REVUE DE STATISTIQUE APPLIQUÉE

R. MEIGNIEZ

Le test des intervalles intermaximaux

Revue de statistique appliquée, tome 7, nº 2 (1959), p. 87-95
<http://www.numdam.org/item?id=RSA_1959__7_2_87_0>

LE TEST DES INTERVALLES INTERMAXIMAUX

par R. MEIGNIEZ
Psychologue à Ebauches S.A .(Neuchâtel), Suisse (1)

BUT DU TEST :

Il s'agit d'un test non-paramétrique devant mettre en évidence qu'une suite temporelle n'est pas au hasard. Il s'intègre dans la même perspective que le test du nombre de maxima décrit dans une précédente publication (2).

DEFINITIONS ET RAPPELS :

Soit une suite temporelle S_n de n valeurs inégales quelconques : v_1, v_2, ..., v_n. On suppose que la suite est circulaire, c'est-à-dire que v_n précède v_1 (2).

On nomme <u>maximum</u> toute valeur v_q telle que l'on ait simultanément $v_q > v_{q-1}$ et $v_q > v_{q+1}$. (2)

On nomme <u>intervalle</u> le groupe de valeurs compris entre deux maxima consécutifs, ces derniers exclus. La grandeur de l'intervalle est le nombre de valeurs de ce groupe.

Soit k le nombre d'intervalles de S_n. k est donc également le nombre de maxima. Soit E_n (k) l'espérance du nombre de maxima pour une suite circulaire au hasard de n valeurs.

La table des probabilités cumulées des valeurs possibles de k pour une valeur donnée de n a été fournie depuis n = 4 jusque n = 46. Au delà, il a été admis que la quantité :

$$z = \frac{3k + 1,5 - n}{\sqrt{0,4\,n}}$$

obéissait approximativement à une loi normale réduite. (2)

En fonction de ces résultats, on a proposé de considérer les valeurs de k ou de z extrêmes par rapport à ces distributions comme symptomatiques de suites non au hasard. On a cependant relevé que le test ne retient qu'une partie très réduite de l'information présente dans le donné. Aussi présentons nous aujourd'hui un test plus efficace.

Soit N_i le nombre d'intervalles de grandeur i présents dans la suite temporelle circulaire aléatoire S_n. Soit E_n (N_i) l'espérance de cette quantité. Le test

----

(1) L'étude de ce test a été réalisée dans le cadre du Service Psychologique d'Ebauches S. A., dirigé par Philippe de Coulon. Nous les remercions vivement des facilités qu'ils ont procurées à notre travail.

(2) "Le test du nombre de maxima", par R. Meigniez, Revue de Statistique Appliquée, vol. VI n° 2, 1958.

revient à mesurer l'écart entre ces valeurs effectives et leurs espérances, i prenant toutes les valeurs possibles.

Un premier objectif sera donc de calculer la valeur $E_n(N_i)$.

MODE GENERAL DE CALCUL DES $E_n(N_i)$:

Soit la suite circulaire au hasard S'_n. On en extrait au hasard un fragment s de (i + 4) valeurs, qu'on peut coder sous la forme :

$$s = \{ y_1, A, x_1, x_2, \ldots, x_i, B, y_2 \} \tag{1}$$

Définissons par $\underline{a}$ l'évènement : A maximum, aucun x maximum
" $\bar{a}$ " A non-maximum, "
" b " B maximum, "
B non-maximum, "

Nous avons identiquement :

$$P(a) = P(a\bar{b}) + P(ab) \tag{2}$$

$$\text{et : } P(b) = P(\bar{a}b) + P(ab) \tag{3}$$

D'autre part, l'expression :

$$\alpha = P(\bar{a}b) + P(\bar{a}b) + P(ab) \tag{4}$$

exprime la probabilité qu'au moins un des deux évènements a et b se produise.

