DEA d'ALGORITHMIQUE 1998-1999

Analyse d'algorithmes: Modèles combinatoires

Philippe.Flajolet@inria.fr
http://algo.inria.fr/flajolet/Teach/

Partie I. Séries de Dirichlet

On explore ici quelques propriétés des séries de Dirichlet, lesquelles sont à la base de la théorie analytique des nombres. Seule l'inversion de Moebius, équation (3), est nécessaire à la Partie II du problème.

Q1. La fonction *zeta* de Riemann est définie par

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

Déterminer le meilleur s_0 tel que la somme converge pour tout s complexe tel que $\Re(s) > s_0$. ($\Re$ = partie réelle.) Que vaut $\lim_{s \to s_0^+} \zeta(s)$?

Établir (sans lourd détail analytique) la formule dite du produit Eulérien,

$$\zeta(s) = \prod_p \frac{1}{1 - 1/p^s}, \qquad (1)$$

valable pour $s > s_0$. (Ici, $\prod_p$ signifie un produit étendu aux valeurs de p premiers, $p = 2, 3, 5, 7, \ldots$.)

Q2. Utiliser la question précédente pour établir par un argument tres simple d'analyse qu'il existe une infinité de nombres premiers. (Cette preuve ingénieuse du théorème d'Euclide est due à Euler.)

Note, Les propriétés fines de $\zeta(s)$ sont à la base du théorème des nombres premiers (Riemann, puis Hadamard et de la Vallée-Poussin) d'après lequel le nombre de nombres premiers inférieurs à x, soit $\varpi(x)$, est estimé par:

$$\varpi(x) = \frac{x}{\log x}(1 + o(1)), \qquad (2)$$

En admettant (2), montrer que le nombre de premiers dont la représentation binaire comporte exactement n bits est asymptotique à

$$c_1 \frac{2^n}{n},$$

et déterminer c_1.

Q3. Montrer que

$$\frac{1}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s},$$

où ne prend que les valeurs $0, \pm 1$; précisément: $\mu(n) = 0$ dès que n est divisible par un carré (> 1), et

$$\mu(p_1 p_2 \cdots p_r) = (-1)^r,$$

les p_j étant des premiers distincts. La suite des valeurs $\{1, -1, -1, 0, -1, 1, -1, \ldots\}$ est appelée suite de Moebius.

Q4. Soit $\{a_n\}$ une suite de complexes. On appelle *série génératrice de Dirichlet* associée à cette suite la somme

$$\alpha(s) := \sum_{n \geq 1} a_n n^{-s}.$$

Étant donné trois suites a_b, b_n, c_n de séries de Dirichlet associées $\alpha(s), \beta(s), \gamma(s)$. Montrer l'équivalence

$$\alpha(s) = \beta(s) \cdot \gamma(s) \quad \Longleftrightarrow \quad a_n = \sum_{d \mid n} b_d c_{n/d},$$

où la somme est étendue à tous les diviseurs d de n, ce qui est noté $d \mid n$.

En déduire la *formule d'inversion de Moebius*: pour deux suite f_n, g_n quelconques, on a l'équivalence

$$f_n = \sum_{d \mid n} g_d \quad \Longleftrightarrow \quad g_n = \sum_{d \mid n} \mu(d) g_{n/d}. \tag{3}$$

Inverser la relation entre séries génératrices ordinaires:

$$f(z) = \sum_{k \geq 1} g(z^k),$$

en exprimant $g(z)$ en fonction de $f(z)$. On suppose g sans terme constant: $g(0) = 0$.

Q5. [Facultatif] Soit $\Omega = \{\omega_n(z)\}$ un système de fonctions. Les cas

$$\Omega_{OGF} = \{z^n\}_{n=0}^{\infty}, \quad \Omega_{EGF} = \{\frac{z^n}{n!}\}_{n=0}^{\infty}, \quad \Omega_{DGF} = \{n^{-z}\}_{n=1}^{\infty}$$

donnent lieu aux séries ordinaires, exponentielles, et de Dirichlet. On definit la fonction ζ et les coefficients de Moebius par

$$\zeta_\Omega(z) := \sum_n 1 \cdot \omega_n(z), \quad \frac{1}{\zeta_\Omega(z)} := \sum_n \mu_\Omega(n) \omega_n(z).$$

Remplir une matrice 4×3 dont les colonnes correspondent à OGF, EGF, DGF, et les lignes donnent: (i) la fonction zeta, (ii) les coefficients de Moebius, (iii) la relation entre coefficients qui exprime le produit de séries, (iv) la relation d'inversion de type Moebius.

