

Schémas de sélection : de la représentation généalogique au modèle statistique. Justification asymptotique

B Goffinet, B Mangin

*Institut national de la recherche agronomique,
laboratoire de Biométrie et Intelligence artificielle de Toulouse,
31320 Castanet-Tolosan, France*

(Reçu le 23 novembre 1992, accepté le 1^{er} mars 1993)

Résumé – Le but de cet article est de démontrer qu'avec des hypothèses simples de tirage aléatoire d'individus dans des populations on peut donner une justification asymptotique au modèle décrit dans Mangin et Vincourt (1992). La démonstration donnée ici généralise les démonstrations que l'on peut trouver dans la littérature pour des modèles particuliers. Pour aider à la compréhension des notations et démonstrations, les aspects qui seront étudiés dans ce papier seront tout d'abord décrits dans le cadre d'un exemple puis généralisés. On décrit d'abord le modèle à effets fixes engendré par les règles de Mangin et Vincourt (1992), et on précise les conditions supplémentaires naturelles qui ne diminuent pas la dimension de l'espace des paramètres estimables. On précise alors les propriétés des variables aléatoires de tirage dans chacune des populations et la forme explicite des effets aléatoires ainsi engendrés. Enfin, on montre que les règles de Mangin et Vincourt (1992) concernant les covariances et variances des différents niveaux des effets aléatoires sont exactes, soit à distance finie, soit asymptotiquement.

justification asymptotique / modèle aléatoire / croisement / analyse de variance

Summary – Selection schemes : from the genealogical representation to the statistical model. Asymptotic validity. The aim of this paper is to give a proof of the asymptotic validity of the model described by Mangin and Vincourt (1992). This proof generalizes the kind of proof that can be found in the bibliography for particular models. To help the reader with the notation and demonstrations, all topics included in this paper are first described in the framework of an example, and then generalized. We begin with the study of the fixed effect model generated by the rules given by Mangin and Vincourt (1992) and we include the additional conditions which maintain the rank of the estimate parameter space. We give the properties of the random variable used to sample in each population, and the exact structure of the random effects. Finally, we show that the rules given by Mangin and Vincourt (1992), which give the variances and covariances of the various levels of the random effects, are correct for a given dimension or asymptotically.

asymptotic justification / random model / cross / analysis of variance

INTRODUCTION

Le but de cet article est de démontrer qu'avec des hypothèses simples de tirage aléatoire d'individus dans des populations on peut donner une justification asymptotique du type de celle développée par Scheffé (1959), au modèle décrit dans Mangin et Vincourt (1992). Cet article sera abrégé par la suite par [Man].

On peut trouver dans la littérature des démonstrations du même type pour des modèles particuliers. Lefort (1977) a étudié le cas des modèles diallèles et Mallard *et al* (1983) ont traité de modèles hiérarchiques multidimensionnels. Il s'agit de généraliser ce type de démonstration à l'ensemble des modèles décrits dans [Man].

Il s'agit ici de donner une démonstration générale pour un vaste ensemble de modèles. Nous proposons, pour rendre plus accessibles les notations utilisées, de reprendre l'exemple utilisé dans [Man]. L'ensemble des aspects qui seront évoqués dans cet article sera tout d'abord décrit dans le cadre de cet exemple puis généralisé.

Nous étudierons successivement le modèle à effets fixes et nous préciserons les conditions supplémentaires naturelles qui ne diminuent pas la dimension de l'espace des paramètres estimables.

Nous préciserons alors les propriétés des variables aléatoires de tirage dans chacune des populations et la forme explicite des effets aléatoires ainsi engendrés.

Enfin nous montrerons que les règles de [Man] donnant les variances et covariances entre les différents niveaux des effets aléatoires sont exactes, soit à distance finie, soit asymptotiquement.

MODÈLE À EFFETS FIXES

Notations

Nous allons décrire l'ensemble des individus à tous les niveaux du schéma par les 2 systèmes de notations N_1 et N_2 . Le système N_1 permet de faire les démonstrations des résultats présentés dans cet article, et le système N_2 permet de faire le lien avec les concepts utilisés dans [Man].

L'exemple

Nous utilisons ici l'exemple donné en figure 1. Il s'agit d'une partie d'un exemple utilisé dans [Man].

Les individus de la population de départ c_1 seront indicés par u_1 dans la notation N_1 et par δ_1 dans la notation N_2 . Chacun des individus u_1 est autofécondé et produit une descendance notée $c_2(u_1)$, et l'ensemble de ces descendance est le « type de descendance » c_2 . Les individus de ce type de descendance sont indicés en u_1u_2 dans la notation N_1 et δ_2 dans la notation N_2 . Deux individus u_1u_2 et $u_1u'_2$ de la même descendance $c_2(u_1)$ sont croisés et produisent des individus notés $u_1u_2u'_2u_3$ et δ_3 , qui forment la descendance $c_3(u_1, u_2, u'_2)$ appartenant au type de descendance c_3 . Le modèle adopté pour décrire la valeur génétique $g_{u_1u_2u'_2u_3}$ d'un individu de c_3 s'écrit :

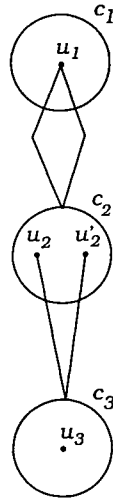


Fig 1. Schéma associé au protocole utilisé dans l'exemple.

$$g_{u_1 u_2 u'_2 u_3} = \mu + a_{u_1}^1 + a_{u_1 u_2}^2 + a_{u_1 u'_2}^3 + a_{u_1 u_2 u'_2}^4 + a_{u_1 u_2 u'_2 u_3}^5$$

dans la notation N_1 et :

$$g_{\delta_3} = \mu + a_{\delta_1}^1 + a_{\delta_2}^2 + a_{\delta'_2}^3 + a_{\delta_2 \delta'_2}^4 + a_{\delta_3}^5$$

dans la notation N_2 .

Nous définirons les paramètres de cette décomposition de g à l'aide des conditions supplémentaires explicitées au paragraphe «Le modèle complet et les conditions supplémentaires».

Notation dans le cas général

De manière générale, le schéma qui conduit à la création des individus que l'on mesure peut se décrire à partir des ancêtres les plus anciens regroupés dans plusieurs populations de départ c_1, c_2, \dots, c_p . Les individus de ces populations de départ sont indicés par u_1, u_2, \dots, u_p dans la notation N_1 et par $\delta_1, \delta_2, \dots, \delta_p$ dans la notation N_2 . Ces individus sont ensuite autofécondés et/ou croisés, soit à l'intérieur d'une population, soit entre 2 populations. Chacun de ces «croisements» produit une descendance, et l'ensemble des descendance produites par le même type de croisement est regroupé dans ce que nous appelons un type de descendance.

