

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ÉTUDE DU CHOIX D'UN MODÈLE D'ARBORESCENCE
EN RÉGRESSION LOGISTIQUE 4-NOMIALE SELON
L'EFFET DE LA VALEUR DES PARAMÈTRES

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN MATHÉMATIQUES

PAR

MARIE-CHRISTINE STAFFORD

NOVEMBRE 2008

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Je voudrais remercier ma directrice de recherche, Mme Pascale Rousseau, pour sa grande disponibilité, son implication et sa patience durant toute la rédaction de ce mémoire. Mes remerciements aussi à Mme Gisèle Legault dont le soutien informatique m'a épargné bien des maux de tête.

TABLE DES MATIÈRES

LISTE DES FIGURES	vii
LISTE DES TABLEAUX	xi
RÉSUMÉ	xiii
INTRODUCTION	1
CHAPITRE I	
RAPPELS SUR LA RÉGRESSION LOGISTIQUE	5
1.1 La régression logistique multinomiale	5
1.2 L'estimation individuelle	9
CHAPITRE II	
ARBORESCENCES DE RÉGRESSION LOGISTIQUE	13
2.1 Estimation des paramètres	22
2.2 Études de simulation	28
2.2.1 Effet du nombre de variables explicatives	29
2.2.2 Effet de la valeur des paramètres	31
2.2.3 Effet de la taille de l'échantillon	37
CHAPITRE III	
ÉQUIVALENCE DE LA VRAISEMBLANCE DE CERTAINS TYPES D'ARBO- RESCENCES SELON LA FORME DES PARAMÈTRES	41
3.1 Théorème de Rosenberg	41
3.2 Effet de la condition de Rosenberg sur les valeurs des paramètres	45
3.3 Cas particuliers	47
3.4 Cas orthogonaux ou quasi-orthogonaux	59
3.5 Équivalence des arborescences A et C et B et D sous certaines conditions	63
CHAPITRE IV	
ÉTUDE DE SIMULATIONS MONTE-CARLO	69
4.1 Notation	70
4.1.1 Structures des arborescences	70

4.1.2	Quelques remarques sur les matrices des paramètres	70
4.1.3	Notation pour les différentes arborescences de même structure	72
4.2	Exemples de suites d'arborescences ordonnées selon la vraisemblance	73
4.2.1	Exemple 1 : $(t, x, y, z) \ t \neq z \neq x, y$	75
4.2.2	Exemple 2 : $(t = 0, x, y, z) \ x \neq y \neq z$	75
4.2.3	Exemple 3 : $(t = 0, x = 0, y, z) \ y \neq z$	76
4.2.4	Exemple 4 : $(t = 0, x = 0, y = 0, z \neq 0)$	76
4.3	Étude des simulations pour le vecteur (t, x, y, z) , avec $x = y = z$	76
4.3.1	Vecteurs de paramètres avec $t = x = y = z$	77
4.3.2	Vecteurs de paramètres avec $t = 0, x = y = z$	77
4.3.3	Vecteurs de paramètres avec $t \leq x = y = z$ et $t > x = y = z$	78
4.4	Étude des simulations pour le vecteur de paramètres $(t = 0, x, y, z)$	79
4.5	Résumé global des résultats des simulations	87
	CONCLUSION	117
	BIBLIOGRAPHIE	119

LISTE DES FIGURES

2.1	Arborescence S	20
2.2	Arborescence T	20
2.3	Structure d'arborescence pour une expérience 4-nomiale	29
2.4	Structure d'arborescence pour une expérience binomiale suivie par une tri-nomiale	29
2.5	Structure d'arborescence pour une tri-nomiale suivie d'une binomiale	30
2.6	Structure d'arborescence pour trois binomiales consécutives	30
2.7	Structure d'arborescence pour une binomiale suivie de deux binomiales parallèles	30
2.8	Graphique du nombre de fois où le bon modèle d'arborescence est sélectionné en fonction du nombre de variables explicatives, pour chaque type d'arborescence (repris de l'article de Rousseau et Sankoff).	32
2.9	Tableau des valeurs des paramètres pour une expérience 4-nomiale (repris de l'article de Rousseau et Sankoff).	33
2.10	Tableau des valeurs des paramètres pour une expérience tri-nomiale suivie d'une binomiale (repris de l'article de Rousseau et Sankoff).	34
2.11	Graphique du nombre de fois où le bon modèle d'arborescence est sélectionné en fonction du poids maximal du coefficient, quand la vraie arborescence est 4-nomiale (repris de l'article de Rousseau et Sankoff).	35

2.12	Graphique du nombre de fois où le bon modèle d'arborescence est sélectionné en fonction du poids maximal du coefficient, quand la vraie arborescence est une tri-nomiale suivie d'une binomiale (repris de l'article de Rousseau et Sankoff).	36
2.13	Graphique du nombre de fois où le bon modèle d'arborescence est sélectionné en fonction de T , quand la vraie arborescence est 4-nomiale (repris de l'article de Rousseau et Sankoff).	38
2.14	Graphique du nombre de fois où le bon modèle d'arborescence est sélectionné en fonction de T , quand la vraie arborescence est une trinomiale suivie d'une binomiale (repris de l'article de Rousseau et Sankoff).	39
3.1	Arborescence A	63
3.2	Arborescence C	63
4.1	Structure d'arborescence A	70
4.2	Structure d'arborescence B	70
4.3	Structure d'arborescence C	70
4.4	Structure d'arborescence D	71
4.5	Structure d'arborescence E	71
4.6	Ordre des arborescences pour le vecteur de paramètres $(t, x, y, z)=(1.2.2,3)$	89
4.7	Ordre des arborescences pour le vecteur de paramètres $(t, x, y, z)=(0.1.2,3)$	90
4.8	Ordre des arborescences pour le vecteur de paramètres $(t, x, y, z)=(0,0,2,3)$	91
4.9	Ordre des arborescences pour le vecteur de paramètres $(t, x, y, z)=(0,0,0,3)$	92
4.10	Sortie d'ordinateur : Ordre des arborescences pour un vecteur de paramètres de la forme (t, x, x, x)	93

- 4.11 Ordre des arborescences pour le vecteur de paramètres $(x, x, x, x)=(3,3,3,3)$ 94
- 4.12 Ordre des arborescences pour le vecteur de paramètres $(x, x, x, x)=(2,2,2,2)$ 95
- 4.13 Ordre des arborescences pour le vecteur de paramètres $(x, x, x, x)=(1,1,1,1)$ 96
- 4.14 Ordre des arborescences pour le vecteur de paramètres $(x, x, x, x)=(.5,.5,.5,.5)$ 97
- 4.15 Ordre des arborescences pour le vecteur de paramètres $(0, x, x, x)=(0,3,3,3)$ 98
- 4.16 Ordre des arborescences pour le vecteur de paramètres $(0, x, x, x)=(0,2,2,2)$ 99
- 4.17 Ordre des arborescences pour le vecteur de paramètres $(0, x, x, x)=(0,1,1,1)$ 100
- 4.18 Ordre des arborescences pour le vecteur de paramètres $(0, x, x, x)=(0,.5,.5,.5)$ 101
- 4.19 Ordre des arborescences pour le vecteur de paramètres $(t, x, x, x)=(3,3,3,3)$ 102
- 4.20 Ordre des arborescences pour le vecteur de paramètres $(t, x, x, x)=(2,3,3,3)$ 103
- 4.21 Ordre des arborescences pour le vecteur de paramètres $(t, x, x, x)=(1,3,3,3)$ 104
- 4.22 Ordre des arborescences pour le vecteur de paramètres $(t, x, x, x)=(.5,3,3,3)$ 105
- 4.23 Ordre des arborescences pour le vecteur de paramètres $(t, x, x, x)=(0,3,3,3)$ 106
- 4.24 Ordre des arborescences pour le vecteur de paramètres $(t, x, x, x)=(1,2,2,2)$ 107
- 4.25 Ordre des arborescences pour le vecteur de paramètres $(t, x, x, x)=(.5,2,2,2)$ 108
- 4.26 Ordre des arborescences pour le vecteur de paramètres $(t, x, x, x)=(0,2,2,2)$ 109
- 4.27 Ordre des arborescences pour le vecteur de paramètres $(t, x, x, x)=(.5,1,1,1)$ 110
- 4.28 Ordre des arborescences pour le vecteur de paramètres $(t, x, x, x)=(0,1,1,1)$ 111
- 4.29 Sortie d'ordinateur : Ordre des arborescences pour un vecteur de paramètres de la forme $(0, 1, y, z)$ 112
- 4.30 Ordre des arborescences pour le vecteur de paramètres $(0, x, y, z) = (0, 2, 3, 3)$ 113

4.31 Ordre des arborescences pour le vecteur de paramètres $(0, x, y, z) = (0, 2, 2, 3)$ 114

4.32 Ordre des arborescences pour le vecteur de paramètres $(0, x, y, z) = (0, 1, 2, 3)$ 115

LISTE DES TABLEAUX

4.1	Tableau résumé de l'équivalence des arborescences selon le nombre de variables explicatives différentes de 0	74
4.2	Sous-groupes d'arborescences pour les vecteurs de paramètres de la forme (t, x, x, x) avec $t < x$	80
4.3	Sous-groupes d'arborescences pour les vecteurs de paramètres de la forme (t, x, x, x) avec $t > x$	81
4.4	Ordre des arborescences pour chacune des valeurs de (t, x, x, x) , pour $t < x$	82
4.5	Ordre des arborescences pour chacune des valeurs de (t, x, x, x) , pour $t > x$	83
4.6	Description des sous-groupes d'arborescences pour un vecteur de paramètres $(t = 0, x = 1, y = 3, z = 3)$	84
4.7	Description des sous-groupes d'arborescences pour un vecteur de paramètres $(t = 0, x = 1, y = 2, z = 3)$	85
4.8	Description des sous-groupes d'arborescences pour un vecteur de paramètres $(t = 0, x = 1, y = 2, z = 3)$, suite.	86

RÉSUMÉ

Ce mémoire traite de modèles d'arborescences en régression logistique 4-nomiale pour rendre compte du cas où les résultats proviennent de séquences d'expérience multinomiales consécutives ou parallèles. Dans le premier chapitre, nous rappelons le modèle général de régression logistique multinomiale et présentons une méthode d'estimation individuelle des paramètres. Le chapitre suivant rapporte les recherches de Rousseau et Sankoff sur les modèles d'arborescences en régression logistique et présente du même coup le cadre dans lequel la présente étude s'inscrit. Le troisième chapitre porte sur différents résultats qui caractérisent les paramètres pour lesquels certaines structures d'arborescences sont équivalentes. Finalement, le dernier chapitre présente une étude de simulations Monte-Carlo effectuée pour comprendre et mettre en évidence les facteurs influençant l'ordre (selon le maximum de vraisemblance) dans lequel les arborescences sont sélectionnées. Ces simulations ont permis d'identifier certains principes auxquels cet ordre obéit, selon la forme du vecteur des paramètres et la grandeur de ces derniers.

Mots-clés : régression logistique, arborescences, modèles réduits.

INTRODUCTION

L'objet principal de ce mémoire est de poursuivre l'étude de la comparaison de l'estimation des paramètres du modèle logistique d'une expérience multinomiale (dit modèle standard) avec l'estimation individuelle des paramètres de ce même modèle. Cette comparaison avait été faite par Bull et Donner (1993) au moyen de l'étude de l'efficacité relative des estimateurs faite lorsque les paramètres prenaient leurs valeurs dans des régions spécifiques. Ces auteurs étaient arrivés à la conclusion qu'il n'y avait pas de perte d'efficacité pour certaines valeurs de paramètres, donc qu'il n'y avait pas de perte d'information en utilisant l'estimation individuelle. Remarquons que la méthode de l'estimation individuelle telle que définie nous portait à croire que nous n'utilisons pas toute l'information contenue dans l'ensemble de données, ce qui aurait dû entraîner une perte d'efficacité dans l'estimation. Bull et Donner avaient conclu qu'il est vraisemblable que l'estimation des paramètres reliés à certaines combinaisons de catégories puisse être aussi performante que l'estimation standard des paramètres lorsque les valeurs de ceux-ci sont dans certaines régions.

Nous pouvons remarquer que les arborescences de modèles de régression logistique utilisent toutes les combinaisons possibles des catégories. Nous avons donc entrepris de poursuivre cette recherche avec les modèles d'arborescences de régression logistique.

Ces modèles d'arborescences de régression logistique furent proposés par Rousseau et Sankoff (1989) dans le but premier d'analyser des données en linguistique. Ces modèles d'arborescences nous permettent de calculer les probabilités d'observer les catégories d'une multinomiale lorsque ces probabilités sont soumises à des contraintes données par des équations de régression logistique, celles-ci étant différentes de celles du modèle dit standard. Le critère de vraisemblance maximale associé à chaque arborescence est utilisé pour estimer le modèle d'arborescence de régression logistique qui aurait généré

l'ensemble de données.

Lors de simulations Monte-Carlo d'ensembles de données générées selon une loi multinomiale, nous avons remarqué que, pour certaines valeurs des paramètres, le critère de vraisemblance maximale ne pouvait discriminer entre certains modèles d'arborescences, c'est-à-dire que la vraisemblance était la même pour certaines arborescences et pour le modèle multinomial standard. Pour ces valeurs de paramètres, il n'y avait donc pas de perte d'information en utilisant ces modèles d'arborescences autres que modèle standard et ces modèles ont donc été considérés comme équivalents au modèle standard. Nous avons aussi trouvé théoriquement les régions des valeurs des paramètres pour lesquelles ces équivalences sont valides.

Nous avons en effet remarqué que l'estimation individuelle est faite à partir d'une équation logistique qui fait partie des équations de régression logistique qui décrivent ces arborescences équivalentes au modèle standard. Nous avons prouvé que lorsque les probabilités d'observer les catégories qui utilisent cette équation sont linéarisables alors ces modèles d'arborescences sont équivalents au modèle standard.

Dans le premier chapitre de ce mémoire, nous présentons donc le modèle de régression logistique multinomiale, ainsi que l'estimation individuelle.

Le deuxième chapitre, en résumant les recherches déjà menées par Rousseau et Sankoff, décrit les caractéristiques générales des modèles d'arborescences en régression logistique multinomiale. Il contient aussi des résultats théoriques sans lesquels l'étude entreprise serait vaine (notamment en ce qui concerne la différence asymptotique entre les vraisemblances du vrai modèle et d'une autre arborescence) et les résultats d'études de simulation menées par les chercheurs.

Le troisième chapitre présente différents résultats qui clarifient pourquoi certaines arborescences sont équivalentes, selon la valeur des paramètres. Nous y énonçons en premier

lieu un théorème qui décrit selon quelles conditions sur les paramètres l'expression

$$\ln \left(1 + \sum_{i=1}^{m-1} e^{X' \alpha_i} \right)$$

est linéarisable. Nous avons utilisé ce résultat pour déterminer certaines valeurs des paramètres qui respectent ces conditions. Nous avons aussi établi le lien entre le fait que certaines probabilités sont linéarisables et l'équivalence de certaines arborescences.

Le dernier chapitre donne les résultats d'une étude de simulations Monte-Carlo menée par ma directrice de recherche, Mme Pascale Rousseau, et moi-même dans le cas 4-nomial avec 4 variables explicatives binaires. Cette étude nous a permis de mettre en évidence certains principes gouvernant l'ordre dans lequel les arborescences sont classées selon leurs vraisemblances et d'identifier des sous-groupes d'arborescences dont les vraisemblances sont proches. Nous avons constaté que l'ordre dépend de la forme du vecteur des paramètres et de la grandeur de ces derniers.

Nous aimerions poursuivre cette recherche en comparant les régions des valeurs des paramètres pour lesquelles les modèles d'arborescences sont équivalents au modèle standard avec les régions des valeurs des paramètres pour lesquelles il semble ne pas y avoir de perte d'information en utilisant l'estimation individuelle. Nous concluons que nos recherches ont amorcé l'étude d'un lien entre les modèles d'arborescences de régression logistique et l'étude de l'efficacité relative de l'estimation individuelle et de l'estimation standard des paramètres d'une loi multinomiale.

CHAPITRE I

RAPPELS SUR LA RÉGRESSION LOGISTIQUE

Dans ce chapitre, nous décrivons en premier lieu le modèle habituel de la régression logistique multinomiale, puis nous présentons une autre approche selon laquelle les paramètres du modèle multinomial sont estimés deux à deux (avec la catégorie de référence).

1.1 La régression logistique multinomiale

La régression logistique multinomiale est une généralisation du modèle de régression multiple au cas où la variable réponse \mathbf{Y} est multinomiale. La variable \mathbf{Y} est donc un vecteur (Y_1, \dots, Y_m) où Y_t dénote le nombre de résultats dans la catégorie t ($1 \leq t \leq m$) et où $\sum_{t=1}^m Y_t = T$. On suppose que la probabilité P_t d'observer la catégorie t dépend d'un vecteur $\mathbf{X}' = (X_0, \dots, X_d)$ de variables explicatives discrètes ou continues conformément à l'équation suivante :

$$\ln\left(\frac{P_t}{P_m}\right) = \mathbf{X}'\boldsymbol{\alpha}_t, \quad (1.1)$$

où m est la catégorie de référence choisie et où $\boldsymbol{\alpha}_t$ ($1 \leq t \leq m-1$) est un vecteur de paramètres de dimensions $(d+1) \times 1$.

Puisque $\sum_{t=1}^m P_t = 1$, et en utilisant l'équation (1.1), on obtient

$$P_t = P_m e^{\mathbf{X}'\boldsymbol{\alpha}_t},$$
$$\sum_{i=1}^{m-1} P_i = \sum_{i=1}^{m-1} P_m e^{\mathbf{X}'\boldsymbol{\alpha}_i}.$$

$$1 - P_m = P_m \sum_{i=1}^{m-1} e^{\mathbf{X}'\alpha_i},$$

d'où

$$P_m = \frac{1}{1 + \sum_{i=1}^{m-1} e^{\mathbf{X}'\alpha_i}}, \quad (1.2)$$

et

$$P_t = \frac{e^{\mathbf{X}'\alpha_t}}{1 + \sum_{i=1}^{m-1} e^{\mathbf{X}'\alpha_i}}. \quad (1.3)$$

Pour estimer les paramètres α_t , on utilise la méthode du maximum de vraisemblance.

Pour un échantillon aléatoire de taille n , on observe les n vecteurs

$$\mathbf{y}_i = (y_{1i}, \dots, y_{mi})$$

et

$$\mathbf{x}_i = (x_{i0}, \dots, x_{id}),$$

où

$$\sum_{t=1}^m y_{ti} = T_i \quad i = 1, \dots, n,$$

avec lesquels on procède à l'estimation des paramètres en utilisant l'équation

$$\ln\left(\frac{P_{ti}}{P_{mi}}\right) = \mathbf{x}'_i \alpha_t \quad (1 \leq i \leq n),$$

où P_{ti} est la probabilité d'observer la catégorie t pour la i -ème observation dépendant de \mathbf{x}_i .

On a la fonction de vraisemblance

$$\begin{aligned} L &= \prod_{i=1}^n \binom{T_i}{y_{1i} \dots y_{mi}} P_{1i}^{y_{1i}} \dots P_{mi}^{y_{mi}} \\ &= \prod_{i=1}^n \binom{T_i}{y_{1i} \dots y_{mi}} \prod_{t=1}^m P_{ti}^{y_{ti}} \end{aligned}$$

dont le logarithme est

$$\ln(L) = \sum_{i=1}^n \left(\ln \binom{T_i}{y_{1i} \dots y_{mi}} + \sum_{t=1}^m y_{ti} \ln(P_{ti}) \right) \quad (1.4)$$

En reportant (1.3) et (1.2) dans (1.4) on obtient l'expression suivante pour le logarithme de la fonction de vraisemblance, où les paramètres à maximiser figurent explicitement :

$$\begin{aligned} \ln(L) &= \sum_{i=1}^n \left[\ln \left(\begin{matrix} T_i \\ y_{1i} \dots y_{mi} \end{matrix} \right) + \sum_{t=1}^{m-1} y_{ti} \ln \left(\frac{e^{\mathbf{x}'_i \alpha_t}}{1 + \sum_{j=1}^{m-1} e^{\mathbf{x}'_i \alpha_j}} \right) + y_{mi} \ln \left(\frac{1}{1 + \sum_{j=1}^{m-1} e^{\mathbf{x}'_i \alpha_j}} \right) \right] \\ &= \sum_{i=1}^n \left[\ln \left(\begin{matrix} T_i \\ y_{1i} \dots y_{mi} \end{matrix} \right) + \sum_{t=1}^{m-1} y_{ti} \mathbf{x}'_i \alpha_t - \sum_{t=1}^m y_{ti} \ln \left(1 + \sum_{j=1}^{m-1} e^{\mathbf{x}'_i \alpha_j} \right) \right]. \end{aligned}$$

Soient α , le vecteur de dimensions $((m-1)(d+1)) \times 1$ suivant :

$$\alpha' = (\alpha'_1, \dots, \alpha'_{m-1})$$

et \mathbf{H} , la matrice de dimensions $(m-1)(d+1) \times (m-1)(d+1)$ suivante :

$$\mathbf{H} = \begin{pmatrix} \frac{\partial^2 \ln(L)}{\partial \alpha_{10}^2} & \frac{\partial^2 \ln(L)}{\partial \alpha_{11} \partial \alpha_{10}} & \dots & \frac{\partial^2 \ln(L)}{\partial \alpha_{md} \partial \alpha_{10}} \\ \frac{\partial^2 \ln(L)}{\partial \alpha_{10} \partial \alpha_{11}} & \frac{\partial^2 \ln(L)}{\partial \alpha_{11} \partial \alpha_{11}} & & \vdots \\ \vdots & & \ddots & \\ \frac{\partial^2 \ln(L)}{\partial \alpha_{10} \partial \alpha_{md}} & \dots & & \frac{\partial^2 \ln(L)}{\partial \alpha_{md}^2} \end{pmatrix}.$$

où $\left(\frac{\partial \ln(L)}{\partial \alpha} \right)$ est le vecteur de dimension $((m-1)(d+1)) \times 1$ suivant :

$$\left(\frac{\partial \ln(L)}{\partial \alpha} \right)' = \left(\frac{\partial \ln(L)}{\partial \alpha_{10}}, \dots, \frac{\partial \ln(L)}{\partial \alpha_{md}} \right).$$

Alors le maximum de la fonction de vraisemblance peut être trouvé en utilisant l'algorithme de Newton-Raphson, en itérant la formule suivante :

$$\alpha^{(s)} = \alpha^{(s-1)} + \mathbf{H}^{-1} \left(\frac{\partial \ln(L)}{\partial \alpha} \right).$$

À l'étape initiale, $\alpha^{(0)}$ est le vecteur $\mathbf{0}$ et on arrête l'algorithme à l'étape u si $|\alpha^{(u+1)} - \alpha^{(u)}| < \varepsilon$. L'estimateur de vraisemblance maximale $\hat{\alpha}$ prend alors la valeur $\alpha^{(u)}$.

Les premières dérivés partielles sont données par l'équation suivante :

$$\frac{\partial \ln(L)}{\partial \alpha_{tk}} = \sum_{i=1}^n \left[y_{ti} x_{ik} - \sum_{s=1}^m y_{si} \left(\frac{x_{ik} e^{\mathbf{x}'_i \alpha_s}}{1 + \sum_{j=1}^{m-1} e^{\mathbf{x}'_i \alpha_j}} \right) \right]$$

$$\begin{aligned}
&= \sum_{i=1}^n \left(y_{ti} x_{ik} - x_{ik} P_{ti} \sum_{s=1}^m y_{si} \right) \\
&= \sum_{i=1}^n x_{ik} (y_{ti} - P_{ti} T_i) .
\end{aligned}$$

Pour les dérivés secondes partielles, on a

$$\begin{aligned}
\frac{\partial^2 \ln(L)}{\partial \alpha_{ik}^2} &= \sum_{i=1}^n -(x_{ik} T_i) \left[\frac{x_{ik} e^{\mathbf{x}'_i \alpha_i} \left(1 + \sum_{j=1}^{m-1} e^{\mathbf{x}'_i \alpha_j} \right) - e^{\mathbf{x}'_i \alpha_i} \left(x_{ik} e^{\mathbf{x}'_i \alpha_i} \right)}{\left(1 + \sum_{j=1}^{m-1} e^{\mathbf{x}'_i \alpha_j} \right)^2} \right] \\
&= \sum_{i=1}^n -(x_{ik}^2 T_i) (P_{ti} - P_{ti}^2) \\
&= - \sum_{i=1}^n T_i x_{ik}^2 P_{ti} (1 - P_{ti}) ,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ln(L)}{\partial \alpha_{il} \partial \alpha_{ik}} &= \sum_{i=1}^n -(x_{ik} T_i) \left[\frac{x_{il} e^{\mathbf{x}'_i \alpha_i} \left(1 + \sum_{j=1}^{m-1} e^{\mathbf{x}'_i \alpha_j} \right) - e^{\mathbf{x}'_i \alpha_i} \left(x_{il} e^{\mathbf{x}'_i \alpha_i} \right)}{\left(1 + \sum_{j=1}^{m-1} e^{\mathbf{x}'_i \alpha_j} \right)^2} \right] \\
&= \sum_{i=1}^n -(x_{ik} x_{il} T_i) (P_{ti} - P_{ti}^2) \\
&= - \sum_{i=1}^n T_i x_{ik} x_{il} P_{ti} (1 - P_{ti}) .
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ln(L)}{\partial \alpha_{sk} \partial \alpha_{tk}} &= \sum_{i=1}^n -x_{ik} T_i e^{\mathbf{x}'_i \alpha_i} \left(\frac{-x_{ik} e^{\mathbf{x}'_i \alpha_s}}{\left(1 + \sum_{j=1}^{m-1} e^{\mathbf{x}'_i \alpha_j} \right)^2} \right) \\
&= \sum_{i=1}^n T_i x_{ik}^2 P_{ti} P_{si} .
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \ln(L)}{\partial \alpha_{sl} \partial \alpha_{tk}} &= \sum_{i=1}^n -x_{ik} T_i e^{\mathbf{x}'_i \alpha_i} \left(\frac{-x_{il} e^{\mathbf{x}'_i \alpha_s}}{\left(1 + \sum_{j=1}^{m-1} e^{\mathbf{x}'_i \alpha_j} \right)^2} \right) \\
&= \sum_{i=1}^n T_i x_{ik} x_{il} P_{ti} P_{si} .
\end{aligned}$$

Puisque $\hat{\alpha}$ est un estimateur de vraisemblance maximale, il est non-biaisé et sa matrice de variance-covariance est asymptotiquement $(-E(\mathbf{H}))^{-1}$. Asymptotiquement, $\hat{\alpha}_t$ suit une loi normale de moyenne $(\alpha_1, \dots, \alpha_{m-1})$ et de variance $(-E(\mathbf{H}))^{-1}$. Notons que les dérivés secondes ne dépendent pas de \mathbf{Y} et que la variable \mathbf{X} n'est pas une variable aléatoire, la matrice \mathbf{H} des dérivés secondes n'est donc pas stochastique et est égale à son espérance.

1.2 L'estimation individuelle

Une autre approche de l'estimation des paramètres du modèle multinomial de régression logistique a été proposée par Begg et Gray (1984). Au lieu d'estimer les $m-1$ vecteurs de paramètres α_t avec une seule fonction de vraisemblance, Begg et Gray ont utilisé $m-1$ modèles logistiques binomiaux pour estimer chacun des vecteurs α_t ($1 \leq t \leq m-1$) avec la catégorie de référence. Pour ce faire, ils ont posé $\theta_t = P(y_t = 1 | \mathbf{x}, y_m + y_t = 1)$ et $\theta_m = P(y_m = 1 | \mathbf{x}, y_m + y_t = 1)$ pour ($1 \leq t \leq m-1$). Chaque paire (θ_t, θ_m) de probabilité correspond à une expérience binomiale dans laquelle l'observation est dans la catégorie t avec une probabilité θ_t et dans la catégorie m avec une probabilité θ_m ($1 \leq t \leq m-1$). Le lien entre les probabilités θ_m et θ_t des $m-1$ modèles binomiaux et les probabilités p_t du modèle de régression logistique multinomiale est défini par les formules

$$\theta_t = \frac{p_t}{p_t + p_m}$$

et

$$\theta_m = 1 - \theta_t = \frac{p_m}{p_t + p_m} ,$$

ce qui entraîne :

$$\frac{\theta_t}{\theta_m} = \frac{p_t}{p_m}$$

d'où

$$\ln \left(\frac{\theta_t}{\theta_m} \right) = \ln \left(\frac{p_t}{p_m} \right) = \mathbf{X}'\alpha_t \quad (1 \leq t \leq m-1) .$$

Les $m-1$ modèles binomiaux utilisent donc les mêmes paramètres que le modèle multinomial.

L'estimation des paramètres α_t ($1 \leq t \leq m-1$) des $m-1$ modèles de régression logistique binomiale se fait par la méthode du maximum de vraisemblance. Si on définit comme dans la section précédente

$$\theta_{ti} = \frac{e^{x'_i \alpha_t}}{1 + e^{x'_i \alpha_t}},$$

avec ($1 \leq t \leq m-1$) et ($1 \leq i \leq n$), où n est la taille de l'échantillon, les $m-1$ fonctions de vraisemblance sont :

$$L_t = \prod_{C_t} \theta_{ti}^{y_{ti}} (1 - \theta_{ti})^{y_{mi}} \text{ où } C_t = \{i \mid y_{ti} + y_{mi} = 1\}.$$

Pour chacune d'entre elles, on doit trouver la valeur de α_t qui maximise L_t , ou, pour simplifier les calculs, qui maximise $\ln(L_t)$. Comme précédemment, on utilise l'algorithme de Newton-Raphson, pour lequel il faut calculer les dérivés partielles premières et secondes.

On trouve

$$\begin{aligned} \frac{\partial \ln(L_t)}{\partial \alpha_{tk}} &= \frac{\partial}{\partial \alpha_{tk}} \sum_{C_t} y_{ti} \ln \left(\frac{e^{x'_i \alpha_t}}{1 + e^{x'_i \alpha_t}} \right) + y_{mi} \ln \left(\frac{1}{1 + e^{x'_i \alpha_t}} \right) \\ &= \sum_{C_t} y_{ti} \left(x_{ik} - \frac{x_{ik} e^{x'_i \alpha_t}}{1 + e^{x'_i \alpha_t}} \right) + y_{mi} (-1) \left(\frac{x_{ik} e^{x'_i \alpha_t}}{1 + e^{x'_i \alpha_t}} \right) \\ &= \sum_{C_t} y_{ti} x_{ik} - y_{ti} x_{ik} \theta_{ti} - y_{mi} x_{ik} \theta_{ti} \\ &= \sum_{C_t} y_{ti} x_{ik} - (y_{ti} + y_{mi}) x_{ik} \theta_{ti} \\ &= \sum_{C_t} y_{ti} x_{ik} - x_{ik} \theta_{ti} \text{ puisque } y_{ti} + y_{mi} = 1 \forall i \in C_t \\ &= \sum_{C_t} x_{ik} (y_{ti} - \theta_{ti}). \end{aligned}$$

ainsi que

$$\begin{aligned} \frac{\partial^2 \ln(L_t)}{\partial \alpha_{tk}^2} &= \frac{\partial}{\partial \alpha_{tk}} \sum_{C_t} x_{ik} \left(y_{ti} - \frac{e^{x'_i \alpha_t}}{1 + e^{x'_i \alpha_t}} \right) \\ &= \sum_{C_t} -x_{ik} \left(\frac{x_{ik} e^{x'_i \alpha_t}}{1 + e^{x'_i \alpha_t}} + e^{x'_i \alpha_t} (-1) \frac{x_{ik} e^{x'_i \alpha_t}}{(1 + e^{x'_i \alpha_t})^2} \right) \\ &= - \sum_{C_t} x_{ik}^2 (\theta_{ti} - \theta_{ti}^2) \\ &= - \sum_{C_t} x_{ik}^2 \theta_{ti} (1 - \theta_{ti}) \end{aligned}$$

et

$$\begin{aligned}
\frac{\partial^2 \ln(L_t)}{\partial \alpha_{ik} \partial \alpha_{il}} &= \frac{\partial}{\partial \alpha_{il}} \sum_{C_t} x_{ik} \left(y_{li} - \frac{e^{x'_i \alpha_l}}{1 + e^{x'_i \alpha_l}} \right) \\
&= \sum_{C_t} -x_{ik} \left(\frac{x_{il} e^{x'_i \alpha_l}}{1 + e^{x'_i \alpha_l}} + e^{x'_i \alpha_l} (-1) \frac{x_{il} e^{x'_i \alpha_l}}{(1 + e^{x'_i \alpha_l})^2} \right) \\
&= - \sum_{C_t} x_{ik} x_{il} (\theta_{li} - \theta_{li}^2) \\
&= - \sum_{C_t} x_{ik} x_{il} \theta_{li} (1 - \theta_{li}) .
\end{aligned}$$

Comme il s'agit d'estimateurs du maximum de vraisemblance, on sait que les estimateurs individuels suivent asymptotiquement une distribution normale dont la moyenne est la vraie valeur des paramètres. Fredette (2001, p.12 à 22) rapporte différentes études visant à comparer l'efficacité relative (le rapport des variances) des estimateurs obtenus par la régression logistique multinomiale et les estimateurs provenant des modèles binomiaux (voir notamment Bull et Donner (1993) et Bull, Greenwood et Donner (1994)). Les résultats varient beaucoup selon la configuration des paramètres, il est donc difficile d'en faire un résumé général. Dans la plupart des cas, comme on pourrait s'y attendre, les estimateurs de la régression logistique multinomiale sont plus efficaces, puisqu'ils utilisent toute l'information disponible. Cependant, pour certains sous-ensembles de paramètres, l'efficacité relative est la même pour les deux estimateurs. Il est également intéressant de noter que généralement l'efficacité relative augmente lorsque la catégorie de référence a une probabilité élevée.

