

Introduction aux modèles graphiques

Version de travail

G rard Letac ¹, H l ne Massam²

¹Laboratoire de Statistique et Probabilit s, Universit  Paul Sabatier, Toulouse.

²Department of Mathematics and Statistics, York University, Toronto

1 Le paradoxe de Simpson.

Si X et Y sont deux variables aléatoires définies sur un même espace, il est traditionnel de noter $X \perp Y$ l'affirmation qu'elles sont indépendantes. Cette notation n'est pas parfaite, car inadaptée aux cas de trois variables ou plus, mais elle est commode ici.

Le paradoxe de Simpson est illustré par une université qui a 48 milliers d'étudiants, répartis en 24 milliers de garçons et 24 milliers de filles, et pour simplifier, répartis en 3 facultés : Lettres, Droit, Science, de 16 milliers d'étudiants chacun. Il y a eu de violentes manifestations, car aux examens de fin d'année, il y a eu 10 milliers de filles recues, 14 milliers de filles collées, alors que chez les garçons il y avait 14 milliers de recus et 10 milliers de collés. Autrement dit, moins de 42% de recues, et plus de 58% de recus. Mais le président de l'université, un statisticien, a su désamorcer la crise en regardant la réalité de plus près dans les trois facultés. En lettres, droit et sciences les tableaux recus collés - garçons filles, en milliers, sont les suivants :

Lettres	Recus	Colles
Filles	3	9
Garçons	1	3

Droit	Recus	Colles
Filles	4	4
Garçons	4	4

Sciences	Recus	Colles
Filles	3	1
Garçons	9	3

Si ces chiffres sont traduits en pourcentages, l'ensemble des étudiants est un espace de probabilité muni de la loi uniforme. Soit X , Y et Z des variables aléatoires qualitatives qui mesurent le résultat de l'examen (pour X = recu ou collé), le genre (pour Y = masculin ou féminin) et la faculté (pour Z = lettres, droit ou sciences), alors on observe que *si on conditionne par Z* alors on a $(X \perp Y)|Z$. Le résultat initial s'explique alors par le fait que les filles ont choisi en majorité des disciplines où le nombre de recus est moins grand.

L'analyse d'une situation statistique doit donc prendre en compte ces phénomènes d'indépendance conditionnelle, soit pour les tester, soit pour les utiliser dans les problèmes d'estimation. Ils sont d'une grande généralité, et le but du cours est de les étudier dans des conditions sensiblement plus simples, à savoir

- Le cas gaussien, dit modèle continu.
- Le cas des tables de contingence, dit modèle discret.

Dans ces cas simples, l'information disponible va être schématisée par un graphe (V, E) non orienté ordinaire dans le cas continu et le cas discret, et en deux couleurs dans le cas mixte. Toutefois, les problèmes d'estimation et de

test mentionnés ci dessus ne connaîtront de solution explicite que si le graphe satisfait à une surprenante condition, celle d'être *triangulé* : jamais de cycle d'ordre > 3 sans qu'il soit possible de le court circuiter par une corde. Cette condition avait déjà été rencontrée en analyse numérique dans les années 60 dans la méthode du pivot et a été redécouverte 20 ans plus tard en statistique. Ces graphes portent aussi d'autres noms dans la littérature (plus une chose a d'utilisateurs en mathématiques, plus elle a de noms) : ils sont appelés graphes *cordaux*, on comprend pourquoi, et aussi *décomposables*, pour de très bonnes raisons qu'on verra apparaître quand il faudra faire des calculs.

Finalement, les modèles utilisés sont des familles exponentielles, ce qui permet d'utiliser une méthode classique pour l'estimation, celle du maximum de vraisemblance. Cette méthode fournit une équation naturelle pour l'estimateur. Dans le cas des modèles graphiques, elle ne sera justement résoluble que si le graphe est triangulé.

Le plan de notre travail est donc d'abord la présentation des modèles graphiques gaussiens aussi dits *continus*, puis celle des modèles discrets. Ensuite : liquidation d'analphabétisme en ce qui concerne les familles exponentielles. Ensuite, des résultats pas commodes à trouver dans la littérature sur les graphes triangulés. Ensuite, application des outils précédents à l'estimation.

2 Modèles graphiques gaussiens

Pour définir ceux ci, nous faisons quelques rappels. Le premier concerne l'algèbre linéaire. On note par B^T la transposée de la matrice B .

Proposition 2.1 : Soit la matrice par blocs $\begin{bmatrix} x_1 & x_{12} \\ x_{21} & x_2 \end{bmatrix}$. On suppose que x_1 et x_2 sont carrées et que x_1^{-1} existe. Alors

$$\begin{bmatrix} x_1 & x_{12} \\ x_{21} & x_2 \end{bmatrix} = \begin{bmatrix} I_1 & 0 \\ x_{21}x_1^{-1} & I_2 \end{bmatrix} \begin{bmatrix} x_1 & 0 \\ 0 & x_2 - x_{21}x_1^{-1}x_{12} \end{bmatrix} \begin{bmatrix} I_1 & x_1^{-1}x_{12} \\ 0 & I_2 \end{bmatrix} \quad (2.1)$$

En particulier

$$\det \begin{bmatrix} x_1 & x_{12} \\ x_{21} & x_2 \end{bmatrix} = \det x_1 \det(x_2 - x_{21}x_1^{-1}x_{12}), \quad (2.2)$$

Si de plus x_2 et la matrice initiale sont inversibles alors

$$\begin{bmatrix} x_1 & x_{12} \\ x_{21} & x_2 \end{bmatrix}^{-1} = \begin{bmatrix} (x_1 - x_{12}x_2^{-1}x_{21})^{-1} & X \\ X^T & (x_2 - x_{21}x_1^{-1}x_{12})^{-1} \end{bmatrix}, \quad (2.3)$$

où $X = -x_1^{-1}x_{12}(x_2 - x_{21}x_1^{-1}x_{12})^{-1} = (x_1 - x_{12}x_2^{-1}x_{21})^{-1}x_{12}x_2^{-1}$.

Démonstration : 2.1 est une vérification. En prenant le déterminant des deux membres de 2.1 et en utilisant le théorème sur le déterminant d'une matrice

triangulaire par blocs, on a 2.2. En prenant l'inverse de 2.1 on a

$$\begin{bmatrix} x_1 & x_{12} \\ x_{21} & x_2 \end{bmatrix}^{-1} = \begin{bmatrix} I_1 & -x_1^{-1}x_{12} \\ 0 & I_2 \end{bmatrix} \begin{bmatrix} x_1^{-1} & 0 \\ 0 & (x_2 - x_{21}x_1^{-1}x_{12})^{-1} \end{bmatrix} \begin{bmatrix} I_1 & 0 \\ -x_{21}x_1^{-1} & I_2 \end{bmatrix}$$

qui fournit en effectuant le bloc (2,2) égal à $(x_2 - x_{21}x_1^{-1}x_{12})^{-1}$ qui était annoncé. Pour le bloc (1,1), il est inutile de refaire les calculs, on raisonne par symétrie.

Voici ensuite un point de vocabulaire sur les v.a. gaussiennes. Si $(X_1, \dots, X_n)^T$ est une variable gaussienne de moyenne $m \in \mathbb{R}^n$ et de matrice de covariance Σ supposée inversible, alors la matrice $W = \Sigma^{-1}$ est appelée la *matrice de précision*. C'est un terme commode, et cette matrice intervient plus souvent ici que la matrice de covariance.

Proposition 2.2 : Soit (X, Y) une variable gaussienne de \mathbb{R}^{p+q} de moyenne (x_0, y_0) , de covariance inversible

$$\Sigma = \begin{bmatrix} \alpha & \beta \\ \beta^T & \gamma \end{bmatrix} \quad (2.4)$$

et de matrice de précision

$$W = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}. \quad (2.5)$$

Alors

1. Il y a équivalence entre $X \perp Y$, $\beta = 0$ et $B = 0$.
2. X est normale de moyenne x_0 de covariance α et de matrice de précision $A - BC^{-1}B^T$.
3. La loi de Y conditionnée par X est gaussienne de moyenne $y_0 + C^{-1}B^T(X - x_0)$, de matrice de précision C et de covariance $\gamma - \beta^T\alpha^{-1}\beta$.

La prochaine proposition est cruciale et est à rapprocher de la précédente.

Proposition 2.3 : Soit (X_1, X_2, X_3) une variable gaussienne de \mathbb{R}^{p+q+r} de matrice de précision

$$W = \begin{bmatrix} W_{11} & W_{12} & W_{13} \\ W_{21} & W_{22} & W_{23} \\ W_{31} & W_{32} & W_{33} \end{bmatrix}. \quad (2.6)$$

Alors $(X_1 \perp X_3)|X_2$ si et seulement si $W_{13} = 0$. Si la matrice de précision W et la covariance Σ de (X_1, X_2, X_3) sont écrites par blocs avec les notations 2.4 et (2.5, où A et α sont carrées de taille p , alors $(X_1 \perp X_3)|X_2$ si et seulement si la matrice $\gamma - \beta^T\alpha\beta$ est diagonale par blocs associés à la décomposition (p, r) de $p + r$. En particulier, dans ces conditions, si $p_{1,2,3}$ est la densité de

(X_1, X_2, X_3) , $p_{1,2}$ est la densité de (X_1, X_2) , $p_{2,3}$ est la densité de (X_2, X_3) , et p_2 est la densité de X_2 on a

$$p_{1,2,3}(x_1, x_2, x_3)p_2(x_2) = p_{1,2}(x_1, x_2)p_{2,3}(x_2, x_3). \quad (2.7)$$

Rappelons qu'un graphe (non orienté) est la donnée d'un ensemble fini V dont les points sont appelés *sommets* (V pour vertex ou vertices) et une partie E de la famille des sous ensembles de taille 2 de V . Les éléments de A sont appelés *arêtes* (E pour edges). Insistons sur les faits suivants, entraînés par cette définition : les arêtes ne sont pas orientées, il n'a entre deux sommets qu'au plus une seule arête (pas d'arêtes dites multiples), il n'y a pas d'arête qui va d'un sommet à lui même : pas de boucle). Si toutes les arêtes possibles sont présentes, on dit que le graphe est *complet*.

