D.E.A. ALGORITHMIQUE

Filière "Combinatoire, algorithmique, modèles et réseaux"
Cours "Modèles aléatoires en combinatoire et réseaux"

Alain Jean-Marie <ajm@lirmm.fr>, Philippe.Flajolet@inria.fr
Voir: `http://algo.inria.fr/flajolet/Teach`
et `http://www.lirmm.fr/~ajm/Cours/03-04/DEA_ALGO`
— 1er mars 2004 —-

Cette épreuve est à faire en 3 heures 30 minutes en tout. (Ceux qui souhaitent compléter leurs réponses au-delà des 3h30m pourront aussi fournir des feuillets supplémentaires marqués comme tels et pourront obtenir quelques points supplémentaires.) Chaque problème compte pour la moitié de la note. La note maximale $\frac{20}{20}$ est attribuable à une copie qui traiterait correctement environ les trois quarts de l'ensemble: vous pouvez donc choisir (un peu)!

Les notes de cours et autres documents peuvent être utilisés.

Pour toute correspondance électronique, merci d'indiquer clairement le champ "`Subject: DEA ALGO`". Envoyez vos réponses au plus tard le *15 mars 2004* à

Alain Jean-Marie,
LIRMM-Université de Montpellier 2,
161 Rue Ada,
34392 Montpellier Cedex 5.

Gardez impérativement une copie personnelle de votre envoi.

Il est demandé de mettre en relief les idées plutôt que les détails de calcul. Plus généralement on se limitera à la description **succinte** des **principales** étapes logiques.

Bonne chance!

Problème I

Combien d'ancêtres aviez vous il y a 1000 ans?

L'objectif de ce problème est d'étudier différents modèles permettant d'estimer $N(t)$, le nombre d'ancêtres qu'un individu avait t années dans le passé.

Pour cela, on considère un *Arbre généalogique*, comme dans la Figure 1. Les nœuds de cet arbre représentent les ancêtres d'un individu, et les arcs les relations de filiation. À chaque nœud correspond une « profondeur » qui est la date de naissance de l'ancêtre. La longueur de l'arc, c'est-à-dire la différence entre les dates de ses extrémités, est l'âge qu'avait l'ancêtre quand est né son descendant direct. Nous dénommerons cette variable par τ. Les deux arcs « en dessous » d'un nœud représentent les deux parents de cet ancêtre.

On suppose que la racine de l'arbre est l'individu considéré, avec la date 0.

Nous allons supposer que les longueurs τ des arcs sont des variables stochastiques *indépendantes*. L'arbre lui-même est donc aléatoire.

Dans tout le problème, $N(t)$ est le nombre d'ancêtres au temps t, c'est à dire le nombre d'arcs de l'arbre qui intersectent une droite horizontale « juste après » la date t (voir la Figure). La condition initiale est donc toujours $N(0) = 2$.