L'ensemble du protocole de croisement est réalisé génération après génération pour aboutir enfin au type de descendance c_K dont les individus sont mesurés. Notons que si l'on a des observations dans plusieurs types de descendance, il y aura autant de modèles que de types de descendance pour lesquels on a des observations.

La valeur génétique d'un individu de c_k sera notée :

$$\begin{aligned} g_{\underline{u}_k}^{[u_k]} & \text{ dans la notation } N_1 \\ g_{\delta_k} & \text{ dans la notation } N_2 \end{aligned}$$

où \underline{u}_k désigne un indice dont les niveaux sont notés u_k et $\underline{\mathcal{U}}$ représente l'ensemble de tous les indices \underline{u}_k permettant de repérer chacun des individus et ses ancêtres. La notation $\underline{u}_k \in \underline{\mathcal{U}}$ désigne la concaténation des niveaux pris par ces indices, concaténation traduisant les protocoles de croisement.

Par la suite, les caractères désignant des indices ou des ensembles d'indices seront soulignés, tandis que ceux désignant des niveaux d'indices ou des ensembles de niveaux d'indices ne le seront pas.

La valeur génétique d'un individu est décomposée en la somme de plusieurs effets notés :

$$\begin{aligned} a_{\underline{u}_k}^l & \text{ dans la notation } N_1 \\ a_{\underline{\delta}_k}^l & \text{ dans la notation } N_2 \end{aligned}$$

où $\underline{\mathcal{U}}^l$ représente l'ensemble de tous les indices caractérisant cet effet dans la notation N_1 , $\underline{\mathcal{D}}^l$ ces indices dans la notation N_2 . Par la suite on désignera aussi cet effet par la notation simplifiée a^l .

Considérons, pour illustrer notre propos, 2 populations de départ c_1 et c_2 constituant la première génération et le schéma suivant :

1) croisement d'un individu de c_1 avec un individu de c_2 ; les individus obtenus constituant la deuxième génération c_3 ;

2) croisement entre eux des individus de cette deuxième génération; les individus obtenus constituant la troisième génération c_4 .

Par exemple, le croisement entre l'individu u_1 de la population c_1 et l'individu u_2 de la population c_2 , produit la descendance $c_2(u_1, u_2)$, et ces descendance sont regroupées dans le type de descendance c_3 . Dans ce type de descendance, un individu est indicé avec la notation N_1 , par $u_1 u_2 u_3$, où u_3 désigne son numéro à l'intérieur de sa descendance, et u_1, u_2 les numéros de ses ancêtres, et par δ_3 , avec la notation N_2 , qui décrit complètement cet individu à l'intérieur de son type de descendance.

Lorsque l'on croise des individus de c_3 avec d'autres individus de c_3 , il faut identifier 4 types de descendance suivant le nombre de parents communs des individus croisés : celui issu des croisements des individus de $c_3(u_1, u_2)$ avec des individus de $c_3(u_1, u_2)$, des croisements de $c_3(u_1, u_2)$ avec $c_3(u_1, u'_2)$, des croisements de $c_3(u_1, u_2)$ avec $c_3(u'_1, u_2)$, et enfin des croisements de $c_3(u_1, u_2)$ avec $c_3(u'_1, u'_2)$. Le plus souvent, un seul de ces types de descendance sera représenté. Par exemple, le croisement d'un individu u_3 de $c_3(u_1, u_2)$ avec un individu u'_3 de $c_3(u_1, u'_2)$ produira la descendance $c_4(u_1, u_2, u'_2, u_3, u'_3)$ dans laquelle un individu sera indicé par u_4 .

$$\underline{u}_k \in \underline{\mathcal{U}} = u_1, u_2, u'_2, u_3, u'_3, u_4$$

Le modèle complet et les conditions supplémentaires

Il s'agit maintenant de préciser quels sont les effets intervenant dans la décomposition de la valeur génétique d'un individu. Pour qu'un effet exprimé avec la notation N_1 ait un sens, il est nécessaire que si un indice \underline{u}_k est présent, alors tous les indices correspondant aux ancêtres qui ont conduit au type de descendance c_k , soient présents. On dira alors que l'indice \underline{u}_k est hiérarchisé dans ces indices. Dans l'exemple, on ne peut avoir l'indice \underline{u}_2 sans l'indice \underline{u}_1 , l'indice \underline{u}_2 est donc hiérarchisé dans l'indice \underline{u}_1 . Nous allons considérer ici que tous les effets satisfaisant à cette condition sont effectivement présents dans le modèle.

Cette définition de l'existence d'un effet est équivalente à celle qui est donnée par l'ensemble des règles de [Man]. Nous montrons en annexe 1 qu'il existe une bijection entre un effet défini par la règle ci-dessus et les sous-graphes considérés dans [Man].

Nous dirons par la suite qu'un indice \underline{u}_k est «de plus basse hiérarchie» dans un effet a^l , s'il n'existe pas dans \mathcal{U}^l d'indices qui soient hiérarchisés dans \underline{u}_k .

Le modèle ainsi engendré par les règles qui viennent d'être décrites est complet car il contient un effet indicé par l'individu mesuré lui-même. Dans l'exemple, il s'agit de $a_{u_1 u_2 u'_2 u_3}^5$ dans la notation N_1 ou de $a_{\delta_3}^5$ dans la notation N_2 . On peut parler de modèle «vrai» en ce sens que l'on pourra toujours décomposer exactement les valeurs génétiques des individus mesurés en une somme des effets a^l . Mais il est surparamétré, c'est-à-dire que l'on ne peut pas estimer l'ensemble des paramètres, et il s'agit donc de définir des conditions supplémentaires pour rendre estimables les paramètres du modèle.

Ces conditions supplémentaires seront données dans la notation N_1 .

L'exemple

Les conditions supplémentaires choisies pour définir les paramètres de ce type de modèle sont simples et permettent de leur donner une interprétation facile :

$$\sum_{u_1} a_{u_1}^1 = 0; \quad \forall u_1, \sum_{u_2} a_{u_1 u_2}^2 = 0 \text{ et } \sum_{u'_2} a_{u_1 u'_2}^3 = 0$$

$$\forall u_1, u'_2, \sum_{u_2, u_2 \neq u'_2} a_{u_1 u_2 u'_2}^4 = 0 \text{ et } \forall u_1, u_2, \sum_{u'_2, u'_2 \neq u_2} a_{u_1 u_2 u'_2}^4 = 0$$

Cas général

La difficulté réside dans le fait qu'un effet peut être caractérisé par un nombre important d'indices dont les niveaux correspondent à des individus d'une même descendance d'un même type de descendance. Certains de ces indices peuvent être de plus basse hiérarchie, d'autres non. Soit \underline{u}_h un indice de plus basse hiérarchie pour un effet a^l . Soit $\underline{\mathcal{U}}_{s_h}^l = (\underline{u}'_h, \underline{u}''_h, \dots)$ l'ensemble des indices, présents dans a^l , dont les niveaux correspondent à des individus de la même descendance s_h que u_h .

