Ecole Polytechnique
Année 1998-1999

Petites classes de Probabilité

Christian Robert

Exercices

1. – On considère une grille $n \times p$ où les déplacements possibles sont de (i, j) vers $(i+1, j)$ (si $i < n$) et vers $(i, j+1)$ (si $j < p$).

(a) Donner le nombre de trajets différents de $(1, 1)$ vers (n, p).

(b) Indiquer comment ce nombre est modifié si le déplacement de (i, j) vers $(i+1, j+1)$ (pour $i < n$ et $j < p$) est également autorisé.

2. – Si n personnes sont réunies dans une pièce, quelle est la probabilité qu'aucune paire d'individus n'ait un anniversaire commun? Déterminer $n_{\min}$ pour que cette probabilité soit inférieure à 0.5.

3. – Soient r objets à répartir dans n cases ($r \geq n$). Donner le nombre K_n^r de répartitions distinctes et le nombre de répartitions où chaque case contient au moins un objet.
Déduire le nombre de solutions (dans $\mathbb{N}^k$) à l'équation $n_1 + \ldots + n_k = r$.
Approcher la probabilité que n objets occupent exactement les n cases à l'aide de la formule de Stirling.

4. – Soient $A_1, \ldots, A_n$ n événements sur un même espace probabilisé (Ω, P). Donner la probabilité

$$P(A_1 \cup \ldots \cup A_n)$$

en fonction des probabilités $P(A_i)$, $P(A_i \cap A_j)$, …, $P(A_1 \cap \ldots \cap A_n)$ *(formule de Poincaré).*

Problème I

Soient N enveloppes associées à N lettres. On apparie au hasard lettres et enveloppes. Soit ϱ_0, probabilité qu'aucune lettre ne soit adressée au bon destinataire.

1. On note A_k l'événement que l'appariement soit correct pour la k-ième enveloppe. Montrer que $P(A_k) = 1/N$ et donner $P(A_{i_1} \cap \ldots \cap A_{i_k})$.
2. Calculer la valeur de ϱ_0 en utilisant la formule de Poincaré et montrer que $\varrho_0 \simeq e^{-1}$.
3. Reprenant l'exercice 3, montrer que la probabilité que k régions données parmi n soient vides si r objets sont à répartir parmi ces n régions est
$$\left(1 - \frac{k}{n}\right)^r .$$
4. Utilisant à nouveau la formule de Poincaré, montrer que la probabilité que toutes les régions soient occupées est
$$\varrho_0 = \sum_{\nu=0}^{n} (-1)^\nu \binom{n}{\nu} \left(1 - \frac{\nu}{n}\right)^r .$$

Reprendre le même raisonnement pour montrer que la probabilité que exactement m régions soient vides est

$$\varrho_m = \binom{n}{m} \sum_{\nu=0}^{n-m} (-1)^\nu \binom{n-m}{\nu} \left(1 - \frac{\nu+m}{n}\right)^r .$$

6. Déduire de
$$\left\{ne^{-(\nu+r)/(n-\nu)}\right\}^\nu < \nu! \binom{n}{\nu} \left(1 - \frac{\nu}{n}\right)^r < \left\{ne^{-r/n}\right\}^\nu$$
que, si le rapport $\lambda = ne^{-r/n}$ reste constant et n tend vers $+\infty$, ϱ_m peut être approchée par $e^{-\lambda}\lambda^m/m!$. (*Cette loi s'appelle la loi de Poisson.*)

7. Appliquer à l'exercice 2 pour un village de 1900 habitants en déterminant la probabilité approchée qu'il y ait dix jours de l'année sans anniversaire.

Problème II

Ce problème étudie une *marche aléatoire* schématisée par une séquence de *pile ou face* où on représente un *pile* par un accroissement de 1 dans le "gain" S_n et un *face* par une diminution de 1 de S_n (avec la convention que $S_0 = 0$). On représente la succession des gains et pertes par un graphe (ou *chemin*) reliant les valeurs successives de S_n en fonction de $n = 1, 2, \ldots$

1. Montrer le principe de réflection: *Si $A = (S_m, m)$ et $B = (S_n, n)$ avec $S_m > 0$, $S_n > 0$ et $n > m$, le nombre de chemins de A à B qui touchent l'axe des x est identique au nombre de chemins joignant $A' = (-S_m, m)$ à B.*

2. Donner la probabilité $P(S_n = r)$. En déduire la probabilité d'un retour à l'origine $P(S_{2t} = 0) = u_{2t}$ et montrer que u_{2t} peut s'approcher par $1/\sqrt{\pi t}$.

3. On considère l'événement que le *premier* retour à l'origine (gain nul) se produise en $2v$, de probabilité f_{2v}. Montrer que
$$u_{2t} = f_2 u_{2t-2} + \ldots + f_{2t} u_0.$$

4. Montrer que la probabilité qu'aucun retour à l'origine ne se soit produit jusqu'à l'étape $2n$ (incluse) est égale à u_{2n}. En déduire que $f_{2n} = u_{2n-2} - u_{2n}$.

5. Une valeur importante est le temps de *dernier retour* à l'origine, T, pour un jeu de $2n$ coups. Déduire des questions précédentes que la probabilité de dernier retour en $2k$ est $\alpha_{n,k} = u_{2k} u_{2n-2k}$.