Nous sommes alors en position de décrire un modèle graphique gaussien. Donnons nous un graphe (V, E) tel que $V = \{1, 2, \dots, n\}$. Le *modèle graphique gaussien* gouverné par ce graphe est l'ensemble des lois normales de \mathbb{R}^n de moyenne inconnue m , non singulières, et dont la matrice de précision $W = (W_{ij})$ est telle que $W_{ij} = 0$ si $i \neq j$ et si $\{i, j\}$ n'est pas dans A . Si le graphe est complet, le modèle est dit *saturé*.

Ces modèles ont été introduits en 1972 par H. Dempster. Ils sont aussi appelés dans la littérature anglo saxonne *covariance selection models*, nom ce que nous trouvons remarquablement peu informatif. Comme le mot modèle graphique est également à toutes les sauces, l'idéal serait des les appeler *modèles de Dempster*. Nous avons toutefois suivi le choix d'Edwards (1995) en parlant de modèles graphiques gaussiens.

Compte tenu de la Proposition 2.3, la définition ci dessus est équivalente à la seule contrainte suivante sur la loi gaussienne de (X_1, \dots, X_n) : Si i et j ne sont pas liés par une arête du graphe (V, E) , alors X_i et X_j conditionnées par le reste des variables aléatoires de (X_1, \dots, X_n) sont indépendantes. Si le modèle est saturé, il n'y a naturellement pas de contraintes du tout, et l'estimation de la covariance par maximum de vraisemblance se fait par une méthode standard qu'on rappellera au chapitre des familles exponentielles.

L'intérêt particulier de ce modèle tient au fait que la contrainte sur la matrice de précision W est linéaire. Toutefois, comme déjà dit et comme on le verra à la section 13, l'estimation explicite par maximum de vraisemblance ne sera possible que si le graphe est triangulé.

Insistons sur le fait que ces modèles graphiques gaussiens ne rendent pas compte de tous les modèles gaussiens qui comprendraient des contraintes d'indépendance conditionnelle. Par exemple dans le cas $n = 5$ on peut imaginer les contraintes $(X_i \perp X_{i+1})|(X_{i+2}, X_{i+3})$ avec $i = 1, 2, 3, 4, 5$ et en prenant les indices modulo 5. Cela ne se traduit pas par une prescription des zéros de la matrice de précision de (X_1, \dots, X_5) mais par une prescription des zéros des 5 matrices de précision des $(X_i, X_{i+1}, X_{i+2}, X_{i+3})$. Le problème de l'estimation explicite de la matrice de covariance ou mieux, de la matrice de préci-

sion W , par maximum de vraisemblance devient impossible. En particulier, la contrainte sur W n'est plus linéaire.

Exercice 2.1. Soit la matrice symétrique définie positive $\begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$ où A et C sont carrées, et soit A_1 une matrice définie positive de même dimension que A . Montrer que

$$S = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}^{-1} + \begin{bmatrix} A_1^{-1} - A^{-1} & 0 \\ 0 & 0 \end{bmatrix}$$

est définie positive et que $S^{-1} = \begin{bmatrix} A_1 & B_1 \\ B_1^T & C_1 \end{bmatrix}$ pour des B_1 et C_1 convenables. Méthode : appliquer la formule 2.1).

Exercice 2.2. Soit la matrice symétrique définie positive $S = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$ où A et C sont carrées, et soit $S^{-1} = \begin{bmatrix} A_1 & B_1 \\ B_1^T & C_1 \end{bmatrix}$. Montrer que $A_1 - A^{-1}$ est positive. Méthode : appliquer la formule 2.3) pour montrer que $A - A_1^{-1}$ est positive.

Exercice 2.3. Soit la matrice définie positive

$$W = \begin{bmatrix} W_{11} & W_{12} & W_{13} \\ W_{21} & W_{22} & 0 \\ W_{31} & 0 & W_{33} \end{bmatrix},$$

et soit Σ son inverse :

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{bmatrix}.$$

Montrer que

$$\det \Sigma \det \Sigma_{11} = \det \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \det \begin{bmatrix} \Sigma_{11} & \Sigma_{13} \\ \Sigma_{31} & \Sigma_{22} \end{bmatrix}.$$

Méthode : faire $x = 1 = x_2 = x_3 = 0$ dans 2.7). Source : Speed et Kiiveri (1986). La Proposition 9.7 contient une importante généralisation.

3 Modèles hiérarchiques et graphiques discrets

Si A est un ensemble fini quelconque et si $(p_a)_{a \in A}$ est une probabilité sur A , soit $(e_a)_{a \in A}$ la base canonique de \mathbb{R}^A et soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes à valeurs dans \mathbb{R}^A et de loi définie par $\Pr(X_n = e_a) = p_a$. Alors la variable aléatoire $S_N = X_1 + \dots + X_N$ de \mathbb{R}^A suit la loi multinomiale :

$$\Pr(S_N = (s_a)_{a \in A}) = \frac{N!}{\prod_{a \in A} s_a!} \prod_{a \in A} p_a^{s_a}$$

si $(s_a)_{a \in A} \in \mathbf{N}^A$ et $\sum_{a \in A} s_a = N$, avec $\Pr(S_N = (s_a)_{a \in A}) = 0$ dans les autres cas (\mathbf{N} est l'ensemble des entiers ≥ 0). La loi des grands nombres garantit que S_N/N tend vers $(p_a)_{a \in A}$. Si maintenant $(p_a)_{a \in A}$ est inconnu, S_N/N est le candidat idéal pour un estimateur de $(p_a)_{a \in A}$ (plus sur ce sujet au chapitre des familles exponentielles). D'où l'idée naturelle en statistique d'observer S_N .

Supposons maintenant que A soit le produit $A_1 \times \cdots \times A_\Delta = \times_{v \in V} A_v$ de Δ ensembles finis (on note $V = \{1, \dots, \Delta\}$ par commodité ici). Un élément $f = (f_a)_{a \in A}$ de \mathbb{R}^A (comme par exemple $(p_a)_{a \in A}$) devient alors une fonction de Δ variables discrètes, plutôt noté

$$f_a = f(a_1, \dots, a_\Delta) = f((a_v)_{v \in V}).$$

Dans l'exemple de l'université de la section 1, on a $V = (\text{Résultat, Genre, Faculté}) = (1, 2, 3)$ et donc $\Delta = 3$, avec $A_1 = (\text{reçu, collé})$, $A_2 = (\text{fille, garçon})$, $A_3 = (\text{lettres, droit, sciences})$.

Dans ce cas, un élément particulier f de \mathbb{R}^A s'écrit non plus comme une ligne ou une colonne de nombres, mais comme un "tableau". Les guillemets sont pour indiquer que ce n'est véritablement un tableau que si $\Delta = 2$. Si $\Delta = 3$, ce serait un tableau en trois dimensions, ce qui n'est guère commode, à moins qu'on ne le représente par des coupes à deux dimensions comme on le voit avec les trois tables de la page 1.

Même dans ce cadre plus spécialisé, on a la possibilité théorique de procéder à l'estimation de la probabilité inconnue $p(a_1, \dots, a_n)$ à l'aide de S_n . Cet S_n se présente sous forme de tableau et est appelé un *tableau de contingence*.

Ceci dit, le nombre de paramètres $p(a_1, \dots, a_n)$ à estimer à l'aide d'un tableau de contingence peut être d'une ampleur décourageante (soit $|A_1| \times \cdots \times |A_\Delta| - 1$) et trop grande par rapport aux nombres n d'observations dont on dispose. En faisant certaines hypothèses raisonnables, on peut diminuer considérablement le nombre de paramètres inconnus. Le plus simple des exemples est celui où $\Delta = 2$ et où on suppose que la variable aléatoire (a_1, a_2) de loi $(p(a_1, a_2))$ est formée de deux variables aléatoires indépendantes. Autrement dit, on postule que $f(a_1, a_2) = \log p(a_1, a_2)$ est la somme d'une fonction de a_1 seul et d'une fonction de a_2 seul. Il n'y a en fait plus que $|A_1| - 1 + |A_2| - 1$ paramètres à estimer.

Les modèles hiérarchiques discrets est une extension de cette idée à un Δ quelconque. Soit $\{V_1, \dots, V_k\}$ une famille de parties de V . Pour de bonnes raisons statistiques, on peut postuler que $f(a_1, a_2, \dots, a_N) = \log p(a_1, a_2, \dots, a_\Delta)$ est une somme de k fonctions g_j , où g_j est une fonction ne dépendant que des variables a_v telles que $v \in V_j$, ou encore : g_j est une fonction définie sur le produit cartésien $\times_{i \in V_j} A_i$. Dans l'exemple précédent on avait $\Delta = k = 2$, $V_1 = \{1\}$ et $V_2 = \{2\}$.

