A5118
Scan A3121
A18241
Kreweman
Sur un probleme...
add to 3 soap

1981, XXVI, fasc. 2, 69-87

A5118
A3121
A18241

~~Quelle revue~~

Publication de l'Institut de Statistique de Université de Paris Vol 26 1981.

PLIS

added ref

A5118
A3121
A18241

SUR UN PROBLEME DE SCRUTIN A PLUS DE DEUX CANDIDATS

G. Kreweras, Institut de Statistique

1. INTRODUCTION

Le problème classique du scrutin, posé par Joseph Bertrand [2] et résolu plus élégamment par Désiré André [1], concerne le recomptage, après mélange, des bulletins d'un vote, uninominal dans lequel un premier candidat a obtenu a_1 voix et un second a_2 voix, avec $a_1 > a_2$: la probabilité pour qu'au cours de ce recomptage le premier candidat demeure constamment en tête au sens strict est

$$(1) \qquad p(a_1,a_2) = \frac{a_1 - a_2}{a_1 + a_2}$$

A18241

Un des traits remarquables de ce problème est un certain contraste entre la simplicité de l'énoncé et de la solution, et l'ingéniosité de la méthode ("principe de réflexion") permettant d'établir le résultat.

Dans ce qui suit nous appellerons le problème évoqué ci-dessus la "version stricte" du problème du scrutin. Pour le résoudre on passe généralement par une "version large" du même problème, laquelle consiste à supposer que $a_1 \geqslant a_2$ et à demander avec quelle probabilité le premier candidat demeure

A5118 A3121

constamment en tête <u>au sens large</u>. La résolution de cette version large consiste en un dénombrement des chemins qui, dans $\mathbb{N} \times \mathbb{N}$, joignent le point (0,0) au point (a_1,a_2) en ne passant que par des points (x_1,x_2) pour lesquels on a $x_1 \geqslant x_2$, dénombrement classique qui donne un nombre de chemins égal à

$$(2) \qquad f(a_1,a_2) = \frac{(a_1+a_2)!(a_1-a_2+1)}{(a_1+1)!a_2!} .$$

La probabilité $q(a_1,a_2)$, réponse à la question de la "version large", est le quotient de (2) par le binomial $(a_1+a_2)!/a_1!a_2!$, c'est-à-dire

$$(3) \qquad q(a_1,a_2) = \frac{a_1-a_2+1}{a_1+1}$$

Il ne s'attache pas en général à la formule (3) un intérêt aussi considérable qu'à la formule équivalente (2). En effet cette formule (2) se généralise classiquement à un nombre quelconque n de candidats, ayant obtenu respectivement $a_1,a_2,\ldots,a_n$ voix, avec $a_1 \geqslant a_2 \geqslant \ldots \geqslant a_n$, grâce au dénombrement des chemins qui, dans $\mathbb{N}^n$, joignent $(0,0,\ldots,0)$ à $(a_1,a_2,\ldots,a_n)$ en ne passant que par des points $(x_1,x_2,\ldots,x_n)$ satisfaisant à $x_1 \geqslant x_2 \geqslant \ldots \geqslant x_n$. La formule généralisant (2) est alors

$$(4) \quad f(a_1,a_2,\ldots,a_n) = \frac{(a_1+a_2+\ldots+a_n)! \prod_{1\leqslant i<j\leqslant n}(a_i-a_j-i+j)}{(a_1+n-1)!(a_2+n-2)! \ldots (a_{n-1}+1)!\, a_n!} .$$

C'est la célèbre formule de Young, établie par celui-ci [7] à l'occasion de ses études sur la représentation du groupe symétrique.

Rappelons deux autres formulations, manifestement équivalentes, de la question à laquelle répond la formule de Young : c'est le nombre d'ordres de construction "acceptables" d'un muret de briques construit sur l'axe des x , adossé à l'axe des y et formé de n couches successives de $a_1,a_2,\ldots,a_n$ briques ; et c'est aussi le nombre de chemins qui, dans le treillis de Young, joignent l'origine au point $(a_1,a_2,\ldots,a_n)$ (cf. [4], p. 35).

