Nouveaux progrès dans l'énumération des modèles mixtes

R. Bayon* N. Lygeros† J.-S. Sereni‡

Résumé : *Un modèle mixte est dit contraint si aucun de ses facteurs fixes n'est emboîté dans un facteur aléatoire. Hess et Iyer se sont intéressés au problème de l'énumération des modèles mixtes (avec ou sans contrainte), à isomorphie près. Dans cet article, en nous basant sur les résultats dans l'énumération des posets nous présentons un nouvel algorithme et obtenons des résultats pour les modèles mixtes de taille au plus 9.*

mots-clefs : énumération, modèles mixtes, poset, proset .

1 Introduction

Le nombre de modèles ANOVA à effets fixes deux à deux non isomorphes est connu pour $n \leq 16$ [1], puisque ceux-ci sont en correspondance bijective avec les posets. Hess et Iyer se sont intéressés au problème de l'énumération des modèles mixtes (avec ou sans contrainte), à isomorphie près. Un modèle mixte est dit contraint si aucun de ses facteurs fixes n'est emboîté dans un facteur aléatoire. Ils ont obtenu des résultats pour les modèles mixtes de taille au plus 5 [7] en prouvant qu'il y a 576 modèles ANOVA à effets fixes deux à deux non isomorphes.

Les modèles mixtes sont des cas particuliers des modèles linéaires généraux pour lesquels les fonctions moyennes et les matrices de covariance des données ont une structure linéaire. Plus précisément les vecteurs d'observations $\mathbf{y}$ ont des moments statistiques qui satisfont :

$$E[\mathbf{y}] = \mathbf{X}\alpha$$

$$Var[\mathbf{y}] = \sum_t \phi_t \mathbf{V_t}$$

ou $\alpha = (\alpha_1, \ldots, \alpha_p)^t$ et $\phi = (\phi_1, \ldots, \phi_q)^t$ sont des parametres inconnus et $\mathbf{X}, V_1, \ldots, V_q$ sont des matrices connues. Tout vecteur moyenne et toute matrice de covariance peuvent s'écrire sous cette forme, mais nous nous intéressons aux modèles pour lesquels les paramètres $\alpha_1, \ldots, \alpha_p, \phi_1, \ldots, \phi_q$ sont fonctionnellement indépendants. De telles structures générales de covariances apparaissent lorsque l'on considère que certains des effets sont aléatoires.

Hess et Iyer ont défini une macro SAS qui permet de déterminer des intervalles de confiance afin de faciliter les calculs de la méthode de Satterthwaite via l'implémentation de Burdick et Graybill [2]. Le domaine de validité de leur macro est réduit aux modèles ayant au plus 5 facteurs. Notre approche qui se base sur les travaux réalisés sur les posets [6, 3, 4], permet actuellement d'étendre leurs résultats aux modèles ayant au plus 9 facteurs.

2 Propositions.

Proposition 2.1 (Hess, Iyer) *Il existe une correspondance bijective entre les modèles à effets fixes et les posets* $(P, <)$.

Preuve: Considérons un poset P à n éléments et un modèle à effets fixes F avec n facteurs. Les éléments $i \neq j$ du poset P tels que $j \leq i$ correspondent aux facteurs j emboîtés dans les facteurs i du modèle à effets fixes F. Les éléments incomparables de P correspondent aux facteurs croisés de F. □

*ENSPS, Parc d'Innovation - Bld S. Brant, F 67400 Illkirch Mail : roman.bayon@ensps.u-strasbg.fr

†Institut Girard Desargues, Département de Mathématiques, Université Lyon 1, 43 Bld du 11 Nov. 1918, 69622 Villeurbanne Cedex Mail : lygeros@igd.univ-lyon1.fr

‡Laboratoire d'Informatique de Marseille, Faculté des Sciences de Luminy, Université de la Méditerranée, Marseille. F-13288, Marseille Cedex 9, FRANCE Mail : jean-sebastien.sereni@dil.univ-mrs.fr

Puisque il y a correspondance entre les posets et les modèles mixtes à effets fixes, l'énumération de ceux-ci à isomorphie près est connue pour $n \leq 16$ [1].

Pour les modèles mixtes, nous introduisons la notion de proset : Partially Reflexive Ordered SET.

Definition 2.2 *Un proset est un ensemble muni d'une relation binaire anti-symétrique et transitive.*

Remarque 2.3 *: Nous pouvons voir un proset P_r comme un poset P pour lequel la relation d'ordre est modifiée, de sorte qu'elle puisse être réflexive pour certains éléments de P seulement. En d'autres termes, certains termes de la diagonale de la matrice d'incidence d'un proset P_r, peuvent être non nuls. Nous dirons dans ce cas que le proset P_r induit le poset P.*

Proposition 2.4 *Il existe une correspondance bijective entre les modèles mixtes (sans contrainte) et les prosets.*

Preuve: Considérons un proset P à n éléments et un modèle mixte (sans contrainte) M à n facteurs. Pour des éléments distincts du proset, la situation est la même que précédemment. Un élément réflexif du proset correspond à un facteur aléatoire de M, alors qu'un élément non réflexif correspond à un facteur fixe de M. $\square$

Nous disposons d'une notion naturelle de dualité pour les prosets.

