# MPRI2 2004–2005

Cours *"Analyse d'algorithmes"*,

Philippe Flajolet, Michèle Soria, et Jean-Marc Steyaert,

Jeudi 17 Février – Examen, durée 3h

### Problème 1. Un petit QCM

Vous devez choisir entre deux options et justifier *en au moins une phrase et en au plus trois phrases* les raisons de chacun de vos choix.

**Q1.** Le nombre d'arbres 2–3 (ici: arbres plans enracinés non-étiquetés dont les sommets ont degré 0, 2, ou 3) de taille $n$ vaut:

[**A** ] $\frac{1}{n}\text{coeff}[w^{n-1}](1+w^2+w^3)^n$
[**B** ] $\text{coeff}[z^n]\left(1+\sqrt[2]{1-2z^2}+\sqrt[3]{1-3z^3}\right)$

**Q2.** La fonction génératrice exponentielle [EGF] des permutations comportant un nombre pair de cycles vaut

[**A** ] $\frac{1}{2}\frac{1}{1-z}+\frac{1}{2}(1-z)$
[**B** ] $\sqrt{\frac{1+z}{1-z}}$

**Q3.** L'ordre de croissance exponentiel du nombre d'arbres unaires-binaires (ici: arbres plans enracinés non-étiquetés dont les sommets ont degré 0, 1, ou 2) est de la forme

[**A** ] $4^n$
[**B** ] $3^n$

**Q4.** Soit $f(z)=\frac{1}{2-\exp(e^z-1)}$. Alors, pour une certaine constante $c$, on a

[**A** ] $[z^n]f(z)\sim c(\log(\log 2+1))^{-n}$
[**B** ] $[z^n]f(z)\sim c(1-\log\log 2)^n$

**Q5.** Soit $g(z)=\exp(e^z-1)$. Alors:

[**A** ] $[z^n]g(z)$ tend vers 0.
[**B** ] $[z^n]g(z)$ tend vers $+\infty$.

**Q6.** La série géneratrice des mots sur l'alphabet $\{a,b\}$ contenant exactement trois fois le motif *aaabbb* est une fraction rationnelle

[**A** ] ayant un pôle dominant qui est simple
[**B** ] ayant un pôle dominant qui est de multiplicité $\geq 2$.

## Problème 2. Allocations et hachage

Soit $\mathcal{A}_n^{(m)}$ l'ensemble de toutes les allocations possibles de $n$ boules distinguées dans $m$ urnes (distiguées). De manière équivalente, $\mathcal{A}_n^{(m)}$ peut être vu comme l'ensemble des mots de longueur $n$ sur un alphabet à $m$ lettres.

**Q1.** Justifier brièvement le fait que la fonction génératrice exponentielle [EGF] de $\mathcal{A}^{(m)} = \cup_n \mathcal{A}_n^{(m)}$ vaut

$$A^{(m)}(z) = (e^z)^m = e^{mz}.$$

Soient de même $\mathcal{B}^{(m)}$ les allocations injectives (chaque urne recoit au plus une boule) et $\mathcal{C}^{(m)}$ les allocations surjectives (chaque urne recoit au moins une boule). Déterminer les EGFs

$$B^{(m)}(z), C^{(m)}(z).$$

**Q2.** Donner des expressions des coefficients,

$$B_n^{(m)} = n![z^n]B^{(m)}(z), \qquad C_n^{(m)} = n![z^n]C^{(m)}(z).$$

Le résultat est "simple" dans le premier cas, donné par une somme dans le second cas. Quel est le comportement asymptotique de $B_n^{(m)}$ et de $C_n^{(m)}$ lorsque $n \to +\infty$ tandis que $m$ reste fixe?

**Q3.** Soit $f(z)$ une fonction analytique en $z = 0$ et à coefficients $\geq 0$. Justifier brièvement l'inégalité

$$[z^n]f(z) \leq \frac{f(r)}{r^n},$$

pour tout $r$ inférieur au rayon $R$ de convergence de la série représentant $f(z)$ au voisnage de $z = 0$.

Supposons qu'il existe une solution au système

$$(\Sigma) : \quad \frac{sf'(s)}{f(s)} = n, \qquad 0 < s < R,$$

alors observer que

$$[z^n]f(z) \leq \frac{f(s)}{s^n}.$$

Quel est l'avantage de cette dernière inégalité? Que se passe-t-il si $s$ est remplacé par une solution approchée $s_0$ au système $(\Sigma)$?

**Q4.** On suppose désormais que $m = n$. Montrer que la probabilité $p_{b,n}$ que chaque urne contient au plus $b$ boules est majorée:

$$p_{b,n} \leq \frac{n!}{n^n} e_b(1)^n \qquad \text{où} \quad e_b(1) = \sum_{j=0}^{b} \frac{1}{j!}.$$

**Q5.** Indiquer un schéma de preuve du fait que

$$\lim_{n \to +\infty} p_{b,n} = 0 \qquad \text{lorsque} \quad b = \lfloor \log n \rfloor.$$

(On ne détaillera pas les calculs.)