Rappelons aussi que la formule de Young a reçu plus tard la très belle reformulation de Frame, Robinson et Thrall par les longueurs d'équerre (hook lengths) [3].

En divisant $f(a_1,a_2,\ldots,a_n)$ par le multinomial $(a_1+\ldots+a_n)!/a_1! \ldots a_n!$, on obtiendrait une expression monôme de la probabilité correspondante $q(a_1,a_2,\ldots,a_n)$, qui répondrait à l'extension à n candidats de la "version large" du problème du scrutin.

L'objet central du présent travail est de généraliser au cas de n candidats la version stricte du problème du scrutin. On suppose que les nombres de voix obtenues sont

$$a_1 > a_2 > \ldots > a_n$$

et l'on exige que les situations transitoires $(x_1, x_2, \ldots, x_n)$ rencontrées pendant le recomptage après mélange satisfassent "autant que possible" à

$$x_1 > x_2 > \ldots > x_n \ .$$

Les mots "autant que possible" signifient que les k dernières inégalités strictes ne sont pas (et ne peuvent pas être) exigées tant qu'aucun des k derniers candidats n'a encore été crédité d'une voix.

Plus précisément, nous montrerons que la probabilité correspondante est

$$(5)\quad p(a_1, a_2, \ldots, a_n) = \prod_{1 \leqslant i < j \leqslant n} \frac{a_i - a_j}{a_i + a_j} \ ,$$

ce qui est une extension remarquable, et à notre connaissance nouvelle, de la formule (1).

Aucun argument probabiliste simple ne semble pouvoir être invoqué pour prouver que

$$p(a_1, a_2, \ldots, a_n) = \prod_{1 \leqslant i < j \leqslant n} p(a_i, a_j) \ .$$

Nous prendrons donc le parti d'établir d'abord l'analogue, en version stricte, de la formule (4), pour arriver ensuite à (5) par division par le multinomial $(a_1+a_2+\ldots+a_n)!/a_1!a_2!\ldots a_n!$.

2. LE TREILLIS DE SYLVESTER

On appelle, d'une manière maintenant traditionnelle, "suite de Young" toute suite finie, décroissante au sens large, d'entiers non-négatifs, éventuellement prolongée par autant qu'on veut de termes nuls.

Nous proposons d'appeler "suite de Sylvester" toute suite, décroissante au sens strict (et par conséquent finie) d'entiers positifs, éventuellement prolongée par autant qu'on veut de termes nuls. Nous choisissons cette dénomination parce que Sylvester a spécialement étudié ces suites([6], cité par [5]).

Exemples :	6 5 5 3 1(0 0 0)	est une suite de Young (mais non de Sylvester)
	7 5 4 1(0 0 0 0)	est une suite de Sylvester (donc également une suite de Young).

Les suites de Young se structurent, par les opérations inf et sup décomposées terme à terme, en un treillis qui a été étudié en détail (cf. [4], p. 32). Il est clair que les suites de Sylvester forment un sous-treillis du treillis de Young, sous-treillis que nous appellerons le treillis de Sylvester.

Etant donné deux suites de Sylvester

$$X' = (x_1' \; x_2' \dots)$$
$$\text{et} \quad X = (x_1 \; x_2 \; \dots),$$

nous dirons que X' est antécédente de X (ou que X est conséquente de X' , ou couvre X') s'il existe un indice k tel que

$$x_k = x_k'+1 \quad \text{et} \quad \forall i \neq k \quad x_i = x_i' \; .$$

Les figures n^{os} 1 et 2 représentent côte à côte par leurs "graphes de couverture" le treillis de Young et le treillis de Sylvester. Les deux figures s'arrêtent à des "niveaux" différents (niveau = somme des termes) parce que le treillis de Sylvester se complique moins vite que le treillis de Young.

