de Gibbs. Effectuer 1000 tirages du couple (y, θ) en prenant 1000 valeurs successives après avoir éliminé les premières valeurs afin de perdre la condition initiale. Commenter le résultat. Est-il sensible au nombre de valeurs éliminées ?

TP 2 : Détection d'un changement ponctuel dans une suite binaire – A remettre avant le lundi de la semaine 6. On observe une suite de longueur n, notée y, composée de signaux binaires correspondant à l'émission d'une source de fréquence inconnue θ_1, susceptible de changer en un instant (point) inconnu de θ_1 à θ_2. Les émissions sont indépendantes les unes des autres. Par exemple, cette suite peut se présenter de la manière suivante :

$$y = 01100\ldots00110\|1110011\ldots1100111$$

Dans cette représentation, le symbole $\|$ marque le changement ayant lieu au point c, un indice entier compris entre 1 et n. Par convention, $c = 1$ correspond à la situation où il n'y a pas de changement et les n tirages sont de fréquence égale à θ_2. Nous notons

$$\theta = (c, \theta_1, \theta_2).$$

Pour $c = 2, \ldots, n$, nous avons

$$p(y_i|\theta) = \theta_1^{y_i}(1-\theta_1)^{1-y_i} \quad i = 1, \ldots, c-1$$

et

$$p(y_i|\theta) = \theta_2^{y_i}(1-\theta_2)^{1-y_i} \quad i = c, \ldots, n.$$

Nous supposons que

$$p(\theta) = p(c)p(\theta_1)p(\theta_2) = \frac{1}{n}\,, \quad 0 \leq \theta_1, \theta_2 \leq 1 \; et \; c = 1, \ldots, n.$$

L'objectif de cet exercice est de proposer (et de tester) un algorithme d'apprentissage du point de changement c, permettant le calcul de la loi a posteriori $p(c|y)$.

1. Soit $c = 1$. Donner l'expression de la loi générative $p(y|c = 1, \theta_1, \theta_2)$.
2. Même question pour $c > 1$.
3. On suppose que θ_1 et θ_2 sont connues. Déduire des questions précédentes que
$$\forall c = 2, \ldots, n, \quad \frac{p(c|y)}{p(c = 1|y)} = \prod_{j=1}^{c-1} \frac{\theta_1^{y_i}(1-\theta_1)^{1-y_i}}{\theta_2^{y_i}(1-\theta_2)^{1-y_i}}\,.$$
Vérifier que l'on peut calculer le rapport $p(c|y)/p(c = 1|y)$ pour tout c en effectuant de l'ordre de $n(n-1)/2$ multiplications.
4. Supposant θ_1 et θ_2 connues, proposer un algorithme permettant de calculer $p(c|y)$ pour tout $c = 1, \ldots, n$ avec une complexité en $O(n)$. Indication : mettre à jour le rapport définit dans la question précédente à l'aide d'une récurrence linéaire.
5. On suppose désormais que θ_1 et θ_2 sont inconnues. A partir de valeurs initiales arbitraires, proposer un algorithme itératif, de type échantillonnage de Gibbs, permettant de calculer $p(c|y)$ pour tout $c = 1, \ldots, n$ en combinant la simulation de la loi $p(c|y, \theta_1, \theta_2)$ et la simulation de valeurs des paramètres θ_1 et θ_2.
6. Tester la convergence de votre algorithme en examinant la sensibilité aux conditions initiales choisies arbitrairement.

7. Analyser les jeux de données envoyés en pièce jointe par l'enseignant. Décrire l'incertitude sur le(s) point(s) de changement pour ces jeux de données (localisation des points des changements et intervalles contenant chacun des points avec une probabilité supérieure à 50%).

TP 3 : Modèle de Poisson et mélanges – A remettre avant le lundi de la semaine 9. On souhaite estimer le nombre moyen d'occurrences d'un phénomène donné, correspondant par exemple au nombre de clics journaliers sur un type de produit spécifique dans un site de vente en ligne. Pour cela, on dispose de n observations entières positives ou nulles, notées $y_1, \ldots, y_n$. Au moins l'une de ces observations est non nulle.

Première partie.

1. On suppose les observations indépendantes et de loi de Poisson de paramètre $\theta > 0$. Déterminer la loi générative des données, y, dans ce modèle.

2. On suppose que la loi a priori est non informative

$$p(\theta) \propto \frac{1}{\theta}, \quad \theta > 0 \,.$$

Déterminer la loi a posteriori du paramètre θ. Quelle est l'espérance de la loi a posteriori ?

3. Rappeler le principe de l'algorithme de Metropolis-Hasting. Ecrire dans un programme en `R` une version de cet algorithme pour simuler la loi a posteriori du paramètre θ. On pourra, par exemple, choisir une loi instrumentale exponentielle.