(I) - Problème (Application des chaînes de Markov en communication numérique)

Il est question d'analyser une séquence aléatoire de symboles 0 et 1 par une machine d'état. Cette machine bascule vers l'état "0" après avoir détecté une sequence de longueur w, composée de w fois le symbole 0. L'objectif est de trouver les probabilités de chacun des deux états après n transitions d'une chaîne de Markov modélisant le comportement de la machine. Le travail consiste à montrer la construction de la matrice de transition et déterminer les probabilités du régime stationnaire de la chaîne de Markov supposée homogène dans le temps.

Formulation mathématique du problème (modélisation)

Une séquence aléatoire $(B_n)_{n=0,1,2,...}$ de sympboles 0 et 1 est analysée par une machine d'état. On note la probabilité d'occurrence du symbole 0 à l'instant n par $p(n) = \mathbb{P}(B_n = 0)$ et celle d'occurence du symbole 1 à l'instant n par $q(n) = \mathbb{P}(B_n = 1) = 1 - p(n)$.
La machine d'état passe à l'état "0" après avoir détecté une séquence inintérompue de w fois le symbole 0, et à l'état "1" après avoir détecté une séquence inintérompue de w fois le symbole 1. Notons par X_n, $n = 0, 1, 2, \ldots$ les variables aléatoires décrivant ce processus, où pour tout n, X_n est à valeur dans $\{0, 1\}$.
Le but est de trouver les probabilités de chacun de ces deux états ("0" et "1") après n transitions.
Notons $x(n)$ (respectivement $y(n)$) la probabilité de l'état "0" (respectivement "1") après n transitions :

$$x(n) = \mathbb{P}(X_n = 0), \quad y(n) = \mathbb{P}(X_n = 1) \tag{1}$$

Utilisation de Chaînes de Markov

Dans un premier temps, donnons un brêf aperçu sur la notion de chaîne de Markov.
Soit X_n une variable aléatoire caractérisant l'état du système étudié en des points de temps discrets $n = 0, 1, 2, \ldots$ La famille des variables aléatoires $\{X_n\}$, formant ainsi un processus stochastique, est appelée processus de Markov si l'occurrence d'un état futur dépend seulement the l'état immédiatement précédent. Cette propriété définissant un processus de Markov $\{X_n\}$ se traduit mathématiquement par la formule suivante :

$$\mathbb{P}(X_{n+1} = x \mid X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \mathbb{P}(X_{n+1} = x \mid X_n = x_n) \tag{2}$$

où $x, x_1, x_2, \ldots, x_n$ sont des états quelconques du système étudié (système modélisé par $\{X_n\}$).
Si le nombre d'état, m, est fini, on note l'ensemble d'état du processus $\{X_n\}$ par
$E := \{1, 2, \ldots, m\}$. Dans ce cas le processus $\{X_n\}$ est appelé une chaîne de Markov. Entre l'instant n et l'instant $n+1$, on définit la matrice des probabilités de transition entre états, de $\{X_n\}$, $\mathbf{P}(n) = (p_{ij}(n))_{i,j\in E}$ par

$$p_{ij}(n) = \mathbb{P}(X_{n+1} = j \mid X_n = i) \tag{3}$$

Si pour tout $i, j \in E$, la probabilité $p_{ij}(n) = p_{ij}$ (donc $\mathbf{P}(n) = \mathbf{P} = (p_{ij})_{i,j\in E}$) est indépendante du temps, alors la chaîne de Markov $\{X_n\}$ est dite homogène dans le temps.
A chaque instant n, on définit la loi de probabilité de X_n par

$$a_j(n) := \mathbb{P}(X_n = j), \quad j \in E, \quad n = 0, 1, 2, \ldots$$

La loi de probabilité $\mathbf{a}(n) := (a_j(n))_{j\in E}$ est donc la loi de la variable aléatoire X_n. La loi de probabilité $\mathbf{a}(0)$ de X_0 est appelée loi initiale de $\{X_n\}$.

