Université Rennes 1
Master 1 Mathématiques année 2010-2011
Nathalie Krell

Fiche de TP n°3 : Chaîne de Markov

1 Un exemple simpliste

Les consommateurs de 3 produits sont répartis respectivement en 50% pour P1, 30% pour P2 et 20% pour P3. Après chaque mois, 60% restent fidèles à P1 contre 70% pour P2 et 90% pour P3. Les autres se réorientent entre les deux autres produits (de manière équiprobable).

[a] Déterminer la distribution initiale ν et la matrice de transition associées (exemple : $P_{12} = 0.2$).

[b] Tracer l'évolution déterministe des répartitions des consommateurs pendant les 12 premiers mois.

[c] Tracer l'évolution de l'opinion d'un individu initialement adepte du produit 1 pendant les douze premiers mois(consultez l'aide pour `grand` option 'markov').

[d] Même question pour un individu pris au hasard dans la population totale selon la loi ν.

[e] Simuler la répartition sur douze mois des opinions de 1000 personnes (qui ne se concertent pas) choisies indépendamment dans la population.

[f] Déterminez la distribution stationnaire de la matrice de transition et refaire la question 2 en prenant la distribution stationnaire comme distribution initiale des consommateurs. Quelles sont les valeurs propres de `P` ? On pourra utiliser les fonctions `spec`, `bdiag`, mais aussi de grandes puissances de P.

2 L'urne d'Ehrenfest

On dispose de m particules que l'on répartit initialement entre deux récipients A et B. À chaque pas de temps, on choisit une particule parmi les m et on la change d'urne. On note X_n le nombre de particules dans l'urne A au temps n. Pour toutes les applications, on prendra $m = 10$.

[a] Expliquer pourquoi (X_n) est une chaîne de Markov sur $\{0, \ldots, m\}$ et donner sa matrice de transition.

[b] Cette chaîne est-elle irréductible ? récurrente ? périodique ?

[c] Écrire une fonction qui trace une trajectoire de cette chaîne.

[d] Vérifier que la mesure invariante μ de cette chaîne est la loi binomiale $\mathcal{B}(m, 1/2)$.

[e] Illustrer le théorème ergodique par simulation *i.e.* illustrer le fait que, pour $l \in \{0, \ldots, m\}$,

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i=l\}} \underset{n\to\infty}{\longrightarrow} \mu(\{l\}) \quad \text{p.s.}$$

[f] On définit T_{ll} le temps de retour en l par $T_{ll} = \inf\{n \geq 1, X_n = l | X_0 = l\}$. Comparer, par simulation, $\mathbb{E}(T_{ll})$ et $\mu(l)$.

[g] La loi de X_n converge-t-elle vers μ ?

[h] Comment simuler les trajectoires d'une chaîne d'Ehrenfest lorsque le nombre de boules atteint 1000000 ?

3 Processus de Galton-Watson

Voici un exemple classique de processus de population. Une cellule se divise, et donne naissance à D cellules, où $D \in \{0, 1, 2\}$: $D = 2$ signifie que la mitose cellulaire s'est bien déroulée, ce qui arrive avec probabilité p_2. Le cas $D = 1$ correspond à l'obtention d'une cellule non viable parmi les deux cellules-filles (probabilité p_1), et $D = 0$ désigne les cas où la cellule mère meurt pendant la division (probabilité p_0). On note m la moyenne de la loi de D, et σ^2 sa variance.

Soit $(D_{n,k})_{n\geq 1,k\geq 1}$ des variables aléatoires i.i.d. de loi celle de D. On part de $Z_0 = 1$ cellule, dont on étudie la descendance. Pour $n \geq 1$, le nombre de cellules à la génération n est ainsi donné par

$$Z_N = \sum_{k=1}^{Z_{n-1}} D_{n,k}.$$

[a] Justifier que $(Z_n)_{n\geq 0}$ forme une chaîne de Markov. On note q la probabilité d'extinction, c'est à dire,

$$q = \mathbb{P}(\exists n \in \mathbb{N}^* | Z_n = 0),$$

et d'après le cours $\mathbb{E}[Z_n] = m^n$.

[b] Soit q_n la probabilité que $Z_n = 0$. Montrer que $(q_n)_{n\geq 0}$ est croissante et préciser sa limite.

