IUFM de Midi-Pyrénées
Préparation au CAPES de Maths
2000-2001

# Problème de probabilités du 22 décembre 2000

Le but de ce problème est l'étude d'une file d'attente simple $X$. Cette file d'attente est construite de la façon suivante. On considère une échelle discrète de temps que l'on indexe par les entiers naturels. A l'instant 0, il y a un nombre aléatoire $X_0$ de personnes présentes dans la file. Pour $n \in \mathbb{N}$, entre les instants $n$ et $n+1$, il y a une probabilité $p \in ]0,1[$ qu'un individu entre dans la file et $q = 1-p$ qu'un individu la quitte. On suppose que les départs et les arrivées des individus sont tous indépendants et sont aussi indépendants du nombre d'individus dans la file à l'instant initial. Pour $n \geq 1$, soit $X_n$ le nombre d'individus dans la file à l'instant $n$. On supposera dans tout le problème que $X_0$ est de carré intégrable. On pose $m = \mathbb{E}(X_0)$ et $\sigma^2 = \text{var} X_0$.
La partie I du problème est consacrée à des préliminaires techniques sur les fonctions génératrices et la convergence en loi. Des résultats de la partie I devront être utilisés dans les parties II et III. Dans la partie II, on étudie le comportement en régime stationnaire de la file d'attente $(X_n)_{n\in\mathbb{N}}$. La dernière partie porte sur le problème d'estimation du paramètre $p$. Chaque partie est notée sur 20. Le problème est donc noté sur 60.

## I) Quelques propriétés des fonctions génératrices

Soit $Z$ une variable aléatoire à valeurs dans $\mathbb{N}$. On définit, pour $s \in ]-1,1[$ sa fonction génératrice

$$G_Z(s) = \mathbb{E}(s^Z) = \sum_{i=0}^{+\infty} \mathbb{P}(Z=i)s^i \tag{1}$$

a) Justifier l'existence de la série entière définissant la fonction génératrice dans (1). Pour un entier naturel $k$ donné, comment est-il possible de calculer $\mathbb{P}(Z=k)$ à partir de la fonction génératrice de $Z$?

b) On suppose que le rayon de convergence $R$ de la série entière définie dans (1) vérifie $R > 1$. Montrer qu'alors $Z$ possède des moments de tout ordre et que pour $l \in \mathbb{N}$,

$$\mathbb{E}\left[Z(Z-1)\cdots(Z-l+1)\right] = G_Z^{(l)}(1). \tag{2}$$

(L'espérance apparaissant dans la dernière équation s'appelle le moment factoriel d'ordre $l$ de $Z$).

c) Soient $Z_1$ et $Z_2$ des variables aléatoires indépendantes et toutes deux à valeurs dans $\mathbb{N}$. Donner l'expression de la fonction génératrice de la variable aléatoire $Z_1 + Z_2$ en fonction de celles de $Z_1$ et $Z_2$.

d) Soit $(Z_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes et de même loi géométrique de paramètre $\theta \in ]0,1[$. C'est-à-dire que l'on a pour $k \in \mathbb{N}$

$$\mathbb{P}(Z_n = k) = \theta^k(1-\theta) \quad (n \in \mathbb{N}). \tag{3}$$

Calculer la fonction génératrice de $Z_0$. Que valent les moments factoriels de $Z_0$. calculer pour, $n \geq 0$, $\mathbb{P}(Z_0 + \cdots + Z_n = 2)$.

e) On rappelle que l'on dit qu'une suite $(U_n)_{n\in\mathbb{N}}$ de variable aléatoires à valeurs dans $\mathbb{N}$ converge en loi vers $U$ si, et seulement si, pour tout $k \in \mathbb{N}$,

$$\lim_{n\to\infty} \mathbb{P}(U_n = k) = \mathbb{P}(U = k).$$

Soit $(U_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires à valeurs dans $\mathbb{N}$. On suppose qu'il existe un réel $\varepsilon > 0$ tel que pour tout $s \in ]-\varepsilon, +\varepsilon[$ $G_{U_n}(s)$ converge, quand $n$ tend vers l'infini, vers $G_U(s)$ où $U$ est une variable aléatoire à valeurs dans $\mathbb{N}$.

i) Montrer que $\lim_{n\to\infty} \mathbb{P}(U_n = 0) = \mathbb{P}(U = 0)$.

ii) On pose pour, $n \in \mathbb{N}$ et $s \in ]-\varepsilon, +\varepsilon[$, $H_{n,0}(s) = G_{U_n}(s)$.
Puis, pour $l \geq 1$

$$H_{n,l}(s) = \frac{H_{n,l-1}(s) - \mathbb{P}(U_n = l-1)}{s} \qquad (s \in ]-\varepsilon, +\varepsilon[\setminus\{0\}).$$

