2.1

1) Une chaîne de Markov à temps discret et à ensemble d'états finis est un processus stochastique (aléatoire) possédant un nombre d'états finis. Le processus peut, à un instant t, se trouver dans un de ces états. La probabilité que le processus soit passé de l'état i à l'état j à t+1 ne dépend pas de t et est notée P_{ij} . P est appelée la matrice de transition.

2) La matrice de transition, notée P, est une matrice carrée dont le terme P_{ij} est la probabilité que le processus soit à l'état j à l'instant t+1 sachant qu'il était à l'état i à l'instant t. La matrice de transition est donc une matrice carrée d'ordre m égal au nombre d'états de la chaîne de Markov.

3) On peut représenter une chaîne de Markov par un graphe orienté valué dont les sommets correspondent aux états de la chaîne et les valeurs des arcs correspondent aux probabilités de passer de l'état d'origine de l'arc à l'état de son extrémité.

4) La puissance de la matrice P permet de connaître avec quelle probabilité le processus se trouve à l'état j à l'instant t alors qu'il était à l'état i à l'instant 0.

Cette probabilité $P_{ij}^{(t)}$ correspond au terme i, j de la t-ième puissance de la matrice de transition P.

Sources :

http://www.nymphomath.ch/graphes-ancien/markov/index.html
https://fr.wikipedia.org/wiki/Chaîne_de_Markov
http://gradus-ad-mathematicam.fr/documents/404_Markov.pdf

2.2 Test d'adéquation du Khi-deux

1) Le principe général du test d'adéquation du Khi-deux est le suivant: c'est un test qui permet de comparer un ensemble de données observées et sa loi de probabilités théoriques en fonction de son degré de liberté.
Le degré de liberté d'un test d'adéquation de Khi-deux est le nombre de valeurs possibles que peut prendre une variable aléatoire auxquelles on soustrait 1 (puisqu'à partir de J-1 premières valeurs, nous pouvons retrouver la dernière valeur). Dans le cas d'une loi continue, on regroupe nos valeurs en intervalles, et le degré de liberté est le nombre d'intervalles moins 1. Avant d'expliquer son principe, nous devons vérifier que pour chaque valeur (ou intervalle de valeurs) que peut prendre notre variable, toutes ont un effectif théorique qui est supérieur à 5. Si ce n'est pas le cas, nous regroupons les différentes valeurs et nous recalculons le degré de liberté jusqu'à rencontrer cette condition.

2) On commence par faire 2 hypothèses H_0 et H_1. L'hypothèse H_0 est généralement : « La distribution observée suit la loi théorique » et H_1 est son opposé : « La distribution observée ne correspond pas à loi théorique ».

3) Ensuite, on fixe un taux d'erreur, une probabilité α que l'hypothèse H_0 soit rejetée alors qu'elle est vraie. Cette valeur vaut généralement 5 % ou 1 %.

Une fois cela fait, nous calculons la somme suivante : $\sum_{i=0}^{k} \frac{(n_i + nt_i)^2}{nt_i}$ où n_i est l'effectif observé pour une classe i, k est le nombre total de classes et nt_i est l'effectif théorique pour la classe i. Nous obtenons un résultat que nous notons d^2 ou X^2. Ensuite, nous regardons la fonction de densité du Khi-deux qui correspond à notre degré de liberté et nous déterminons à partir de quelle valeur la surface sous la courbe (l'intégrale) est supérieure à 1- α . La fonction de densité de la Loi du Khi-deux étant complexe à intégrer, nous pouvons aussi utiliser un tableau qui, pour un α précis et un certain degré de liberté, nous donne cette valeur (que l'on appellera s).

4) Une fois cette valeur trouvée, on regarde si d^2 est inférieure à s. Si c'est le cas, nous acceptons l'hypothèse H_0, sinon nous la rejetons au profit de l'hypothèse H_1 avec une marge d'erreur de 5 %.

Sources :

http://www.bibmath.net/dico/index.php?action=affiche&quoi=c/chideuxtest.html

http://www.iutbayonne.univ-pau.fr/~grau/2A/stat/cadre5.html

https://fr.wikipedia.org/wiki/Test_du_χ²

http://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-m-inf-khi2.pdf

3.1 QUELQUES FONCTIONS

Les fonctions et méthodes réalisées pour les questions 1, 2, 3 et 4 se trouvent dans un fichier joint nommé « projet.r ».

Le package R nommé « expm » est nécessaire pour la fonction de la question 1.