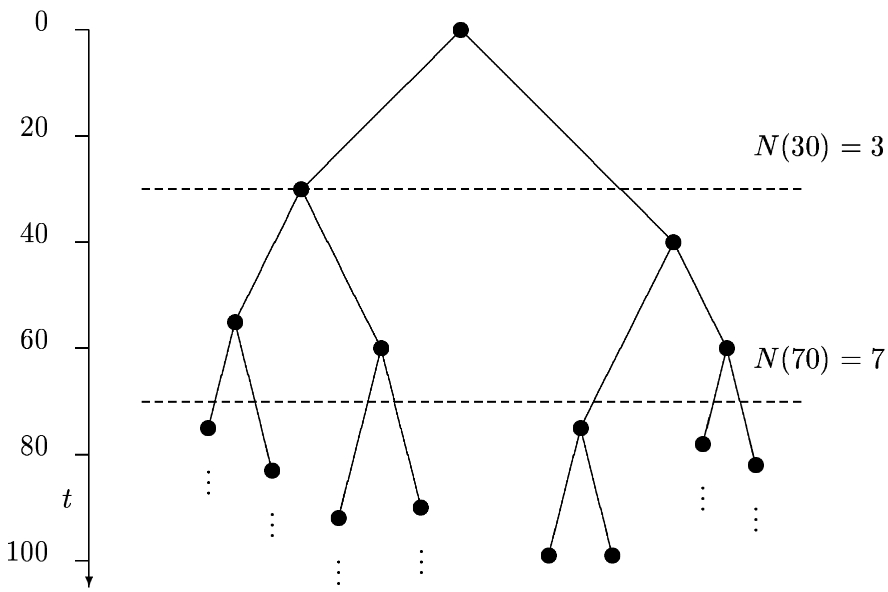

FIGURE 1. Un arbre généalogique

Soient les notations:

$$m(t) = \mathbb{E}N(t), \qquad \sigma^2(t) = \mathbb{E}N^2(t) - m^2(t), \qquad c(t) = \frac{\sigma^2(t)}{m^2(t)} = \frac{\mathbb{E}N^2(t)}{m^2(t)} - 1.$$

On va chercher à déterminer la « vitesse de croissance asymptotique » de l'arbre:

$$v = \lim_{t\to\infty} \frac{\log(m(t))}{t} .$$

Modèle 1 Dans ce premier modèle, on suppose que le temps τ est déterministe avec valeur T.

a/ Calculer $N(t)$ pour tout $t \geq 0$.

b/ Calculer la vitesse de croissance asymptotique v_1.

Modèle 2 Dans ce modèle, on suppose que le temps τ a une distribution géométrique de paramètre $1-\alpha$:

$$\mathbb{P}\{\tau = k\} = \alpha(1-\alpha)^{k-1}, \qquad k \geq 1.$$

a/ Calculer $\mathbb{P}\{\tau > k\}$, $k \in N$. Calculer $\mathbb{E}\tau$.

b/ Montrer que la distribution géométrique est « sans mémoire », c'est-à-dire, que

$$\mathbb{P}\{\tau > k + l | \tau > k\} = \mathbb{P}\{\tau > l\}, \qquad \forall k, l \in \mathbb{N} .$$

c/ Montrer que $N(t)$ est en fait dans ce cas un processus de Galton-Watson pour une certaine distribution du nombre de descendants, qu'on précisera.

d/ En déduire une récurrence sur G_t, la fonction génératrice de la variable $N(t)$: $G_t(z) = \mathbb{E}(z^{N(t)})$. Calculer $m(t)$ et $c(t)$.

e/ Déterminer la vitesse de croissance asymptotique v_2.

Modèle 3 Dans ce modèle, on suppose que le temps τ est fixe, mais est différent pour les hommes et les femmes. On suppose que pour les femmes, $\tau = a$ et pour les hommes, $\tau = b$ (a et b sont des entiers strictement positifs).

a/ Établir une relation récursive pour $N(t)$.

b/ Soit $G(z)$ la série génératrice $\sum_{t=0}^{\infty} N(t)z^t$. Montrer que $G(z)$ converge au voisinage de l'origine. En utilisant a/, calculer $G(z)$.

c/ Calculer la vitesse de croissance asymptotique v_3 dans le cas $a = 20$, $b = 30$.

Modèle 4 Dans ce modèle on suppose que le temps τ a des distributions différentes pour les hommes et pour les femmes. On suppose que pour femmes, la distribution de τ est géométrique avec paramètre α_f, et pour hommes, est géométrique avec paramètre α_h. Soient $N_f(t)$ et $N_h(t)$ les nombres d'ancêtres femmes et hommes, respectivement, au temps t.

a/ Montrer que $\{(N_f(t), N_h(t)), t \in \mathbb{N}\}$ est un processus de Galton-Watson multi-type.

b/ Soit $G_t(x, y) = \mathbb{E}(x^{N_f(t)} y^{N_h(t)})$, la fonction génératrice jointe des variables $N_f(t)$ et $N_h(t)$. Trouver une formule de récurrence pour G_t. Que vaut la fonction $G_0(x, y)$?

c/ Soient $m_f(t) = \mathbb{E}(N_f(t))$ et $m_h(t) = \mathbb{E}(N_h(t))$. Montrer que le vecteur

$$\begin{pmatrix} m_f(t) \\ m_h(t) \end{pmatrix}$$

satisfait une récurrence linéaire.

d/ Calculer la vitesse de croissance asymptotique v_4.