6. Utiliser l'approximation de 2. pour approcher $\alpha_{10,k}$ ($k = 1, \ldots, 10$). Pour n assez grand, en déduire que la probabilité que S_k reste positif au plus $x\%$ des coups est (approximativement) $\frac{2}{\pi} \arcsin \sqrt{x}$.

Exercices

1. – Un examen comporte 15 questions, chacune admettant 3 réponses possibles. Les étudiants répondent à chaque question indépendamment; on suppose que les étudiants ayant préparé l'examen sont en proportion 70% et répondent correctement à une question avec une probabilité de 0.8, les 30% d'étudiants restant choisissant les réponses au hasard. Il faut au moins 8 bonnes réponses pour réussir l'examen.

(a) Si un étudiant échoue, quelle est la probabilité qu'il ait préparé l'examen?

(b) Soit M le nombre moyen de bonnes réponses pour un étudiant ayant préparé l'examen. Si un étudiant obtient cette note M, quelle est la probabilité qu'il n'ait pas préparé l'examen?

2. – Une mouche pond N œufs où N est distribué suivant une loi de Poisson $\mathcal{P}(\lambda)$. Chaque œuf éclot avec probabilité p.

(a) A N donné, quelle est la loi de X, nombre d'œufs éclos? En déduire la loi marginale de X et sa fonction génératrice.

(b) Sachant que x œufs ont éclos, quelle est la loi de N?

3. – Une population de N individus comporte un nombre inconnu N_A d'individus malades. On suppose que N_A peut prendre indifféremment les valeurs $0, \ldots, N$. On a sélectionné n individus dans cette population et obtenu n_A individus malades.

(a) Donner la loi de n_A conditionnellement à N, N_A et n.

(b) Actualiser la distribution de probabilité de N_A en fonction de ces observations.

(c) Donner la probabilité que le prochain individu sélectionné soit malade. (*Ce résultat est parfois appelé loi de succession de Laplace dans le cas particulier où* $n_A = 0$.)

4. – Un jeu télévisé consiste à choisir une porte parmi trois qui donnent sur deux tickets de métro et une Ferrari. Le joueur ayant choisi sa porte, le meneur de jeu ouvre l'une des deux autres portes qui donne sur un ticket de métro. Expliquer pourquoi cette "information" ne modifie pas les chances du joueur mais que, par contre, il a toujours intérêt à changer avec la seconde porte qui n'a pas été ouverte.

Problème I. – Lemme de Borel-Cantelli et décompositions binaires.

Sur $[0, 1]$, on définit les fonctions dyadiques, d_n $(n \geq 1)$: $d_n(\omega)$ est le n-ième chiffre dans la décomposition binaire de $\omega \in [0, 1]$. Pour les décompositions multiples, comme $1 = 1.0 = 0.1111\ldots$ ou $1/2 = 0.1 = 0.0111\ldots$, on choisit la décomposition infinie, c'est-à-dire $d_n(1) = 1$ pour tout n. L'intervalle $[0, 1]$ est muni de la loi uniforme, i.e. P telle que $P([a, b]) = b - a$ pour tout intervalle $[a, b]$ de $[0, 1]$.

1. Montrer que

$$P(\omega : d_j(\omega) = u_j) = \frac{1}{2} \qquad \text{et} \qquad P(\omega : d_i(\omega) = u_i, i = 1, \ldots, n) = \frac{1}{2^n}$$

quelques soient n, j, u_j et u_i. En déduire que

$$P(\omega : \sum_{i=1}^{n} d_i(\omega) = k) = \binom{n}{k} \frac{1}{2^n} \quad 0 \leq k \leq n.$$

2. Soient les fonctions l_n définies sur $[0, 1]$ par

$$l_n(\omega) = k \quad \text{si} \quad d_n(\omega) = \ldots = d_{n+k-1}(\omega) = 0.$$

Montrer que

$$P(\omega : l_n(\omega) = k) = \frac{1}{2^{k+1}} \qquad \text{et} \qquad P(\omega : l_n(\omega) \geq r) = \frac{1}{2^r}.$$

3. Soit une suite croissante de réels r_n. Montrer que, si $\sum_n 2^{-r_n}$ converge et si $A_n = \{\omega : l_n(\omega) \geq r_n\}$, alors

$$P(\limsup A_n) = 0.$$

Considérant le cas particulier $r_n = (1 + \epsilon) \log_2 n$, en déduire que

$$P(\omega : \limsup \frac{l_n(\omega)}{\log_2 n} \leq 1) = 1.$$

4. Montrer que, si $B_n = \{\omega : l_n(\omega) = 0\}$,

$$P(\limsup B_n) = 0.$$

5. Considérant une suite croissante (r_n) de réels positifs telle que la série $\sum_n 2^{-r_n} r_n^{-1}$ diverge, montrer que $P(\limsup A_n) = 1$. En prenant $r_n = \log_2 n$, déduire que

$$P(\omega : \limsup \frac{l_n(\omega)}{\log_2 n} = 1) = 1.$$

Exercices

1. – Pour estimer la taille d'une population animale, on installe successivement r trappes. Soit q la probabilité qu'un animal soit capturé par une trappe et N la taille de la population avant l'expérience.

a. Supposant que les animaux capturés ne sont pas relachés, donner la probabilité que les effectifs de capture soient $n_1, \ldots, n_r$.

b. Une méthode alternative consiste à relacher les animaux après *marquage*. Donner la probabilité que les effectifs des captures soient $n_1, \ldots, n_r$ et que les effectifs d'animaux *recapturés* soient $m_1, \ldots, m_{r-1}$.

c. On considère maintenant n_1 et n_2 fixés à l'avance. Donner la distribution de m_1, nombre de recaptures dans la seconde trappe, et en déduire un estimateur intuitif de N.

