Quick 1. *Indications de correction*

1 Permis à points

1.1 Problème et hypothèses

Dans un pays fort, fort lointain, les accidents de voiture sont si ravageurs qu'un permis à point va être mis en place par les autorités à compter de l'an prochain. Au départ, tous les titulaires du permis se verront attribuer le nombre maximal de points, soit 2 points. Chaque infraction grave entraîne le retrait d'un point. Lorsqu'un conducteur n'a plus de points, son permis lui est retiré temporairement. Au bout d'une année sans infraction, un conducteur peut récupérer 1 point (dans la limite du maximum de 2 points).

Des statistiques observées dans un pays voisin, pionner dans le domaine du permis à points, ont permis d'estimer les quantités suivantes. Un conducteur ayant le nombre maximum de points a 80% de chances de ne commettre aucune infraction dans l'année. Dans 10% des cas, il commet 1 seule infraction grave. Les conducteurs n'ayant plus qu'un seul point, en raison du stress engendré par la crainte de perdre leur permis, sont plus erratiques et ont une chance sur 4 de commettre 1 ou plusieurs infraction(s) grave(s). Les conducteurs sans permis sont échaudés et ne conduisent pas.

On fait l'hypothèse que la population des conducteurs est stable.

1.2 Modélisation markovienne

Question 1.1 : Modèle (2pts)

Modéliser un conducteur de ce pays par une chaîne de Markov $\{X_n\}_{n\geqslant 0}$.

Réponse

On se place à temps discret $n \in \mathbb{N}$, où n modélise l'année considérée [1].

On note X_n le nombre de points du conducteur considéré au 1er janvier [2] de l'année n.

X_n est une chaîne de Markov homogène car d'après l'énoncé, le risque d'infraction grave ne dépend pas de l'historique des infraction du conducteur, mais seulement de son nombre de points actuel. [3]

Son espace d'états est $\mathcal{E} = \{0, 1, 2\}$.

On notera $\pi_{\mathbf{n}}$ le vecteur des probabilités transitoires : $\pi_n(i) = \mathbb{P}[X_n = i]$.

Question 1.2 : Année 0 (1pt)

Quelle est sa distribution initiale π_0 ?

Réponse

L'année 0 (mise en place du permis à points) tout le monde reçoit 2 points, d'où $\mathbb{P}[X_0 = 2] = 1$, soit $\pi_0 = (0, 0, 1)$.

Question 1.3 : Dynamique (3pts)

Donner son graphe de transition ainsi que sa matrice de transition. (Justifier vos réponses)

Réponse

1. unité de temps indiquée par l'énoncé
2. n'importe quelle date fixe ferait l'affaire
3. ici la difficulté était de considérer que l'administration traite le dossier d'un conducteur 1 seule fois par an, afin d'avoir un modèle à temps discret.

- Dans l'état 0 on ne conduit pas $\Rightarrow$ on récupère forcément 1 point l'année suivante :

$$P_{01} = \mathbb{P}[X_{n+1} = 1 | X_n = 0] = 1.$$

- Dans l'état 1, on peut
  - soit perdre 1 point (1 chance sur 4) :

$$P_{10} = \mathbb{P}[X_{n+1} = 0 | X_n = 1] = \frac{1}{4}$$

  - soit en récupérer 1 (aucune infraction) :

$$P_{12} = \mathbb{P}[X_{n+1} = 2 | X_n = 1] = \frac{3}{4}.$$

- Dans l'état 2, on peut :
  - commettre 1 infraction et perdre 1 pt :

$$P_{21} = \mathbb{P}[X_{n+1} = 1 | X_n = 2] = \frac{10}{100}$$

  - ne commettre aucune infraction et garder son capital de points :

$$P_{22} = \mathbb{P}[X_{n+1} = 2 | X_n = 2] = \frac{80}{100}$$

  - commettre plusieurs infractions [4], et perdre tous ses points :

$$P_{20} = \mathbb{P}[X_{n+1} = 0 | X_n = 2] = 1 - \frac{80}{100} - \frac{10}{100} = \frac{1}{10}.$$

On en déduit la matrice de transition P suivante :

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0.25 & 0 & 0.75 \\ 0.1 & 0.1 & 0.8 \end{pmatrix} \tag{1}$$

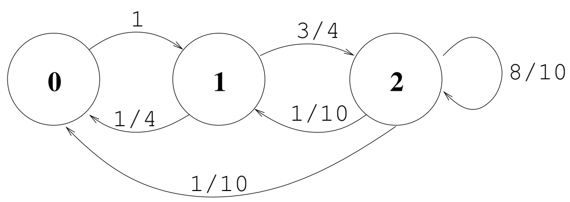

Figure 1 – Graphe de transition

1.3 Calculs de performance

Question 1.4 : Régime transitoire (2pts)

Combien de conducteurs ont encore leurs 2 points au bout d'1 an (année 1) ? Combien sont privés de leur permis ? Mêmes questions pour l'année suivante (année 2).