On obtient alors par addition de (2) et (3) et soustraction de α :

$$P(ab) = P(a) + P(b) - \alpha \tag{5}$$

Or, P (ab) est précisément la probabilité que le fragment s détermine un intervalle i, et il est d'autre part clair que l'on a :

$$E_n(N_i) = n\,P(ab), \tag{6}$$

puisque dans la suite circulaire S_n on a la possibilité d'extraire n fragments différents de la grandeur de s, soit de i + 4 valeurs.

En conséquence, le calcul de $E_n(N_i)$ se ramène à celui de P (ab) et donc à à la résolution de l'expression (5).

LES PERMUTATIONS CONCAVES :

La notion de permutation concave nous permettra de donner un contenu algebrique aux termes de l'expression (5).

Soit un fragment de suite de r valeurs différentes. Nous nommons permutation concave de ces r valeurs toute permutation non circulaire qui ne comporte pas de maximum interne. Autrement dit, la valeur la plus élevée se trouve obligatoirement en première ou en dernière position parmi les r valeurs du fragment.

Il est clair qu'on peut obtenir toutes les permutations concaves de r valeurs en ajoutant la plus grande de ces valeurs successivement à droite et à gauche de chacune des permutations concaves des r-1 autres valeurs, ce qui double le nombre de ces dernières permutations.

Si donc U_r est le nombre des permutations concaves de r valeurs, et puisqu'on a visiblement $U_2 = 2$, on aura :

$$U_3 = 4, \quad U_4 = 8, \quad \ldots, \quad U_r = 2^{r-1} \tag{7}$$

Donc, si P_U (R) est la probabilité qu'un fragment R de r valeurs au hasard présente une permutation concave, on aura d'après (7), et étant donné que le nombre de toutes les permutations quelconques possibles de r éléments est r! :

$$P_U\,(R) = \frac{2^{r-1}}{r!} \tag{8}$$

APPLICATION AU CALCUL DE P (a) ET P (b) :

Considérons le fragment :

$$s_1 = \{A, x_1, x_2, \ldots, x_i, B\}$$

fragment extrait de s (voir (1))

On a, d'après (8) :

$$P_U\,(s_1) = \frac{2^{i+1}}{(i+2)!} \tag{9}$$

De même si l'on considère le fragment :

$s_2 = \{y_1, A, x_1, x_2, \ldots, x_i, B\}$ également extrait de s, on a :

$$P_U\,(s_2) = \frac{2^{i+1}}{(i+3)!} \tag{10}$$

Or, si nous considérons la suite fragmentaire s, nous voyons que $P_U\,(s_1)$ représente la probabilité qu'elle réalise au hasard le cas: A maximum ou non ; A, x_1, x_2, ..., B sans maximum interne.

De même, $P_U\,(s_2)$ représente la probabilité que la suite s réalise au hasard le cas : A non maximum ; A, x_1, x_2, ..., B sans maximum interne.

P (a) représentant pour la suite s la probabilité de l'évènement A maximum, aucun x maximum, on aura donc finalement logiquement :

$$P\,(a) = P_U\,(s_1) - P_U\,(s_2),$$

soit, d'après (9) et (10) :

$$P\,(a) = \frac{2^{i+1}}{(i+2)!} - \frac{2^{i+2}}{(i+3)!} \tag{11}$$

formule qui, par raison de symétrie, exprime également la valeur de P (b).

APPLICATION DE (8) AU CALCUL DE α :

La même méthode s'applique au calcul de α.

D'après l'expression (1) de s, on voit que P_U (s) est la probabilité de : A et B non maxima ; A, x_1, ..., x_i, B, sans maximum interne. Parallèlement, l'inspection du fragment s_1 nous montre que P_u (s_1) est la probabilité de : A $\binom{ou}{et}$ B maximum ou non ; A, x_1, ..., x_i, B, sans maximum interne. On en conclut, d'après la définition de α dans (4) que :

$$\alpha = P_u (s_1) - P_u (s)$$

Soit, d'après (8) :

$$\alpha = \frac{2^{i+1}}{(i+2)!} - \frac{2^{i+3}}{(i+4)!} \tag{12}$$

ESPERANCE DU NOMBRE D'INTERVALLES DE GRANDEUR i :

