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

mettre la référence du pdf.
Étudions d'abord la probabilité de transition.
La probabilité d'obtenir un élément x_j sachant que l'élément précédent de la chaîne de Markov est x_i est pour $i \neq j$ la probabilité que cet élément soit généré selon la loi q et accepté.

$$P(x_j|x_i) = \alpha(x_j, x_i) q(x_j|x_i)$$

avec

$$\begin{aligned}
\alpha(x_j, x_i) &= min\left\{1, \frac{f(x_j)}{f(x_i)} \frac{q(x_i|x_j)}{q(x_j|x_i)}\right\} \\
&= min\left\{1, \frac{cP_X(x_j)}{cP_X(x_i)} \frac{q(x_i|x_j)}{q(x_j|x_i)}\right\}
\end{aligned}$$

La probabilité d'obtenir à nouveau l'élément x_i sachant que l'élément précédent de la chaîne de Markov est également x_i est la somme de la probabilité que l'élément x_i soit généré selon la loi q et accepté et de la probabilité que tout autre élément soit généré et refusé.

$$P(x_i|x_i) = \alpha(x_i, x_i) q(x_i|x_i) + \sum_k (1 - \alpha(x_k, x_i)) q(x_k|x_i)$$

Dans le cas où l'élément généré est différent du précédent, on a :

$$
\begin{aligned}
P(x_j|x_i)\pi_0(x_i) &= \alpha(x_j, x_i)q(x_j|x_i)\pi_0(x_i) \\
&= min\left\{1, \frac{cP_X(x_j)}{cP_X(x_i)}\frac{q(x_i|x_j)}{q(x_j|x_i)}\right\} q(x_j|x_i)\pi_0(x_i) \\
&= \frac{\pi_0(x_i)}{cP_X(x_i)} min\left\{cP_X(x_i)q(x_j|x_i), cP_X(x_j)q(x_i|x_j)\right\} \\
&\text{en posant } x_i \leftarrow x_j \text{ et } x_j \leftarrow x_i \\
&= \frac{\pi_0(x_j)}{cP_X(x_j)} min\left\{cP_X(x_j)q(x_i|x_j), cP_X(x_i)q(x_j|x_i)\right\} \\
&= min\left\{1, \frac{cP_X(x_i)}{cP_X(x_j)}\frac{q(x_j|x_i)}{q(x_i|x_j)}\right\} q(x_i|x_j)\pi_0(x_j) \\
&= \alpha(x_i, x_j)q(x_i|x_j)\pi_0(x_j) \\
&= P(x_i|x_j)\pi_0(x_j)
\end{aligned}
$$

Dans le cas où l'élément généré est le même que le précédent, x_i étant égal à x_j il est évident que

$$P(x_i|x_j)\pi_0(x_j) = P(x_j|x_i)\pi_0(x_i)$$

car

$$P(x_i|x_i)\pi_0(x_i) = P(x_i|x_i)\pi_0(x_i)$$

2 Deuxième partie : décryptage d'une séquence codée

Question 1

Question 2

Question 3

Question 4

Question 5