Trivialement si un V_j est contenu dans un $V_{j'}$ la contrainte apportée par V_j est inexistante. Pour cette raison nous dirons qu'une famille de parties $\{V_1, \dots, V_k\}$ de V est une *famille de Sperner* si $j \neq j'$ implique que $V_j \not\subset V_{j'}$ est impossible (Sperner faisait de la topologie combinatoire, et le lien entre

cette discipline et les modèles graphiques est un sujet d'étude contemporain). Lauritzen, Speed et Vijayan (1984) disent *hypergraphe* ou encore *classe génératrice*. Notez qu'en général, les V_j ne sont pas disjoints. Voici alors une définition importante :

Soit $A = A_1 \times \cdots \times A_N = \times_{v \in V} A_v$ le produit de Δ ensembles finis avec $V = \{1, \dots, \Delta\}$ et soit une famille de parties $\{V_1, \dots, V_k\}$ de V qui soit de Sperner. On appelle *modèle hiérarchique discret* engendré par $\{V_1, \dots, V_k\}$ l'ensemble des probabilités $(p_a)_{a \in A}$ telles que $p_a > 0$ pour tout $a \in A$ et telles que

$$\log p_a = \sum_{j=1}^k g_j(a) \quad (3.8)$$

où la fonction $a \mapsto g_j(a)$ ne dépend que des variables a_v telles que $v \in V_j$. Si la famille de Sperner est réduite à $\{V\}$, le modèle graphique discret est dit *saturé*.

Ouvrons une parenthèse sur l'unicité de la représentation d'une fonction f de \mathbb{R}^A sous la forme $\sum_{j=1}^k g_j(a)$ où les g_j sont comme ci dessus. Même dans le cas simple $f(a_1, a_2) = g_1(a_1) + g_2(a_2)$ il n'y a pas unicité, puisque si c est un réel quelconque alors $g_1^*(a_1) = g_1(a_1) + c$ et $g_2^*(a_2) = g_2(a_2) - c$ sont des fonctions de a_1 seul et de a_2 seul telles que $f = g_1^* + g_2^*$. Ce point est clarifié par la proposition suivante.

Proposition 3.1 Soit $A = A_1 \times \cdots \times A_\Delta = \times_{v \in V} A_v$ le produit de Δ ensembles finis avec $V = \{1, \dots, \Delta\}$. L'espace \mathbb{R}^A est muni de la structure euclidienne canonique. Si $S \subset V$ on note E_S le sous espace de \mathbb{R}^A des f tels que $f(a_1, \dots, a_n)$ ne dépende que des variables a_v telles que $v \in S$. On note par E_S^\perp son orthogonal dans \mathbb{R}^A et par F_S l'intersection de E_S avec tous les $E_{S'}^\perp$ où $S' \subset S$ et où $S' \neq S$. Alors les espaces F_S quand $S \subset V$ sont en somme directe orthogonale. L'espace F^\emptyset est l'espace des constantes et, si S est non vide, F_S est l'espace des $f \in E_S$ tels que pour tout $i \in S$ on ait $\sum_{a_i \in A_i} f(a) = 0$. On a aussi

$$\dim E_S = \prod_{v \in V} |A_v|, \quad \dim F_S = \prod_{v \in V} (|A_v| - 1), \quad E_S = \oplus_{S' \subset S} F_{S'}. \quad (3.9)$$

En particulier, si $\{V_1, \dots, V_k\}$ est une famille de Sperner, on a

$$E_{V_1} + \cdots + E_{V_k} = \oplus' F_S \quad (3.10)$$

où \oplus' indique que la somme est prise pour tous les S qui sont sous ensemble d'au moins un V_j .

Remarques : Avec la notation de l'égalité 3.10) on voit que p est dans le modèle hiérarchique gouverné par $\{V_1, \dots, V_k\}$ si et seulement si $\log p \in E_{V_1} + \cdots + E_{V_k}$. Notez que dire que S est sous ensemble d'au moins un V_j est plus fort que de dire que S est sous ensemble de $\cup_j V_j$. Croire à la réciproque est une erreur fréquente.

Pour définir les modèles graphiques discrets, voici des points de vocabulaire de la théorie des graphes. Si (V, E) est un graphe non orienté et si $V_1 \subset V$, soit E_1 l'ensemble des arêtes $\{i, j\}$ de E telles que i et j soient dans E . Le graphe (V_1, E_1) est alors appelé le graphe *induit* de (V, E) sur V_1 (Dans la littérature il y a un objet plus général, appelé *sous graphe* (V_1, E'_1) où E'_1 est simplement astreint à être un sous ensemble de E_1). Une *clique* de (V, E) est alors une partie V_1 de V telle que le graphe induit de (V, E) sur V_1 soit *complet*. Une clique V_1 est dite *maximale* si $V_1 \subset V'_1$ et $V'_1 \neq V_1$ entraîne que V'_1 n'est pas une clique. Il est évident et fondamental d'observer que l'ensemble des cliques maximales d'un graphe (V, E) forme une famille de Sperner de V .

(Il serait plus simple d'appeler clique tout court ce qu'on appelle clique maximale et sous graphes complets les cliques ordinaires, mais de pénibles discussions avec des théoriciens des graphes m'ont montré que c'était peine perdue).

Soit alors un modèle hiérarchique discret gouverné par la famille de Sperner $\{V_1, \dots, V_k\}$ On appelle *graphe associé* le graphe dont les arêtes sont tous les sous ensembles à deux éléments d'au moins un V_j . On dit aussi *graphe d'interaction*. Voici trois exemples de familles de Sperner sur $\{1, 2, 3, 4\}$. Le lecteur est invité à dessiner les graphes correspondants

1. $V_1 = \{2, 3, 4\}$, $V_2 = \{1, 2\}$.
2. $V_1 = \{2, 3, 4\}$, $V_2 = \{1, 2\}$, $V_3 = \{1, 4\}$,
3. $V_1 = \{2, 3, 4\}$, $V_2 = \{1, 2, 4\}$.

On voit que les exemples 2) et 3) ont même graphe : celui ci n'est donc pas caractéristique de la famille de Sperner, et donc du modèle hiérarchique qu'elle gouverne.

On dira alors que le modèle hiérarchique gouverné par la famille de Sperner $\{V_1, \dots, V_k\}$ est *graphique* si les cliques maximales du graphe associé sont les $\{V_1, \dots, V_k\}$. La connaissance du graphe d'un modèle graphique discret donne la connaissance de ses cliques maximales, donc de la famille de Sperner qui le gouverne : le graphe détermine le modèle graphique, à la différence du modèle hiérarchique. Dans les trois exemples précédents, on voit que le 1) et le 3) sont graphiques et que le 2) ne l'est pas.

4 Familles exponentielles naturelles et générales.

Comme expliqué à la section 1, les modèles que nous considérons ici appartiennent à la grande famille des modèles exponentiels, et héritent de leurs propriétés générales, qu'on va passer en revue maintenant.

Une famille exponentielle naturelle est définie (on va voir comment dans un instant) sur un espace vectoriel réel de dimension finie E . Résistons ici à la tentation de confondre E et \mathbb{R}^n , ce qui nous lierait trop à un système particulier de coordonnées. L'espace dual E^* est l'espace des formes linéaires θ sur E . On notera la valeur $\theta(x)$ prise par la forme $\theta \in E^*$ sur le vecteur

$x \in E$ plutôt par $\langle \theta, x \rangle$. Si E est muni d'une structure euclidienne gouvernée par le produit scalaire \langle, \rangle on sait que E^* s'identifie canoniquement à E et la notation $\langle \theta, x \rangle$ est compatible avec cette identification.

Ensuite, on considère une mesure μ (positive) sur E , pas nécessairement bornée, et sa *transformée de Laplace*, c'est à dire la fonction définie sur E^* à valeurs dans $[0, \infty]$ définie par

$$L_\mu(\theta) = \int_E e^{\langle \theta, x \rangle} \mu(dx).$$

On note $k_\mu = \log L_\mu$ qu'on appelle *fonction de cumulants* de μ . On note

$$D(\mu) = \{\theta \in E^*; L_\mu(\theta) < \infty\}$$

et $\Theta(\mu)$ l'intérieur de $D(\mu)$. On appelle ce dernier ensemble le *domaine* de μ .

Proposition 4.1 Soit μ une mesure sur E . Alors

1. Les ensembles $D(\mu)$ et $\Theta(\mu)$ sont convexes, k_μ est convexe sur $D(\mu)$, et analytique réelle sur $\Theta(\mu)$.
2. Si de plus μ n'est pas concentrée sur un hyperplan affine de E et si $\Theta(\mu)$ n'est pas vide, alors k_μ est strictement convexe sur $\Theta(\mu)$, et sa différentielle sur cet ensemble satisfait

$$k'_\mu(\theta) = \frac{L'_\mu(\theta)}{L_\mu(\theta)} = \int_E x e^{\langle \theta, x \rangle - k_\mu(\theta)} \mu(dx). \quad (4.11)$$

Une fois pour toutes on note alors $\mathcal{M}(E)$ l'ensemble des mesures μ sur E telles que μ ne soit pas concentrée sur un hyperplan affine et telles que $\Theta(\mu)$ soit non vide. Si $\mu \in \mathcal{M}(E)$ et si $\theta \in \Theta(\mu)$, alors

$$P(\theta, \mu)(dx) = e^{\langle \theta, x \rangle - k_\mu(\theta)} \mu(dx)$$

est une probabilité sur E . Leur ensemble

$$F = F(\mu) = \{P(\theta, \mu); \theta \in \Theta(\mu)\}$$

est appelé *famille exponentielle naturelle* engendrée par μ . Notez que $F(\mu)$ est un ensemble, alors que $\theta \mapsto P(\theta, \mu)$ est une paramétrisation de cet ensemble, qui admet d'autres paramétrisations utiles. Par exemple, $F(\mu) = F(\mu')$ si et seulement si il existe $\theta_0 \in E^*$ et $b \in \mathbb{R}$ tels que $\mu'(dx) = e^{\langle \theta_0, x \rangle + b} \mu(dx)$. La paramétrisation correspondante serait obtenue de la précédente par une simple translation $\theta \mapsto \theta - \theta_0$.