Quelle est l'implication de ce dernier résultat pour la gestion de tables de hachage, lorsqu'on dispose d'un système physique (e.g. disque magnétique) doté de "pages" capables de contenir un nombre $b_0$ fixé d'enregistrements?

## Problème 3. Graphes de fonctions

Un graphe fonctionnel de taille $n$ est le graphe d'une fonction de [1..n] vers [1..n]. Un tel graphe est dit aléatoire lorsque les $n^n$ fonctions sont équiprobables.

On rappelle que la série génératrice exponentielle des graphes fonctionnels est

$$g(z) \equiv \sum g_n \frac{z^n}{n!} = \frac{1}{1 - a(z)}$$

où $a(z)$ est la série génératrice exponentielle des arbres de Cayley (arbres généraux étiquetés non planaires).

**Q1.** Exprimer la série génératrice bivariée des graphes fonctionnels dans lesquels on marque les cycles de longueur $k$ (i.e. formés de $k$ arbres de Cayley).

En déduire la valeur asymptotique moyenne (lorsque $n \to \infty$) du nombre cycles de longueur $k$ (fixé) dans un graphe fonctionnel aléatoire de taille $n$.

**Q2.** Donner l'expression de la probabilité pour qu'un graphe fonctionnel soit constitué de $k$ composantes connexes. Calculer la valeur asymptotique de cette probabilité, lorsque $n \to \infty$.

On va considérer à présent les graphes de fonctions de [1..n] vers [1..n] telles que chaque point est l'image de 2 ou 0 points.

**Q3.** Montrer que la série génératrice exponentielle des arbres binaires (0 ou 2 fils) étiquetés non planaires satisfait l'équation

$$b(z) = z + \frac{z}{2} b(z).$$

Calculer la valeur exacte et la valeur asymptotique du nombre d'arbres de taille $n$.

**Q4.** Exprimer la série génératrice des graphes de fonctions de [1..n] vers [1..n] telles que chaque point est l'image de 2 ou 0 points.

Que vaut asymptotiquement le nombre moyen de composantes connexes dans ces graphes de fonctions. Même question pour le nombre d'arbres.

**Q5.** On considère la série génératrice

$$f(z) = \frac{1}{1 - b(z)}$$

Donner une (ou plusieurs) interprétation combinatoire de cette série génératrice de dénombrement de F-structures. Calculer la valeur asymptotique du nombre moyen d'arbres dans une F-structure de taille $n$.

## Problème 4. Histoires de motifs

Dans tout le problème, on considère des mots écrits sur l'alphabet à deux lettres $A = \{a, b\}$. On appellera longueur d'un mot son nombre de lettres. Deux mots seront particulièrement distingués et appelés *motifs* : $v = aba$ et $w = bba$.

**Q1.** Exprimer la s.g.o. des nombres d'occurrences $\mathrm{Occ}(v, n)$ du motif $v$ dans les mots de longueur $n$. Que peut-on dire de celle de $w$ ? Plus généralement de celle d'un motif de longueur $\ell$ ?

**Q2.** Soit $S_v$ l'ensemble des mots qui ne contiennent pas le motif $v$. Quelle est la nature de $S_v$ ? Proposer une méthode pour calculer sa s.g.o. $h_v(z)$(selon la longueur). Faire de même pour le motif $w$. Comparer les deux résultats. Expliquer combinatoirement.

**Q3.** On veut maintenant calculer la série bivariée $f_v(z, u) = \Sigma_{n,k}\mathrm{Occ}(v, n, k)u^k z^n$ où $\mathrm{Occ}(v, n, k)$ est le nombre de mots de longueur $n$ qui contiennent exactement $k$ occurrences du motif $v$. Quelle est la nature de cette série Comment peut-on la calculer ? Donner une expression pour cette série ; en déduire la variance de $\mathrm{Occ}(v, n)$.

**Q4.** Si l'on se livre au même calcul pour le motif $w$, quelque chose doit-il changer ? Expliquer pourquoi.

**Q5.** On fixe maintenant le paramètre $k$ : disons que $k$ varie ente 10 et 100. Quel est le comportement asymptotique en $n$ du nombre de mots de longueur $n$ qui contiennent exactement $k$ occurrences du motif $v$ ? Est-ce le même pour $w$ ?
Quelle est la nature de la loi-limite des distributions lorsque $k$ grandit avec $n$ ?

**Q6.** On interprête maintenant les deux motifs $v$ et $w$ comme étant des marqueurs biologiques dans des séquences : par exemple, $v$ correspond à un début de zone codante pour une protéine et $w$ à la fin d'une telle zone. Comment peut-on calculer la distribution des séquences de longueur $n$ qui contiennent $k$ zones codantes ? Quelle est sa nature ? Quelles difficultés sont à craindre dans la pratique ?