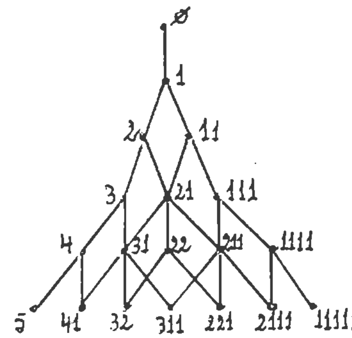

Fig. 1

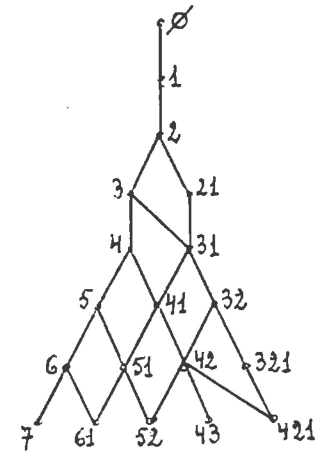

Fig. 2

De même qu'un recomptage de bulletins, dans la version <u>large</u> du problème du scrutin, est un cheminement d'antécédent en conséquent dans le treillis de Young, de même un recomptage dans la version <u>stricte</u> est un cheminement dans le treillis de Sylvester.

Soit $A = (a_1, a_2, \dots, a_n)$ un point du treillis de Sylvester. Tout cheminement qui arrive en A passe par l'un des antécédents Z de A ; si donc on appelle g(A) le nombre de cheminements aboutissant en A, il est clair que

$$(6) \qquad g(A) = \sum_{Z \text{ ant. de } A} g(Z) \; .$$

La formule que nous nous proposons d'établir est

$$(7)\quad g(a_1,a_2,\ldots,a_n) = \frac{(a_1+a_2+\ldots+a_n)!}{a_1!a_2!\ \ldots\ a_n!}\ \cdot \prod_{1\leqslant i<j\leqslant n}\frac{a_i-a_j}{a_i+a_j}\ .$$

La méthode utilisée sera la récurrence par rapport au niveau $m=a_1+a_2+\ldots+a_n$ de A, puisque tous les antécédents de A sont de niveau $m-1$, et que la formule se vérifie trivialement pour les premiers niveaux.

Donnons-nous A au niveau m du treillis de Sylvester et supposons (7) établie pour tous les antécédents de A appartenant au niveau m-1 de ce treillis. Pour passer de A à un antécédent il faut diminuer d'une unité un terme et un seul de A, par exemple le k-ième. Mais il peut se faire que pour certaines valeurs de k la suite A_k ainsi obtenue ne soit pas de Sylvester : cela arrive si et seulement si $a_k-1 = a_{k+1}$. Si néanmoins on calcule pour une telle suite A_k le second membre de (7), le facteur a_i-a_j devient nul pour $i=k$ et $j=k+1$, donc la contribution d'une telle suite à une somme de termes $g(A_k)$ est nulle.

Il en résulte que, pour conduire la démonstration par récurrence, on peut remplacer (6) par

$$(8)\qquad g(A) = \sum_{k=1}^{n} g(A_k)$$

où l'on désignera par A_k la suite (de Sylvester ou non) obtenue en remplaçant a_k par a_k-1.