3.2 ACTIVITES D'UN ETUDIANT

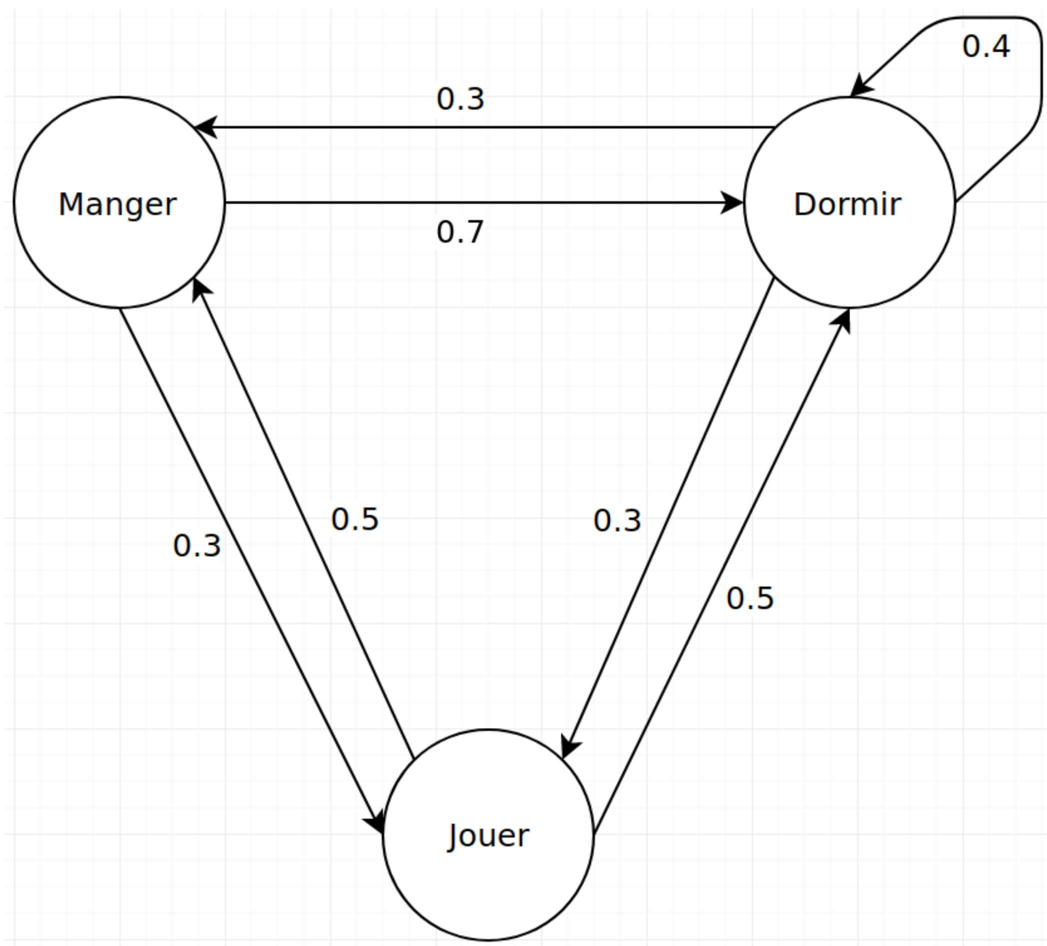
1) Le système proposé correspond bien à une chaîne de Markov car :

- Il possède un nombre d'états finis : manger, dormir, jouer.

- Le temps n'influe pas sur la probabilité de passer d'un état i à j à l'instant t + 1.

- La matrice de transition réalisée depuis ce système à la question 3) est une matrice stochastique (ou matrice de Markov) dont la somme des probabilités sur une ligne vaut forcément 1.

2) La chaîne de markov peut être représentée par le graphe suivant :



3) La matrice de transition de cette chaîne de Markov est :

$$\boldsymbol{P}=\begin{bmatrix} 0 & 7/10 & 3/10 \\ 3/10 & 2/5 & 3/10 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

4)

La transposée de la matrice de transition P est $^{t}\boldsymbol{P}=\begin{bmatrix} 0 & 3/10 & 1/2 \\ 7/10 & 2/5 & 1/2 \\ 3/10 & 3/10 & 0 \end{bmatrix}$ Le vecteur propre associé à la valeur propre 1 de $^{t}\boldsymbol{P}$ répond à l'équation

$$\begin{bmatrix} 0 & 3/10 & 1/2 \\ 7/10 & 2/5 & 1/2 \\ 3/10 & 3/10 & 0 \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

On peut remplacer ceci en résolvant le système d'équation suivant :

$$\left\{ \begin{array}{c} 0\,x_1+\frac{3}{10}x_2+\frac{1}{2}x_3=x_1 \\ \frac{7}{10}x_1+\frac{2}{5}x_2+\frac{1}{2}x_3=x_2 \\ \frac{3}{10}x_1+\frac{3}{10}x_2+0\,x_3=x_3 \end{array} \right\}$$

On transforme ce système en système d'équations où chacune des équations vaut 0 et on multiplie chaque ligne par 10 pour se débarrasser des fractions.