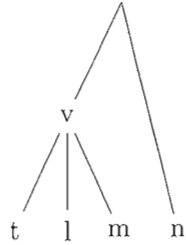
CHAPITRE II

ARBORESCENCES DE RÉGRESSION LOGISTIQUE

Le chapitre précédent rend compte du cas où les résultats observés proviennent pour chaque valeur de x_i d'une seule expérience multinomiale avec m catégories $M(T_i; P_{1i}, \dots, P_{mi})$. Rousseau et Sankoff (1989) ont considéré le cas où les résultats observés proviennent d'une séquence de plus petites expériences successives et/ou parallèles, que l'on décrit alors au moyen d'arborescences.

Afin de comprendre facilement les modèles d'arborescences de régression logistique, nous allons d'abord décrire à l'aide d'un exemple un modèle d'arborescence de régression logistique.

Supposons qu'il y ait quatre catégories étiquetées n, t, l, m où n est la catégorie des individus qui appliquent une règle de grammaire A dans un contexte décrit par un vecteur \mathbf{X} de variable explicatives, et t, l, m sont trois façons de contourner la règle de grammaire. On suppose que dans le contexte décrit par ce vecteur \mathbf{X} , l'individu décide, dans un premier temps et avec probabilité $p(n)$, s'il applique la règle de grammaire ou décide, avec probabilité $p_v = 1 - p(n)$, de ne pas l'appliquer. Si l'individu décide d'appliquer la règle de grammaire alors l'expérience s'arrête; s'il décide de ne pas appliquer la règle de grammaire, alors dans une deuxième étape il décide avec probabilités $p_t = p(t|v)$, $p_l = p(l|v)$, $p_m = p(m|v)$ de choisir la variante t, l ou m respectivement, étant donné que l'on a observé la non-application de la règle de grammaire. Ce modèle de décision peut être représenté par l'arborescence suivante :



où la probabilité $p(t)$ d'observer la catégorie t est $p(t) = p_v p_t$; de même pour les catégories l et m . De plus, les probabilités $p(n)$, p_v , p_t , p_l , p_m suivent les équations de régression logistique suivantes

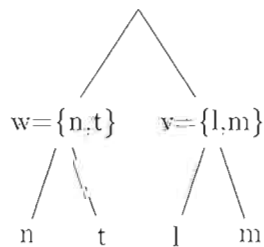
$$\ln \left(\frac{p(n)}{p_v} \right) = \alpha_1' \mathbf{X} ,$$

$$\ln \left(\frac{p_t}{p_m} \right) = \alpha_2' \mathbf{X} ,$$

$$\ln \left(\frac{p_l}{p_m} \right) = \alpha_3' \mathbf{X} .$$

Il est évident que si le vecteur \mathbf{X} est une constante alors ce modèle revient au modèle logistique standard. Nous prouverons que le vecteur \mathbf{X} doit être composé d'au moins 2 variables explicatives afin que ces modèles d'arborences ne décrivent pas le même type d'expérience.

Un autre modèle d'arborescence aurait pu être celui où l'individu décide de choisir dans un premier temps l'ensemble des catégories n et t ou l'ensemble des catégories l et m . L'arborescence qui décrit ce modèle est :



avec les équations

$$\ln \left(\frac{p_w}{p_v} \right) = \alpha_1' \mathbf{X} ,$$

$$\ln \left(\frac{p_t}{p_n} \right) = \alpha_2' \mathbf{X} ,$$

$$\ln \left(\frac{p_l}{p_m} \right) = \alpha_3' \mathbf{X} .$$

Nous pouvons prouver facilement qu'avec 4 catégories, il y a 26 modèles d'arborescences différents.

Nous allons maintenant décrire d'une façon générale les modèles d'arborescences de régression logistique pour m catégories, avec $m > 2$. Pour permettre de présenter un résumé détaillé des résultats de Rousseau et Sankoff, commençons par introduire leur notation qui rendra plus précise la description des modèles et des processus d'inférence dans le cas général de m catégories.

On a donc comme dans le cas général pour chaque cellule (on entend par cellule une valeur du vecteur de variables indépendantes et le vecteur de variables dépendantes qui lui est associé) un vecteur \mathbf{X} de variables explicatives et un vecteur $(Y(1), \dots, Y(m))$ de résultats finaux avec $\sum_{t=1}^m Y(t) = T$. Les séquences possibles de multinomiales successives ou parallèles (permettant de faire le choix d'une catégorie parmi les m catégories finales) sont représentées par une arborescence. Chaque arborescence a une racine V_0 avec $m_0 \geq 2$ branches incidentes auxquelles sont rattachés, pour chacune, un noeud V qui est soit un noeud terminal (une feuille) ou un autre noeud non terminal avec $m_V \geq 2$ autres branches incidentes. Chaque noeud V correspond à un sous-ensemble S_V qui contient les catégories pouvant être choisies si on se trouve au noeud V , et si le noeud V_k a pour ancêtre le noeud V_l , alors $S_{V_k} \subset S_{V_l}$. On a aussi que pour deux noeuds quelconques V_k et V_l , on a

$$S_{V_k} \cap S_{V_l} = \phi ,$$

ou

$$S_{V_k} \subset S_{V_l} ,$$

ou

$$S_{V_l} \subset S_{V_k} .$$

Dans le premier cas, V_k n'est ni descendant ni ancêtre de V_l , dans le second, V_k est un descendant de V_l et dans le dernier, V_k est ancêtre de V_l . On associe à chaque noeud non terminal V un ensemble de $m_V - 1$ équations de la forme de l'équation (1.1) permettant de calculer les probabilités m_V -nomiales $(P^V(1), \dots, P^V(m_V))$ dépendant de \mathbf{X} . Explicitement, on aurait donc pour le noeud V , en supposant que m_V est la catégorie de référence, les équations suivantes :

$$\ln \left(\frac{P^V(j)}{P^V(m_V)} \right) = \mathbf{X}' \boldsymbol{\alpha}_j^V \quad (1 \leq j \leq m_V - 1)$$

À la racine de l'arborescence, pour chaque \mathbf{X} , une expérience multinomiale $M_{m_0}(T; P^0(1), \dots, P^0(m_0))$ est menée, qui permet de choisir entre m_0 sous-ensembles de catégories finales. Ces sous-ensembles sont mutuellement exclusifs et leur union couvre toutes les catégories. Le nombre de fois que S_j , le j -ème de ces sous-ensemble, est observé est noté $Y^0(j)$ ($1 \leq j \leq m_0$). Si le sous-ensemble S_j contient une seule catégorie, disons la catégorie t , alors $Y^0(j) = Y(t)$, le nombre de fois que la catégorie t a été observée, et on associe S_j à une feuille de l'arborescence, notée t . Si S_j contient plus d'une catégorie, alors $Y^0(j) = \sum_{t \in S_j} Y(t)$ et on lui associe un noeud non-terminal noté V_j . Une nouvelle expérience multinomiale $M(T^V; P^V(1), \dots, P^V(m_V))$ est alors menée en V_j , avec $T^V \equiv Y^0(j)$ et où m_V correspond au nombre de noeuds colinéaires à V_j , c'est-à-dire au nombre de sous-ensembles inclus dans S_j (mutuellement exclusifs et dont l'union couvre S_j) entre lesquels le choix s'effectue.

On procède de cette façon pour tous les noeuds non terminaux de l'arborescence, c'est-à-dire que pour tout noeud non terminal V (correspondant à un sous-ensemble S_V de catégories) . une expérience multinomiale $M_{m_V}(T^V; P_1^V, \dots, P_{m_V}^V)$ est menée, pour laquelle le nombre d'essais T^V a été déterminée par les résultats d'une séquence d'expériences préalables et dont les résultats $(Y^V(0), \dots, Y^V(m_V))$ sont associés chacun à

une branche incidente. Ces résultats sont appelés résultats intermédiaires et sont notés $Y^V(j)$ ($1 \leq j \leq m_V$). Ils correspondent chacun à un sous-ensemble S_j de S_V (comme précédemment, les sous-ensembles S_j sont disjoints et leur union recouvre S_V). $Y^V(j)$ est le nombre de fois que le sous-ensemble S_j est observé et $Y^V(j) = \sum_{t \in S_j} Y(t)$, $Y(t)$ (sans exposant) étant la notation pour indiquer le nombre de fois que la catégorie t ($1 \leq t \leq m$) a été observée. On note de même $P^V(j)$, pour la probabilité intermédiaire d'observer le sous-ensemble S_j , alors que $P(t)$ est la probabilité finale d'observer la catégorie t , et $T^V = \sum_{t \in S_V} Y(t)$ pour le nombre d'essais au noeud V , alors que T est le nombre d'essais global. Chaque expérience multinomiale à un noeud de l'arborescence correspond à un choix entre des sous-ensembles de catégories finales, et ces sous-ensembles sont de plus en plus petits comme on descend dans l'arborescence. À la fin, il ne reste qu'une catégorie, disons t , dans le sous-ensemble et ce dernier est associé à un noeud terminal.

Notons que chaque résultat final provient, selon un chemin unique, de résultats intermédiaires et que la probabilité d'observer la catégorie t est, comme dans l'exemple précédent, le produit des probabilités d'observer les résultats intermédiaires. Une autre propriété des arborescences est que, pour m catégories, le nombre total de paramètres du modèle est le même pour chaque arborescence. En effet, on a :

$$m = 1 + \sum_{\text{noeud } V} (m_V - 1),$$

et on sait que le nombre de paramètres de l'arborescence (de l'ensemble du modèle) est

$$\sum_{\text{noeud } V} (m_V - 1)(d + 1)$$

si \mathbf{X} , le vecteur des variables indépendantes, est de dimension $(d + 1) \times 1$.

Les auteurs ont montré que les probabilités des résultats finaux de deux modèles (c'est-à-dire de deux arborescences) différents ne peuvent pas être exactement les mêmes pour presque tout l'espace des paramètres, ce qui est prouvé plus loin dans le théorème (1) dont la preuve requiert le lemme suivant :

Lemme 1 *Soient S et T , deux arborescences identiques ayant la racine V_0 . Supposons que S est une sous-arborescence de S' et T une sous-arborescence de T' (S' et T' sont des*

arborescences différentes) et que p_{V_0} et q_{V_0} sont respectivement les probabilités d'observer le résultat intermédiaire correspondant à V_0 dans S' et T' . Soient $\mathbf{X}_1, \dots, \mathbf{X}_n$, les n vecteurs de variables indépendantes du plan d'expérience. Alors si $p_{V_0} \neq q_{V_0}$ pour au moins un \mathbf{X}_i et que la matrice $(\mathbf{X}_1 \dots \mathbf{X}_n)$ est de rang $d + 1$, il existe une catégorie t pour laquelle les probabilités $p(t)$ et $q(t)$ associées à la feuille t dans S' et T' ne peuvent être identiquement égales pour tous les \mathbf{X}_i .

Preuve : Soit V_0, V_1, \dots, V_s , le chemin allant de V_0 à la feuille V_s dans S et T , où V_s est associée au résultat s . Pour $j = 1, \dots, s$, on dénote respectivement par p_j et q_j les probabilités d'obtenir, dans les arborescences S et T , le résultat intermédiaire correspondant à V_j , sachant qu'on a obtenu le résultat correspondant à V_{j-1} . Les probabilités d'observer la catégorie s dans les arborescences S' et T' sont alors

$$p(s) = p_{V_0} p_1 \dots p_s \quad (2.1)$$

$$q(s) = q_{V_0} q_1 \dots q_s \quad (2.2)$$

On va démontrer que $p_j = q_j$ $j = 1, \dots, s$ sous les conditions du lemme, et donc que $q_{V_0} = p_{V_0}$, ce qui contredit l'hypothèse.

Soit s et t , deux catégories correspondant aux feuilles terminales V_s et V_t ayant le même ancêtre commun immédiat V_{s-1} (de telles feuilles existent nécessairement). Sans perte de généralité, on peut écrire les équations logistiques associés à V_{s-1} de la façon suivante, en utilisant (2.1) et (2.2) :

$$\ln\left(\frac{p(t)}{p(s)}\right) = \ln\left(\frac{p_1 \dots p_{s-1} p_t}{p_1 \dots p_{s-1} p_s}\right) = \ln\left(\frac{p_t}{p_s}\right) = \mathbf{X}_i' \alpha_t$$

et

$$\ln\left(\frac{q(t)}{q(s)}\right) = \ln\left(\frac{q_1 \dots q_{s-1} q_t}{q_1 \dots q_{s-1} q_s}\right) = \ln\left(\frac{q_t}{q_s}\right) = \mathbf{X}_i' \beta_t .$$

Si $p(t)$ est identiquement égal à $q(t)$ pour tout t et pour tout \mathbf{X}_i , on a $\alpha_t = \beta_t$, puisque la matrice $(\mathbf{X}_1 \dots \mathbf{X}_n)$ est de rang $d + 1$. On peut montrer de même que pour toute autre catégorie u ayant pour ancêtre immédiat V_{s-1} , $\alpha_u = \beta_u$. On a donc que $p_s = q_s$, que $p_t = q_t$, et que, pour toute catégorie u ayant pour ancêtre immédiat V_{s-1} , $p_u = q_u$.

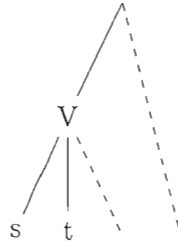
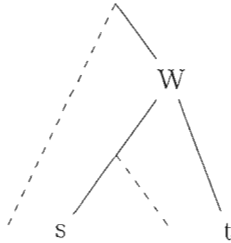
L'argument peut être répété pour tous les ancêtres immédiats des noeuds terminaux (les avant-derniers noeuds), puis pour tous les noeuds intermédiaires en remontant dans l'arborescence. (Puisqu'en effet, si V est un noeud avant-dernier $p_V = \sum_{t \in S_V} p(t) = \sum_{t \in S_V} q(t) = q_V$.) On a alors que $p_j = q_j$ $j = 1, \dots, s$, et donc que p_{V_0} doit être égal à q_{V_0} , ce qui contredit l'hypothèse. \square

Avec ce lemme, on peut maintenant présenter et démontrer le premier résultat important trouvé par les auteurs, qui fait l'objet du théorème suivant :

Théorème 1 *Soient S et T , deux arborescences différentes, et soient $p(t)$ et $q(t)$ ($1 \leq t \leq m$), les probabilités que le résultat t soit observé dans les arborescences S et T respectivement, lesquelles probabilités sont déterminées par les équations de la régression logistique (de la forme de l'équation (1.1)), avec respectivement les paramètres α_t et β_t . Alors si n , le nombre de cellules différentes X_1, \dots, X_n est plus grand que $d + 1$ et si le rang de la matrice $(X_1 \dots X_n)$ est $d + 1$, on a que pour au moins une cellule X_i et au moins une catégorie t , $p(t) \neq q(t)$ pour presque toutes les valeurs que peuvent prendre α_t et β_t .*

Preuve : Supposons qu'il existe un noeud V dans l'arborescence S ayant uniquement des feuilles comme descendants. Parmi lesquelles deux correspondent aux catégories s et t . Supposons également que W , l'ancêtre commun des catégories s et t dans l'arborescence T n'est pas un ancêtre immédiat pour au moins une des feuilles associées à s et t . S'il n'existe pas de noeud V dans S et W dans T satisfaisant ces conditions, on tronque l'arborescence de tout noeud V n'ayant que des feuilles comme descendants en remplaçant ce noeud V et ses descendants par une nouvelle feuille s^* . On obtient alors une nouvelle arborescence S^* . On procède de la même façon dans T . En continuant ainsi, on devrait à la longue trouver un noeud V satisfaisant aux conditions. autrement S et T auraient la même structure.

Supposons, sans perte de généralité, que les feuilles de V correspondent aux catégories $1, \dots, r$, parmi lesquelles se trouvent s et t . On a les équations suivantes pour le noeud

Figure 2.1 Arborescence S Figure 2.2 Arborescence T

V :

$$\ln \left(\frac{p(k)}{p(r)} \right) = \mathbf{X}_i' \alpha_k \quad k = 1, \dots, r-1 \quad (2.3)$$

Les chemins dans l'arborescence T (ou T^*) menant du noeud W aux feuilles s et t sont respectivement W, V_1, \dots, V_s et W, W_1, \dots, W_t . La manière dont s et t ont été choisis assure qu'un de ces chemins comporte au moins un noeud non terminal entre W et s ou entre W et t . Considérons le chemin allant de W à la feuille s . Notons par q_1 , la probabilité d'observer V_1 sachant qu'on a observé W , et par q_j ($2 \leq j \leq s$), les probabilités d'observer V_j sachant que l'on a observé V_{j-1} . Considérons le chemin allant de W à la feuille t . Notons par q'_j ($1 \leq j \leq t$), les probabilités d'observer W_j sachant que

l'on a observé W_{j-1} (en considérant W comme W_0). On a

$$\ln\left(\frac{q(s)}{q(t)}\right) = \ln\left(\frac{q_1 \cdots q_s}{q'_1 \cdots q'_t}\right) \quad (2.4)$$

$$= \ln\left(\frac{q_1}{q'_1}\right) + \sum_{j=2}^s \ln(q_j) - \sum_{j=2}^t \ln(q'_j) \quad (2.5)$$

$$= \mathbf{X}'_i \beta_1 + \sum_{j=2}^s \ln(q_j) - \sum_{j=2}^t \ln(q'_j)'. \quad (2.6)$$

L'équation(2.6) est obtenue à partir de l'équation (2.5) en utilisant l'égalité

$$\ln\left(\frac{q_1}{q'_1}\right) = \mathbf{X}'_i \beta_1 .$$

En effet, V_1 et W_1 ont tous deux W comme ancêtre immédiat et on considère (sans perte de généralité) que W_1 est le noeud de référence pour les équations logistiques associées au noeud W . Remarquons aussi qu'à cause de la façon dont s et t ont été choisis, au moins une des deux sommations de (2.6) est non vide et qu'il y a donc un terme non-linéaire.

Supposons maintenant que $p(t) = q(t)$ pour tout t ($1 \leq t \leq m$) et pour tout \mathbf{X}_i . Supposons aussi que la feuille t joue le rôle de la catégorie de référence pour les équations logistiques associées au noeud V (équations du type de l'équation(2.3)) dans l'arborescence S (ou S^*), s étant une des catégories k ($k = 1, \dots, r-1$). On aurait alors, en utilisant (2.6) et puisqu'on vient de supposer que

$$\ln\left(\frac{p(s)}{p(t)}\right) = \ln\left(\frac{q(s)}{q(t)}\right) ,$$

que

$$\mathbf{X}'_i \alpha_s = \mathbf{X}'_i \beta_1 + \sum_{j=1}^s \ln(q_j) - \sum_{j=1}^t \ln(q'_j)'. \quad (2.7)$$

d'où

$$\mathbf{X}'_i (\alpha_s - \beta_1) = \sum_{j=1}^s \ln(q_j) - \sum_{j=1}^t \ln(q'_j)'. \quad (2.7)$$

Pour que l'égalité soit vraie pour tout $\mathbf{X}_i \neq 0$, il faut que le membre de droite soit égal à 0, puisqu'il n'est pas linéaire et il faut aussi par conséquent que $\alpha_s = \beta_1$. Cette condition n'est pas satisfaite pour presque tout l'espace des paramètres, puisque q_j et

q'_j dépendent de paramètres α différents. Donc il existe au moins une catégorie t et au moins un \mathbf{X}_i pour lesquels $p(t) \neq q(t)$ pour les arborescences S et T (ou S^* et T^*).

Finalement, si on a créé des arborescences S^* et T^* pour lesquelles le théorème est vrai, alors il existe une catégorie t (qui est une feuille de l'arborescence S^* et de l'arborescence T^*) pour laquelle $p(t) \neq q(t)$. Si cette catégorie de S^* et T^* est aussi une catégorie finale de S et T , on a le résultat souhaité; et si cette catégorie a été obtenue en remplaçant un noeud et ses descendants dans S et T , alors on se trouve dans les conditions du lemme (1) (avec ce noeud jouant le rôle de V_0) et la véracité du théorème pour les arbres S^* et T^* assure sa véracité pour les arbres S et T . \square

2.1 Estimation des paramètres

Avant de présenter le deuxième résultat important de Rousseau et Sankoff, il nous faut considérer la manière dont l'estimation des paramètres est effectuée. À partir de l'information disponible, c'est-à-dire à partir des valeurs prises par les vecteurs \mathbf{X}_i ($1 \leq i \leq n$) des variables indépendantes et les vecteurs \mathbf{Y}_i ($1 \leq i \leq n$) du nombre d'observations par catégorie pour chaque cellule, il faut estimer la structure de l'arborescence et la valeur des paramètres α_t . Supposons que le nombre d'observations par catégorie (Y_{i1}, \dots, Y_{im}) ($1 \leq i \leq n$) de la variable \mathbf{Y}_i est le résultat d'une séquence de multinomiales parallèles et/ou successives déterminée par une même arborescence A connue et identique pour chaque i ($1 \leq i \leq n$). Nous allons définir une fonction $f_A(\mathbf{Y}_i)$ qui a la même forme pour chaque i . Pour simplifier la notation, nous n'utiliserons pas l'indice i dans ce qui suit.

Définissons donc, pour une arborescence A dont la structure est connue, la fonction

$$f_A(\mathbf{Y}) = \prod_{(V \text{ noeud non terminal de } A)} f(V)$$

où $f(V)$ est la fonction de masse (multinomiale) au noeud V , soit

$$f(V) = \binom{T^V}{Y^V(1) \dots Y^V(m_V)} p^V(1)^{Y^V(1)} \dots p^V(m_V)^{Y^V(m_V)}$$

On peut voir facilement que les termes $\binom{T^V}{Y^V(1) \dots Y^V(m_V)}$ dans $f_A(\mathbf{Y})$ vont se simplifier en un seul facteur $\binom{T}{Y(1) \dots Y(t)}$, où T est le nombre total d'essais et $Y(t)$, ($1 \leq t \leq m$),

le nombre d'observations pour la catégorie t . Ainsi, l'expression peut s'écrire de la façon suivante :

$$f_A(\mathbf{Y}) = \binom{T}{Y(1) \dots Y(t)} \prod_{(V \text{ noeud non terminal de } A)} p^V(1)^{Y^V(1)} \dots p^V(m_V)^{Y^V(m_V)}$$

Le lemme suivant devrait nous permettre de montrer que $f_A(\mathbf{Y})$ est la fonction de probabilité d'une multinomiale où les paramètres sont les probabilités finales et les nombres finaux d'observations par catégorie.

Lemme 2 *Soit, dans l'arborescence A , pour une feuille t quelconque ($1 \leq t \leq m$) le chemin $V_0 \dots V_u$ allant de la racine V_0 à la feuille t (t serait le noeud V_{u+1}). Si on définit la probabilité finale de la catégorie t de la façon suivante :*

$$p(t) = p^{V_0}(t) \dots p^{V_u}(t) ,$$

où $p^{V_j}(t)$ ($0 \leq j \leq u$) est la probabilité d'observer V_{j+1} sachant qu'on a observé V_j , alors :

$$f(A) = \binom{T}{Y(1) \dots Y(t)} \prod_{t=1}^m p(t)^{Y(t)} ,$$

c'est-à-dire que pour l'arborescence A

$$\prod_{(V \text{ noeud non terminal de } T)} p^V(1)^{Y^V(1)} \dots p^V(m_V)^{Y^V(m_V)} = \prod_{t=1}^m p(t)^{Y(t)} .$$

Preuve : Nous allons démontrer que

$$\ln \left(\prod_{(V \text{ noeud non terminal de } T)} p^V(1)^{Y^V(1)} \dots p^V(m_V)^{Y^V(m_V)} \right) = \ln \left(\prod_{t=1}^m p(t)^{Y(t)} \right) ,$$

c'est-à-dire que

$$\sum_V \sum_{j=1}^{m_V} Y^V(j) \ln(p^V(j)) = \sum_{t=1}^m Y(t) \ln p(t) ,$$

Notons $q_V = \prod_{j=1}^{m_V} p^V(j)^{Y^V(j)}$. On a alors

$$\ln(q_V) = \sum_{j=1}^{m_V} Y^V(j) \ln(p^V(j)) . \quad (2.8)$$

Nous allons donc démontrer que :

$$\sum_V \ln(q_V) = \sum_{t=1}^m Y(t) \ln p(t) .$$

Soit V_k un noeud de l'arborescence A n'ayant que des feuilles comme descendant (un tel V_k existe toujours). Sans perte de généralité, supposons que ces feuilles correspondent aux r catégories $1, \dots, r$ pour $r \geq 2$. Alors

$$\begin{aligned} \sum_{t=1}^r Y(t) \ln(p(t)) &= \sum_{t=1}^r Y(t) \ln \left(\frac{p(t)}{p^{V_k}(t)} \right) + \sum_{t=1}^r Y(t) \ln(p^{V_k}(t)) \\ &= \sum_{t=1}^r Y(t) \ln \left(\frac{p^{V_0}(t) \dots p^{V_k}(t)}{p^{V_k}(t)} \right) + \ln(q_{V_k}) \\ &= \sum_{t=1}^r Y(t) \ln (p^{V_0}(t) \dots p^{V_{k-1}}(t)) + \ln(q_{V_k}) \\ &= \ln (p^{V_0}(r) \dots p^{V_{k-1}}(r)) \sum_{t=1}^r Y(t) + \ln(q_{V_k}) , \end{aligned}$$

car pour tous les descendants de V_k , $p^{V_j}(t) = p^{V_j}(s)$ pour $(1 \leq s, t \leq r)$ et $(0 \leq j \leq k-1)$, et donc $\ln (p^{V_0}(t) \dots p^{V_{k-1}}(t))$ a la même valeur pour tout t ($1 \leq t \leq r$). On a ensuite,

$$\sum_{t=1}^r Y(t) \ln(p(t)) = Y^{(V_{k-1})}(k) \ln (p^{V_0}(r) \dots p^{V_{k-1}}(r)) + \ln(q_{V_k})$$

où $Y^{(V_{k-1})}(k)$ est le nombre de résultats intermédiaires observés à la suite de l'expérience multinomiale menée au noeud V_{k-1} pour le sous-ensemble correspondant à V_k .

On a donc

$$\sum_{t=1}^m Y(t) \ln(p(t)) = \sum_{t>r} Y(t) \ln(p(t)) + Y(t^*) \ln(p(t^*)) + \ln(q_{V_k})$$

où t^* est la feuille de l'arborescence A^* construite à partir de A en remplaçant V_k et ses r descendants par t^* . Donc si le lemme est vrai pour l'arborescence A^* , il doit être vrai pour A . On peut répéter le même argument pour A^* (qui est une arborescence avec moins de feuilles), puis pour la nouvelle arborescence obtenue à partir de A^* et ainsi de suite en obtenant des arborescences avec de moins en moins de feuilles à chaque fois, jusqu'à ce que toutes les branches aient été enlevées de la somme et remplacées par des

$\ln(q_V)$. On a donc que

$$\sum_{t=1}^m Y(t) \ln p(t) = \sum_V \ln(q_V)$$

c'est-à-dire que

$$f_A(\mathbf{Y}) = \binom{T}{Y(1) \dots Y(t)} \prod_{t=1}^m p(t)^{Y(t)}$$

où $p(t) = p^{V_0}(t) \dots p^{V_u}(t)$ ($1 \leq t \leq m$) \square

La fonction $f(A)$ est donc la fonction de masse d'une multinomiale et peut être utilisée pour estimer les paramètres par la méthode du maximum de vraisemblance. Ainsi, la fonction de vraisemblance de l'arborescence A est, en utilisant le lemme (2)

$$L(A) = \prod_{i=1}^n \binom{T_i}{Y(1i) \dots Y(ti)} \prod_{t=1}^m p_i(t)^{Y_i(t)}, \quad (2.9)$$

c'est-à-dire que le logarithme de la fonction de masse de l'arborescence est, à une constante près,

$$\ln(L(A)) = \sum_{i=1}^n \sum_{t=1}^m Y_i(t) \ln p_i(t) = \sum_{i=1}^n \sum_V \ln(q_{iV}) = \sum_V \sum_{i=1}^n \ln(q_{iV})$$

où $\ln(q_{iV})$ est donné par l'équation (2.8). Les paramètres α sont estimés en maximisant $\sum_i \ln(q_{iV})$ pour chaque V , car les paramètres à un noeud V^s sont différents des paramètres à un noeud V^t pour $s \neq t$. Les propriétés des estimateurs de vraisemblance maximale des α , notés $\hat{\alpha}$, sont telles que présentées dans la section sur la régression logistique multinomiale.

Si on ne connaît pas la structure de l'arborescence qui a généré les données, on calcule la vraisemblance maximale pour chaque arborescence possible. On choisit le modèle d'arborescence pour lequel la vraisemblance est maximale.

Introduisons le corollaire suivant pour présenter le deuxième théorème important de Rousseau et Sankoff.

Corollaire 1 *Soit une arborescence A avec m feuilles, le chemin $V_0 \dots V_u$ allant de la racine V_0 à la feuille t (t serait le noeud V_{u+1}) et les probabilités*

$$p(t) = p^{V_0}(t) \dots p^{V_u}(t)$$

d'obtenir la catégorie t tels que définis dans le lemme (2). Soient $\hat{p}^{V_j}(t)$ ($1 \leq j \leq u$) et ($1 \leq t \leq m$), les estimateurs du maximum de vraisemblance de $p^{V_j}(t)$. Alors, en utilisant le lemme (2) et la définition de l'estimateur du maximum de vraisemblance, l'estimateur du maximum de vraisemblance de $p(t)$ est :

$$\hat{p}(t) = \hat{p}^{V_0}(t) \dots \hat{p}^{V_u}(t)$$

et

$$\max \ln(L(A)) = \sum_{i=1}^n \sum_{t=1}^m Y_i(t) \ln \hat{p}_i(t) + C$$

où C est la constante obtenue à cause des termes combinatoires figurant dans l'équation (2.9).

Ce qui amène le résultat suivant :

Théorème 2 Soit une arborescence A qui détermine un vecteur de résultats de dimension m . Pour la i -ème cellule du plan d'expérience, soit $v_i > 0$ ($1 \leq i \leq n$) tel que $\sum_{i=1}^n v_i = 1$. Si $T = \sum_{i=1}^n T_i$ est la taille totale de l'échantillon, et T_i est proportionnel à T , c'est-à-dire $T_i = v_i T$ ($1 \leq i \leq n$), alors pour toute arborescence B différente de A et pour tout n suffisamment grand, $\lim_{T \rightarrow \infty} T^{-1} (\max \ln(L(A)) - \max \ln(L(B))) = C_B$ pour une constante $C_B > 0$ avec probabilité 1.

Preuve : Nous allons démontrer pour commencer que lorsque que $T \rightarrow \infty$. les estimateurs définis comme dans le lemme(2) obtenus à partir de la vraie arborescence

$$p(t) = p^{V_0}(t) \dots p^{V_u}(t) \rightarrow p(t)$$

pour toute catégorie t ($1 \leq t \leq m$) et pour toute cellule i ($1 \leq i \leq n$). Pour prouver ceci, nous aurons besoin du fait que $T^V \rightarrow \infty$ lorsque $T \rightarrow \infty$, pour pouvoir utiliser les propriétés de convergence asymptotique des estimateurs du maximum de vraisemblance $\hat{p}^V(t)$.

Soit $M_{m_0}(T_i; p_i^{V_0}(1), \dots, p_i^{V_0}(m_0))$. l'expérience multinomiale associée avec la racine de

l'arborescence pour la cellule i . Puisque $T_i = v_i T$, $T_i \rightarrow \infty$ si $T \rightarrow \infty$ et donc

$$\lim_{T \rightarrow \infty} \frac{Y_i^{V_0}(j)}{T_i} = p_i^{V_0}(j)$$

avec probabilité 1 pour tout j ($1 \leq j \leq m_0$) et pour tout i ($1 \leq i \leq n$). Si le j -ème noeud descendant directement de la racine est un noeud non terminal V , alors $T_i^V = Y_i^{V_0}(j)$ tend vers ∞ puisqu'on a supposé $p_i^{V_0}(j) > 0$ et $T_i \rightarrow \infty$. Donc, pour tout j ($1 \leq j \leq m_V$) et pour tout i ($1 \leq i \leq n$), on a

$$\lim_{T \rightarrow \infty} \frac{Y_i^V(j)}{T_i^V} = p_i^V(j)$$

et ainsi de suite pour tout noeud non terminal de l'arborescence.