Cet ensemble peut être vide. Les conditions supplémentaires s'écrivent :

$$\forall u_k; \underline{u}_k \in \{\underline{u}^l - \underline{u}_h\}, \sum_{\substack{u_h, u_h \notin \underline{u}^l \\ \underline{u}_k \in \underline{u}^l}} a^l_{\{u_k\}} = 0$$

Sous les conditions d'orthogonalité décrites ci-dessous, on montre dans l'annexe 2, que ces conditions supplémentaires rendent estimables les paramètres sans affecter le caractère complet du modèle.

Conditions d'orthogonalité : le nombre de descendants de tout individu ou de tout couple d'individus est le même quel que soit le niveau de l'indice ou des indices caractérisant ce ou ces individus dans leurs propres descendance.

Dans un croisement entre 2 types de descendance, le croisement concerne tous les individus des 2 types de descendance.

Remarque : ces conditions font référence aux individus avant les tirages aléatoires décrits dans le paragraphe suivant, et non pas aux individus effectivement mesurés après ce tirage, pour lesquels elles ne sont pas nécessaires.

Les variables aléatoires de tirage

Définition des variables aléatoires de tirage

De la même façon que dans Scheffé (1959), Lefort (1977) ou Mallard *et al* (1983), les individus qui sont utilisés à chaque niveau de la généalogie sont tirés par un tirage aléatoire sans remise dans la descendance à laquelle ils appartiennent.

La descendance dans laquelle un individu est indicé par u_k , est indicée par un ensemble de niveaux des indices hiérarchisant u_k . Pour ne pas alourdir la notation, on dira qu'un individu est indicé par u_k dans la descendance s_k . Le tirage aléatoire dans cette descendance se fera avec la variable aléatoire $X^{s_k}_{i_k, u_k}$, où i_k désigne le numéro du tirage dans s_k .

Chacune de ces variables est une indicatrice qui prend la valeur 1 si l'individu u_k est tiré au tirage i_k et 0 sinon avec équiprobabilité pour les différents u_k de la descendance s_k . Nous ferons l'hypothèse que l'ensemble des variables X^s est indépendant de l'ensemble des variables $X^{s'}$ si s et s' sont 2 descendance différentes, qu'elles soient ou non du même type de descendance. À l'intérieur d'une même descendance s_k , $\sum_{u_k} X^{s_k}_{i_k, u_k} = 1$ puisque l'ancêtre numéroté i_k est obligatoirement l'un des u_k ; les autres caractéristiques de la loi conjointe des X^{s_k} seront précisées au fur et à mesure de leur nécessité.

Dans notre exemple, un premier tirage est effectué dans la population de départ c_1 . Nous noterons $X^{c_1}_{i_1, u_1}$ les variables aléatoires indicatrices qui prennent la valeur 1 si l'individu u_1 est tiré au tirage numéro i_1 , et 0 sinon. Au deuxième niveau de la généalogie, il y a autant de descendance $c_2(u_1)$ de type c_2 que d'individus dans c_1 . Les variables aléatoires indicatrices sont alors notés $X^{c_2(u_1)}_{i_2, u_2}$. On définit de la même façon les indicatrices $X^{c_3(u_1, u_2, u'_2)}_{i_3, u_3}$.

L'hypothèse d'indépendance des indicatrices X^s et $X^{s'}$ de 2 descendance s et s' concerne par exemple $s = c_2(u_1)$ et $s' = c_2(u'_1)$, ou bien 2 descendance

n'appartenant pas au même type de descendance, par exemple $s = c_2(u_1)$ et $s' = c_2(u_1, u_2, u'_2)$. Par contre, bien évidemment, dans une même descendance si par exemple $X_{i_2, u_2}^{c_2(u_1)} = 1$, on a $X_{i'_2, u_2}^{c_2(u_1)} = 0$ pour tout $i'_2 \neq i_2$ puisque l'individu u_2 a été sorti au tirage i_2 .

Écriture des effets aléatoires du modèle

L'exemple

Quand on procède, à chaque étape de la création des individus, à des tirages aléatoires comme décrits au dessus, on obtient la valeur génétique des individus mesurés, indicés par les numéros de tirage

$$G_{i_1 i_2 i'_2 i_3} = \sum_{u_1} \sum_{u_2} \sum_{u'_2} \sum_{u_3} X_{i_1 u_1}^{c_1} X_{i_2 u_2}^{c_2(u_1)} X_{i'_2 u'_2}^{c_2(u_1)} X_{i_3 u_3}^{c_3(u_1, u_2, u'_2)} g_{u_1 u_2 u'_2 u_3}$$

On obtient aussi la décomposition de $G_{i_1 i_2 i'_2 i_3}$ comme somme des variables aléatoires.

$$\begin{aligned} A_{i_1}^1 &= \sum_{u_1} X_{i_1 u_1}^{c_1} a_{u_1}^1; \\ A_{i_1 i_2}^2 &= \sum_{u_1} \sum_{u_2} X_{i_1 u_1}^{c_1} X_{i_2 u_2}^{c_2(u_1)} a_{u_1 u_2}^2; \\ A_{i_1 i'_2}^3 &= \sum_{u_1} \sum_{u'_2} X_{i_1 u_1}^{c_1} X_{i'_2 u'_2}^{c_2(u_1)} a_{u_1 u'_2}^3; \\ A_{i_1 i_2 i'_2}^4 &= \sum_{u_1} \sum_{u_2} \sum_{u'_2} X_{i_1 u_1}^{c_1} X_{i_2 u_2}^{c_2(u_1)} X_{i'_2 u'_2}^{c_2(u_1)} a_{u_1 u_2 u'_2}^4; \\ A_{i_1 i_2 i'_2 i_3}^5 &= \sum_{u_1} \sum_{u_2} \sum_{u'_2} \sum_{u_3} X_{i_1 u_1}^{c_1} X_{i_2 u_2}^{c_2(u_1)} X_{i'_2 u'_2}^{c_2(u_1)} X_{i_3 u_3}^{c_3(u_1, u_2, u'_2)} a_{u_1 u_2 u'_2 u_3}^5 \end{aligned}$$

Dans la notation N_2 ces effets s'écrivent :

$$A_{\lambda_1}^1; A_{\lambda_2}^2; A_{\lambda'_2}^3; A_{\lambda_2 \lambda'_2}^4; A_{\lambda_3}^5$$

où λ_1 désigne l'individu sorti au tirage i_1 dans la population c_1 , λ_2 l'individu sorti au tirage i_2 dans la descendance $c_2(i_1)$, etc.