En résumé, une chaîne de Markov $\{X_n\}$ est complètement caractérisée par le triplet $(E, \mathbf{a}(0), \mathbf{P}(n))$, où E est son espace d'état (supposé ici fini de cardinal m), $\mathbf{a}(0)$ sa loi initiale et $(\mathbf{P}(n))_{n=1,2,...}$ ses matrices de transition.
Dans le cas où $\{X_n\}$ est homogène dans le temps, son triplet caractérisant est $(E, \mathbf{a}(0), \mathbf{P})$, où $\mathbf{P}$ est la matrice de transition (indépendante du temps).

$\mathbf{Q_1}$(I) : Montrer, en utilisant les équations (2) et (3) ainsi que les règles élémentaires de calcul des probabilités, que pour une chaîne de Markov $\{X_n\}$ caractérisée par $(E, \mathbf{a}(0), \mathbf{P}(n))$, nous avons

$$\mathbf{a}(1) = \mathbf{a}(0)\mathbf{P}(0) \quad \text{et} \quad \mathbf{a}(2) = \mathbf{a}(0)\mathbf{P}(0)\mathbf{P}(1)$$

En déduire que, pour tout $n = 1, 2, \ldots,$

$$\mathbf{a}(n+1) = \mathbf{a}(0)\prod_{k=0}^{n}\mathbf{P}(k)$$

$\mathbf{Q_2}$(I) : En supposant que $\{X_n\}$ est homgène dans le temps, déduire la relation suivante :

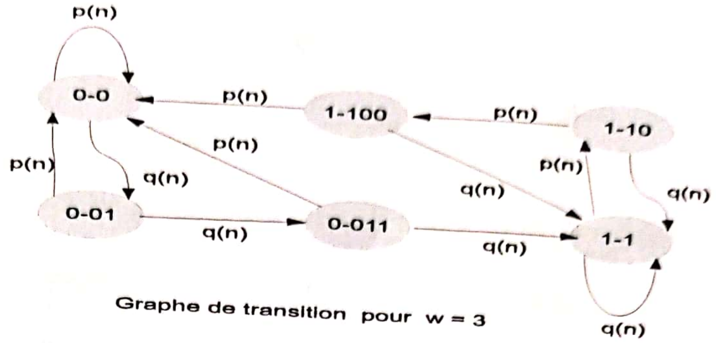
$$\mathbf{a}(n) = \mathbf{a}(0)\mathbf{P}^n, \quad \text{pour tout} \quad n = 1, 2, \ldots$$

Construction de la matrice de transition correspondant à notre problème

Si par exemple l'état courant du système étudié (la machine d'état) est "0", alors la probabilité qu'il devienne "1" au prochain instant (à la prochaine transition) dépend de différents facteurs, et non seulement de l'état courant et du dernier symbole de la séquence aléatoire B_n. Expliquons cela dans un exemple où $w = 3$ (c'est-à-dire le cas où l'état bascule après l'occurence de 3 symboles identiques). Dans ce cas, si l'état à l'instant n était "1" et à l'instant $n+1$, il est devenu "0" (Ce qui veut dire que la fin de la séquence déterminante des symboles est $\ldots, 0, 0, 0$), alors il ne peut pas revenir à létat "1" aux instants $n+2$ et $n+3$, mais premièrement à l'instant $n+4$. Ainsi, la relation (2) n'est pas satisfaite.
Pour surmonter ce problème, il est convenable de considérer plus d'états à la place des deux états "0" et "1" seulement. Cela nous amène à construire une nouvelle chaîne de Markov $\{Y_n\}$, à partir de X_n, de la manière suivante :
Définissons de nouvelles variables aléatoires Y_n avec $2w$ valeurs possibles :