Réponse

Au bout d'1 an, il y a eu 1 seule transition, c'est donc π_1 que l'on cherche.

$$\pi_1 = \pi_0 P = (0.1, 0.1, 0.8)$$

On cherche la 3^{e} coordonnée : 80% des conducteurs ont encore leurs 2 points au bout d'1 an.

4. la somme des probabilités ci-dessus ne fait pas 1

Question 1.5 : Classification (2pts)

Cette chaîne est-elle irréductible ? Apériodique ? (Prouver votre réponse.)

Réponse

- La chaîne est irréductible car le graphe de transition est connexe (par exemple, de l'état 2 on peut aller partout en 1 étape, et il existe au moins un chemin de probabilité strictement positive de 1 vers 2 (par exemple $P_{12} > 0$) et de 0 vers 2 (0 vers 1 puis 1 vers 2). Plus généralement on peut montrer que pour tout couple d'états (i, j) il existe au moins un chemin de probabilité non nulle : si $i < j$ cela correspond à une suite consécutive de $(j - i)$ infractions graves, et pour $j < i$ à une suite de $i - j$ années consécutives sans infraction.
- La chaîne est apériodique : par exemple il existe un cycle de longueur 1 ($P_{22} > 0$) donc le PGCD de tous les cycles vaut 1.

Donc la CMTD est *ergodique*.

Question 1.6 : Ergodicité (5pts)

Quelle est, après quelques années, la proportion de conducteurs ayant le nombre maximal de points ?

Réponse

La chaîne est ergodique, à espaces d'états fini, donc elle admet une distribution stationnaire. La proportion de conducteurs ayant 2 points, après quelques années de stabilisation, est égale à la probabilité stationnaire d'être dans l'état 2 (propriété d'*ergodicité*).

Calculons donc la loi stationnaire π de $\{X_n\}$. Elle vérifie l'équation d'équilibre

$$\pi = \pi P.$$

En développant :

$$\pi(0) = \frac{1}{4}\pi(1) + \frac{1}{10}\pi(2) \tag{2}$$

$$\pi(1) = \pi(0) + \frac{1}{10}\pi(2) \tag{3}$$

$$\pi(2) = \frac{3}{4}\pi(1) + \frac{8}{10}\pi(2) \tag{4}$$

De l'éq. (4) on déduit (par substitution)

$$\Leftrightarrow \pi(2) = \frac{15}{4}\pi(1) \tag{5}$$

De même en plongeant cette expression dans (3) on obtient

$$\pi(1) = \pi(0) + \frac{1}{10}\frac{15}{4}\pi(1) = \pi(0) + \frac{3}{8}\pi(1)$$

$$\Leftrightarrow \pi(1) = \frac{8}{5}\pi(0) \tag{6}$$

On reporte dans (7) pour obtenir

$$\pi(2) = \frac{15}{4}\frac{8}{5}\pi(0) = 6\,\pi(0). \tag{7}$$

On détermine ensuite $\pi(0)$ grâce à l'équation de normalisation [5] :

$$\pi(0) + \pi(1) + \pi(2) = 1 \tag{8}$$

$$\Leftrightarrow \pi(0)\left(1 + \frac{8}{5} + 6\right) = 1 \qquad \text{en utilisant (7) et (6)} \tag{9}$$

$$\pi(0) = \frac{5}{43} \approx 0.11 \tag{10}$$

5. L'équation (2), elle, ne sert qu'à vérifier ces résultats car elle n'apporte pas d'information supplémentaire.

On en déduit $\pi = (\frac{5}{43}, \frac{8}{43}, \frac{30}{43})$

Donc la proportion de conducteurs ayant tous leurs points tend vers $\pi(2) = \frac{30}{43} \approx 0.7$.

1.4 Intervalle de confiance

Au bout d'un an, on s'aperçoit qu'il faut réévaluer les probabilités d'infraction car le pays voisin les a grossièrement sous-estimées (ou alors, le dispositif en place depuis quelques années a porté ses fruits...).

On cherche d'abord à estimer une probabilité p de commettre au moins une infraction grave pour les conducteurs n'ayant qu'un seul et dernier point de permis. On souhaite que cette probabilité soit estimée à 10^{-2} près.

Question 1.7 : Estimation d'intervalle de confiance (4 pts)

Combien de conducteurs à 1 seul point doit-on interroger pour avoir une estimation correcte ?

Réponse

D'après le TCL, si l'on interroge N conducteurs et que l'on moyenne leurs réponses $Z_1, \dots Z_N$ ($Z_i = 1$ en cas d'infraction dans l'année, 0 sinon), l'intervalle de confiance[6] à 95% sur la moyenne vaut :

$$\widehat{\mu_N} \pm 1.96 \frac{\widehat{\sigma_N}}{\sqrt{N}} \tag{11}$$

où $\widehat{\mu_N} = \frac{1}{N}\sum_{i=1}^{N} Z_i$ est la moyenne empirique et $\widehat{\sigma_N}$ l'écart-type empirique[7].