Une autre paramétrisation de $F(\mu)$ plus importante est obtenue à partir de la formule 4.11), qui peut être reformulée en disant que $k'_\mu(\theta)$ est la moyenne $m \in E$ de la probabilité $P(\theta, \mu)$. Ensuite, puisque maintenant k_μ est strictement convexe sur $\Theta(\mu)$, alors $\theta \mapsto k'_\mu(\theta) = m$ devient une fonction *injective*

(d'après les propriétés des fonctions convexes) de $\Theta(\mu)$ à valeurs dans E . L'ensemble $M(F) = M(F(\mu))$ des valeurs qu'elle prend s'appelle le *domaine des moyennes* de $F(\mu)$, et on note désormais par ψ_μ la fonction réciproque de k'_μ . La fonction $m \mapsto \psi_\mu(m)$ est donc une bijection de $M(F)$ et de $\Theta(\mu)$ telle que

$$\psi_\mu(k'_\mu(\theta)) = \theta, \quad k'_\mu(\psi_\mu(m)) = m.$$

Par conséquent l'application de $M(F)$ dans F définie par $m \mapsto P(m, F) = P(\psi_\mu(m), \mu)$ est une bijection appelée *paramétrisation par la moyenne* de la famille exponentielle naturelle F . Le théorème des fonctions implicites entraîne que M_F est un *ouvert* de E homéomorphe à E . Dans les cas classiques qui nous intéressent dans ce cours, c'est un convexe, mais pas en général.

Étudions maintenant les propriétés de ces familles du point de vue de la convolution. Si $\mu \in \mathcal{M}(E)$ on appelle ensemble de Jorgensen de μ l'ensemble $\Lambda(\mu)$ des $\lambda > 0$ ayant la propriété suivante : il existe un $\mu_\lambda \in \mathcal{M}(E)$ tel que $\Theta(\mu_\lambda) = \Theta(\mu)$ et tel que $L_{\mu_\lambda} = (L_\mu)^\lambda$. Observons que $\Lambda(\mu)$ contient toujours l'ensemble \mathbb{N}^* des entiers > 0 . En effet même si μ n'est pas une mesure bornée, le fait que $\mu \in \mathcal{M}(E)$ entraîne que ses puissances de convolution μ^{*n} existent et sont dans $\mathcal{M}(E)$, avec $(L_\mu)^n$ pour transformées de Laplace. Donc $\mu_\lambda = \mu^{*n}$ pour $\lambda = n \in \mathbb{N}^*$. Quel lien y a-t-il entre les familles exponentielles naturelles $F_\lambda = F(\mu_\lambda)$ et $F = F(\mu)$? On voit facilement à partir des définitions que pour λ dans l'ensemble de Jorgensen, et en particulier pour λ entier on a :

$$M(F_\lambda) = \lambda M(F), \quad \psi_{\mu_\lambda}(m) = \psi_\mu\left(\frac{m}{\lambda}\right) \quad (4.12)$$

Une chose simple mais importante est que

$$P(\theta, \mu)^{*n} = P(\theta, \mu^{*n}).$$

Pour le voir, il suffit de calculer les transformées de Laplace des deux membres : Prenons α pour paramètre de la transformée de Laplace, puis que θ est déjà pris : pour tout $\alpha \in \Theta(\mu) - \theta$ on a

$$L_{P(\theta, \mu)}(\alpha) = \frac{L_\mu(\theta + \alpha)}{L_\mu(\theta)}, \quad L_{P(\theta, \mu^{*n})}(\alpha) = \frac{L_{\mu^{*n}}(\theta + \alpha)}{L_{\mu^{*n}}(\theta)} = \left(\frac{L_\mu(\theta + \alpha)}{L_\mu(\theta)} \right)^n = L_{(P(\theta, \mu))^{*n}}(\alpha).$$

Une conséquence théorique est alors la suivante : si X_1, \dots, X_n sont indépendantes et de même loi $P(\theta, \mu)$ où θ est inconnu, alors la loi de $X_1 + \dots + X_n$ appartient à la famille exponentielle naturelle $F(\mu^{*n})$ avec le même paramètre inconnu. Si bien que dans les problèmes d'estimation du paramètre inconnu θ dont nous allons parler maintenant, on peut imaginer qu'on fait l'estimation avec *une seule* observation.

Proposition 4.2. Soit $\mu \in \mathcal{M}(E)$ et X une variable aléatoire de loi $P(\theta, \mu)$ où θ est inconnu. Alors si la valeur de X est dans le domaine des moyennes

$M(F(\mu))$, l'estimateur du *maximum de vraisemblance* pour θ est défini et est égal à $\psi_\mu(X)$.

Dans la pratique, on va appliquer ce résultat au cas où $X = X_1 + \dots + X_n$ comme ci dessus. D'après 4.12), si la moyenne empirique $\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$ est dans M_F alors l'estimateur de θ du maximum de vraisemblance est $\psi_\mu(\bar{X}_n)$. Subsiste la difficulté que pour que cet estimateur existe, il faut que \bar{X}_n soit dans $M(F)$. Pour nous réconforter, il y a la loi forte des grands nombres, qui dit ici que la limite presque sûre de la suite $(\bar{X}_n)_{n \geq 1}$ est $k'_\mu(\theta) \in M(F)$. Comme on a vu que $M(F)$ est un ouvert un tel résultat garantit l'existence d'un n_0 (aléatoire il est vrai) tel que \bar{X}_{n_0} appartienne à $M(F)$.

Les familles exponentielles ont également une propriété concernant les statistiques suffisantes qui découle de la proposition suivante :

Proposition 4.3. Soit $\mu \in \mathcal{M}(E)$, soit $\lambda_1, \dots, \lambda_n$ dans l'ensemble de Jorgensen $\Lambda(\mu)$ et soit $\theta \in \Theta(\mu)$ fixé. On considère des variables aléatoires indépendantes X_1, \dots, X_n telles que X_j soit de loi $P(\theta, \mu_{\lambda_j})$. On pose $S = X_1 + \dots + X_n$ et $\lambda = \lambda_1 + \dots + \lambda_n$. Alors S est de loi $P(\theta, \mu_\lambda)$ et la loi conditionnelle de (X_1, \dots, X_n) sachant $S = s$ ne dépend pas de θ .

Le cas particulier important de la proposition précédente est évidemment le cas $\lambda_j = 1$ pour tout j . Il entraîne que pour estimer θ à partir de l'échantillon (X_1, \dots, X_n) il est inutile de stocker individuellement les valeurs des (X_1, \dots, X_n) : toute l'information disponible sur θ à partir des observations (X_1, \dots, X_n) est contenue dans $S = X_1 + \dots + X_n$. La proposition 4.3 entraîne que la statistique S est *exhaustive*, ou *suffisante*, pour θ .

Exemples.

1) Loi de Bernoulli, loi binomiale. On prend $E = \mathbb{R}$ et $\mu = \delta_0 + \delta_1$. Alors $\Lambda(\mu) = \mathbb{N}^*$, $\mu^n = \sum_{k=0}^n C_n^k \delta_k$, et $F(\mu^{*n}) = \{B(n, p); p \in]0, 1[\}$, où $B(n, p)$ est la loi binomiale.

2) Loi gamma. On prend $E = \mathbb{R}$ et $\mu(dx) = \mathbf{1}_{]0, \infty[}(x) dx$ Alors $\Lambda(\mu) =]0, \infty[$

$$\mu_\lambda(dx) = \frac{x^{\lambda-1}}{\Gamma(\lambda)} \mathbf{1}_{]0, \infty[}(x) dx$$

et $F(\mu_\lambda) =$

.....

Expliquons maintenant ce qu'est une famille exponentielle générale. C'est la donnée de deux objets : d'abord d'un espace mesuré $(\Omega, \mathcal{A}, \nu)$ où la mesure ν n'est pas nécessairement bornée ; ensuite d'une application mesurable t de Ω dans un espace vectoriel de dimension finie E tel que l'image $\mu = t \circ \nu$ dans E de la mesure ν par l'application t soit un élément de $\mathcal{M}(E)$. Dans ces conditions, l'ensemble $F(t, \nu)$ des probabilités sur Ω qui sont de la forme

$$P(\theta, t, \nu)(d\omega) = e^{\langle \theta, t(\omega) \rangle - k_\mu(\theta)} \nu(d\omega)$$

pour des $\theta \in \Theta(\mu)$, est appelé la *famille exponentielle générale* engendrée par (t, ν) . La famille exponentielle naturelle $F(\mu)$, faite des images des $P(\theta, t, \nu)(d\omega)$ par $\omega \mapsto t(\omega)$, est appelée la famille exponentielle naturelle *associée* à $F(t, \nu)$. Pour un θ fixé, alors $t(\omega)$ est une variable aléatoire de loi $P(\theta, \mu)$ et est appelée *la statistique exhaustive associée*.