Donc, pour toute cellule i ($1 \leq i \leq n$) et pour toute catégorie t ($1 \leq t \leq m$),

$$\begin{aligned} \lim_{T \rightarrow \infty} \hat{p}_i(t) &= \prod_{j=0}^u \lim_{T_i \rightarrow \infty} \hat{p}_i^{V_j}(t) \\ &= p_i^{V_0}(t) \dots p_i^{V_u}(t) \\ &= p_i(t) . \end{aligned}$$

Considérons maintenant l'estimation effectuée avec les mêmes données, mais en utilisant une autre structure (incorrecte) d'arborescence B . Comme $T \rightarrow \infty$, les estimés de tous les paramètres vont converger même si les données ont été générées par l'arborescence B . Notons $\hat{q}_i(t)$ et $q_i(t)$ les estimateurs et la limite de ces estimateurs lorsque $T \rightarrow \infty$ pour les probabilités finales des catégories 1 à m pour la cellule i . Alors :

$$\begin{aligned} \lim_{T \rightarrow \infty} T^{-1}(\max \ln(L(A)) - \max \ln(L(B))) &= \lim_{T \rightarrow \infty} \sum_{i=1}^n \frac{v_i}{v_i T} \sum_{t=1}^m Y_i(t) (\ln(\hat{p}_i(t)) - \ln(\hat{q}_i(t))) \\ &= \sum_{i=1}^n v_i \sum_{t=1}^m \lim_{T \rightarrow \infty} \frac{Y_i(t)}{T_i} \left(\ln \left(\frac{\hat{p}_i(t)}{\hat{q}_i(t)} \right) \right) \\ &= \sum_{i=1}^n v_i \sum_{t=1}^m p_i(t) \ln \left(\frac{p_i(t)}{q_i(t)} \right) \\ &= C_B \end{aligned}$$

On a que $C_B > 0$ puisqu'on sait, par le théorème (1), que les $q_i(t)$ ne sont pas identiques

aux $p_i(t)$ pour tout i et pour tout t , et l'unique minimum de

$$\sum_{t=1}^m p_i(t) \ln \left(\frac{p_i(t)}{q_i(t)} \right)$$

est zéro et est atteint quand $q_i(t) \equiv p_i(t)$ ($1 \leq t \leq m$) et ($1 \leq i \leq n$). \square

2.2 Études de simulation

On sait par le théorème (2) que la vraisemblance de la vraie arborescence est asymptotiquement plus grande que celle de n'importe quelle autre arborescence pour presque tout l'espace des paramètres. Pour un nombre d'essais T suffisamment grand, la probabilité P de choisir la vraie arborescence devrait donc être asymptotiquement proche de 1 pour presque tous les paramètres. Pour déterminer quelle doit être la grandeur de T pour que P soit suffisamment proche de 1 et comment P est affectée par les autres caractéristiques du plan d'expérience, les auteurs ont mené des études de simulation.

Les caractéristiques étudiées étaient le nombre de variables explicatives, la dispersion des valeurs prises par les paramètres et la taille de l'échantillon. Chacune de ces caractéristiques a été étudiée séparément, afin de mieux mettre en évidence l'effet de chacune d'elle, les autres étant maintenues fixes. Les études ont porté sur le cas 4-nomial, qui a l'avantage d'offrir suffisamment de variation dans la structure possible des arborescences sans être trop lourd quant aux calculs. Comme il a été mentionné précédemment, il y a 26 arborescences différentes pour le cas 4-nomial, cependant ces arborescences n'ont en fait que les 5 structures possibles illustrées par les figures (2.3), (2.4), (2.5), (2.6) et (2.7); les 26 arborescences différentes étant obtenues par les différentes positions occupées dans l'arborescence par chacun des résultats A, B, C, D. (On considère que deux arborescences sont différentes si les différentes positions occupées par un résultat correspondent à une différence dans la séquence d'expériences qui permet d'arriver à ce résultat.) On peut donc effectuer les simulations uniquement pour les 5 structures différentes d'arborescences.

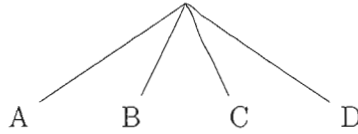


Figure 2.3 Structure d'arborescence pour une expérience 4-nomiale

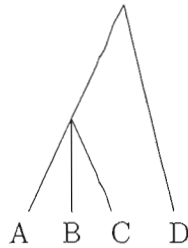


Figure 2.4 Structure d'arborescence pour une expérience binomiale suivie par une trinomial

2.2.1 Effet du nombre de variables explicatives

Cette section présente les résultats obtenus par les études de simulation de Rousseau et Sankoff en faisant varier d , le nombre de variables explicatives. On sait que s'il n'y a pas de variables explicatives ($d = 0$) toutes les arborescences auront la même vraisemblance. Les auteurs ont voulu voir comment le nombre de variables influence la détermination de la structure de l'arborescence. (Nous reviendrons sur ces résultats au chapitre 3, avec des explications et des précisions sur l'effet de $d = 0, 1, 2, 3$ ou 4). Pour les simulations, le modèle utilisé pour l'inférence avait le même nombre de variables explicatives que celui utilisé pour générer les données : la détermination du nombre de variables explicatives à partir des données n'étant pas étudiée. Le plan d'expérience était un plan d'expérience factoriel 2^d . Le nombre d'essais par cellule, T_i , a été fixé à 50 pour chaque cellule i , ce qui laisse une variabilité intéressante pour P . (Un T_i trop grand dissimulerait l'effet de d à cause des propriétés asymptotiques de la taille de l'échantillon, alors qu'un T_i trop petit pourrait faire confondre l'effet de d avec la variabilité de l'estimation de la structure

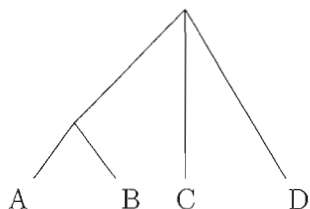


Figure 2.5 Structure d'arborescence pour une tri-nomiale suivie d'une binomiale

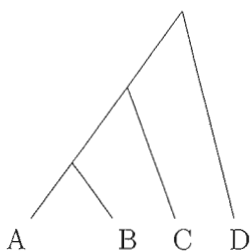


Figure 2.6 Structure d'arborescence pour trois binomiales consécutives

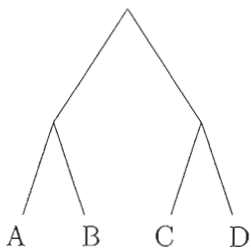


Figure 2.7 Structure d'arborescence pour une binomiale suivie de deux binomiales parallèles

de l'arborescence). Les valeurs des paramètres α influençant beaucoup la précision de l'estimation de la vraisemblance (les simulations pour étudier cet effet sont présentées plus loin), les auteurs ont choisi, pour chaque arborescence, un ensemble de valeurs parmi celles qui permettaient le mieux d'identifier la structure de l'arborescence (en utilisant les résultats des simulations menées pour déterminer l'effet des paramètres).

Pour une arborescence donnée, avec un ensemble approprié de valeurs des α fixé et T_i fixé

à 50, 10 ensembles de données ont été générés pour chaque valeur de d ($d = 2, 3, 4$). Pour chaque arborescence, les auteurs ont établi le nombre de fois où la vraie arborescence a été choisie pour chaque nombre d de variables explicatives. Précisons que les variables explicatives étaient des variables binaires. Les auteurs ont constaté que pour chaque arborescence, il faut au moins 3 variables explicatives et souvent 4 pour que les résultats soient satisfaisants, et ce même si les valeurs des paramètres ont été choisies intuitivement pour que la bonne arborescence ait plus de chance d'être correctement déterminée. La figure (2.8) reproduit d'après l'article de Rousseau et Sankoff le graphique du nombre de fois où la structure d'arborescence est correctement déterminée selon le nombre de variables explicatives pour chaque type d'arborescence. À part cette évidence qu'il faille un nombre minimale de variables explicatives pour déterminer correctement la bonne structure d'arborescence, il est difficile de tirer des conclusions bien nettes des résultats obtenus par les auteurs quant à l'effet du nombre de variables explicatives et ses liens éventuels avec la structure de l'arborescence.

2.2.2 Effet de la valeur des paramètres

Si pour chaque cellule i ($1 \leq i \leq n$) les paramètres α valent tous 0, alors on se trouve dans le cas où l'effet des variables explicatives est nul : les probabilités $p(t)$ ($1 \leq t \leq m$) sont toutes égales et les arborescences ont toutes la même vraisemblance. Si, pour un sous-ensemble des α , les paramètres contenus dans ce sous-ensemble valent tous 0, nous verrons qu'il est possible que les probabilités finales de deux arborescences différentes S et T , disons respectivement $p(t)$ et $q(t)$, soient les mêmes et donc que leurs vraisemblances soient les mêmes (la caractérisation "presque tout", dans les théorèmes présentés plus haut rend compte de ce cas). Si, par ailleurs, les α sont proches de 0, l'ensemble des probabilités finales $p(t)$ ($1 \leq t \leq m$) de l'arborescence S ne diffère qu'un peu de l'ensemble des probabilités finales $q(t)$ ($1 \leq t \leq m$) de l'arborescence T , et les vraisemblances de S et T auront presque la même valeur. La question qui se pose est donc quelle proportion des α doivent être différents de 0 et à quel point ils doivent l'être pour que la probabilité de choisir la bonne arborescence soit satisfaisante. Nous présenterons des

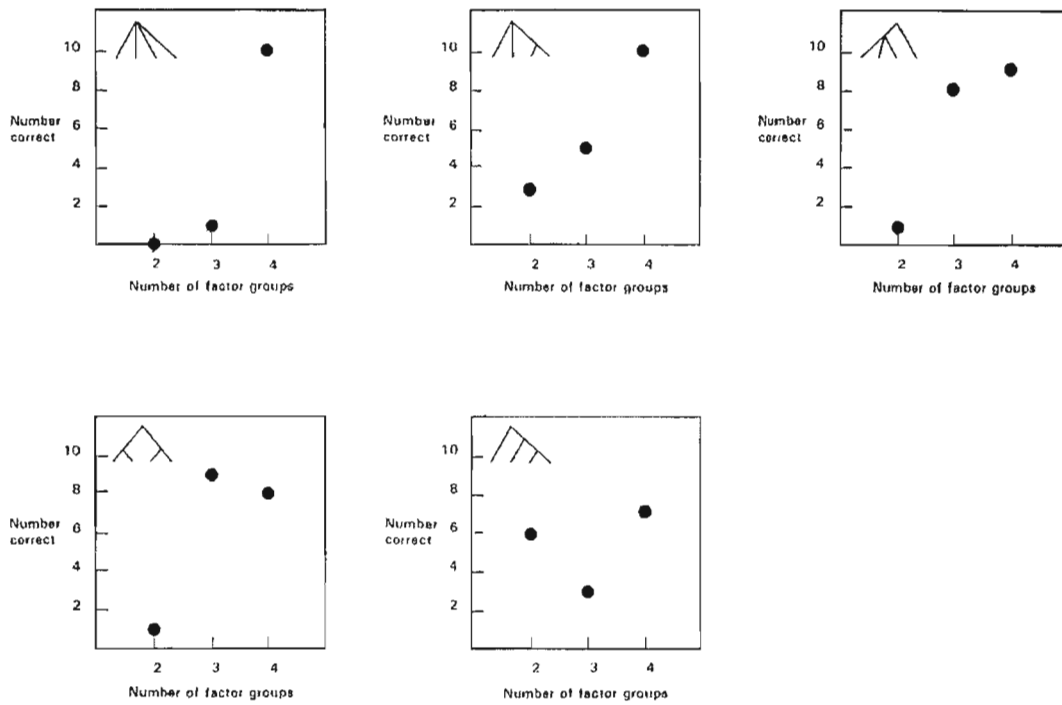


Figure 2.8 Graphique du nombre de fois où le bon modèle d'arborescence est sélectionné en fonction du nombre de variables explicatives, pour chaque type d'arborescence (repris de l'article de Rousseau et Sankoff).

		Level				Level				Level			
		1	2	3	4	1	2	3	4	1	2	3	4
outcome	A	.1	.1	.1	.7	.1	.1	.1	.6	.1	.1	.1	.5
	B	.1	.1	.7	.1	.1	.1	.6	.1	.1	.1	.5	.1
	C	.1	.7	.1	.1	.2	.6	.1	.1	.2	.5	.2	.2
	D	.7	.1	.1	.1	.6	.2	.2	.2	.5	.2	.2	.2
outcome	A	.1	.1	.1	.4	.10	.10	.10	.39	.10	.10	.10	.35
	B	.2	.2	.4	.1	.25	.25	.39	.10	.25	.25	.35	.10
	C	.3	.4	.2	.2	.26	.39	.25	.25	.30	.35	.25	.25
	D	.4	.3	.3	.3	.39	.26	.26	.26	.35	.30	.30	.30

		(1)			
outcome	A	.2	.2	.3	.3
	B	.2	.2	.3	.3
	C	.3	.3	.2	.2
	D	.3	.3	.2	.2

Table 2. Four-nomial model. Sets of (transformed) coefficients for studying the effects of parameter spread. $N_i = 50$ and the untransformed coefficient $\alpha_{10} = 0$ throughout.

Figure 2.9 Tableau des valeurs des paramètres pour une expérience 4-nominale (repris de l'article de Rousseau et Sankoff).

résultats théoriques sur cette question au chapitre 3, qui confirmeront ceux des études de simulation effectuées par les auteurs.

Les simulations ont donc été effectuées comme dans la sous-section précédente en utilisant un plan d'expérience factoriel de 2^4 pour maximiser une bonne détermination de la structure de l'arborescence, avec T_i , le nombre d'essais par cellule, fixé comme précédemment à 50, pour assurer suffisamment de variabilité à P . Pour chaque arborescence, avec un plan d'expérience factoriel 2^4 et $T_i = 50$, 10 ensembles de données ont été générés pour chaque différent ensemble de valeurs des paramètres. Les tableaux (2.9) et (2.10), reproduits, ainsi que les graphiques plus bas, à partir de l'article de Rousseau et Sankoff, donnent les valeurs des paramètres pour l'arborescence décrivant une expérience 4-nominale et l'arborescence associée à une expérience tri-nominale suivie d'une binomiale respectivement, alors que les graphiques (2.11) et (2.12) donnent les résultats pour ces mêmes arborescences (dans le même ordre).

		Level				Level				
		1	2	3	4	1	2	3	4	
outcome	A	.1	.1	.8	.1	.1	.1	.7	.1	
	B	.1	.8	.1	.1	.2	.7	.1	.2	
	2	.8	.1	.1	.8	.7	.2	.2	.7	
	C	.1	.9	.5	.5	.3	.7	.5	.5	
outcome	D	.9	.1	.5	.5	.7	.3	.5	.5	
	A	.2	.2	.6	.2	.2	.2	.5	.2	
	B	.2	.6	.2	.2	.3	.5	.2	.3	
	2	.6	.2	.2	.6	.5	.3	.3	.5	
outcome	C	.4	.6	.5	.5	.4	.6	.5	.5	
	D	.6	.4	.5	.5	.6	.4	.5	.5	
	outcome	A	.3	.3	.4	.3				
		B	.3	.4	.3	.3				
2		.4	.3	.3	.4					
C		.4	.6	.5	.5					
outcome	D	.6	.4	.5	.5					

Table 3. Tri-nomial followed by a binomial. Sets of (transformed) coefficients for studying the effects of parameter spread. $N_i = 50$ and the untransformed coefficient $\alpha_{i0} = 0$ throughout.

Figure 2.10 Tableau des valeurs des paramètres pour une expérience tri-nomiale suivie d'une binomiale (repris de l'article de Rousseau et Sankoff).

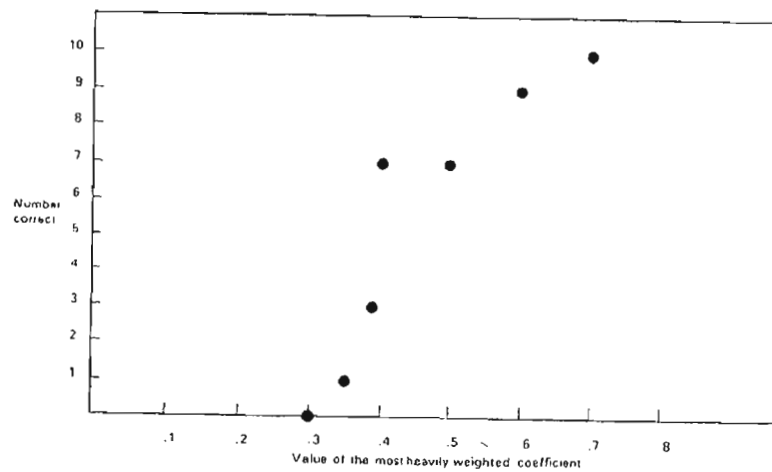


Figure 5. Performance of method for simulated data. True tree consists of one 4-nomial experiment. See Table 2 for experimental design, $N = 50$.

Figure 2.11 Graphique du nombre de fois où le bon modèle d'arborescence est sélectionné en fonction du poids maximal du coefficient, quand la vraie arborescence est 4-nomiale (repris de l'article de Rousseau et Sankoff).

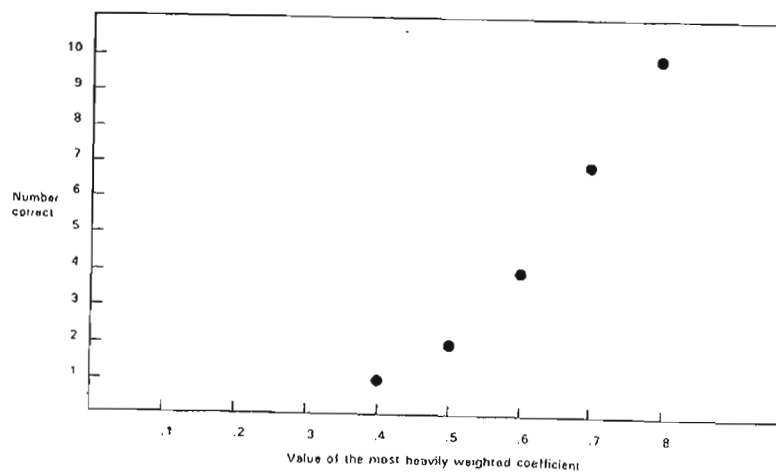


Figure 6. Performance of method for simulated data. True tree consists of one 3-nomial followed by a binomial experiment. See Table 3 for experimental design, $N = 50$.

Figure 2.12 Graphique du nombre de fois où le bon modèle d'arborescence est sélectionné en fonction du poids maximal du coefficient, quand la vraie arborescence est une tri-nomiale suivie d'une binomiale (repris de l'article de Rousseau et Sankoff).

Ces ensembles de valeur indiquent une dépendance plus ou moins grande de P envers les variables explicatives (selon l'ordre de grandeur des paramètres), laquelle détermine les différences entre les probabilités $p(t)$ ($1 \leq t \leq m$). On peut noter que dans les premiers ensembles de données des deux tableaux (2.9) et (2.10), les valeurs des paramètres sont toutes nulles sauf pour un paramètre, qui est plus ou moins fortement pondérée. Pour une telle structure, les probabilités $p(t)$ diffèrent de façon marquée pour chaque cellule. Intuitivement, c'est la situation opposée à des probabilités toutes égales, et on suppose donc que la probabilité de déterminer la vraie arborescence devrait être élevée pour ces ensembles de données. En parcourant les tableaux, on voit que les valeurs des paramètres sont de moins en moins différenciées et donc que les différences entre les probabilités $p(t)$ tendent à être plus petites. On s'attend donc à ce que la probabilité P de déterminer la bonne arborescence diminue. Les graphiques (2.11) et (2.12) du nombre de fois où la bonne arborescence a été déterminée versus le poids de la valeur la plus fortement pondérée soutiennent cette idée.

L'influence de la différence de grandeur entre les valeurs des paramètres semble donc importante, cependant, il faut noter que les auteurs expriment le souhait de caractériser davantage l'ensemble des valeurs des α qui influencerait significativement la probabilité de choisir la bonne arborescence. C'est ce que nous ferons au chapitre 3.

2.2.3 Effet de la taille de l'échantillon

Les dernières simulations faites par Rousseau et Sankoff visaient à déterminer l'influence de la taille T de l'échantillon sur la probabilité P de déterminer la vraie structure de l'arborescence. Pour une arborescence comportant une seule expérience 4-nomiale, avec un plan d'expérience factoriel 2^4 et un ensemble fixé de valeurs des α . 10 ensembles de données ont été générées pour $T_i = 30, 40, 50, 60, 70, 80, 90, 100$. L'ensemble des valeurs des α a été choisi en utilisant les résultats de la sous-section précédente pour permettre de bien déterminer la structure (la probabilité P était de 0,7 pour $T_i = 50$). Pour les autres ensembles de valeurs des paramètres, P était très proche de 1 ou près de 0, et il aurait alors fallu des valeurs de T_i très élevées pour améliorer P . Pour les valeurs des α

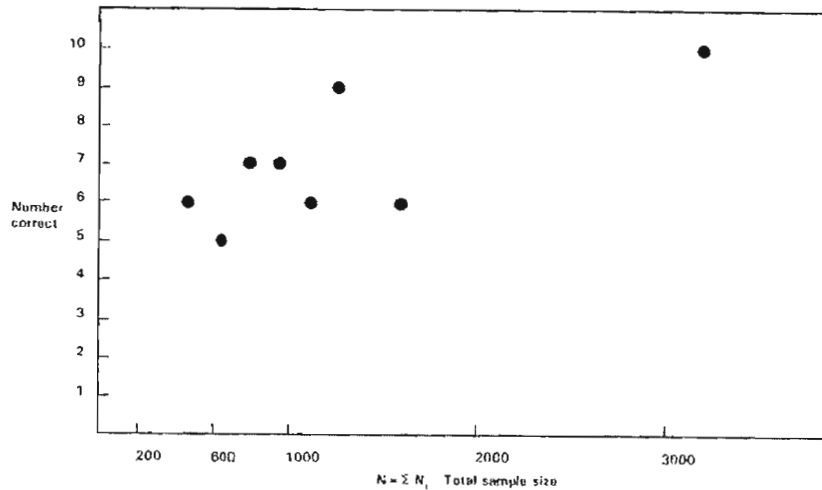


Figure 2.13 Graphique du nombre de fois où le bon modèle d'arborescence est sélectionné en fonction de T , quand la vraie arborescence est 4-nomiale (repris de l'article de Rousseau et Sankoff).

choisies, le nombre de fois sur 10 où la vraie arborescence est trouvée est stable autour de 7 pour $T_i = 50$, atteint 9 pour $T_i = 70$ et n'est jamais inférieure à 5 pour de petites valeurs de T_i . La figure (2.13), tirée de l'article de Rousseau et Sankoff, donne le nombre de fois sur 10 où la structure de l'arborescence a été correctement déterminée selon T pour une arborescence 4-nomiale.

Des simulations du même genre furent effectuées pour une arborescence dans laquelle une expérience trinominale est suivie par une binomiale, avec un ensemble de valeurs pour les paramètres fixées en utilisant les résultats de la sous-section précédente. Pour cette structure d'arborescence, la méthode donne de bons résultats pour $T_i = 50$, s'améliore rapidement pour $T_i > 50$ et devient rapidement moins efficace pour $T_i < 50$ (voir la figure (2.14), d'après l'article de Rousseau et Sankoff).

Les résultats obtenus dans cette sous-section montrent donc, comme on pouvait s'y attendre, que la taille de l'échantillon a un effet sur la détermination correcte de la

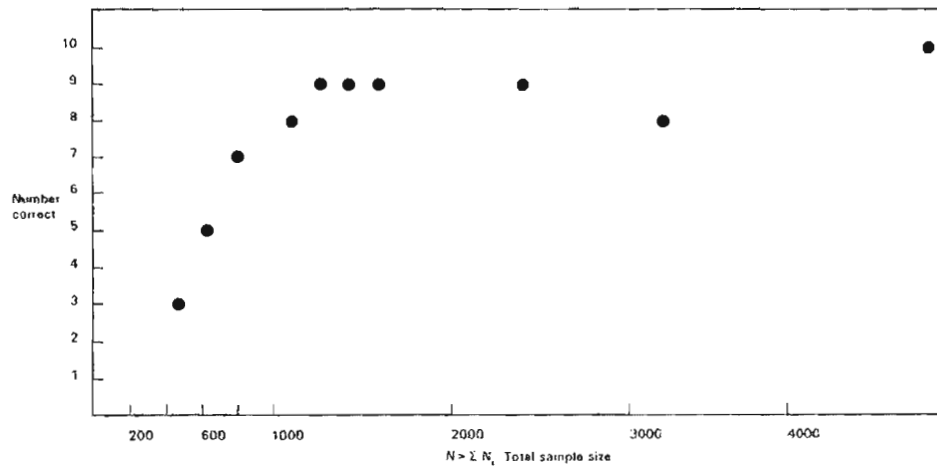


Figure 2.14 Graphique du nombre de fois où le bon modèle d'arborescence est sélectionné en fonction de T , quand la vraie arborescence est une trinomiale suivie d'une binomiale (repris de l'article de Rousseau et Sankoff).

structure de l'arborescence, et que l'efficacité de la méthode est proportionnelle à la taille de l'échantillon.

CHAPITRE III

ÉQUIVALENCE DE LA VRAISEMBLANCE DE CERTAINS TYPES D'ARBORESCENCES SELON LA FORME DES PARAMÈTRES

Nous avons vu avec l'équation (2.7) qui figure dans la preuve du théorème (1) du chapitre 2 que si $\ln(q_j)$ est linéarisable, alors les probabilités $p(t)$ et $q(t)$, associées aux arborescences S et T respectivement, peuvent être égales. Dans ce chapitre, nous présentons donc un théorème qui permet de déterminer les valeurs des paramètres pour lesquelles le logarithme de certaines probabilités est linéarisable. Nous donnons ensuite quelques exemples de ces valeurs et nous montrons les équivalences entre les arborescences qui surviennent lorsque certaines probabilités sont linéarisables.

3.1 Théorème de Rosenberg

Le théorème qui suit, dû au professeur Paul Rosenberg du Centre de recherches en mathématiques de l'Université de Montréal, montre sous quelles conditions l'expression

$$\ln \left(1 + \sum_{i=1}^{m-1} e^{\mathbf{X}'\alpha_i} \right),$$

où \mathbf{X} est le vecteur composé de d variables explicatives valant 0 ou 1 et où α_i ($1 \leq i \leq m-1$) est le vecteur contenant les d paramètres, est linéarisable. Introduisons d'abord la notation suivante :

Soit

$$\mathbf{A}_\alpha = \begin{pmatrix} \alpha_{1'} \\ \vdots \\ \alpha_{m-1'} \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1d} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{(m-1)1} & \alpha_{(m-1)2} & \dots & \alpha_{(m-1)d} \end{pmatrix},$$

une matrice réelle $(m-1) \times d$. Soit une fonction $f : \{0,1\}^d \rightarrow \mathbb{R}$ dépendant de \mathbf{A}_α et définie comme suit :

$$f_{\mathbf{A}_\alpha} = \ln \left(1 + \sum_{i=1}^{m-1} e^{\mathbf{X}'\alpha_i} \right).$$

Si $f_{\mathbf{A}_\alpha}$ est linéarisable alors il existe des $\gamma_0, \gamma_1, \dots, \gamma_d$ tels que

$$f_{\mathbf{A}_\alpha}(x_1, \dots, x_d) = \gamma_0 + \sum_{j=1}^d \gamma_j x_j \quad \forall x_j \in \{0,1\} \quad (1 \leq j \leq m-1).$$

Le théorème suivant donne les conditions pour lesquelles ces $\gamma_0, \gamma_1, \dots, \gamma_d$ existent :

Théorème 3 (Théorème de Rosenberg) *Pour une matrice \mathbf{A}_α donnée, il existe des $\gamma_0, \gamma_1, \dots, \gamma_d$ pour lesquels $f_{\mathbf{A}_\alpha}$ est linéarisable si et seulement si pour chaque sous-ensemble \mathbf{J} de $\{1, \dots, d\}$ de cardinalité $(1 < c \leq d)$ on a*

$${}^{(m)}c^{-1} \left(1 + \sum_{i=1}^{m-1} \left(\prod_{j \in \mathbf{J}} a_{ij} \right) \right) = \prod_{j \in \mathbf{J}} (1 + a_j^*), \quad (3.1)$$

où a_{ij} et a_j^* sont définis comme suit

$$a_{ij} := e^{\alpha_{ij}}$$

et

$$a_j^* := \sum_{i=1}^{m-1} a_{ij}.$$

Preuve : On sait que chaque fonction $f : \{0,1\}^d \rightarrow \mathbb{R}$ peut être représentée par un unique polynôme

$$p(x_1, \dots, x_d) = \sum_{S \subseteq \{1, \dots, d\}} a_S \prod_{i \in S} x_i. \quad (3.2)$$

où les a_S sont des réels appropriés. Ceux-ci peuvent être trouvés en définissant premièrement

$$a_\phi = f(0, \dots, 0) = \ln \left(1 + \sum_{i=1}^{m-1} e^{0' \alpha_i} \right) = \ln(1 + m - 1) = \ln(m) .$$

Ensuite, si $k \leq d$ et a_S a été trouvé pour tout $S \subseteq \{1, \dots, d\}$ pour lequel $|S| < k$ alors pour $J = \{j_1, \dots, j_k\} \subseteq \{1, \dots, d\}$ ($|J| = k$) on a

$$a_J = f(t_1, \dots, t_d) - \sum_{S \subset J} a_S ,$$

$$\text{où } t_i = \begin{cases} 1 & \text{si } i \in J \\ 0 & \text{sinon} \end{cases} .$$

(En effet, en définissant les t_i de cette façon, on fait disparaître les coefficients a_S pour tous les ensembles S avec $S \subseteq \{1, \dots, d\}$ sauf J et les ensembles inclus dans J . Puisqu'on a

$$\begin{aligned} p(t_1, \dots, t_d) &= f(t_1, \dots, t_d) \\ &= \sum_{S \subseteq \{1, \dots, d\}} a_S \prod_{i \in S} x_i . \end{aligned}$$

mais que les x_i associés à a_S prennent la valeur 1 pour l'ensemble J et les ensembles S tels que $S \subset J$ et la valeur 0 pour les autres, on a donc

$$f(t_1, \dots, t_d) = \sum_{S \subset J} a_S + a_J ,$$

d'où l'équation ci-dessus.)

Supposons que f_{A_α} est linéarisable. alors le polynôme p a forme

$$\begin{aligned} p(x_1, \dots, x_d) &= f_{A_\alpha}(x_1, \dots, x_d) \\ &= \gamma_0 + \sum_{j=1}^{m-1} \gamma_j x_j . \end{aligned}$$

Alors $\gamma_0 = f_{A_\alpha}(0, \dots, 0) = a_\phi = \ln(m)$ et de même de (3.2) on a :

$$\begin{aligned}
\gamma_j &= a_{\{j\}} \\
&= f_{A_\alpha}(0, \dots, 0, 1, 0, \dots, 0) - a_\phi \quad (\text{où } 1 \text{ est à la } j\text{-ème position}) \\
&= \ln\left(1 + \sum_{i=1}^{m-1} e^{\alpha_{ij}}\right) - \ln(m) \\
&= \ln\left(\frac{1}{m} \left(1 + \sum_{i=1}^{m-1} e^{\alpha_{ij}}\right)\right) \\
&= \ln\left(\frac{1}{m} \left(1 + \sum_{i=1}^{m-1} a_{ij}\right)\right) \\
&= \ln\left(\frac{1 + a_j^*}{m}\right). \tag{3.3}
\end{aligned}$$

On a ensuite $a_S = 0$ pour tout $S \subseteq \{1, \dots, d\}$ pour lequel $|S| > 1$ (puisqu'on ne veut pas de termes mixtes ou de puissances supérieure ou égale à deux). Il faut que pour tout $J \subseteq \{1, \dots, d\}$ on ait

$$f(t_1, \dots, t_d) = \sum_{j \in J} a_{\{j\}} + a_\phi. \tag{3.4}$$

$$\text{avec } t_i = \begin{cases} 1 & \text{si } i \in J \\ 0 & \text{sinon} \end{cases}.$$

Par la définition de $f(t_1, \dots, t_d)$, le membre de gauche de (3.4) vaut

$$\begin{aligned}
\ln\left(1 + \sum_{i=1}^{m-1} e^{\sum_{j \in J} \alpha_{ij}}\right) &= \ln\left(1 + \sum_{i=1}^{m-1} \prod_{j \in J} e^{\alpha_{ij}}\right) \\
&= \ln\left(1 + \sum_{i=1}^{m-1} \prod_{j \in J} a_{ij}\right). \tag{3.5}
\end{aligned}$$

et par (3.3), le membre de droite de (3.4) est égal à

$$\begin{aligned}
\sum_{j \in J} \ln \left(\frac{1 + a_j^*}{m} \right) + \ln(m) &= \left(\sum_{j \in J} \ln(1 + a_j^*) - \ln(m) \right) + \ln(m) \\
&= \sum_{j \in J} \ln(1 + a_j^*) - |J| \ln(m) + \ln(m) \\
&= \ln \left(\prod_{j \in J} (1 + a_j^*) \right) + \ln(m)^{-(|J|-1)} \\
&= \ln \left(\frac{\prod_{j \in J} (1 + a_j^*)}{m^{(|J|-1)}} \right). \tag{3.6}
\end{aligned}$$

Si on élève à la puissance e les deux côtés de (3.4), par (3.5) et (3.6) on obtient :

$$m^{|J|-1} \left(1 + \sum_{i=1}^{m-1} \prod_{j \in J} a_{ij} \right) = \prod_{j \in J} (1 + a_j^*) \quad \square$$

3.2 Effet de la condition de Rosenberg sur les valeurs des paramètres

Nous allons maintenant montrer comment les conditions du théorème (3) pour que $\ln \left(1 + \sum_{i=1}^{m-1} e^{\mathbf{X}' \boldsymbol{\alpha}_i} \right)$ soit linéarisable se traduisent quant aux valeurs des paramètres dans le contexte de nos modèles d'arborences. Nous allons énoncer ces conditions pour $2 \leq m \leq 4$ et $1 \leq d \leq 4$. Pour rendre plus claire la notation, réécrivons les conditions comme suit, dans une notation plus proche de celle de l'algèbre linéaire : Soit

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{pmatrix} = (\mathbf{a}_1 \ \mathbf{a}_2 \ \mathbf{a}_3 \ \mathbf{a}_4).$$

Alors pour $c=2$ on peut réécrire ainsi l'équation (3.1) avec $1 \leq j, k \leq 4$:

$$m^{c-1} (1 + \mathbf{a}'_j \mathbf{a}_k) = (1 + \mathbf{1}' \mathbf{a}_j) (1 + \mathbf{1}' \mathbf{a}_k),$$

où $\mathbf{1}' = (1, \dots, 1)$, le vecteur $\mathbf{1}$ de dimension appropriée.

