Projet de sujet de Probabilités 2011
Concours d'entrée à l'ÉNS Cachan 3e année

Les différentes parties du problème ne sont pas indépendantes

Notations

On note $\mathcal{N}(0,1)$ la loi normale de moyenne 0 et de variance 1. Si X est une variable aléatoire, on note $\mathbb{E}(X)$ son espérance et $\mathrm{Var}(X)$ sa variance. On rappelle que X est une variable aléatoire de Poisson de paramètre λ lorsque

$$\mathbb{P}(X = n) = e^{-\lambda}\frac{\lambda^n}{n!} \quad \text{pour tout } n \in \mathbb{N}.$$

On dit que X suit la loi de Bernoulli de paramètre $p \in [0,1]$ lorsque

$$p = \mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0).$$

Toutes les variables aléatoires sont définies sur un espace de probabilité commun $(\Omega, \mathcal{A}, \mathbb{P})$. Si μ est une loi de probabilité sur $\mathbb{N}$, alors pour tout $k \in \mathbb{N}$, on note $\mu(k)$ la probabilité affectée par μ au singleton $\{k\}$. On note enfin $\mathbb{E}(X \,|\, Y)$ l'espérance conditionnelle de X sachant Y.

Si E est un ensemble fini, on note $|E|$ son cardinal. Pour tout $n \in \mathbb{N}^*$, on note $\mathcal{B}_n$ l'ensemble des partitions de $\{1,\ldots,n\}$. Pour tout $B \in \mathcal{B}_n$ et tout $b \subset \{1,\ldots,n\}$ non vide, on note $b \in B$ lorsque b fait partie de la partition B, et on dit que b est un *bloc* de B. La partition B est constituée de $|B|$ blocs. Pour tous $b, b' \in B$ avec $b \neq b'$, on a $b \cap b' = \varnothing$. On a également $|b| > 0$ pour tout $b \in B$ et

$$\{1,\ldots,n\} = \bigcup_{b \in B} b.$$

On note enfin B^{b+} l'élément de $\mathcal{B}_{n+1}$ obtenu à partir de B en remplaçant le bloc $b \in B$ par $b \cup \{n+1\}$. Notons que $|B^{b+}| = |B|$.

Un modèle de croissance

Le but de ce sujet est d'étudier l'évolution au cours du temps d'une population répartie en groupes. À l'instant $n \in \mathbb{N}^*$, la population comporte n individus, numérotés de 1 à n, et les groupes qu'ils constituent sont modélisés par une variable aléatoire B_n à valeurs dans $\mathcal{B}_n$. Chaque bloc correspond à un groupe.

On pose $B_1 = \{\{1\}\}$. À l'instant $n+1$, un nouvel individu rejoint la population et décide **soit** de rejoindre l'un des $|B_n|$ groupes déjà constitués avec une probabilité proportionnelle au nombre de membres du groupe (grégarisme), **soit** de créer un nouveau groupe (individualisme).

On modélise par $\theta/(n+\theta)$ la probabilité qu'a l'individu $n+1$ de créer un nouveau groupe, où $\theta > 0$ est un paramètre réel fixé et inconnu (identique pour tous). La population ne fait que croître, et comporte n individus au temps n. Le numéro d'un individu correspond exactement à son temps d'arrivée dans la population. Les individus ne meurent pas et ne quittent pas leur groupe. La suite $(B_{k+1})_{k\in\mathbb{N}}$ est une chaîne de Markov d'espace d'états $\cup_{n=1}^{\infty}\mathcal{B}_n$.

Rappels

On dit qu'un suite de variables aléatoires $(X_k)_{k\in\mathbb{N}}$ à valeurs dans un ensemble au plus dénombrable E est une *chaîne de Markov* lorsque pour tout $k \in \mathbb{N}$ et tous $x_0, \ldots, x_k, x_{k+1}$ dans E, on a

$$\mathbb{P}(X_{k+1} = x_{k+1} \,|\, X_0 = x_0, \ldots, X_k = x_k) = \mathbb{P}(X_{k+1} = x_{k+1} \,|\, X_k = x_k).$$

La loi de X_0 est appelée *loi initiale* et l'ensemble E *espace d'états*. On dit que la chaîne est *homogène* lorsque pour tout $k \in \mathbb{N}$ et tous x, y dans E, la quantité $\mathbb{P}(X_{k+1} = y \,|\, X_k = x)$ ne dépend pas de k.

