Processus stochastiques: cryptanalyse

Stegen Thomas s154315
Adrien Minne s154340
Delaunoy Arnaud s153059

1 Première partie : chaines de Markov pour la modélisation du langage et MCMC

1.1 Chaine de Markov pour la modélisation du langage

Question 1

L'élément (i,j) de la matrice de transition correspond à la probabilité de passer de l'état i à l'état j. Il correspond donc à la probabilité que la lettre i soit suivie de la lettre j dans la séquence. Dès lors, soit θ l'élément (i,j) de la matrice de transition, θ est le paramètre d'une loi de Bernouilli avec comme possibilités :

— l'élément i est suivi de j (avec une probabilité θ)
— l'élément i n'est pas suivi de j (avec une probabilité $1-\theta$)

La méthode du maximum de vraisemblance consiste à maximiser $P(\mathbf{D_n}|\theta)$ avec $\mathbf{D_n}$ l'échantillon de donnée, ici seq1 et n le nombre de données. Pour une variable de Bernouilli, on a : Soit,

$$x_i = \begin{cases} 1 & \text{si i est suivi de j} \\ 0 & \text{si i n'est pas suivi de j} \end{cases}$$

et m le nombre d'occurrences de la lettre i.

$$\begin{aligned} P(\mathbf{D_n}|\theta) &= \prod_{i=1}^{m}(x_i\theta + (1-x_i)(1-\theta)) \\ &= \theta^{n_1}(1-\theta)^{n_0} \end{aligned}$$

Avec n_0 le nombre de fois où $x_i = 0$ et n_1 le nombre de fois où $x_i = 1$. Déterminons maintenant le θ maximisant cette fonction :

$$\begin{aligned} \frac{\partial P(\mathbf{D_n}|\theta)}{\partial\theta} &= n_1\theta^{n_1-1}(1-\theta)^{n_0} - n_1\theta^{n_1}(1-\theta)^{n_0-1} \\ &= \theta^{n_1-1}(1-\theta)^{n_0-1}(n_1(1-\theta) - n_0\theta) \\ &= \theta^{n_1-1}(1-\theta)^{n_0-1}(n_1 - \theta(n_1+n_0)) \end{aligned}$$

La valeur de θ maximisant la fonction $P(\mathbf{D_n}|\theta)$ est donc :

$$\theta_{i,j} = \frac{n_1}{n_0+n_1} = \frac{\text{nombre d'occurrences de i suivies de j}}{\text{nombre d'occurrences de i}}$$

Question 2

Question 3

Question 4

Question 5

1.2 Algorithme MCMC

Question 1

Pour prouver que π_0 est une distribution stationnaire de la chaine de Markov, il suffit de prouver que $\pi_0 = \pi_0 * Q$. On sait que les équations de balances détaillées $\pi_0(i)Q_{i,j} = \pi_0(j)Q_{j,i}$

sont satisfaites. Passant en notation indicielle, on doit donc montrer que :

$$\begin{aligned}\pi_0(i) &= \sum_{k=0}^{N} \pi_0(k) * Q_{k,i} \\ &= \sum_{k=0}^{N} \pi_0(i) * Q_{i,k} \\ &= \pi_0(i) \sum_{k=0}^{N} \pi_0(i) Q_{i,k} \\ &= \pi_0(i)\end{aligned}$$

Cette distribution stationnaire est unique si la matrice de transition Q est irréductible.

Question 2

2 Deuxième partie : décryptage d'une séquence codée

Question 1

Question 2

Question 3

Question 4

Question 5