Exemples.

- 1) La loi normale de moyenne et de covariance inconnues

.....

COMMENT CALCULER PRATIQUEMENT LA FONCTION VARIANCE D'UNE FAMILLE EXPONENTIELLE NATURELLE ? Si $\mu \in \mathcal{M}(E)$, si $M_F = k'_\mu(\Theta(\mu)) \subset E^*$ est le domaine des moyennes de la famille exponentielle naturelle $F = F(\mu)$, et si ψ_μ est la bijection réciproque de M_F dans $\Theta(\mu)$ de $\theta \mapsto k'_\mu(\theta)$ alors

$$V_F(m) = k''_\mu(\psi_\mu(m)) = (\psi'_\mu(m))^{-1}$$

est l'opérateur de covariance de la probabilité $P(\psi_\mu(m), \mu) = P(m, F)$. Notons $\mathcal{Q}(E^*) = L_s(E^*, E)$ l'espace des endomorphismes symétriques de E^* vers E (naturellement identifié à l'espace des formes quadratiques sur E^*). Alors l'application de M_F vers $L_s(E^*, E)$ définie par $m \mapsto V_F(m)$ est appelée la *fonction variance* de F . Les lignes qui suivent cherchent à donner quelques méthodes pratiques pour calculer V_F dans le cas où E n'est pas \mathbb{R} mais un espace vectoriel.

CAS OÙ E EST \mathbb{R}^n . Si on identifie le dual de \mathbb{R}^n à lui même en munissant \mathbb{R}^n de la structure euclidienne canonique, alors on identifiera $k'_\mu(\theta)$ à la suite des dérivées partielles

$$\left(\frac{\partial}{\partial \theta_1} k_\mu(\theta), \dots, \frac{\partial}{\partial \theta_n} k_\mu(\theta) \right) = (m_1, \dots, m_n).$$

On résout ensuite en θ ce système en fonction de m pour obtenir $\theta = \psi_\mu(m)$ sous la forme

$$(\theta_1, \dots, \theta_n) = (\psi_\mu(m)_1, \dots, \psi_\mu(m)_n).$$

A ce point, il y a deux voies :

- Ou bien on décide d'utiliser $V_F(m) = (\psi'_\mu(m))^{-1}$. Dans ce cas on calcule $\psi'_\mu(m)$, qui est donc représenté par la matrice carrée symétrique $(\frac{\partial}{\partial m_i} \psi_\mu(m)_j)$ et (aspect pénible de cette méthode) on calcule l'inverse $V_F(m)$ de cette matrice carrée.
- Ou bien on décide d'utiliser $V_F(m) = k''_\mu(\psi_\mu(m))$. Dans ce cas on calcule la matrice hessienne

$$k''_\mu(\theta) = \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} k_\mu(\theta) \right]$$

et on substitue ensuite $\psi_\mu(m)$ à θ dans cette expression. Dans la pratique, cette deuxième méthode est souvent plus rapide. En général, le pénible

calcul d'inversion de $\theta = \psi_\mu(m)$ a conduit à introduire des fonctions auxiliaires $f_v(\theta) = g_v(m) = f_v(\psi_\mu(m))$ qui apparaissent souvent explicitement dans l'expression de $k_\mu''(\theta)$ et le calcul n'est pas si décourageant après tout.

CAS OÙ μ EST CONCENTRÉE SUR UN CÔNE CONVEXE. Si la famille exponentielle est réelle, il arrive souvent qu'elle soit concentrée sur $[0, \infty)$ et qu'alors $\Theta(\mu)$ soit une demi droite $(-\infty, a)$. C'est une source d'erreurs de trainer un paramètre systématiquement négatif. A plusieurs dimensions, cette situation se reproduit si μ est concentrée sur un cône convexe saillant C de dual $C^* = \{y \in E^*; \langle y, x \rangle \geq 0\}$. Dans ce cas $\Theta(\mu)$ est souvent un translaté de $-C^*$ et le signe moins, maudit, réapparaît. On diminue les chances d'erreur en revenant à la transformée de Laplace des physiciens $S_\mu(y) = \int_0^\infty e^{-yx} \mu(dx)$ plutôt que $L_\mu(\theta) = \int_0^\infty e^{\theta x} \mu(dx)$.

Dans E on définit donc

$$S_\mu(y) = \int_E e^{-\langle y, x \rangle} \mu(dx)$$

où $y \in E^*$. On définit aussi $s_\mu(y) = -\log S_\mu(y)$ sur $-\Theta(\mu)$ et on note par ρ_μ la fonction réciproque de M_F vers $-\Theta(\mu)$ de $y \mapsto m = s'_\mu(y)$.

D'où les relations avec les formes classiques tirées de L_μ :

$$S_\mu(-\theta) = L_\mu(\theta); \quad -s_\mu(-\theta) = k_\mu(\theta); \quad s'_\mu(-\theta) = k'_\mu(\theta); \quad -\rho_\mu(m) = \psi_\mu(m).$$

Ceci conduit à l'importante relation

$$V_F(m) = -s''_\mu(\rho_\mu(m)). \quad (4.13)$$

A partir de celle ci, les méthodes pour $V_F(m) = k''_\mu(\psi_\mu(m))$ s'appliquent.

CAS OÙ LA DUALITÉ ENTRE E ET E^* EST PEU CLASSIQUE. Il arrive souvent que E et E^* soient clairement identifiés à \mathbb{R}^n mais que leur mise en dualité se fasse par une forme bilinéaire $\varphi(x, y)$ qui n'est pas $x_1y_1 + \dots + x_ny_n$. Dans ce cas on a

$$L_\mu(\theta) = e^{k_\mu(\theta)} = \int_E e^{\varphi(\theta, x)} \mu(dx), \quad S_\mu(y) = e^{-s_\mu(y)} = \int_E e^{-\varphi(y, x)} \mu(dx).$$

C'est le cas quand il est question d'espace hyperbolique, avec $\varphi(x, y) = -x_1y_1 + x_2y_2 + \dots + x_ny_n$, ou quand E est l'espace des matrices (r, r) symétriques réelles $x = (x_{ij})$ avec $\varphi(x, y) = \text{tr } xy$: dans ce cas E s'identifie à \mathbb{R}^n avec $n = r(r+1)/2$ mais pour $i < j$ le coefficient de $x_{ij}y_{ij}$ est 2.

Le calcul de $\psi_\mu(m)$ ou de $\rho_\mu(m)$ à partir des dérivées partielles de k_μ ou de s_μ doit se faire ainsi : on observe que par définition de la différentielle on a pour tout h dans E assez petit

$$\begin{aligned}
k_\mu(\theta + h) - k_\mu(\theta) &= \sum_{k=1}^n \frac{\partial}{\partial \theta_k} k_\mu(\theta) h_k + o(h) \\
&= k'_\mu(\theta)(h) + o(h) \\
&= \varphi(k'_\mu(\theta), h) + o(h) = \varphi(m, h) + o(h)
\end{aligned}$$

Pour calculer m en fonction des dérivées partielles de k_μ il faut tirer parti de l'égalité $\sum_{k=1}^n \frac{\partial}{\partial \theta_k} k_\mu(\theta) h_k = \varphi(m, h)$ valable pour tout $h \in E$. On introduit alors la matrice symétrique Q inversible d'ordre n telle que $\varphi(x, y) = xQy^T$ (on a écrit les éléments de \mathbb{R}^n comme des vecteurs lignes) et on obtient

$$k'_\mu(\theta) = (m_1, \dots, m_n) = \left(\frac{\partial}{\partial \theta_1} k_\mu(\theta), \dots, \frac{\partial}{\partial \theta_n} k_\mu(\theta) \right) Q^{-1}. \quad (4.14)$$

$$s'_\mu(y) = (m_1, \dots, m_n) = \left(\frac{\partial}{\partial y_1} s_\mu(y), \dots, \frac{\partial}{\partial y_n} s_\mu(y) \right) Q^{-1}. \quad (4.15)$$

Le calcul de $\theta = \psi_\mu(m)$ ou de $y = \rho_\mu(m)$ devra alors se faire à partir des équations 4.14 ou 4.15. Voici enfin l'expression de la matrice représentative dans \mathbb{R}^n de la fonction variance

$$V_F(m) = Q^{-1} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} k_\mu(\psi(m)) \right] Q^{-1} = -Q^{-1} \left[\frac{\partial^2}{\partial y_i \partial y_j} s_\mu(\rho(m)) \right] Q^{-1}. \quad (4.16)$$

On trouvera un exemple de cette technique dans le calcul de la fonction variance des lois de Wishart associées au graphe de la marguerite à la section 11.

5 Restriction du paramètre canonique à un sous espace affine.

Une situation qui se présente souvent en statistique, en particulier dans les modèles graphiques, est celle d'une famille exponentielle (qu'on peut prendre naturelle sans perte réelle de généralité), donc de la forme $F(\mu)$, à laquelle on impose ensuite au paramètre canonique d'appartenir à un sous espace affine H de E^* . Ceci se retrouve dans une grande variété de situations. Avant de les mentionner, observons que cela pose deux problèmes théoriques :

- Comment tester l'hypothèse que le paramètre inconnu θ appartient à H ?
- Sachant que le paramètre inconnu θ est dans H , comment l'estimer par maximum de vraisemblance? En particulier, en quoi les outils de la famille exponentielle complète peuvent ils être utiles?

