Probabilités Statistiques TP
Chaînes de Markov

Les chaînes de Markov vont naturellement amener à manipuler des matrices. Le package qui sera adapté pour cela est **LinearAlgebra** (qu'il faut donc charger, en plus des autres **Statistics**, **plots**).
On déclare une matrice à l'aide de l'instruction **Matrix**. Par exemple **Matrix(2,2,[1,3,2,5])** définit la matrice $\begin{pmatrix} 1 & 3 \\ 2 & 5 \end{pmatrix}$. On définit un vecteur ligne avec la commande $\mathbf{Vector_{row}}$: par exemple $\mathbf{Vector_{row}}([1,3,6])$ définit le vecteur $(1,3,6)$. Deux matrices, ou un vecteur et une matrice, se multiplient à l'aide de l'opérateur **.**. On peut extraire une ligne d'une matrice A à l'aide de la commande **Row(A,i)** (qui fournit un vecteur-ligne correspondant à la i-ème ligne; de même la commande **Column(A,i)** fournit un vecteur colonne). La commande **Eigenvectors(A)** fournit les valeurs propres et les vecteurs propres d'une matrice A, et **Transpose(A)** fournit la transposée d'une matrice A.

1 Petit modèle

On suppose que les gens sont initialement répartis respectivement en : 50% dans l'état 1, 30% dans l'état 2, et 20% dans l'état 3. Après chaque mois, 60% des gens dans l'état 1 y restent, 70% de ceux dans l'état 2 y restent, et 90% de ceux dans l'état 3. Les autres se répartissent de façon équiprobable dans les autres états.

1. On modélise la situation à l'aide d'une chaîne de Markov. On note μ_0 la distribution initiale, et P la matrice de transition. Déterminer μ_0 et P et les déclarer dans maple, sous forme respectivement d'un vecteur ligne et d'une matrice. (**Remarque** : on a fortement intérêt à rentrer les coefficients comme des fractions : **pour maple, 0.3 et** $\frac{3}{10}$ **ne sont pas traités de manière équivalente**.

2. La chaîne est-elle récurrente ? irréductible ? apériodique ? Que peut-on affirmer sur l'existence et l'unicité de probabilité invariante ?

3. Simuler la chaîne pendant 5 ans (i.e. 60 mois), c'est-à-dire commencer par tirer $X_1(\omega)$ selon la loi μ_0, puis faire une boucle, et tirer $X_i(\omega)$ selon une loi qui dépend de $X_{i-1}(\omega)$. Tracer $X_i(\omega)$ en fonction de i.
   Pour tirer une variable selon une loi discrète donnée, on peut utiliser
   RandomVariable(ProbabilityTable(List)) qui si *List* est une liste de m probabilités p_i définit une variable aléatoire qui vaut i avec la probabilité p_i.
   Attention : *List* doit vraiment être une liste... pour transformer un vecteur en liste, utiliser la commande **convert**.
   Attention **ProbabilityTable** est (un peu) buggée : si vous avez une variable **i** qui traine dans votre session, il risque d'y avoir conflit. Dans ce cas, commencez par faire ***unassign('i')*** avant d'utiliser **ProbabilityTable**

4. Déterminer la probabilité invariante μ de P.

5. Comparer la proportion des X_i qui sont dans l'état 1 avec la valeur limite théorique (si on avait simulé la chaîne pendant un temps infini).

6. Tracer sur un même graphique les trois courbes donnant les trois coordonnées du vecteur $\mu_0 P^n$ en fonction de n (penser à faire des **evalf**), ainsi que les trois droites horizontales correspondant aux valeurs limites.

7. Représenter $|\mu_0 P^n - \mu|_1$ en fonction de n. La convergence est-elle polynomiale ? exponentielle ?
(On peut calculer la norme 1 explicitement, ou utiliser **Norm(V,1)** qui si V est un vecteur donne sa norme 1.
On pourra utiliser les commandes **logplot** et **loglogplot**.

2 Modèle de diffusion d'Ehrenfest

10 particules sont placés dans 2 récipients A et B. A chaque instant on choisit une particule au hasard (uniformément) et on la change de récipient. On rappelle qu'on obtient ainsi le modèle d'Ehrenfest, dont la probabilité invariante est la loi binomiale $\mu = Bin(10, 0.5)$.

1. Écrire une procédure qui prend en paramètre deux entiers n et k_0, et simule la chaîne de Markov correspondant au modère pendant n unités de temps, en partant d'une distribution initiale qui est un Dirac en k_0.

2. Exécuter la procédure avec $n = 1000$ et un k_0 de votre choix, et tracer sur un même graphique l'histogramme empirique des X_i obtenus (on pourra utiliser les options **discrete=true, frequencyscale=relative** pour l'histogramme) et la distribution de la probabilité invariante (dont le graphe est donné à l'aide de la commande ***DensityPlot***).

3. A-t-on convergence de X_n vers la probabilité invariante ?

4. Si μ_0 désigne la distribution initiale, et si P désigne la matrice de transition de la chaîne, on définit $\mu_n = \frac{(\mu_0 + \mu_0 P + \ldots + \mu_0 P^n)}{n}$. Tracer sur un même graphique, pour μ_0 correspondant à un Dirac de votre choix, les courbes donnant $|\mu_0 P^n - \mu|_1$ et $|\mu_n - \mu|_1$ en fonction de n. Qu'observe-t-on ?

5. On définit le temps de retour en l par $T_l = inf\{n > 0, X_n = l | X_0 = l\}$. Écrire une procédure qui simule la chaîne partant de l et retourne le temps de retour en l. Effectuer la procédure 500 fois et calculer la moyenne empirique des temps de retour obtenus. Comparer avec la valeur théorique $\frac{1}{\mu(l)}$.