$Y_n = 1$ (état "0-0") si $X_n = 0$ et le dernier symbole apparu est 0,
$Y_n = 2$ (état "0-01") si $X_n = 0$ et les deux derniers symboles sont $0, 1$,
$Y_n = 3$ (état "0-011") si $X_n = 0$ et les trois derniers symboles sont $0, 1, 1$,
...
$Y_n = w$ (état "$0 - 0\overbrace{11\ldots1}^{w-1}$") si $X_n = 0$ et les w derniers symboles sont $0, \overbrace{1, 1\ldots1}^{w-1}$,
$Y_n = w+1$ (état "1-1") si $X_n = 1$ et le dernier symbole apparu est 1,
$Y_n = w+2$ (état "1-10") si $X_n = 1$ et les deux derniers symboles sont $1, 0$,
$Y_n = w+3$ (état "1-100") si $X_n = 1$ et les trois derniers symboles sont $1, 0, 0$,
...
$Y_n = 2w$ (état "$1 - 1\overbrace{00\ldots0}^{w-1}$") si $X_n = 1$ et les w derniers symboles sont $1, \overbrace{0, 0, \ldots 0}^{w-1}$,

Notons que si $Y_n = w$ et 1 vient comme symbole suivant, alors $X_{n+1} = 1$ et $Y_{n+1} = w+1$. Et si $Y_n = 2w$ et le symbole suivant qui vient est 0, alors $X_{n+1} = 0$ et $Y_{n+1} = 1$.
Nous avons ainsi construit une chaîne de Markov $\{Y_n\}$. Son graphe de transition pour $w = 3$ est donné dans la figure suivante :



Graphe de transition pour w = 3

On numérotte les états, du cas $w = 3$ (voir graphe de transition), "0-0", "0-01", "0-011", "1-1", "1-10" et "1-100" respectivement par 0,1,2,3,4 et 5.

$\mathbf{Q_3(I)}$: Ecrire la matrice de transition $\mathbf{P}(n)$ de la chaîne de Markov Y_n, pour le cas $w = 3$.

$\mathbf{Q_4(I)}$: En notant $b_j(n) := I\!P(Y_n = j), \quad j = 1, 2, \ldots, 2w, \quad n = 0, 1, 2, \ldots,$ montrer que les probabilités désirées $x(n)$ et $y(n)$, définies dans (1), peuvent s'écrire :

$$x(n) = \sum_{i=1}^{w} b_i(n), \quad y(n) = \sum_{i=w+1}^{2w} b_i(n)$$

$\mathbf{Q_5(I)}$: Montrer que les probabilités auxiliaires $\mathbf{b}(n) = (b_j(n))_{j=1,2,\ldots,2w}$ sont solution du système d'équations

$$\mathbf{b}(n+1) = \mathbf{b}(n)\mathbf{P}(n), \quad n = 0, 1, 2, \ldots, \tag{4}$$

où $\mathbf{P}(n)$ est la matrice de transition de notre chaîne de Markov Y_n, qui s'écrit en blocs sous la forme

$$\mathbf{P}(n) = \left(\begin{array}{c|c} \mathbf{A(n)} & \mathbf{B(n)} \\ \hline \mathbf{C(n)} & \mathbf{D(n)} \end{array}\right)$$

où $\mathbf{A(n)}, \mathbf{B(n)}, \mathbf{C(n)}$ et $\mathbf{D(n)}$ sont des matrices carrées d'ordre w, avec

$$\mathbf{A(n)} = \begin{pmatrix} p(n) & q(n) & 0 & 0 & \ldots \\ p(n) & 0 & q(n) & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p(n) & 0 & 0 & 0 & \ldots \end{pmatrix} \quad \text{et} \quad \mathbf{D(n)} = \begin{pmatrix} q(n) & p(n) & 0 & 0 & \ldots \\ q(n) & 0 & p(n) & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ q(n) & 0 & 0 & 0 & \ldots \end{pmatrix}$$

Donner les matrices $\mathbf{B(n)}$ et $\mathbf{C(n)}$.

