Revue de statistique appliquée

D. Dugué

Sur un test indépendant de la distribution parente

Revue de statistique appliquée, tome 14, n° 4 (1966), p. 69-73
<http://www.numdam.org/item?id=RSA_1966__14_4_69_0>

SUR UN TEST INDÉPENDANT DE LA DISTRIBUTION PARENTE

D. DUGUÉ

Directeur de l'Institut de Statistique de l'Université de Paris

Les résultats ci-après, présentés par Monsieur le Professeur Dugué, sont dus à Monsieur Pierre Dufresne, Ingénieur à Electricité de France décédé en 1964, qui les avait présentés en 1958 au 83e Congrès des Sociétés Savantes (Dijon 1958).

L'intérêt qu'il portait à la théorie des jeux (et plus particulièrement à ceux qui peuvent être tirés des problèmes de dépouillement d'urnes) l'avait amené à perfectionner des méthodes d'une grande élégance s'inspirant du fameux triangle de Pascal.

Soit une urne contenant a bulletins au nom du candidat A et b bulletins au nom du candidat B. L'urne étant ainsi constituée, on procède au dépouillement en tirant les bulletins un à un sans les remettre dans l'urne, bien entendu, et on a ainsi une suite de lettres de la forme :

A B B A B . . .

ce que l'on appelle aujourd'hui un mot.

Dans un tel dépouillement, on dira qu'il y a un dépassement quand "la majorité change de camp", c'est-à-dire quand A par exemple ayant la majorité (ce qui veut dire que le nombre de bulletins A tirés jusqu'à un certain moment est strictement supérieur au nombre de bulletins B tirés jusqu'au même moment), deux tirages successifs donnent à B la majorité.

Soit par exemple le dépouillement :

A B B A A B A B . . .

Il y a un dépassement au profit de B au 3e bulletin, un autre au profit de A au 5e et un autre au profit de A au 7e.

Nous appellerons N(x ; a, b) le nombre de dépouillements d'une urne contenant a bulletins A et b bulletins B pour lesquels il y a x dépassements. Il y a en tout $\binom{a+b}{a}$ dépouillements possibles et également probables. La probabilité d'avoir x dépassements est donc :

$$\frac{N(x\,;\,a\,,\,b)\;a!\,b!}{(a+b)!}$$

S'il y a x dépassements, on voit facilement qu'il y a au moins $2x + 1$ bulletins (donc $a + b \geq 2x + 1$) car chaque dépassement est précédé d'au moins deux bulletins au nom du même candidat et le premier dépassement ne peut se produire qu'au troisième bulletin.

Si $a + b = 2x + 1$, minimum possible pour $a + b$, le dépouillement doit être tel que l'on ait des successions de paires de bulletins au même nom A ou B avec changement de majorité à chaque paire ; on doit donc avoir $a = x$, $b = x + 1$ ou $a + 1$, $b = x$. On voit aisément que dans ce cas $N(x ; x, x + 1) = N(x; x + 1, x) = 1$.

On doit de même avoir $x \leqslant a$ et $x \leqslant b$ car entre deux dépassements consécutifs et avant le premier dépassement, il y a au moins un bulletin A et au moins un bulletin B.

Le problème se pose de calculer $N(x ; a, b)$ pour $a + b > 2x + 1$.

La solution peut être obtenue à partir du lemme suivant de Dufresne :

(i) Si $a + b > 2x + 1$ et $|a - b| \neq 1$

$$N(x ; a, b) = N(x ; a - 1, b) + N(x ; a, b - 1)$$

(ii) Si $a + b > 2x + 1$ avec $a = b + 1$

$$N(x ; b + 1, b) = N(x ; b + 1, b - 1) + \frac{1}{2} [N(x ; b, b) + N(x - 1 ; b, b)]$$

La démonstration de ce lemme est très simple.

Si $|a - b| \neq 1$ le dernier bulletin tiré de l'urne ne peut amener de renversement de la majorité et par conséquent ne peut augmenter le nombre de dépassements acquis après l'avant dernier bulletin.

Ce dernier bulletin est un bulletin A ou un bulletin B. Les classes des dépouillements se terminant par A et de ceux se terminant par B sont évidemment disjointes. Dans la première, il y a $N(x; a-1, b)$ dépouillements à x dépassements. Dans la seconde $N(x ; a, b - 1)$ dépouillements à x dépassements. Ce qui établit le (i).

Si $a = b + 1$, la classe des dépouillements à x dépassements peut se décomposer en trois classes disjointes.

(1) Dépouillements se terminant par B et ayant x dépassements leur nombre est $N(x ; b + 1, b - 1)$.

(2) Dépouillements se terminant par un A et ou l'avant dernier bulletin est un B et par conséquent dans lesquels le dernier bulletin n'amène aucun dépassement. Il faut donc qu'après avoir dépouillé b bulletins A et b bulletins B, le dernier étant un B on ait x dépassements, ce qui donne $N(x ; b, b)/2$ dépouillements dans cette classe.

(3) Dépouillements se terminant par un A et où l'avant dernier bulletin est un A et par conséquent dans lequel le dernier bulletin amène un dépassement. Dans cette classe, après avoir dépouillé b bulletins A et b bulletins B le dernier bulletin étant un A on doit avoir $x - 1$ dépassement, cette classe contient donc $N(x - 1) ; b, b)/2$ dépouillements. C'est la démonstration du (ii).