Partie I

Cette partie est consacrée à une première étude de la chaîne de Markov $(B_{k+1})_{k\in\mathbb{N}}$.

1. Montrer que pour tout $n \in \mathbb{N}^*$ et tout $B \in \mathcal{B}_n$,

$$\sum_{b\in B} |b| = n$$

2. Montrer que pour tout $n \in \mathbb{N}^*$ et tout $B \in \mathcal{B}_n$ et $B' \in \mathcal{B}_{n+1}$,

$$\mathbb{P}(B_{n+1} = B' \,|\, B_n = B) = \begin{cases} \dfrac{|b|}{\theta+n} & \text{si } B' = B^{b+} \text{ pour un } b \in B \\[2ex] \dfrac{\theta}{\theta+n} & \text{si } B' = B \cup \{\{n+1\}\} \\[2ex] 0 & \text{sinon} \end{cases}$$

3. La chaîne de Markov $(B_{k+1})_{k\in\mathbb{N}}$ est-elle homogène ? Préciser sa loi initiale
4. Préciser l'évolution de la population dans le cas extrême où $\theta = 0$
5. Préciser l'évolution de la population dans le cas extrême où $\theta = +\infty$

Partie II

Cette partie est consacrée à la loi de la partition aléatoire B_n ainsi qu'à ses blocs.

1. Établir par récurrence que pour tout $n \geq 1$ et tout $B \in \mathcal{B}_n$,
$$\mathbb{P}(B_n = B) = \frac{\theta^{|B|}}{(\theta+0)(\theta+1)\cdots(\theta+n-1)} \prod_{b \in B} (|b|-1)!$$
2. Soit B une partition de $\{1, \ldots, n\}$ en blocs $b_1, \ldots, b_{|B|}$. Pour tout $1 \leq i \leq n$, soit a_i le nombre de blocs de taille i de B. Montrer que
$$a_1 + 2a_2 + \cdots + na_n = n$$
3. Avec les notations de la question précédente, établir que le nombre de partitions $B \in \mathcal{B}_n$ à nombres $a_1, \ldots, a_n$ prescrits vaut
$$\frac{n!}{\prod_{i=1}^n (i!)^{a_i} a_i!}$$
4. Pour tout $n \geq 1$ et tout $1 \leq i \leq n$, soit $C_{n,i}$ le nombre de blocs de taille i de B_n, de sorte que $n = C_{n,1} + 2C_{n,2} + \cdots + nC_{n,n}$. Déduire des questions précédentes que pour tous entiers $0 \leq a_1, \ldots, a_n \leq n$ tels que $a_1 + 2a_2 + \cdots + na_n = n$,
$$\mathbb{P}(C_{n,1} = a_1, \ldots, C_{n,n} = a_n) = \frac{n!}{(\theta+0)(\theta+1)\cdots(\theta+n-1)} \prod_{i=1}^n \frac{1}{a_i!} \left(\frac{\theta}{i}\right)^{a_i}$$
5. Pour tout $n \geq 1$, on pose $C_n = (C_{n,1}, \ldots, C_{n,n})$ où $C_{n,i}$ est comme précédemment. Montrer que la suite de vecteurs aléatoires $(C_{k+1})_{k \in \mathbb{N}}$ est une chaîne de Markov non homogène dont on précisera l'espace d'états.

Partie III

Cette partie est consacrée à l'étude du nombre de blocs $|B_n|$ de B_n.

1. Montrer que la suite $(|B_{k+1}|)_{k \in \mathbb{N}}$ est une chaîne de Markov non homogène
2. Montrer que pour tout $n \geq 1$, $|B_n| = \xi_1 + \cdots + \xi_n$ où $\xi_1, \ldots, \xi_n$ sont des variables aléatoires indépendantes de lois de Bernoulli vérifiant pour tout $1 \leq k \leq n$
$$\mathbb{P}(\xi_k = 1) = 1 - \mathbb{P}(\xi_k = 0) = \frac{\theta}{\theta+k-1}$$
3. Montrer que pour tout $n \geq 1$, on a
$$\mathbb{E}(|B_n|) = \sum_{k=0}^{n-1} \frac{\theta}{\theta+k} \quad \text{et} \quad \mathrm{Var}(|B_n|) = \sum_{k=1}^{n-1} \frac{\theta k}{(\theta+k)^2}$$
La suite $(|B_{k+1}|)_{k \in \mathbb{N}}$ constitue-t-elle une martingale ?
4. Montrer que $\mathbb{E}(|B_n|) \sim \theta \ln(n)$ et $\mathrm{Var}(|B_n|) \sim \theta \ln(n)$ quand $n \to +\infty$
5. Établir que
$$\frac{|B_n|}{\ln(n)} \xrightarrow[n\to\infty]{\mathbb{P}} \theta \qquad (\star)$$
où $\xrightarrow{\mathbb{P}}$ désigne la convergence en probabilité.