En remplaçant dans (5) les termes du second membre par leur valeur en (11) et (12), il vient :

$$P (ab) = \frac{2^{i+1} i (i+3)}{(i+4)!} \tag{13}$$

Et, d'après (6) :

$$E_n (N_i) = \frac{n\ 2^{i+1}\ i\ (i+3)}{(i+4)!} \tag{14}$$

On peut chercher la probabilité correspondante à cette espérance. Soit donc P_i la probabilité qu'un intervalle pris au hasard soit de grandeur i. On a établi dans le travail cité en note (2) que, E_n (k) étant l'espérance du nombre total d'intervalles pour une suite circulaire au hasard de n valeurs, on a :

$$E_n (k) = \frac{n}{3} \tag{15}$$

On aura donc :

$$P_i = \frac{E_n (N_i)}{E_n (k)} = \frac{3 . 2^{i+1} i (i+3)}{(i+4)!} \tag{16}$$

On constate que les P_i sont indépendants de n, sauf exceptions que nous allons envisager.

Cas particuliers de P_i :

En effet, notre raisonnement repose sur une structure de s (cf. (1)) qui implique la relation :

$$i + 4 \leqslant n\rho \tag{17}$$

Il nous reste donc à calculer P_i pour les cas extrêmes, très peu probables pour n grand : i = n-1 et i = n-3 (le cas i = n-2 est absurde, comme impliquant la contiguité de deux valeurs maxima).

a) Cas i = n-1 :

Un seul maximum est possible : la plus forte valeur de S_n.

Pour le fragment s_3 constitué par les n-1 autres valeurs, on a selon (8) :

$$P_u (s_3) = \frac{2^{n-2}}{(n-1)!}$$

Comme on peut extraire un seul tel fragment, on a en raisonnant comme ci-dessus :

$$E_n\,(N_{n-1}) = 1 \times \frac{2^{n-2}}{(n-1)\,!} \tag{18}$$

Et, d'après (15) :

$$P_{n-1} = \frac{E_n\,(N_{n-1})}{E_n\,(k)} = \frac{3\,.\,2^{n-2}}{n\,!} \tag{19}$$

b) Cas i = n - 3 :
On observe simplement que :

$$P_{n-3} = 1 - P_{n-1} - \sum_{i=1}^{n-4} P_i \tag{20}$$

$$\text{Or, } \sum_{i=1}^{n-4} P_i = 3 \sum_{i=1}^{n-4} \left(\frac{2^{i+1}}{(i+2)\,!} - \frac{2^{i+3}}{(i+3)\,!} + \frac{2^{i+3}}{(i+4)\,!} \right)$$

Soit, en développant :

$$\sum_{i=1}^{n-4} P_i = 3 \left[\left(\frac{2^2}{3\,!} + \frac{2^3}{4\,!} + \ldots + \frac{2^{n-3}}{(n-2)\,!} \right) - \left(\frac{2^4}{4\,!} + \frac{2^5}{5\,!} + \ldots + \frac{2^{n-1}}{(n-1)\,!} \right) + \left(\frac{2^4}{5\,!} + \frac{2^5}{6\,!} + \ldots + \frac{2^{n-1}}{n\,!} \right) \right]$$

Ce qui donne, après regroupement sur les dénominateurs identiques :

$$\sum_{i=1}^{n-4} P_i = 3 \left[\frac{1}{3} - \frac{2^{n-2}}{(n-1)\,!} + \frac{2^{n-1}}{(n)\,!} \right] = 1 - \frac{3\,(n-2)\,2^{n-2}}{n!} \tag{21}$$

En remplaçant dans (20), d'après (19) et (21), il vient :

$$P_{n-3} = 3\,(n-3)\,\frac{2^{n-2}}{n!} \tag{22}$$

On voit donc que P_{n-1} et P_{n-3} dépendent bien de n, mais décroissent très rapidement lorsque cette variable croit, ce qui les rend pratiquement négligeables.