[c] Dans le cas (dit sous-critique) $m < 1$, par exemple pour

$$(p_0, p_1, p_2) = (1/5, 7/10, 1/10),$$

écrire une fonctions calculant un N-échantillon de $(Z_0, Z_1, ..., Z_n)$, où $n = 100$ et N est assez grand (justifier sa valeur), et illustrant graphiquement la convergence de la suite des probabilités d'extinction $(q_n)_n$ vers sa valeur limite. Que semble valoir q ? Enfin, tracez sur un même graphique une suite d'estimations empiriques des $(\mathbb{E}[Z_n])_{n\geq 0}$ contre les valeurs théoriques $(m^n)_{n\geq 0}$. Dans le cas sur-critique $m > 1$, on a convergence presque-sûre et dans $\mathbb{L}^2$ de Z_n/m^n vers une variable aléatoire W, de moyenne 1 et de variance égale à $\sigma^2/(m^2 - m)$.

[d] On fixe $(p_0, p_1, p_2) = (1/10, 7/10, 1/5)$. Pour $n = 100$ et N bien choisi, écrire une fonction qui simule une approximation empirique de la loi de W : cette fonction estimera la probabilité que $W = 0$, ainsi que sa variance et sa moyenne, et les comparera aux valeurs proposées par la théorie. Ensuite, vous représenterez graphiquement la loi de W par un histogramme correctement calibré. Dans le cas critique $m = 1$, il y a extinction presque-sûre, à la vitesse :

$$\mathbb{P}(Z_n = 0) \sim \frac{2}{\sigma^2 n}.$$

De plus, les lois des Z_n/n, conditionnés à la non-extinction (i.e. à $Z_n > 0$), convergent vers une loi exponentielle de paramètre $\sigma^2/2$.

[e] Illustrez ces deux faits selon votre convenance, en justifiant votre démarche et vos choix de représentation. Pour cela, considérez

$$(p_0, p_1, p_2) = (3/20, 7/10, 3/20).$$

4 Marche aléatoire

Simuler une marche aléatoire simple à valeur dans $\mathbb{Z}$. Tracer la distance à 0 de la chaîne en fonction du temps.

Maintenant on s'intéresse à une marche aléatoire simple à valeur dans $\mathbb{Z}^3$. Tracer la distance à 0 de la chaîne en fonction du temps.

5 Agrégation limitée par diffusion interne

Un fût de déchets radioactifs est enterré secrètement dans le Cantal. Au bout de quelques années, il devient poreux et laisse échapper son contenu. Pour éviter une contamination excessive, on a disposé des pièges à particules qui capturent la première particule qui passe dessus mais deviennent ensuite inertes (une nouvelle particule passe dessus sans être arrêtée). On suppose que le milieu est isotrope : une particule radioactive se déplace de la même manière dans toutes les directions. Cette particule continuera à se déplacer tant qu'elle passe sur des pièges qui ont déjà été activés et est capturée par le premier piège libre qu'elle rencontre. On souhaite connaître la forme typique des zones qui seront contaminées par cette fuite.

On modélise donc la trajectoire d'une particule radioactive sortant du fût par une marche aléatoire simple symétrique sur $\mathbb{Z}$ issue de 0 et arrêtée lorsqu'elle sort de A_0. On note A_1 l'ensemble aléatoire composé de 0 et du point où la particule est sortie. Il est évident que, pour tout x voisin de 0, A_1 est égal à $\{0, x\}$ avec probabilité 1/2. On itère ensuite le procédé. Étant donné un ensemble $A_n \subset \mathbb{Z}$, on considère une marche aléatoire symétrique $(S_k)_{k\geq 0}$ issue de 0 arrêtée lorsqu'elle sort de A_n. On définit alors A_{n+1} comme l'ensemble des éléments de A_n et du point où est sortie la marche S.

Notons $G_n = \min A_n$ et $D_n = \max A_n$. L'ensemble A_n est de la forme $A_n = \{G_n, G_n + 1, \ldots, D_n - 1, D_n\}$. Puisque le cardinal de A_n est $n + 1$, on a $D_n - G_n = n$. Ainsi, A_n est caractérisé par $X_n = D_n + G_n$.

On peut montrer que si $|X_n| = 0$ alors $|X_{n+1}| = 1$ et, que, si $|X_n| > 0$, alors

$$|X_{n+1}| = \begin{cases} |X_n| - 1 & \text{avec probabilité } \dfrac{1}{2} + \dfrac{|X_n|}{2(n+2)}, \\ |X_n| + 1 & \text{avec probabilité } \dfrac{1}{2} - \dfrac{|X_n|}{2(n+2)}. \end{cases}$$

[a] Modéliser cette chaîne de Markov.