De même, on peut écrire pour $c=3$, avec $1 \leq j, k, l \leq 4$

$$m^{c-1} \left(1 + \sum_{i=1}^{m-1} a_{ij} a_{ik} a_{il} \right) = (1 + \mathbf{1}'\mathbf{a}_j)(1 + \mathbf{1}'\mathbf{a}_k)(1 + \mathbf{1}'\mathbf{a}_l) ,$$

et pour $c=4$

$$m^{c-1} \left(1 + \sum_{i=1}^{m-1} a_{i1} a_{i2} a_{i3} a_{i4} \right) = (1 + \mathbf{1}'\mathbf{a}_1)(1 + \mathbf{1}'\mathbf{a}_2)(1 + \mathbf{1}'\mathbf{a}_3)(1 + \mathbf{1}'\mathbf{a}_4) .$$

Présentons maintenant quelques résultats simples, mais fort utiles pour la suite :

Lemme 3 *Si tous les éléments de \mathbf{A} ont la valeur 1, c'est-à-dire si $\alpha_{ij} = 0$ pour $1 \leq i \leq m-1$ et $1 \leq j \leq d$ alors l'équation (3.1) est vérifiée.*

Preuve : En effet, on a alors pour le membre de gauche de l'équation

$$1 + \sum_{i=1}^{m-1} \prod_{j \in J} a_{ij} = 1 + \sum_{i=1}^{m-1} \underbrace{1 \times 1 \times \dots \times 1}_{|J| \text{ fois}} = 1 + \sum_{i=1}^{m-1} 1 = m ,$$

et on a pour le membre de droite

$$1 + \mathbf{1}'\mathbf{a}_k = 1 + \mathbf{1}'\mathbf{1} = 1 + m - 1 = m .$$

En remplaçant les expressions qui précèdent dans (3.1), on a effectivement :

$$m^{c-1} m = \underbrace{m m \dots m}_{c \text{ fois}} \square$$

Lemme 4 *Si \mathbf{a}_l est quelconque et si $\mathbf{a}_k = \mathbf{1}$ pour tout $k \neq l$ ($1 \leq k \leq d$), alors l'équation (3.1) est vérifiée.*

Preuve : En effet dans ce cas le membre de gauche de (3.1) vaut

$$m^{c-1} \left(1 + \sum_{i=1}^{m-1} a_{ik_1} a_{ik_2} \dots a_{ik_c} \right) .$$

On vérifie l'égalité pour tout c et pour tous les sous-ensembles $\{k_1, \dots, k_c\}$ de J , $1 < c \leq d$. Alors, soit le sous-ensemble n'inclut pas l , auquel cas l'égalité est vérifiée par le lemme (3), soit le sous-ensemble inclut l et alors le membre de gauche de la condition de Rosenberg s'écrit

$$\begin{aligned} m^{c-1} \left(1 + \sum_{i=1}^{m-1} a_{ik_1} a_{ik_2} \dots a_{ik_c} \right) &= m^{c-1} \left(1 + \sum_{i=1}^{m-1} 1 \dots 1 a_{il} 1 \dots 1 \right) \\ &= m^{c-1} \left(1 + \sum_{i=1}^{m-1} a_{il} \right) \\ &= m^{c-1} (1 + \mathbf{1}'\mathbf{a}_l) . \end{aligned}$$

Le membre de droite devient

$$\begin{aligned} (1 + \mathbf{a}_l') \prod_{k \neq j} (1 + \mathbf{a}_k') &= (1 + \mathbf{1}'\mathbf{a}_l) \underbrace{(1 + m - 1) \dots (1 + m - 1)}_{(c-1) \text{ fois}} \\ &= m^{c-1} (1 + \mathbf{1}'\mathbf{a}_l) . \end{aligned}$$

ainsi la condition de l'équation (3.1) est vérifiée. \square

Le théorème de Rosenberg s'applique pour $d > 1$. Le lemme suivant traite le cas $d = 1$.

Lemme 5 Si $d=1$, alors $\ln \left(1 + \sum_{i=1}^{m-1} e^{\alpha_i x} \right)$ est linéarisable.

Preuve : Comme x vaut 0 ou 1, on a :

$$\begin{aligned} \ln \left(1 + \sum_{i=1}^{m-1} e^{\alpha_i x} \right) &= \left(\ln \left(1 + \sum_{i=1}^{m-1} e^{\alpha_i} \right) \right) x + (\ln m)(1 - x) \\ &= \left(\ln \left(1 + \sum_{i=1}^{m-1} e^{\alpha_i} \right) - \ln m \right) x + \ln m \quad \square \end{aligned}$$

3.3 Cas particuliers

Nous allons maintenant trouver les valeurs des paramètres lorsque la condition de linéarisation est satisfaite pour les cas $m = 2, 3, 4$ catégories et $d = 2, 3, 4$ variables. Nous examinerons particulièrement combien d'éléments des paramètres peuvent être différents de 0 pour que l'expression qui nous intéresse soit linéarisable. et quelles conditions

doivent satisfaire les éléments différents de 0. Notre attention se concentrera sur des exemples avec des paramètres $\alpha_{ij} = 0$, c'est-à-dire $a_{ij} = 1$, puisque les simulations dont nous rendons compte dans le chapitre 4 portent sur des ensembles de données générés avec de tels paramètres.

Cas $m = 2, d = 2$

Étant donné qu'il n'y a qu'une équation de régression logistique avec $d = 2$ variables explicatives, la matrice \mathbf{A} est un vecteur de dimensions 1×2 . Posons

$$\mathbf{A} = (a \ b) .$$

Il n'y a qu'un cas à examiner pour la condition de Rosenberg, soit $c = d = 2$. On a donc

$$\begin{aligned} 2^{2-1}(1+ab) &= (1+a)(1+b) &\iff 2+2ab &= 1+a+b+ab \\ & &\iff 1+ab &= a+b \\ & &\iff ab-a &= b-1 \\ & &\iff a(b-1) &= (b-1) \\ & &\iff a \text{ ou } b &= 1 . \end{aligned} \tag{3.7}$$

Donc, pour que l'équation (3.1) soit vérifiée il faut avoir $a = 1$ ou $b = 1$, c'est-à-dire $\alpha_1 = 0$ ou $\alpha_2 = 0$. Il ne peut y avoir qu'une seule variable explicative dont l'effet est non nul.

Cas $m = 2, d = 3$

Étant donné qu'il n'y a qu'une équation de régression logistique avec $d = 3$ variables explicatives, la matrice \mathbf{A} est un vecteur de dimensions 1×3 . Posons

$$\mathbf{A} = (a \ b \ f) .$$

La condition de Rosenberg doit être vérifiée pour $c \leq d = 3$, donc pour $c = 2$ et $c = 3$.

Condition de Rosenberg pour $c = 2$

En utilisant le résultat donné par l'équation (3.7), on trouve qu'un seul paramètre parmi

a, b et f doit être différent de 1, c'est-à-dire un seul $\alpha_i \neq 0$ et donc qu'une seule variable explicative a un effet non nul.

Condition de Rosenberg pour $c = 3$

Si un seul parmi a, b et $f \neq 1$, alors l'équation (3.1) est vérifiée par le lemme (4).

Cas $m = 2, d = 4$

Étant donné qu'il n'y a qu'une équation de régression logistique avec $d = 4$ variables explicatives, la matrice \mathbf{A} est un vecteur de dimensions 1×4 . Posons

$$\mathbf{A} = (a \ b \ f \ g) .$$

La condition de Rosenberg doit être vérifiée pour $c = 2, 3, 4$.

Condition de Rosenberg pour $c = 2$

Par (3.7) il faut avoir un seul parmi a, b, f et $g \neq 1$.

Condition de Rosenberg pour $c = 3$ et $c = 4$

Si un seul parmi a, b, f et $g \neq 1$, alors l'équation (3.1) est vérifiée par le lemme (4).

Examinons maintenant le cas avec $m = 3$ catégories et $d = 2$ variables explicatives.

Cas $m = 3, d = 2$

Il y a maintenant 2 équations de régression logistique avec avec 2 variables explicatives.

La matrice \mathbf{A} est donc de dimensions 2×2 . Posons

$$\mathbf{A} = \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} .$$

La condition de Rosenberg doit être vérifiée pour $c = 2$.

Condition de Rosenberg pour $c = 2$

La condition de Rosenberg s'écrit alors

$$\begin{aligned}
3^{2-1}(1 + a_1b_1 + a_2b_2) &= (1 + a_1 + a_2)(1 + b_1 + b_2) & (3.8) \\
&\iff \\
3 + 3a_1b_1 + 3a_2b_2 &= 1 + a_1 + a_2 + b_1 + b_2 + a_1b_1 + a_1b_2 + a_2b_1 + a_2b_2 \\
&\iff \\
2a_2b_2 - b_2 - a_1b_2 &= -2 + a_1 + a_2 + b_1 - 2a_1b_1 + a_2b_1 \\
&\iff \\
b_2(2a_2 - 1 - a_1) &= a_1 + a_2 - 2 - b_1(2a_1 - 1 - a_2) \\
&\iff \\
b_2 &= \frac{1'a - 2 - b_1(2a_1 - 1 - a_2)}{2a_2 - 1 - a_1} . & (3.9)
\end{aligned}$$

Donc, si

$$b_2 = \frac{1'a - 2 - b_1(2a_1 - 1 - a_2)}{2a_2 - 1 - a_1} ;$$

la condition de Rosenberg est vérifiée. Voyons quelques exemples du cas $m = 3, d = 2$ selon les valeurs que peuvent prendre a_1, a_2, b_1 . Ces exemples traitent essentiellement du cas où certains $\alpha_{ij} = 0$ (et donc les a_i ou b_i valent 1 et les variables explicatives ont un effet nul) et du cas où $\alpha_{ij} = \alpha_{kj}$ (même effet des variables explicatives).

$$\underline{a_1 = a_2 \neq 1}$$

$$\Rightarrow b_2 = \frac{2a_1 - 2 - b_1(a_1 - 1)}{a_1 - 1} = \frac{2(a_1 - 1) - b_1(a_1 - 1)}{a_1 - 1} = 2 - b_1 ;$$

et on a donc

$$\mathbf{A} = \begin{pmatrix} a_1 & b_1 \\ a_1 & 2 - b_1 \end{pmatrix} .$$

$a_1 = a_2 = 1$ Par le lemme (4), on a \mathbf{b} quelconque et on a donc

$$\mathbf{A} = \begin{pmatrix} 1 & b_1 \\ 1 & b_2 \end{pmatrix} .$$

$a_1 = 1, a_2$ quelconque

$$\Rightarrow b_2 = \frac{a_2 - 1 - b_1(1 - a_2)}{2(a_2 - 1)} = \frac{a_2 - 1 + b_1(a_2 - 1)}{2(a_2 - 1)}, = \frac{1 + b_1}{2}$$

et on a donc

$$\mathbf{A} = \begin{pmatrix} 1 & b_1 \\ a_2 & \frac{1+b_1}{2} \end{pmatrix}.$$

$a_1 = b_1 = 1$ ou $a_1 = b_2 = 1$

Par ce qui précède on a

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ a_2 & 1 \end{pmatrix}.$$

Étudions maintenant le cas où $m = 3$ catégories et $d = 3$ variables explicatives.

Cas $m = 3, d = 3$

Il y a maintenant 2 équations de régression logistique avec avec 3 variables explicatives.

La matrice \mathbf{A} est donc de dimensions 2×3 . Posons

$$\mathbf{A} = \begin{pmatrix} a_1 & b_1 & f_1 \\ a_2 & b_2 & f_2 \end{pmatrix}.$$

Compte tenu des simulations décrites dans le chapitre 4, nous avons concentré notre attention sur les cas où les $\alpha_{ij} = 0$, c'est-à-dire où a_{ij}, b_{ij} ou $f_{ij} = 1$.

Considérons par exemple la matrice \mathbf{A} suivante :

$$\mathbf{A} = \begin{pmatrix} 1 & b_1 & f_1 \\ 1 & b_2 & \frac{1'b-2-f_1(2b_1-1-b_2)}{2b_2-1-b_1} \end{pmatrix}.$$

La condition de Rosenberg doit être vérifiée pour $c = 2$ et $c = 3$. En utilisant ce qui précède, nous arrivons au résultat suivant :

Pour $c = 2$, l'équation (3.1) est vérifiée pour la colonne 1 avec les colonnes 2 et 3 par le lemme (4) et on sait par (3.9) qu'elle est aussi vérifiée pour les colonnes 2 et 3.

Pour $c = 3$, il faut qu'avec

$$f_2 = \frac{\mathbf{1}'\mathbf{b} - 2 - f_1(2b_1 - 1 - b_2)}{2b_2 - 1 - b_1},$$

l'équation suivante soit vérifiée :

$$\begin{aligned} 3^2(1 + 1b_1f_1 + 1b_2f_2) &= (3)(1 + b_1 + b_2)(1 + f_1 + f_2) \\ &\iff \\ 3(1 + b_1f_1 + b_2f_2) &= (1 + b_1 + b_2) + (1 + f_1 + f_2), \end{aligned} \quad (3.10)$$

ce qui est la même équation que (3.8) en remplaçant a_i par b_i et b_i par f_i $i = 1, 2$; les conditions de Rosenberg sont donc respectées.

Si on regarde les cas particuliers vus précédemment dans le contexte où $d = 3$, c'est-à-dire les cas où certains a_{ij}, b_{ij} ou f_{ij} valent 1 et donc où les α_{ij} correspondant valent 0, on trouve les matrices \mathbf{A} suivantes pour lesquelles la condition de Rosenberg est vérifiée :

$b_1 = b_2 \neq 1$

$$\mathbf{A} = \begin{pmatrix} 1 & b_1 & f_1 \\ 1 & b_2 & 2 - f_1 \end{pmatrix}.$$

$b_1 = b_2 = 1$ Par le lemme (4), on a f quelconque et on a donc

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & f_1 \\ 1 & 1 & f_2 \end{pmatrix}.$$

$b_1 = 1, b_2$ quelconque

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & f_1 \\ 1 & b_2 & \frac{1+f_1}{2} \end{pmatrix}.$$

$b_1 = f_1 = 1$ ou $b_1 = f_2 = 1$ Par ce qui précède on a

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & b_2 & 1 \end{pmatrix}.$$

Cas $m = 3, d = 4$

Le cas où il y a 2 équations de régression logistique et 4 variables explicatives est examiné en considérant que certains $\alpha_{ij} = 0$ et donc certains $a_{ij} = 1$. Posons

$$\mathbf{A} = \begin{pmatrix} a_1 & b_1 & f_1 & g_1 \\ a_2 & b_2 & f_2 & g_2 \end{pmatrix}.$$

Considérons la matrice A suivante

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & f_1 & g_1 \\ 1 & 1 & f_2 & \frac{1'f - 2 - g_1(2f_1 - 1 - f_2)}{2f_2 - 1 - f_1} \end{pmatrix}.$$

Pour $c = 2$ et $c = 3$, on voit facilement, par ce qui précède, que l'équation (3.1) est vérifiée, puisqu'on a la même matrice que dans le cas $d = 3$ à une colonne de 1 près, et on a vu par le lemme (4) que l'égalité est toujours vérifié lorsqu'une seule colonne est différente de $\mathbf{1}$. Pour $c = 4$, il faut qu'avec

$$g_2 = \frac{1'f - 2 - g_1(2f_1 - 1 - f_2)}{2f_2 - 1 - f_1},$$

l'équation suivante soit vérifiée :

$$\begin{aligned} 3^3(1 + 1f_1g_1 + 1f_2g_2) &= (3)(3)(1 + f_1 + f_2)(1 + g_1 + g_2) \\ &\iff \\ 3(1 + f_1g_1 + f_2g_2) &= (1 + f_1 + f_2)(1 + g_1 + g_2) \end{aligned} \quad (3.11)$$

ce qui est la même équation que (3.8) en remplaçant a_i par f_i , et b_i par g_i $i = 1, 2$; les conditions de Rosenberg sont donc respectées.

Si on regarde maintenant rapidement d'autres cas particuliers vus précédemment (avec certains $\alpha_{ij} = 0$, c'est-à-dire avec certains $a_{ij} = 1$) dans le contexte où $d = 4$, on obtient les matrices suivantes qui respectent la condition de Rosenberg :

$f_1 = f_2 \neq 1$

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & f_1 & g_1 \\ 1 & 1 & f_2 & 2 - g_1 \end{pmatrix}.$$

$f_1 = f_2 = 1$ Par le lemme (4), on a \mathbf{f} quelconque et on a donc

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 & g_1 \\ 1 & 1 & 1 & g_2 \end{pmatrix}.$$

$f_1 = 1, f_2$ quelconque

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 & g_1 \\ 1 & 1 & f_2 & \frac{1+g_1}{2} \end{pmatrix}.$$

$f_1 = g_1 = 1$ ou $f_1 = g_2 = 1$

Par ce qui précède on a

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & f_2 & 1 \end{pmatrix}.$$

On se rend compte que lorsque des $\alpha_{ij} = 0$ ($a_{ij} = 1$), certains ensembles de paramètres qui satisfont la condition de Rosenberg nous ramènent à des cas avec seulement 1 ou 2 variables explicatives avec des effets non nuls.

Considérons maintenant le cas où il y a $m = 4$ catégories et $d = 2$ variables explicatives.

Cas $m = 4, d = 2$

Soit

$$\mathbf{A} = \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \end{pmatrix}.$$

La condition de Rosenberg doit être vérifiée pour $c = 2$.

Condition de Rosenberg pour $c = 2$

$$4^{2-1}(1 + a_1b_1 + a_2b_2 + a_3b_3) = (1 + a_1 + a_2 + a_3)(1 + b_1 + b_2 + b_3) \quad (3.12)$$

$$\iff$$

$$4 + 4a_1b_1 + 4a_2b_2 + 4a_3b_3 = 1 + a_1 + a_2 + a_3 + b_1 + b_2 + b_3 + a_1b_1 + a_1b_2 + a_1b_3 + a_2b_1 + a_2b_2 + a_2b_3 + a_3b_1 + a_3b_2 + a_3b_3$$

$$\iff$$

$$3a_3b_3 - a_1b_3 - a_2b_3 - b_3 = -3 + a_1 + a_2 + a_3 + b_1 + b_2 - 3a_1b_1 + a_1b_2 + a_2b_1 + a_2b_2 - 3a_2b_2 + a_3b_1 + a_3b_2$$

$$\iff$$

$$b_3(-a_1 - a_2 + 3a_3 - 1) = a_1 + a_2 + a_3 - 3 - b_1(3a_1 - a_2 - a_3 - 1) - b_2(-a_1 + 3a_2 - a_3 - 1)$$

$$\iff$$

$$b_3 = \frac{1'a - 3 - b_1(3a_1 - a_2 - a_3 - 1) - b_2(-a_1 + 3a_2 - a_3 - 1)}{-a_1 - a_2 + 3a_3 - 1}. \quad (3.13)$$

Donc, si

$$b_3 = \frac{1'a - 3 - b_1(3a_1 - a_2 - a_3 - 1) - b_2(-a_1 + 3a_2 - a_3 - 1)}{-a_1 - a_2 + 3a_3 - 1}.$$

l'équation (3.1) est vérifiée.

Examinons pour ce cas également quelques exemples quand a_1, a_2, a_3, b_1 et b_2 prennent des valeurs particulières (généralement 1, c'est-à-dire que les α_{ij} correspondant valent 0).

$a_1 = a_2 = a_3 \neq 1$

On a

$$\begin{aligned} b_3 &= \frac{3a_1 - 3 - b_1(a_1 - 1) - b_2(a_1 - 1)}{a_1 - 1} \\ &= \frac{(a_1 - 1)(3 - b_1 - b_2)}{a_1 - 1} \\ &= 3 - b_1 - b_2, \end{aligned}$$

et donc

$$\mathbf{A} = \begin{pmatrix} a_1 & b_1 \\ a_1 & b_2 \\ a_1 & 3 - b_1 - b_2 \end{pmatrix}.$$

$$\underline{a_1 = a_2 = a_3 = 1}$$

Par le lemme (4), on a b quelconque et on a donc

$$\mathbf{A} = \begin{pmatrix} 1 & b_1 \\ 1 & b_2 \\ 1 & b_3 \end{pmatrix}.$$

$$\underline{a_1 = a_2 = 1, a_3 \text{ quelconque}}$$

\Rightarrow

$$\begin{aligned} b_3 &= \frac{a_3 - 1 - b_1(1 - a_3) - b_2(1 - a_3)}{3a_3 - 3} \\ &= \frac{(a_3 - 1)(b_1 + b_2 - 1)}{3(a_3 - 1)} \\ &= \frac{b_1 + b_2 + 1}{3}, \end{aligned}$$

et on a donc

$$\mathbf{A} = \begin{pmatrix} 1 & b_1 \\ 1 & b_2 \\ a_3 & \frac{b_1 + b_2 + 1}{3} \end{pmatrix}.$$

$$\underline{a_1 = a_2 = b_1 = b_2 = 1 \text{ ou}}$$

$$\underline{a_1 = a_2 = b_1 = b_3 = 1 \text{ ou}}$$

$$\underline{a_1 = a_2 = b_2 = b_3 = 1}$$

Par ce qui précède on a

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ a_3 & 1 \end{pmatrix}.$$

$$\underline{a_1 = a_3 = 1, a_2 \text{ quelconque}}$$

⇒

$$\begin{aligned} b_3 &= \frac{a_2 - 1 - b_1(1 - a_2) - b_2(3a_2 - 3)}{1 - a_2} \\ &= \frac{(-b_1 + 3b_2 - 1)(1 - a_2)}{1 - a_2} \\ &= -b_1 + 3b_2 - 1, \end{aligned}$$

et on a donc

$$\mathbf{A} = \begin{pmatrix} 1 & b_1 \\ a_2 & b_2 \\ 1 & -b_1 + 3b_2 - 1 \end{pmatrix}.$$

$$\underline{a_1 = a_3 = b_1 = b_2 = 1 \text{ ou}}$$

$$\underline{a_1 = a_3 = b_1 = b_3 = 1 \text{ ou}}$$

$$\underline{a_1 = a_3 = b_2 = b_3 = 1}$$

Comme précédemment et en utilisant le résultat ci-haut, on trouve :

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ a_2 & 1 \\ 1 & 1 \end{pmatrix},$$

$$\underline{a_2 = a_3 = 1, a_1 \text{ quelconque}}$$

⇒

$$\begin{aligned} b_3 &= \frac{a_1 - 1 - b_1(3a_1 - 3) - b_2(1 - a_1)}{1 - a_1} \\ &= \frac{(1 - a_1)(3b_1 - b_2 - 1)}{1 - a_1} \\ &= 3b_1 - b_2 - 1, \end{aligned}$$

et on a donc

$$\mathbf{A} = \begin{pmatrix} a_1 & b_1 \\ 1 & b_2 \\ 1 & 3b_1 - b_2 - 1 \end{pmatrix}.$$

$$\underline{a_2 = a_3 = b_1 = b_2 = 1} \text{ ou}$$

$$\underline{a_2 = a_3 = b_1 = b_3 = 1} \text{ ou}$$

$$\underline{a_2 = a_3 = b_2 = b_3 = 1}$$

On obtient :

$$\mathbf{A} = \begin{pmatrix} a_1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Étudions maintenant le cas avec $m = 4, d = 3$.

Cas $m = 4, d = 3$

Il y a 3 équations de régression logistique et 3 variables explicatives. la matrice \mathbf{A} est donc de dimensions 3×3 . Posons

$$\mathbf{A} = \begin{pmatrix} a_1 & b_1 & f_1 \\ a_2 & b_2 & f_2 \\ a_3 & b_3 & f_3 \end{pmatrix}.$$

Comme précédemment, nous allons voir que la matrice

$$\mathbf{A} = \begin{pmatrix} 1 & b_1 & f_1 \\ 1 & b_2 & f_2 \\ 1 & b_3 & f_3 \end{pmatrix}.$$

où

$$f_3 = \frac{1'b - 3 - f_1(3b_1 - b_2 - b_3 - 1) - f_2(-b_1 + 3b_2 - b_3 - 1)}{-b_1 - b_2 + 3b_3 - 1}.$$

satisfait l'équation (3.1). Nous avons vu plus haut que la condition est satisfaite pour $c = 2$ pour les 2e et 3e colonnes, et nous savons qu'elle l'est aussi pour les 1ère et 2e colonnes et pour les 1ère et 3e colonnes puisque la première colonne est composée

uniquement de 1. Pour $c = 3$ nous avons que l'équation (3.1) est vérifiée si :

$$\begin{aligned} 4^2(1 + 1b_1f_1 + 1b_2f_2 + 1b_3f_3) &= (4)(1 + b_1 + b_2 + b_3)(1 + f_1 + f_2 + f_3) \\ &\iff \\ 4(1 + b_1f_1 + b_2f_2 + b_3f_3) &= (1 + b_1 + b_2 + b_3)(1 + f_1 + f_2 + f_3), \end{aligned} \quad (3.14)$$

ce qui est la même équation que (3.12) en remplaçant a_i par b_i et b_i par f_i pour $i = 1, 2, 3$; l'égalité est donc respectée si on a f_3 de la forme spécifiée plus haut.

En se rappelant les matrices les plus simples vues pour le cas $d=2$, on a en particulier les matrices suivantes qui satisfont la condition de Rosenberg :

$$\underline{b_1 = b_2 = b_3 \neq 1}$$

$$\mathbf{A} = \begin{pmatrix} 1 & b_1 & f_1 \\ 1 & b_1 & f_2 \\ 1 & b_1 & 3 - f_1 - f_2 \end{pmatrix},$$

et

$$\underline{b_1 = b_2 = b_3 = 1}$$

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & f_1 \\ 1 & 1 & f_2 \\ 1 & 1 & f_3 \end{pmatrix}.$$

3.4 Cas orthogonaux ou quasi-orthogonaux

Les simulations dont nous parlerons plus en détails dans le prochain chapitre ont été faites en utilisant des matrices de paramètres (pour $d = 4$) de la forme

$$\mathbf{A} = \begin{pmatrix} a_1 & 1 & 1 & g_1 \\ a_1 & 1 & f_2 & 1 \\ a_1 & b_3 & 1 & 1 \end{pmatrix}$$

Il serait intéressant de savoir pour quels types de paramètres des matrices de la forme

$$\mathbf{A} = \begin{pmatrix} a_1 & 1 & f_1 \\ a_1 & b_2 & 1 \\ a_1 & 1 & 1 \end{pmatrix}$$

ou

$$\mathbf{A} = \begin{pmatrix} a_1 & 1 & f_1 \\ a_1 & 1 & 1 \\ a_1 & b_3 & 1 \end{pmatrix},$$

ou

$$\mathbf{A} = \begin{pmatrix} a_1 & 1 & 1 \\ a_1 & 1 & f_2 \\ a_1 & b_3 & 1 \end{pmatrix},$$

respectent les conditions de Rosenberg. Les équations étant les mêmes dans les 3 cas, à l'étiquette des paramètres près, nous ne traiterons que le cas

$$\mathbf{A} = \begin{pmatrix} a_1 & 1 & f_1 \\ a_1 & b_2 & 1 \\ a_1 & 1 & 1 \end{pmatrix}.$$

Nous trouvons que la matrice \mathbf{A} doit être de l'un des 3 types suivants pour que la condition de Rosenberg soit respectée :

$$\mathbf{A} = \begin{pmatrix} a_1 & 1 & 1 \\ a_1 & 1 & 1 \\ a_1 & 1 & 1 \end{pmatrix},$$

ou

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & b_2 & 1 \\ 1 & 1 & 1 \end{pmatrix},$$

ou

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & f_1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

En effet, avec $c = 2$ pour les 1ère et 2e colonnes, on a

$$4^{2-1}(1 + a_1 + a_1b_2 + a_1) = (1 + 3a_1)(3 + b_2) \quad (3.15)$$

$$\iff$$

$$4(1 + 2a_1 + a_1b_2) = 3 + b_2 + 9a_1 + 3a_1b_2$$

$$\iff$$

$$1 + a_1b_2 = b_2 + a_1$$

$$\iff$$

$$b_2(1 - a_1) = 1 - a_1 .$$

Donc il faut que $a_1 = 1$ ou $b_2 = 1$. En vérifiant la condition de Rosenberg pour les 1ère et 3e colonnes, on trouve aussi que $a_1 = 1$ ou $f_1 = 1$. Pour les colonnes 2 et 3 on a :

$$4^{2-1}(1 + f_1 + b_2 + 1) = (3 + b_2)(3 + f_1) \quad (3.16)$$

$$\iff$$

$$8 + 4b_2 + 4f_1 = 9 + 3b_2 + 3f_1 + b_2f_1$$

$$\iff$$

$$1 + b_2f_1 = b_2 + f_1$$

$$\iff$$

$$b_2(1 - f_1) = 1 - f_1 ; \quad (3.17)$$

et donc $b_2 = 1$ ou $f_1 = 1$. En utilisant le lemme (4), on obtient donc le résultat ci-haut.

Nous allons maintenant étudier le cas où il y a $m = 4$ catégories et $d = 4$ variables explicatives.

Cas $m = 4, d = 4$

Posons

$$\mathbf{A} = \begin{pmatrix} a_1 & b_1 & f_1 & g_1 \\ a_2 & b_2 & f_2 & g_2 \\ a_3 & b_3 & f_3 & g_3 \end{pmatrix} .$$

Comme dans les autres cas et selon le même argument, on peut voir facilement que la matrice

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & f_1 & g_1 \\ 1 & 1 & f_2 & g_2 \\ 1 & 1 & f_3 & g_3 \end{pmatrix},$$

où

$$g_3 = \frac{1'f - 3 - g_1(3f_1 - f_2 - f_3 - 1) - g_2(-f_1 + 3f_2 - f_3 - 1)}{-f_1 - f_2 + 3f_3 - 1},$$

satisfait l'équation (3.1). Comme dans le cas $d = 3$, les matrices suivantes respectent les conditions pour que l'expression qui nous intéresse soit linéarisable :

$$\underline{f_1 = f_2 = f_3 \neq 1}$$

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & f_1 & g_1 \\ 1 & 1 & f_1 & g_2 \\ 1 & 1 & f_1 & 3 - g_1 - g_2 \end{pmatrix},$$

et

$$\underline{f_1 = f_2 = f_3 = 1}$$

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 & g_1 \\ 1 & 1 & 1 & g_2 \\ 1 & 1 & 1 & g_3 \end{pmatrix}.$$

Intéressons-nous maintenant aux matrices de paramètres de nos simulations, qui ont la forme

$$\mathbf{A} = \begin{pmatrix} a_1 & 1 & 1 & g_1 \\ a_1 & 1 & f_2 & 1 \\ a_1 & b_3 & 1 & 1 \end{pmatrix}$$

Nous voulons savoir pour quelles valeurs des paramètres de telles matrices respectent les conditions de Rosenberg. Pour $c = 2$, nous avons des équations de la forme (3.15) lorsque nous regardons les colonnes 1 et 2 ou 1 et 3 ou 1 et 4. Nous avons donc $a_1 = 1$ ou $b_3 = 1$, $a_1 = 1$ ou $f_2 = 1$ et $a_1 = 1$ ou $g_1 = 1$. En regardant les colonnes 2 et 3, 2 et 4 et 3 et 4, nous avons des équations de la forme (3.16) qui entraînent $b_3 = 1$ ou $f_2 = 1$, $b_3 = 1$ ou $g_1 = 1$ et $f_2 = 1$ ou $g_1 = 1$. Donc, pour que l'équation (3.1) soit vérifiée

pour une matrice de cette forme, il faut qu'un seul paramètre parmi a_1, b_3, f_2 et g_1 soit différent de 1. Si tel est le cas, alors par le lemme (4) l'équation (3.1) est vérifiée pour $c = 3$ et $c = 4$.

3.5 Équivalence des arborescences A et C et B et D sous certaines conditions

Les études de simulation que l'on verra au chapitre 4 ont montré que pour plusieurs ensembles de paramètres, l'arborescence A (figure 3.1) et certaines arborescences C (figure 3.2) ont des vraisemblances quasi égales.

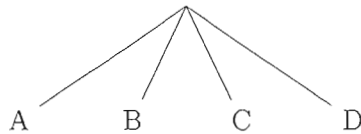


Figure 3.1 Arborescence A

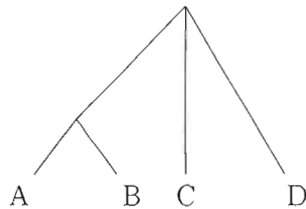
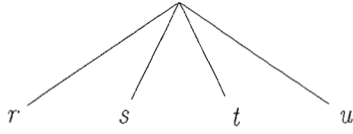


Figure 3.2 Arborescence C

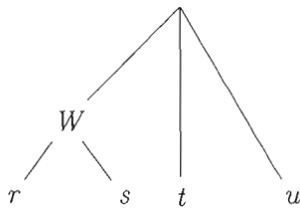
Ceci tend à suggérer que, pour ces ensembles de paramètres, les deux types d'arborescences modélisent des probabilités non identifiables. Le théorème qui suit fait le lien entre la possibilité de trouver des paramètres pour lesquels les probabilités des deux arborescences sont les mêmes et le fait que le logarithme de la probabilité intermédiaire associée à la catégorie de référence est linéarisable.

Théorème 4 *Soit l'arborescence S suivante :*



ayant r comme catégorie de référence et avec les paramètres α_s, α_t et α_u associés respectivement aux probabilités finales $p(s), p(t)$ et $p(u)$.

Soit l'arborescence T suivante :



pour laquelle r (et W) est la catégorie de référence, avec les probabilités finales $q(r), q(s), q(t), q(u)$ et les probabilités intermédiaires q_W, q_r, q_s . Alors, on peut trouver des paramètres β_s, β_t et β_u associés à l'arborescence T pour lesquels $p(i) = q(i) \ i = r, s, t, u \iff \ln(q_r)$ est linéarisable.