Probabilités cumulées $P_{i>m}$:

Il est intéressant pour les besoins pratiques de connaître la probabilité cumulée pour les valeurs extrêmes de P_i. Soit m un entier $\leqslant$ n-4. Soit $P_{i>m}$ la probabilité qu'un intervalle extrait au hasard soit d'une grandeur qui dépasse m.

Observons le dernier membre de (21). Il est clair, d'après le raisonnement qui y aboutit, qu'on obtiendrait de la même manière :

$$\sum_{i=1}^{m} P_i = 1 - \frac{3\,.\,2^{m+2}\,(m+2)}{(m+4)\,!} \tag{23}$$

en substituant simplement m à n-4.

D'où :

$$P_{i>m} = 1 - \sum_{i=1}^{m} P_i = \frac{3\,.\,2^{m+2}\,(m+2)}{(m+4)\,!} \tag{24}$$

On vérifie aisément par (24) que $P_{i>0} = 1$

Valeurs numériques des P_i :

La formule (16) nous permet ce calcul, avec la réserve que $i \leqslant n-4$ doit être vérifié. Nous donnons ci-dessous quelques valeurs fondamentales, et, à titre de comparaison, les valeurs correspondantes pour la loi de Poisson, de variable i-1 et de paramètre 1 (valeur pour laquelle l'espérance de i est la même dans les deux distributions.). On remarquera que la loi de Poisson fournit une approximation très grossière de la loi des intervalles intermaximaux.

TABLEAU I

	i = 1	i = 2	i = 3	i = 4	i = 5	i = 6	i = 7	i = 8	Total
P_i	0,4000	0,3333	0,1714	0,0667	0,0211	0,0057	0,0013	0,0003	0,9998
Poisson	0,3679	0,3679	0,1839	0,0613	0,0153	0,0031	0,0005	0,0001	1,0000

Vérification expérimentale de la loi des P_i :

En extrayant d'une urne au hasard et jusqu'à épuisement 198 billes numérotées de 1 à 198, et en répétant 100 fois l'opération, on a obtenu 100 suites au hasard. On a considéré les 58 premiers intervalles de chaque suite et déterminé les N_i. On a considéré qu'en faisant le total des N_i pour chaque valeur de i dans les 100 suites on obtiendrait une approximation valable de ce qu'on aurait trouvé dans une suite circulaire de 5 800 intervalles. Le tableau II compare ces résultats expérimentaux aux valeurs théoriques 5 800 P_i.

TABLEAU II

	i = 1	i = 2	i = 3	i = 4	i = 5	i = 6	i = 7	i = 8	$\underbrace{i = 9 \ \& +}$	Total
N_i	2 298	1 962	981	406	113	28	9	3	0	5 800
5 800 P_i	2 320	1 933	994	387	123	33	8	2	0	5 800

L'adéquation théorico-expérimentale parait bonne. La première tentation qui vient à l'esprit est de calculer un χ^2 sur la base du tableau II. Mais cette procédure soulèverait de graves objections pour des raisons de non-indépendance.

En effet, supposons pour illustrer qu'au cours de l'extraction de nos billes numérotées de l'urne on en vienne à former un intervalle de grande taille (par exemple, i = 8). Il y a une forte probabilité que le dernier maximum sorti qui borne cet intervalle possède une valeur numérique élevée par rapport aux autres valeurs de l'urne, sans quoi cette suite de grandeur 8 deviendrait encore moins probable (raisonnement qui s'applique d'ailleurs aussi au maxima de gauche de la suite). Or, la probabilité forte que le maxima de droite soit élevé augmente à son tour la probabilité que l'intervalle suivant à sortir soit lui-même de grande taille. Autrement dit, la grandeur d'un intervalle est une fonction aléatoire de celle de l'intervalle précédent, et nous avons affaire à une variété de chaîne de Markov. Les cases du tableau II ne concernent donc pas en droit des fréquences d'occurence indépendantes, et la technique du χ^2 ne parait pas applicable.