2. – Soit une population à k modalités en proportions $p_1, \ldots, p_k$. On en tire n individus (avec remise).

a. Déterminer la loi de $(n_1, \ldots, n_k)$, nombres d'individus prenant les modalités $1, \ldots, k$. En déduire la loi marginale de n_1.

b. Même question si le tirage s'opère sans remise.

3. – La ville de Londres étant divisée en 576 districts, on relève le nombre d'impacts de bombes subi par chaque district pendant la Seconde Guerre Mondiale. Les résultats sont donnés dans le tableau suivant:

k	0	1	2	3	4	5 et plus
N_k	229	211	93	35	7	1

a. Comment tester l'adéquation de ces observations à une loi de Poisson? (On rappelle que, si $Z \sim \chi_4^2$, $P(Z \leq 7.8) = 0.9$ et $P(Z \leq 9.5) = 0.95$.)

b. Si l'adéquation à une loi de Poisson est acceptable, donner la probabilité que la dernière observation vaille 6.

4. – Trouver la probabilité (approximative) que, sur 500 personnes, $k = 0, 1, 2, 3$ aient leur anniversaire le 20 janvier.

5. – Une compagnie de téléphone déservant 2000 abonnés cherche à déterminer le nombre N de lignes nécessaires pour qu'au plus un appel sur cent ne soit pas immédiatement satisfait. En moyenne, chaque utilisateur donne un appel de deux minutes toutes les heures. Déterminer $N_{\min}$ en utilisant les approximations de Poisson et de De Moivre.

6. – Soit $N_n \sim \mathcal{B}(n, p)$. On note

$$N_p^* = \frac{N_n - np}{\sqrt{np(1-p)}}.$$

a. Montrer que, si (x_n) est une suite croissante vers ∞, on a, pour ν fixé,

$$\frac{P(N_n^* \geq x_n + \nu)}{P(N_n^* \geq x_n)}$$

qui tend vers 0 quand n tend vers ∞.

b. En déduire que, si $x_n^3/\sqrt{n}$ tend vers 0 et x_n tend vers ∞, il vient

$$P(N_n^* \geq x_n) \approx \frac{1}{\sqrt{2\pi}} \frac{e^{-x_n^2/2}}{x_n}.$$

c. Construire la fonction génératrice de N_n^*, $M_n(t) = \mathbb{E}[\exp(tN_n^*)]$ et déduire de l'inégalité de Chernoff,

$$P(N_n^* \geq s) \leq \inf_{t>0} M_n(t)\, e^{-ts},$$

une majoration de $P(N_n^* \geq s)$.

Problème. – Premier temps de passage d'un processus.

On modélise l'arrivée de voitures à un passage clouté par une succession de secondes indépendantes où chaque seconde a la probabilité p de voir passer une voiture.

1. Sachant qu'un piéton a besoin de trois secondes pour traverser, déterminer la probabilité qu'un piéton attende 0, 1, 2 ou 3 secondes.
2. Etant donné une suite de variables de Bernoulli indépendantes, $X_1, \ldots, X_n, \ldots$, on note $\mathcal{E}_n = \mathcal{E}(X_1, \ldots, X_n)$ un événement impliquant les n premiers éléments de la suite. On suppose que $\mathcal{E}$ vérifie la *propriété de récurrence* suivante:

$$P(\mathcal{E}(X_1, \ldots, X_n) \text{ et } \mathcal{E}(X_1, \ldots, X_{n+m})) = P(\mathcal{E}(X_1, \ldots, X_n))P(\mathcal{E}(X_1, \ldots, X_m)).$$

Montrer que, si $u_n = P(\mathcal{E}_n)$ et si f_n est la probabilité d'une première occurence de $\mathcal{E}$ pour $X_1, \ldots, X_n$, on a

$$u_n = f_1 u_{n-1} + f_2 u_{n-2} + \ldots + f_n. \tag{1}$$

De même, montrer que, si $f_n^{(r)}$ est la probabilité d'une r-ième occurence de $\mathcal{E}$ en n, on a

$$f_n^{(r)} = f_1 f_{n-1}^{(r-1)} + f_2 f_{n-2}^{(r-1)} + \ldots + f_{n-1} f_1^{(r-1)}.$$

En déduire que, si $f = \sum_{n=1}^{\infty} f_n$,

$$f^r = \sum_{n=1}^{\infty} f_n^{(r)}.$$

Interpréter ce résultat.

3. Soient
$$U(s) = 1 + \sum_{n=1}^{\infty} u_n s^n \quad \text{et} \quad Q(s) = \sum_{n=1}^{\infty} f_n s^n.$$
Montrer que ces séries sont définies pour $0 \leq s < 1$. Déduire de (1) que
$$U(s) = \frac{1}{1 - Q(s)}$$
et que $u_n = f_n^{(1)} + f_n^{(2)} + \ldots$
4. Déterminer $\mathcal{E}$ pour le modèle du passage de voitures. Montrer que, si $q = 1 - p$,
$$U(s) = \frac{1 - s - qp^3 s^4}{(1 - s)(1 - p^3 s^3)}$$
et
$$Q(s) = \frac{p^3 s^3}{1 - qs(1 + ps + p^2 s^2)}$$
en utilisant l'égalité $u_n + u_{n-1} q + u_{n-2} q^2 = q^3$ $(n \geq 3)$.
5. Retrouver f_0, f_1, f_2 et déterminer le temps moyen d'attente.

