J. PELLAUMAIL

Loi de répartition d'informations arrivant aléatoirement sur un disque

Publications des séminaires de mathématiques et informatique de Rennes, 1975, fascicule 1
« Séminaires de Rennes », , p. 29-35
<http://www.numdam.org/item?id=PSMIR_1975___1_A8_0>

LOI DE REPARTITION D'INFORMATIONS ARRIVANT ALEATOIREMENT SUR UN DISQUE

par J. PELLAUMAIL

A - INTRODUCTION

On considère un disque, par exemple une mémoire périphérique d'un ordinateur. Pour simplifier, on suppose que les informations qui arrivent sur ce disque sont indépendantes : si une information trouve une case occupée elle passe aux cases suivantes jusqu'à ce qu'elle se trouve une place. Le problème est d'étudier comment se répartissent ces informations.

Le plan adopté est le suivant :

B - Position du problème
C - Un problème voisin
D - Mise en équation
E - Résolution du système
F - Etude de la solution
G - Loi du nombre de cases à parcourir à l'instant t
H - Retour au problème initial

B - POSITION DU PROBLEME

Un disque comprend m cases disposées sur un cercle (m étant très grand). Chacune de ces cases est destinée à être remplie par une information. Ces informations arrivent, les unes après les autres, aléatoirement sur le disque considéré : plus précisément, les informations successives sont indépendantes les unes des autres et chaque information a la même probabilité d'arriver dans n'importe quelle case.

Si une information arrive dans une case déjà occupée, elle passe à la case suivante, et ainsi de suite, jusqu'à ce qu'elle trouve une case vide qu'elle occupe alors.

Soit x un réel tel que $0 < x < 1$ et tel que xm soit un entier ; le problème est d'étudier l'état du disque quand il a reçu xm informations ; notamment, quel est le nombre de cases qu'une information a du franchir avant de se trouver une place ?

C - UN PROBLEME VOISIN

Pour faciliter l'étude, on va considérer un problème un peu différent mais qui donnera à peu près les résultats cherchés pour n assez grand ; ce problème voisin est d'ailleurs intéressant en lui-même.

On considère une infinité de cases numérotées de 1 à l'infini. A partir de l'instant $t = 0$, des informations arrivent successivement sur les cases numérotées de 1 à n ; ces diverses informations sont indépendantes les unes des autres, le délai entre deux informations suit une loi exponentielle de paramètre b et, si une information arrive dans une case déjà occupée, elle passe aux cases suviantes jusqu'à ce qu'elle se trouve une place.

On se propose d'étudier l'état de ce système à l'instant t.

D - MISE EN EQUATION

On pose :

$p_o(t)$ = probabilité pour que la première case soit vide à l'instant t.

et pour $k \geq 1$:

$p_k(t)$ = probabilité pour que les k premières cases soient occupées et que la $(k+1)$ième soit vide à l'instant t.

On a $p_o(o) = 1$ et, pour $k \geq 1$, $p_k(o) = 0$.

De plus : $\quad p_o(t+\Delta t) = p_o(t)\ [1-b\ \Delta t] + \sigma(\Delta t)$

(où (Δt) désigne un infiniment petit par rapport à Δt)

ce qui donne $p'_o(t) = -b.p_o(t)$

Enfin, pour $1 \leq k \leq (n-1)$

$$p_k(t+\Delta t) = p_k(t).\left[1-(k+1).b.\Delta t\right] + \sum_{j=0}^{k-1} p_j(t).p_{k-j-1}(t).(j+1).b.\Delta t + \sigma(\Delta t)$$

en effet, la probabilité pour qu'il y ait, à l'instant t, les k premières cases occupées, sauf la (j+1)ième, et la (k+1)ième vide, est égale à $p_j(t).\, p_{k-j-1}(t)$ (pour k petit par rapport à n).