Modèle 5 Dans ce modèle, on essaie de prendre en compte le fait que les ancêtres ne sont pas tous distincts. On suppose que la variable τ a une distribution géométrique (comme dans le modèle 2), mais que l'évolution de N est:

$$N(t+1) \;=\; N(t) \;+\; \sum_{i=1}^{M(t)} B_{t,i},$$

où $M(t)$ a une distribution binomiale de paramètres $N(t)$ et δ. C'est-à-dire que, <u>sachant que</u> $N(t) = n$,

$$\mathbb{P}\{M(t) = k\} \;=\; \binom{n}{k} \delta^k (1-\delta)^{n-k}.$$

a/ Interpréter ce nouveau modèle.

b/ Calculer G_{t+1} en fonction de G_t. Est-ce que ce modèle est réellement différent du modèle 2?

Facultatif Calculer $m(1000)$ pour ces modèles (Un système de calcul formel pourra être utile...). Valeurs: $T = 25$ pour le modèle 1, $\mathbb{E}\tau = 25$ pour le modèle 2, $b = \mathbb{E}\tau_f = 20$ et $a = \mathbb{E}\tau_h = 30$ pour les modèles 3 et 4. Un commentaire sur la validité de ces modèles?

Problème II

Variations combinatoires et analyse de singularités

On examine dans ce problème divers modèles combinatoires qui se traitent asymptotiquement par l'analyse de singularités. Les groupes de questions Q1–Q2, Q3–Q4, Q5–Q8 sont largement indépendants.

Q1. Compositions binaires. On appelle composition binaire d'un entier n une *suite* $(n_1, n_2, \ldots, n_k)$ d'entiers (k quelconque) telle que la somme des n_j vaut n et telle que chaque n_j appartienne à l'ensemble $\mathcal{B} = \{1, 2, 4, 8, \ldots\}$ des puissances de 2. La suite des nombres de compositions binaires commence par $1, 1, 2, 3, 6$ pour $n = 0, 1, 2, 3, 4$. Soit $\mathcal{C}$ la classe des compositions binaires, $C(z)$ sa fonction génératrice ordinaire, et $C_n = [z^n]C(z)$.

Établir brièvement que le rayon de convergence ρ de $C(z)$ vérifie *a priori* l'inégalité $\frac{1}{2} \leq \rho \leq 1$. Exprimer $C(z)$ en fonction de la série

$$B(z) := \sum_{j=0}^{\infty} z^{2^j}.$$

En déduire que ρ vérifie $\frac{1}{2} < \rho < 1$ et qu'il existe une formule asymptotique

$$C_n \sim \lambda \rho^{-n} \qquad (n \to \infty),$$

pour une certain $\lambda > 0$ (qu'il n'est pas nécessaire de calculer).

Q2. Justifier brièvement le fait que la série génératrice double

$$C(z, u) := \frac{1}{1 - uB(z)}$$

est telle que $[z^n u^k]C(z, u)$ compte le nombre de compositions binaires de taille n ayant k sommants. Montrer que

$$[z^n] \left(\frac{\partial}{\partial u} C(z, u) \right)_{u=1} \sim \mu n \rho^{-n},$$

pour une certain $\mu > 0$. Que peut-on en déduire quant à l'espérance de la variable aléatoire "nombre de sommants" prise sur l'ensemble des compositions binaires de taille n munies de la distribution uniforme?

Q3. Partitions binaires. On appelle partition binaire d'un entier n un *multiensemble* $\{n_1, n_2, \ldots, n_k\}$ d'entiers (k quelconque) tel que la somme des n_j vaut n et tel que chaque n_j appartienne à l'ensemble $\mathcal{B} = \{1, 2, 4, 8, \ldots\}$. Les nombres correspondants sont $1, 1, 2, 2, 4$ pour $n = 0, 1, 2, 3, 4$. On *fixe* alors un entier $\ell > 0$ et l'on impose que chaque n_j vérifie $n_j < 2^\ell$. Soit $\mathcal{P}^{[\ell]}$ la classe correspondante et $P^{[\ell]}(z)$ sa fonction génératrice ordinaire.