Preuve :

\Rightarrow Si on a des paramètres α_s, α_t et α_u associés à l'arborescence S et des paramètres β_s, β_t et β_u associés à l'arborescence T tels que $p(i) = q(i) \ i = r, s, t, u$ alors on a

$$\begin{aligned}
 \alpha'_t \mathbf{X} &= \ln \left(\frac{p(t)}{p(r)} \right) \\
 &= \ln \left(\frac{q(t)}{q(r)} \right) \\
 &= \ln \left(\frac{q(t)}{q_W q_r} \right) \\
 &= \ln \left(\frac{q(t)}{q_W} \right) - \ln(q_r) \\
 &= \beta'_t \mathbf{X} - \ln(q_r) \text{ (puisque } W \text{ est la catégorie de référence)}.
 \end{aligned}$$

D'où $\ln(q_r) = (\beta'_t - \alpha'_t) \mathbf{X}$ et donc $\ln(q_r)$ est linéarisable.

\Leftarrow Supposons que $\ln(q_r)$ est linéarisable. et donc que $\ln(q_r) = \mathbf{c}'_r \mathbf{X}$ où \mathbf{c}'_r est un vecteur

de constantes. Il faut maintenant montrer qu'on peut trouver des paramètres β_s, β_t et β_u associés à l'arborescence T tels que $p(i) = q(i)$ $i = r, s, t, u$.

On peut poser d'abord $\beta_s = \alpha_s$ pour les paramètres associés aux catégories s et r issues du noeud W . Ainsi on obtient

$$\ln \left(\frac{p(s)}{p(r)} \right) = \alpha'_s \mathbf{X} = \beta'_s \mathbf{X} = \ln \left(\frac{q(s)}{q(r)} \right),$$

et donc

$$p(s) = q(s) \left(\frac{p(r)}{q(r)} \right). \quad (3.18)$$

Pour la catégorie t ayant la racine pour ancêtre immédiat, le paramètre β_t associé à l'arborescence T doit satisfaire :

$$\begin{aligned} \ln \left(\frac{q(t)}{q(r)} \right) &= \ln \left(\frac{q(t)}{q_W q_r} \right) \\ &= \ln \left(\frac{q(t)}{q_W} \right) - \ln(q_r) \\ &= \beta'_t \mathbf{X} - \mathbf{c}'_r \mathbf{X} \text{ (puisque } \ln(q_r) \text{ est linéarisable)} \\ &= (\beta'_t - \mathbf{c}'_r) \mathbf{X}. \end{aligned}$$

En posant $\beta_t = \alpha_t + \mathbf{c}_r$, on obtient :

$$\ln \left(\frac{p(t)}{p(r)} \right) = \ln \left(\frac{q(t)}{q(r)} \right),$$

et donc

$$p(t) = q(t) \left(\frac{p(r)}{q(r)} \right). \quad (3.19)$$

De même, pour la catégorie u également issue directement de la racine, si on prend $\beta_u = \alpha_u + \mathbf{c}_r$, on a :

$$\begin{aligned} \ln \left(\frac{q(u)}{q(r)} \right) &= \ln \left(\frac{q(u)}{q_W q_r} \right) \\ &= \ln \left(\frac{q(u)}{q_W} \right) - \ln(q_r) \\ &= \beta'_u \mathbf{X} - \mathbf{c}'_r \mathbf{X} \\ &= (\beta'_u - \mathbf{c}'_r) \mathbf{X} \\ &= \alpha'_u \mathbf{X} \\ &= \ln \left(\frac{p(u)}{p(r)} \right), \end{aligned}$$

et donc :

$$p(u) = q(u) \left(\frac{p(r)}{q(r)} \right). \quad (3.20)$$

Par (3.18), (3.19) et (3.20), on a

$$p(i) = q(i) \left(\frac{p(r)}{q(r)} \right) \text{ pour } i = s, t, u.$$

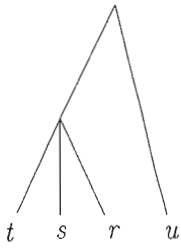
Il suffit donc de démontrer que $p(r) = q(r)$ et ainsi $p(i) = q(i)$ $i = r, s, t, u$.

Or

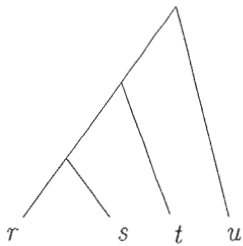
$$\begin{aligned} q(r) &= q_w q_r \\ &= \left(\frac{1}{1 + e^{\beta'_t \mathbf{X}} + e^{\beta'_u \mathbf{X}}} \right) \left(\frac{1}{1 + e^{\beta'_s \mathbf{X}}} \right) \\ &= \frac{1}{1 + e^{\beta'_s \mathbf{X}} + (e^{\beta'_t \mathbf{X}} + e^{\beta'_u \mathbf{X}}) (1 + e^{\beta'_s \mathbf{X}})} \\ &= \frac{1}{1 + e^{\alpha'_s \mathbf{X}} + (e^{\beta'_t \mathbf{X}} + e^{\beta'_u \mathbf{X}}) (1 + e^{\beta'_s \mathbf{X}})} \\ &= \frac{1}{1 + e^{\alpha'_s \mathbf{X}} + (e^{(\alpha'_t + \mathbf{c}'_r) \mathbf{X}} + e^{(\alpha'_u + \mathbf{c}'_r) \mathbf{X}}) (1 + e^{\beta'_s \mathbf{X}})} \quad (\text{car } \beta_i = \alpha'_i + \mathbf{c}'_r \text{ pour } i = t, u) \\ &= \frac{1}{1 + e^{\alpha'_s \mathbf{X}} + (e^{\alpha'_t \mathbf{X}} + e^{\alpha'_u \mathbf{X}}) e^{\mathbf{c}'_r \mathbf{X}} (1 + e^{\beta'_s \mathbf{X}})} \\ &= \frac{1}{1 + e^{\alpha'_s \mathbf{X}} + (e^{\alpha'_t \mathbf{X}} + e^{\alpha'_u \mathbf{X}}) q_r (1/q_r)} \quad (\text{car } 1/q_r = 1 + e^{\beta'_s \mathbf{X}}) \\ &= \frac{1}{1 + e^{\alpha'_s \mathbf{X}} + e^{\alpha'_t \mathbf{X}} + e^{\alpha'_u \mathbf{X}}} \\ &= p(r) \quad \square \end{aligned}$$

De même, on a remarqué dans les études de simulations que l'arborescence B et certaines arborescences D semblent équivalentes pour certaines valeurs des paramètres. Nous déduisons donc du théorème ci-haut le corollaire suivant :

Corollaire 2 *Soit l'arborescence S suivante :*

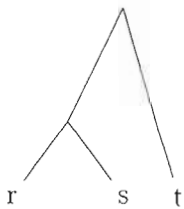
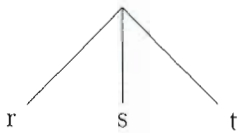


pour laquelle r est la catégorie de référence et l'arborescence T suivante :



qui a aussi la catégorie r comme catégorie de référence. Alors les deux arborescences sont équivalentes $\iff \ln(q_r)$ est linéarisable.

Preuve On déduit aisément du théorème (4) que les sous-arborescences



sont équivalentes, car seule la catégorie u , qui occupe la même position dans les deux arborescences, n'apparaît pas. Il suffit donc d'appliquer le théorème (4) aux sous-arborescences formées des catégories r, s, t . \square

CHAPITRE IV

ÉTUDE DE SIMULATIONS MONTE-CARLO

Pour mieux comprendre quels facteurs influençaient la détermination du type d'arborescence, une étude de simulation Monte-Carlo a été menée. Le plan d'expérience était défini par les éléments suivants :

1. il y avait 4 variables explicatives, selon un modèle réduit.
2. les ensembles de données étaient générés suivant une loi multinomiale.
3. les paramètres utilisés pour générer les ensembles de données pouvaient prendre les valeurs 0 ; 0,5 ; 1 ; 2 et 3.
4. le nombre d'essais de la multinomiale était 10 000 pour chaque cellule, une cellule étant définie par les valeurs 1 ou -1 que prennent les 4 variables explicatives binaires X_1 , X_2 , X_3 et X_4 .
5. pour chaque ensemble de paramètres, 50 jeux de données ont été simulés.

Les données étaient simulées à partir d'une expérience 4-nomiale (comme la figure (4.1) ci-dessous). On calculait ensuite le maximum de vraisemblance pour chaque arborescence pour chaque jeu de données et la médiane du rang de chacune des arborescences sur les 50 jeux de données. L'étude portait donc sur la capacité asymptotique de la vraisemblance à distinguer le vrai modèle d'arborescence (plutôt que sur la capacité d'estimer les paramètres).

4.1 Notation

4.1.1 Structures des arborescences

Les structures des arborescences étaient les suivantes :



Figure 4.1 Structure d'arborescence *A*

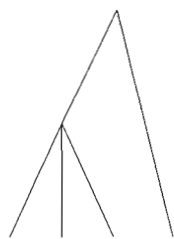


Figure 4.2 Structure d'arborescence *B*

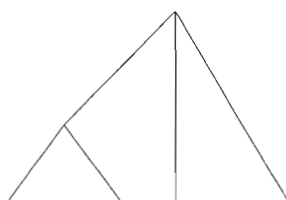


Figure 4.3 Structure d'arborescence *C*

4.1.2 Quelques remarques sur les matrices des paramètres

Les matrices de paramètres avaient la forme suivante :

$$\begin{pmatrix} a_{11} & 0 & 0 & a_{14} \\ a_{21} & 0 & a_{23} & 0 \\ a_{31} & a_{32} & 0 & 0 \end{pmatrix}$$

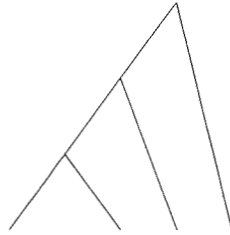


Figure 4.4 Structure d'arborescence D

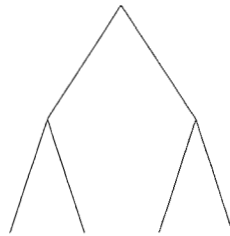


Figure 4.5 Structure d'arborescence E

Dans le cas qui nous occupe, les éléments de la j -ème colonne d'une telle matrice sont les effets de la j -ème variable explicative et l'élément (i, j) est l'effet de la j -ème variable explicative sur la i -ème catégorie.

Puisque les a_{i1} , $i = 1, 2, 3$ étaient toujours égaux dans nos simulations, cette matrice peut se réécrire de la façon suivante :

$$\begin{pmatrix} t & 0 & 0 & z \\ t & 0 & y & 0 \\ t & x & 0 & 0 \end{pmatrix}$$

Nous désignerons donc sans ambiguïté cette matrice par le vecteur (t, x, y, z) .

Les remarques suivantes aideront à comprendre les résultats des simulations ainsi que l'ordre des arborescences dans ces simulations.

- Si $t, x, y, z \neq 0$ alors toutes les catégories sont distinctes les unes des autres sur 2 variables explicatives.
- Si $t = 0$ et $x, y, z \neq 0$ alors les catégories 1,2 et 3 sont distinctes entre elles sur

- 2 variables explicatives et distinctes de la 4ème catégorie sur seulement 1 variable explicative. On dira qu'il n'y a que 3 variables explicatives avec effets non nuls.
- Si $x = 0$ et $t, y, z \neq 0$ alors la catégorie 3 est distincte de la catégorie 4 sur une seule variables explicative, tandis que les catégories 1,2,4 sont distinctes entre elles sur 2 variables explicatives. On dira aussi qu'il n'y a que 3 variables explicatives avec des effets non nuls. (Il en est de même si $y = 0$ et $t, x, z \neq 0$ ou si $z = 0$ et $t, x, y \neq 0$)
 - Si $t = x = 0$ et $y, z \neq 0$, alors les catégories 3 et 4 ne sont pas distinctes l'une de l'autre, et elles sont distinctes des autres catégories sur une seule variables explicative. Il n'y a que 2 variables explicatives avec effets non nuls. (Le cas est le même pour les catégories $t = y = 0$ et $x, z \neq 0$ et pour $t = z = 0$ et $x, y \neq 0$)

4.1.3 Notation pour les différentes arborescences de même structure

Dans les sections suivantes, nous utiliserons sans confusion possible, deux notations différentes pour désigner les différentes arborescences de même structure, afin de faciliter la lecture des résultats.

Par exemple, nous pourrions noter B_j , l'arborescence ayant la catégorie j issue directement de la racine ; C_{jk} , l'arborescence ayant les catégories j et k issues directement de la racine ; D_{jk} , l'arborescence pour laquelle la catégorie k est issue directement de la racine et la catégorie j est issue du nœud intermédiaire directement issu de la racine ; E_{jk} , l'arborescence avec les catégories j et k issues du même noeud intermédiaire. Remarquons que deux arborescences de même structure sont distinctes si le même chemin conduit à une catégorie différente. Ainsi, si on spécifie pour l'arborescence B quelle catégorie est issue directement de la racine, il n'est pas nécessaire de préciser la configuration des 3 autres catégories ; la description ci-haut est donc suffisamment détaillée pour toutes les structures d'arborescences.

Nous pourrions aussi désigner chaque catégorie à l'aide de la variable qui la distingue des autres catégories : la catégorie x sera par exemple celle que l'effet de x distingue des autres. Ainsi, nous pourrions représenter par B_x l'arborescence B dont la catégorie x est

observée à la première expérience, c'est-à-dire l'arborescence pour laquelle la catégorie influencée par x a été observée à la première expérience. Nous parlerons, selon le même raisonnement, des arborescences C_{xy} , D_{xy} ou E_{xy} .

Nous écrirons $A = C_{xy}$ si ces deux arborescences sont équivalentes, $C_{xy} \sim D_{xy}$ si l'ordre de ces arborescences est semblable et $C_{xy} \gtrsim D_{xy}$ si l'ordre de ces arborescences (selon leur vraisemblance) est en général semblable mais C_{xy} est souvent (un peu) plus grand.

Le tableau (4.1) donne la liste des arborescences équivalentes et celles qui sont quasi équivalentes, c'est-à-dire dont les rangs sont près l'un de l'autre pour certaines valeurs des paramètres. Dans ce tableau, x, y, z désignent la valeur du paramètre qui distingue la catégorie qui lui est associée. Rappelons que ce paramètre est aussi l'effet d'une variable explicative sur la catégorie, et cette variable est différente pour chaque catégorie, compte tenu la forme de notre matrice de paramètres. Si la variable explicative qui distingue la catégorie est nulle, la catégorie est notée par 0. Si x est l'effet de la variable qui distingue une catégorie j des autres, alors au lieu de noter B_j , nous noterons B_x . Si x est l'effet de la variable qui distingue une catégorie j des autres et si y est l'effet de la variable qui distingue une catégorie k des autres, alors au lieu de noter C_{jk} , nous noterons C_{xy} et ainsi de suite. Quand il est important de préciser les valeurs de tous les paramètres, nous noterons par exemple $C_{xy}(z, 0)$ pour l'arborescence C avec les catégories associées à x et y issues directement de la racine et pour laquelle z est quelconque et t vaut 0.

4.2 Exemples de suites d'arborescences ordonnées selon la vraisemblance

Dans cette section, nous allons illustrer à l'aide des simulations Monte-Carlo comment il est possible de distinguer clairement les différents modèles d'arborescences, et que ceci se reflète dans la différence entre les ordres de grandeur des vraisemblances des arborescences. Ainsi, le critère de vraisemblance de chaque arborescence est vraiment un critère qui permet de distinguer les arborescences les unes des autres et qui peut servir à estimer le modèle d'arborescence. Nous allons aussi illustrer le fait que certaines

Nombre de catégories d variables explicatives $\neq 0$	Arborescences équivalentes (x, y, z quelconques)	Arborescences quasi équivalentes (x, y, z quelconques)
$d = 3$	$A = C_{xy}(z, 0)$ $B_z = D_{xz}(y, 0)$	si $t \rightarrow 0$ alors $C_{xy}(z, t) \rightarrow A$ si $t \rightarrow 0$ alors $D_{xz}(y, t) \rightarrow B_z$ si $t < 1$ alors $C_{xy}(z, t) \sim A$ si $t < 1$ alors $D_{0z}(x, y) \sim D_{tz}(0, x) \sim B_z$
$d = 2$	$A = C_{xy}(0, 0) = B_z(x, 0, 0)$ $= D_{0z}(0, y) = E_{z,0}(x, 0)$	si $t, x \rightarrow 0$ alors $B_z(x, t, y) \rightarrow A$ si $t, x \rightarrow 0$ alors $D_{tz}(x, y) \rightarrow A$ si $t, x \rightarrow 0$ alors $E_{tz}(x, y) \rightarrow A$
$d = 1$	Toutes les arborescences sont équivalentes	
$d = 0$	Toutes les arborescences sont équivalentes	

Tableau 4.1 Tableau résumé de l'équivalence des arborescences selon le nombre de variables explicatives différentes de 0

arborescences sont équivalentes lorsque certaines variables explicatives ont des effets nuls.

Dans cette section, nous utilisons la notation B_j pour indiquer que la catégorie j est reliée directement à la racine dans l'arborescence B . et ainsi de suite comme nous l'avons vu précédemment. Si la valeur d'un paramètre est 0 alors on notera la catégorie correspondante (celle pour laquelle l'effet de la variable explicative qui la distingue des autres est nul) par la catégorie 0 : ainsi $C_{uv}(r0)$ $u, v \neq 0$, désigne l'arborescence ayant la catégorie 0 issue du nœud intermédiaire. (Les autres structures sont décrites de manière similaire.)

4.2.1 Exemple 1 : $(t, x, y, z) \ t \neq z \neq x, y$

La figure (4.6) donne en abscisse le rang de la vraisemblance de chaque arborescence et en ordonnée la structure de chaque arborescence. Ainsi pour $(t = 1, x = y = 2, z = 3)$, la vraisemblance de l'arborescence A était la plus grande parmi toutes les vraisemblances des 26 arborescences. Le rang de l'arborescence A est donc premier, 50 fois sur 50; celui de l'arborescence $B1$ est deuxième, 50 fois sur 50, etc. On peut remarquer que les arborescences $C13$ et $C12$ se partagent le rang 3 et 4, 50 fois sur 50. En effet, les catégories 2 et 3 ayant la même valeur de paramètre (la valeur 2), les deux arborescences $C12$ et $C13$ représentent des modèles symétriques. Leurs vraisemblances devraient théoriquement être égales et elles devraient se partager ainsi le rang 3 et 4 pour la moitié des 50 simulations. On constate que numériquement, les vraisemblances diffèrent légèrement et que l'arborescence $C12$ arrive au 3e rang 33 fois et l'arborescence $C13$ arrive au 3e rang 17 fois.

Toutes les fois qu'il y a des blocs de rangs dans le tableau, c'est que les arborescences faisant partie de ce bloc de rangs sont symétriques et devraient théoriquement avoir des vraisemblances égales.

4.2.2 Exemple 2 : $(t = 0, x, y, z) \ x \neq y \neq z$

Nous avons prouvé que si $d = 3$ (l'effet d'une variable explicative est nul), et si $u, v \neq 0$ alors les arborescences $Cuv(r, 0)$ sont équivalentes à l'arborescence A . Il y a trois arborescences $Cuv(r, 0)$, $u, v \neq 0$, il y a donc 4 arborescences équivalentes : A , $C12$, $C13$, $C23$. Nous voyons dans la figure (4.7) que les arborescences A et Cuv , $u, v \neq 4$ se partagent le premier rang pour le quart des 50 simulations. On peut juger sommairement que le rang de A est distribué uniformément entre le rang 1 et 4 (il en va de même pour chacune des arborescences Cuv). Nous observons donc un bloc en bas à gauche du tableau.

Sous ces conditions, nous avons aussi prouvé que l'arborescence Bj est équivalente à

l'arborescence D_{uj} , $u, j \neq 4$. Par exemple, les trois arborescences $B1$, $D21$, $D31$ sont équivalentes entre elles : ces trois arborescences se partagent les rangs 5,6,7 d'une façon que l'on juge sommairement comme étant uniforme, c'est-à-dire que chacune constitue le tiers des 50 simulaitons. (De même, $B2$, $B3$, appartiennent à des blocs de fréquences à l'intérieur desquels les arborescences sont équivalentes.)

4.2.3 Exemple 3 : $(t = 0, x = 0, y, z) \ y \neq z$

Nous avons prouvé que si $d = 2$ (les effets de deux variables explicatives sont nuls) alors les arborescences A , $Cuv(00)$, $D0u(0v)$ et $Eu0(v0)$ sont toutes équivalentes. Ces 16 arborescences se partagent donc toutes également les rangs 1 à 16. On voit avec la figure (4.8) que la fréquence des rangs est assez bien distribuée uniformément entre les rangs 1 à 16.

4.2.4 Exemple 4 : $(t = 0, x = 0, y = 0, z \neq 0)$

Nous avons prouvé que si $d = 1$ (les effets de trois variables explicatives sont nuls) alors toutes les arborescences sont équivalentes. La figure (4.9) illustre bien ce fait : les rangs sont distribués uniformément entre 1 et 26 pour chaque arborescence.

4.3 Etude des simulations pour le vecteur (t, x, y, z) , avec $x = y = z$

L'ordre des arborescences selon les vraisemblances pour les différentes valeurs du vecteur (t, x, y, z) est assez régulier. Nous allons d'abord étudier le comportement de l'ordre des arborescences pour le vecteur (t, x, y, z) où $x = y = z$. Cette étude nous permettra de déduire un ordre que nous généraliserons aux autres suites d'arborescences pour différentes valeurs de (t, x, y, z) .

La figure (4.10) montre les résultats de sortie du programme d'ordinateur. Chaque cellule (t, x) du tableau donne la suite des arborescences qui ont été ordonnées selon la médiane du rang de leurs vraisemblances : la cellule (t, x) représentant le vecteur de paramètres (t, x, x, x) , $x, t = 0; 0, 5; 1; 2; 3$. Ce tableau nécessite des explications. car sa

lecture n'est pas aisée et ne fait pas ressortir d'une façon évidente des motifs réguliers dans la suite des arborescences. Nous avons néanmoins pu en extraire de l'information et en déduire que le comportement de la suite des arborescences est régulier. D'autres tableaux, pour différentes valeurs du vecteur (t, x, y, z) , nous ont permis d'identifier avec certitude des sous-groupes ordonnés des arborescences. Nous allons maintenant présenter le cheminement qui nous a permis de le faire.

4.3.1 Vecteurs de paramètres avec $t = x = y = z$

Si le vecteur des paramètres est (x, x, x, x) , $x > 0$, alors l'ordre des arborescences est $ABCDE$. Les figures (4.11), (4.12), (4.13) et (4.14) donnent la fréquence du rang des arborescences selon leur vraisemblance pour $x = 3$, $x = 2$, $x = 1$ et $x = 0,5$ respectivement. La figure (4.11) illustre très clairement qu'il y a une nette distinction entre la vraisemblance des différentes structures d'arborescences lorsque $t = x = y = z = 3$. Cette nette distinction existe aussi pour les valeurs de $t = x = y = z = 2, 1$ (voir les figures (4.12) et (4.13) respectivement) : quant à $t = x = y = z = 0,5$ (figure (4.14)) la distinction entre les arborescences C et D est perdue : les médianes des vraisemblances des arborescences C et D ont des valeurs similaires. Ceci peut s'expliquer par le fait que si $t = x = y = z = 0$ ($x = 0,5$ est proche de 0), alors tous les arborescences sont équivalentes.

4.3.2 Vecteurs de paramètres avec $t = 0, x = y = z$

Si le vecteur des paramètres est $(0, x, x, x)$, avec $x = 0,5, 1, 2$ ou 3 , nous avons prouvé que les arborescences Cxx et A ne sont pas identifiables entre elles et que les arborescences Bx et Dxx ne le sont pas non plus. Ainsi on écrira $A = Cxx$, $Bx = Dxx$. (Le tableau (4.1) donne le résumé des équivalences entre les arborescences.)

Les figures (4.15), (4.16), (4.17) et (4.18) montrent la fréquence du rang des arborescences selon leur vraisemblance pour $t = 0$ et $x = 3, x = 2, x = 1$ et $x = 0,5$ respectivement. Ces figures illustrent bien qu'il n'y a plus de distinction entre les arborescences A et

$Cxx(0, x)$ quelle que soit la valeur de x . Il n'y a pas non plus de différences entre les arborescences Bj et Dij lorsque i et j correspondent aux catégories 1, 2 ou 3 et la distinction entre les autres arborescences est très nette pour les figures (4.15) et (4.16), soit quand les vecteurs de paramètres sont $(0,3,3,3)$ et $(0,2,2,2)$. Lorsque $x = 1$ ou 0.5 (figures (4.17) et (4.18)) , on identifie encore les deux blocs de A et de Cxx et de Bj et de Dij pour i et j valant entre 1 et 3, mais la distinction entre les arborescences $Cx0$ et E a disparu pour $x = 1$, et pour $x = 0,5$, ce sont les arborescences $Cx0$, $Dx0$ et E qu'il est impossible de distinguer, les médianes des vraisemblances de chacune de ces arborescences étant similaires. La seule explication que nous pouvons donner est que si la catégorie 0 est observée dès la première expérience, alors ces arborescences sont toutes équivalentes dans leur différence avec l'arborescence A .

4.3.3 Vecteurs de paramètres avec $t \leq x = y = z$ et $t > x = y = z$

Les figures (4.19). (4.20). (4.21), (4.22) et (4.23) donnent la fréquence du rang des arborescences selon leur vraisemblance pour $t = 3$, $t = 2$, $t = 1$, $t = 0,5$ et $t = 0$ respectivement lorsque $x = y = z = 3$. Les figures (4.24),(4.25) et (4.26) montrent la fréquence du rang des arborescences selon leur vraisemblance pour $t = 1$, $t = 0,5$ et $t = 0$ respectivement lorsque $x = y = z = 2$ et les figures (4.27) et (4.28) présentent la même chose pour $t = 0,5$ et $t = 0$ lorsque $x = y = z = 1$. On voit sur toutes une nette distinction entre les vraisemblances des arborescences, sauf pour certaines valeurs particulières.

L'étude de la figure (4.10) et des figures (4.19) à (4.28) nous a permis de déduire une régularité dans le classement des arborescences et de répartir celles-ci en sous-groupes ordonnés : les rangs des arborescences d'un sous-groupe dit supérieur étant tous plus grands que les rangs des arborescences d'un sous-groupe dit inférieur. Les descriptions des sous-groupes des arborescences pour les vecteurs (t, x, x, x) , pour $t < x$ et $t > x$ sont données plus loin dans les tableaux (4.2) et (4.3) respectivement. Afin de confirmer cette subdivision du rang des arborescences, nous avons donné dans le tableau (4.4) la suite des arborescences pour chaque cellule (t, x) , pour $t < x$ et dans le tableau (4.5) cette

même suite pour chaque cellule (t, x) pour $t > x$. Les tableaux (4.4) et (4.5) ont eux aussi été déduits de la figure (4.10). Ils présentent la suite ordonnée des arborescences selon leurs vraisemblances en tenant compte des différents sous-groupes identifiables, quand il y a lieu de le faire.

4.4 Étude des simulations pour le vecteur de paramètres $(t = 0, x, y, z)$

La figure (4.29) contient les résultats de sortie d'un programme d'ordinateur pour le vecteur de paramètres $(t = 0, x = 1, u, v)$. Chaque cellule (u, v) du tableau de la figure (4.29) donne la suite des arborescences qui ont été ordonnées selon le rang médian de leurs vraisemblances. Les cellules (u, v) montrent la suite des arborescences pour le vecteur de paramètres $(t = 0, x = 1, u, v)$, avec $u, v = 0; 0, 5; 1; 2; 3$. Afin de faire ressortir plus facilement des motifs dans la suite des arborescences, nous avons utilisé la notation $x = 1, y = 2, z = 3$.

Les figures (4.30), (4.31) et (4.32) donnent la fréquence du rang des arborescences selon leur vraisemblance pour différentes valeurs du vecteur $(t = 0, x, y, z)$. Ces figures illustrent bien le fait qu'il y a une nette distinction entre la vraisemblance des différents arborescences non équivalentes lorsque $t = 0$. La figure (4.32), avec $x \neq y \neq z$ fait ressortir une démarcation particulièrement forte entre les différentes arborescences. On peut remarquer que l'arborescence avec la vraisemblance maximale est parfois A , parfois Cij , avec $i, j = 1, 2$ ou 3 : en effet, nous avons vu que ces arborescences sont équivalentes si $t = 0$.

Nous avons pu extraire de la figure (4.29) des sous-groupes d'arborescences plus généraux que pour le vecteur de paramètres de forme (t, x, x, x) . Les descriptions de ces sous-groupes d'arborescences pour le vecteur $(t, x, y, z) = (0, 1, 3, 3)$, c'est-à-dire pour la cellule (3,3) et pour le vecteur $(t, x, y, z) = (0, 1, 2, 3)$, c'est-à-dire pour la cellule (2,3) apparaissent respectivement dans les tableaux (4.6) et (4.7 et 4.8).

Sous-groupe	Description générale des arborescences du sous-groupe	Suite des arborescences	Description détaillée de la structure des arborescences
1	x issue directement de la racine t aussi loin que possible de la racine	$A C_{xx} B_x D_{xx}$ (si $t = 2$ ou 3 et $x = 3$ alors l'ordre est $ABCD$)	Ordre : l'ordre des arborescences est $ACBD$ pour $t \leq 1$ <u>Première expérience</u> : Les catégories avec le plus grand effet, soit x ($x > t$), sont observées. <u>Dernière expérience</u> : La catégorie avec le plus petit effet, soit t ($t < x$), est observée.
2	t issue directement de la racine pour B et C t issue du noeud intermédiaire lui-même issu directement de la racine pour D	$B_t D_{tx} C_{tx}$ (si $x = 3$ et $t \leq 1$ ou si $x = 2$ et $t = 0$ alors l'ordre est DBC)	Ordre : BDC est généralement suivi, ou sinon DBC pour $x = 3$ et $t \leq 1$ ou $x = 2$ et $t = 0$ <u>Première expérience (2e pour D)</u> : la catégorie avec le plus petit effet, soit t , est observée. <u>Dernière expérience</u> : Les catégories avec le plus grand effet (x) sont observées.
3	Arborescences de structure E et D avec t issue directement de la racine	$E_{xt} D_{xt}$	<u>Première expérience</u> : La catégorie t est observée pour D . <u>Dernière expérience</u> : les catégories avec les plus grands effets (x) sont observées.

Tableau 4.2 Sous-groupes d'arborescences pour les vecteurs de paramètres de la forme (t, x, x, x) avec $t < x$

Sous-groupe	Description générale des arborescences du sous-groupe	Suite des arborescences	Description détaillée de la structure des arborescences
1	t issue directement de la racine x aussi loin que possible de la racine	$A B_t C_{xt} D_{xt}$ (si $t = 3$ et $x = 2$ alors l'ordre est $AB_t C_{xt} B_x D_{xt}$)	Ordre : l'ordre des arborescences est $ABCD$ pour $t \leq 1$ <u>Première expérience</u> : La catégorie avec le plus grand effet, soit t ($t > x$), est observée. <u>Dernière expérience</u> : Les catégories avec les plus petits effets, soient x ($x < t$), sont observées.
2	x issue directement de la racine t issue du noeud intermédiaire lui-même issu directement de la racine	$B_x D_{tx} C_{xx}$	Ordre : BDC est généralement suivi, sauf pour $t = 3$ et $x = 2$ et pour $t < 1$ <u>Première expérience</u> : la catégorie avec le plus petit effet, soit x , est observée. La catégorie t est aussi observée à la 1ère expérience pour C et à la 2e pour B et D . <u>Dernière expérience</u> : Les catégories avec le plus petit effet (x) sont observées.
3	Arborescences de structure E et D avec t aussi loin que possible de la racine	$D_{xx} E_{xx}$	<u>Première expérience</u> : La catégorie x est observée pour D . <u>Dernière expérience</u> : La catégorie avec le plus petit effet (t) est observée.