L'utilisation de ce lemme jointe aux valeurs initiales pour $N(x\,;\,a,b)$ permet de donner les formules générales :

Si $b = x$ on devra avoir $a > b$ (à cause de $a + b > 2x + 1$). Il n'y a qu'un seul dépouillement possible, car comme nous l'avons remarqué, entre deux dépassements consécutifs, il y a au moins un bulletin B et il y a aussi au moins un bulletin B avant le premier dépassement.

Si donc, il y a exactement autant de dépassements que de bulletins B les derniers bulletins après le dernier dépassement doivent tous être des A.

Il en résulte qu'il n'y a qu'un seul dépouillement possible dans ce cas. Donc :

$$N(x\,;\,a,\,x) = 1$$

De même, naturellement

$$N(x\,;\,x,\,b) = 1$$

Quant à $N(0\,;\,a,\,b)$ son calcul relève du problème du scrutin ordinaire, a étant supérieur ou égal à b quel est le nombre de dépouillements ou A ne perd jamais la majorité ? Par les méthodes habituelles (principe de symétrie de Désiré André par exemple), on peut montrer que si $a \neq b$,

$$N(0\,;\,a,\,b) = \binom{a+b}{b} - \binom{a+b}{b-1}$$

Tableaux donnant Différentes Valeurs de $N(x\,;\,a;\,b)$

$N(0\,;\,a,\,b)$

a + b \ b	0	1	2	3	4	5	6	7	8	9	10	11
1	1	1										
2	1	2	1									
3	1	2	2	1								
4	1	3	4	3	1							
5	1	4	5	5	4	1						
6	1	5	9	10	9	5	1					
7	1	6	14	14	14	14	6	1				
8	1	7	20	28	28	28	20	7	1			
9	1	8	27	48	42	42	48	27	8	1		
10	1	9	35	75	90	84	90	75	35	9	1	
11	1	10	44	110	165	132	132	165	110	44	10	1
12	1	11	54	154	275	297	264	297	275	154	54	11

N(1 ; a, b)

b / a + b	0	1	2	3	4	5	6	7	8	9	10	11
3		1	1								–	
4		1	2	1								
5		1	4	4	1							
6		1	5	8	5	1						
7		1	6	14	14	6	1					
8		1	7	20	28	20	7	1				
9		1	8	27	48	48	27	8	1			
10		1	9	35	75	96	75	35	9	1		
11		1	10	44	110	165	165	110	44	10	1	
12		1	11	54	154	275	330	275	154	54	11	1

N(2 ; a, b)

5			1	1								
6			1	2	1							
7			1	6	6	1						
8			1	7	12	7	1					
9			1	8	27	27	8	1				
10			1	9	35	54	35	9	1			
11			1	10	44	110	110	44	10	1		
12			1	11	54	154	220	154	54	11	1	

N(3 ; a, b)

7				1	1							
8				1	2	1						
9				1	8	8	1					
10				1	9	16	9	1				
11				1	10	44	44	10	1			
12				1	11	54	88	54	11	1		

N(4 ; a, b)

9					1	1						
10					1	2	1					
11					1	10	10	1				
12					1	11	20	11	1			

$$N(5\ ;\ a,\ b)$$

b / a + b						5	6	7				
11						1	1					
12						1	2	1				

et

$$N(0\ ;\ a,\ a) = 2\ \frac{2a!}{a!\,(a+1)!}$$

Ces remarques permettent d'arriver aux formules générales données par M. P. Dufresne :

$$N(x\ ;\ a,\ b) = \binom{a+b}{a+x} - \binom{a+b}{a+x+1} \qquad \text{si} \quad a > b > x$$

et

$$N(b\ ;\ a,\ b) = 1 \qquad \text{si} \quad a > b$$

ou

$$N(x\ ;\ a,\ a) = 2\binom{2a-1}{a+x} - 2\binom{2a-1}{a+x+1} \qquad \text{si} \quad a > x+1$$

et

$$N(a-1\ ;\ a,\ a) = 2$$

Ces formules ont évidemment des applications immédiates au mouvement brownien et à l'analyse séquentielle. On peut en tirer un test de rang. Soient a résultats d'une série et b résultats d'une autre portant sur des variables aléatoires dont on veut tester l'hypothèse qu'elles ont même fonction de répartition de la probabilité, cette fonction étant supposée continue. Selon les méthodes classiques dans les tests de rang, on rangera les a + b résultats par ordre de grandeur croissant et dans la série des a + b résultats, on notera le nombre de dépassements au sens que nous avons donné à ce mot au début de cet article. Selon que la probabilité du nombre de dépassements observée sera inférieure ou supérieure au seuil de signification adopté, on rejetera ou on acceptera l'hypothèse. Il est donc bon d'avoir une table des probabilités des différentes valeurs des dépassements pour des valeurs de a et b données.

On peut facilement déduire des résultats précédents, les formules suivantes :

<u>Cas 1</u> : $a > b$

$$\text{Prob}\,[x \leqslant c] = 1 - \frac{a!}{(a+c+1)!}\,\frac{b!}{(b-c-1)!} \qquad \text{si} \quad 0 \leqslant c < b$$

$$= 1 \qquad \text{si} \quad c = b$$

<u>Cas 2</u> : $a = b$

$$\text{Prob}\,[x \leqslant c] = 1 - \frac{a!}{(a+c+1)!}\,\frac{(a-1)!}{(a-c-2)!} \qquad \text{si} \quad 0 \leqslant c < a-1$$

$$\text{Prob}\,[x \leqslant c] = 1 \qquad \text{si} \quad c = a-1$$

qui sont d'une remarquable simplicité.