Cas général

Notons $\underline{\mathcal{I}}$ l'ensemble des indices correspondant aux indices de tirage des indices de \mathcal{U} . On notera de la même façon, $\underline{\mathcal{I}}^t$ celui correspondant à $\underline{\mathcal{U}}^t$.

On obtient en général :

$$G_{i_k \in \underline{\mathcal{I}}}^{[i_K]} = \sum_{u_k: \underline{u}_k \in \underline{\mathcal{U}}} \prod_{i_k \in \underline{\mathcal{I}}^t} X_{i_k, u_k}^{s_k} g_{\underline{u}_k \in \underline{\mathcal{U}}^t}^{[u_k]}$$

Cette notation représente une somme multiple pour tous les niveaux de tous les indices de \underline{u} . Cette somme est de plus une suite ordonnée de sommations de façon à ce que les indices hiérarchisant l'indice \underline{u}_k le précèdent toujours dans la somme.

La décomposition de $g_{\substack{[u_k] \\ \underline{u}_k \in \underline{u}}}$ conduit à décomposer $G_{\substack{[i_k] \\ i_k \in \underline{i}}}$ en une somme d'effets aléatoires qui s'écrivent :

$$A_{\substack{[i_k] \\ i_k \in \underline{i}^l}}^l = \sum_{u_k: \underline{u}_k \in \underline{u}^l} \prod_{i_k \in \underline{i}^l} X_{i_k, u_k}^{s_k} a_{\substack{[u_k] \\ \underline{u}_k \in \underline{u}^l}}^l$$

dans la notation N_1 , et en utilisant la notation N_2 :

$$A_{\substack{[\lambda_l] \\ \lambda_k \in \underline{\lambda}^l}}^l$$

où $\underline{\lambda}^l$ est l'ensemble des indices après tirage correspondant à \underline{D}^l .

On utilisera aussi les notations simplifiées $a_{\underline{\lambda}^l}^l$ et A^l .

PROPRIÉTÉS DES EFFETS ALÉATOIRES DU MODÈLE

Les différents niveaux d'un même effet aléatoire ont la même distribution. En effet, les variables aléatoires $X_{i_k, u_k}^{s_k}$ ont la même distribution pour toutes les valeurs du niveau i_k . Cela montre en particulier que les différents niveaux de chaque effet ont la même espérance et la même variance. On fera de plus l'hypothèse que, lorsque la taille des populations tends vers l'infini, ces variances restent finies.

Nous allons commencer par démontrer que les espérances de ces variables aléatoires sont nulles; puis nous préciserons les couples de niveaux d'effets qui donnent une covariance nulle.

Espérance

L'exemple

Prenons dans l'exemple l'effet $A_{i_1 i_2 i_2'}^4$. Nous allons en calculer l'espérance et mettre en évidence l'idée de la démonstration donnée dans le cas général.

$$A_{i_1 i_2 i_2'}^4 = \sum_{u_1} \sum_{u_2} \sum_{u_2'} X_{i_1 u_1}^{c_1} X_{i_2 u_2}^{c_2(u_1)} X_{i_2' u_2'}^{c_2(u_1)} a_{u_1 u_2 u_2'}^4$$

Notons $n(c_1)$ le nombre d'individus dans la population c_1 , $n(u_1)$ le nombre d'individus de la descendance $c_2(u_1)$ et $\text{Esp}(X|Y)$ l'espérance conditionnelle de X sachant Y , alors :

$$\begin{aligned}
 \text{Esp}(A_{i_1 i_2 i'_2}^4) &= \sum_{u_1} \sum_{u_2} \sum_{u'_2} \text{Esp}(X_{i_1 u_1}^{c_1}) \text{Esp}(X_{i_2 u_2}^{c_2(u_1)} X_{i'_2 u'_2}^{c_2(u_1)}) a_{u_1 u_2 u'_2}^4 \\
 &= \frac{1}{n(c_1)} \sum_{u_1} \sum_{u_2} \sum_{u'_2} \text{Esp}(X_{i'_2 u'_2}^{c_2(u_1)} (\text{Esp}(X_{i_2 u_2}^{c_2(u_1)} | X_{i'_2 u'_2}^{c_2(u_1)}))) a_{u_1 u_2 u'_2}^4 \\
 &= \frac{1}{n(c_1)} \sum_{u_1} \frac{1}{n(u_1)} \sum_{u'_2} \sum_{u_2} \text{Esp}(X_{i_2 u_2}^{c_2(u_1)} | X_{i'_2 u'_2}^{c_2(u_1)} = 1) a_{u_1 u_2 u'_2}^4 \\
 &= \frac{1}{n(c_1)} \sum_{u_1} \frac{1}{n(u_1)(n(u_1) - 1)} \sum_{u'_2} \sum_{u_2, u_2 \neq u'_2} a_{u_1 u_2 u'_2}^4
 \end{aligned}$$

Cette espérance est nulle d'après les conditions supplémentaires. La démonstration pour les autres effets est encore plus simple.

Cas général

Dans tout effet a^l , on peut toujours trouver un indice \underline{u}_h est un ensemble $\underline{\mathcal{U}}_{s_h}^l$ (comme défini dans le paragraphe « Cas général ») tels que tous les indices de $\underline{\mathcal{U}}_{s_h}^l$ soient de plus basse hiérarchie pour cet effet. Cette propriété se démontre simplement en utilisant le fait que le nombre d'indices est fini et qu'aucun individu ne peut être son propre descendant. Pour l'effet a^4 de notre exemple, il s'agit de \underline{u}_2 et de $\underline{\mathcal{U}}_{c_2(u_1)}^4 = \{u'_2\}$.

Notons \underline{i}_h et $\underline{\mathcal{I}}_{s_h}^l$ les indices des tirages correspondant respectivement à \underline{u}_h et $\underline{\mathcal{U}}_{s_h}^l$. On peut écrire l'effet A^l de la manière suivante :

$$A_{i_k \in \underline{\mathcal{I}}^l}^l = \sum_{u_k: \underline{u}_k \in \check{\underline{\mathcal{U}}}^l} \sum_{u'_h: \underline{u}'_h \in \underline{\mathcal{U}}_{s_h}^l} \sum_{u_h} \left(\prod_{i_k \in \check{\underline{\mathcal{I}}}^l} X_{i_k, u_k}^{s_k} \right) \left(\prod_{i'_h \in \underline{\mathcal{I}}_{s_h}^l} X_{i'_h, u'_h}^{s_h} X_{i_h, u_h}^{s_h} \right) a_{\underline{u}_k \in \underline{\mathcal{U}}^l}^l$$

où $\check{\underline{\mathcal{U}}}^l$ est le complémentaire dans $\underline{\mathcal{U}}$ de l'ensemble $\{\underline{u}_h, \underline{\mathcal{U}}_{s_h}^l\}$ des indices de plus basse hiérarchie et $\check{\underline{\mathcal{I}}}^l$ le complémentaire dans $\underline{\mathcal{I}}^l$ de l'ensemble $\{\underline{i}_h, \underline{\mathcal{I}}_{s_h}^l\}$ des indices de tirage de plus basse hiérarchie. Soit $n(s_h)$ le nombre d'individus de la descendance s_h et q le nombre d'indices dans $\underline{\mathcal{U}}_{s_h}^l$. Pour calculer l'espérance de $A_{i_k \in \underline{\mathcal{I}}^l}^l$ on utilisera

l'indépendance des variables aléatoires de tirages lorsque les tirages ont lieu dans des descendance différentes puis le fait que :

$$\text{Esp}(X_{i_h, u_h}^{s_h} | \prod_{i'_h \in \underline{\mathcal{I}}_{s_h}^l} X_{i'_h, u'_h}^{s_h} = 1) = \frac{1}{n(s_h) - q}$$