Par hypothèse de récurrence on peut appliquer la formule (7) au calcul de $g(A_k)$. Le remplacement de a_k par a_k-1 a pour effet de remplacer le facteur multinomial du second membre de (7) par

$$\frac{(a_1+a_2+\ldots+a_n-1)!}{a_1!\ldots(a_k-1)!\ldots a_n!} = a_k\,\frac{(a_1+\ldots+a_n-1)!}{a_1!\ldots a_k!\ldots a_n!}\ ,$$

c'est-à-dire de le multiplier par $a_k/(a_1+\ldots+a_n)$.
Le calcul de $g(A_1)+g(A_2)+\ldots+g(A_n)$ fait ainsi apparaître le facteur commun $(a_1+\ldots+a_n-1)!/a_1!\ldots a_n!$; il se ramène donc à calculer la somme $\sum_{k=1}^{n} a_k P_k$, où P_k désigne ce que devient le produit

$$(9)\qquad P = \prod_{1\leqslant i<j\leqslant n}\frac{a_i-a_j}{a_i+a_j}$$

lorsqu'on y remplace a_k par a_k-1.

Pour montrer que la somme finale est bien celle qu'indique la formule (7), il suffira de montrer que

$$(10)\qquad a_1P_1+a_2P_2+\ldots+a_nP_n = (a_1+a_2+\ldots+a_n)\,P\ .$$

Or (10) se présente comme une identité entre fractions rationnelles de n variables, que rien n'empêche de continuer à appeler $a_1\ a_2\ \dots\ a_n$. L'établissement de cette identité exige une organisation soigneuse du calcul algébrique, laquelle est présentée dans le § suivant.

Bien entendu un tel calcul n'est pas nécessaire si l'on n'a besoin d'établir l'identité que pour une très petite valeur de n, par exemple n=3. Elle s'écrit en effet alors

$$a\,\frac{(a-1-b)(a-1-c)(b-c)}{(a-1+b)(a-1+c)(b+c)} + b\,\frac{(a-b+1)(a-c)(b-1-c)}{(a+b-1)(a+c)(b-1+c)}$$

$$+\,c\,\frac{(a-b)(a-c+1)(b-c+1)}{(a+b)(a+c-1)(b+c-1)} = (a+b+c)\frac{(a-b)(a-c)(b-c)}{(a+b)(a+c)(b+c)}$$

Un calcul direct (encore que déjà un peu lourd) permet alors de vérifier cette identité.

3. DEMONSTRATION GENERALE DE L'IDENTITE (10)

La fraction rationnelle P définie par (9) a pour dénominateur

$$\prod_{1\leq i<j\leq n} (a_i+a_j) .$$

Dans ce produit de n(n-1)/2 binômes a_i+a_j , il sera commode, en vue du calcul de P_k, de distinguer d'une part les (n-1)(n-2)/2 facteurs pour lesquels ni i ni j ne sont égaux à k, d'autre part les n-1 autres facteurs, qui donnent

$$(a_1+a_k)(a_2+a_k)\dots(a_{k-1}+a_k) \times (a_k+a_{k+1})(a_k+a_{k+2})\dots(a_k+a_n) .$$

Le remplacement de a_k par a_k-1 diminuera de 1 chacun de ces facteurs, et par conséquent le choix d'un dénominateur commun pour le calcul de la somme $a_1P_1+\dots+a_nP_n$ conduira à adopter

$$D = \prod_{1\leq i<j\leq n} (a_i+a_j)(a_i+a_j-1) .$$

Le numérateur de a_kP_k deviendra dans ces conditions, comme on s'en assure facilement,

$$(11)\quad N_k = a_k \prod_{\substack{i\neq k\\ j\neq k\\ 1\leq i<j\leq n}} (a_i-a_j)(a_i+a_j-1) \prod_{i\leq i<k} (a_i-a_k+1)(a_i+k)$$

$$\cdot \prod_{k<j\leq n} (a_k-a_j-1)(a_k+a_j)$$

N_k est un polynôme de degré n^2-n+1 des n variables $a_1\ \dots\ a_n$; il en est donc de même de la somme $N_1 + N_2 + \dots + N_n$. Pour achever d'établir (10), il suffit de s'assurer

(1°) que cette somme est divisible par chacune des $n(n-1)/2$ différences a_i-a_j, ainsi que par chacun des $n(n-1)/2$ facteurs a_i+a_j-1 du dénominateur D

(2°) qu'après toutes ces divisions, qui reviennent à diviser par un polynôme de degré n^2-n, le quotient , nécessairement de degré 1 , est simplement égal à $a_1+a_2+\dots+a_n$.