[b] **Proposition 1** *Soit S une marche aléatoire simple sur $\mathbb{Z}$ issue de 0 et soit a et b deux entiers distincts, le premier négatif, le second positif. On note $T_i = \inf\{n \geq 0,\ S_n = i\}$ pour $i = a, b$ et $T = T_a \wedge T_b$. Alors, T_a et T_b sont finis presque sûrement, T est intégrable et*

$$\mathbb{P}(T = T_a) = 1 - \mathbb{P}(T = T_b) = \frac{b}{b-a} \quad \text{et} \quad \mathbb{E}(T) = -ab.$$

On pourra vérifier cette proposition par la simulation.

[c] **Theorème 2** *Le processus $(|X_n|)_{n\geq 0}$ est une chaîne de Markov inhomogène et $(|X_n|/n)_n$ converge vers 0 presque sûrement.*

Montrer la convergence p.s. pour la simulation.

Pour cet exercice on pourra se référer au texte d'agrégation fait par Florent Malrieu disponible sur sa page web.

6 Modèle de Wright-Fisher sans mutations

On modélise le nombre d'allèles A dans une population de $2N$ gènes par la chaîne de Markov de matrice de transition suivante :

$$p_{ij} = C_{2N}^{j} \left(\frac{i}{2N}\right)^{j} \left(1 - \frac{i}{2N}\right)^{2N-j}, \quad \text{pour } 0 \leq i, j \leq 2N.$$

La loi de X_{n+1} sachant que $X_n = i$ est la loi binomiale $B(2N, i/2N)$. On pourra utilisez les fonctions `binomial` et `grand` option `markov`.

[a] Représentez plusieurs trajectoires de cette chaîne de Markov pour N =10, 20, 100.

[b] Confrontez l'estimation numérique de la probabilité que la chaîne soit absorbée en $2N$ au résultat théorique pour N =10, 20, 100..

[c] Proposez une estimation de l'espérance du temps de d'absorption de la chaîne pour N =10, 20, 100. N'oubliez pas l'intervalle de confiance !

7 Modèle de Wright-Fisher avec mutations

On remplace le modèle précédent par la chaîne suivante, avec u et v dans $]0,1[$,

$$p_{ij} = C_{2N}^{j}\left(\frac{i}{2N}(1-u)+\left(1-\frac{i}{2N}\right)v\right)^{j}\left(\left(1-\frac{i}{2N}\right)(1-v)+\frac{i}{2N}u\right)^{2N-j}, \quad \text{pour } 0 \leq i,j \leq 2N.$$

[a] Calculer la mesure invariante π pour $N = 10$ et 20. À partir de quelle valeur de N, le logiciel ne permet plus le calcul de π ?

[b] Proposez une méthode s'appuyant sur le théorème ergodique pour donner une estimation des coefficients de π. Confrontez les résultats de la simulation à ceux de la question précédente pour $N = 10$ et 20.

8 Urnes de Polya

Une urne contient v boules vertes et r boules rouges. A chaque étape, on tire une boule au hasard, on la replace dans l'urne et on ajoute c boules de la couleur de cette boule dans l'urne. Soit X_n la proportion de boules vertes dans l'urne. X_n est une martingale. D'après le théorème de convergence des martingales, elle converge presque surement vers une variable aléatoire X_∞. On s'intéresse à la loi de X_∞.

Proposition 3 *Dans le cas $v = 1$, $r = 1$ et $c = 1$, X_∞ est de loi uniforme sur* $[0, 1]$.

Proposition 4 *Dans le cas général, X_∞ est de loi Béta de paramètres $\frac{v}{c}$ et $\frac{r}{s}$.*

[a] Programmer une fonction qui prend en paramètre le nombre de boules vertes, le nombre de boules rouges, le nombre c de boules que l'on rajoute à chaque étape et le temps de jeu et qui simule X_n .

[b] Faire à présent un programme illustrant la convergence en loi de X_n. (On pourra par exemple simuler m expériences indépendantes et tracer un histogramme qui indique combien d'expériences sont dans l'intervalle $[x, x + dx]$ au temps $t = 100$.)