MPRI

Cette épreuve est à faire en 3 heures en tout. Il est demandé de mettre en relief les idées plutôt que les détails de calcul. La notation tiendra compte de la qualité de présentation des arguments.
You can write your exam paper in English !
Tous les documents-papiers peuvent être utilisés. Bonne chance !

Problème 1 : Un petit QCM

Vous devez choisir entre deux options et justifier *en au moins une et au plus trois phrase(s) et/ou formule(s)* les raisons de chacun de vos choix.

Q1. Soient $\mathcal{A}$ la famille des arbres planaires non-étiquetés enracinés dont les degrés des sommets sont restreints à $\{0, 2, 3, 4, 5, \ldots\}$. La série génératrice (ordinaire) de dénombrement de $\mathcal{A}$ est
[**A**] $A(z) = \exp\left(1 + z + \log \frac{1}{1 - \sqrt{1 - 6z} + z^2}\right)$
[**B**] $A(z) = \frac{1}{4}\left(1 + z - \sqrt{1 - 6z + z^2}\right)$.

Q2. Le nombre d'arbres 2–3 (arbres plans enracinés non-étiquetés dont les sommets ont degré 0, 2, ou 3) de taille n vaut :
[**A**] $\frac{1}{n+1}\binom{2n}{n}$
[**B**] $\frac{1}{n}[y^{n-1}](1 + y^2 + y^3)^n$

Q3. On rappelle que $[z^n]\frac{1}{\sqrt{1-z}} \sim \frac{1}{\sqrt{\pi n}}$ et $[z^n] \log \frac{1}{1-z} = \frac{1}{n}$.
Soit $k_n = [z^n]\frac{1}{\sqrt{1-3z}} \log \frac{1}{1-2z}$. Alors :
[**A**] $k_n \sim \log 3 \frac{3^n}{\sqrt{\pi n}}$
[**B**] $k_n \sim \sqrt{2}\frac{2^n}{n}$

Q4. Soit $A(z)$ la série génératrice ordinaire d'une classe $\mathcal{A}$. Soit $\mathcal{C}$ la classe des paires ordonnées (α, β) d'éléments de $\mathcal{A}$ telles que $|\alpha| = |\beta|$. Alors la série génératrice ordinaire de $\mathcal{C}$ vérifie :
[**A**] $C(z) = \sum_n (A_n z^n)^2$
[**B**] $C(z) = \left(\sum_n A_n z^n\right)^2$

Q5. Le nombre de compositions de l'entier n en sommants *tous pairs* vaut :
[**A**] $2^{n/2-1}$
[**B**] 2^{n-3}

Q6. 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**] $\sqrt{\frac{1+z}{1-z}}$

Q7. 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 ≥ 2.

Q8. Le nombre de polynômes unitaires irréductibles sur $Z/5Z$ (F_5) vaut asymptotiquement :
[**A**] $5^n/n$
[**B**] $5^n \log n$

Problème 2 : Contraintes d'ordre sur les structures étiquetées.

Une construction qui est bien adaptée à la prise en compte des contraintes d'ordre dans le produit étiqueté $\mathcal{B}\star\mathcal{C}$ est le produit étiqueté dit Min-contraint, qu'on note $\mathcal{B}^{\Box}\star\mathcal{C}$. Ce produit désigne le sous-ensemble du produit étiqueté $\mathcal{B}\star\mathcal{C}$ formé des structures où la plus petite étiquette est toujours dans la composante de $\mathcal{B}$.

Pour certaines applications, il est plus adapté d'imposer des contraintes sur l'étiquette maximale (plutôt que sur l'étiquette minimale). Le produit étiqueté dit Max-contraint est désigné par $\mathcal{B}^{\Diamond}\star\mathcal{C}$ et il possède des propriétés duales du produit étiqueté Min-Contraint.

1) Montrer que le nombre A_n de structures de taille n de la classe $\mathcal{A} = \mathcal{B}^{\Box}\star\mathcal{C}$ vérifie

$$A_n = \sum_{k=1}^{n} \binom{n-1}{k-1} B_k C_{n-k}.$$

Montrer que la série génératrice exponentielle relative au produit $\mathcal{A} = \mathcal{B}^{\Box}\star\mathcal{C}$ est

$$A(z) = \int_0^z \frac{dB(t)}{dt} \cdot C(t)dt.$$

2) *Records dans les permutations.* Etant donnée une permutation de $[1 \ldots n]$ de la forme $\sigma = \sigma(1)\,\sigma(2)\,\ldots\,\sigma(n)$, un record de cette permutation est (par définition) un indice j pour lequel on a $\sigma(k) < \sigma(j)$ pour tous les indices $k < j$. [Remarque : Un tel élément est "supérieur" à tous ses prédécesseurs, d'où le nom...].
On désigne par $\mathcal{P}$ la classe des permutations et par $\mathcal{P}^{(k)}$ la classe des permutations possédant k records.

