# TP de Modélisation Stochastique: Illustrations pour l'épreuve (II)

## 1 Leçon sur les processus de branchement (thème 14)

(Programme **Galton.m**) On se propose d'étudier l'évolution d'une population de taille $Z_n$ à la génération $n$, modélisée par le processus de branchement (Galton-Watson) suivant:

- $Z_0 = 1$.
- Chaque individu a pour probabilité $p_k$ d'avoir $k$ descendants, avec $k = 0, 1, \ldots$

On suppose que la loi de fécondité est une loi géométrique translatée de paramètre $p$ avec $p \in ]0,1[$, soit $p_k = p(1-p)^k$. On note $m$ l'espérance de la loi de fécondité: $m = \sum_{k=1}^{+\infty} k p_k = \frac{1-p}{p}$.

1. Dans le cas où $m \leq 1$ (soit $p \geq 0.5$), représenter une évolution de la population ($Z_n$ en fonction de $n$); on la chosira pas trop courte.
2. Dans le cas où $m > 1$ (soit $p < 0.5$), représenter une évolution de la population ($Z_n$ en fonction de $n$), puis représenter plusieurs trajectoires de $m^{-n} Z_n$.

*Les preuves, détails et explications sont donnés dans la leçon 13 des Thèmes de probabilités et statistiques. Le cas de la loi géométrique est détaillé dans le livre de Karlin, Initiation aux processus aléatoires, Dunod, Chapitre 11, et plus précisément p.324-327*

## 2 Leçon sur la convergence d'une martingale (thème 6)

(Programme **Galton.m** modifié légèrement) Même situation que la précédente. On utilise le fait que $m^{-n} Z_n$ est une martingale.

1. Dans le cas où $m > 1$ (soit $p < 0.5$), représenter une évolution de la population ($Z_n$ en fonction de $n$), puis représenter plusieurs trajectoires de $m^{-n} Z_n$.
2. Pour déterminer empiriquement la distribution de la variable aléatoire limite de la martingale, on génère 1000 fois $m^{-50} Z_{50}$ que l'on représente par un histogramme (noter la probabilité de 0 qui est $1/m$).

*Les preuves, détails et explications sont donnés dans la leçon 13 des Thèmes de probabilités et statistiques, précisément section 6. Le cas de la loi géométrique est détaillé dans le livre de Karlin, Initiation aux processus aléatoires, Dunod, Chapitre 11, et plus précisément p.324-327*

## 3 Leçon sur les chaînes de Markov à espace d'états fini (thème 3)

(Programme **Viemort.m**) On considère la dynamique de population suivante, avec $Z_n$ désignant la taille de la population à la génération $n$ et avec $p \in ]0,1[$:

i. $Z_0 \in \{0, 1, .., N\}$ et plus généralement $Z_n \in \{0, 1, .., N\}$.

ii. Pour $k \in \{1, 2, .., N-1\}$, $P(Z_{n+1} = k+1 \mid Z_n = k) = p$ et $P(Z_{n+1} = k-1 \mid Z_n = k) = 1-p$.

iii. Pour $k = 0$, $P(Z_{n+1} = 1 \mid Z_n = 0) = p$ et $P(Z_{n+1} = 0 \mid Z_n = 0) = 1-p$.

iv. Pour $k = N$, $P(Z_{n+1} = N \mid Z_n = N) = p$ et $P(Z_{n+1} = N-1 \mid Z_n = N) = 1-p$.

1. Représenter plusieurs évolutions de la population avec $p < 0.5$, $p = 0.5$ et $p > 0.5$ (on représentera $Z_0, Z_1, .., Z_{1000}$).
2. Représenter par un histogramme une approximation de la mesure stationnaire vers laquelle converge le processus, en choisissant $p < 0.5$, $p = 0.5$ et $p > 0.5$ (on utilisera le théorème ergodique et $Z_0, Z_1, .., Z_{10000}$)

*On pourra utiliser les leçons 10 et 11 des Thèmes de probabilités et statistiques. L'exemple est traité dans le livre de Karlin, Initiation aux processus aléatoires, Dunod, Chapitre 7*

# 4 Leçon sur les chaînes de Markov à espace d'états dénombrable (thème 12)

(Programme **Viemort.m** légèrement modifié) On reprend l'exercice précédent, mais cette fois-ci la chaîne vit sur $\mathbb{N}$, c'est-à-dire que seules les définitions ii et iii sont conservées et $Z_0 \in \mathbb{N}$.