Régime stationnaire d'une chaîne de Markov homogène

On se place dans le cas où la chaîne de Markov $\{Y_n\}$ est homogène dans le temps et on s'intéresse aux probabilités $x(n)$ et $y(n)$ lorsque n tend vers $+\infty$.
En notant $\pi_j := \lim\limits_{n\longrightarrow\infty} b_j(n), \quad j = 1, 2, \ldots, 2w$, la loi statinnaire $\pi := (\pi_1, \pi_2, \ldots, \pi_{2w})$ de $\{Y_n\}$, est

solution du système d'équations, avec contrainte, suivant :

$$(S) = \begin{cases} \pi & = \pi \mathbf{P} \\ \sum_{i=1}^{2w} \pi_i & = 1 \end{cases}$$

Q$_6$(I) : Résoudre le système (S) en considérant $w = 3$, $p(n) = p$ et $q(n) = q = 1 - p$, où $p \in]0, 1[$ est une probabilité donnée.

Q$_7$(I) : En déduire, pour $w = 3$, les probabilités limites $\lim_{n \longrightarrow \infty} x(n)$ et $\lim_{n \longrightarrow \infty} y(n)$.

(II) - Estimation d'une espérance par simulation Monte Carlo

Considérons une variable aléatoire réelle absolument continue X de loi de probabilité définie par une densité f qui est nulle sur $] - \infty, a[\cup]b, +\infty[$, où a et b sont deux réels tels que $a < b$. On s'intéresse à l'estimation, par la méthode Monte Carlo, de la quantité $\theta = \mathbb{E}(g(X))$, on suppose que la variance de $g(X)$ est finie, pour une fonction continue $g \; : \; \mathbb{R} \longrightarrow \mathbb{R}$. Pour cela, on considère deux échantillons, indépendants et identiquement distribués, de X, de tailles respectives N_1 et N_2 :

$$\left(X_1^{(1)}, X_2^{(1)}, \ldots, X_{N_1}^{(1)}\right) \quad \text{et} \quad \left(X_1^{(2)}, X_2^{(2)}, \ldots, X_{N_2}^{(2)}\right),$$

puis on propose les deux estimateurs de Monte Carlo standard

$$\hat{\theta}_i = \frac{1}{N_i} \sum_{k=1}^{N_i} g(X_k^{(i)}), \; i = 1, 2,$$

que l'on souhaite comparer à la famille d'estimateurs $(\hat{\theta}(a))_{a \in]0,1[}$, définie par :

$$\hat{\theta}(a) = a\hat{\theta}_1 + (1 - a)\hat{\theta}_2, \quad 0 < a < 1$$

Q$_1$(II) : Calculer $\mathbb{E}(\hat{\theta}_i)$ et $Var(\hat{\theta}_i)$, $\forall \, i = 1, 2$.

Q$_2$(II) : Calculer $\mathbb{E}(\hat{\theta}(a))$ et $Var(\hat{\theta}(a))$.

Q$_3$(II) : Pour quelle valeur de a, l'estimateur $\hat{\theta}(a)$ est plus intéressant qu'un estimateur de Monte Carlo standard $\hat{\theta}_i$.

(II) - OPTIMISATION

I) Soit $f : \mathbb{R}^n \to \mathbb{R}$, d un vecteur de $\mathbb{R}^n$, $x \in \mathbb{R}^n$, on dit que f est différentiable en x dans la direction d si la limite $\lim_{t\to 0, t>0} \frac{1}{t}(f(x+td) - f(x))$ existe et on note :

$$f'(x,d) = \lim_{t\to 0,\ t>0} \frac{1}{t}(f(x+td) - f(x)).$$

$f'(x,d)$ est appelée dérivée directionnelle de f en x dans la direction d.