ceci donne :

$$p'_k(t) = -(k+1).b.p_k(t) + \sum_{j=0}^{k-1} b.(j+1).p_j(t).p_{k-j-1}(t)$$

E - RESOLUTION DU SYSTEME

Soit $(d_k)_{k \geq 0}$ la suite de réels définie par récurrence par :

$$d_o = d_1 = 1$$

$$d_k = \frac{1}{k} \sum_{j=0}^{k-1} (j+1).d_j.d_{k-j-1}$$

On vérifie immédiatement que la solution (unique) du système précédent est donnée, pour $k \leq (n-1)$, par :

$$p_k(t) = d_k.(bt)^k\, e^{-(k+1)bt} = e^{-bt}.d_k.(bt\, e^{-bt})^k$$

F - ETUDE DE LA SOLUTION

Soit h(t) la fonction définie sur $\mathbb{R}^+$ par $h(t) = bt\, e^{-bt}$; $h(\mathbb{R}^+) = [0, \frac{1}{e}]$. Puisque la famille $\{p_k\}$ définie une loi de probabilité, on doit avoir

$$\sum_{j=0}^{k} p_j(t) \leq 1 \qquad \text{(quel que soit t)}$$

donc $\displaystyle\sum_{j=0}^{k} d_j\, u^j \leq e^{bt}$ pour $0 \leq u \leq \frac{1}{e}$

ceci étant vrai pour tout k, la série entière de terme général $d_k.u^k$ doit avoir un rayon de convergence au moins égal à 1/e.

On peut alors considérer la fonction f, analytique dans le domaine complexe $|u| < 1/e$, définie par $f(u) = \sum_{k=0}^{\infty} d_k.u^k$.

En dérivant terme à terme, on vérifie que, au voisinage du point u = 0, la fonction f satisfait à l'équation différentielle suivante :

$$f'(u) = f(u).\left[f(u) + u.f'(u)\right] .$$

On en déduit que, sur l'intervalle $[0, 1/e]$, la fonction f est identique à la fonction réciproque de la fonction g définie sur $[1, e]$ par $g(x) = \frac{\text{Log } x}{x}$

On a donc, <u>pour $0 \leq t \leq 1/b$</u> , $1 = e^{-bt}.\, f(bt\, e^{-bt})$, soit $\sum_{k=0}^{\infty} p_k(t) = 1$ (ce qu'on ne pouvait pas affirmer a priori).

Par contre, pour $t > \frac{1}{b}$, $\sum_{k=0}^{\infty} p_k(t) < 1$.

Notons que $f'(\frac{1}{e}) = \infty$; on en déduit que le rayon de convergence de la série entière de terme général $d_k\, u^k$ est exactement $\frac{1}{e}$.

La fonction génératrice $\sum_{k=0}^{\infty} p_k(t).z^k = G(z,t)$ associée à la famille $\{p_k(t)\}_{k \geq 0}$, loi de probabilité pour $0 \leq t \leq \frac{1}{b}$, vaut $e^{-bt}.\, f\left(tb\, e^{-bt} z\right)$.

La moyenne de cette loi de probabilité vaut donc $bt\, e^{-2bt}\, f\left(bt\, e^{-bt}\right)$ soit, compte-tenu de ce qui précède, $\frac{bt}{1-bt}$; on pouvait, a priori, espérer ce résultat sans pour autant l'affirmer.

G - LOI DU NOMBRE DE CASES A PARCOURIR A L'INSTANT t

On considère le disque à l'instant t avec $0 \leq t \leq 1/b$: une information arrive à cet instant : combien de cases va-t-elle parcourir avant de trouver une case vide ?

Soit $q_k(t)$ la probabilité pour que l'information qui arrive à l'instant t ait exactement k cases à parcourir avant de trouver une case vide ; $q_o(t)$ est donc la probabilité pour que l'information arrive dans une case vide.

Pour $k \geq 1$, soit $r_k(t)$ la probabilité pour qu'une information, qui arrive à l'instant t, arrive dans un ensemble de k cases consécutives occupées (k cases exactement en comptant la case où arrive l'information et les cases avant ou après).

On a : $r_k(t) = \left[1 - q_o(t)\right] \; k.p_k(t) \; (\frac{1-bt}{bt})$

puisque $\frac{k\,p_k(t)}{\sum_{j=0}^{\infty} j.p_j(t)} = \frac{1-bt}{bt} \cdot k.p_k(t)$ est le pourcentage d'ensembles de k cases consécutives occupées parmi ces ensembles consécutifs.