Pouvez-vous interpréter

$$[z^n]\frac{1}{2 - \frac{1}{1-z}}, \quad [z^n]\frac{1}{2 - e^z}, \quad [n^{-z}]\frac{1}{2 - \zeta(z)},$$

où l'on utilise la généralisation naturelle de la fonction coefficient $[\cdot]$? (Une telle théorie transverse a été développée par Rota.)

Partie II. Polynômes

Cette partie du problème montre que des objets algébriques peuvent se traitent par les méthodes de la combinatoire énumérative. Les développements qui suivent sont à la base de l'analyse d'algorithmes probabilistes en calcul formel et en théorie des codes.

Rappel. Pour un nombre premier p, on considère le corps $F = F_p$ formé par les entiers pris modulo p. On travaille dans l'anneau $F[x]$ des polynômes à coefficients dans F. On rappelle que, dans cet anneau, il existe une division euclidienne et une propriété de factorisation unique. Les élements de $F[x]$ qui sont indécomposables (l'équivalent des nombres premiers) sont appelés polynômes irreductibles: f est irréductible si $f = g \cdot h$ implique $g = 1$ ou $h = 1$.

Fixons p. On désigne par $\mathcal{P}$ la classe des polynômes unitaires (dont le coefficient de tête vaut 1) et par $\mathcal{I}$ la sous-classe des polynômes unitaires et irréductibles. Par convention, la taille (au sens combinatoire) d'un polynôme est son degré (au sens algébrique). La sous-classe $\mathcal{Q} \subset \mathcal{P}$ représente les polynômes unitaires "sans carrés", c-à-d, non divisibles par un carré: f est sans carré si $f = g \cdot h^2$ implique $h = 1$.

Q1. Montrer que la série génératrice de $\mathcal{P}$ vaut

$$P(z) = \frac{1}{1 - pz}.$$

Observer que tout $f(x) \in F[x]$ s'écrit $f(x) = g(x) \cdot h(x)^2$ où g est sans carré et h quelconque. D'autre part, sii $\mathcal{H}$ est une classe de polynômes de série génératrice $H(z)$, la série génératrice de l'ensemble $\{h(x)^2 \ / \ h(x) \in \mathcal{H}\}$ vaut $H(z^2)$. En déduire la série génératrice de $\mathcal{Q}$. Calculer la probabilité qu'un polynôme pris uniformément dans $\mathcal{P}_n$ soit sans carré.

Q2. Montrer que

$$\begin{aligned} P(z) &= \prod_{n=1}^{\infty} (1 - z^n)^{-I_n} \\ &= \exp\left(\sum_{k=1}^{\infty} \frac{1}{k} I(z^k) \right). \end{aligned}$$

Exprimer la relation correspondante entre $zP'(z)/P(z)$ et $J(z) = zI'(z)$. En déduire la relation

$$I_n = \frac{1}{n} \sum_{d \ | \ n} \mu(d) p^{n/d},$$

où μ est le coefficient de Moebius.

Q3. Montrer brièvement que

$$I_n = \frac{p^n}{n}(1 + o(1)).$$

Comment se compare la situation des polynômes sur un corps fini à celle des entiers?

On suppose qu'on dispose de deux procédures: `draw(n)` tire au hasard un élément de $\mathcal{P}_n$ et `irred(f)` répond vrai ou faux selon que f est irréductible ou non.

Combien d'essais en moyenne effectue l'algorithme de tirage d'un polynôme irréductible suivant?

```
repeat
    f := draw(n);
until irred(f);
```

Quelle est la variance du nombre de tirages?