Tableau 4.3 Sous-groupes d'arborescences pour les vecteurs de paramètres de la forme (t, x, x, x) avec $t > x$

	$x = 0$	$x = 0,5$	$x = 1$	$x = 2$	$x = 3$
$t = 3$					$ABCDE$
$t = 2$				$ABCDE$	$A B_x C_{xx} D_{xx}$ $B_t D_{tx} C_{tx}$ $E_{xt} D_{xt}$
$t = 1$			$ABCDE$	$A C_{xx} \sim B_x D_{xx}$ $B_t D_{tx} C_{tx}$ $E_{xt} D_{xt}$	$A C_{xx} B_x D_{xx}$ $D_{tx} \gtrsim B_t C_{tx}$ $E_{xt} \sim D_{xt}$
$t = 0,5$		$ABC \sim DE$	$A C_{xx} \sim B_x D_{xx}$ $B_t D_{tx} \sim C_{tx}$ $E_{xt} \sim D_{xt}$	$A C_{xx} B_x D_{xx}$ $B_t D_{tx} C_{tx}$ $E_{xt} D_{xt}$	$A C_{xx} B_x D_{xx}$ $D_{tx} \sim B_t C_{tx}$ $E_{xt} D_{xt}$
$t = 0$		$A = C_{xx} B_x = D_{xx}$ $B_t D_{tx}$ $C_{tx} \sim E_{tx} \sim D_{tx}$	$A = C_{xx} B_x = D_{xx}$ $B_t D_{tx}$ $C_{tx} \sim E_{tx} D_{tx}$	$A = C_{xx} B_x = D_{xx}$ $D_{tx} \sim B_t C_{tx}$ $E_{xt} D_{xt}$	$A = C_{xx} B_x = D_{xx}$ $D_{tx} \sim B_t C_{tx}$ $E_{xt} D_{xt}$

Tableau 4.4 Ordre des arborescences pour chacune des valeurs de (t, x, x, x) , pour $t < x$

	$x = 0$	$x = 0,5$	$x = 1$	$x = 2$	$x = 3$
$t = 3$		$AB_t C_{xt} \sim D_{xt}$ $B_x \sim D_{tx}$ $C_{xx} \lesssim D_{xx} \sim E_{xx}$	$AB_t C_{xt} \sim D_{xt}$ $B_x D_{tx}$ $C_{xx} \sim D_{xx} E_{xx}$	$AB_t C_{xt} \sim B_x$ $D_{xt} D_{tx}$ $C_{xx} D_{xx} E_{xx}$	ABCDE
$t = 2$		$AB_t C_{xt} \sim D_{xt}$ $B_x \sim D_{tx}$ $C_{xx} \sim D_{xx} \sim E_{xx}$	$AB_t C_{xt} D_{xt}$ $B_x D_{tx}$ $C_{xx} D_{xx} E_{xx}$	ABCDE	
$t = 1$		$AB_t C_{xt}$ $D_{xt} \sim B_x D_{tx}$ $C_{xx} \sim D_{xx} E_{xx}$	ABCDEF		
$t = 0,5$		ABC ~ DE			
$t = 0$					

Tableau 4.5 Ordre des arborescences pour chacune des valeurs de (t, x, x, x) , pour $t > x$

Sous-groupe	Description générale des arborescences	Suite des arborescences	Description détaillée de la structure des arborescences
1	Arborescences équivalentes à l'arborescence A	$AC_{zz}C_{zx}$	Arborescence A et arborescences C avec les catégories xz ou zz issues directement de la racine
2	Arborescences équivalentes à (ou quasi-équivalentes) l'arborescence B_z	$D_{zz}B_zD_{xz} \sim D_{0z}$	<u>Première expérience</u> : la catégorie avec l'effet le plus grand (z) est observée. <u>Dernière expérience</u> : la catégorie 0 est observée.
3	Arborescences équivalentes à l'arborescence B_x	$D_{zx} B_x$	<u>Première expérience</u> : la catégorie avec le 2e effet le plus grand (x) est observée. <u>Dernière expérience</u> : la catégorie 0 est observée.
4	Arborescences caractérisées par la dernière expérience et avec t issue directement de la racine	Groupe 1 : $C_{z0}B_0E_{zx}D_{z0}$ Groupe 2 : $D_{0x}D_{x0}C_{x0}E_{x0}$	<u>Première expérience</u> : la catégorie $t = 0$ est observée. <u>Dernière expérience</u> . <u>Groupe 1</u> : les catégories avec les effets x et z sont observées. <u>Groupe 2</u> : les catégories avec les effets z et z sont observées.

Tableau 4.6 Description des sous-groupes d'arborescences pour un vecteur de paramètres ($t = 0$, $x = 1$, $y = 3$, $z = 3$).

Sous-groupe	Description générale des arborescences	Suite des arborescences	Description détaillée de la structure des arborescences
1	Arborescences équivalentes à l'arborescence A	$C_{zx}C_{yx}AC_{zy}$	Arborescence A et arborescences C avec les catégories zx , yx ou zy issues directement de la racine
2	Arborescences équivalentes à l'arborescence B_z	$D_{xz}B_zD_{yz}$	<u>Première expérience</u> : la catégorie avec l'effet le plus grand (z) est observée. <u>Dernière expérience</u> : la catégorie 0 est observée.
3	Arborescences équivalentes à l'arborescence B_y	$D_{xy}D_{zy}B_y$	<u>Première expérience</u> : la catégorie avec le 2e effet le plus grand (y) est observée. <u>Dernière expérience</u> : la catégorie 0 est observée.
4	Arborescences équivalentes à l'arborescence B_x et arborescences D_{0z} et D_{0y}	$D_{yx}D_{zx}B_x$ $D_{0z}D_{0y}$	<u>Première expérience</u> : la catégorie avec le 3e effet le plus grand (x) est observée. <u>Dernière expérience</u> : la catégorie 0 est observée. Les arborescences D_{0z} et D_{0y} se rattachent au même sous-groupe.

Tableau 4.7 Description des sous-groupes d'arborescences pour un vecteur de paramètres ($t = 0, x = 1, y = 2, z = 3$)

Sous-groupe	Description générale des arborescences	Suite des arborescences	Description détaillée de la structure des arborescences
5	Arborescences caractérisées par la dernière expérience et avec t issue directement de la racine (sauf pour E)	<p>Groupe 1 : $C_{z0}E_{yx}B_0$</p> <p>Groupe 2 : $C_{y0}E_{y0}D_{z0}D_{y0}$</p> <p>Groupe 3 : $D_{0x}D_{x0}C_{x0}E_{x0}$</p>	<p><u>Première expérience</u> la catégorie $t = 0$ est observée en premier</p> <p><u>Dernière expérience :</u> <u>Groupe 1 :</u> les catégories avec les 2 effets les plus petits (x et y) sont observées.</p> <p><u>Groupe 2 :</u> les catégories avec les effets x et z sont observées (avec une exception).</p> <p><u>Groupe 3 :</u> les catégories avec les effets les plus grands (y et z) sont observées.</p>

Tableau 4.8 Description des sous-groupes d'arborescences pour un vecteur de paramètres ($t = 0, x = 1, y = 2, z = 3$), suite.

4.5 Résumé global des résultats des simulations

Si nous voulons maintenant résumer les résultats de nos simulations, les observations suivantes peuvent être faites :

- Si les effets sont tous de même grandeur et différents de 0 ($t = x = y = z \neq 0$), alors l'ordre des arborescences selon la vraisemblance est $ABCDE$;
- Si les effets sont tous de même grandeur et valent tous 0 ($t = x = y = z = 0$), alors les arborescences sont toutes équivalentes ;
- Si les effets sont de grandeurs différentes avec un vecteur de paramètres (t, x, y, z) de la forme $t = 0 < x < y < z$, alors on peut séparer les arborescences en sous-groupes $i = 1, \dots, 5$. Chaque sous-groupe i contient des arborescences pour lesquelles la vraisemblance est supérieure à celle des arborescences du sous-groupe $i + 1$. L'ordre des arborescences à l'intérieur de ces sous-groupes n'a pas encore été déterminé, mais il semble certain que cet ordre dépende de la valeur des paramètres. Les sous-groupes sont, dans le cas qui nous occupe :

1. arborescences $A = Cyz = Cxz = Cxy$.
2. arborescences de type $Bz = Dyz = Dxz$.
3. arborescences de type $By = Dzy = Dxy$.
4. arborescences de type $Bx = Dyx = Dzx$, ainsi que $D0z$ et $D0y$.
5. arborescences $Ci0, Di0$ où $i = x, y, z$: arborescences $B0, D0x$ et toutes les arborescences E .

Si des éléments du vecteur (t, x, y, z) sont égaux, par exemple si $x = y$, alors les sous-groupes avec les arborescences Bx et By (3 et 4) contiennent les mêmes arborescences : il y a ainsi un sous-groupe de moins ;

- Si l'effet de la catégorie t est 0 et les effets des autres catégories sont identiques (donc si on a un vecteur de paramètres de la forme $(0, x, x, x)$), les sous-groupes 2,3,4 définis plus haut (ceux contenant Bi et les arborescences équivalentes, avec $i = x, y, z$) se mêlent et ne forme plus qu'un groupe indistinct. Si la valeur de x est assez grande, le dernier groupe tend à se scinder en plusieurs sous-groupes constitués des arborescences

$Cx0$, $Dx0$ et E nettement séparées les unes des autres ;

– Si le vecteur des paramètres est de la forme (t, x, x, x) , avec $t \neq 0$, on observe les sous-groupes suivants pour les arborescences quand elles sont classées selon leurs vraisemblances :

1. arborescence A
2. arborescences Cxx
3. arborescences Bx
4. arborescences Dxx
5. arborescences Bt et Dtx
6. arborescences Cxt, E et Dxt généralement groupées par arborescences de même structure, mais dont l'ordre varie.

Remarquons que l'ordre des sous-groupes 2 et 3 peut être interverti, comme c'est le cas lorsque $t = 2$ et $x = 3$, et que certains groupes peuvent se mélanger, particulièrement quand valeurs de t et de x sont faibles.

Nous avons donc pu mettre en évidence certains principes dans la façon dont les arborescences se classent selon l'ordre de leurs vraisemblances. Si certaines équivalences devraient être les mêmes quelle que soit l'arborescence utilisée pour générer les données, comme l'équivalence entre les arborescences A et C et B et D appropriées en présence de modèles réduits, il pourrait être intéressant de mener d'autres études avec une autre arborescence de départ, pour voir comment l'ordre est affecté (notamment en ce qui concerne les arborescences de structure E , qui semblent les plus éloignées de l'arborescence A).

```

PARAMETRE NUMERO 33
  1.0 -1.0 .0 .0 .0 .0 3.0 -3.0 .0
  1.0 -1.0 .0 .0 2.0 -2.0 .0 .0 .0
  1.0 -1.0 2.0 -2.0 .0 .0 .0 .0 .0

NOMBRE D OCURENCES PAR ENVIRONNEMENT 10000

RANG ET FREQ DE RANG DE CHAQUE ARBRE
RANG DES ARBRES
  1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 MED SD
STRUCTURE
D34 - - - - - - - - - - - - - - - - - - - - 23 27 FIN -120279. 274.
D24 - - - - - - - - - - - - - - - - - - - - 27 23 FIN -120238. 287.
E34 - - - - - - - - - - - - - - - - - - - - 25 25 - - FIN -119369. 283.
E24 - - - - - - - - - - - - - - - - - - - - 25 25 - - FIN -119367. 296.
C34 - - - - - - - - - - - - - - - - - - - - 25 25 - - FIN -118749. 279.
C24 - - - - - - - - - - - - - - - - - - - - 25 25 - - FIN -118739. 297.
D14 - - - - - - - - - - - - - - - - - - - - 5 45 - - - - FIN -118223. 299.
E23 - - - - - - - - - - - - - - - - - - - - 45 5 - - - - FIN -118176. 299.
D43 - - - - - - - - - - - - - - - - - - - - 19 31 - - - - FIN -117606. 288.
D42 - - - - - - - - - - - - - - - - - - - - 31 19 - - - - FIN -117569. 301.
B4 - - - - - - - - - - - - - - - - - - - - 18 32 - - - - FIN -116529. 284.
C14 - - - - - - - - - - - - - - - - - - - - 32 18 - - - - FIN -116514. 299.
D41 - - - - - - - - - - - - - - - - - - - - 14 7 29 - - - - FIN -115888. 305.
D32 - - - - - - - - - - - - - - - - - - - - 19 20 11 - - - - FIN -115848. 301.
D23 - - - - - - - - - - - - - - - - - - - - 17 23 10 - - - - FIN -115846. 283.
D13 - - - - - - - - - - - - - - - - - - - - 20 30 - - - - FIN -115231. 301.
D12 - - - - - - - - - - - - - - - - - - - - 30 20 - - - - FIN -115227. 303.
B3 - - - - - - - - 5 15 30 - - - - - - - - - - - - - - FIN -114724. 290.
B2 - - - - - - - - 7 25 18 - - - - - - - - - - - - - - FIN -114710. 297.
C23 - - - - - - - - 38 10 2 - - - - - - - - - - - - - - FIN -114628. 299.
D21 - - - - - - - - 28 22 - - - - - - - - - - - - - - FIN -114347. 312.
D31 - - - - - - - - 22 28 - - - - - - - - - - - - - - FIN -114337. 309.
C13 - - - 17 33 - - - - - - - - - - - - - - - - - - - FIN -113815. 304.
C12 - - - 33 17 - - - - - - - - - - - - - - - - - - - FIN -113795. 304.
B1 - - 50 - - - - - - - - - - - - - - - - - - - - - - - FIN -113556. 315.
A 50 - - - - - - - - - - - - - - - - - - - - - - - - - - FIN -112664. 301.
NUMERO DE PARAMETRES NO 34

```

Figure 4.6 Ordre des arborescences pour le vecteur de paramètres $(t, x, y, z) = (1, 2, 2, 3)$

PARAMETRE NUMERO 40																															
.0	.0	.0	.0	.0	.0	.0	3.0	-3.0	.0																						
.0	.0	.0	.0	2.0	-2.0	.0	.0	.0	.0																						
.0	.0	1.0	-1.0	.0	.0	.0	.0	.0	.0																						
NOMBRE D OCCURENCES PAR ENVIRONNEMENT									10000																						
RANG ET FREQ DE RANG DE CHAQUE ARBRE																															
RANG DES ARBRES																															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	MED	SD			
STRUCTORD																															
E34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1 49	FIN	-119428.	366.		
C34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5 45	-	FIN	-119402.	367.	
D34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	45 4 1	FIN	-119337.	362.		
D43	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	50	-	-	FIN	-117803.	362.
D24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	50	-	-	FIN	-117585.	347.
D14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1 49	-	-	FIN	-116849.	344.
E24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	49 1	-	-	FIN	-116652.	341.
C24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1 49	-	-	FIN	-116471.	347.
B4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	49 1	-	-	FIN	-116302.	342.
E23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	50	-	-	FIN	-116071.	328.
C14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4 46	-	-	FIN	-115806.	340.
D42	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	46 4	-	-	FIN	-115737.	349.
D41	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	14 1 35	-	-	FIN	-115511.	344.
B3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	9 22 13 '6	-	-	FIN	-115425.	349.
D13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	13 12 21 4	-	-	FIN	-115425.	349.
D23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	14 16 15 5	-	-	FIN	-115424.	349.
B2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	11 24 15	-	-	FIN	-114914.	344.
D12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	13 22 15	-	-	FIN	-114914.	344.
D32	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	26 4 20	-	-	FIN	-114913.	343.
D21	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	15 16 19	-	-	FIN	-114686.	337.
B1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	10 26 14	-	-	FIN	-114686.	337.
D31	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	25 8 17	-	-	FIN	-114605.	337.
C12	7	15	18	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-114505.	340.
A	7	17	17	9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-114505.	340.
C23	25	3	6	16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-114504.	340.
C13	11	15	9	15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-114504.	340.
NUMERO DE PARAMETRES NO									41																						

Figure 4.7 Ordre des arborescences pour le vecteur de paramètres $(t, x, y, z) = (0, 1, 2, 3)$

PARAMETRE NUMERO 50																																		
.0	.0	.0	.0	.0	.0	.0	3.0	-3.0	.0																									
.0	.0	.0	.0	.0	2.0	-2.0	.0	.0	.0																									
.0	.0	.0	.0	.0	.0	.0	.0	.0	.0																									
NOMBRE D OCURENCES PAR ENVIRONNEMENT 10000																																		
RANG ET FREQ DE RANG DE CHAQUE ARBRE																																		
RANG DES ARBRES																																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	MEAN	SD						
STRUBTURD																																		
E34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	23	27	FIN	-120669.	340.				
C34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	27	23	FIN	-120669.	340.				
D34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	28	22	-	-	FIN	-119433.	339.	
D43	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	22	28	-	-	FIN	-119413.	343.
D13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	9	11	11	5	5	9	-	-	-	-	-	-	-	-	FIN	-116988.	344.
B3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6	13	9	6	11	3	-	-	-	-	-	-	-	-	FIN	-116988.	344.
D23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	14	7	10	9	3	7	-	-	-	-	-	-	-	-	FIN	-116988.	344.
D14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5	4	10	6	5	20	-	-	-	-	-	-	-	-	FIN	-116984.	340.
B4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4	12	3	5	21	5	-	-	-	-	-	-	-	-	FIN	-116984.	340.
D24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	10	3	7	19	5	6	-	-	-	-	-	-	-	-	FIN	-116984.	341.
D41	3	2	3	3	1	3	3	1	1	3	7	2	2	3	7	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-116265.	346.
C14	1	3	2	4	2	7	1	-	3	7	-	1	2	1	9	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-116264.	346.
C13	3	5	3	2	1	-	6	-	2	5	5	1	2	6	6	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-116264.	346.
E23	3	2	6	2	4	-	1	2	3	3	1	2	2	6	1	12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-116264.	346.
D31	4	4	4	1	-	1	2	3	2	4	1	8	2	4	6	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-116264.	346.
E24	8	3	3	1	3	3	2	1	2	1	2	3	3	6	2	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-116264.	346.
C23	4	6	2	4	3	4	2	2	6	1	4	3	1	1	5	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-116264.	346.
B1	-	1	-	1	6	3	3	5	5	3	3	5	9	4	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-116263.	346.
D21	1	-	-	6	1	3	8	3	4	3	3	6	6	3	2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-116263.	346.
D32	6	2	2	5	4	5	3	5	3	3	4	-	3	3	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-116263.	346.
A	-	2	-	2	6	5	5	5	6	3	6	4	1	4	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-116263.	346.
C12	1	-	3	-	5	6	5	9	2	7	4	2	4	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-116263.	346.
D42	5	5	8	6	1	1	2	4	-	2	5	1	1	4	3	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-116263.	346.
C24	5	6	7	1	6	1	1	2	5	1	2	5	3	1	3	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-116263.	346.
B2	2	6	1	6	3	5	4	5	4	-	2	3	5	2	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-116262.	346.
D12	4	3	6	6	4	3	2	3	2	4	1	4	4	2	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-116262.	346.
NUMERO DE PARAMETRES NO 51																																		

Figure 4.8 Ordre des arborescences pour le vecteur de paramètres $(t, x, y, z)=(0,0,2,3)$

```

PARAMDTRD NUMDRO 125
.0 .0 .0 .0 .0 .0 3.0 -3.0 .0
.0 .0 .0 .0 .0 .0 .0 .0 .0
.0 .0 .0 .0 .0 .0 .0 .0 .0

NOMBRD C OBBURDNBS PAR ENVIRONNMDMT 10000

RANG DT PRDQ CU RANG CD BEAQUO ARERD
RANG CDS ARERDS
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 MDC SC
STRUBTURD
C23 - 1 3 2 3 3 4 2 3 3 1 - 2 2 - 1 1 2 2 - - - 2 3 5 5 PIN -135402. 274.
E23 - 4 2 3 1 2 3 5 4 1 - 1 2 1 1 - 2 2 1 1 - 1 2 4 4 3 PIN -135402. 274.
D23 6 1 3 2 1 2 3 2 1 1 1 2 2 - - 3 1 1 1 2 1 1 7 4 - 2 PIN -135402. 274.
D12 6 1 1 1 5 1 1 2 2 3 1 - - - 1 2 3 1 2 3 3 3 1 1 6 - PIN -135402. 274.
B2 - 6 1 - 2 5 2 1 2 1 - 4 - - - 1 2 3 2 2 3 2 3 1 1 6 PIN -135402. 274.
D32 3 1 6 3 1 2 4 2 1 2 1 1 1 - 1 - 1 1 1 3 4 2 3 2 2 2 PIN -135402. 274.
D42 6 - 3 2 2 3 1 1 2 2 - 1 1 1 - - 2 - - 3 2 3 5 3 2 5 PIN -135402. 274.
A - - - - 3 1 1 3 2 3 1 1 2 6 3 1 7 5 3 2 2 1 - 2 - 1 PIN -135402. 274.
C13 - - - 1 1 - 2 4 1 4 1 3 4 4 3 6 3 - 2 3 3 - 4 - 1 - PIN -135402. 274.
C12 - 2 1 3 - - 2 2 2 1 1 3 5 2 3 5 1 6 5 3 2 - - 1 - - PIN -135402. 274.
D31 - 1 - 2 2 2 2 4 2 1 5 4 2 3 6 2 5 2 1 1 3 - - - - PIN -135402. 274.
D21 3 - 2 1 1 2 3 2 4 3 7 - 2 4 5 5 1 2 2 1 - - - - PIN -135402. 274.
D41 - - 1 1 2 1 4 2 1 6 2 5 2 8 5 2 3 1 1 3 - - - - PIN -135402. 274.
C14 1 - 1 - 1 2 2 2 4 1 1 4 4 7 3 2 4 6 3 1 - - 1 - - PIN -135402. 274.
B1 - 1 2 1 - 3 1 1 2 2 3 4 8 2 2 7 3 3 2 - - 3 - - - PIN -135402. 274.
E24 2 4 3 2 1 2 1 1 - 1 3 1 1 - 5 - 1 2 1 1 4 4 2 5 3 - PIN -135401. 274.
B3 - 5 2 - 4 3 2 1 3 2 3 2 - - 1 1 - 1 4 3 2 5 1 1 1 3 PIN -135401. 274.
D13 4 1 1 2 4 4 1 2 2 3 - 1 2 1 1 1 1 3 2 1 5 2 2 - 3 1 PIN -135401. 274.
C24 1 5 2 2 3 1 - 1 3 1 1 1 2 3 - 2 - - 2 2 1 4 3 3 5 2 PIN -135401. 274.
D24 3 1 1 5 3 2 1 1 - - 2 1 - 2 1 - 5 1 2 2 2 3 4 4 - 4 PIN -135401. 274.
B4 1 4 - - 2 4 1 1 - 2 3 4 2 1 - 3 - 2 1 5 1 7 - 1 2 3 PIN -135401. 274.
D14 3 1 1 3 3 1 - - 2 1 5 3 1 - 5 1 - 1 3 2 7 1 - 1 4 1 PIN -135401. 274.
C34 - 3 4 5 3 - 4 3 2 2 - 2 - - 2 1 2 - 2 2 - 1 1 2 5 4 PIN -135401. 274.
E34 2 2 6 3 2 1 2 4 1 1 2 - 2 2 - 3 - 4 - - 1 - 1 4 5 2 PIN -135401. 274.
D43 5 3 - 3 - 2 1 1 1 2 1 2 1 2 1 2 1 2 1 2 2 3 4 4 4 - 2 PIN -135401. 274.
D34 4 3 4 3 - 1 2 - 3 2 4 1 1 - - - - 3 2 1 3 4 4 1 4 PIN -135401. 274.
DNSMELD CD PARAMDTRDS NO 26
    
```

Figure 4.9 Ordre des arborescences pour le vecteur de paramètres $(t, x, y, z)=(0,0,0,3)$

```

EN ORDONNEE: VALEUR DU PARAMETRE POUR CATEGORIES 1
3
D..E..C..D..B...B...C..D. AB...C..D..B...D..D..C..E AB...C..D..B...D..C..D..E AB...C..B...D..D..C..D..E AB...C..D..E.....
00.00.00.30.0...3...30.03 .3...35.53.5...35.55.55.5 .3...31.13.1...31.11.11.1 .3...32.2...23.32.22.22.2 .3...33.33.33.....
30.30.30.00.300.000.00.00 .555.55.55.355.55.35.35.3 .111.11.11.311.11.31.31.3 .222.22.322.22.22.32.32.3 .333.33.33.33.....
3
.A.....
.....
.....
5.....
5.....
1.....
2.....
2
D..E..C..D..B...D..C..AB. AB...C..D..B...D..D..C..E AB...C..D..B...D..C..D..E AB...C..D..E..... AB...C..D..B...D..C..E..D
00.00.00.20.0...02.20..2 .2...25.52.5...25.55.55.5 .2...21.12.1...21.11.11.1 .2...22.22.22..... .3...33.33.2...23.23.33.3
20.20.20.00.200.00.00.00 .555.55.55.255.55.25.25.2 .111.11.11.211.11.21.21.2 .222.22.22.22..... .233.23.23.333.33.33.23.3
2
.....
.....
.....
5.....
5.....
0.....
1
B...D..D..C..E..C..D..AB. AB...C..D..B...D..D..C..E AB...C..D..E..... AC..B...D..B...D..C..E..D AC..B...D..B...C..E..D
0...10.00.00.00.10.01..1 .1...15.51.5...15.55.55.5 .1...11.11.11..... .22.2...22.1...12.12.22.2 .33.3...33.13.1...13.33.3
100.00.10.10.10.00.00.00 .555.55.55.155.55.15.15.1 .111.11.11.11..... .12.122.12.222.22.22.12.2 .13.133.13.33.333.33.13.3
1
.....
.....
5.....
5.....
0.....
5
D..E..C..D..B...C..D..B.. AB...C..D..E..... AB...C..D..B...D..C..D..E AC..B...D..B...D..C..E..D AC..B...D..B...C..D..E
50.50.50.00.500.00.00.000 .555.55.55.55..... .511.51.51.111.11.11.11.5 .52.522.52.222.22.22.52.2 .53.533.53.33.333.33.33.5
5
.A.....
.....
.....
1.....
5.....
3.....
2.....
0
D..B...C..AE..... AC..B...D..B...D..E..C..D C..AB...D..B...D..C..E..D C..AD..B...D..B...C..E..D C..AB...D..B...C..D..E
00.0...00.00..... .55.5...55.0...05.55.05.5 11...11.11.0...01.01.11.1 22..22.2...02.0...02.22.2 33..3...33.03.0...03.30.3
00.000.00.00..... .05.055.05.555.55.05.55.5 01...011.01.111.11.11.01.1 02..02.022.22.222.22.02.2 03..033.03.33.333.33.33.0
0
.....
.....
0.....
5.....
1.....
2.....
3.....
0
5
1
2
3
VALEUR DU PARAMETRE DE LA CATEGORIE 4
VALEUR DU PARAMETRE DE LA CATEGORIE 3
VALEUR DU PARAMETRE DE LA CATEGORIE 2

```

Figure 4.10 Sortie d'ordinateur : Ordre des arborescences pour un vecteur de paramètres de la forme (t, x, x, x)

PARAMETRE NUMERO 1																													
3.0	-3.0	.0	.0	.0	.0	3.0	-3.0	.0																					
3.0	-3.0	.0	.0	3.0	-3.0	.0	.0	.0																					
3.0	-3.0	3.0	-3.0	.0	.0	.0	.0	.0																					
NOMBRE D OCURENCES PAR ENVIRONNEMENT 10000																													
RANG ET FREQ DE RANG DE CHAQUE ARBRE																													
RANG DES ARBRES																													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	MED	SD	
STRUBTURD																													
E24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	10	17	23	FIN	-126849.	138.
E23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	15	18	17	FIN	-126813.	117.
E34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	25	15	10	FIN	-126768.	144.
D13	-	-	-	-	-	-	-	-	-	-	5	1	5	3	4	3	3	3	9	7	4	3	-	-	-	-	FIN	-125381.	109.
D31	-	-	-	-	-	-	-	-	-	-	2	4	2	5	3	5	3	7	5	5	5	4	-	-	-	-	FIN	-125369.	102.
D41	-	-	-	-	-	-	-	-	-	-	3	7	3	3	3	5	6	2	3	2	4	9	-	-	-	-	FIN	-125368.	114.
D14	-	-	-	-	-	-	-	-	-	-	6	2	1	5	5	4	4	4	3	5	9	2	-	-	-	-	FIN	-125366.	122.
D24	-	-	-	-	-	-	-	-	-	-	2	2	5	3	8	4	9	6	1	3	3	4	-	-	-	-	FIN	-125354.	122.
D23	-	-	-	-	-	-	-	-	-	-	3	7	6	2	4	3	3	8	4	6	1	3	-	-	-	-	FIN	-125348.	119.
D43	-	-	-	-	-	-	-	-	-	-	7	5	6	4	2	4	6	-	6	4	1	5	-	-	-	-	FIN	-125342.	130.
D32	-	-	-	-	-	-	-	-	-	-	3	6	2	4	5	7	-	5	5	7	2	4	-	-	-	-	FIN	-125342.	139.
D12	-	-	-	-	-	-	-	-	-	-	6	6	5	3	1	4	3	5	5	3	4	5	-	-	-	-	FIN	-125340.	136.
D42	-	-	-	-	-	-	-	-	-	-	4	5	3	7	7	2	4	2	1	2	8	5	-	-	-	-	FIN	-125337.	148.
D21	-	-	-	-	-	-	-	-	-	-	5	3	4	5	3	5	5	4	5	3	4	4	-	-	-	-	FIN	-125332.	133.
D34	-	-	-	-	-	-	-	-	-	-	4	2	8	6	5	4	4	4	3	3	5	2	-	-	-	-	FIN	-125328.	125.
C13	-	-	-	-	6	5	8	11	10	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-124032.	107.
C24	-	-	-	-	5	7	14	11	5	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-123995.	134.
C14	-	-	-	-	6	12	5	5	11	11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-123994.	114.
C12	-	-	-	-	11	5	10	11	5	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-123980.	136.
C23	-	-	-	-	10	8	6	5	13	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-123964.	125.
C34	-	-	-	-	12	13	7	7	6	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-123953.	123.
B1	-	11	14	13	12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-121238.	115.
B3	-	12	16	10	12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-121220.	119.
B2	-	18	12	5	15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-121216.	140.
B4	-	9	8	22	11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-121210.	124.
A	50	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-116604.	124.
NUMERO DE PARAMETRES NO 2																													

Figure 4.11 Ordre des arborescences pour le vecteur de paramètres $(x, x, x, x)=(3,3,3,3)$

PARAMTRD NUMDRO 157

2.0	-2.0	.0	.0	.0	.0	2.0	-2.0	.0
2.0	-2.0	.0	.0	2.0	-2.0	.0	.0	.0
2.0	-2.0	2.0	-2.0	.0	.0	.0	.0	.0

NOMERD C OBBURDNBDS PAR DNVRONNDMDNT 10000

RANG DT FRDQ CU RANG CD BBAQUD ARERD

STRUBTURD	RANG CDS ARERDS																										MDC	SC		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26				
E24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	12	18	20	FIN	-137450.	198.	
E23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	21	16	13	FIN	-137445.	203.	
E34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	17	16	17	FIN	-137437.	197.	
D34	-	-	-	-	-	-	-	-	-	-	-	3	3	1	-	7	8	3	6	3	3	6	7	-	-	-	-	FIN	-135835.	207.
D43	-	-	-	-	-	-	-	-	-	-	-	2	4	5	4	3	3	7	2	6	3	7	4	-	-	-	-	FIN	-135819.	204.
D24	-	-	-	-	-	-	-	-	-	-	-	6	1	3	4	5	2	1	7	6	3	7	5	-	-	-	-	FIN	-135817.	199.
D41	-	-	-	-	-	-	-	-	-	-	-	-	7	6	2	5	5	5	3	4	3	6	4	-	-	-	-	FIN	-135802.	195.
D23	-	-	-	-	-	-	-	-	-	-	-	4	5	5	6	6	4	2	7	1	4	2	4	-	-	-	-	FIN	-135798.	200.
D13	-	-	-	-	-	-	-	-	-	-	-	5	2	1	6	7	3	2	5	7	5	1	6	-	-	-	-	FIN	-135790.	207.
D32	-	-	-	-	-	-	-	-	-	-	-	7	5	5	2	4	3	5	4	5	3	4	3	-	-	-	-	FIN	-135778.	199.
D31	-	-	-	-	-	-	-	-	-	-	-	4	7	4	3	2	8	3	3	4	6	4	2	-	-	-	-	FIN	-135776.	198.
D42	-	-	-	-	-	-	-	-	-	-	-	4	4	5	3	2	5	7	4	2	7	4	3	-	-	-	-	FIN	-135775.	187.
D12	-	-	-	-	-	-	-	-	-	-	-	5	6	6	10	-	2	4	3	4	2	3	5	-	-	-	-	FIN	-135773.	176.
D14	-	-	-	-	-	-	-	-	-	-	-	1	5	5	5	7	3	4	2	4	6	3	5	-	-	-	-	FIN	-135771.	201.
D21	-	-	-	-	-	-	-	-	-	-	-	9	1	4	5	2	4	7	4	4	5	3	2	-	-	-	-	FIN	-135765.	186.
C34	-	-	-	-	-	6	7	9	8	11	9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-134884.	209.
C24	-	-	-	-	-	6	9	8	7	10	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-134831.	200.
C14	-	-	-	-	-	9	10	5	8	9	9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-134830.	198.
C23	-	-	-	-	-	12	7	10	8	7	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-134816.	202.
C12	-	-	-	-	-	11	7	7	10	7	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-134813.	180.
C13	-	-	-	-	-	6	10	11	9	6	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-134805.	201.
B3	-	11	10	14	15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-133223.	211.
B4	-	7	15	14	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-133203.	208.
B2	-	16	16	8	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-133184.	189.
B1	-	16	9	14	11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-133163.	207.
A	50	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-130421.	211.