L'étape (1°) se parcourt en fixant i et j et en constatant d'abord que les n-2 termes N_k pour lesquels k n'est égal ni à i ni à j sont bien divisibles chacun par a_i+a_j-1 et par a_i-a_j ; il suffit donc de vérifier ces divisibilités pour la somme N_i+N_j des deux termes restants.

La formule (11) conduit à écrire (en changeant les indices muets)

$$N_i = a_i \cdot \prod_{\substack{i\leqslant\lambda<\mu\leqslant n\\ \lambda\neq i\\ \mu\neq i}} (a_\lambda-a_\mu)(a_\lambda+a_\mu-1) \prod_{1\leqslant\lambda<i} (a_\lambda-a_i)(a_\lambda+a_i)$$

$$\cdot \prod_{1<\mu\leqslant n} (a_i-a_\mu-1)(a_i+a_\mu)$$

$$= a_i\ T_i\ U_i\ V_i$$

et de même

$$N_j = a_j\ T_j\ U_j\ V_j \quad .$$

En ce qui concerne T_i et T_j, on peut remarquer que, tant que ni λ ni μ ne sont égaux ni à i ni à j, les facteurs correspondants sont présents à la fois dans T_i et T_j et jouent le rôle de coefficients constants quand a_i et a_j varient seuls. Il suffit donc de ne retenir dans T_i et T_j que les facteurs respectifs T'_i et T'_j introduisant l'une de ces variables, et de travailler sur

$$N'_i = a_i\ T'_i\ U_i\ V_i \quad \text{et} \quad N'_j = a_j\ T'_j\ U_j\ V_j$$

Or on peut écrire

$$T'_i = \prod_{\alpha<i} (a_\alpha-a_j)(a_\alpha+a_j-1) \ . \prod_{i<\beta<j} (a_\beta-a_j)(a_\beta+a_j-1)$$

$$\cdot \prod_{\gamma>j} (a_j-a_\gamma)(a_j+a_\gamma-1) \qquad ,$$

écrire U_i sous la forme $U_i = \prod_{\alpha<i} (a_\alpha-a_i+1)(a_\alpha+a_i)$,

et enfin décomposer V_i en trois parties en écrivant

$$V_i = \prod_{i<\beta<j} (a_i-a_\beta-1)(a_i+a_\beta\ ) \ . \ [(a_i-a_j-1)(a_i+a_j)]$$

$$\cdot \prod_{\gamma>j} (a_i-a_\gamma-1)(a_i+a_\gamma)$$

On écrit ensuite, de manière analogue :

$$T'_j = \prod_{\alpha < i}(a_\alpha - a_i)(a_\alpha + a_i - 1) \cdot \prod_{i < \beta < j}(a_i - a_\beta)(a_i + a_\beta - 1)$$

$$\cdot \prod_{\gamma > j}(a_i - a_\gamma)(a_i + a_\gamma - 1)$$

$$U_j = \prod_{\alpha < i}(a_\alpha - a_j + 1)(a_\alpha + a_j) \cdot [(a_i - a_j + 1)(a_i + a_j)]$$

$$\cdot \prod_{i < \beta < i}(a_\beta - a_j + 1)(a_\beta + a_j)$$

$$V_j = \prod_{\gamma > j}(a_j - a_\gamma - 1)(a_j + a_\gamma) \; .$$

Regardons d'abord les facteurs qui, dans $a_i\ T'_i\ U_i\ V_i$ ou dans $a_j\ T'_j\ U_j\ V_j$, sont groupés sous des symboles Π.