Problème I. – Précision d'un sondage.

Lors d'un référendum, la proportion p de votants favorables à l'objet du référendum est inconnue et on cherche à l'approcher par un sondage d'opinion.

1. Si, parmi $n = 989$ personnes interrogées, N_n sont favorables, donner la loi de N_n. On observe $N_n = 539$. Si $p = 1/2$, quelle est la probabilité d'observer au moins ce nombre, i.e. $P(N_n \geq 539)$?
2. Si $p = 0.55$, quelle est la borne I telle que $P(N_n \geq I) = 0.95$?
3. On cherche un intervalle $[a, b]$ sur p dépendant de N_n qui contienne p avec probabilité 0.95 *au moins*. Donner la valeur de cet intervalle pour $N_n = 539$.
4. Déterminer une région exacte à 95% pour p à partir de l'approximation
$$\frac{N_n - np}{\sqrt{np(1-p)}} \approx \mathcal{N}(0,1).$$
5. Donner la taille minimale de l'échantillon de sondage pour atteindre une précision d'au moins 1% sur p (avec probabilité 95%).

Problème II. – Processus de branchement

Soient $X_1, \ldots, X_n, \ldots$ des variables entières de même fonction génératrice $f(s) = \mathbb{E}[s^{X_i}]$.

1. Soit N suivant une loi de Poisson $\mathcal{P}(\lambda)$. On définit $S = X_1 + \ldots + X_N$. Montrer que S est *infiniment divisible*, i.e. que, pour tout $m \in \mathbb{N}$, elle peut se représenter comme la somme de m variables Y_i^m de même loi.
2. On suppose que les X_i sont des variables aléatoires décrivant le nombre de descendants d'une paramécie. Les descendants directs des paramécies de la génération t forment la génération $t+1$ et on note Z^t la taille de la génération t. Ecrire Z^{t+1} en fonction des descendants de chaque paramécie de la génération t, $X_1^t, \ldots, X_{Z^t}^t$.
3. Partant d'une paramécie à l'instant 0, $Z^0 = 1$, montrer que la fonction génératrice de Z^t est $f^t(s) = f(f^{t-1}(s))$ $(T > 1)$. Quelle est cette fonction si $X_i \sim \mathcal{B}e(p)$?
4. Une caractéristique importante concerne les chances de survie de la population. Soit ϱ la probabilité que la population s'éteigne et ϱ_t celle qu'elle ait disparu à la génération t, $\varrho_t = P(Z^t = 0)$ (ou avant). Exprimer ϱ_t en fonction de f et déduire de la relation de récurrence entre ϱ_t et ϱ_{t-1} une équation définissant ϱ.
5. Déduire de la convexité de f que le fait que ϱ soit égale à 1 dépend de $f'(1)$. Interpréter en termes de $\mathbb{E}[X_1]$.
6. Montrer que, si $\varrho < 1$, $f^t(s)$ tend vers ϱ pour tout $0 < s < 1$ et interpréter.
7. Considérant Y^t, nombre de descendants de la paramécie jusqu'à la génération t,

déduire de

$$Y^t = 1 + Z^1 + \ldots + Z^t$$

que la fonction génératrice g_t de Y^t est telle que $g_t(s) = sf(g_{t-1}(s))$.

Exercices

1. – Soit X_n une chaîne de Markov sur $\mathcal{X}$ fini ou dénombrable de matrice de transition $P = (p_{xy})$ et on note $P^n = (p^n_{xy})$. On définit la *période* de $x \in \mathcal{X}$, d_x, comme le p.g.c.d. des n tels que $p^n_{xx} > 0$. Par convention, $d_x = 0$ si $p^n_{xx} = 0$ pour tout n.

(a) Soient x et y dans la même classe irréductible de la chaîne; montrer que x et y ont même période.

(b) Soit E une classe irréductible et $x \in E$ de période $d_x > 0$. On définit

$$C_k(x) = \{y \in E;\ \text{si } p^n_{xy} > 0,\ n \equiv k(d_x)\}$$

avec $0 \le k \le d_x - 1$. Montrer que $C_k(x)$ est indépendant de x.

(c) Montrer que $C_0, \dots, C_{d_x-1}$ forment une partition de E. Donner la forme de P sur cette partition.

2. – Etudier la chaîne de Markov de matrice de transition

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/3 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

en termes d'états absorbants, irréductibles et de loi stationnaire.

3. – On se donne une suite (π_k) de nombres positifs telle que la série associée soit convergente et une suite doublement indicée, (q_{mn}) telle que

$$\sum_{n=0}^{\infty} q_{nm} = 1 \qquad \forall m \in \mathbb{N}.$$

On définit une suite (X_k) de variables aléatoires sur $\mathbb{N}$ de la manière suivante: étant donné X_k, on génére Y_k suivant la loi de probabilité $P(Y_k = n | X_k = m) = q_{nm}$ et on prend

$$X_{k+1} = \begin{cases} Y_k & \text{avec probabilité } \varrho, \\ X_k & \text{avec probabilité } 1 - \varrho, \end{cases}$$

où

$$\varrho = \frac{\pi_{Y_k}}{\pi_{X_k}} \frac{q_{X_k Y_k}}{q_{Y_k X_k}} \wedge 1.$$

(a) Montrer que cette construction définit une chaîne de Markov et exprimer les composantes de la matrice de transition.