DNSDMELD CD PARAMTRDS NO 58

Figure 4.12 Ordre des arborescences pour le vecteur de paramètres $(x, x, x, x)=(2,2,2,2)$

```

PARAMTRD NUMDRO 313
1.0 -1.0 .0 .0 .0 .0 1.0 -1.0 .0
1.0 -1.0 .0 .0 1.0 -1.0 .0 .0 .0
1.0 -1.0 1.0 -1.0 .0 .0 .0 .0 .0

NOMERD C OBBURDNBS PAR ENVIRONNMENT 10000

RANG DT FRDQ CU RANG CD BHAQUD ANERD
RANG CDS ARERDS
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 MDC SC
STRUBTURD
E34 - - - - - - - - - - - - - - - - - - - - - 14 21 15 PIN -178574. 225.
E24 - - - - - - - - - - - - - - - - - - - - - 13 18 19 PIN -178532. 220.
E23 - - - - - - - - - - - - - - - - - - - - - 23 11 16 PIN -178517. 232.
D41 - - - - - - - - - - - 2 5 3 6 6 4 7 - 6 3 2 3 3 - - - PIN -178012. 227.
D31 - - - - - - - - - - - 1 3 1 5 6 6 5 5 4 3 4 3 4 - - - PIN -178005. 225.
D34 - - - - - - - - - - - 1 4 2 2 1 3 6 5 3 4 8 9 2 - - - PIN -178005. 222.
D14 - - - - - - - - - - - 1 6 2 2 5 3 7 2 10 6 2 4 - - - PIN -178003. 226.
D24 - - - - - - - - - - - 3 2 3 8 2 2 7 3 6 4 6 4 - - - PIN -177998. 221.
D21 - - - - - - - - - - - 1 3 3 6 5 6 5 5 3 3 3 5 2 - - - PIN -177993. 234.
D42 - - - - - - - - - - - 5 5 3 4 3 3 4 6 3 6 3 5 - - - PIN -177992. 229.
D43 - - - - - - - - - - - 2 4 6 3 3 3 2 4 4 4 2 13 - - - PIN -177992. 224.
D13 - - - - - - - - - - - 7 2 4 5 3 4 3 4 3 6 5 4 - - - PIN -177986. 226.
D23 - - - - - - - - - - - 4 6 4 - 4 5 2 8 6 1 8 2 - - - PIN -177981. 229.
D32 - - - - - - - 1 1 - - 3 7 5 2 5 3 5 5 2 3 3 5 - - - PIN -177978. 236.
D12 - - - - - - - - - - - 1 6 6 3 8 6 4 5 2 3 3 1 2 - - - PIN -177977. 229.
C14 - - - - - 8 9 7 11 7 6 - 1 1 - - - - - - - - - - - PIN -177863. 227.
C34 - - - - - 6 6 4 7 14 12 1 - - - - - - - - - - - - - PIN -177861. 227.
C12 - - - - - 8 12 12 8 7 3 - - - - - - - - - - - - - PIN -177852. 228.
C13 - - - - - 8 10 10 7 3 11 - 1 - - - - - - - - - - - PIN -177851. 226.
C24 - - - - - 7 6 8 9 12 7 - 1 - - - - - - - - - - - PIN -177830. 224.
C23 - - - - - 13 7 8 7 7 5 3 - - - - - - - - - - - - - PIN -177820. 236.
B1 - 11 16 12 11 - - - - - - - - - - - - - - - - - - PIN -177527. 234.
B4 - 11 9 21 9 - - - - - - - - - - - - - - - - - - PIN -177525. 225.
B2 - 15 15 8 12 - - - - - - - - - - - - - - - - - - PIN -177520. 235.
B3 - 13 10 9 18 - - - - - - - - - - - - - - - - - - PIN -177514. 225.
A 50 - - - - - - - - - - - - - - - - - - - - - - - PIN -177089. 231.
DNSDMELD CD PARAMTRDS NO 14

```

Figure 4.13 Ordre des arborescences pour le vecteur de paramètres $(x, x, x, x)=(1,1,1,1)$

PARAMETRE NUMERO 469																																	
.5	-.5	.0	.0	.0	.0	.5	-.5	.0																									
.5	-.5	.0	.0	.5	-.5	.0	.0	.0																									
.5	-.5	.5	-.5	.0	.0	.0	.0	.0																									
NOMER C OBBURDNBS PAR ENVIRONNEMNT										10000																							
RANG DT PRDQ CU RANG CD BHAQUD ARERD																																	
RANG CDS ARERDS																																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	MDC	SC					
STRUBTURD																																	
B24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	14	13	23	PIN	-208044.	136.				
B23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	18	15	16	PIN	-208033.	139.			
E24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	16	22	11	PIN	-208032.	140.			
D12	-	-	-	-	3	1	2	1	-	4	1	3	1	4	6	4	3	5	1	3	6	2	-	-	-	-	-	-	-	PIN	-207987.	141.	
D31	-	-	-	-	-	1	4	3	3	3	5	2	6	1	1	1	4	5	3	3	5	-	-	-	-	-	-	-	-	PIN	-207985.	138.	
D23	-	-	-	-	-	1	1	3	-	2	-	1	5	5	2	5	6	3	1	4	4	7	-	-	-	-	-	-	-	-	PIN	-207985.	140.
D32	-	-	-	-	-	1	1	2	3	-	2	2	3	-	4	3	5	5	1	6	3	8	1	-	-	-	-	-	-	-	PIN	-207984.	141.
D42	-	-	-	-	-	1	1	4	3	1	6	1	3	4	1	1	2	1	5	4	6	5	1	-	-	-	-	-	-	-	PIN	-207982.	139.
D21	-	-	-	-	-	3	-	3	5	3	1	4	4	2	2	2	3	2	4	4	3	5	-	-	-	-	-	-	-	-	PIN	-207979.	140.
D13	-	-	-	-	-	1	1	1	-	3	2	1	2	2	6	3	3	1	2	10	6	6	-	-	-	-	-	-	-	-	PIN	-207977.	138.
C13	-	-	-	-	1	3	2	5	3	5	1	6	3	3	2	3	-	5	2	3	2	1	-	-	-	-	-	-	-	-	PIN	-207974.	138.
C12	-	-	-	-	2	7	5	3	3	6	3	6	4	1	2	1	3	-	2	1	1	-	-	-	-	-	-	-	-	-	PIN	-207973.	141.
D24	-	-	-	-	-	1	3	1	2	5	1	4	3	3	6	4	3	5	3	1	3	2	-	-	-	-	-	-	-	-	PIN	-207973.	140.
C24	-	-	-	-	-	7	7	3	4	2	6	1	7	2	2	1	2	3	-	3	-	-	-	-	-	-	-	-	-	-	PIN	-207972.	139.
D43	-	-	-	-	-	1	4	2	2	1	2	3	1	3	1	3	4	3	5	7	3	1	4	-	-	-	-	-	-	-	PIN	-207970.	141.
D41	-	-	-	-	-	2	4	4	2	4	4	2	2	2	4	4	1	2	3	1	6	3	-	-	-	-	-	-	-	-	PIN	-207969.	139.
C23	-	-	-	-	1	3	4	5	1	1	6	4	2	4	3	1	4	4	3	2	2	-	-	-	-	-	-	-	-	-	PIN	-207968.	140.
D14	-	-	-	-	1	4	1	1	5	1	4	3	5	5	3	2	5	2	3	1	1	2	-	-	-	-	-	-	-	-	PIN	-207966.	140.
D34	-	-	-	-	1	2	6	3	3	1	1	2	1	2	3	4	3	5	7	1	5	-	-	-	-	-	-	-	-	-	PIN	-207964.	140.
C34	-	-	-	-	1	7	4	6	5	3	4	2	5	4	2	2	2	-	2	-	1	-	-	-	-	-	-	-	-	-	PIN	-207961.	141.
C14	-	-	-	1	1	11	6	4	5	6	1	3	-	2	3	1	2	-	1	2	-	1	-	-	-	-	-	-	-	-	PIN	-207957.	140.
B2	-	9	10	13	15	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	PIN	-207939.	141.
B1	-	14	18	5	12	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	PIN	-207937.	140.
B3	-	13	8	16	10	2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	PIN	-207934.	141.
B4	-	14	14	15	5	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	PIN	-207923.	141.
A	50	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	PIN	-207899.	141.
DMSMELD CD PARAMETRES NO										170																							

Figure 4.14 Ordre des arborescences pour le vecteur de paramètres $(x,x,x,x)=(.5,.5,.5,.5)$

PARAMETRE NUMERO 5									
.0	.0	.0	.0	.0	.0	.0	3.0	-3.0	.0
.0	.0	.0	.0	3.0	-3.0	.0	.0	.0	.0
.0	.0	3.0	-3.0	.0	.0	.0	.0	.0	.0

NOMBRE D OCURENCES PAR ENVIRONNEMENT 10000									

RANG ET PREQ DE RANG DE CHAQUE ARBRE																															
RANG DES ARBRES																															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	MEQ	SD			
STRUCTURD																															
E24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	16	16	18	FIN	-105393.	267.	
E34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	16	15	19	FIN	-105355.	228.	
E23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	18	19	13	FIN	-105327.	239.	
D24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-104042.	303.
D34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-104014.	286.
D14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-103996.	285.
C24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-102476.	287.
C34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-102440.	268.
C14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-102425.	270.
B4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	16	34	-	-	-	-	-	-	-	-	-	-	FIN	-99686.	290.
D42	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	20	9	14	7	-	-	-	-	-	-	-	-	-	FIN	-99563.	286.
D43	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	13	16	17	4	-	-	-	-	-	-	-	-	-	FIN	-99552.	279.
D41	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	17	25	3	5	-	-	-	-	-	-	-	-	-	FIN	-99493.	279.
D12	-	-	-	-	7	2	4	7	2	3	11	7	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-97220.	258.
B2	-	-	-	-	2	8	3	4	4	4	6	13	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-97219.	258.
D32	-	-	-	-	4	3	6	1	6	5	8	5	12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-97219.	258.
B3	-	-	-	-	3	10	-	7	10	6	2	10	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-97162.	244.
D23	-	-	-	-	5	-	8	10	5	8	9	1	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-97162.	244.
D13	-	-	-	-	5	3	5	6	8	9	3	3	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-97162.	244.
B1	-	-	-	-	4	14	6	1	7	7	2	4	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-97151.	245.
D31	-	-	-	-	10	4	10	4	5	6	6	2	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-97151.	246.
D21	-	-	-	-	10	6	8	10	3	2	3	5	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-97151.	245.
C12	22	11	7	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-95105.	258.
C13	10	10	14	16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-95105.	258.
C23	9	16	12	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-95104.	258.
A	9	13	17	11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-95104.	258.

NUMERO DE PARAMETRES NO 6

Figure 4.15 Ordre des arborescences pour le vecteur de paramètres $(0, x, x, x) - (0, 3, 3, 3)$

```

PARAMTRD NUMDRO 160
.0 .0 .0 .0 .0 2.0 -2.0 .0
.0 .0 .0 .0 2.0 -2.0 .0 .0 .0
.0 .0 2.0 -2.0 .0 .0 .0 .0 .0

NOMERD C OBBURDNBDS PAR ENVIRONNEMNT 10000

RANG DT PRDQ CU RANG CD BBAQUD ARERD
RANG CDS ARERDS
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 MDC SC
STRUBTURD
D34 - - - - - - - - - - - - - - - - - 13 21 16 FIN -135519. 256.
D14 - - - - - - - - - - - - - - - - - 18 13 19 FIN -135483. 242.
D24 - - - - - - - - - - - - - - - - - 19 16 15 FIN -135473. 236.
E34 - - - - - - - - - - - - - - - - - 12 18 20 - - - FIN -134500. 264.
E23 - - - - - - - - - - - - - - - - - 21 15 14 - - - FIN -134439. 242.
E24 - - - - - - - - - - - - - - - - - 1 - 16 17 16 - - - FIN -134432. 222.
C34 - - - - - - - - - - - - - - - - - 14 15 21 - - - - - FIN -134096. 268.
C14 - - - - - - - - - - - - - - - - - 22 11 16 1 - - - - - FIN -134059. 247.
C24 - - - - - - - - - - - - - - - - - 14 23 13 - - - - - FIN -134036. 230.
B4 - - - - - - - - - - - - - - - - - 7 29 14 - - - - - FIN -132653. 242.
D43 - - - - - - - - - - - - - - - - - 13 16 8 13 - - - - - FIN -132641. 272.
D41 - - - - - - - - - - - - - - - - - 20 12 5 13 - - - - - FIN -132617. 255.
D42 - - - - - - - - - - - - - - - - - 17 15 8 10 - - - - - FIN -132604. 233.
D13 - - - - 2 3 5 7 9 7 9 1 7 - - - - - - - - - - - - - FIN -130191. 256.
D23 - - - - 5 2 4 9 5 8 5 7 5 - - - - - - - - - - - - - FIN -130191. 256.
B3 - - - - 3 5 2 6 9 8 3 9 5 - - - - - - - - - - - - - FIN -130191. 256.
D12 - - - - 11 3 9 4 4 6 4 5 4 - - - - - - - - - - - - - FIN -130181. 247.
B2 - - - - 4 12 8 6 5 2 2 6 5 - - - - - - - - - - - - - FIN -130181. 247.
D32 - - - - 9 9 6 4 4 5 7 2 4 - - - - - - - - - - - - - FIN -130181. 247.
D21 - - - - 5 2 9 4 5 5 7 5 8 - - - - - - - - - - - - - FIN -130180. 251.
B1 - - - - 4 10 2 6 6 2 8 10 2 - - - - - - - - - - - - - FIN -130180. 251.
D31 - - - - 7 4 5 4 3 7 5 5 10 - - - - - - - - - - - - - FIN -130180. 251.
A 6 24 16 4 - - - - - - - - - - - - - - - - - - - - - - - FIN -129438. 254.
C13 12 9 15 14 - - - - - - - - - - - - - - - - - - - - - - - FIN -129438. 254.
C12 10 10 13 17 - - - - - - - - - - - - - - - - - - - - - - - FIN -129438. 254.
C23 22 7 6 15 - - - - - - - - - - - - - - - - - - - - - - - FIN -129438. 254.
ONSOMELD CD PARAMTRDS NO 61

```

Figure 4.16 Ordre des arborescences pour le vecteur de paramètres $(0, x, x, x)=(0,2,2,2)$

```

PARAMTRD NUMDRO 315
.0 .0 .0 .0 .0 .0 1.0 -1.0 .0
.0 .0 .0 .0 1.0 -1.0 .0 .0 .0
.0 .0 1.0 -1.0 .0 .0 .0 .0 .0

NOMERD C OBBDNBDN PAR DMVIRONNMDNT 10000

RANG DT FRDQ CU RANG CD BHAQUD ARERD
RANG CDS ARERDS
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 MDC SC
STROBTURD
D34 - - - - - - - - - - - - - - - - - - - - - - 15 14 21 PIN -186086. 250.
D14 - - - - - - - - - - - - - - - - - - - - - - 1 - 16 20 13 PIN -186083. 251.
D24 - - - - - - - - - - - - - - - - - - - - - - 18 16 16 PIN -186051. 263.
E24 - - - - - - - - - - - - - - - - - - - - - - 1 15 5 12 6 11 - - - PIN -185916. 263.
C24 - - - - - - - - - - - - - - - - - - - - - - 15 14 6 9 4 1 - - - PIN -185914. 263.
X23 - - - - - - - - - - - - - - - - - - - - - - 6 14 1 11 4 14 - - - PIN -185908. 250.
C14 - - - - - - - - - - - - - - - - - - - - - - 14 7 13 2 12 2 - - - PIN -185908. 251.
E34 - - - - - - - - - - - - - - - - - - - - - - 7 7 1 15 2 18 - - - PIN -185907. 249.
C34 - - - - - - - - - - - - - - - - - - - - - - 1 7 6 16 3 16 1 - - - PIN -185905. 249.
D42 - - - - - - - - - - - - - - - - - - - - - - 15 23 11 - - - 1 - - - PIN -185753. 264.
D41 - - - - - - - - - - - - - - - - - - - - - - 21 11 18 - - - - - - - PIN -185738. 254.
D43 - - - - - - - - - - - - - - - - - - - - - - 14 16 20 - - - - - - - PIN -185735. 252.
B4 - - - - - - - - - - - - - - - - - - - - - - 50 - - - - - - - - - PIN -185582. 260.
D12 - - - - - 8 5 6 6 5 4 6 5 5 - - - - - - - - - - - - - - - PIN -185256. 260.
B2 - - - - - 3 11 5 3 7 4 5 6 6 - - - - - - - - - - - - - - - PIN -185256. 260.
D32 - - - - - 8 3 8 6 2 7 6 5 5 - - - - - - - - - - - - - - - PIN -185255. 260.
B1 - - - - - 3 7 6 4 14 5 1 5 5 - - - - - - - - - - - - - - - PIN -185248. 255.
D21 - - - - - 8 3 5 13 3 8 6 3 1 - - - - - - - - - - - - - - - PIN -185247. 255.
D31 - - - - - 5 6 5 7 7 10 3 3 4 - - - - - - - - - - - - - - - PIN -185247. 255.
D13 - - - - - 6 3 6 5 2 5 8 9 6 - - - - - - - - - - - - - - - PIN -185244. 255.
D23 - - - - - 7 5 3 6 - 5 7 7 10 - - - - - - - - - - - - - - - PIN -185243. 255.
B3 - - - - - 2 7 6 - 10 2 8 7 8 - - - - - - - - - - - - - - - PIN -185243. 255.
C13 15 15 6 14 - - - - - - - - - - - - - - - - - - - - - - - PIN -185120. 259.
C12 17 17 5 11 - - - - - - - - - - - - - - - - - - - - - - - PIN -185120. 259.
C23 13 7 10 20 - - - - - - - - - - - - - - - - - - - - - - - PIN -185120. 259.
A 5 11 29 5 - - - - - - - - - - - - - - - - - - - - - - - PIN -185120. 259.
DN5DMELD CD PARAMTRDS NO 16

```

Figure 4.17 Ordre des arborescences pour le vecteur de paramètres $(0, x, x, x) = (0.1, 1, 1)$

```

PARAMDTRD NUMDRO 470
.0 .0 .0 .0 .0 .0 .5 -.5 .0
.0 .0 .0 .0 .5 -.5 .0 .0 .0
.0 .0 .5 -.5 .0 .0 .0 .0 .0

NOMSRD C OBBURDNBDS PAR DNVIRONNMDMT 10000

RANG DT PRDQ CU RANG CD BRAQUO ARERD
RANG COS ARERDS
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 MDC SC
STROBTURD
D14 - - - - - - - - - - - - - - - - - - - - - - 2 3 1 3 7 8 8 18 FIN -211221. 163.
E23 - - - - - - - - - - - - - - - - - - - - - - 2 4 3 2 7 6 3 5 13 1 4 FIN -211217. 163.
C14 - - - - - - - - - - - - - - - - - - - - - - 1 4 5 3 5 6 1 8 5 12 - FIN -211216. 163.
D34 - - - - - - - - - - - - - - - - - - - - - - 1 1 3 2 1 9 9 5 5 14 FIN -211214. 164.
D24 - - - - - - - - - - - - - - - - - - - - - - 6 2 1 9 9 1 11 11 FIN -211211. 160.
C34 - - - - - - - - - - - - - - - - - - - - - - 1 8 4 7 5 7 5 3 4 6 - FIN -211205. 167.
E34 - - - - - - - - - - - - - - - - - - - - - - 5 3 9 4 6 7 3 3 7 1 2 FIN -211205. 167.
D41 - - - - - - - - - - - - - - - - - - - - - - 11 7 8 6 8 3 1 5 1 - - - FIN -211202. 163.
E24 - - - - - - - - - - - - - - - - - - - - - - 3 10 4 4 8 8 2 3 5 3 - FIN -211199. 160.
C24 - - - - - - - - - - - - - - - - - - - - - - 8 5 5 4 7 8 5 2 2 3 1 FIN -211199. 160.
D43 - - - - - - - - - - - - - - - - - - - - - - 2 17 10 4 6 5 2 2 2 - - - FIN -211192. 167.
D42 - - - - - - - - - - - - - - - - - - - - - - 2 19 13 3 7 2 - 2 3 - - - FIN -211187. 160.
B4 - - - - - - - - - - - - - - - - - - - - - - 46 3 - - - - - - - - - FIN -211171. 164.
D31 - - - - - 3 6 5 6 2 9 5 4 10 - - - - - - - - - - - - - - - - - - - - - - FIN -211146. 165.
B1 - - - - - 8 5 1 - 13 3 5 9 5 1 - - - - - - - - - - - - - - - - - - - - - - FIN -211146. 165.
D21 - - - - - 4 3 8 11 2 5 8 5 4 - - - - - - - - - - - - - - - - - - - - - - FIN -211146. 165.
D23 - - - - - 7 2 12 4 4 7 6 3 5 - - - - - - - - - - - - - - - - - - - - - - FIN -211144. 165.
D13 - - - - - 5 11 5 5 6 5 5 4 4 - - - - - - - - - - - - - - - - - - - - - - FIN -211144. 165.
B3 - - - - - 8 8 4 7 6 4 2 6 5 - - - - - - - - - - - - - - - - - - - - - - FIN -211144. 165.
D32 - - - 1 - 4 6 3 5 5 5 9 6 6 - - - - - - - - - - - - - - - - - - - - - - FIN -211142. 164.
B2 - - - 1 3 6 6 5 7 4 5 6 7 - - - - - - - - - - - - - - - - - - - - - - FIN -211142. 164.
D12 1 - - - 7 2 5 6 5 8 5 7 4 - - - - - - - - - - - - - - - - - - - - - - FIN -211142. 164.
C13 9 12 11 17 - 1 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - FIN -211133. 165.
C23 17 3 17 12 1 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - FIN -211133. 165.
A 5 23 16 5 - - 1 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - FIN -211133. 165.
C12 18 12 5 15 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - FIN -211133. 165.
DNSDMELD CD PARAMDTRDS MO 171

```

Figure 4.18 Ordre des arborescences pour le vecteur de paramètres $(0, x, x, x) = (0, .5, .5, .5)$

PARAMETRE NUMERO 1																																	
3.0	-3.0	.0	.0	.0	.0	3.0	-3.0	.0																									
3.0	-3.0	.0	.0	3.0	-3.0	.0	.0	.0																									
3.0	-3.0	3.0	-3.0	.0	.0	.0	.0	.0																									
NOMBRE D OCURENCES PAR ENVIRONNEMENT 10000																																	
RANG ET FREQ DE RANG DE CHAQUE ARBRE																																	
RANG DES ARBRES																																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	MED	SD					
STRUBTURD																																	
E24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	10	17	23	FIN	-126849.	138.	
E23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	15	18	17	FIN	-126813.	117.
E34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	25	15	10	FIN	-126768.	144.
D13	-	-	-	-	-	-	-	-	-	-	-	5	1	5	3	4	3	3	3	9	7	4	3	-	-	-	-	-	-	-	FIN	-125381.	109.
D31	-	-	-	-	-	-	-	-	-	-	-	2	4	2	5	3	5	3	7	5	5	4	-	-	-	-	-	-	-	-	FIN	-125369.	102.
D41	-	-	-	-	-	-	-	-	-	-	-	3	7	3	3	3	5	6	2	3	2	4	9	-	-	-	-	-	-	-	FIN	-125368.	114.
D14	-	-	-	-	-	-	-	-	-	-	-	6	2	1	5	5	4	4	4	3	5	9	2	-	-	-	-	-	-	-	FIN	-125366.	122.
D24	-	-	-	-	-	-	-	-	-	-	-	2	2	5	3	8	4	9	6	1	3	3	4	-	-	-	-	-	-	-	FIN	-125354.	122.
D23	-	-	-	-	-	-	-	-	-	-	-	3	7	6	2	4	3	3	8	4	6	1	3	-	-	-	-	-	-	-	FIN	-125348.	119.
D43	-	-	-	-	-	-	-	-	-	-	-	7	5	6	4	2	4	6	-	6	4	1	5	-	-	-	-	-	-	-	FIN	-125342.	130.
D32	-	-	-	-	-	-	-	-	-	-	-	3	6	2	4	5	7	-	5	5	7	2	4	-	-	-	-	-	-	-	FIN	-125342.	139.
D12	-	-	-	-	-	-	-	-	-	-	-	6	6	5	3	1	4	3	5	5	3	4	5	-	-	-	-	-	-	-	FIN	-125340.	136.
D42	-	-	-	-	-	-	-	-	-	-	-	4	5	3	7	7	2	4	2	1	2	8	5	-	-	-	-	-	-	-	FIN	-125337.	140.
D21	-	-	-	-	-	-	-	-	-	-	-	5	3	4	5	3	5	5	4	5	3	4	4	-	-	-	-	-	-	-	FIN	-125332.	133.
D34	-	-	-	-	-	-	-	-	-	-	-	4	2	8	6	5	4	4	4	3	3	5	2	-	-	-	-	-	-	-	FIN	-125328.	125.
C13	-	-	-	-	-	6	5	8	11	10	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-124032.	107.
C24	-	-	-	-	-	5	7	14	11	5	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-123995.	134.
C14	-	-	-	-	-	6	12	5	5	11	11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-123994.	114.
C12	-	-	-	-	-	11	5	10	11	5	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-123980.	136.
C23	-	-	-	-	-	10	8	6	5	13	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-123964.	125.
C34	-	-	-	-	-	12	13	7	7	6	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-123953.	123.
B1	-	11	14	13	12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-121238.	115.
B3	-	12	16	10	12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-121220.	119.
B2	-	18	12	5	15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-121216.	140.
B4	-	9	8	22	11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-121210.	124.
A	50	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-116604.	124.
NUMERO DE PARAMETRES NO 2																																	

Figure 4.19 Ordre des arborescences pour le vecteur de paramètres $(t, x, x, x) - (3, 3, 3, 3)$

```

PARAMETRE NUMERO 2
2.0 -2.0 .0 .0 .0 .0 3.0 -3.0 .0
2.0 -2.0 .0 .0 3.0 -3.0 0 .0 .0
2.0 -2.0 3.0 -3.0 .0 .0 .0 .0 .0

NOMBRE D OCURENCES PAR ENVIRONNEMENT 10000

RANG ET FREQ DE RANG DE CHAQUE ARBRE
RANG DES ARBRES
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 MRD SD
STRUCTURE
D34 - - - - - - - - - - - - - - - - - - - - 11 23 16 FIN -121839. 177.
D14 - - - - - - - - - - - - - - - - - - - - 20 12 18 FIN -121831. 186.
D24 - - - - - - - - - - - - - - - - - - - - 19 15 16 FIN -121793. 179.
E34 - - - - - - - - - - - - - - - - - - - - 12 16 22 - - - FIN -120895. 179.
E23 - - - - - - - - - - - - - - - - - - - - 17 19 14 - - - FIN -120819. 166.
E24 - - - - - - - - - - - - - - - - - - - - 21 15 14 - - - FIN -120785. 186.
C34 - - - - - - - - - - - - - - - - - - - - 12 19 19 - - - FIN -119589. 173.
C24 - - - - - - - - - - - - - - - - - - - - 20 14 16 - - - FIN -119557. 180.
C14 - - - - - - - - - - - - - - - - - - - - 18 17 15 - - - FIN -119554. 180.
D43 - - - - - - - - - - - - - - - - - - - - 12 25 13 - - - FIN -118995. 185.
D42 - - - - - - - - - - - - - - - - - - - - 19 11 20 - - - FIN -118973. 199.
D41 - - - - - - - - - - - - - - - - - - - - 19 14 17 - - - FIN -118941. 193.
B4 - - - - - - - - - - - - - - - - - - - - 50 - - - - - FIN -117389. 179.
D23 - - - - - - - - - 9 5 8 11 5 12 - - - - - - - - - - - FIN -116779. 196.
D32 - - - - - - - - - 11 7 4 11 8 9 - - - - - - - - - - - FIN -116761. 172.
D12 - - - - - - - - - 7 7 10 6 11 9 - - - - - - - - - - - FIN -116759. 186.
D21 - - - - - - - - - 6 6 11 6 12 9 - - - - - - - - - - - FIN -116735. 198.
D13 - - - - - - - - - 8 11 11 8 7 5 - - - - - - - - - - - FIN -116733. 192.
D31 - - - - - - - - - 9 14 6 8 7 6 - - - - - - - - - - - FIN -116708. 212.
C23 - - - - - 21 8 21 - - - - - - - - - - - - - - - - - FIN -115234. 181.
C12 - - - - - 9 25 16 - - - - - - - - - - - - - - - - - FIN -115204. 194.
C13 - - - - - 20 17 13 - - - - - - - - - - - - - - - - - FIN -115190. 196.
B2 - 17 12 21 - - - - - - - - - - - - - - - - - - - - - FIN -114405. 171.
B3 - 14 24 12 - - - - - - - - - - - - - - - - - - - - - FIN -114402. 183.
B1 - 19 14 17 - - - - - - - - - - - - - - - - - - - - - FIN -114367. 196.
A 50 - - - - - - - - - - - - - - - - - - - - - - - - - - - FIN -111064. 175.
NUMERO DE PARAMETRES NO 3

```

Figure 4.20 Ordre des arborescences pour le vecteur de paramètres $(t, x, x, x) = (2.3.3.3)$

```

PARAMETRE NUMERO 3
1.0 -1.0 .0 .0 .0 .0 3.0 -3.0 .0
1.0 -1.0 .0 .0 3.0 -3.0 .0 .0 .0
1.0 -1.0 3.0 -3.0 .0 .0 .0 .0 .0

NOMBRE D OCURENCES PAR ENVIRONNEMENT 10000

RANG ET FREQ DE RANG DE CHAQUE ARBRE
RANG DES ARBRES
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 MED SD
STRUCTURE
D34 - - - - - - - - - - - - - - - - - 3 7 9 11 10 10 FIN -110279. 267.
D14 - - - - - - - - - - - - - - - - - 3 3 5 13 11 15 FIN -110270. 258.
D24 - - - - - - - - - - - - - - - - - 1 8 5 13 11 12 FIN -110222. 294.
E23 - - - - - - - - - - - - - - - - - 14 12 8 5 6 5 FIN -110171. 235.
E34 - - - - - - - - - - - - - - - - - 16 12 9 1 7 5 FIN -110149. 239.
E24 - - - - - - - - - - - - - - - - - 13 8 14 7 5 3 FIN -110121. 242.
C14 - - - - - - - - - - - - - - - - - 17 13 20 - - - - - FIN -108259. 249.
C34 - - - - - - - - - - - - - - - - - 18 20 12 - - - - - FIN -108245. 254.
C24 - - - - - - - - - - - - - - - - - 15 17 18 - - - - - FIN -108193. 276.
B4 - - - - - - - - - - - - - - - - - 30 20 - - - - - FIN -105950. 271.
D43 - - - - - - - - - - - - - - - - - 16 22 7 5 - - - - - FIN -105776. 262.
D41 - - - - - - - - - - - - - - - - - 12 18 8 12 - - - - - FIN -105772. 268.
D42 - - - - - - - - - - - - - - - - - 22 10 5 13 - - - - - FIN -105733. 275.
D21 - - - - - - - 8 7 5 8 11 11 - - - - - - - - - - - FIN -103297. 248.
D12 - - - - - - - 9 7 5 10 11 8 - - - - - - - - - - - FIN -103297. 263.
D32 - - - - - - - 5 11 9 6 10 9 - - - - - - - - - - - FIN -103290. 253.
D31 - - - - - - - 6 10 8 8 8 10 - - - - - - - - - - - FIN -103290. 236.
D13 - - - - - - - 11 8 13 7 6 5 - - - - - - - - - - - FIN -103251. 235.
D23 - - - - - - - 11 7 10 11 4 7 - - - - - - - - - - - FIN -103239. 246.
B1 - - - - - 12 18 20 - - - - - - - - - - - - - - - - - FIN -102601. 245.
B2 - - - - - 14 17 19 - - - - - - - - - - - - - - - - - FIN -102585. 264.
B3 - - - - - 24 15 11 - - - - - - - - - - - - - - - - - FIN -102580. 238.
C12 - 15 18 17 - - - - - - - - - - - - - - - - - - - - - FIN -101409. 259.
C13 - 17 18 15 - - - - - - - - - - - - - - - - - - - - - FIN -101403. 240.
C23 - 18 14 18 - - - - - - - - - - - - - - - - - - - - - FIN -101374. 253.
A 50 - - - - - - - - - - - - - - - - - - - - - - - - - - - FIN -100179. 254.
NUMERO DE PARAMETRES NO 4

```

Figure 4.21 Ordre des arborescences pour le vecteur de paramètres $(t, x, x, x)=(1,3,3,3)$

```

PARAMETRE NUMERO 4
.5 -.5 .0 .0 .0 3.0 -3.0 .0
.5 -.5 .0 .0 3.0 -3.0 .0 .0
.5 -.5 3.0 -3.0 .0 .0 .0 .0

NOMBRE D OCURENCES PAR ENVIRONNEMENT 10000

RANG ET FREQ DE RANG DE CHAQUE ARBRE
RANG DES ARBRES
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 MED SD
STRUCTURD
E24 - - - - - - - - - - - - - - - - - - - - 15 20 15 PIN -106670. 304.
E34 - - - - - - - - - - - - - - - - - - - - 11 19 20 PIN -106658. 268.
E23 - - - - - - - - - - - - - - - - - - - - 24 11 15 PIN -106582. 271.
D24 - - - - - - - - - - - - - - - - - - - - 18 13 19 - - - PIN -105734. 337.
D34 - - - - - - - - - - - - - - - - - - - - 9 22 19 - - - PIN -105729. 332.
D14 - - - - - - - - - - - - - - - - - - - - 23 15 12 - - - PIN -105700. 320.
C24 - - - - - - - - - - - - - - - - - - - - 13 15 22 - - - PIN -104074. 320.
C34 - - - - - - - - - - - - - - - - - - - - 11 24 15 - - - PIN -104057. 319.
C14 - - - - - - - - - - - - - - - - - - - - 26 11 13 - - - PIN -103987. 302.
B4 - - - - - - - - - - - - - - - - - - - - 1 6 43 - - - PIN -101344. 335.
D42 - - - - - - - - - - - - - - - - - - - - 10 20 19 1 - - - PIN -101265. 325.
D43 - - - - - - - - - - - - - - - - - - - - 14 19 15 2 - - - PIN -101222. 338.
D41 - - - - - - - - - - - - - - - - - - - - 26 10 10 4 - - - PIN -101173. 313.
D32 - - - - - - - - 1 7 6 10 8 9 9 - - - - - - - - - - - PIN -98848. 315.
D12 - - - - - - - - 7 9 7 9 9 9 - - - - - - - - - - - PIN -98837. 313.
D23 - - - - - - - - 5 10 10 6 8 11 - - - - - - - - - - - PIN -98812. 303.
D13 - - - - - - - - 9 5 7 9 12 8 - - - - - - - - - - - PIN -98802. 301.
D31 - - - - - - - - 4 8 8 11 6 3 10 - - - - - - - - - - - PIN -98796. 289.
D21 - - - - - - - - 1 10 10 5 12 9 3 - - - - - - - - - - - PIN -98787. 284.
B2 - - - - - 16 15 16 3 - - - - - - - - - - - - - - - - PIN -98654. 312.
B3 - - - - - 15 17 16 - 2 - - - - - - - - - - - - - - - PIN -98623. 301.
B1 - - - - - 19 18 12 1 - - - - - - - - - - - - - - - - PIN -98606. 287.
C12 - 18 13 19 - - - - - - - - - - - - - - - - - - - - - PIN -96820. 309.
C23 - 13 21 16 - - - - - - - - - - - - - - - - - - - - - PIN -96817. 314.
C13 - 19 16 15 - - - - - - - - - - - - - - - - - - - - - PIN -96815. 312.
A 50 - - - - - - - - - - - - - - - - - - - - - - - PIN -96490. 310.
NOMBRE DE PARAMETRES NO 5