L'utilisation des conditions supplémentaires permet alors d'obtenir la nullité de l'espérance de $A_{i_k \in \underline{\mathcal{I}}^l}^l$.

Détermination des couples de niveaux d'effets qui donnent une covariance non nulle

Il s'agit d'étudier les covariances existant entre les niveaux d'un même effet A^l ou bien entre les niveaux de 2 effets différents A^l et $A^{l'}$.

L'exemple

Classiquement, comme dans Lefort (1977) ou Mallard *et al* (1983), on développe l'expression du produit de 2 effets pour obtenir les résultats concernant la covariance entre leurs niveaux. Il est possible d'utiliser cette technique dans le cadre de l'exemple, mais elle ne semble pas efficace pour donner une démonstration dans le cas général.

Remarquons tout d'abord que, dans l'exemple, il existe des covariances non nulles entre des niveaux de deux effets différents du modèle. Il s'agit en particulier des niveaux de $A_{\lambda_2}^2$ et $A_{\lambda_2}^3$ qui ont une covariance non nulle puisque les individus λ_2 qui définissent les niveaux de ces effets sont les mêmes. Mais il existe aussi des covariances entre les niveaux d'un même effet, par exemple les niveaux $A_{\lambda_2 \lambda'_2}^4$ et $A_{\lambda'_2 \lambda_2}^4$.

Cependant, la covariance est nulle entre la plupart des niveaux de A^2 et A^3 , et de même entre la plupart des niveaux de A^4 . Ce qu'il faut démontrer pour valider les caractéristiques du modèle décrit dans [Man], c'est que la covariance intra- ou intereffet est nulle ou tend vers zéro avec la taille de la population, lorsque l'on ne peut pas passer de la suite d'indices qui définit un niveau du premier effet à la suite des indices qui permet de définir un niveau du second effet par permutation des indices.

On peut distinguer 2 situations. Cette covariance est structurellement nulle par exemple entre $A_{\lambda_2 \lambda'_2}^4$ et $A_{\lambda_2}^3$ alors qu'elle tend vers zéro avec la taille de la population dans le cas qui est décrit ci-dessous. Pratiquement la covariance sera considérée comme nulle dans les 2 cas, et il n'est donc pas nécessaire de les distinguer.

Nous allons le démontrer dans le détail pour 2 niveaux de l'effet A^4 . Considérons par exemple $A_{\lambda_2 \lambda'_2}^4$ et $A_{\lambda_2 \lambda''_2}^4$, qui s'écrivent $A_{i_1 i_2 i'_2}^4$ et $A_{i_1 i_2 i''_2}^4$ dans la notation N_1 avec la correspondance $\lambda_2 = (i_1 i_2)$, $\lambda'_2 = (i_1 i'_2)$ et $\lambda''_2 = (i_1 i''_2)$:

$$\begin{aligned} &Esp(A_{\lambda_2 \lambda'_2}^4 A_{\lambda_2 \lambda''_2}^4) \\ &= \sum_{u_1} Esp(X_{i_1 u_1}^{c_1}) \sum_{u_2} \sum_{u'_2} \sum_{u''_2} Esp(X_{i_2 u_2}^{c_2(u_1)} X_{i'_2 u'_2}^{c_2(u_1)} X_{i''_2 u''_2}^{c_2(u_1)}) a_{u_1 u_2 u'_2}^4 a_{u_1 u_2 u''_2}^4 \\ &= \frac{1}{n(c_1)} \sum_{u_1} \left(\sum_{u_2} \sum_{u'_2} Prob(X_{i_2 u_2}^{c_2(u_1)} = 1, X_{i'_2 u'_2}^{c_2(u_1)} = 1) \right) a_{u_1 u_2 u'_2}^4 \\ &\quad \sum_{u''_2} Esp(X_{i''_2 u''_2}^{c_2(u_1)} | X_{i_2 u_2}^{c_2(u_1)} = 1, X_{i'_2 u'_2}^{c_2(u_1)} = 1) a_{u_1 u_2 u''_2}^4 \end{aligned}$$

où $n(c_1)$ est le nombre d'individus dans la population c_1 .

En utilisant la condition supplémentaire $\sum_{u_2, u_2' \neq u_2''} a_{u_1 u_2 u_2''}^4 = 0$ on obtient :

$$Esp(A_{\lambda_2 \lambda_2'}^4 A_{\lambda_2 \lambda_2''}^4) = \frac{1}{n(c_1)} \sum_{u_1} \frac{1}{n(u_1)(n(u_1) - 1)} \sum_{u_2} \sum_{u_2'} a_{u_1 u_2 u_2'}^4 \frac{-a_{u_1 u_2 u_2'}^4}{n(u_1) - 2}$$

où $n(u_1)$ est le nombre d'individus de la descendance $c_2(u_1)$. Cette covariance tend vers zéro lorsque le nombre d'individus $n(u_1)$ tend vers l'infini.

Le cas général

Il s'agit essentiellement ici de démontrer le résultat qui est énoncé dans [Man] : la covariance entre le niveau d'un même effet ou de 2 effets est nulle ou tend vers zéro avec la taille de la population lorsqu'il n'existe pas de permutations entre les niveaux de ces effets. Ce résultat est énoncé plus formellement dans la proposition ci-dessous.

Proposition

La covariance entre le niveau \mathcal{L}^l de l'effet A^l et le niveau $\mathcal{L}^{l'}$ de l'effet $A^{l'}$ est nulle lorsque l'on ne peut pas passer de la suite des niveaux d'indices \mathcal{L}^l , qui définit un niveau de l'effet A^l , à la suite des niveaux d'indices $\mathcal{L}^{l'}$ qui permet de définir un niveau de l'effet $A^{l'}$, par permutation de ces niveaux d'indices.

Démonstration

Soit λ_h^l un niveau d'un indice de \mathcal{L}^l qui n'est pas l'un des niveaux d'un indice de $\mathcal{L}^{l'}$ ni un niveau d'un indice d'un ancêtre d'un des indices de $\mathcal{L}^{l'}$.