(b) Montrer que la chaîne est irréductible sous l'une des deux conditions suivantes:

   (i) $q_{mn} > 0$ si $\pi_n > 0$ et $\pi_m > 0$.

   (ii) $q_{m(m-1)} > 0$ et $q_{m(m+1)} > 0$ et le support de π, i.e. l'ensemble des m tels que $\pi_m > 0$, est de la forme $\{M, M+1, \ldots, M+T\}$.

(c) Déduire l'intérêt de cette construction pour la *simulation* de la loi π de probabilités

$$\pi(k) = \frac{\pi_k}{\sum_{t=1}^{\infty} \pi_t}$$

du fait que π est la loi stationnaire de X_k sous la condition (i) ci-dessus. *(Méthode de Metropolis–Hastings)*

Problème – Modèle de diffusion de Bernouilli-Laplace.

On considère deux récipients de même volume remplis de ϱ molécules de type A et de ϱ molécules de type B (au total). On modélise les échanges entre les deux récipients comme une suite de permutations discrètes d'une molécule du premier récipient et d'une molécule du second récipient. On note X_k le nombre de molécules de type A à l'étape k dans le premier récipient. (Notons que X_k caractérise entièrement le système.)

1. Montrer que la suite X_k forme une chaîne de Markov et donner sa matrice de transition.

2. Montrer que la chaîne (X_k) est irréductible et récurrente.

3. On cherche à caractériser la loi stationnaire de ce modèle de diffusion et, plus généralement, des chaînes de Markov de matrices de transition (p_{ij}) telles que

$$p_{ij} = 0 \qquad \text{si} \qquad |i-j| > 1.$$

   Montrer que la loi stationnaire, de vecteur de probabilités (π_k), satisfait:

$$\begin{aligned}
\pi_0 &= \pi_0 p_{00} + \pi_1 p_{10}, \\
\pi_1 &= \pi_0 p_{01} + \pi_1 p_{11} + \pi_2 p_{21}, \\
&\cdots \\
\pi_\varrho &= \pi_\varrho p_{\varrho\varrho} + \pi_{\varrho-1} p_{\varrho-1\varrho-1}.
\end{aligned}$$

   En déduire que les π_k satisfont

$$\begin{aligned}
\pi_1 &= \frac{p_{01}}{p_{10}} \pi_0, \\
\pi_2 &= \frac{p_{01} p_{12}}{p_{10} p_{21}} \pi_0, \\
\pi_3 &= \frac{p_{01} p_{12} p_{23}}{p_{10} p_{21} p_{32}} \pi_0, \\
&\cdots
\end{aligned}$$

4. En déduire que, pour le modèle de Bernouilli-Laplace, la loi stationnaire π vérifie

$$\pi_k = \binom{\varrho}{k}^2 \pi_0 = \binom{\varrho}{k}^2 \Big/ \binom{2\varrho}{\varrho},$$

c'est-à-dire que π est une loi hypergéométrique, $\mathcal{H}(2\varrho, \varrho, 1/2)$.

5. Utiliser le résultat de 3. pour montrer que la loi stationnaire du modèle de Ehrenfest est une loi binomiale $\mathcal{B}(\varrho, 1/2)$.

Exercices

1. – Soit (X_n) chaîne de Markov irreductible sur $\mathcal{X}$ fini ou dénombrable. Montrer que l'une des deux propositions

(i) Tous les états de $\mathcal{X}$ sont transients,

$$P_i(\cup_j\{X_n = j \ i.o.\}) = 0, \qquad \sum_n p_{ij}^{(n)} < \infty,$$

pour tout couple (i,j).

(ii) Tous les états de $\mathcal{X}$ sont récurrents,

$$P_i(\cap_j\{X_n = j \ i.o.\}) = 1, \qquad \sum_n p_{ij}^{(n)} = \infty,$$

pour tout couple (i,j).

est nécessairement satisfaite.

2. – Sur $\mathbb{R}^k$, on considère le réseau $\mathcal{S}$ des points de coordonnées entières. On définit (X_n) *marche aléatoire* symétrique sur $\mathcal{S}$ par un déplacement de X_n à l'un des $2k$ voisins avec probabilité $1/2k$. On suppose $x_0 = 0$.

(a) Soit $k = 2$. Déduire de

$$P(X_{2n} = 0) = 4^{-2n} \sum_{m=0}^{n} \frac{2n!}{(m!(n-m)!)^2} = 4^{-2n} \binom{2n}{n}^2$$

que cette marche est récurrente.

(b) Pour la marche aléatoire correspondante sur $\mathbb{R}^3$, déduire de

$$\begin{aligned} P(X_{2n} = 0) &= 2^{-2n} \binom{2n}{n} \sum_{j,k} \left\{ \frac{3^{-n} n!}{j!k!(n-j-k)!} \right\}^2 \\ &\le 2^{-2n} \binom{2n}{n} \max_{j,k} \frac{3^{-n} n!}{j!k!(n-j-k)!} \le C n^{-3/2} \end{aligned}$$

que cette marche aléatoire est transiente.