Definition 2.5 *Un proset P' est le dual du proset P si, et seulement si :*
- *P et P' induisent le meme poset.*
- *Les relations de réflexivité de P' sont complémentaires de celles de P.*

En termes de matrices d'incidence, P' est construit en inversant les 0 et les 1 de la diagonale de la matrice d'incidence de P.

Remarque 2.6 *: Soit n un entier impair. La dualité nous assure que le nombre de modèles mixtes sans contrainte de taille n est pair.*

Hess et Iyer sont parvenus à énumérer, à isomorphie près, les modèles mixtes avec et sans contrainte de taille au plus 5. Notre algorithme nous a permis d'étendre ce résultat pour $n \leq 9$.

3 Enumération des modèles mixtes

Nous représentons un proset $M = (x_1, x_2, ..., x_n, \prec)$ par sa matrice d'incidence $(m_{ij}) \in \{0,1\}^{n^2}$ avec $m_{ij} = 1$ si, et seulement si, $x_i \prec x_j$. Si cette représentation est idéale informatiquement, elle conduit à indicer les sommets et force ainsi à les ordonner. Notre algorithme consiste donc à restreindre les permutations parmi toutes les permutations possibles pour retrouver l'unicité de la représentation des prosets.

Algorithme de Hess et Iyer :
1. Génération des posets de taille n.
2. Pour chaque poset, génération de 2^n matrices A, via toutes les combinaisons possibles de 0 et 1 sur la diagonale de la matrice.
3. Si ce sont les modèles mixtes contraints qui sont désirés alors suppression de toutes les matrices A qui violent la contrainte, en calculant $A' = \sum_{i=0}^{i=n-1} A^k$ et en testant si $A'(i,j) \geq 1$ pour tout couple (i,j) tel que $A(i,i) = 1$ et $A(j,j) = 0$.
4. Suppression de tous les modèles isomorphes.

Notre algorithme :
1. Génération des posets de taille n sous forme canonique, avec l'algorithme de Chaunier et Lygeros [3, 4].
2. Pour chaque poset, génération des 2^{n-1} matrices d'incidence A obtenues en placant les 0 et les 1 sur la diagonale de sorte que le nombre de 1 soit inferieur a $n/2$. Ainsi nous générons seulement un seul proset par couple de prosets duaux, ce qui réduit le nombre de tests d'isomorphie.

3. Suppression de tous les modèles isomorphes.

4. Incrémentation de 2 du nombre de modèles mixtes sans contrainte. Puis nous testons pour le modèle mixte, et pour son dual si la contrainte est vérifiée.

Remarque 3.1 *: Pour vérifier si un proset valide ou non la contrainte, il suffit de tester si un élément réflexif est adjacent à un élément non réflexif. Comme notre algorithme génère des prosets sous forme canonique (via l'algorithme de génération de posets), le test de la contrainte est élémentaire.*

4 Résultats et interprétation.

Nous rappelons tout d'abord les résultats connus sur l'énumeration des posets à isomorphie près, puis donnons les résultats obtenus pour les modèles mixtes.

Facteurs	Modèles Fixes	Nom
$n = 1$	1	
$n = 2$	2	
$n = 3$	5	
$n = 4$	16	
$n = 5$	63	
$n = 6$	318	
$n = 7$	2.045	Wright 1972
$n = 8$	16.999	Das 1977
$n = 9$	183.231	Mohring 1984
$n = 10$	2.567.284	Culberson, Rawlins 1990
$n = 11$	46.749.427	Culberson, Rawlins 1990
$n = 12$	1.104.891.746	C. Chaunier, N. Lygeros 1991
$n = 13$	33.823.827.452	C. Chaunier, N. Lygeros 1994
$n = 14$	1.338.193.159.771	N. Lygeros, P. Zimmermann 2000
$n = 15$	68.275.077.901.156	G. Brinkmann, B. D. McKay 2002
$n = 16$	44.831.306.651.195.087	G. Brinkmann, B. D. McKay 2002

Facteurs	Modèles Mixtes (Non Contraints)	Nom
$n = 1$	2	A. Hess, H. Iyer 1999
$n = 2$	7	A. Hess, H. Iyer 1999
$n = 3$	32	A. Hess, H. Iyer 1999
$n = 4$	192	A. Hess, H. Iyer 1999
$n = 5$	1.490	A. Hess, H. Iyer 1999
$n = 6$	15.067	R.Bayon, N. Lygeros, J.-S. Sereni 2002
$n = 7$	198.296	R.Bayon, N. Lygeros, J.-S. Sereni 2002
$n = 8$	3.398.105	R.Bayon, N. Lygeros, J.-S. Sereni 2002
$n = 9$	75.784.592	R.Bayon, N. Lygeros, J.-S. Sereni 2002