La largeur de l'intervalle est donc $2 \times 1.96 \frac{\widehat{\sigma_N}}{\sqrt{N}}$. Puisque l'on veut une approximation à $\pm 10^{-2}$ sur la moyenne empirique cela donne l'inéquation suivante :

$$1.96 \frac{\widehat{\sigma_N}}{\sqrt{N}} \geqslant 10^{-2} \tag{12}$$

On ne connaît pas la variance puisque l'on cherche à dimensionner l'expérience. Cependant on peut la majorer (pire cas) par le maximum de la variance d'une loi de Bernoulli[8]. On prend donc $\widehat{\sigma_N} = 0.5$ ce qui donne

$$\sqrt{N} \geqslant \frac{1.96}{2 \times 10^{-2}} \quad \Leftrightarrow N \geqslant 98^2 = 9604 \tag{13}$$

Il faut donc interroger environ 9600 conducteurs pour avoir une estimation de p à 10^{-2} près (en pratique, un peu moins si la probabilité p est très faible, ce qui réduit la variance).

1.5 Modèle à mémoire

En réalité, il existe plusieurs sortes de comportements. Un conducteur possédant 1 point de permis est soit un conducteur attentif qui vient exceptionnellement de perdre 1 point, soit un conducteur distrait qui qui était suspendu et vient de récupérer un précieux point. Heureusement, dans la vie rien n'est jamais figé et les gens peuvent changer de catégorie (dans les 2 sens) selon le modèle suivant :

- un conducteur à 2 points se sent confiant. Il ne commet aucune infraction avec probabilité 0.8, et 1 seule infraction avec probabilité 0.1, *quel que soit le nombre de points qu'il avait l'année d'avant.*
- un conducteur qui avait 2 points l'an passé et qui en a perdu 1 seul, redouble de prudence. Il ne commet aucune infraction avec probabilité 0.9.
- un conducteur qui était suspendu (0 pt) et qui vient de récupérer 1 pt est plus enthousiaste. Il risque de commettre 1 ou plusieurs infraction(s) avec probabilité 0.4.

Question 1.8 : Mémoire

6. niveau de condiance *classique* ; on aurait aussi pu choisir 99% par exemple, ce qui change juste la valeur de quantile
7. cf. par exemple l'ouvrage de J.Y. LeBoudec, p. 43
8. La variance d'une loi de Bernoulli $\mathcal{B}(p)$ vaut $p(1-p)$. Elle peut donc être majorée par son maximum $\frac{1}{4}$ *(lorsque* $p = \frac{1}{2}$*)*. L'écart type vaut donc au pire 1/2.

Expliquer pourquoi le nombre de points d'un conducteur donné n'est pas une chaîne de Markov.

Question 1.9 : Nouvelle CMTD

Proposer un nouveau modèle Markovien Y_n tenant compte de cet effet "mémoire".

Question 1.10 : Dynamique du modèle

Donner au choix le graphe ou la matrice de transition de Y_n.

Question 1.11 : Nouveaux résultats

Justifier que Y_n est ergodique et en déduire, à long terme, la proportion moyenne de conducteurs qui ont 2 points.

1.6 Généralisation

On veut déterminer le nombre de points idéal K. Pour cela on généralise le modèle de la façon suivante :

- le nombre d'infractions à 1 point commise dans l'année est une variable aléatoire de loi de Poisson de paramètre λ_i lorsque le conducteur possède i points[9], pour $1 \leqslant i \leqslant K$.
- un conducteur au permis suspendu qui conduit illégalement et commet une infraction (ou plusieurs) va en prison pour 1 an. À sa sortie, il doit repasser son permis, ce qu'il réussit avec une probabilité p. Il bénéficiera alors de K points.

Question 1.12 : Espace d'états

Reprendre le modèle de la question 1.2 et l'adapter au nouveau problème. Quel est son espace d'états ? Quel est son graphe de transition ?

Question 1.13 : Probabilités de transition

Calculer, en fonction des λ_i et de p, pour les états 0,1, $K-1$ et K les probabilités de transition. Donner l'allure générale de la matrice de transition.

Question 1.14 : Simulation

Écrire, en langage R (ou C selon votre convenance) un programme de simulation du régime transitoire de ce système pour $K = 12$ (en supposant la matrice P pré-remplie).

Question 1.15 : Plan d'expérience

On suppose les matrices de transition P_K fournies pour toutes les valeurs de K entre 1 et 15. Proposer une expérience de simulation permettant de comparer ces options et donner le code correspondant.

9. on suppose les λ_i estimés statistiquement dans un pays voisin