Si $a_j = a_i$, les lers Π de T'_i et T'_j deviennent identiques ; de même si $a_j = 1 - a_i$; même conclusion pour les 3èmes Π de T'_i et T'_j . Pour les deuxièmes Π , les facteurs deviennent opposés, ce qui introduit j-i-1 changements de signe. U_i devient identique au premier Π de U_j. Les facteurs du premier Π de V_i deviennent opposés à ceux du dernier Π de U_j, ce qui fait de nouveau j-i-1 changements de signe ; enfin le troisième Π de V_i devient identique à V_j . Les six symboles Π de chacun des deux produits N'_i et N'_j représentent donc, chaque fois que $a_j = a_i$ ou $a_j = 1 - a_i$, des facteurs identiques.

Restent à prendre en compte les facteurs non groupés sous des Π , c'est-à-dire les facteurs initiaux et les facteurs entre crochets, à savoir

pour N'_i : $\quad a_i[(a_i - a_j - 1)(a_i + a_j)]$

pour N'_j : $\quad a_j[(a_i - a_j + 1)(a_i + a_j)]$.

Ceux-là deviennent manifestement opposés si l'on remplace a_j soit par a_i, soit par $1 - a_i$.

En fin de compte on a bien prouvé la divisibilité par $(a_i - a_j)(a_i + a_j - 1)$ de $N'_i + N'_j$, donc de $N_i + N_j$, donc de $N_1 + N_2 + \ldots + N_n$.

L'étape (2°) est beaucoup plus simple : en effet, pour montrer que le quotient de $N_1 + \ldots + N_n$ par $\Pi(a_i - a_j)(a_i + a_j - 1)$ est bien égal à $a_1 + \ldots + a_n$, il suffit de raisonner comme suit :

(a) ce quotient, par suite du caractère alterné (symétrique gauche) de P, est nécessairement symétrique, c'est-à-dire de la forme $C_n(a_1 + \ldots + a_n) + D_n$.

(b) Les systèmes de valeurs particuliers tels que $a_i = a + 1 - i$, qui ont l'avantage d'annuler tous les P_k sauf P_n, permettent de s'assurer aisément que $C_n = 1$ et $D_n = 0$.

La formule (7) se trouve ainsi définitivement établie, et par conséquent aussi la formule (5).

4. CAS PARTICULIERS ET PROBLEMES OUVERTS

Nous avons rappelé plus haut la signification de $f(a_1,a_2,\ldots,a_n)$ comme nombre d'ordres de construction acceptables d'un muret de briques. "Acceptables" voulait dire, puisqu'il s'agissait de murets "adossés" à l'axe des y dans sa disposition habituelle, qu'aucune brique ne pouvait être mise en place avant sa voisine du dessous ni avant sa voisine de gauche ; cette dernière exigence ne valait évidemment pas pour la colonnette élevée le long même de l'axe des y.

La fonction $g(a_1,a_2,\ldots,a_n)$ peut manifestement recevoir une signification analogue, à condition d'adosser le muret non plus à l'axe des y mais à la ligne brisée représentée sur la figure 3, et de définir toujours le profil à atteindre par les nombres $a_1\ a_2 \ldots a_n$ de briques des couches successives considérées de bas en haut (nous dirons "muret à profil brisé").

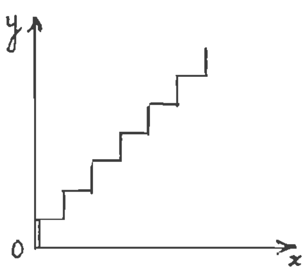

Fig. 3

Il est intéressant de considérer les cas particuliers où les a_i sont en dégression arithmétique, notamment si celle-ci est de raison 2 (profil symétrique) ou de raison 1 (profil à pic à droite).

Dans le premier cas on peut poser

$$a_i = a-2(i-1) ,$$

avec $a \geqslant 2(n-1)$ et appeler h la moyenne arithmétique $a-n+1$ de cette dégression.