(a) Si $\mathcal{Z}$ est la classe formée d'une seule structure avec un seul atome étiqueté, décrire la classe

$$\mathcal{Q} := \{\mathcal{Z}\}^{\Diamond}\star\mathcal{P},$$

et montrer que

$$Q(z) = \log\frac{1}{1-z}.$$

(b) Montrer que tout élément de $\mathcal{P}^{(k)}$ peut s'écrire comme une suite non ordonnée de k éléments de $\mathcal{Q}$ (avec re-étiquetage). Décrire la construction [de la classe étiquetée] qui permet de construire $\mathcal{P}^{(k)}$ à partir de $\mathcal{Q}$.

(c) En déduire la série génératrice de la classe $\mathcal{P}^{(k)}$ est

$$\frac{1}{k!}\left(\log\frac{1}{1-z}\right)^k.$$

(d) Avez-vous déjà a vu cette série génératrice quelque part ? Commentez !

3) *Arbres croissants.* Un arbre croissant est un arbre binaire plan enraciné et étiqueté (possiblement vide) dans lequel seuls les noeuds internes portent des étiquettes (avec les contraintes usuelles sur les étiquettes), qui croissent le long des branches à partir de la racine. Les noeuds externes ne portent pas d'étiquettes.

(a) Dessiner un arbre croissant de taille 7.

(b) Montrer que la classe $\mathcal{I}$ des arbres croissants admet la définition récursive suivante

$$\mathcal{I} = \{\epsilon\} + \left(\mathcal{Z}^{\Box} \star \mathcal{I} \star \mathcal{I}\right)$$

(c) En déduire que la série génératrice $I(z)$ satisfait une équation différentielle $I'(z) = I^2(z)$ avec $I(0) = 1$. En déduire que $I_n = n!$.

(d) On veut retrouver ce résultat par une preuve "bijective". A une permutation σ élément de $\mathcal{P}$ écrite comme précédemment sous forme de mot $\sigma = \sigma(1)\,\sigma(2)\,\ldots\,\sigma(n)$, on associe sa "factorisation", de la forme $\sigma = \sigma_G \cdot \min(\sigma) \cdot \sigma_D$, où $\min(\sigma)$ est la plus petite étiquette de la permutation, σ_G, σ_D les facteurs gauche et droits de la permutation. Puis l'arbre binaire $\beta(\sigma)$ associé à la permutation est juste l'arbre binaire de racine $\min(\sigma)$ qui a pour sous-arbre gauche $\beta(\sigma_G)$ et comme sous-arbre droit $\beta(\sigma_D)$.
Montrer que l' application β definit une bijection de $\mathcal{P}$ dans $\mathcal{I}$ qui respecte la taille.

4) *Permutations Alternantes.* On désigne par $\mathcal{J}$ la classe des arbres croissants qui ne possèdent pas de noeud à un fils (tout noeud interne a donc 0 ou 2 fils). Remarquons qu'un tel arbre a un nombre impair de noeuds internes.

(a) Montrer que la classe $\mathcal{J}$ admet la définition récursive suivante

$$\mathcal{J} = \mathcal{Z} + \left(\mathcal{Z}^{\Box} \star \mathcal{J} \star \mathcal{J}\right)$$

(b) En déduire que la série génératrice $J(z)$ vérifie une équation différentielle $J'(z) = 1 + J^2(z)$ avec $J(0) = 0$. En déduire que $J(z) = \tan(z)$

Une permutation est alternante si

$$\sigma(1) > \sigma(2) < \sigma(3) > \sigma(4) \ldots\ldots$$

On considère ici seulement les permutations d'ordre impair. Dans ce cas, on a $\sigma(1) > \sigma(2) < \sigma(3) > \sigma(4) \ldots\ldots < \sigma(n)$. On désigne par $\mathcal{A}$ la classe des permutations alternantes d'ordre impair.

(c) Montrer que l'image par β de la classe $\mathcal{A}$ est la classe $\mathcal{J}$. Qu'en déduit-on pour le nombre des permutations alternantes d'ordre impair ?

Problême 3 : Arbres compacts

Dans tout le problème, on considère des arbres binaires dont les sommets sont soit des feuilles (degré 0) soit des sommets internes ayant deux descendants (degré 2). On appellera taille d'un arbre son nombre de sommets internes. (Ainsi la feuille est-elle de taille 0).
Le nombre de cellules occupées, appelé encombrement de l'arbre, pour un arbre de taille n est donc $2n+1$. Certaines cellules étant identiques, il devient loisible de les partager pour gagner de la place. Par exemple on peut partager toutes les feuilles qui sont identiques, en n'en gardant qu'un seul exemplaire ; puis on peut itérer le processus pour les sous-arbres de taille 1, 2, etc.(voir exemples au tableau). On diminue ainsi l'encombrement de la représentation de l'arbre initial en le représentant comme un graphe dirigé acyclique (dag). On appelle coefficient de compactage le rapport entre l'encombrement de l'arbre initial et celui du dag qui le représente.

