Université de Nice - M1 IMEA - Méthode de Monte-Carlo

DEVOIR MAISON NO 2
DÉCRYPTAGE PAR ALGORITHME DE METROPOLIS

RENDRE LES PROGRAMMES SOUS FORME DE FICHIER INFORMATIQUE. POUR LES AUTRES QUESTIONS, RENDRE UN DEVOIR PAPIER.

La phrase suivante est tirée d'une œvre littéraire bien connue mais elle a été codée. Le but de ce decoir est de décrypter ce code.

> g,hpobv ,gpochb :chbp ;hgpo'og,hpo"b ? :'sbyo"bos,gpbymbkohgoabhgbox,eebroch ogn'm' vo ?'po :n' yovy bpo gvb : : lbgvro ?'y :'ochb :chbpo gpv'gvpo'mbsohgoe,gp bhyoch opbovy,hm' vo'os,vbo"bo :h ro ? h po :o' : :'opn'ppb, yko"bhuoxbhybpo ? :hpov'y"roabo :boybgs,gvy' o"bog,hmb'hoqo :obv' vobgos ,e ?'lg bo"nhgos'e'y'"bobvo ?'y :' vosx ! !,gp

Les caractères utilisés sont : $\mathcal{A}$ ={'a', 'b', 'c', ...,'x', 'y', 'z', '.', ' ' (espace), '-', "' (guillemet simple), '"' (guillemet double à l'anglaise), ',', ';', '?', '!', ':'}. Cet ensemble est de cardinal 36, on l'identifiera dans la suite à $[36] = \{1, 2, \ldots, 36\}$ (chaque caractère est identifié par son numéro). Le codage consiste en une permutation des caractères, c'est à dire en une application g bijective $[36] \to [36]$. Ainsi, chaque lettre 'a' de la phrase d'origine a été remplacée par la lettre 'g(a)'. Le décodage consiste donc en la recherche de g^{-1}. Vous trouverez tous les fichiers utiles sur http ://math.unice.fr/∼rubentha/cours.html (fichier contenant la phrase codée, des instructions utiles en scilab, la matrice de transition M).

(1) Soit $E = \{f : [36] \to [36]\,,$ bijective$\}$. Soit Q la transition markovienne sur E décrite de la manière suivante.

- Quand on est en $f \in E$. On tire $\{X, Y\}$ suivant la loi uniforme sur $\mathcal{P}_2([36])$ (les parties à deux éléments de [36]).
- On saute en $f^{(X,Y)}$ définie par

$$f^{(X,Y)}(z) = \begin{cases} f(z) & \text{si } z \notin \{X, Y\}\,, \\ f(X) & \text{si } z = Y\,, \\ f(Y) & \text{si } z = X\,. \end{cases}$$

On se donne une matrice de Markov M de taille 36×36 qui représente les probabilités de transition d'un caractère à l'autre quand on regarde les caractères les uns après les autres un texte écrit en français (cette matrice est simplement calibrée sur un volume de : *Les trois mousquetaires*, A. Dumas). Par exemple, $M(2, 1) \approx 0, 14$ parce qu'un 'a' est suivi d'un 'b' dans 14% des cas. Pour des raisons techniques, il faudra remplacer les 0 de la matrice par des nombres très petits.

Notons c_1, c_2, ..., c_{314} la suite des numéros des caractères apparaissant dans la phrase codée. On s'intéresse à la loi π sur E telle que, pour toute f :

$$\pi(f) = \prod_{i=1}^{313} M(f(c_i), f(c_{i+1}))\,.$$

On remarque que plus $\pi(f)$ est grande, plus f est un décodage plausible. Construire une chaîne de Markov $(F_n)_{n\geq 0}$ telle que F_n converge en loi vers π quand n tend vers l'infini (on n'oubliera pas de vérifier les hypothèse nécessaires).

(2) Écrire un programme qui simule la chaîne de Markov ci-dessus.

(3) La chaîne (F_n) se promène dans les points fournissant un décodage plausible. Donner une version plausible de la phrase d'origine, c'est à dire avant qu'elle ait été codée.

Indications :

– Attendre au moins 30000 itérations de la chaîne pour qu'il se passe quelque chose. Relancer plusieurs fois la chaîne si le résultat n'est pas satisfaisant au premier essai.

– Si vous avez des problèmes de division par 0 quand vous calculez des rapports $\pi(f)/\pi(g)$, pensez à calculer $\log(\pi(f))$ et $\log(\pi(g))$ comme des sommes, puis à faire $\exp(\log(\pi(f)) - \log(\pi(g))$.