M. GIRAULT

Loi de probabilité du plus grand intervalle dans une partition « au hasard »

Revue de statistique appliquée, tome 10, nº 2 (1962), p. 63-65
<http://www.numdam.org/item?id=RSA_1962__10_2_63_0>

LOI DE PROBABILITÉ DU PLUS GRAND INTERVALLE DANS UNE PARTITION "AU HASARD"

M. GIRAULT

Professeur à l'Institut de Statistique

1 - ENONCE DU PROBLEME -

Sur le segment [0, 1] on considère n points $\{A_i\}$ d'abscisses $\{X_i\}_{i=1,2\ldots n}$ où les X_i sont des variables aléatoires indépendantes, chacune d'elles obéissant à la loi uniforme sur [0, 1]. Ces n points définissent une partition du segment [0-1] en n + 1 intervalles aléatoires qu'on appelle une "partition au hasard". On se propose de déterminer la loi de probabilité du plus grand intervalle.

Cette question, telle qu'elle vient d'être formulée peut paraître strictement théorique ; pourtant elle nous a été posée à plusieurs reprises récemment par des "chercheurs opérationnels" qui l'ont rencontrée à propos de problèmes concrets. Nous avons des raisons de penser que cette brève étude a sa place dans la présente revue.

Quelques remarques générales

Si n est le nombre de points, la partition comprend (n + 1) segments aléatoires. Soit U la grandeur du plus grand segment :

On a $$\frac{1}{n+1} \leqslant U \leqslant 1$$

Désignons par $F_n(u)$ la fonction de répartition de U, grandeur maximum des segments obtenus en répartissant n points et soit $f_n(u)\,du = dF_n(u)$ la loi élémentaire de la même variable aléatoire.

2 - RELATION DE RECURRENCE -

Pour calculer $F_{n+1}(u)$ en fonction de F_n, exprimons la probabilité que $OA_1 = u$ soit le plus grand intervalle de la partition par (n + 1) points. [On suppose ici que le premier segment OA_1 est le plus grand, et que ce segment est obtenu avec le premier point tiré (A_1).

Bien entendu on a pris u entre $1/(n+2)$ et 1.

L'évènement dont on veut calculer la probabilité s'exprime par les conditions suivantes

a) $u < OA_1 < u + du$

b) les n autres points (A_2, A_3.. A_n) tombent sur A_1I

c) Dans la partition "au hasard" du segment A_1I par les n points, le plus grand intervalle obtenu est inférieur ou égal à u.

On s'assure sans difficulté que la répartition conditionnelle des n points sur A_1I est une répartition "au hasard" sur le segment de longueur 1 - u. La fonction de répartition du plus grand intervalle V de cette partition est

$$\text{Prob}\,\{V < v\} = \text{Prob}\left\{\frac{V}{1-u} < \frac{v}{1-u}\right\} = F_n\left(\frac{v}{1-u}\right)$$

Finalement la probabilité de l'évènement (a.b.c) est :

$$du\,.(1-u)^n\,.\,F_n\left(\frac{u}{1-u}\right)$$

Il est alors immédiat de calculer $f_{n+1}(u)\,du$, probabilité que le plus grand intervalle d'une partition par (n + 1) points soit compris entre u et u + du. Il faut et il suffit que les conditions (a).(b).(c) soient réalisées mais où u est la longueur de l'un quelconque des (n + 2) intervalles obtenus et où A_1 est l'un quelconque des (n + 1) points. Ces (n + 2)(n + 1) possibilités sont également probables, d'où

$$f_{n+1}(u)\,du = (n+2)(n+1)(1-u)^n\,F_n\left(\frac{u}{1-u}\right)du$$

3 - EXPRESSION GENERALE DE $F_n(u)$ -

On calcule facilement f_n puis F_n pour n = 1, 2, 3 ce qui montre que $f_n(u)$ et $F_n(u)$ prennent des formes analytiques différentes selon les valeurs de u. Ce calcul suggère la formule générale suivante

$$F_n(u) = [(n+1)\,u - 1]^n \quad \text{pour} \quad \frac{1}{n+1} \leqslant u \leqslant \frac{1}{n}$$

$$= [(n+1)\,u - 1]^n - C^1_{n+1}\,[nu - 1]^n \quad \text{pour} \quad \frac{1}{n} \leqslant u \leqslant \frac{1}{n-1}$$

$$= [(n+1)u - 1]^n - C^1_{n+1}\,[nu - 1]^n + C^2_{n+1}\,[(n-1)]u - 1^n$$

$$\text{pour} \quad \frac{1}{n-1} \leqslant u \leqslant \frac{1}{n-2}$$

etc.