Nous allons nous intéresser d'abord au second problème. Pour commencer, nous allons faire deux hypothèses simplificatrices, qui ne changent rien à la généralité, mais choqueront les puristes. On va supposer d'abord que H est un sous espace vectoriel (ce qui revient à changer la mesure génératrice $\mu(dx)$ en $\exp(\langle \theta_0, x \rangle) \mu(dx)$), ensuite que E a une structure euclidienne (ce qui permet d'identifier E et E^* , de considérer H comme un sous espace de E et de faire intervenir son espace orthogonal H^\perp . Si ce n'était pas euclidien, cet objet H^\perp devrait être remplacé par $H^0 = \{x \in E; \langle \theta, x \rangle = 0 \forall \theta \in H\}$ et on devrait considérer l'espace quotient E/H^0 à la place de $H \subset E$ pour plaire à Bourbaki.

Soit donc E euclidien, $\mu \in \mathcal{M}(E)$ soit H un sous espace vectoriel de E . On suppose que $\Theta_{\sharp} = \Theta(\mu) \cap H$ est non vide. Si $F = F(\mu)$, on veut étudier le modèle

$$F_H = \{P(\theta, \mu); \theta \in \Theta_{\sharp}\}$$

du point de vue des familles exponentielles. La proposition suivante clarifie.

Proposition 5.1. Avec les notations et hypothèses ci dessus, notons par t la projection orthogonale de E sur H . Alors :

1. Le modèle F_H est une famille exponentielle *générale* gouvernée par (μ, t) .
2. La mesure projetée $t \circ \mu = \mu_{\sharp}$ sur H existe et engendre la famille exponentielle associée $F(\mu_{\sharp}) = F_{\sharp}$.
3. La transformée de Laplace de μ_{\sharp} est la restriction de L_{μ} à H .
4. Pour tout q dans le domaine des moyennes $M_{\sharp} \subset H$ de F_{\sharp} , il existe un unique $u(q)$ dans H^\perp tel que

$$\psi_{\mu_{\sharp}}(q) = \psi_{\mu}(q + u(q)), \quad (5.17)$$

en d'autres termes, tel que l'équation $k'_{\mu}(\theta) = q + u(q)$ ait une solution dans Θ_{\sharp} .

Démonstration.

Si $\theta \in \Theta_{\sharp}$, alors

$$\exp(\langle \theta, x \rangle - k_{\mu}(\theta)) \mu(dx) \stackrel{(1)}{=} \exp(\langle t(\theta), x \rangle - k_{\mu}(\theta)) \mu(dx) \stackrel{(2)}{=} \exp(\langle \theta, t(x) \rangle - k_{\mu}(\theta)) \mu(dx)$$

où (1) vient de $t(\theta) = \theta$ et (2) vient du fait que $t^* = t$. Ceci montre 1,2 et 3. En différentiant $k_{\mu}(t(\theta)) = k_{\mu_{\sharp}}(\theta)$ dans H l'égalité, on obtient

$$t(k'_{\mu}(t(\theta))) = k'_{\mu_{\sharp}}(\theta).$$

Donc si $q \in M_{\sharp} \subset H$ on sait qu'il existe un et un seul $\theta = \psi_{\mu_{\sharp}}(q) \in \Theta_{\sharp}$ tel que $q = k'_{\mu_{\sharp}}(\theta)$. Donc $q = t(k'_{\mu}(t(\psi_{\mu_{\sharp}}(q))))$, donc il existe $u(q) \in H^\perp$ tel que $q + u_q = k'_{\mu}(t(\psi_{\mu_{\sharp}}(q)))$. Puisque $\theta \mapsto k'_{\mu}(\theta)$ est injectif, $u(q)$ est unique, et 5.17 est montré.

Le premier problème est alors résolu généralement à l'aide d'un objet appelé déviance. Pour la famille exponentielle initiale $F(\mu)$ et pour les observations (X_1, \dots, X_n) de moyenne empirique \bar{X}_n , telles que \bar{X}_n est dans le domaine des moyennes, on considère le *logarithme de la vraisemblance*, c'est à dire

$$\langle \psi_\mu(\bar{X}_n), \bar{X}_n \rangle - k_\mu(\psi_\mu(\bar{X}_n)). \quad (5.18)$$

Remarquons que pour une fois la fonction t qui définit une famille exponentielle générale est linéaire et on peut écrire

$$t(\bar{X}_n) = \frac{1}{n}(t(X_1) + \dots + t(X_n)).$$

Donc le log de la vraisemblance pour la famille F_H ou pour la famille F_\sharp est

$$\langle \psi_{\mu_\sharp}(t(\bar{X}_n)), t(\bar{X}_n) \rangle - k_{\mu_\sharp}(\psi_{\mu_\sharp}(t(\bar{X}_n))),$$

dont une présentation un peu plus agréable, grâce à la proposition 5.1, est :

$$\langle \psi_\mu(t(\bar{X}_n) + u(t(\bar{X}_n))), \bar{X}_n \rangle - k_\mu(\psi_\mu(t(\bar{X}_n) + u(t(\bar{X}_n)))), \quad (5.19)$$

On appelle alors *déviance* de F_H par rapport à $F(\mu)$ la différence de ces deux log de vraisemblance 5.18) et 5.19) multipliée par 2, c'est à dire le nombre

$$2[\langle \psi_\mu(\bar{X}_n) - \psi_\mu(t(\bar{X}_n) + u(t(\bar{X}_n))), \bar{X}_n \rangle - k_\mu(\psi_\mu(\bar{X}_n)) + k_\mu(\psi_\mu(t(\bar{X}_n) + u(t(\bar{X}_n))))]. \quad (5.20)$$

La multiplication par 2 est là, comme on s'en doute, pour des raisons historiques : si on applique la théorie précédente au cas où F_μ est la famille des lois normales dans \mathbb{R}^r de moyenne inconnue et de covariance I_n la déviance n'est autre que le carré de la distance de \bar{X}_n au sous espace H .

Il est clair que la déviance en général est toujours un nombre positif. En effet si $x \in M_{F(\mu)}$, alors $\theta \mapsto \langle \theta, x \rangle - k_\mu(\theta)$ est maximisée en $\theta = \psi_\mu(x)$. Appliquons cette remarque à $x = \bar{X}_n$ et à $\theta = \psi_\mu(t(\bar{X}_n) + u(t(\bar{X}_n)))$, cela montre que la déviance est positive. Cette statistique est alors utilisée comme test du fait que θ appartient ou non à H . Si la déviance est grande, on rejettera l'hypothèse.

Les formules précédentes montre l'importance du calcul de la fonction $q \mapsto u(q)$ de M_\sharp dans H^\perp . Voyons en détail un exemple plus intéressant que les lois normales

On considère l'espace E des matrices symétriques réelles $(n+1, n+1)$, muni de la structure euclidienne $\langle x, y \rangle = \text{tr}(xy)$. Notons par E_+ le cône ouvert des matrices symétriques définies positives et par \bar{E}_+ le cône fermé des matrices symétriques positives. Soit H le sous espace de E des matrices de la forme

$$q = \begin{bmatrix} a & b_1 & \dots & b_n \\ b_1 & c_1 & \dots & 0 \\ \dots & 0 & \dots & 0 \\ b_n & 0 & \dots & c_n \end{bmatrix}. \quad (5.21)$$

Prenons alors pour famille exponentielle sur E la famille des lois de Wishart de paramètre de forme p engendrée par la mesure μ_p dont la transformée de Laplace est $\det(-\theta)^{-p}$ et est définie sur l'ensemble $\Theta(\mu_p) = -E_+$. Il est connu que dans ce cas, $\theta \mapsto k_{\mu_p}(\theta) = -p \log \det(-\theta)$ a pour gradient $p(-\theta)^{-1}$.