3. – Soit $\mathbb{P}$ matrice de transition, de matrice limite (stationnaire) A.

(a) Montrer que la *matrice fondamentale*

$$\mathbb{Z} = (I - (\mathbb{P} - A))^{-1}$$

existe, que

$$\mathbb{Z} = I + \sum_{n=1}^{\infty} (\mathbb{P}^n - A).$$

$\pi\mathbb{Z} = \pi$ et $\mathbb{P}\mathbb{Z} = \mathbb{Z}\mathbb{P}$, où π représente une ligne de A, i.e. la loi stationnaire associée à $\mathbb{P}$.

(b) Soit $N_j(n)$ le nombre de fois où la chaîne visite l'état j durant les n premiers instants. Montrer que, pour toute distribution initiale μ,

$$\lim_{n\to\infty} \mathbb{E}_\mu[N_j(n)] - n\pi_j = \mu(\mathbb{Z} - A).$$

En déduire que, pour tout couple (u, v),

$$\lim_{n\to\infty} \mathbb{E}_u[N_j(n)] - \mathbb{E}_v[N_j(n)] = z_{uj} - z_{vj}.$$

(c) Soit f_j le nombre de pas avant une visite de j. Montrer que la matrice $\mathbb{M}$ des $m_{ij} = \mathbb{E}_i[f_j]$ vérifie

$$\mathbb{M} = \mathbb{P}(\mathbb{M} - \mathbb{M}_d) + E,$$

où $\mathbb{M}_d$ est la matrice diagonale de même diagonale que $\mathbb{M}$ et E est la matrice faite de 1. Montrer que, pour chaque couple de lois initiales, (μ, ν),

$$\mathbb{E}_\mu[f_i] - \mathbb{E}_\nu[f_i] = (\mu - \nu)(I - \mathbb{Z})\mathbb{D},$$

où $\mathbb{D}$ est la matrice diagonale $\mathrm{diag}(1/\pi_i)$.

(d) Soit h fonction à valeurs dans $\{1, \ldots, r\}$, avec $h(i) = h_i$ et (X_n) chaîne de Markov irreductible, montrer que

$$\lim_{n\to\infty} \frac{1}{n} \mathrm{Var}\left(\sum_{t=1}^{n} h(x_t)\right) = \sum_{i,j} h_i c_{ij} h_j,$$

où $c_ij = \pi_i z_{ij} + \pi_j z_{ji} - \pi_i \delta_{ij} - \pi_i \pi_j$ et δ_{ij} est le symbole de Kroenecker.

Problème I. – Files d'attente.

Une file d'attente est caractérisée par la loi du service et la loi d'arrivée des clients. On la représente généralement par la notation “$A/B/n$” où A caractérise la loi d'arrivée, B la loi de service et n le nombre de serveurs.

1. $M/G/1$: ce modèle est décrit par une arrivée des clients suivant un processus de Poisson de paramètre λ et un temps de service quelconque, régi par une fonction de répartition F. On note X_n le nombre de clients en attente quand le n-ème client commence à être servi.

a) On rappelle que le nombre moyen de clients arrivant sur une période t suit une loi de Poisson de paramètre λt. Montrer que les X_n forment une chaîne de Markov et que la matrice de transition fait intervenir les probabilités

$$a_k = \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^k}{k!} dF(t).$$

b) Déduire de l'exercice sur les processus de branchement [II - PC(4)] que, si $\mu = \sum_k k a_k > 1$, la chaîne de Markov est transiente et que, si $\mu \leq 1$, elle est récurrente.

c) Montrer de plus que, si $\mu = 1$, la chaîne est *récurrente nulle* et n'admet pas de loi stationnaire.

2. $M/M/1$: dans ce cas, la loi du service est aussi régie par une loi exponentielle, i.e. $F(t) = 1 - e^t$.

   a) Donner les probabilités a_k et en déduire que $\mu = \lambda$.

   b) Montrer que la loi stationnaire de la chaîne est donnée par $\pi_j = (1 - \lambda)\lambda^j$ si $\lambda < 1$.

3. $M/M/\infty$: ce modèle a une infinité de serveurs et on considère

$$X_{n+1} = \sum_{k=1}^{X_n} \xi_k + Y_{n+1}$$

où les ξ_k sont indépendants de loi de Bernoulli, $\mathcal{B}e(p)$, et les Y_n sont des variables de Poisson $\mathcal{P}(\lambda)$.

   a) Montrer que ce processus forme une chaîne de Markov et exprimer les probabilités de transition.

   b) Donner $\mathbb{E}[X_1|X_0 = x]$ et en déduire que la chaîne est récurrente pour tout $p < 1$. Montrer que $\mathbb{E}[X_n|X_0 = x] = \nu + p^n(x - \nu)$ où $\nu = \lambda/(1 - p)$.

   c) Montrer que la loi de Poisson $\mathcal{P}(\nu)$ est la loi stationnaire de cette chaîne.

Problème II. – Chaînes duales.