Partie IV

On étudie dans cette partie le nombre de blocs de taille n de B_n, noté $C_{n,n}$.

1. Montrer que $C_{n,n}$ vaut 0 ou 1 et que $C_{n,n} = 1$ si et seulement si $|B_n| = 1$
2. Établir que
$$\mathbb{P}(C_{n,n} = 1) = \prod_{k=1}^{n-1} \frac{k}{\theta + k}$$
3. En déduire que $\lim_{n\to\infty} \mathbb{P}(C_{n,n} = 1) = 0$
4. En déduire que presque-sûrement, $C_{n,n} = 0$ à partir d'un certain rang sur n.

Partie V

Dans cette partie, on étudie le nombre de blocs de taille 1 de B_n, noté $C_{n,1}$.

1. Montrer que $(C_{k+1,1})_{k\in\mathbb{N}}$ est une chaîne de Markov d'espace d'état $\mathbb{N}$ et que pour tout $n \geq 1$, tout $0 \leq a \leq n$ et tout $0 \leq a' \leq n+1$, on a
$$\mathbb{P}(C_{n+1,1} = a' \mid C_{n,1} = a) = \begin{cases} \dfrac{\theta}{\theta+n} & \text{si } a' = a+1 \\[2ex] \dfrac{a}{\theta+n} & \text{si } a' = a-1 \\[2ex] 1 - \dfrac{\theta}{\theta+n} - \dfrac{a}{\theta+n} & \text{si } a' = a \\[2ex] 0 & \text{sinon} \end{cases}$$
2. En déduire que pour tout $n \geq 1$,
$$\mathbb{E}(C_{n+1,1} \mid C_{n,1} = a) = \frac{a(\theta + n - 1) + \theta}{\theta + n}$$
3. En déduire que pour tout $n \geq 1$,
$$\mathbb{E}(C_{n,1}) = \frac{n\theta}{n + \theta - 1}$$
4. Si X et Y sont des variables aléatoires à valeurs dans $\mathbb{N}$ avec X de carré intégrable, on définit $\mathrm{Var}(X \mid Y) = \mathbb{E}(X^2|Y) - \mathbb{E}(X|Y)^2$. Montrer que $\mathrm{Var}(X \mid Y) \geq 0$ et que
$$\mathrm{Var}(X) = \mathbb{E}(\mathrm{Var}(X \mid Y)) + \mathrm{Var}(\mathbb{E}(X \mid Y))$$
5. En déduire que pour tout $n \geq 1$,
$$\mathrm{Var}(C_{n,1}) = \frac{n(n-1)(n-2+2\theta)\theta}{(n+\theta-2)(n+\theta-1)^2}$$

Partie VI

Cette partie est consacrée à l'étude d'une distance sur l'ensemble des lois de probabilités sur $\mathbb{N}$. Si μ et ν sont des lois de probabilité sur $\mathbb{N}$, on définit

$$d_V(\mu,\nu) = \sup_{A \subset \mathbb{N}} |\mu(A) - \nu(A)|.$$

Si X et Y sont des variables aléatoires sur $\mathbb{N}$ de lois respectives μ et ν, on pose par commodité $d_V(X,Y) = d_V(\mu,\nu)$.