.....

$$u_q = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & b_1 b_n / a \\ \dots & 0 & \dots & 0 \\ 0 & b_n b_1 / a & \dots & 0 \end{bmatrix}. \quad (5.22)$$

(et donc $u(q)$ est défini par $u_{ij} = -b_i b_j / a$ pour $1 \leq i < j \leq n$),

$$(q + u(q))^{-1} = \begin{bmatrix} \frac{-(n-1)}{a} + \sum_{i=1}^n \frac{c_i}{ac_i - b_i^2} & \frac{-b_1}{ac_1 - b_1^2} & \dots & \frac{-b_n}{ac_n - b_n^2} \\ \frac{-b_1}{ac_1 - b_1^2} & \frac{a}{ac_1 - b_1^2} & \dots & 0 \\ \dots & 0 & \dots & 0 \\ \frac{-b_n}{ac_n - b_n^2} & 0 & \dots & \frac{a}{ac_n - b_n^2} \end{bmatrix}. \quad (5.23)$$

6 Les graphes triangulés sont les graphes décomposables

Étant donné un graphe $G = (V, E)$ (fini, et non orienté, comme ci dessus), on va noter $a \sim b$ lorsque a et b sont des sommets du graphe et que $\{a, b\}$ est une arête. On dit aussi que a et b sont *adjacents*, et que b est un *voisin* de a .. On appelle *chemin de longueur* $n \geq 0$ toute suite (v_0, v_1, \dots, v_n) de longueur $n + 1$ telle que $v_{j-1} \sim v_j$ pour tout $j = 1, 2, \dots, n$. Les sommets v_0 et v_n sont appelés les *extrémités* du chemin. Notez qu'un chemin de longueur 0 existe et est réduit à un sommet. Un *cycle* est un chemin (v_0, v_1, \dots, v_n) de longueur $n \geq 3$ tel que (v_1, \dots, v_n) soient *distincts* et tel que $v_n = v_0$. Le graphe G est alors dit *triangulé* si pour tout cycle de G (v_0, v_1, \dots, v_n) de longueur $n \geq 4$ alors il existe i et j avec $1 \leq i < j \leq n$ tels que $1 < j - i < n - 1$, et $v_i \sim v_j$. L'arête $\{v_i, v_j\}$ est alors appelée une *corde* du cycle. Si $U \subset V$, soit E_U l'ensemble des arêtes $\{a, b\}$ de E telles que a et b soient dans U . Le graphe $G_U = (U, E_U)$ est alors appelé *graphe induit* de G sur U . C'est un sous graphe de G , mais attention, comme déjà dit en section 3, tous les sous graphes ne sont pas des graphes induits. Notez que si un graphe est triangulé, tout graphe induit est aussi triangulé.

Il faut donner encore d'autres définitions. Si a et b sont dans V il est facile de voir que la relation $a R b$ définie par le fait qu'il existe un chemin d'origine a et d'extrémité b est une relation d'équivalence. Ses classes sont appelées les composantes connexes de G . S'il n'y a qu'une seule classe, on dit que G est *connexe*.

Soit maintenant a et b non adjacents dans V . Un *séparateur* de a et b est une partie S de V telle que tout chemin (v_0, v_1, \dots, v_n) d'extrémités $v_0 = a$ et

$v_n = b$ est tel qu'il existe j avec $1 \leq j \leq n - 1$ tel que $v_j \in S$. Bien entendu, tout ensemble plus grand que S est encore un séparateur, et comme G est un graphe fini, la notion de *séparateur minimal* est bien définie. Notons que ni a ni b ne peuvent être dans un séparateur minimal de a et b . Plus généralement, si A et B sont des parties disjointes de V , on dit que la partie S de V est un séparateur de A et B si S est un séparateur de a et b pour tout couple (a, b) de $A \times B$.

Voici d'abord une première caractérisation non triviale des graphes triangulés :

Théorème 6.1. Soit $G = (V, E)$ un graphe fini non orienté. Alors les deux affirmations suivantes sont équivalentes :

1. G est triangulé.
2. Pour tout couple de sommets a et b , tout séparateur minimal S de a et b est tel que le graphe induit G_S est complet.

Démonstration.

$2 \Rightarrow 1$: Soit (v_0, v_1, \dots, v_n) un cycle de longueur $n \geq 4$ et soit S un séparateur minimal de v_0 et de v_2 . Alors S contient v_1 et un v_j , avec $2 < j < n$. Donc $\{v_1, v_j\}$ est une corde du cycle. Le graphe est donc bien triangulé.

$1 \Rightarrow 2$: Soit S un séparateur minimal de a et b . Soit A et B les composantes connexes du graphe induit $G_{E \setminus S}$ contenant a et b respectivement. Pour tout s de S il existe certainement un a_s dans A adjacent à s : sinon $S \setminus \{s\}$ serait aussi un séparateur de a et b , ce qui contredit la minimalité de S . De même il existe b_s dans B adjacent à s . Montrons alors que si s et s' sont dans S , alors ils sont adjacents. Puisque A est connexe, il existe un chemin d'extrémités a_s et $a_{s'}$ dans G_A , disons $(a_s, a_1, \dots, a_{n-1}, a_{s'})$. Notons que sa longueur n peut être 0. Choisissons ce chemin pour que cette longueur n soit minimale (plus précisément, s et s' étant fixés dans S , on choisit même a_s et $a_{s'}$ pour que cette longueur soit minimale). De même, il existe un chemin $(b_{s'}, b_1, \dots, b_{m-1}, b_s)$ dans G_B , avec m minimal. Donc

$$(s, a_s, a_1, \dots, a_{n-1}, a_{s'}, s', b_{s'}, b_1, \dots, b_{m-1}, b_s, s)$$

est un cycle de longueur $n + m + 4$. Il y a donc une corde. Celle ci ne peut être de la forme $\{a_i, a_j\}$ (convention : $a_n = a_{s'}$) : cela contredirait la minimalité de n . De même la corde $\{b_i, b_j\}$ est exclue. Quant à la corde $\{a_i, b_j\}$, elle est exclue par définition d'un séparateur. Les cordes du genre $\{a_i, s\}$, $\{a_i, s'\}$, $\{b_i, s\}$, $\{b_i, s'\}$ sont exclues par minimalité de n ou m . La seule corde possible est $\{s, s'\}$, ce qui achève la démonstration.

Ce théorème nous conduit à une propriété importante des graphes triangulés :

Lemme 6.2. Soit $G = (V, E)$ un graphe triangulé. Alors, si G n'est pas complet, il existe une partition (A, S, B) de V qui soit propre (c'est à dire

formée d'ensembles non vides), telle que S soit un séparateur de A et B et telle que le graphe induit G_S soit complet.

Démonstration. Si G n'est pas complet, soit a et b deux sommets non adjacents, et soit S un séparateur minimal de a et b . Soit A la composante connexe contenant a du graphe induit $G_{V \setminus S}$ et soit $B = V \setminus (S \cup A)$. L'ensemble B contient b et n'est donc pas vide. Par le théorème 6.1, S est complet. Reste à voir que S sépare A et B . S'il existait $(a', b') \in A \times B$ et un chemin C d'extrémités a' et b' qui évite S on distingue deux cas : ou bien b' est dans la composante connexe de $G_{V \setminus S}$ contenant b : ce chemin C permettrait de fabriquer un chemin de a à b qui évite S : c'est impossible. Ou bien b' n'est pas dans la composante connexe de $G_{V \setminus S}$ qui contient b . Comme b' est dans la composante connexe de G qui contient a , et non dans la composante connexe de $G_{V \setminus S}$, cela entraîne que S est un séparateur de a et b' et il faut bien que le chemin C coupe S , d'où la contradiction. Le lemme est montré.

Le lemme 6.2 nous conduit à une caractérisation des graphes triangulés, qui commence par la subtile définition de graphe décomposable. Si $G = (V, S)$ est un graphe quelconque (fini et non orienté bien sûr), appelons *décomposition* de G la donnée d'une partition propre (A, S, B) de V telle que G_S soit complet et soit un séparateur de A et B . La Proposition 6.2 dit que si G est triangulé non complet, il admet une décomposition. Voici alors la *réursive* définition suivante. On dit que le graphe G est *décomposable* si il est ou bien complet, ou bien tel qu'il admette une décomposition (A, S, B) telle que *de plus* les graphes induits $G_{A \cup S}$ et $G_{S \cup B}$ soient déjà décomposables.

Cette définition en est vraiment une, bien qu'abstraite : si on connaît tous les graphes décomposables à n sommets (et c'est le cas pour $n = 1 \dots$) alors on connaît les graphes à $n + 1$ sommets. En fait :

Théorème 6.3. Un graphe est décomposable si et seulement si il est triangulé.

Démonstration. Soit $G = (V, E)$. Nous montrons l'équivalence par récurrence sur $n = |V|$. C'est trivial pour $n = 1$. Supposons l'équivalence vraie pour tous les V de taille $< n$ et montrons la pour n . Sans perte de généralité on peut supposer que G n'est pas complet.

\Leftarrow : Puisque G est triangulé non complet, le lemme 6.2 montre qu'il existe une décomposition (A, S, B) . Comme $G_{A \cup S}$ et $G_{S \cup B}$ sont triangulés, l'hypothèse de récurrence entraîne qu'ils sont décomposables, et donc G est décomposable.

\Rightarrow : Puisque G est décomposable non complet, par définition il existe une décomposition (A, S, B) . Les graphes $G_{A \cup S}$ et $G_{S \cup B}$ sont décomposables et donc triangulés par l'hypothèse de récurrence. Si C est un cycle de G , alors ou bien il est dans $G_{A \cup S}$, qui est triangulé, et il a une corde. Ou bien il est dans $G_{B \cup S}$, qui est triangulé, et il a une corde. Ou bien C comprend un point a de A et un point b de B , avec $C = (a, \dots, b, \dots, a)$. Le cycle se fractionne en un chemin de a à b , qui coupe donc S en un point s , et un chemin de b à a en un point

s' . Comme S est complet, $\{s, s'\}$ est une corde de C et G est bien triangulé.

7 Enumérations parfaites dans les graphes triangulés

Énonçons un lemme important concernant les graphes triangulés. Pour cela, on dit qu'un sommet s d'un graphe $G = (V, E)$ est *simplicial* si l'ensemble des voisins de s , c'est à dire l'ensemble $U = \{v \in V ; v \sim s\}$, est tel que le graphe induit G_U est complet.

Lemme 7.1. (Lemme de Dirac) Soit $G = (V, E)$ un graphe triangulé *non complet*. Alors il existe deux sommets simpliciaux non adjacents.