Soit $f(x|\theta)$ une densité conditionnelle sur x et $\pi(\theta)$ une densité *marginale* sur θ. On introduit

$$m(x) = \int f(x|\theta)\pi(\theta)d\theta, \qquad \pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)},$$

loi marginale de x et loi conditionnelle de θ. Soient les noyaux de transition

$$R(\theta, \eta) = \int f(x|\theta)\pi(\eta|x)dx, \qquad \tilde{R}(x, y) = \int \pi(\theta|x)f(y|\theta)d\theta.$$

(a) Montrer que la chaîne de Markov associée à R est récurrente si, et seulement si, la chaîne de Markov associée à $\tilde{R}$ est récurrente.

(b) Montrer que la mesure π est invariante pour la chaîne de Markov associée à R.

(c) Donner les transitions dans le cas où $f(x|\theta)$ correspond à la loi $\mathcal{N}(\theta, 1)$ et $\pi(\theta) \propto \exp(-(\theta - \mu)^2/2\sigma)$.

Exercices

1. – Soit $x = (y^t, z^t)^t \in \mathbb{R}^p$ avec $y \in \mathbb{R}^q$ $(q < p)$. Montrer que, si x suit une loi de densité $f(||y-\theta||^2 + ||z||^2)$ par rapport à la mesure de Lebesgue, la loi de

$$T(x) = \frac{||y-\theta||^2}{||z||^2}\frac{p-q}{q}$$

est la loi de Fisher $\mathcal{F}_{q,p-q}$, quelque soit f.

2. – Soient $x, y \sim \mathcal{U}_{[0,1]}$. Montrer que, pour $\alpha \in]-1,1[$, les lois de fonction de répartition

$$H_\alpha(x,y) = xy[1 + \alpha(1-x)(1-y)]$$

sont compatibles avec ces lois marginales et en déduire que les lois marginales ne sont jamais suffisantes à caractériser la loi d'un couple (x, y).

3. – Soient x, y de densités conditionnelles $f_{X|Y}(x|y)$ et $f_{Y|X}(y|x)$. Montrer que, dans ce cas, la loi du couple est parfaitement déterminée. Appliquer au cas où $x|y \sim \mathcal{G}a(2, \theta + y)$ et $y|x \sim \mathcal{E}xp(x)$. Etendre au cas où $Z = (Z_1, \ldots, Z_n)$ et où les lois conditionnelles $f_1(z_1|z_2, \ldots, z_n)$, …, $f_1(z_n|z_1, \ldots, z_{n-1})$ sont connues.

4. – Soient x, y i.i.d. $\mathcal{N}(0,1)$. On définit (R, θ) comme les coordonnées polaires de (x, y).

(a) Montrer que $x+y$ et $x-y$ sont indépendants et en déduire la loi de R^2 sachant $x = y$.

(b) Montrer que R et θ sont indépendants et en déduire la loi de R^2 sachant $\theta = \pi/4$ ou $5\pi/4$.

(c) Etant donné que $x = y$ correspond à $\theta = \pi/4$ ou $5\pi/4$, expliquer ce paradoxe.

5. – Soit $x \sim \mathcal{E}xp(\theta)$. Donner la loi de x conditionné par $\min(x, a)$, où a est un réel positif.

6. – Calculer

$$\int_0^1 \left\{ \int_0^1 f(x,y)dy \right\} dx, \quad \int_0^1 \left\{ \int_0^1 f(x,y)dx \right\} dy, \quad \int_{[0,1]^2} f(x,y)dydx,$$

pour

$$f(x,y) = \begin{cases} (x-1/2)^{-3} & \text{si } 0 < y < |x-1/2| \\ 0 & \text{sinon} \end{cases}$$

$$f(x,y) = (1-xy)^{-p} \quad (p > 0).$$

Problème I. – Lois usuelles.