Nous avons cependant pu montrer que cette objection n'est pas absolue, en démontrant expérimentalement que la quantité :

$$K^{(2)} = \sum_{i=1} \frac{(NP_i - N_i)^2}{NP_i} \qquad (25)$$

se comporte avec une bonne approximation comme une loi de χ^2, au moins à partir d'une certaine valeur de N, nombre d'intervalles total du fragment de suite considéré. On rappelle que N_i représente le nombre d'intervalles de grandeur i, et P_i la probabilité théorique calculée plus haut qu'un intervalle pris au hasard soit de grandeur i. Il est à noter que sur ce plan expérimental la circularité des suites n'est plus prise en considération.

APPROXIMATION DE $K^{(2)}$ A L'AIDE DE LA DISTRIBUTION DE χ^2

Pour chacune de nos 100 suites aléatoires de 58 intervalles obtenues par tirage de billes numérotées, on a calculé la valeur $K^{(2)}$ à la manière d'un χ^2, selon l'expression (25). Pour cela, on a utilisé les 4 classes : i = 1, i = 2, i = 3 et i ⩾ 4. Les "fréquences attendues" NP_i dans chacune de ces classes étaient respectivement 23,2 ; 19,3 ; 9,9 ; 5,5.

En possession de nos 100 $K^{(2)}$, nous avons lu dans la table des χ^2 avec 3 degrés de liberté, les 100 seuils de probabilité correspondants. L'hypothèse était que, N étant élevé, les effets markoviens s'en trouveraient amoindris. Nous avons ainsi obtenu le tableau suivant :

TABLEAU III

Seuils lus dans la table	1.00	.95	.90	.80	.70	.50	.30	.20	.10	.05	.00	T
Nbre. théor. de $K^{(2)}$	5	5	10	10	20	20	10	10	5	5		100
Nbre. expér. de $K^{(2)}$	6	3	11	8	23	21	6	13	5	4		100

On voit que l'adéquation n'est pas mauvaise, et qu'en particulier les variations tendent à se compenser de proche en proche. A gauche et à droite de .50, on trouve respectivement 51 et 49 valeurs. Si l'on détaille les extrémités, la concordance devient encore plus frappante :

TABLEAU IV

Seuils lus dans la table	1.00	.99	.98	.95	.05	.02	.01	.00
Nbre. théor. de $K^{(2)}$	1	1	3		3	1	1	
Nbre. expér. de $K^{(2)}$	1	1	4		2	1	1	

Notre conclusion, c'est qu'au moins à partir d'un nombre assez élevé d'intervalles il est possible d'utiliser $K^{(2)}$ pour tester si une suite est ou non aléatoire du point de vue des tendances locales qu'elle peut ou non renfermer. Il conviendrait d'investiguer plus à fond cette possibilité en fonction de la grandeur de N.

UTILISATION PRATIQUE DU TEST :

Résumons tout d'abord la démarche. Soit une suite de N intervalles dont on se demande si elle est l'objet de variations locales graduelles significatives, ou au contraire encore d'une absence significative de telles variations graduelles. On pourra par exemple avoir à se poser des problèmes de mélange.

On comptera donc N_i, nombre d'intervalles de chaque grandeur, i = 1, i = 2, ..., i ≥ m, présents dans la suite. Soit t le nombre des classes ainsi distinguées.

On comparera pour chacune des t classes le N_i obtenu avec la valeur théorique NP_i les P_i étant fournis par le tableau I, et l'on calculera $K^{(2)}$ selon la formule (25). On entrera dans la table des χ^2 avec t-1 degrés de liberté avec une valeur de χ^2 égale à celle de $K^{(2)}$. Le seuil de probabilité ainsi obtenu pourra être considéré comme une approximation valable du seuil auquel nous pouvons dire que notre suite n'est pas aléatoire sous l'angle considéré.