Q4. Soit $\omega(f)$ le nombre de facteurs irréductibles de f comptés avec multiplicité, par exemple,

$$\omega\Big((1+x)(3+x+x^3)^5(4+x^7)^2(2+2x+x^9)\Big) = 1+5+2+1 = 9.$$

Soit $P_{n,k}$ le nombre de $f \in \mathcal{P}_n$ tels que $\omega(f) = k$ et $P(z,u) := \sum_{n,k} P_{n,k} u^k z^n$. Montrer que

$$\begin{aligned} P(z,u) &= \prod_{n=1}^{\infty} (1-uz^n)^{-I_n} \\ &= \exp\left(\sum_{k=1}^{\infty} \frac{u^k}{k} I(z^k)\right). \end{aligned}$$

Montrer que l'espérance de ω prise sur l'ensemble des polynômes $\mathcal{P}_n$ vaut exactement

$$M_n = \mathrm{coeff}[z^n] \left(\frac{\partial}{\partial u} P\left(\frac{z}{p}, u\right)\right)_{u=1}.$$

Déterminer un équivalent asymptotique de

$$[z^n]\left(I(\frac{z}{p}) + I(\frac{z^2}{p^2}) + I(\frac{z^3}{p^3}) + \cdots\right)$$

et en déduire que

$$M_n = \log n(1+o(1)).$$

(Se limiter aux grandes étapes de la démonstration.)

[Facultatif] Qu'est-ce que ce résultat suggère quant au nombre moyen de facteurs premiers d'un entier $\leq x$?

Q5. [Facultatif] Montrer que le nombre d'éléments de $\mathcal{P}_n$ qui ont n_1 facteurs irréductibles de degré 1, n_2 de degré 2, etc, vaut (en marquant la dépendance en p):

$$G(n_1, n_2, \ldots; p) = \prod_{j=1}^{n} \binom{I_{n_j} + n_j - 1}{n_j}.$$

Les I_n sont bien sûr ceux qui sont relatifs à la caractéristique p du corps: $I_n = I_n(p)$. Calculer pour n *fixé* et $p \to +\infty$ la limite

$$g(n_1, n_2, \ldots) = \lim_{p\to+\infty} \frac{G(n_1, n_2, \ldots; p)}{p^n}.$$

Montrer que $g(n_1, n_2, \ldots)$ est aussi le nombre de permutations de n qui ont n_1 cycles de taille 1, n_2 cycles de taille 2, etc. En résumé la décomposition en facteurs irréductibles d'un polynôme de degre fixé n sur un corps de grande cardinalité ressemble à la décomposition en cycles d'une permutation de la même taille n.

Partie III. Explorations de calcul formel

Il s'agit là d'un exercice à préparer en binome. Le résultat peut fort bien être une feuille de travail Maple, mais éditée convenablement et brièvement commentée, ou bien un "compte-rendu" d'expérience manuscrit.

Le problème traité ici est une simplification d'une situation qui s'est présentée réellement il y a quelques années. Les statisticiens construisent des arbres de classification appelés hiérarchies. *Le problème consiste à analyser les propriétés des hiérarchies aléatoires sur la base du principe qu'une classification produite dans le monde réel (par ex. par un programme d'analyse statistique) a d'autant plus de "chances" d'être significative qu'elle s'écarte des lois du hasard. D'où l'intérêt de quantifier l'aléa!*

Les hiérarchies sont des arbres dont les sommets externes sont les n objets à classifier (la taille vaut alors n) et dont les sommets internes (représentant les "classes") ont degré tous ≥ 2, car il faut que la classification soit non triviale. Par exemple:

```
C=chat, H=homme, L=lion, O=orang-outang, T=tigre, V=vache

              .               n = 5 (nombre de sommets externes)
             /|\
            / | \             3 sommets internes (classes)
           .  V  .
          /|\   /\
         T C L  H O
```

Les hiérarchies sont donc "non-planaires" (pas de distinction gauche/droite) et enracinées, de par l'origine du problème. Les objets à classifier sont assimilés aux étiquettes $1, \ldots, n$. En d'autres termes:

> *La classe $\mathcal{H}$ des hiérarchies est la classe des arbres enracinés non-planaires, étiquetés aux sommets externes seulement (qui déterminent la taille) et dont les sommets internes ont tous degrés ≥ 2.*

Symboliquement:

$$\mathcal{H} = \mathcal{Z} + \texttt{Set}(\mathcal{H}, \text{card} \geq 2).$$

III.1 Dénombrement asymptotique des hiérarchies

Q1. On commence par se convaincre que la série génératrice exponentielle des hierarchies vérifie l'équation

$$H(z) = z + \left(e^{H(z)} - 1 - H(z)\right).$$

Maple sait trouver pour nous une solution au moyen de la fonction "spéciale" `LambertW` (W) qui est la solution de

$$z = We^{W}.$$

définie et positive pour $z \in [0, e^{-1}]$. (En terme de la fonction d'arbre du cours, $T(z)$, on a $W(z) = -T(-z)$; essayer `series(LambertW(z),z=0,10);`)

```
eqH:=H=z+exp(H)-1-H; Hz:=solve(eqH,H);
```

Calculer les 30 premiers coefficients

```
Hs:=series(Hz,z=0,31);
for i from 0 to 30 do
    c[i]:=coeff(Hs,z,i)*1.0 od;
```

et deviner une loi très approximative de croissance de

$$c_i = [z^i]\, H(z).$$

On pourra s'aider de

```
plot([seq([i,c[i+1]/c[i]],i=5..30)],0..30,0..3);
```

Le nombre de hiérarchies de taille n vaut $H_n = n!c_n$.