Exprimer $P^{[\ell]}(z)$ comme une fraction rationnelle, puis en déduire un équivalent asymptotique de $[z^n]P^{[\ell]}(z)$.

Estimer asymptotiquement le nombre moyen de sommants dans une partition de $\mathcal{P}^{[\ell]}$ ayant taille n. (Il peut être commode de traiter un cadre plus général.)

Q4. On reprend la question précédente en imposant de plus que tous les sommands n_j soient distincts. Soit $Q^{[\ell]}(z)$ la fonction génératrice ordinaire correspondante. Déterminer un équivalent asymptotique de $[z^n]Q^{[\ell]}(z)$. Montrer et commenter dans ce contexte l'identité

$$\frac{1}{1-z} = (1+z)(1+z^2)(1+z^4)(1+z^8)\cdots.$$

Q5. Arbres généraux. On considère la classe $\mathcal{G}$ des arbres plans "généraux" (au sens que tous les degrés sont permis) et non étiquetés. Commenter très brièvement les raisons des faits suivants: la série génératrice ordinaire vérifie

$$G(z) = \frac{z}{1-G(z)}, \qquad G(z) = \frac{1}{2}\left(1-\sqrt{1-4z}\right),$$

et $G_n \equiv [z^n]G(z)$ vaut

$$G_n = \frac{1}{n}\binom{2n-2}{n-1} \sim \frac{4^{n-1}}{\sqrt{\pi n^3}}.$$

Soit $G^{[k]}(z)$ la série des arbres dont le degré de la racine vaut k (au sens où la racine possède k descendants). Déterminer $G^{[0]}(z), G^{[1]}(z), G^{[2]}(z)$ directement, et donner la forme générale de $G^{[k]}(z)$.

Quels paramètres le degré de la racine pris sur $\mathcal{G}$ induit-il sur les chemins et sur les arbres binaires associés lorsqu'on utilise les bijections standard?

Q6. On rappelle que d'après l'analyse de singularités, on a

$$[z^n] - \lambda\sqrt{1-z} \sim \frac{\lambda}{\sqrt{\pi n^3}},$$

cette estimation restant valable pour les fonctions ayant ce même type de comportement singulier (cf cours). Retrouver de cette manière la forme asymptotique de G_n donnée en (Q5). Montrer par un calcul simple qu'à k fixé, on a $(n \to \infty)$

$$\frac{G_n^{[k]}}{G_n} \sim C(k-1)\omega^k,$$

pour des constantes C, ω qu'on déterminera.

Quel est asymptotiquement le degré moyen à la racine d'un arbre de grande taille? (On pourra utiliser la série génératrice double $z/(1-uG(z))$). Ce résultat est-il logiquement impliqué par l'estimation asymptotique de $G_n^{[k]}/G_n$?

Q7. Quelle est la probabilité qu'un arbre de $\mathcal{G}$ possède k sommets sur sa branche (issue de la racine) la plus à gauche? (Ici, on prend encore k fixé et $n \to \infty$). Qu'en est-il de la branche la plus à droite?

Q8. On considère désormais la famille $\mathcal{H}$ des arbres de $\mathcal{G}$ dont les degrés des sommets sont contraints à être dans l'ensemble $\{0\} \cup \mathcal{B}$ où, de nouveau, $\mathcal{B} = \{1, 2, 4, 8, \ldots\}$. Justifier rapidement le fait que le nombre H_n des tels arbres de taille n vérifie une formule asymptotique du type

$$H_n \sim D\xi^n n^{-3/2}.$$

Déterminer asymptotiquement la probabilité qu'un arbre de taille n dans $\mathcal{H}$ ait une racine de degré 2^k pour $k \geq 0$ fixé quelconque.