De même, on pose pour $s \in ]-\varepsilon, +\varepsilon[$, $H_0(s) = G_U(s)$ et pour $l \geq 1$

$$H_l(s) = \frac{H_{l-1}(s) - \mathbb{P}(U = l-1)}{s} \qquad (s \in ]-\varepsilon, +\varepsilon[\setminus\{0\}).$$

Montrer par récurrence sur l'entier $l \in \mathbb{N}$ que pour tout $n \in \mathbb{N}$ et $s \in ]-\varepsilon, +\varepsilon[\setminus\{0\}$, on a

$$H_{n,l}(s) = \sum_{j\geq l} \mathbb{P}(U_n = j)s^{j-l} \tag{4}$$

Puis que les fonctions $H_{n,l}$ et $H_l$ sont prolongeables par continuité en 0. On appelera encore $H_{n,l}$ et $H_l$ les fonctions prolongées. Pour $s \in ]-\varepsilon, \varepsilon[$ fixé montrer par récurrence sur $l \geq 0$ que

$$\lim_{n\to\infty} H_{n,l}(s) = H_l(s).$$

iii) Déduire de la question ii) que la suite $(U_n)$ converge en loi vers $U$.

f) Soit $(Z_j^{(n)})_{1\leq j\leq n, n\in\mathbb{N}^*}$ un tableau triangulaire de variables aléatoires toutes indépendantes. On suppose que pour $n \geq 1$ et $j = 1, \cdots, n$, $Z_j^{(n)}$ suit la loi géométrique de paramètre $\frac{\theta}{n}$. En utilisant les questions d) et e), montrer que $Z_1^{(n)} + \cdots + Z_n^{(n)}$ converge en loi vers une loi à préciser.

## II) Etude du régime stationnaire de la file

a) Quelques géneralités

i) On rappelle que pour $x \in \mathbb{R}$ on a $x^+ = \sup(x, 0)$. Montrer que la suite $(X_n)$ satisfait l'équation de récurrence suivante :

$$X_{n+1} = (X_n + \varepsilon_n)^+, \ \ (n \in \mathbb{N}). \tag{5}$$

Où $(\varepsilon)_{n\in\mathbb{N}}$ est une suite de variables aléatoires indépendantes, identiquement distribuées (i.i.d.), indépendante de $X_0$ et à valeurs dans $\{-1, 1\}$. Préciser la loi de $\varepsilon_1$.

ii) Soit $S_{-1} = 0$ et pour $n \geq 0$, $S_n = \sum_{j=0}^{n} \varepsilon_j + X_0$. Montrer l'égalité :

$$X_{n+1} = S_n - \inf_{-1 \leq j \leq n} S_j \quad (n \geq 0). \tag{6}$$

b) On suppose que $p > \frac{1}{2}$.

0) **Par la suite on utilisera la définition suivante :**
Soit $(U_n)$ une suite de variables aléatoires. On dit que $(U_n)$ tend en probabilité vers l'infini (ou diverge en probabilité vers $+\infty$), quand $n$ tend vers l'infini, si pour tout réel $A > 0$, on a

$$\lim_{n \to \infty} \mathbb{P}(U_n \leq A) = 0.$$

i) Calculer l'espérance et la variance de $S_n$. Quelles sont les limites en probabilité des suites $(n^{-1} S_n)$ et $(S_n)$ (quand $n$ tend vers l'infini)?

ii) Montrer l'inégalité pour $n \geq 1$

$$X_{n+1} \geq S_n. \tag{7}$$

En déduire que la suite $(X_n)$ diverge en probabilité vers $+\infty$. Interpréter ce résultat.

c) On suppose que $p < \frac{1}{2}$.

i) Montrer que la loi de la suite $(X_n)$ satisfait la récurrence suivante :

$$\begin{aligned} \mathbb{P}(X_{n+1} = x) &= p\mathbb{P}(X_n = x - 1) + q\mathbb{P}(X_n = x + 1) \quad (x \in \mathbb{N}^*, n \in \mathbb{N}), \\ \mathbb{P}(X_{n+1} = 0) &= q[\mathbb{P}(X_n = 0) + \mathbb{P}(X_n = 1)] \quad (n \in \mathbb{N}). \end{aligned}$$

ii) Déduire de la question précédente l'équation de récurrence suivante valide pour tout $s \in ]-1, 1[\backslash\{0\}$ :

$$G_{X_{n+1}}(s) = qG_{X_n}(0)(1 - \frac{1}{s}) + (ps + \frac{q}{s})G_{X_n}(s). \tag{8}$$