1. Représenter plusieurs évolutions de la population avec $p < 0.5$, $p = 0.5$ et $p > 0.5$ (on représentera $Z_0, Z_1, .., Z_{1000}$).

2. Illustrer le fait que pour $p < 0.5$ le temps de retour en 0 est presque sûrement fini pour $p \leq 0.5$. En calculant la moyenne empirique de $n$ temps de premier retour en 0 pour $p = 0.5$, avec $n = 10, 20, .., 200$, on essayera de montrer que la loi des grands nombres n'est pas vérifiée.

*Voir la leçon 11 (section 7) des Thèmes de probabilités et statistiques. Voir aussi le livre de Karlin, Initiation aux processus aléatoires, Dunod, Chapitre 7*

# 5 Leçon sur le théorème d'arret pour une martingale (thème 5)

(Programme **Viemort.m** légèrement modifié) Même situation que la précédente. On suppose que l'on a $Z_0 = z$ avec $z > 0$. On note $T_a = \inf\{n \mid Z_n \notin ]0, z+a[\}$, où $a \in \mathbb{N}^*$ ($T_a$ désigne la première génération à partir de laquelle la population devient nulle ou bien elle dépasse un certain seuil).

1. Représenter plusieurs évolutions de la population avec $p < 0.5$, $p = 0.5$ et $p > 0.5$ (on représentera $Z_0, Z_1, .., Z_{1000}$).

2. Déterminer empiriquement $P(Z_{T_a} = 0)$ en simulant 1000 fois $Z_{T_a}$. On vérifie que $P(Z_{T_a} = 0) = \dfrac{a}{z+a}$ pour $p = 0.5$.

*Voir la leçon 9 (section 3) des Thèmes de probabilités et statistiques. Voir aussi le livre de Karlin, Initiation aux processus aléatoires, Dunod, Chapitre 7*

# 6 Leçon sur le modèle linéaire gaussien (thème 6)

(Programme **Modelineaire.m**) On va étudier un jeu de donnée fourni dans Stixbox: on l'obtient par getdata(5) et représente les poids du corps et du cerveau pour différents animaux.

1. Après avoir récuperer les données, on fait une première régression. On teste si la régression est légitime, puis on représente les résidus de manière à éliminer deux valeurs abérantes.

2. On recommence une régression sur les données sans valeurs abérantes (animaux préhistoriques). On teste alors la légitimité de la régression.

*L'exemple est traité dans Robust regression and outlier detection, Rousseeuw et Leroy, Wiley, p. 57. On pourra reprendre aussi le cours let l'exemple de la leçon 6 des Thèmes de probabilités et statistiques.*

# 7 Leçon sur les tests du Khi-deux (thème 7)

(Programme **Testkhi2.m**) On reprend l'exercice de vie et de mort sur $\{0, 1, .., N\}$ (on choisira par exemple $N = 10$) avec $p = 0.5$ (voir chaînes de Markov sur espace d'états fini).

1. Représenter plusieurs évolutions de la population (on représentera $Z_0, Z_1, .., Z_{1000}$). Représenter par un histogramme une approximation de la mesure stationnaire vers laquelle converge le processus (on utilisera le théorème ergodique et $Z_0, Z_1, .., Z_{10000}$).

2. On teste par un test du Khi-deux que la distribution de la mesure stationnaire est uniforme sur $\{0, 1, .., N\}$. On fera le test à partir de $Z_0, Z_1, .., Z_n$ pour $n = 100, 1000, 10000$ sur les $N + 1$ classes naturelles.

*Pour le théorème ergodique on consultera la leçon 10 des Thèmes de probabilités et statistiques. La leçon 7 explicite tout ce qui concerne le test du Khi-deux.*