Précisons dans notre exemple.

Dans l'étude de la covariance de A^1 et A^4 , on pourra choisir λ_2' et λ_2 lorsque $A^l = A^4, A^{l'} = A^1, \mathcal{L}^l = \{\lambda_2, \lambda_2'\}$ et $\mathcal{L}^{l'} = \{\lambda_1\}$. Mais λ_1 ne convient que si il n'est pas l'ancêtre commun à λ_2 et λ_2' .

Dans l'étude de la covariance de A^4 avec A^4 on pourra choisir λ_2'' lorsque $A^l = A^4, A^{l'} = A^4, \mathcal{L}^l = \{\lambda_2, \lambda_2''\}$ et $\mathcal{L}^{l'} = \{\lambda_2, \lambda_2'\}$. Mais on aurait pu aussi choisir λ_2' si l'on avait interverti \mathcal{L}^l avec $\mathcal{L}^{l'}$. Cependant λ_2 ne convient pas car il est présent dans $\mathcal{L}^{l'}$ et \mathcal{L}^l .

L'annexe 3 montre qu'il existe toujours un niveau d'indice de ce type lorsque l'on ne peut pas passer de la suite des niveaux d'indices \mathcal{L}^l qui définit un niveau de l'effet A^l , à la suite des niveaux d'indices $\mathcal{L}^{l'}$ qui permet de définir un niveau de l'effet $A^{l'}$, par permutation des niveaux d'indices.

Soient i_h le niveau d'un indice i_h de plus basse hiérarchie de λ_h^l dans la notation N_1, \underline{u}_h son indice correspondant avant tirage, $\mathcal{U}_{s_h}^l$ l'ensemble des indices présents dans \underline{u}_h^l appartenant à la même descendance s_h que les individus indicés par les niveaux de \underline{u}_h .

Nous allons considérer dans \underline{u}_h^l les ensembles d'indices $\underline{u}_h, \mathcal{U}_{s_h}^l, \text{Desc}(\mathcal{U}_{s_h}^l)$ qui représentent l'ensemble des indices de descendants de niveaux d'indices de \underline{u}_h^l (un

niveau de l'indice \underline{u}_h ne peut avoir de descendant puisque \underline{u}_h est de plus basse hiérarchie), ainsi que $\check{\underline{U}}^l$ le complémentaire dans \underline{U}^l des indices précédents.

De même, nous considérons $\underline{U}_{s_h}^{l'}$, $\underline{Desc}(\underline{U}_{s_h}^{l'})$ et $\check{\underline{U}}^{l'}$.

On notera : $\check{\underline{U}}^{l,l'}$, $\underline{U}_{s_h}^{l,l'}$ et $\underline{Desc}(\underline{U}_{s_h}^{l,l'})$ les unions d'ensembles d'indices correspondants pour les 2 effets et $\underline{I}_{s_h}^{l,l'}$ l'ensemble des indices de tirage correspondant à $\underline{U}_{s_h}^{l,l'}$. Remarquons qu'aucun des indices de $\underline{I}_{s_h}^{l,l'}$ n'a un niveau i_h , car alors λ_h serait présent dans $\mathcal{L}^{l'}$.

L'espérance du produit des variables aléatoires de tirage qui interviennent dans $Esp(A_{\mathcal{L}^l}^l, A_{\mathcal{L}^{l'}}^{l'})$ peut se scinder en 3 produits d'espérances qui correspondent aux sommes sur les indices de $\check{\underline{U}}^{l,l'}$ (pour le premier produit), de \underline{u}_h et de $\underline{U}_{s_h}^{l,l'}$ (pour le deuxième produit) et de $\underline{Desc}(\underline{U}_{s_h}^{l,l'})$ (pour le dernier produit).

1) Si $\underline{U}_{s_h}^{l,l'}$ est vide, alors une démonstration analogue à celle faite dans le paragraphe « Cas général » permet d'obtenir que la covariance est nulle. Il faut pour cela chercher un indice de plus basse hiérarchie dans $\underline{Desc}(\underline{U}_{s_h}^{l,l'})$ s'il n'est pas vide, ou prendre \underline{u}_h sinon.

2) Si $\underline{U}_{s_h}^{l,l'}$ n'est pas vide on écrit :

$$\begin{aligned}
 Esp(A_{\mathcal{L}^l}^l, A_{\mathcal{L}^{l'}}^{l'}) &= \sum_{\substack{u_k; \underline{u}_k \in \check{\underline{U}}^{l,l'} \\ u'_h; \underline{u}'_h \in \underline{U}_{s_h}^{l,l'}}} \underbrace{(\dots)}_{\text{premier produit}} \\
 &\quad \sum_{u_h} \text{Prob} \left(\prod_{\substack{i'_h \in \underline{I}_{s_h}^{l,l'} \\ i'_h \in \underline{I}_{s_h}^{l,l'}}} X_{i'_h, u'_h}^{s_h} = 1 \right) \\
 &\quad \sum_{u_h} Esp(X_{i_h, u_h}^{s_h} \mid \prod_{\substack{i'_h \in \underline{I}_{s_h}^{l,l'} \\ i'_h \in \underline{I}_{s_h}^{l,l'}}} X_{i'_h, u'_h}^{s_h} = 1) \\
 &\quad \sum_{\substack{u_{k'}; \underline{u}_{k'} \in \underline{Desc}(\underline{U}_{s_h}^{l,l'}) \\ u_k \in \underline{U}^l}} \underbrace{(\dots)}_{\text{dernier produit}} a_k^{l'} \prod_{\substack{u'_k \in \underline{U}^{l'} \\ u_k \in \underline{U}^l}} a_k^l
 \end{aligned}$$

On termine la démonstration, en remarquant que le dernier produit ne dépend pas du niveau de l'indice \underline{u}_h , on peut donc intervenir la somme sur u_h avec la somme multiple sur les indices de $\underline{Desc}(\underline{U}_{s_h}^{l,l'})$. On utilise donc :

$$\begin{aligned}
 \sum_{u_h} Esp(X_{i_h, u_h}^{s_h} \mid \prod_{\substack{i'_h \in \underline{I}_{s_h}^{l,l'} \\ i'_h \in \underline{I}_{s_h}^{l,l'}}} X_{i'_h, u'_h}^{s_h} = 1) a_k^{l'} \prod_{\substack{u'_k \in \underline{U}^{l'} \\ u_k \in \underline{U}^l}} a_k^l \\
 = \frac{1}{n(s_h) - q} a_k^{l'} \prod_{\substack{u'_k \in \underline{U}^{l'} \\ u_k \in \underline{U}^l}} a_k^l
 \end{aligned}$$

où $q - 1$ est le nombre d'indices dans $\underline{U}_{s_h}^{l,l'}$ et $n(s_h)$ le nombre d'individus de la descendance s_h . L'utilisation des conditions supplémentaires et le fait de faire tendre $n(s_h)$ vers l'infini permettent d'obtenir soit que la covariance est nulle soit qu'elle tend vers zéro.