Or, $q_o(t)$ est le pourcentage de cases vides à l'instant t, soit $q_o(t) = (1-bt)$

Enfin, on a, pour $k \geq 1$:

$$q_k(t) = \sum_{j=k}^{\infty} \frac{1}{j} r_j(t) = (1-bt) \sum_{j=k}^{\infty} p_j(t)$$

puisque, pour que l'information ait k cases à parcourir, il faut qu'elle arrive dans un ensemble comprenant plus de k cases consécutives occupées et qu'elle arrive dans la "bonne" case de cet ensemble.

Par ailleurs, soit $s_k(t)$ la probabilité pour qu'une information, qui est arrivée avant à l'instant t, ait parcouru k cases exactement avant de se trouver une place. On a :

$$s_k(t+\Delta t) = \frac{t}{t+\Delta t} \cdot s_k(t) + \frac{\Delta t}{t+\Delta t} \cdot q_k(t) + \sigma(\Delta t)$$

soit $s'_k(t) = \frac{1}{t} \cdot \left[q_k(t) - s_k(t)\right]$ et $s_o(o) = 1$ et $s_k(o) = 0$ si $k \geq 1$

ce qui donne : $s_k(t) = \frac{1}{t} \int_o^t q_k(u)\,du = \frac{1}{bt} \sum_{j=k}^{\infty} \frac{1}{j+1} \cdot d_j \cdot \left(bt\, e^{-bt}\right)^{j+1}$

H - RETOUR AU PROBLEME INITIAL

Dans le cadre du problème initial exposé au paragraphe B, on peut écrire les équations "discrétisées" associées aux équations données au paragraphe D, c'est-à-dire que, dans ces équations, on remplace $(b\,\Delta t)$ par $\frac{1}{m}$ et (bt) par le pourcentage x de cases occupées.

Avec une erreur de l'ordre de $\frac{1}{m}$ les résultats obtenus pour le deuxième problème étudié sont donc valables pour le problème initial exposé au paragraphe B en remplaçant bt par le pourcentage x de cases occupées.

On considère donc le problème initial exposé au paragraphe B. On suppose que le nombre m de cases du disque étudié est "grand". Soit x le pourcentage de cases occupées. On considère :

a) $\hat{p}_k(x)$ *la probabilité pour qu'il y ait exactement k cases consécutives occupées à la suite d'une case vide sachant que cette case est vide*

b) $\hat{q}_k(x)$ *la probabilité pour qu'une information, qui arrive alors qu'il y a x % de cases occupées, ait k cases exactement à parcourir pour se trouver une place*

c) $\hat{r}_k(x)$ *la probabilité pour qu'une information, qui arrive alors qu'il y a x % de cases occupées, arrive dans un ensemble de k cases consécutives occupées*

d) $\hat{s}_k(x)$ *la probabilité que l'une quelconque des informations en place ait du parcourir k cases exactement avant de se trouver une place si on considère le disque alors qu'il y a x % de cases occupées.*

On a alors : $\hat{p}_k(x) = e^{-x} \cdot d_k \cdot (x\, e^{-x})^k$

si $(d_k)_{k>0}$ *est la suite de nombres définie par récurrence par*

$$d_o = 1 \text{ et } d_k = \frac{1}{k} \sum_{j=0}^{k-1} (j+1).d_j.d_{k-j-1}$$

$$\hat{q}_k(x) = (1-x) . \sum_{j=k}^{\infty} \hat{p}_j(x) \quad \text{pour } k \geq 0 \text{ donc } \hat{q}_o(x) = 1-x$$

$$\hat{r}_k(x) = k . (1-x) . \hat{p}_k(x)$$

$$\hat{s}_k(x) = \frac{1}{x} . \int_o^x \hat{q}_k(u)\, du = \frac{1}{x} . \sum_{j=k}^{\infty} \frac{1}{j+1} . d_j . (x\, e^{-x})^{j+1}$$

Avant de terminer, signalons qu'il est facile d'écrire des équations généralisant celles qui précèdent dans le cas où les informations arrivent par paquets consécutifs, la taille de ces paquets suivant une loi connue. Par contre, dans ce cas, la résolution de ces équations semble plus difficile.

- : -