Etablir une équation de récurrence satisfaite par la suite $(G_{X_n}(0))$ et montrer que cette équation s'obtient aussi en faisant tendre $s$ vers 0 dans (8).

iii) On suppose que $X_0$ suit la loi géométrique de paramètre $p/q$. Montrer qu'alors, il en est de même pour $X_n$ $(n \in \mathbb{N})$.

iv) On ne suppose plus que $X_0$ suit la loi géométrique de paramètre $p/q$, mais on fait l'hypothèse qu'il existe $\varepsilon > 1$, tel que, pour $s \in ]-\varepsilon, \varepsilon[$, $\lim_{n\to\infty} G_{X_n}(s)$ existe. Montrer en utilisant I) que la suite $(X_n)$ converge en loi. Quelle est la loi limite?

## III) Estimation de $p$

On se replace dans le cadre de la question II.c.iii) et l'on considère $N$ files d'attente différentes et indépendantes ayant toutes la même dynamique que $X$. On note $X_j^k$ $j \geq 1$ et $k = 1, \ldots N$ le nombre d'individus dans la $k$-ième file à l'instant $j$. Dans tout le problème on fixe un instant $j > 0$. En d'autres termes, à $j$ fixé, $(X_j^k)_{k=1,\dots N}$ est une suite i.i.d. de loi géométrique de paramètre $p/q$.

Soit $(W_n)$ une suite de variables aléatoires réelles et $W$ une variable aléatoire réelle qui possède une densité. On rappelle que l'on dit que la suite $(W_n)$ converge en loi vers $W$ si la suite des fonctions de répartition des variables $W_n$ converge ponctuellement vers la fonction de répartition de $W$.

a) On pose $\eta_N = \frac{1}{N}\sum_{k=1}^N X_j^k$. Calculer $\mathbb{E}(\eta_N)$ et var $\eta_N$. Quelle est la limite en probabilité de la suite $\eta_N$? Dans la suite, on notera $f(p)$ cette limite. Montrer que $\sqrt{N}(\eta_N - f(p))$ converge en loi vers une loi gaussienne centrée et de variance $\Delta^2(p)$ à déterminer.

b) Etudier les variations de la fonction $f(p)$ sur $]0, 1/2[$. Montrer qu'il existe une fonction $g$ continue dérivable et strictement croissante sur $\mathbb{R}_*^+$ qui vérifie pour tout $\tau > 0$ et $\theta \in ]0, 1/2[$, $f(g(\tau)) = \tau$ et $g(f(\theta)) = \theta$ .

c) Pour estimer $p$ à partir des observations $X_j^1, \dots, X_j^N$ on propose d'utiliser la variable aléatoire $\widehat{p}_N = g(\eta_N)$.

   0) **Résultat préliminaire**
   *On admettra le résultat suivant :* Soit $(W_n)$ une suite de variables aléatoires réelles. On suppose que $(W_n)$ converge en loi, quand $n$ tend vers l'infini, vers une variable aléatoire $W$. Soient $(\chi_n)$ une suite de variables aléatoires qui converge en probabilité vers 0 et $(\alpha_n)$ une suite de variables aléatoires qui converge en probabilité vers 1. Alors, $(W_n + \chi_n)$ et $(\alpha_n W_n)$ convergent aussi en loi vers $W$.

   i) Commenter le procédé de construction de $\widehat{p}_N$.

   ii) Montrer que $\widehat{p}_N$ converge en probabilité, quand $N$ tend vers l'infini, vers $p$.

   iii) Soit $\tau > 0$. Montrer qu'il existe un nombre réel $\tau^* > 0$ qui vérifie l'égalité :

$$g(\tau) - p = (1-2p)^2(\tau - f(p)) - \frac{2(\tau - f(p))^2}{(1+2\tau^*)^3}. \tag{9}$$

   iv) Quelle est la limite en loi de la suite $\sqrt{N}(1-2p)^2(\eta_N - f(p))$?

d) En utilisant les résultats obtenus dans les questions précédentes, montrer à l'aide de l'inégalité de Tchebycheff, que $\sqrt{N}(\eta_N - f(p))^2$ converge en probabilité vers 0. En déduire, en utilisant la formule (9) et le point III.c.o), que $\sqrt{N}(\widehat{p}_N - p)$ converge en loi vers une loi normale centrée de variance à préciser.

e) En utilisant II.c.0) et III.d), Construire une suite d'intervalles aléatoires $[a_N, b_N]$ qui ne dépend pas de $p$ et qui satisfait :

$$\lim_{N\to\infty} \mathbb{P}(p \in [a_N, b_N]) = 0,95. \tag{10}$$