On obtient :

$$\left\{ \begin{array}{c} -10\,x_1+3\,x_2+5\,x_3=0 \\ 7\,x_1-6\,x_2+5\,x_3=0 \\ 3\,x_1+3\,x_2-10\,x_3=0 \end{array} \right\}$$

On résout ensuite ce système et on obtient qu'un vecteur propre associé à la valeur propre 1 est de la forme

$$\begin{bmatrix} \frac{15}{13}x_3 \\ \frac{85}{39}x_3 \\ x_3 \end{bmatrix} \quad ,où\, x_3 \in \mathbb{R}$$

On doit ensuite faire en sorte que la somme des coordonnées de ce vecteur vaille 3.

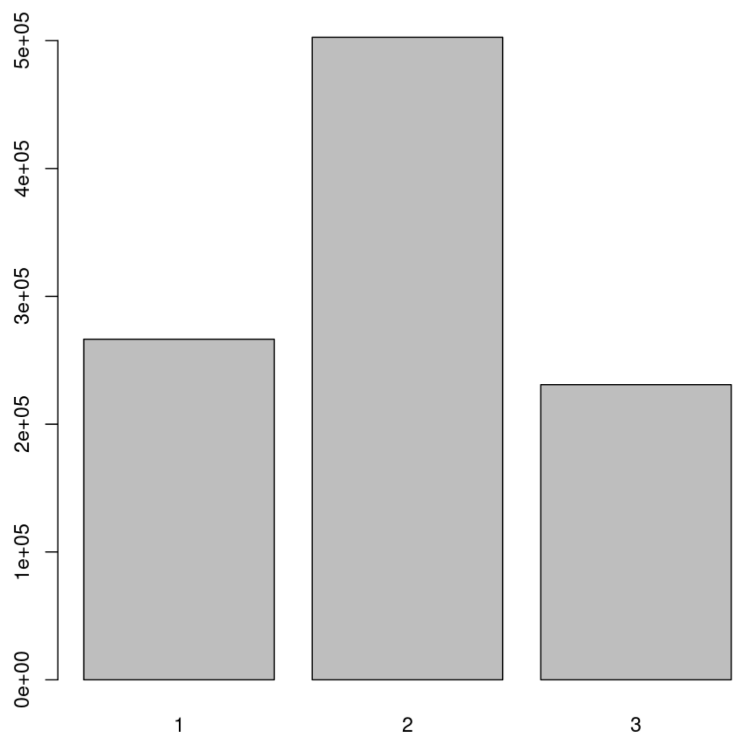
On résout donc $\frac{15}{13}x_3+\frac{85}{39}x_3+x_3=1$ et on obtient $x_3=\frac{39}{169}$

Ce qui fait que le vecteur propre qui nous intéresse est $\begin{bmatrix} \frac{45}{169} \\ \frac{85}{169} \\ \frac{39}{169} \end{bmatrix}$

5) Avec l'état initial à 1 :

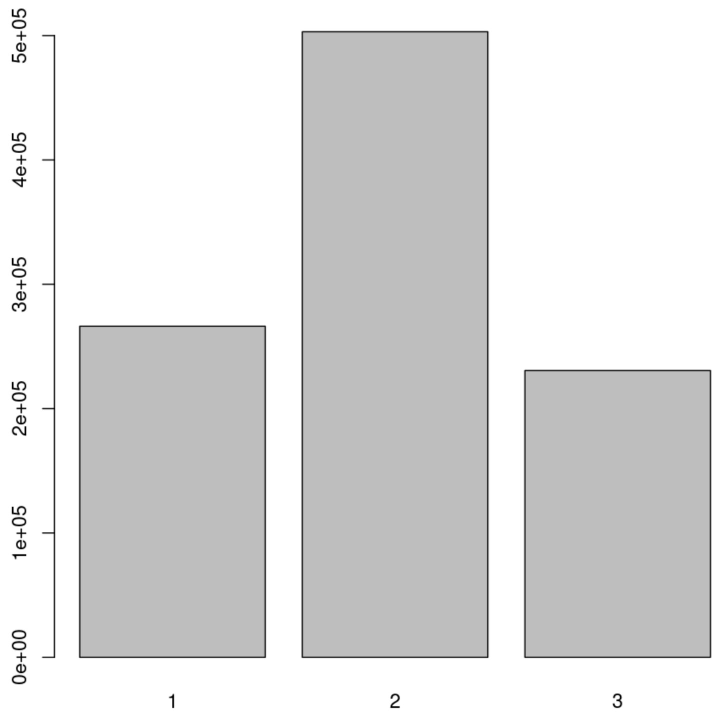
Avec n = 10^6, et en comptant l'étape initiale et n+1 on obtient :

étape	occurrences
1	266461
2	502617
3	230924



Avec l’état initial à 2 :
Avec n = 10^6, et en comptant l’étape initiale et n+1 on obtient :