Prenons un exemple simple. On surveille l'évolution d'un indice économique annuel concernant un certain produit, indice qu'il a été possible d'obtenir pour les 58 dernières années. On se demande à ce propos si l'indice varie de façon indépendante tous les ans, ou si une même source de variation peut être supposée à l'œuvre sur plusieurs années successives. Dans ce cas, on suppose que la non-indépendance des années successives peut se manifester aussi bien par une gradualisation des variations de l'indice que par des réactions d'alternance tendant à faire se succéder une valeur forte à une valeur faible et réciproquement.

L'hypothèse testée est donc celle de l'indépendance des indices successifs, et s'oppose à celle de la non-indépendance locale sous toutes ses formes.

On se contente de donner ci-après le début et la fin supposés de la suite des 58 indices :

.157 .124 .112 .135 .146 .113 .169 .152 .147 .136 .148 .109
.173 .161 .... .084 .105 .096 .117 (.157)

On a souligné ensemble les valeurs qui consituent un même intervalle. .157, qui figure en tête de la suite, est répété entre parenthèses comme valeur supplémentaire à la fin de la suite, car, du fait que celle-ci est supposée circulaire, il joue le rôle d'un maximum, entouré par .117 et .124.

On compte le nombre d'intervalles de chaque grandeur : ce sont les valeurs dites expérimentales du tableau qui suit. Les valeurs dites théoriques sont obtenues en multipliant par 58 les probabilités P_i du tableau I.

D'où le tableau résumant le calcul de $K^{(2)}$:

	i = 1	i = 2	i = 3	i ≥ 4	Somme
Théorique	23,20	19,33	9,94	5,53	58
Expérimental	33	17	4	4	58
Différence	9,80	2,33	5,94	1,53	
$(\text{Différence})^{(2)}$	96,04	5,43	35,28	2,34	
$(\text{Diff.})^{(2)}/\text{Théor.}$	4,13	0,28	3,54	0,42	$K^{(2)} = 8,37$

On voit conc que le calcul de $K^{(2)}$ est identique à celui d'un χ^2. La valeur obtenue de $K^{(2)}$ (ici, 8,37) est ensuite lue dans une table de χ^2 avec un nombre de degrés de liberté égal au nombre de classes moins un (ici, 4 - 1 = 3), et l'on obtient en regard le seuil de probabilité auquel l'hypothèse de l'indépendance locale des indices successifs peut être admise étant données nos valeurs expérimentales.

Or, nous lisons dans la table que pour 8,37 et 3 degrés de liberté on se trouve placé entre les seuils de probabilité 02 et .05. Ceci signifie qu'on n'a qu'entre 2 % et 5 % de chances de ne pas se tromper en affirmant l'indépendance locale de notre suite. L'hypothèse de l'indépendance est donc rejetée à un seuil de P = .05, souvent considéré comme un seuil "raisonnable" (alors qu'elle ne serait pas rejetée, par exemple, à un seuil de .01).

CONCLUSION

Le test des intervalles intermaximaux est dans son principe parent de celui de Olmstead(1) qui considère l'importance des suites croissantes et décroisantes présentes dans la suite totale donnée. Autrement dit, cet auteur considère le nombre de valeurs qui séparent un maximum du minimum suivant, ou un minimum du maximum qui lui succède.

Il serait sans doute utile de se livrer à une comparaison du test d'Olmstead avec celui des intervalles intermaximaux.

Il est à noter que le test des intervalles intermaximaux peut également s'appliquer sous forme de test des intervalles interminimaux, par raison de symétrie. Si on entend en effet par minimum toute valeur inférieure à celle qui la précède et à celle qui la suit, l'ensemble de nos considérations s'applique immédiatement à ce nouveau champ, en remplaçant simplement la notion de permutation concave par celle de permutation convêxe.

Dans cette perspective, il y aurait lieu de trouver le moyen de combiner l'application successive des deux points de vue aux mêmes données. Nous pensons qu'on accroitrait ainsi la sensibilité du test sans pour autant compliquer la procédure.

----

(1) Olmstead, "Distribution of sample arrangements for runs up and down", cité par E. Morice, in Revue de Stat. Appl., 1956, vol IV, n° 4, pp 105-106.