On peut résumer le résultat par l'expression

3.(1) $$\boxed{F_n(u) = \sum_{k=0} (-1)^k\,C^k_{n+1}\,[(n-k+1)\,u - 1]^n}$$

où la somme Σ est étendue aux valeurs entières positives de k(k = 0 ; k = 1 etc.) qui donnent à [(n - k + 1) u - 1] des valeurs > 0.

4 - DEMONSTRATION -

4.(1) Cas n = 1. Il est immédiat de voir que le plus grand segment obéit à une loi uniforme sur l'intervalle [1/2, 1].

Sa fonction de répartition est $F_1(u) = 2u - 1$ pour $1/2 < u < 1$ ce qui est conforme à l'expression générale 3.(1).

4.(2) Récurrence. En appliquant la relation de récurrence (2) on démontre que la relation 3.(1) s'étend du rang n au rang n + 1 :

Il faut le montrer pour tout intervalle $\left(\frac{1}{m+1}, \frac{1}{m}\right)$ de variation de u. Exprimons $\underset{n+1}{f(u)}\, du$ pour $\frac{1}{n-h+1} \leqslant u \leqslant \frac{1}{n-h+1}$. Dans cette hypothèse $\frac{u}{1-u}$ est compris entre $\frac{1}{n-h+1}$ et $\frac{1}{n-h}$ d'où

$$\underset{n+1}{f(u)}\, du = (n+2)(n+1)(1-u)^n\, du \sum_{k=0}^{h} (-1)^k C_{n+1}^k \left[(n-k+1)\frac{u}{1-u} - 1\right]^n$$

$$= \sum_{k=0}^{h} (-1)^k (n+2)(n+1)\, C_{n+1}^k \left[(n-k+1)\, u - (1-u)\right]^n$$

$$= \sum_{k=0}^{h} (-1)^k (n+2)(n+1)\, C_{n+1}^k \,.\, \left[(n-k+2)\, u - 1)\right]^n$$

Par intégration, on obtient toujours pour $\frac{1}{n-h+2} \leqslant u \leqslant \frac{1}{n-h+1}$

$$\underset{n+1}{F(u)} = \sum_{k=0}^{h} (-1)^k \frac{n+2}{n-k+2} C_{n+1}^k \left[(n-k+2)\, u - 1\right]^{n+1}$$

or

$$\frac{n+2}{n-k+2}\, C_{n+1}^k = C_{n+2}^k$$

et finalement

$$\underset{n+1}{F(u)} = \sum_{k=0}^{h} (-1)^k C_{n+2}^k \left[(n-k+2)\, u - 1\right]^{n+1} \quad \text{pour } \frac{1}{n-h+1} \leqslant u \leqslant \frac{1}{n-h}$$

expression conforme à la formule générale 3.(1).

5 - VALEURS NUMERIQUES -

Nous complétons cette étude en donnant des extraits des fonctions de répartition de U_n pour n = 4 ; n = 8 et n = 15.

n = 4. (4 points ou 5 segments)

F(1/5) = 0	F(1/3) = 0,136
F(1/4) = 0,004	F(1/2) = 0,689

n = 8. (8 points ou 9 segments)

F(1/6) = 0,0025	F(1/4) = 0,238	F(1/2) = 0,965
F(1/5) = 0,0399	F(1/3) = 0,654	

n = 15. (15 points ou 16 segments)

F(1/9) = 0,0015	F(1/7) = 0,065	F(1/5) = 0,493
F(1/8) = 0,021	F(1/6) = 0,2186	F(1/9) = 0,8698
		F(1/3) = 0,964
		F(1/2) = 0,99951