1) Montrer que si f est différentiable en x alors f est différentiable en x dans toute direction $d \neq 0$ et que $f'(x,d) = \nabla f(x).d$

2) $d \in \mathbb{R}^n$ est dit direction de descente de f en un point $x \in X$ s'il existe $\delta > 0$ tel que $f(x+td) < f(x)$ pour tout $0 < t < \delta$.
Soit d une direction de descente de f. Montrer que si f admet une dérivée directionnelle en x_0 dans la direction d alors $f'(x_0, d) \leq 0$

3) La méthode de Gradient à pas optimal consiste à construire une suite minimisante (x_k), partant de x_0 choisi, prendre en chaque point x_k : $d_k = -\nabla f(x_k)$, déterminer t_k qui minimise la fonction $\phi_k(t) = f(x_k + td_k)$ pour $t > 0$ et poser $x_{k+1} = x_k + t_k d_k$
Montrer que d_k est une direction de descente et que d_k et d_{k+1} sont orthogonaux

4) Soit A une matrice carré définie positive, b un vecteur de $\mathbb{R}$ donnés et f la fonction quadratique définie par $f(x) = \frac{1}{2}x^T Ax - bx$. Donner d_k et t_k pour la méthode de gradient à pas optimal.

5) Soit $X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$ et $f(X) = x_1^2 + 3x_2^2$, partant de $X_0 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$ calculer X_1, X_2, X_3 par la méthode de gradient à pas optimal puis déduire la forme générale de X_k. Que dire de la convergence ? Est ce que le choix de X_0 influe sur la vitesse de convergence (justifier)

II) Soient $x_i > 0$ et $\alpha_i > 0$, $i = 1, \cdots, n$ tel que $\sum_{i=1}^n \alpha_i = 1$.

1) Montrer que

$$x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n} \leq \alpha_1 x_1 + \alpha_2 x_2 + \cdots \alpha_n x_n$$

avec égalité si $x_1 = x_2 = \cdots = x_n$

2) En déduire que : $\frac{1}{n}(x_1 + x_2 + \cdots + x_n) \geq (x_1 x_2 \cdots x_n)^{1/n}$.

3) On veut fabriquer une boite sous forme d'un parallélépipède (ouverte en haut). La surface des 5 faces est fixée égale à S. Donner les dimensions a_1, a_2, a_3 des cotés qui donnent une boite de volume maximal.

III) Une société de transport de voyageurs exploite une ligne de bus Fes-Marrakech-Agadir, elle ne peut pendre ou faire descendre des voyageurs en cours de route. Les billets de voyage sont vendus à trois tarifs différents :

Tarif 1 : plein tarif, billet remboursable jusqu'au jour du départ.

Tarif 2 : billet non remboursable et non échangeable.

Tarif 3 : billet acheté au moins 3 semaines avant le départ, non remboursable et non échangeable.

les tarifs en fonction des trajets sont donnés dans le tableau suivant :

	Fes-Marrakech	Marrakech-Agadir	Fes-Agadir
tarif1	300	160	360
tarif2	220	130	180
tarif3	100	80	140

Chaque bus a une capacité maximum de 30 places. D'après l'expérience des années passées, les prévisions de demandes en places pour un jour donné sont :

	Fes-Marrakech	Marrakech-Agadir	Fes-Agadir
tarif1	4	8	3
tarif2	8	10	8
tarif3	22	20	18

La société doit proposer à la vente des places de chaque catégorie de tarif pour chacun des trajets.

1) Ecrire le modèle mathématique qui donnera une répartition assurant un bénéfice maximum pour le jour concerné.

2) Donner une méthode numérique permettant de résoudre ce type de problème et décrire son principe de base.

3) donner le nom de logiciels qui peuvent être utilisés pour traiter ce problème et en donner une brève description.

(III) - Analyse Numérique : Décomposition en valeurs singulières d'une matrice et application aux problèmes des moindres carrés

Dans ce problème on note $\mathcal{M}_{n,p}(I\!K)$ l'ensemble des matrices n lignes p colonnes à coefficients dans $I\!K = I\!R$ ou $\mathbb{C}$.
Pour une matrice $A \in \mathcal{M}_{n,p}(I\!K)$, on note A^* sa matrice adjointe, $C = A^*A$ et on suppose que le rang de A est égal à r : $rg(A) = r$
On notera le produit scalaire dans $I\!K^n$ (u,v) ou v^*u pour u, v dans $I\!K^n$, $\|.\|$ désigne la norme associée et aussi la norme maricielle définie par $\|A\| = \sup_{x \in I\!K^n, x \neq 0} \frac{\|Ax\|}{\|x\|}$.
On rappelle qu'une matrice carrée unitaire (orthogonale dans le cas réel) est inversible et verifie $AA^* = A^*A = I$.
Une matrice normale($AA^* = A^*A$) est diagonalisable par une matrice unitaire.