1. Montrer que d_V est une distance sur l'ensemble des lois de probabilité sur $\mathbb{N}$
2. Établir que si μ et ν sont des lois de probabilité sur $\mathbb{N}$ alors

$$2d_V(\mu,\nu) = \sum_{k\in\mathbb{N}} |\mu(k) - \nu(k)|$$

3. Établir que si μ et ν sont des lois de probabilité sur $\mathbb{N}$ alors

$$2d_V(\mu,\nu) = \sup_{f\in\mathcal{F}} \left| \int f\, d\mu - \int f\, d\nu \right|$$

   où $\mathcal{F}$ désigne l'ensemble des fonctions définies sur $\mathbb{N}$ et à valeurs dans $[-1,1]$

4. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires sur $\mathbb{N}$ et μ une loi de probabilité sur $\mathbb{N}$. Pour tout $n \in \mathbb{N}$, on note μ_n la loi de X_n. Montrer que les propriétés suivantes sont équivalentes :
   (a) X_n converge en loi vers μ
   (b) $\lim_{n\to\infty} \mu_n(k) = \mu(k)$ pour tout $k \in \mathbb{N}$
   (c) $\lim_{n\to\infty} d_V(\mu_n,\mu) = 0$
5. Montrer que si ξ est une variable aléatoire de Bernoulli de paramètre $p \in [0,1]$ et P une variable aléatoire de Poisson de même moyenne alors

$$d_V(\xi, P) \leq p^2$$

6. Montrer que si $(X_k)_{1\leq k\leq n}$ et $(Y_k)_{1\leq k\leq n}$ sont des suites de variables aléatoires indépendantes à valeurs dans $\mathbb{N}$ alors

$$d_V(X_1 + \cdots + X_n, Y_1 + \cdots + Y_n) \leq \sum_{k=1}^{n} d_V(X_k, Y_k)$$

7. En déduire que si $\xi_1,\ldots,\xi_n$ sont des variables aléatoires indépendantes de loi de Bernoulli de paramètres respectifs $p_1,\ldots,p_n$ et si P_n est une variable aléatoire de Poisson de moyenne $p_1 + \cdots + p_n$ alors

$$d_V(\xi_1 + \cdots + \xi_n, P_n) \leq \sum_{k=1}^{n} p_k^2 \qquad (\star\star)$$

Partie VII

Cette partie est consacrée à l'étude des fluctuations asymptotiques pour la convergence en probabilité de $|B_n|/\ln(n)$ vers θ obtenue en $(\star)$. Elle met en œuvre la distance d_V étudiée précédemment. On admettra qu'on peut remplacer, dans le membre de droite de $(\star\star)$, la quantité $\sum_{k=1}^{n} p_k^2$ par la quantité plus petite

$$(1-\exp(-p_1-\cdots-p_n))\,\frac{p_1^2+\cdots+p_n^2}{p_1+\cdots+p_n}.$$

D'autre part, afin d'alléger les notations, on pose

$$\lambda_n = \mathbb{E}(|B_n|) \quad \text{et} \quad \sigma_n^2 = \mathrm{Var}(|B_n|)$$

et on note P_n une variable aléatoire de Poisson de paramètre λ_n.

1. Établir que
$$\lim_{n\to\infty} d_V(|B_n|, P_n) = 0$$
2. Établir que pour toute fonction $f : \mathbb{R} \to \mathbb{R}$ continue et bornée,
$$\left|\mathbb{E}\left(f\left(\frac{|B_n|-\lambda_n}{\sigma_n}\right)\right) - \mathbb{E}\left(f\left(\frac{P_n-\lambda_n}{\sigma_n}\right)\right)\right| \leq 2\,(\sup|f|)\ d_V(|B_n|, P_n)$$
3. En utilisant $\lim_{n\to\infty} \lambda_n/\sigma_n^2 = 1$, établir que
$$\frac{P_n-\lambda_n}{\sigma_n} \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0,1)$$
4. Établir que
$$\frac{|B_n|-\lambda_n}{\sigma_n} \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0,1)$$
5. Établir que si $(X_n)_{n\in\mathbb{N}}$ est une suite de variables aléatoires convergeant en loi vers une loi de probabilité μ et si $(a_n)_{n\in\mathbb{N}}$ et $(b_n)_{n\in\mathbb{N}}$ sont deux suites de réels convergeant vers a et b respectivement alors $(a_nX_n+b_n)_{n\in\mathbb{N}}$ converge en loi vers la loi de $aX+b$ où X suit la loi μ
6. En déduire que
$$\frac{|B_n|-\theta\ln(n)}{\sqrt{\theta\ln(n)}} \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0,1)$$
7. Soit $\alpha\in]0,1[$. Déduire de la question précédente un intervalle de confiance asymptotique pour θ au niveau de confiance $1-\alpha$, en fonction d'un quantile de la loi $\mathcal{N}(0,1)$.

– Fin de l'épreuve –