CONCLUSIONS

Les développements ont été faits dans le cas où une seule variable est observée sur un seul type de descendance. Sa généralisation aux cas multivariés où plusieurs types de descendance sont observés simultanément nécessiterait des notations encore plus lourdes, mais ne poserait pas de problèmes.

Le modèle obtenu ainsi peut être très lourd et n'est pas toujours le plus efficace. Dans certains cas l'utilisateur peut être amené à diminuer le nombre d'effets de son modèle. Par exemple, dans les cas diallèles, il peut faire l'hypothèse d'absence d'effets réciproques. Dans ces cas-là, la validité du modèle obtenu dépendra de la validité des hypothèses ajoutées, mais souvent le modèle moins paramétré sera plus efficace même s'il n'est pas parfaitement valide. Quelques idées sur l'efficacité des modèles sont données dans Bouchez et Goffinet (1990).

Enfin, on ajoute fréquemment l'hypothèse de normalité des variables aléatoires. Il s'agit d'une hypothèse qui n'est justifiée en rien par des arguments statistiques du type de ceux employés ici. Cette hypothèse est utile par exemple lorsque l'on veut déterminer les stratégies optimales de sélection (Goffinet et Elsen, 1984). L'estimation des paramètres de variance et covariance obtenues avec ces hypothèses en utilisant des techniques de maximum de vraisemblance restera le plus souvent raisonnable, même si celles-ci s'avèrent inexactes.

RÉFÉRENCES

- Berge C (1983) *Graphes*. Gauthier-Villars, Paris
- Bouchez A, Goffinet B (1990) Evaluation of selection index : application to the choice of an indirect multitrait selection index for soybean breeding. *Theor Appl Genet* 79, 261-267
- Goffinet B, Elsen JM (1984) Critère optimal de sélection : quelques résultats généraux. *Génét Sél Evol* 16, 307-318
- Lefort G (1977) Remarques sur la modélisation et l'interprétation des dispositifs diallèles. *Ann Génét Sél Anim* 27, 171-202
- Mallard J, Masson JP, Douaire M (1983) Interaction et modèle mixte. *Génét Sél Evol* 15, 379-394
- Mangin B, Vincourt P (1992) Schémas de sélection : de la représentation généalogique au modèle statistique. Élaboration du modèle. *Genet Sel Evol* 24, 71-84
- Scheffé H (1959) *The analysis of variance*. Wiley & Sons, New York

ANNEXES

Annexe 1. Bijection entre les effets définis dans « le modèle complet et les conditions supplémentaires » et les sous-graphes considérés dans [Man]

Soient :

– \underline{U} , l'ensemble des indices nécessaires pour décrire un individu du type de descendance c_K ;

– E , l'ensemble des sommets du graphe du cercle c_K auquel appartient cet individu.

Prenons un effet du modèle, défini dans la notation N_1 par l'ensemble d'indices :

$$\underline{U}^l = \{\underline{u}_k \in \underline{U}; \text{ tous les indices hiérarchisant } \underline{u}_k \in \underline{U}^l\}$$

et dans la notation N_2 par l'ensemble d'indices \underline{D}^l .

Prenons le graphe G de la relation «être descendant par une manipulation élémentaire» définie sur l'ensemble E , et considérons un sous-graphe G^l dont l'ensemble des sommets E^l contient au moins la racine de G et dont les sommets ont un demi-degré intérieur égal à celui qu'ils avaient dans G . Les concepts de la théorie des graphes utilisés ici sont issus de Berge (1983), et sont précisés dans [Man]. Définissons :

$$S^l = \{c_{k/l} \in E^l; c_{k/l} \text{ est de demi-degré extérieur égal à zéro } \}$$

Dans la proposition 1, nous démontrons qu'il existe une bijection notée b entre \underline{U} et E .

Nous montrons alors, qu'un $b(\underline{D}^l)$ est un S^l (propositions 2 et 3). Puis qu'un $b^{-1}(S^l)$ est un \underline{D}^l (propositions 4).

Les propositions 2 et 3 prouvent qu'il y a une bijection de l'ensemble des \underline{D}^l vers l'ensemble des S^l . La proposition 4 prouve qu'il y a une injection de l'ensemble des S^l vers l'ensemble \underline{D}^l . Ces 2 ensembles étant de dimension finie les 2 injections étant inverses l'une de l'autre, on peut conclure que la bijection b qui existe entre \underline{U} et E permet d'induire une bijection entre l'ensemble des \underline{D}^l et l'ensemble des S^l , ce qui permet de conclure à l'équivalence des 2 définitions.

Proposition 1

Il existe entre \underline{U} et E une bijection que nous noterons b , qui associe l'indice de l'ancêtre au sommet représentant son type de géniteur.

Démonstration

L'indice \underline{u}_K est associé au sommet $c_{K/1}$. Prenons l'indice $\underline{u}_k \in \underline{U}$ pour $k \neq K$, il appartient au type de descendance c_k et donc au cercle c_k du schéma. Il est utilisé par la suite, soit comme un type de géniteur correspondant à un sommet $c_{k/l}$ soit à un sommet $c_{\bar{k}/m}$ d'un cercle $c_{\bar{k}}$ équivalent à c_k (c'est-à-dire un cercle représentant le même type de descendance). D'autre part, s'il est utilisé par 2 ou plusieurs fois comme ancêtre appartenant à 2 ou plusieurs types de géniteur, les sommets qui correspondent à ces types de géniteur sont égalés dans le graphe. À un ancêtre indicé par \underline{u}_k correspond un unique sommet du graphe où un ensemble de sommets tous égaux. Réciproquement, prenons tous les sommets égaux au sommet $c_{k/l}$; ils appartiennent au cercle c_k et donc à un type de descendance c_k et ils leur correspondent un indice \underline{u}_k dans \underline{U} . S'ils sont égaux à un sommet $c_{\bar{k}/l}$, ils appartiennent au type de descendance c_k , si le cercle c_k est équivalent au cercle $c_{\bar{k}}$ et l'on conclut.

Proposition 2

Pour qu'un ensemble S de sommets de E soit égal à un S^l d'un des sous-graphes considérés, il faut et il suffit qu'aucun chemin n'existe dans G entre 2 quelconques de ses sommets.

Démonstration

Par construction, un ensemble de sommets S^l ne possède aucun chemin dans G entre ses sommets.