```

Figure 4.22 Ordre des arborescences pour le vecteur de paramètres $(t, x, x, x) = (.5, 3, 3, 3)$

PARAMETRE NUMERO 5																																		
.0	.0	.0	.0	.0	.0	3.0	-3.0	.0																										
.0	.0	.0	.0	3.0	-3.0	.0	.0	.0																										
.0	.0	3.0	-3.0	.0	.0	.0	.0	.0																										
NOMBRE D OCURENCES PAR ENVIRONNEMENT 10000																																		
RANG ET FREQ DE RANG DE CHAQUE ARBRE																																		
RANG DES ARBRES																																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	MED	SD						
STRUBTURD																																		
E24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	16	16	18	PIN	-105393.	267.			
E34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	16	15	19	PIN	-105355.	228.			
E23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	18	19	13	PIN	-105327.	239.			
D24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	12	15	23	-	-	PIN	-104042.	303.
D34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	18	16	16	-	-	PIN	-104014.	286.
D14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	20	19	11	-	-	PIN	-103996.	285.
C24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	15	15	20	-	-	-	-	-	-	-	-	-	PIN	-102476.	287.
C34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	15	16	19	-	-	-	-	-	-	-	-	-	PIN	-102440.	268.
C14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	20	19	11	-	-	-	-	-	-	-	-	-	PIN	-102425.	270.
B4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	16	34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	PIN	-99686.	290.
D42	-	-	-	-	-	-	-	-	-	-	-	-	-	-	20	9	14	7	-	-	-	-	-	-	-	-	-	-	-	-	-	PIN	-99563.	286.
D43	-	-	-	-	-	-	-	-	-	-	-	-	-	-	13	16	17	4	-	-	-	-	-	-	-	-	-	-	-	-	-	PIN	-99552.	279.
D41	-	-	-	-	-	-	-	-	-	-	-	-	-	-	17	25	3	5	-	-	-	-	-	-	-	-	-	-	-	-	-	PIN	-99493.	279.
D12	-	-	-	-	7	2	4	7	2	3	11	7	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	PIN	-97220.	258.
B2	-	-	-	-	2	8	3	4	4	4	6	13	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	PIN	-97219.	258.
D32	-	-	-	-	4	3	6	1	6	5	8	5	12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	PIN	-97219.	258.
B3	-	-	-	-	3	10	-	7	10	6	2	10	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	PIN	-97162.	244.
D23	-	-	-	-	5	-	8	10	5	8	9	1	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	PIN	-97162.	244.
D13	-	-	-	-	5	3	5	6	8	9	3	3	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	PIN	-97162.	244.
B1	-	-	-	-	4	14	6	1	7	7	2	4	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	PIN	-97151.	245.
D31	-	-	-	-	10	4	10	4	5	6	6	2	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	PIN	-97151.	246.
D21	-	-	-	-	10	6	8	10	3	2	3	5	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	PIN	-97151.	245.
C12	22	11	7	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	PIN	-95105.	258.
C13	10	10	14	16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	PIN	-95105.	258.
C23	9	16	12	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	PIN	-95104.	258.
A	9	13	17	11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	PIN	-95104.	258.
NUMERO DE PARAMETRES NO 6																																		

Figure 4.23 Ordre des arborescences pour le vecteur de paramètres $(t, x, x, x)=(0.3.3.3)$

PARAMTRD NUMDRO 158																																		
1.0	-1.0	.0	.0	.0	.0	2.0	-2.0	.0																										
1.0	-1.0	.0	.0	2.0	-2.0	.0	.0	.0																										
1.0	-1.0	2.0	-2.0	.0	.0	.0	.0	.0																										
NOMERD C OBBURDNBS PAR ENVIRONNEMNT										10000																								
RANG DT FRDQ CU RANG CD BRAQUD ARERD																																		
RANG CDS ARERDS																																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	MDC	SC						
STRUBTURD																																		
D34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	14	19	17	FIN	-137546.	274.				
D24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	11	16	23	FIN	-137535.	270.			
D14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	25	15	10	FIN	-137507.	247.			
E34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	17	15	18	-	-	FIN	-137059.	277.	
E24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	10	16	24	-	-	FIN	-137025.	260.	
E23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	23	19	8	-	-	FIN	-137004.	236.	
C34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	17	21	12	-	-	FIN	-136137.	282.	
C14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	24	16	10	-	-	FIN	-136121.	243.	
C24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	9	13	28	-	-	FIN	-136117.	276.	
D41	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	20	16	14	-	-	FIN	-135529.	243.	
D43	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	20	17	13	-	-	FIN	-135516.	285.	
D42	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	10	17	23	-	-	FIN	-135485.	282.	
B4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	50	-	-	-	-	FIN	-134773.	271.	
D13	-	-	-	-	-	-	-	-	9	8	12	8	6	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-133782.	265.
D31	-	-	-	-	-	-	-	-	13	8	9	8	7	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-133769.	252.
D23	-	-	-	-	-	-	-	-	9	5	6	11	6	13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-133766.	270.
D32	-	-	-	-	-	-	-	-	7	8	10	8	9	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-133755.	264.
D12	-	-	-	-	-	-	-	-	4	9	5	7	13	12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-133742.	258.
D21	-	-	-	-	-	-	-	-	8	12	8	8	9	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-133740.	256.
B3	-	-	4	8	11	13	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-132809.	275.	
B1	-	3	7	5	11	13	11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-132874.	260.	
B2	-	1	2	3	9	13	22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-132854.	271.	
C13	-	19	11	10	7	2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-132817.	260.	
C23	-	12	19	10	4	4	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-132816.	266.	
C12	-	15	7	14	8	5	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-132813.	262.	
A	50	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-131331.	277.	
DNSDMELD CD PARAMTRDS NO										59																								

Figure 4.24 Ordre des arborescences pour le vecteur de paramètres $(t, x, x, x) = (1, 2, 2, 2)$

```

PARAMETRD NUMDRO 159
.5 -.5 .0 .0 .0 2.0 -2.0 .0
.5 -.5 .0 .0 2.0 -2.0 .0 .0
.5 -.5 2.0 -2.0 .0 .0 .0 .0

NOMBRD C OBBDNBD8 PAR ENVIRONNEMENT 10000

RANG DT PRDQ CU RANG CD BHAQUD ARERD
RANG CDS ARERDS
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 MDC SC
STRUBTURD
D24 - - - - - - - - - - - - - - - - - - - - - - 12 15 23 PIN -136270. 275.
D34 - - - - - - - - - - - - - - - - - - - - - - 17 17 16 PIN -136265. 261.
D14 - - - - - - - - - - - - - - - - - - - - - - 21 18 11 PIN -136255. 279.
E24 - - - - - - - - - - - - - - - - - - - - - - 13 12 25 - - - PIN -135350. 289.
E34 - - - - - - - - - - - - - - - - - - - - - - 18 18 14 - - - PIN -135334. 265.
E23 - - - - - - - - - - - - - - - - - - - - - - 19 20 11 - - - PIN -135310. 271.
C34 - - - - - - - - - - - - - - - - - - - - - - 18 19 13 - - - PIN -134814. 262.
C24 - - - - - - - - - - - - - - - - - - - - - - 14 12 24 - - - PIN -134803. 283.
C14 - - - - - - - - - - - - - - - - - - - - - - 18 19 13 - - - PIN -134785. 279.
D43 - - - - - - - - - - - - - - - - - - - - - - 3 15 16 16 - - - PIN -133612. 277.
D42 - - - - - - - - - - - - - - - - - - - - - - 15 14 21 - - - PIN -133578. 292.
D41 - - - - - - - - - - - - - - - - - - - - - - 1 16 20 13 - - - PIN -133559. 291.
B4 - - - - - - - - - - - - - - - - - - - - - - 46 4 - - - PIN -133454. 286.
D23 - - - - - - - 9 8 9 8 8 8 - - - - - - - - - - - - - - - - - - - - - - PIN -131449. 277.
D32 - - - - - - - 9 7 7 8 8 11 - - - - - - - - - - - - - - - - - - - - - - PIN -131444. 276.
D13 - - - - - - - 6 9 12 11 4 8 - - - - - - - - - - - - - - - - - - - - - - PIN -131435. 276.
D21 - - - - - - - 11 9 5 10 12 3 - - - - - - - - - - - - - - - - - - - - - - PIN -131426. 289.
D31 - - - - - - - 10 9 9 3 7 12 - - - - - - - - - - - - - - - - - - - - - - PIN -131422. 292.
D12 - - - - - - - 5 8 8 10 11 8 - - - - - - - - - - - - - - - - - - - - - - PIN -131422. 283.
B3 - - - - - 18 17 15 - - - - - - - - - - - - - - - - - - - - - - - - - - - - PIN -131188. 276.
B1 - - - - - 19 16 15 - - - - - - - - - - - - - - - - - - - - - - - - - - - - PIN -131174. 289.
B2 - - - - - 13 17 20 - - - - - - - - - - - - - - - - - - - - - - - - - - - - PIN -131172. 277.
C13 - - - - - 16 10 24 - - - - - - - - - - - - - - - - - - - - - - - - - - - - PIN -130487. 289.
C12 - - - - - 16 22 12 - - - - - - - - - - - - - - - - - - - - - - - - - - - - PIN -130485. 290.
C23 - - - - - 18 18 14 - - - - - - - - - - - - - - - - - - - - - - - - - - - - PIN -130484. 284.
A 50 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - PIN -130056. 284.
DNSEMELD CD PARAMETRDE NO 60
    
```

Figure 4.25 Ordre des arborescences pour le vecteur de paramètres $(t, x, x, x) = (.5, 2, 2, 2)$

```

PARAMTRD NUMDRO 160
.0 .0 .0 .0 .0 .0 2.0 -2.0 .0
.0 .0 .0 .0 2.0 -2.0 .0 .0 .0
.0 0 2.0 -2.0 .0 .0 .0 .0 .0

NOMERD C OBBURDNBDS PAR DNVIRONNDMNT 10000

RANG DT PRDQ CU RANG CD DRAQUD ARERD
RANG CDS ARERDS
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 MDC SC
STRUBTURD
D34 - - - - - - - - - - - - - - - - - - - - - - - 13 21 16 FIN -135519. 256.
D14 - - - - - - - - - - - - - - - - - - - - - - - 18 13 19 FIN -135483. 242.
D24 - - - - - - - - - - - - - - - - - - - - - - - 19 16 15 FIN -135473. 236.
E34 - - - - - - - - - - - - - - - - - - - - - - - 12 18 20 - - - FIN -134500. 264.
E23 - - - - - - - - - - - - - - - - - - - - - - - 21 15 14 - - - FIN -134439. 242.
E24 - - - - - - - - - - - - - - - - - - - - - - - 1 16 17 16 - - - FIN -134432. 222.
C34 - - - - - - - - - - - - - - - - - - - - - - - 14 15 21 - - - FIN -134096. 268.
C14 - - - - - - - - - - - - - - - - - - - - - - - 22 11 16 1 - - - FIN -134059. 247.
C24 - - - - - - - - - - - - - - - - - - - - - - - 14 23 13 - - - FIN -134036. 230.
B4 - - - - - - - - - - - - - - - - - - - - - - - 7 29 14 - - - FIN -132653. 242.
D43 - - - - - - - - - - - - - - - - - - - - - - - 13 16 8 13 - - - FIN -132641. 272.
D41 - - - - - - - - - - - - - - - - - - - - - - - 20 12 5 13 - - - FIN -132617. 255.
D42 - - - - - - - - - - - - - - - - - - - - - - - 17 15 8 10 - - - FIN -132604. 233.
D13 - - - - - 2 3 5 7 9 7 9 1 7 - - - - - - - - - - - FIN -130391. 256.
D23 - - - - - 5 2 4 9 5 8 5 7 5 - - - - - - - - - - - FIN -130391. 256.
B3 - - - - - 3 5 2 6 9 8 3 9 5 - - - - - - - - - - - FIN -130391. 256.
D12 - - - - - 11 3 9 4 4 6 4 5 4 - - - - - - - - - - - FIN -130381. 247.
B2 - - - - - 4 12 8 6 5 2 2 6 5 - - - - - - - - - - - FIN -130381. 247.
D32 - - - - - 9 9 6 4 4 5 7 2 4 - - - - - - - - - - - FIN -130381. 247.
D21 - - - - - 5 2 9 4 5 5 7 5 8 - - - - - - - - - - - FIN -130380. 251.
B1 - - - - - 4 10 2 6 6 2 8 10 2 - - - - - - - - - - - FIN -130380. 251.
D31 - - - - - 7 4 5 4 3 7 5 5 10 - - - - - - - - - - - FIN -130380. 251.
A 6 24 16 4 - - - - - - - - - - - - - - - - - - - - - FIN -129438. 254.
C13 12 9 15 14 - - - - - - - - - - - - - - - - - - - - - FIN -129438. 254.
C12 10 10 13 17 - - - - - - - - - - - - - - - - - - - - - FIN -129438. 254.
C23 22 7 6 15 - - - - - - - - - - - - - - - - - - - - - FIN -129438. 254.
DNSSMELD CD PARAMTRDS NO 61

```

Figure 4.26 Ordre des arborescences pour le vecteur de paramètres $(t, x, x, x) = (0, 2, 2, 2)$

```

PARAMETRE NOMDRO 314
.5 -.5 .0 .0 .0 .0 1.0 -1.0 .0
.5 -.5 .0 .0 1.0 -1.0 .0 .0 .0
.5 -.5 1.0 -1.0 .0 .0 .0 .0 .0

NOMERD C OBBURDNBS PAR ENVIRONNEMENT 10000

RANG DT PRDQ CU RANG CD BHAQUD ARERD
RANG CDS ARERDS
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 MDC SC
STRUCTURE
E34 - - - - - - - - - - - - - - - - - - - - - - 5 4 4 9 13 15 FIN -184206. 194.
E23 - - - - - - - - - - - - - - - - - - - - - - 2 10 4 10 7 17 FIN -184171. 208.
D34 - - - - - - - - - - - - - - - - - - - - - - 16 10 7 7 9 1 FIN -184167. 200.
E24 - - - - - - - - - - - - - - - - - - - - - - 8 10 7 11 14 FIN -184150. 208.
D14 - - - - - - - - - - - - - - - - - - - - - - 15 8 11 8 6 2 FIN -184132. 216.
D24 - - - - - - - - - - - - - - - - - - - - - - 12 10 14 9 4 1 FIN -184131. 211.
C34 - - - - - - - - - - - - - - - - - - - - - - 1 6 13 12 18 - - - - - FIN -183991. 198.
C14 - - - - - - - - - - - - - - - - - - - - - - 1 13 20 16 - - - - - FIN -183966. 212.
C24 - - - - - - - - - - - - - - - - - - - - - - 8 12 14 16 - - - - - FIN -183959. 212.
D43 - - - - - - - - - - - - - - - - - - - - - - 19 16 9 4 2 - - - - - FIN -183910. 200.
D41 - - - - - - - - - - - - - - - - - - - - - - 13 21 10 4 2 - - - - - FIN -183900. 208.
D42 - - - - - - - - - - - - - - - - - - - - - - 18 12 16 4 - - - - - FIN -183896. 213.
B4 - - - - - - - - - - - - - - - - - - - - - - 1 49 - - - - - - - - - FIN -183648. 214.
D21 - - - - - - - - - - - - - - - - - - - - - - 10 4 7 15 4 10 - - - - - FIN -183558. 208.
D23 - - - - - - - - - - - - - - - - - - - - - - 9 12 8 5 7 9 - - - - - FIN -183554. 201.
D13 - - - - - - - - - - - - - - - - - - - - - - 8 8 8 9 11 6 - - - - - FIN -183550. 200.
D12 - - - - - - - - - - - - - - - - - - - - - - 8 6 10 10 9 7 - - - - - FIN -183547. 204.
D32 - - - - - - - - - - - - - - - - - - - - - - 8 13 4 5 10 9 1 - - - - - FIN -183547. 206.
D31 - - - - - - - - - - - - - - - - - - - - - - 7 7 13 6 9 8 - - - - - FIN -183526. 207.
C12 - - 5 6 9 10 11 9 - - - - - - - - - - - - - - - - - - - - - - - - - - - FIN -183425. 207.
B2 - - 13 5 11 6 10 5 - - - - - - - - - - - - - - - - - - - - - - - - - - - FIN -183414. 208.
B1 - - 9 9 7 9 6 10 - - - - - - - - - - - - - - - - - - - - - - - - - - - FIN -183414. 208.
C23 - - 5 15 5 8 6 11 - - - - - - - - - - - - - - - - - - - - - - - - - - - FIN -183413. 206.
B3 - - 10 11 10 7 8 4 - - - - - - - - - - - - - - - - - - - - - - - - - - - FIN -183412. 203.
C13 - - 8 4 8 10 9 11 - - - - - - - - - - - - - - - - - - - - - - - - - - - FIN -183406. 206.
A 50 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - FIN -183202. 210.
DNSMELD CD PARAMETRES NO 15

```

Figure 4.27 Ordre des arborescences pour le vecteur de paramètres $(t, x, x, x) = (.5, 1, 1, 1)$

PARAMTRD NUMDRO 315																														
.0	.0	.0	.0	.0	.0	.0	1.0	-1.0	.0																					
.0	.0	.0	.0	.0	1.0	-1.0	.0	.0	.0																					
.0	.0	1.0	-1.0	.0	.0	.0	.0	.0	.0																					

NOMERD C OBBURDNBDS PAR DNVIRONNDMNT 10000																																	
RANG DT PRDQ CU RANG CD BHAQUID ARERD																																	
RANG CDS ARERDS																																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	MDC	SC					
STRUBTURD																																	
D34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	15	14	21	FIN	-186086.	250.			
D14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	16	20	13	FIN	-186083.	251.
D24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	18	16	16	FIN	-186051.	263.		
E24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	15	5	12	6	11	-	-	-	-	FIN	-185916.	263.		
C24	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	15	1	14	6	9	4	1	-	-	-	FIN	-185914.	263.		
E23	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6	14	1	11	4	14	-	-	-	FIN	-185908.	250.		
C14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	14	7	13	2	12	2	-	-	-	FIN	-185908.	251.		
E34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	7	7	1	15	2	18	-	-	-	FIN	-185907.	249.		
C34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	7	6	16	3	16	1	-	-	FIN	-185905.	249.		
D42	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	15	23	11	-	-	-	-	-	-	-	-	-	-	FIN	-185753.	264.		
D41	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	21	11	18	-	-	-	-	-	-	-	-	-	-	FIN	-185738.	254.		
D43	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	14	16	20	-	-	-	-	-	-	-	-	-	-	FIN	-185735.	252.		
B4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	50	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-185582.	260.		
D12	-	-	-	-	0	5	6	6	5	4	6	5	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-185256.	260.		
B2	-	-	-	-	3	11	5	3	7	4	5	6	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-185256.	260.		
D32	-	-	-	-	6	3	8	6	2	7	6	5	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-185255.	260.		
B1	-	-	-	-	3	7	6	4	14	5	1	5	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-185248.	255.		
D21	-	-	-	-	8	3	5	13	3	8	6	3	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-185247.	255.		
D31	-	-	-	-	5	6	5	7	7	10	3	3	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-185247.	255.		
D13	-	-	-	-	6	3	6	5	2	5	8	9	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-185244.	255.		
D23	-	-	-	-	7	5	3	6	-	5	7	7	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-185243.	255.		
B3	-	-	-	-	2	7	6	-	10	2	8	7	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-185243.	255.		
C13	15	15	6	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-185120.	259.		
C12	17	17	5	11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-185120.	259.		
C23	13	7	10	20	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-185120.	259.		
A	5	11	29	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	FIN	-185120.	259.		
DNSDMBLD CD PARAMTRDS NO 16																																	

Figure 4.28 Ordre des arborescences pour le vecteur de paramètres $(t, x, x, x)=(0,1,1,1)$


```

PARAMETRE NUMERO 10
.0 .0 .0 .0 .0 .0 3.0 -3.0 .0
.0 .0 .0 .0 3.0 -3.0 .0 .0 .0
.0 .0 2.0 -2.0 .0 .0 .0 .0 .0

NOMBRE D OCURENCES PAR ENVIRONNEMENT 10000

RANG ET FREQ DE RANG DE CHAQUE ARBRE
RANG DES ARBRES
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 MED SD
STRUCTORD
D34 - - - - - - - - - - - - - - - - - - - - - - - 50 FIN -108092. 283.
E34 - - - - - - - - - - - - - - - - - - - - - - - 50 FIN -107269. 277.
C34 - - - - - - - - - - - - - - - - - - - - - - - 50 FIN -106623. 286.
E24 - - - - - - - - - - - - - - - - - - - - - - - 30 20 FIN -105800. 283.
E23 - - - - - - - - - - - - - - - - - - - - - - - 1 19 30 FIN -105794. 293.
D14 - - - - - - - - - - - - - - - - - - - - - - - 19 31 FIN -105480. 320.
D24 - - - - - - - - - - - - - - - - - - - - - - - 31 18 1 FIN -105478. 312.
C14 - - - - - - - - - - - - - - - - - - - - - - - 20 30 FIN -103919. 313.
C24 - - - - - - - - - - - - - - - - - - - - - - - 1 29 20 FIN -103910. 302.
D43 - - - - - - - - - - - - - - - - - - - - - - - 49 1 FIN -103622. 307.
B4 - - - - - - - - - - - - - - - - - - - - - - - 50 FIN -102871. 305.
D41 - - - - - - - - - - - - - - - - - - - - - - - 21 29 FIN -102091. 313.
D42 - - - - - - - - - - - - - - - - - - - - - - - 29 21 FIN -102079. 304.
D13 - - - - - - - - - - - - - - - - - - - - - - - 20 8 22 FIN -101067. 282.
B3 - - - - - - - - - - - - - - - - - - - - - - - 12 27 11 FIN -101066. 282.
D23 - - - - - - - - - - - - - - - - - - - - - - - 18 15 17 FIN -101066. 282.
D31 - - - - - 7 2 14 11 4 12 - - - - - - - - - - - FIN -100115. 310.
B1 - - - - - 6 14 3 5 16 6 - - - - - - - - - - - FIN -100115. 310.
D21 - - - - - 10 7 6 11 7 9 - - - - - - - - - - - FIN -100115. 310.
D32 - - - - - 9 5 13 6 4 13 - - - - - - - - - - - FIN -100099. 304.
D12 - - - - - 11 7 9 7 8 8 - - - - - - - - - - - FIN -100098. 304.
B2 - - - - - 7 15 5 10 11 2 - - - - - - - - - - - FIN -100098. 304.
C12 12 12 12 14 - - - - - - - - - - - - - - - - - FIN -99012. 302.
A 10 18 16 6 - - - - - - - - - - - - - - - - - FIN -99012. 302.
C23 13 14 6 17 - - - - - - - - - - - - - - - - - FIN -99011. 302.
C13 15 6 16 13 - - - - - - - - - - - - - - - - - FIN -99011. 302.

NUMERO DE PARAMETRES NO 11

```

Figure 4.30 Ordre des arborescences pour le vecteur de paramètres $(0, x, y, z) = (0, 2, 3, 3)$


```

PARAMETRE NUMERO 35
.0 .0 .0 .0 .0 .0 .0 3.0 -3.0 .0
.0 .0 .0 .0 .0 2.0 -2.0 .0 .0 .0
.0 .0 2.0 -2.0 .0 .0 .0 .0 .0

NOMBRE D OCURENCES PAR ENVIRONNEMENT 10000

RANG ET FREQ DE RANG DE CHAQUE ARBRE
RANG DES ARBRES
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 MED SD
STRUCTURE
D24 - - - - - - - - - - - - - - - - - - - - - - 20 30 PIN -116757. 264.
D34 - - - - - - - - - - - - - - - - - - - - - - 30 20 PIN -116705. 265.
E24 - - - - - - - - - - - - - - - - - - - - - - 15 35 - - - PIN -115816. 259.
E34 - - - - - - - - - - - - - - - - - - - - - - 35 15 - - - PIN -115741. 260.
C24 - - - - - - - - - - - - - - - - - - - - - - 17 33 - - - PIN -115265. 265.
C34 - - - - - - - - - - - - - - - - - - - - - - 33 17 - - - PIN -115231. 268.
D14 - - - - - - - - - - - - - - - - - - - - - - 50 - - - - - PIN -114665. 239.
E23 - - - - - - - - - - - - - - - - - - - - - - 50 - - - - - PIN -114134. 202.
C14 - - - - - - - - - - - - - - - - - - - - - - 7 13 30 - - - - - PIN -113157. 221.
D42 - - - - - - - - - - - - - - - - - - - - - - 6 11 20 13 - - - - - PIN -113084. 273.
D43 - - - - - - - - - - - - - - - - - - - - - - 9 17 17 7 - - - - - PIN -113077. 279.
B4 - - - - - - - - - - - - - - - - - - - - - - 35 15 - - - - - PIN -112978. 254.
D41 - - - - - - - - - - - - - - - - - - - - - - 50 - - - - - PIN -112113. 227.
D32 - - - - - - - - - - - - - - - - - - - - - - 7 2 11 13 5 12 - - - - - PIN -110885. 250.
B2 - - - - - - - - - - - - - - - - - - - - - - 6 10 4 12 13 5 - - - - - PIN -110884. 250.
D12 - - - - - - - - - - - - - - - - - - - - - - 7 8 5 5 12 13 - - - - - PIN -110884. 250.
D23 - - - - - - - - - - - - - - - - - - - - - - 14 4 12 6 2 12 - - - - - PIN -110843. 252.
D13 - - - - - - - - - - - - - - - - - - - - - - 9 9 12 10 6 4 - - - - - PIN -110841. 253.
E3 - - - - - - - - - - - - - - - - - - - - - - 7 17 6 4 12 4 - - - - - PIN -110841. 253.
D21 - - - - - 20 15 15 - - - - - - - - - - - - - - - - - - - - - - PIN -110164. 239.
B1 - - - - - 10 24 16 - - - - - - - - - - - - - - - - - - - - - - PIN -110163. 239.
D31 - - - - - 20 11 19 - - - - - - - - - - - - - - - - - - - - - - PIN -220163. 239.
C12 13 13 10 14 - - - - - - - - - - - - - - - - - - - - - - PIN -109499. 245.
C23 14 7 12 17 - - - - - - - - - - - - - - - - - - - - - - PIN -109499. 245.
A 8 20 15 7 - - - - - - - - - - - - - - - - - - - - - - PIN -109499. 245.
C13 15 10 13 12 - - - - - - - - - - - - - - - - - - - - - - PIN -109498. 245.
NUMERO DE PARAMETRES NO 36
    
```

Figure 4.31 Ordre des arborescences pour le vecteur de paramètres $(0, x, y, z) = (0, 2, 2, 3)$

```

PARAMETRE NUMERO 40
.0 .0 .0 .0 .0 .0 3.0 -3.0 .0
.0 .0 .0 .0 2.0 -2.0 .0 .0 .0
.0 .0 1.0 -1.0 .0 .0 .0 .0 .0

NOMBRE D OCURENCES PAR ENVIRONNEMENT 10000

RANG ET FREQ DE RANG DE CHAQUE ARBRE
RANG DES ARBRES
 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 MED SD
STRUBTURD
E34 - - - - - - - - - - - - - - - - - - - - - 1 49 PIN -119428. 366.
C34 - - - - - - - - - - - - - - - - - - - - - 5 45 - PIN -119402. 367.
D34 - - - - - - - - - - - - - - - - - - - - - 45 4 1 PIN -119337. 362.
D43 - - - - - - - - - - - - - - - - - - - - - 50 - - - PIN -117803. 362.
D24 - - - - - - - - - - - - - - - - - - - - - 50 - - - PIN -117585. 347.
D14 - - - - - - - - - - - - - - - - - - - - - 1 49 - - - PIN -116849. 344.
E24 - - - - - - - - - - - - - - - - - - - - - 49 1 - - - PIN -116652. 341.
C24 - - - - - - - - - - - - - - - - - - - - - 1 49 - - - PIN -116471. 347.
B4 - - - - - - - - - - - - - - - - - - - - - 49 1 - - - PIN -116302. 342.
E23 - - - - - - - - - - - - - - - - - - - - - 50 - - - PIN -116071. 328.
C14 - - - - - - - - - - - - - - - - - - - - - 4 46 - - - PIN -115806. 340.
D42 - - - - - - - - - - - - - - - - - - - - - 46 4 - - - PIN -115737. 349.
D41 - - - - - - - - - - - - - - - - - - - - - 14 - 1 35 - - - PIN -115511. 344.
D3 - - - - - - - - - - - - - - - - - - - - - 9 22 13 6 - - - PIN -115425. 349.
D13 - - - - - - - - - - - - - - - - - - - - - 13 12 21 4 - - - PIN -115425. 349.
D23 - - - - - - - - - - - - - - - - - - - - - 14 16 15 5 - - - PIN -115424. 349.
B2 - - - - - - - - - - - - - - - - - - - - - 11 24 15 - - - PIN -114914. 344.
D12 - - - - - - - - - - - - - - - - - - - - - 13 22 15 - - - PIN -114914. 344.
D32 - - - - - - - - - - - - - - - - - - - - - 26 4 20 - - - PIN -114913. 343.
D21 - - - - - 15 16 19 - - - - - - - - - - - - - - - PIN -114686. 337.
B1 - - - - - 10 26 14 - - - - - - - - - - - - - - - PIN -114686. 337.
D31 - - - - - 25 8 17 - - - - - - - - - - - - - - - PIN -114685. 337.
C12 7 15 18 10 - - - - - - - - - - - - - - - - - - - PIN -114505. 340.
A 7 17 17 9 - - - - - - - - - - - - - - - - - - - PIN -114505. 340.
C23 25 3 6 16 - - - - - - - - - - - - - - - - - - - PIN -114504. 340.
C13 11 15 9 15 - - - - - - - - - - - - - - - - - - - PIN -114504. 340.
NUMERO DE PARAMETRES NO 41

```

Figure 4.32 Ordre des arborescences pour le vecteur de paramètres $(0, x, y, z) = (0, 1, 2, 3)$

CONCLUSION

Ce mémoire visait à mieux comprendre les facteurs influençant le choix d'une structure d'arborescence dans le cas de modèles d'arborescences décrivant des séquences d'expériences multinomiales consécutives et/ou parallèles et utilisant le modèle logistique. L'étude de simulation menée a permis de mettre en évidence un ordre relativement stable lorsque les arborescences sont classées selon la grandeur de leurs vraisemblances et d'identifier des sous-groupes d'arborescences équivalentes ou quasi-équivalentes. Nous avons observé que cet ordre dépendait de la forme du vecteur des paramètres et de la valeur de ces derniers. Les théorèmes du chapitre 3 quant à l'effet des paramètres sur l'équivalence des arborescences ont aussi été vérifiés dans un cas concret et ont permis d'éclairer certains résultats. Ils pourraient également être utiles dans des études ultérieures ou peut-être être développés pour éclairer certains résultats plus incertains de la présente étude.

Il serait également pertinent de mener d'autres simulations avec différentes arborescences de départ, pour voir à quel point les résultats obtenus sont variables et aussi, comme nos simulations utilisaient des modèles réduits ou quasi-réduits, de savoir si les résultats sont stables pour des paramètres proches de 0 au lieu de 0, le tout dans l'espoir de réussir à mettre en évidence les cas pour lesquels on peut utiliser indifféremment des structures d'arborescences distinctes, ou passer d'une structure à une autre selon les besoins d'une étude sans perte de précision trop grande.

Il est intéressant d'identifier certaines arborescences au modèle standard, ce qui laisse supposer qu'il n'y a pas de perte d'information lorsqu'on utilise ces arborescences équivalentes. Une étape ultérieure serait de comparer les régions des valeurs des paramètres pour lesquels nous avons trouvé ce résultat avec celles de l'étude de Bull et Donner pour examiner ce qui concorde et voir s'il est possible d'en dégager des principes généraux.

BIBLIOGRAPHIE

- Begg C. et Gray R., « Calculation of polychotomous logistic regression parameters using individualized regressions », *Biometrika*, vol. 71-1, p.11-18, 1984.
- Bull S. et Donner A., « A characterization of the efficiency of individualized logistic regressions », *The Canadian Journal of Statistics*, vol. 21, No. 1, p.71-78, 1993.
- Bull S., Greenwood C. et Donner A., « Efficiency of reduced logistic regression models », *The Canadian Journal of Statistics*, vol. 22, No. 3, p.319-334, 1994.
- Fredette J., « Caractérisation de l'efficacité de l'estimateur individuel et des modèles d'arbres en régression logistique multinomiale », Mémoire de maîtrise. Montréal, Université du Québec à Montréal, 2001.
- Richard H., « Modèle logistique, modèle log-linéaire et famille de tests d'ajustement basée sur les critères de divergence », Mémoire de maîtrise, Montréal, Université du Québec à Montréal, 1999.
- Rosenberg, Paul, Professeur, Centre de recherches en mathématiques de l'Université de Montréal, communication privée.
- Rousseau P. et Sankoff D., « The Inference of Hierarchical Schemes for Multinomial Data », *Journal of Classification*, vol.6, p.73-95, 1989.