1. Montrer que les valeurs propres de C sont réelles et positives et que C admet exactement r valeurs propres strictement positives que l'on notera $\lambda_1 \geq \ldots \geq \lambda_r > 0$ et $\lambda_{r+1} = \ldots = \lambda_p = 0$. En déduire qu'il existe une base orthonormée $(u_1, \ldots, u_p)$ de $I\!K^p$ de vecteurs propres associés.

2. On note $\mu_i = \sqrt{\lambda_i}$ les valeurs singulières de la matrice A et les vecteurs $v_i = \frac{1}{\mu_i} A u_i$ pour $i = 1, \ldots, r$
   (a) Montrer que la matrice A admet la décomposition en valeur singulière :
   $$A = \sum_{i=1}^{r} \mu_i v_i u_i^* \tag{5}$$
   (b) Soient $(w_1, \ldots, w_r)$ et $(t_1, \ldots, t_r)$ deux systèmes orthonrmés de $I\!K^r$ et $\alpha_1, \ldots, \alpha_r$ r, scalaires de $I\!K$. Montrer que la matrice définie par
   $$M = \sum_{i=1}^{r} \alpha_i w_i t_i^* \tag{6}$$
   est de rang r et admet $(\alpha_1, \ldots, \alpha_r, 0, \ldots, 0)$ pour valeurs singulières.

3. On considère la matrice $\Sigma = (\sigma_{i,j}) \in \mathcal{M}_{n,p}(I\!R)$ telles $\sigma_{i,i} = \mu_i$ et tous les autres coefficients sont nuls. Montrer qu'il existe deux matrices unitaires $P \in \mathcal{M}_n(I\!K)$ et $Q \in \mathcal{M}_p(I\!K)$ telles que $A = P\Sigma Q^*$

4. On suppose dans cette question que $n = p = r$.
   À partir de la décomposition (5) de A, donner la décomposition de A^{-1} et en déduire que si δb est une perturbation du second membre du système linéaire
   $$Ax = b$$
   alors on a l'estimation d'erreur relative sur la solution x est donnée par :
   $$\frac{\|\delta x\|}{\|x\|} \leq \frac{\mu_1}{\mu_n} \frac{\|\delta b\|}{\|b\|}$$
   (On écrira le système perturbé : $A(x + \delta x)) = b + \delta b$)

Soit $A \in \mathcal{M}_{n,p}(I\!R)$ admettant une décomposition en valeur singulière $A = P\Sigma Q^*$ et un vecteur $b \in I\!R^p$, on considère le problème des moindres carrés :

$$\min_{x \in \mathbb{R}^p} \|Ax - b\|^2$$

dont une solution sera notée x^*. (7)

5. Etudier l'existence et l'unicité de solution pour le problème (7) suivant le rang de la matrice A. (On explicitera les solutions dans chaque cas)
6. Démontrer que le problème (7) admet une unique solution, de norme minimale, x^* donnée par :

$$x^* = \sum_{i=1}^{r} \frac{(b, v_i)}{\mu_i} u_i \tag{8}$$

   où une solution de norme minimale est la solution du problème :

$$\min_{x \in \{x^* \text{ Solution de } (7)\}} \|x\| \tag{9}$$

7. Donnez une estimation de l'erreur relative sur la solution obtenue pour une perturbation δb sur la donnée b si on suppose $rg(A) = p$
8. Ecrire une procédure permettant de résoudre numériquement le problème (7). On suppose que l'on dispose d'une fonction SVD qui retourne la décomposition SVDde A, $A = P\Sigma Q^*$ d'une matrice quelconque $A \in \mathcal{M}_{n,p}(I\!R)$ dont on précisera les paramètres nécessaires.