Q2. On sait que le rayon de convergence d'une série entière à coefficients positifs est égal à la distance entre l'origine et la singularité positive la plus proche. Ici, la singularité correspond à une pente verticale (de dérivée infinie), comme on peut s'en convaincre par un `plot`.

```
plot(Hz,z=0..0.5); plot(diff(Hz,z),z=0..0.5);
H1z:=normal(diff(Hz,z));
```

Par une suite de tracés de ce type, on devra obtenir une approximation à 0.001 près de la singularité positive,

$$\rho = ??? \pm 0.001.$$

Combien de chiffres environ comporte la représentation décimale de H_{1000}?

Q3. On se propose de déterminer maintenant exactement la singularité positive ρ. En supposant qu'elle provient de l'annulation du denominateur de $\frac{d}{dz}H(z)$, on déterminera alors sa valeur exacte (utliser `solve`):

$$\rho = ???$$

On peut alors tenter de comprendre de quelle manière $H(z)$ devient singulière lorsque z tend vers ρ par valeurs inférieures. Il est commode de poser $z = \rho - h$ et faire $h \to 0^+$, ce qui s'obtient par un développement en série.

```
map(simplify,series(subs(z=rho-h,Hz),h=0));
```

On suppose que c_n est asymptotiquement equivalent a ce que l'on obtient en ne retenant que les deux premiers termes, soit

$$c_n \sim [z^n]\left(c_0 - c_1\sqrt{\rho - z}\right).$$

En d'autres termes, on suppose, et cela est justifié par l'analyse de singularité que: *une bonne approximation de la fonction donne une bonne approximation des coefficients.* En déduire une formule asymptotique pour c_n de la forme

$$c_n \sim K_1 K_2^n n^{K_3}, \quad K_1, K_2, K_3 = ???.$$

Vérifier comme suit:

```
for i from 2 to 30 do
        evalf(coeff(Hs,z,i)/rho^(-i+1/2)*i^(3/2)*sqrt(Pi))
od;
```

Avec quelle précision estimez-vous pouvoir prévoir la valeur de H_{1000}?

III.2 Nombre d'étapes de classification. On s'intéresse au nombre de sommets internes de l'arbre (le nombre de classes). Soient $H_{n,k}$ le nombre de hiérarchies de taille n ayant k sommets internes et n sommets externes. On se convaincra (ou bien, on admettra) que la série double

$$H(z,u) = \sum_{n,k} H_{n,k} u^k \frac{z^n}{n!}$$

vérifie l'equation

$$H(z,u) = z + u\left(e^{H(z,u)} - 1 - H(z,u)\right).$$

Q4. Résoudre cette équation en termes de la fonction W de Lambert. On pose

$$G(z) := \left(\frac{\partial}{\partial u} F(z,u)\right)_{u=1}$$

Exprimer de même $G(z)$ (la dérivation s'obtient par `diff`).

Soit $d_n := [z^n]G(z)$, de sorte que le nombre moyen de sommets internes dans une hiérarchie de taille n vaut

$$\mu_n = \frac{d_n}{c_n}.$$

On tabulera μ_n pour $n = 2\,..\,30$.

Deviner une formule approximative décrivant la croissance asymptotique de μ_n.

Chercher a améliorer l'estimation, empiriquement (en devinant simultanément une valeur approchée du terme d'erreur), et/ou en évaluant

$$\gamma := \lim_{z \to \rho^-} \frac{G(z)}{z \frac{d}{dz} H(z)}.$$

Estimer au mieux μ_{1000}.

Un programme de classification produit sur $n = 1000$ données une hierarchie qui comprend 273 classes (sommets internes). La hiérarchie ainsi constituée ressemble-t-elle à une hiérarchie aléatoire?

[ ] OUI   [ ] NON   [ ] INCERTAIN