s_2 de P_n; si la région λ a $r+1$ côtés, alors Δ s'identifie à une séquence de r dissections (certaines pouvant être réduite à une unique arête).
Soit $D(z) = \sum D_n z^n$ la série génératrice ordinaire des dissections, où D_n est le nombre de dissections sur P_n. Montrer que si l'on marque par z^2 la dissection formée d'une seule arête, on a

$$D(z) = z^2 + \frac{D^2(z)}{z} + \ldots + \frac{D^r(z)}{z^{r-1}} + \ldots$$

Question 7. Soit $D(z) = zy(z)$, montrer que $y(z)$ vérifie l'équation

$$y(z) = z\frac{1-y}{1-2y}$$

Question 8. Montrer que $n[z^n]y(z)$ représente le nombre de n-uplets de compositions d'entiers dont la somme totale vaut $n-1$ (par exemple $((1,2,5),(1,1,2,4)(2,3))$ est un triplet de compositions d'entiers dont la somme totale vaut 21).

Question 9. On considère la série bivariée $D(z,u) = \sum D_{n,k} u^k z^n$, où $D_{n,k}$ est le nombre de dissections sur P_n avec k régions. Montrer que $D(z,u)$ vérifie:

$$D(z,u) = z^2 + u\left(\frac{D^2(z,u)}{z} + \ldots + \frac{D^r(z,u)}{z^{r-1}} + \ldots\right)$$

En déduire que

$$D_{n,k} = \frac{1}{k}\binom{n-3}{k-1}\binom{n+k-2}{k-1}$$

Question 10. Montrer que le nombre de graphes non croisés sur P_n est $G_n = 2^n D_n$, et que le nombre de graphes non croisés sur P_n avec k arêtes vaut:

$$G_{n,k} = \sum_{j=0}^{k} \binom{n}{j} D_{n,j+1}.$$

Problème III (10 points)

Dans cet exercice on s'intéresse à des distributions biaisées sur les permutations, qui suivent un modèle introduit par le mathématicien-biologiste Warren Ewens pour la génétique des populations.

I. Modèle d'Ewens classique sur les permutations

L'objectif est de définir une probabilité non-uniforme sur les permutations de taille n. On a un paramètre θ qui est un réel strictement positif, et on veut que la probabilité d'une permutation σ de taille n soit proportionnelle à $\theta^{\chi(\sigma)}$, où $\chi(\sigma)$ est le nombre de cycles de σ. Autrement dit,

$$\mathbb{P}_n(\sigma) = \frac{\theta^{\chi(\sigma)}}{w_n(\theta)},$$

où $w_n(\theta) = \sum_{\sigma \in \mathcal{S}_n} \theta^{\chi(\sigma)}$ est le facteur de normalisation pour que les probabilités somment bien à 1.

Question 1. Pour quelle(s) valeur(s) de θ retrouve-t-on la distribution uniforme ?

Question 2. Rappelez, en justifiant, une expression simple pour la série génératrice exponentielle bivariée $P(z,u)$ qui compte le nombre de cycles dans une permutation :

$$P(z,u) = \sum_{n,k\geq 0} \frac{p_{n,k}}{n!} z^n u^k,$$

où $p_{n,k}$ est le nombre de permutations de taille n aillant exactement k cycles.

Question 3. En déduire une expression simple pour la série génératrice exponentielle $W(z)$ des $w_n(\theta)$ définie par

$$W(z) = \sum_{n\geq 0} \frac{w_n(\theta)}{n!} z^n$$

Question 4. En déduire que pour tout $n \geq 1$ on a

$$w_n(\theta) = \theta(\theta+1)\cdots(\theta+n-1).$$

On pourra utiliser le fait que pour $n \geq 1$,

$$[z^n](1+z)^a = \frac{a(a-1)\dots(a-n+1)}{n!}.$$

II. Asymptotique du nombre de cycles

On s'intéresse maintenant au nombre moyen de cycles dans le modèle d'Ewens, pour θ fixé. On note $\mathbb{E}_n[\chi]$ cette quantité qui est donc définie par

$$\mathbb{E}_n[\chi] = \sum_{\sigma\in\mathcal{S}_n} \chi(\sigma)\,\mathbb{P}_n(\sigma).$$

Question 5. En utilisant les séries génératrices, montrez que pour $n \geq 1$,

$$\mathbb{E}_n[\chi] = \frac{n!}{w_n(\theta)} \left([z^n]\theta \frac{d}{du} P(z,u)\Big|_{u=\theta} \right)$$

Question 6. A partir de la formule trouvée pour $W(z)$ à la question **Q3**, donnez un équivalent asymptotique à $\frac{1}{n!} w_{n(\theta)}$ par le théorème de transfert.

Question 7. En déduire un équivalent asymptotique à $\mathbb{E}_n[\chi]$ par le théorème de transfert.

III. Modèle d'Ewens sur les points fixes

Dans cette partie on reprend l'idée de biaiser la distribution sur les permutations, mais cette fois-ci on veut que la probabilité de σ soit proportionnelle à $\theta^{\xi(\sigma)}$, où $\xi(\sigma)$ est le nombre de points fixes de σ.

Question 8. En vous inspirant de ce qui a été fait pour les cycles dans la partie I, montrez que pour $n \geq 1$,

$$\mathbb{E}_n[\xi] = \frac{\theta[z^n]\frac{d}{du}Q(z,u)\Big|_{u=\theta}}{[z^n]Q(z,\theta)},$$

où $Q(z,u)$ est la série génératrice exponentielle bivariée qui compte les permutations selon leur nombre de points fixes.

Question 9. Déduire de l'expression du cours de $Q(z,u)$ et de la question précédente que $\mathbb{E}_n[\xi] \to \theta$ quand $n \to \infty$.