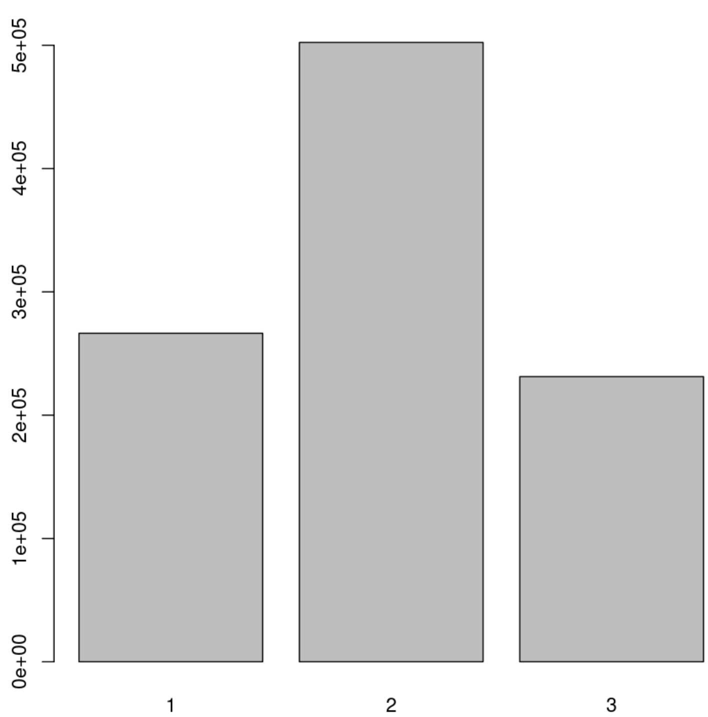
étape	occurrences
1	266261
2	503079
3	230662



Avec l’état initial à 3 :

Avec n = 10^6, et en comptant l’étape initiale et n+1 on obtient :

étape	occurrences
1	266434
2	502293
3	231275



Quand nous effectuons les tests d'adéquation (grâce à notre code R), nous voyons que les tests sont positifs et que nos 3 séries statistiques semblent donc suivre la loi de probabilité donnée par le vecteur propre.

6) En simulant une trajectoire avec n = 10^6 on obtient les valeurs suivantes :

Pour l'état initial 1 :

La durée moyenne d'un retour à l'état 1 est de 3.752912

Pour l'état initial 2 :

La durée moyenne d'un retour à l'état 2 est de 1.987765

Pour l'état initial 3 :

La durée moyenne d'un retour à l'état 3 est de 4.323858

Pour les 3 durées moyennes, cela semble être l'inverse. En effet $\frac{169}{45} \simeq 3{,}7529$,

$\frac{169}{85} \simeq 1{,}9878$ et $\frac{169}{39} \simeq 4{,}3238$

7)

$$\boldsymbol{P}^{50} = \begin{bmatrix} 0.2662722 & 0.2662722 & 0.2662722 \\ 0.5029586 & 0.5029586 & 0.5029586 \\ 0.2307692 & 0.2307692 & 0.2307692 \end{bmatrix}$$

8) Nous avons lancé 1000 simulations partant de l'état 1 avec une simulation de 50, nous avons testé l'adéquation à la première colonne de $\boldsymbol{P}^{50}$ des résultats et nous avons noté et regroupé les 1000 résultats, à l'aide de R : Nous obtenons :

```
results
FALSE  TRUE
    3   997
```

A la vue de la proportion de bons tests par rapport aux mauvais tests, et au choix du nombre alpha qui est de 0,05, il semble y avoir adéquation entre cette première colonne et les résultats de la simulation.

Nous avons lancé 1000 simulations partant de l'état 2 avec une simulation de 50, nous avons testé l'adéquation à la première colonne de $\boldsymbol{P}^{50}$ des résultats et nous avons noté et regroupé les 1000 résultats, à l'aide de R : Nous obtenons :

```
results
FALSE  TRUE
    5   995
```

A la vue de la proportion de bons tests par rapport aux mauvais tests, et au choix du nombre alpha, il semble y avoir adéquation entre cette première colonne et les résultats de la simulation.

Nous avons lancé 1000 simulations partant de l'état 3 avec une simulation de 50, nous avons testé l'adéquation à la première colonne de $\boldsymbol{P}^{50}$ des résultats et nous avons noté et regroupé les 1000 résultats, à l'aide de R : Nous obtenons :

```
results
FALSE  TRUE
   11   989
```

A la vue de la proportion de bons tests par rapport aux mauvais tests, et au choix du nombre alpha, il semble y avoir adéquation entre cette première colonne et les résultats de la simulation.

9) $\boldsymbol{P}^{\boldsymbol{n}}$ semble tendre vers notre vecteur propre « concaténé » 3 fois (pour obtenir une matrice 3x3).