Q1. Montrer par des exemples commentés que ce rapport peut varier de $1/2$ à $n/2^n$.

On va maintenant étudier le coefficient de compactage moyen lorsqu'on fixe la hauteur maximale h des sous-arbres partagés.

Q2. On considère le cas $h = 0$; Construire la s.g.o. des encombrements des arbres binaires compactés par les feuilles. En déduire leur coefficient de compactage.

Q3. Exprimer les arbres binaires en fonction de leurs sous-arbres terminaux de type *feuille* —$\square$— et de taille 1 —$\circ(\square, \square)$. En déduire la s.g.o. des encombrements des arbres binaires compactés à hauteur 1. En déduire leur coefficient de compactage.
Q4. Comment procéder pour le faire en compactant à la hauteur $h = 2$? Indiquer la méthode, et calculer si possible le coefficient de compactage. Peut-on aller plus loin et comment ?
Q5. On considère maintenant les familles $\mathcal{B}_{\leq h}$ des arbres binaires de hauteur au plus h. Exprimer les récurrences qui permettent de calculer les s.g.o. de ces familles. Quel est leur rayon de convergence ?
Q6. Lorsqu'on compacte un arbre binaire jusqu'à la hauteur h, le compacté ne contiendra pas forcément tous les arbres de $\mathcal{B}_{\leq h}$, mais certainement un certain nombre. Peut-on ainsi encadrer l'encombrement d'un arbre compacté à hauteur h ? Il faudra introduire, outre la s.g.o. des arbres de hauteur au plus h, celle du nombre de sommets des arbres binaires de hauteur au moins h.
Q7. Calculer la probabilité popur qu'un sommet dans un arbre binaire de taille n soit à hauteur au moins h. En déduire un encadrement pour le taux de compactage à hauteur h.
Q8. Déduire de ces bornes que le taux moyen de compactage (pour h quelconque) tend vers 0, quand la taille des arbres tend vers l'infini.
Q9. Pourrait-on estimer de la même façon les longueurs de cheminement des arbres et de leurs compactés ?

Problème 4 : Temps d'attente d'un motif

On étudie ici le temps d'attente d'un mot donné H dans une séquence de taille n, où n vaut typiquement de 8000 à 20000.

(a) Une approximation courante pour la probabilité de trouver un mot H dans une séquence de taille n est :

$$F_n(\mathrm{H}) \sim 1 - (1 - P(\mathrm{H}))^{n-m+1} \tag{1}$$

où m est la taille du mot et $P(\mathrm{H})$ la probabilité de H dans un modèle de Bernoulli. En effet, la probabilité de ne pas trouver H en une position quelconque du texte est $1-P(\mathrm{H})$, et $n-m+1$ positions sont possibles. En considérant que toutes ces positions sont indépendantes, on obtient (1).
La série génératrice des textes contenant au moins une fois H est

$$F_{\mathrm{H}}(z) = \frac{P(\mathrm{H})z^m}{(1-z)A(z) + P(\mathrm{H})z^m} \frac{1}{(1-z)} \tag{2}$$

où A est le polynôme d'autocorrélation. On a : $F_n(\mathrm{H}) = [z^n]F_{\mathrm{H}}(z)$ et on sait que $F_n(\mathrm{H}) \sim \alpha\rho^{-n}$ où ρ est la plus petite racine positive de $(1-z)A(z) + P(\mathrm{H})z^m = 0$. Calculer une approximation de ρ par bootstrapping. Montrer qu'elle coïncide avec (1) si $A(z) = 1$, c.a.d. si le motif est non-chevauchant.

(b) On généralise (1) à un ensemble de motifs $\mathcal{H}$. Une chaine recouvrante est une suite de mots de $\mathcal{H}$ se recouvrant de manière non triviale (au moins une lettre). Par exemple, si $\mathcal{H} = \{AACAA, AATCA\}$, le mot
$AACAATCAACAACAA$ est une chaine recouvrante, mais $AACAAAACAA$ ou $AACAAATTAACAAA$ ne le sont pas. On montre que la série génératrice des textes contenant au moins une fois un élément de $\mathcal{H}$ est

$$F_{\mathcal{H}}(z) = \frac{1}{1-z} - \frac{1}{1 - z + \sum_{w \text{ chaine recouvrante}} P(w)z^{|w|}}$$

On note $P(\mathcal{H}) = \sum_{w \in \mathcal{H}} P(w)$. On pourra vérifier que cette expression coincide avec (2) lorsque $\mathcal{H}$ se réduit à un mot. On généralisera l'approximation de (a) en calculant la racine ρ de $(1-z) + \sum_{w \text{ chaine recouvrante}} P(w)z^{|w|} = 0$.