Facteurs	Modèles Mixtes (Contraints)	Nom
$n = 1$	2	A. Hess, H. Iyer 1999
$n = 2$	6	A. Hess, H. Iyer 1999
$n = 3$	22	A. Hess, H. Iyer 1999
$n = 4$	101	A. Hess, H. Iyer 1999
$n = 5$	576	A. Hess, H. Iyer 1999
$n = 6$	4.162	R.Bayon, N. Lygeros, J.-S. Sereni 2002
$n = 7$	38.280	R.Bayon, N. Lygeros, J.-S. Sereni 2002
$n = 8$	451.411	R.Bayon, N. Lygeros, J.-S. Sereni 2002
$n = 9$	6.847.662	R.Bayon, N. Lygeros, J.-S. Sereni 2002

Interprétation du nombre de modèles ANOVA à effets fixes avec n éléments, via les ordres des groupes d'automorphismes des posets associés [5, 8]. $N_{P_m} \times K$ signifie qu'il y a M posets de groupe d'automorphismes d'ordre m générant chacun K modèles mixtes (sans contrainte).

$M_1 = 2 = 1_{P_1} \times 2$

$M_2 = 7 = 1_{P_1} \times 4 + 1_{P_2} \times 3$

$M_3 = 32 = 2_{P_1} \times 8 + 2_{P_2} \times 6 + 1_{P_6} \times 4$

$$M_4 = 192 = 5_{P_1} \times 16 + (6_{P_2} \times 12 + 1_{P_2} \times 10) + 1_{P_4} \times 9 + 2_{P_6} \times 8 + 1_{P_{24}} \times 5$$

Nous voyons ici que l'ordre du groupe d'automorphisme n'est pas forcément suffisant pour déterminer le nombre de modèles mixtes engendrés. Plus précisément :

Soit un poset $(P, <)$ de taille n et de groupe d'automorphisme G d'ordre a, alors :

(i) si $a = 1$, P engendre 2^n modèles mixtes (sans contrainte).

(ii) si $a = n!$, P engendre $n + 1$ modèles mixtes (sans contrainte).

(iii) si l'on connaît la composition du groupe d'automorphisme G, on peut alors donner le nombre de modèles mixtes (sans contrainte) engendrés par P à l'aide du théorème de Pólya. Ce nombre est :

$$\mathcal{Z}_G(2) = \frac{1}{a} \sum_{g \in G} 2^{\lambda_1(g) + \lambda_2(g) + \ldots + \lambda_n(g)}$$

où $\lambda_i(g)$ est le nombre de cycles de longueur i dans la décomposition de g en cycles disjoints.

Enfin, le résultat de Prömel [9] sur les posets rigides à savoir que presque tous les posets ont un groupe d'automorphismes trivial, nous permet d'affirmer explicitement le caractère efficace de notre méthode. En effet par un transport de structure, il est facile de voir que presque tous les prosets sont rigides et que proportionellement les prosets ont une densité effective supérieure pour un n donné.

Références

[1] G. Brinkmann and B. D. McKay. *Posets on up to 16 Points.* Order 19(2) : p. 147-179 ; Jan 2002.

[2] R. K. Burdick and F. A. Graybill. *Confidence intervals on variance components.* Marcel Dekker, Inc. New-York ; 1992.

[3] C. Chaunier and N. Lygeros. *The number of orders with thirteen elements.* Order, vol.9 : p. 203-204 ; 1992.

[4] C. Chaunier and N. Lygeros. *Le nombre de posets à isomorphie près ayant 12 éléments.* Theoretical Computer Science, n123 : p. 89-94 ; février 1994.

[5] C. Chaunier and N. Lygeros. *Posets minimaux ayant un groupe d'automorphismes d'ordre premier.* C.R.Acad.Sci.Paris, t.318, s.I : p. 695-698 ; 1994.

[6] R. Fraïssé and N. Lygeros. *Petits posets : dénombrement, représentabilité par cercles et compenseur.* C.R.Acad.Sci.Paris, t.313, s.I : p. 417-420 ; septembre 1991.

[7] A. Hess and H. Iyer. *Enumeration of mixed linear models and SAS macro for computation of confidence intervels for variance components.* Applied Statistics in Agriculture Conference at Kansas State University 2001.

[8] N. Lygeros and M. Mizony. *Construction de posets dont le groupe d'automorphismes est isomorphe à un groupe donné* . C.R.Acad.Sci.Paris, t.322, s.I : p. 203-206 ; 1996.

[9] H. J. Prömel. *Counting unlabeled structures.* . J. Comb. Theory, série A, 44 : p. 83-93 ; 1987.