Réciproquement : soit un ensemble S pour lequel aucun chemin n'existe entre ses sommets. Construisons le sous-graphe de G , à partir de tous les chemins existant dans G entre l'un des sommets et la racine de G . Il est évident qu'un tel sous-graphe a la racine G dans ses sommets et que S est l'ensemble de ses sommets de demi-degré égale à zéro. Le fait qu'il possède aussi la propriété pour les demi-degrés intérieurs de chacun de ses sommets découle de l'absence de circuit dans G .

Considérons maintenant un $\underline{\mathcal{D}}^l$ associé à un effet du modèle.

Proposition 3

$b(\underline{\mathcal{D}}^l)$ possède la propriété de la proposition 2.

Démonstration

Supposons que la propriété soit fautive, alors il existe c et c' appartenant tous deux à $b(\underline{\mathcal{D}}^l)$, et un chemin dans G entre les deux; orientons-le par exemple de c à c' . Considérons maintenant $\underline{\delta} = b^{-1}(c)$ et $\underline{\delta}' = b^{-1}(c')$. L'existence du chemin implique que $\underline{\delta}$ est un indice descendant de $\underline{\delta}'$, ce qui est contradictoire avec le fait que $\underline{\delta}'$ est de plus basse hiérarchie.

Remarque : par construction, aucun des indices de $\underline{\mathcal{D}}^l$ n'est ancêtre l'un de l'autre. Et par définition, tous les ensembles d'indices possédant cette propriété définissent un effet du modèle.

Considérons un des S^l et son ensemble d'indices associé $b^{-1}(S^l)$.

Proposition 4

$b^{-1}(S^l)$ possède la propriété de la remarque.

Démonstration

Supposons que cela soit faux. Alors il existe $\underline{\delta}$ et $\underline{\delta}'$ dans $b^{-1}(S^l)$ tels que par exemple $\underline{\delta}$ est un indice ancêtre de $\underline{\delta}'$. Ceci implique que $c = b(\underline{\delta})$ qui appartient à S^l n'est pas de demi-degré extérieur égal à zéro, ce qui est contradictoire.

Annexe 2. Les conditions supplémentaires proposées rendent les paramètres estimables

Définissons, pour chaque ensemble \underline{U}^l , une quantité $\tilde{g}^{[u_k]}_{\underline{u}_k \in \underline{U}^l}$ par récurrence :

$$\tilde{g}^{[u_k]}_{\underline{u}_k \in \underline{U}^l} = \frac{1}{n(s_h) - q} \sum_{\substack{u_h, u_h \notin \underline{U}^{l'} \\ \underline{u}_k \in \underline{U}^{l'}}} \tilde{g}^{[u_k]}_{\underline{u}_k \in \underline{U}^{l'}}$$

avec :

$$g^{[u_k]}_{\underline{u}_k \in \underline{U}} = g^{[u_k]}_{\underline{u}_k \in \underline{U}}$$

où \underline{u}_k est un indice tel que $\underline{U}^{l'} = \{\underline{U}^l, \underline{u}_j\}$ représente l'ensemble des indices d'un effet $a^{l'}$, s_h est le numéro de la descendance formée de $n(s_h)$ individus, dans laquelle se trouve les niveaux de $\underline{u}_h, \underline{U}^{l'}_{s_h} = \{\underline{u}'_h, \underline{u}''_h, \dots\}$, l'ensemble des q indices correspondant à des individus de la même descendance que les individus indicés par \underline{u}_h dans $\underline{U}^{l'}$.

Bien qu'il existe plusieurs façons d'obtenir par récurrence la quantité $\tilde{g}^{[u_k]}_{\underline{u}_k \in \underline{U}^l}$ pour les ensembles \underline{U}^l , tels que :

$$\underline{U}^l = \{\underline{U}^{l'}, \underline{u}_h\} = \{\underline{U}^{l''}, \underline{u}'_h\}$$

On peut démontrer que cette quantité ne dépend pas du choix de l'indice \underline{u}_h . L'idée de la démonstration est de faire une récurrence sur le nombre d'indices. Si cette propriété est vraie pour les effets dont le nombre d'indices est n , elle est vraie pour les effets dont le nombre d'indices est $n - 1$, pourvu que la condition d'orthogonalité soit satisfaite.

Écrivons alors :

$$a^l_{\underline{u}_k \in \underline{U}^l} = \tilde{g}^{[u_k]}_{\underline{u}_k \in \underline{U}^l} - \sum_{u_h: \underline{u}_h \in \underline{K}^l} \tilde{g}^{[u_k]}_{\underline{u}_k \in \{\underline{U}^l - u_h\}} + \sum_{u_h, u'_h: (\underline{u}_h, \underline{u}'_h) \in (\underline{K}^l \times \underline{K}^l)} \tilde{g}^{[u_k]}_{\underline{u}_k \in \{\underline{U}^l - \underline{u}_h - \underline{u}'_h\}} - \dots$$

où \underline{K}^l désigne l'ensemble des indices de plus basse hiérarchie de \underline{U}^l et $\sum_{u_h: \underline{u}_h \in \underline{K}^l}$ est

la somme multiple pour tous les niveaux de tous les indices de \underline{K}^l .

On vérifie que $a^l_{\underline{u}_k \in \underline{U}^l}$ ainsi défini vérifie les conditions supplémentaires.

Puis on conclut en vérifiant que la somme des effets a^l rentrant dans la décomposition de $g^{[u_k]}_{\underline{u}_k \in \underline{U}}$ est égale à $g^{[u_k]}_{\underline{u}_k \in \underline{U}}$ (chaque terme de $\tilde{g}^l_{\underline{u}_k \in \underline{U}^l}$ se trouvant autant de fois avec le signe moins qu'avec le signe plus).

Annexe 3. Existence d'un indice de type λ'_h

Soient \underline{L}^l et $\underline{L}^{l'}$ 2 ensembles de niveaux d'indices pour lesquels il n'existe pas de permutation.

Alors, il existe au moins un niveau λ_h^l qui est présent dans \mathcal{L}^l et qui n'est pas présent dans $\mathcal{L}^{l'}$. Ce niveau λ_h^l peut s'écrire aussi en utilisant les indices \underline{i}_k :

$$\lambda_h^l = (i_1^{h,l}, i_2^{h,l}, \dots)$$

Le niveau λ_h^l n'est pas présent dans $\mathcal{L}^{l'}$ mais peut être un ancêtre d'un niveau de $\mathcal{L}^{l'}$, soit $\lambda_{h'}^{l'}$, avec :

$$\lambda_{h'}^{l'} = (\lambda_h^l, i_m^{h,l'}, \dots)$$

Dans ce cas, $\lambda_{h'}^{l'}$ n'est présent ni lui-même, ni comme ancêtre d'un niveau de \mathcal{L}^l . Sinon λ_h^l ne serait pas un niveau d'indice de plus basse hiérarchie dans l'effet A^l et ne pourrait pas être dans \mathcal{L}^l .