Pour $p > 0$, on définit

$$\Gamma(p) = \int_0^\infty t^{p-1}e^{-t}dt$$

1. En déduire que
$$f_p(t|\theta) = \theta^p t^{p-1} e^{-\theta t}/\Gamma(p)$$
définit une densité de probabilité sur $\mathbb{R}_+$. Cette loi est la loi *gamma*, $\mathcal{G}a(p,\theta)$.
   a) Montrer que $\mathcal{G}a(1,\theta)$ est la loi exponentielle, $\mathcal{E}xp(\theta)$.
   b) Montrer que, si $x \sim \mathcal{G}a(p,\theta)$, $x\theta \sim \mathcal{G}a(p,1)$.
   c) Déduire de $\Gamma(p+1) = p\Gamma(p)$ que $\Gamma(n+1) = n!$ pour $n \in \mathbb{N}$.
2. Soit $x \sim \mathcal{N}(0,1)$ une v.a. normale.
   a) Montrer que $x^2 \sim \mathcal{G}a(1/2, 1/2)$.
   b) Montrer que la fonction génératrice de $\mathcal{G}a(p,\theta)$ est $\varphi(t) = (\theta/\theta - t)^p$ pour $t < \theta$ et en déduire que, si $y_1 \sim \mathcal{G}a(p,\theta)$, $y_2 \sim \mathcal{G}a(q,\theta)$, indépendantes, $y_1 + y_2 \sim \mathcal{G}a(p+q,\theta)$.
   c) Soient $x_1, \ldots, x_n$ des v.a. i.i.d. $\mathcal{N}(0,1)$. Exprimer la loi de $z = x_1^2 + \ldots + x_n^2$, dite *loi du khi-deux*, ξ_n^2.
   d) Déduire de a) que $\mathcal{G}amma(1/2) = \sqrt{\pi}$.
3. Soient $y_1 \sim \mathcal{G}a(p,\theta)$ et $y_2 \sim \mathcal{G}a(q,\theta)$.
   a) Utiliser 2.b) pour trouver la loi de $y_1/(y_1+y_2)$, appelée *loi béta de première espèce*, $\mathcal{B}e(p,q)$. On montrera d'abord que cette loi est indépendante de θ. On note $B(p,q)$ le coefficient de normalisation de la densité.
   b) En déduire la loi de y_1/y_2, dite *loi béta de seconde espèce*, $\mathcal{B}e^2(p,q)$, et celle de $(y_1/p)/(y_2/q)$, dite *loi de Fisher*, $\mathcal{F}_{p,q}$. Comment relier cette loi à la loi normale?
   c) Si $z \sim \mathcal{B}e(p,q)$, exprimer $\mathbb{E}[z^r]$ ($r \in \mathbb{R}$) en fonction des $B(p',q')$.
4. Soit $t \sim \mathcal{B}e(p,1)$. Donner la loi de θ/t, dite *loi de Pareto*, $\mathcal{P}a(p,\theta)$, et montrer que c'est aussi la loi de $\theta(1+w)$ si $w \sim \mathcal{B}e^2(1,p)$.
5. Soient $x \sim \mathcal{N}(0,1)$ et $y \sim \chi_p^2$.
   a) Donner la loi de $z = x/\sqrt{y/p}$. dite *loi de Student*, $\mathcal{T}_p(0,1)$.
   b) En déduire la loi de $z\sigma + \theta$, $\mathcal{T}_p(\theta,\sigma)$.
   c) La loi $\mathcal{T}_1(\theta,\sigma)$ est dite *loi de Cauchy*, $\mathcal{C}(\theta,\sigma)$. Montrer que cette loi n'a pas de moment.

Problème II. – Statistiques d'ordre.

Soient $x_1, \ldots, x_n$ des v.a. i.i.d. de fonction de répartition $F(x)$, définies sur $(\Omega, \mathcal{A}, P)$.

1. Pour $x \in \mathbb{R}$, $r \in \{1, \ldots, n\}$, on définit $A_r(x)$ comme l'ensemble des $\omega \in \Omega$ tels que r au moins des n v.a. $x_1(\omega), \ldots, x_n(\omega)$ sont inférieures à x. Montrer que $A_r(x)$ est un événement de $\mathcal{A}$.
2. Montrer que
$$\begin{aligned} F_r(x) = P[A_r(x)] &= \sum_{i=r}^{n} \binom{n}{i} F(x)^i (1 - F(x))^{n-i} \\ &= \frac{1}{B(r, n-r+1)} \int_0^{F(x)} t^{r-1}(1-t)^{n-r} dr. \end{aligned}$$

3. Montrer que $F_r(x)$ est une fonction de répartition et interpréter la v.a. associée à F_r, notée $x_{(r)}$.
4. Si les x_i ont une densité f par rapport à la mesure de Lebesgue, montrer que la densité de $x_{(r)}$ est

$$f_r(x) = \frac{1}{B(r, n-r+1)} F(x)^{r-1} f(x) (1-F(x))^{n-r}.$$

5. On suppose que les x_i sont distribués suivant $\mathcal{U}_{[0,1]}$, loi uniforme sur $[0,1]$. Montrer que la loi de $x_{(r)}$ est la loi $\mathcal{B}e(r, n-r+1)$.
6. Sous l'hypothèse de 4., donner la loi du couple $(x_{(1)}, x_{(n)})$ et appliquer au cas où les x_i suivent une loi $\mathcal{U}_{[-\theta,\theta]}$.

Problème III. – Processus de Poisson.

On considère des v.a. positives x_i, $i \in \mathbb{N}^*$, indépendantes, définies sur $(\Omega, \mathcal{A}, P)$.

1. Montrer que, si les x_i satisfont la propriété de non-vieillissement

$$\forall t, u \qquad 0 < P(x_i \geq u) = P(x_i \geq u + t | x_i \geq t) < 1,$$

elles sont nécessairement exponentielles, $\mathcal{E}xp(\theta)$.

2. Montrer que $\min x_i = x_{(1)}$ et $x_1 + \ldots + x_n$ suivent une loi $\mathcal{E}xp(n\theta)$.
3. Si $S_n = x_1 + \ldots + x_n$ et $S_0 = 0$, on définit

$$N_t = \sum_{n \in \mathbb{N}} n \mathbb{I}_{\{S_n \leq t < S_{n+1}\}} + (+\infty) \mathbb{I}_{\{S_n \leq t \ \forall n\}} .$$

Montrer que $N_t \geq n$ équivaut à $S_n \leq t$ et en déduire que $P(N_t = +\infty) = 0$. Quelle est la loi de N_t?

4. Soient t arbitraire et

$$Z_t = \begin{cases} t - S_{N_t} & \text{si } t \geq S_1, \\ t & \text{sinon,} \end{cases} \qquad W_t = X_{N_t+1} - Z_t.$$

   a) Interpréter ces v.a. et donner la loi de (Z_t, W_t).
   b) En déduire que ces deux v.a. sont indépendantes, que $W_t \sim \mathcal{E}xp(\theta)$ et que Z_t suit “presque” une loi exponentielle $\mathcal{E}xp(\theta)$ (pour t grand).