On constate alors que le calcul de $g(a_1,a_2,\ldots,a_n)$ par la formule (7) donne pour résultat, après les simplifications et groupements que permet la forme particulière des a_i ,

$$g(a,a-2,\ldots,a-2n+2)$$

$$= \frac{(nh)!1!2!\ldots(n-1)!}{h!(h+1)!(h+2)!\ldots(h+n-1)!}$$

Or cette dernière expression est exactement la même que celle qui résulterait de l'application de la formule de Young (4) à une suite formée de n termes égaux à h .

Il en découle le curieux résultat suivant : tout muret à profil brisé symétrique a exactement le même nombre d'ordres de construction acceptables que le muret rectangulaire ayant même hauteur n et même nombre de briques nh . Les figures 4 et 5 montrent deux tels murets (n=3, h=5), qui ont l'un et l'autre 6006 ordres de construction acceptables.

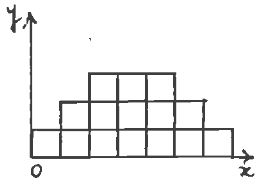

Fig. 4

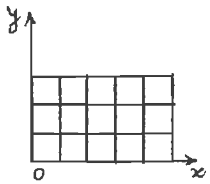

Fig. 5

Il serait intéressant de pouvoir établir ce résultat par une méthode bijective.

Dans le cas où les a_i sont en dégression arithmétique de raison 1, un autre résultat curieux apparaît si l'on pose $a_i = n-i+1$, $\forall\, i \in \{1,2,\dots,n\}$.

On constate alors (et l'on prouve facilement, en utilisant au besoin dans le calcul une récurrence sur n) que

$$g(n,n-1,\dots,2,1) = f(n,n-1,\dots,2,1).2^{-n(n-1)/2} .$$

Cela signifie par exemple, pour n=5, que le muret de la fig. 7 a $2^{5.4/2}$ (=1024) fois plus d'ordres de construction acceptables que celui de la fig. 6 : ici 286 pour l'un et 292864 pour l'autre.

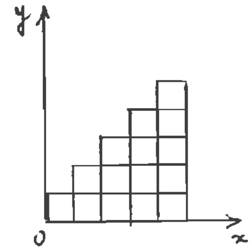

Fig.6

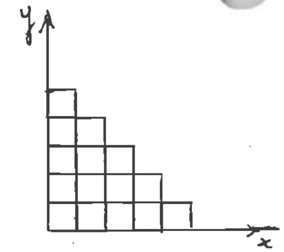

Fig.7

D'où à nouveau un problème combinatoire ouvert : trouver une partition "naturelle" des 292864 ordres de droite soit en 1024 classes de 286 ordres, soit en 286 classes de 1024 ordres. Les valeurs numériques correspondant aux premières valeurs de n sont

	n:	1	2	3	4	5	6
A3121	à gauche :	1	1	2	12	286	33592
A5118 →	à droite :	1	2	16	768	292864	1100742656

Bibliographie

[1] ANDRE D., Solution directe du problème résolu par M. Bertrand, C.R.Ac.Sc., Paris, vol. 105 (1887),436.

[2] BERTRAND J., Solution d'un problème, C.R.Ac.Sc., Paris, Vol. 105 (1887), 369.

[3] FRAME J.S., ROBINSON G. de B., THRALL R.M., The hook graphs of the symmetric group, Canadian J. of Math., vol. 6 (1954), 316-323.

[4] KREWERAS G., Cahiers du B.U.R.O., vol. 6 (1965).

[5] MACMAHON P.A., Combinatory Analysis, Cambridge (1915) ; réimpr. Chelsea, New-York (1950).

[6] SYLVESTER J.J., Collected Mathematical Papers, vol. IV (1912).

[7] YOUNG A., On quantitative substitutional analysis, Proc. London Math. Soc., 2, Vol. 28 (1927),255-292.

Reçu en Mai 1981