Démonstration. On procède par récurrence sur la taille n de l'ensemble V des sommets. C'est évident si $n = 1$ ou 2 . Supposons le résultat vrai jusqu'à $n - 1$ et montrons le pour n . Puisque G n'est pas complet, il existe deux sommets non adjacents a et b auxquels on applique la Proposition 6.1 : soit S un séparateur minimal de a et b , A et B les composantes connexes contenant respectivement a et b dans $G_{V \setminus S}$.

Ce serait une erreur de penser que parce que l'hypothèse de récurrence garantit que G_A et G_B ont des sommets simpliciaux, la récurrence est étendue. La raison est qu'un sommet simplicial pour G_A ne l'est pas nécessairement pour G . Toutefois, un sommet s simplicial pour $G_{A \cup S}$ est simplicial pour G . En effet si $s \in A \cup S$, et G_S étant complet d'après la Proposition 6.1, on a $\{v \in V ; v \sim s\} = \{v \in A \cup S ; v \sim s\}$.

Alors : ou bien $G_{A \cup S}$ et $G_{B \cup S}$ sont complets, et a et b sont simpliciaux ; ou bien un des deux, disons $G_{A \cup S}$, n'est pas complet. L'hypothèse de récurrence garantit alors l'existence de deux points simpliciaux non adjacents dans $A \cup S$. Ils sont simpliciaux au sens $G_{A \cup S}$, mais nous avons vu que cela entraînait que c'est aussi au sens G . La démonstration du lemme de Dirac est complète.

Voici un troisième groupe de caractérisation des graphes triangulés :

Théorème 7.2. Soit $G = (V, E)$ un graphe fini connexe non orienté à n sommets, et soit \mathcal{C} l'ensemble de ses k cliques maximales. Alors les trois affirmations suivantes sont équivalentes :

1. G est triangulé.
2. Il existe une bijection $j \mapsto v_j$ de $\{1, \dots, n\}$ et de V telle que pour tout $j > 1$ le sommet v_j soit simplicial par rapport au graphe induit $G_{\{v_1, \dots, v_j\}}$.
3. Il existe une bijection $q \mapsto C_q$ de $\{1, \dots, k\}$ et de \mathcal{C} telle que si on note $H_q = C_1 \cup \dots \cup C_q$ et $S_q = C_q \cap H_{q-1}$, (avec $H_0 = \emptyset$), alors pour tout $q = 2, 3, \dots, k$ il existe $p \leq q - 1$ tel que $S_q \subset C_p$.

De plus, au 2) v_n peut être pris arbitrairement parmi les sommets simpliciaux de G ; au 3) C_1 peut être pris arbitrairement parmi \mathcal{C} .

Remarques. 1) Une énumération des *sommets* satisfaisant le 2) du théorème 7.1 est dite *parfaite*. Une énumération des *cliques* maximales satisfaisant le 3) du théorème 7.1 est dite *parfaite*. On parle aussi d'ordres parfaits des sommets et des cliques.

2) Les $S_q = H_{q-1} \cap C_q$ sont appelés les séparateurs de cette énumération parfaite. On verra au Théorème 8.3 que leurs valeurs et leurs multiplicités ne dépendent pas de l'énumération parfaite qu'on a choisie.

3) Avec la notion d'énumération parfaite des sommets, nous touchons à l'origine historique de l'intérêt pour les graphes triangulés : à la fin des années 50, le développement des ordinateurs et du calcul numérique fait considérer de grands systèmes linéaires de Cramer $AX = Y$ à n inconnues et n équations, où la matrice A comprend beaucoup de zéros, une situation qu'on veut exploiter au maximum si on applique la méthode de Gauss, qui consiste simplement à prendre une inconnue x_i , à l'exprimer en fonction des autres x_j et de y_i et à remplacer dans le système $AX = Y$ et obtenir un système de Cramer d'ordre $n - 1$, disons $A_1 X_1 = Y_1$. Pour être plus précis et en particulier éviter des annulations de déterminants, supposons que $A = (a_{ij})$ soit symétrique définie positive. Examinons le passage de la matrice A à la matrice A_1 si on a choisi d'éliminer l'inconnue x_n . Puisque

$$x_n = \frac{y_n}{a_{nn}} - \frac{1}{a_{nn}} \sum_{j=1}^{n-1} a_{jn} x_j,$$

alors le coefficient (i, j) dans la matrice $A_1 = (b_{ij})$ d'ordre $n - 1$ est

$$b_{ij} = a_{ij} - \frac{a_{in} a_{jn}}{a_{nn}}.$$

Notons alors par E l'ensemble des couples $\{i, j\}$ de $V = \{1, \dots, n\}$ tels que $a_{ij} \neq 0$ (avec $i \neq j$). Donc, une condition suffisante pour que $b_{ij} = 0$ est que $a_{ij} = 0$ et que a_{in} ou $a_{jn} = 0$. Mais si n est un sommet simplicial de $G = (V, E)$, cela entraîne que si $a_{ij} = 0$ alors a_{in} ou $a_{jn} = 0$. Donc $a_{ij} = 0$ entraîne $b_{ij} = 0$. En d'autres termes si n est un sommet simplicial alors l'élimination de x_n n'introduit pas de nouveaux coefficients non nuls dans A_1 . Si maintenant le graphe G est triangulé, et que $V = (1, \dots, n)$ soit un ordre parfait, on voit que le processus ci dessus pourra être réitéré jusqu'à la solution complète du système linéaire $AX = Y$ sans jamais avoir détruit les précieux zéros de la matrice initiale A .

Démonstration.

$1 \Rightarrow 2$: On procède par récurrence sur $n = |V|$. C'est trivial pour $n = 1$. Si vrai pour $n - 1$, notons par v_n un sommet simplicial (qui existe, d'après le lemme 7.1). Puisque le graphe induit $G_{V \setminus \{x\}}$ est triangulé, la récurrence est immédiate.

$2 \Rightarrow 1$: Soit $(v_{i_0}, v_{i_1}, \dots, v_{i_k})$ un cycle de longueur $k \geq 4$. Sans perte de généralité on suppose que $i_k = i_0$ est $> i_j$ pour $j = 1, 2, \dots, k - 1$. Comme v_{i_k}

est simplicial et que v_{i_1} et $v_{i_{k-1}}$ sont ses voisins, alors $(v_{i_1}, v_{i_{k-1}})$ est une corde et G est donc triangulé.

$3 \Rightarrow 2$: On utilise le lemme suivant :

Lemme 7.3. Soit $G = (V, E)$ un graphe connexe non orienté, soit C une clique maximale et soit H l'union de toute les cliques maximales de G différentes de C . Alors ou bien $C \subset H$, ou bien $(C \setminus H, C \cap H, H \setminus C)$ est une décomposition, c'est à dire que $C \cap H$ est un séparateur de $C \setminus H$ et $H \setminus C$.

Démonstration. Soit $u \in C \setminus H$ et $v \in H \setminus C$ et montrons que $u \sim v$ est impossible. En effet, si $u \sim v$ soit C' une clique maximale contenant $\{u, v\}$. Puisque $v \in H \setminus C$ la clique C' ne peut être la clique C et donc $u \in H$: contradiction. Soit alors $a \in C \setminus H$ et $b \in H \setminus C$ et soit un chemin de a à b (qui existe, puisque G est connexe). Soit v le premier élément de ce chemin situé dans H et soit u celui qui le précède : d'après la remarque précédente, ou bien u ou bien v est dans $C \cap H$ et le lemme est montré.

Nous démontrons maintenant $3 \Rightarrow 2$: Notons $R_q = C_q \setminus H_{q-1}$. Cet ensemble n'est jamais vide car sinon on aurait $C_q \subset S_q \subset C_p$ et donc $C_q = C_p$. De plus les R_1, \dots, R_k sont deux deux disjoints et forment une partition de V . Soit $r_q > 0$ le nombre d'éléments de R_q et notons $\sigma_q = r_1 + r_2 + \dots + r_q$. Numérotons alors V par $(1, 2, \dots, n)$ de sorte que pour tout $q = 1, \dots, k$ on ait

$$R_q = \{\sigma_{q-1} + 1, \dots, \sigma_q\} = \{i; \sigma_{q-1} < i \leq \sigma_q\}.$$

Il y a bien entendu plusieurs choix possibles pour cela, $r_1! \dots r_k!$ en fait. Montrons que cette énumération sur V est parfaite. Soit i un sommet quelconque. Il existe un unique q tel que $i \in R_q$. Les sommets adjacents à i sont de trois sortes

1. Tous les sommets de C_q sauf i ,
2. certains sommets de $V \setminus H_q$
3. certains sommets de $H_{q-1} \setminus C_q$.

En fait la catégorie 3 est vide : on le voit en appliquant le lemme 7.3 au graphe induit par H_q et à la clique C_q . Les sommets de la catégorie 2 ont des numéros $> \sigma_q \geq i$. Donc les sommets adjacents à i et situés dans $\{1, \dots, i\}$ sont dans C_q et forment donc un ensemble complet ; c'est dire que i est simplicial par rapport à $\{1, \dots, i\}$ et $3 \Rightarrow 2$ est montré.

$2 \Rightarrow 3$: Cette partie est plus difficile.

(a) On suppose que $v_i = i$ pour tout $i = 1, \dots, n$ sans perte de généralité. On définit pour tout i l'ensemble B_i formé par i et ceux des sommets adjacents à i qui sont de plus dans $\{1, 2, \dots, i\}$. Comme i est simplicial par rapport à $\{1, 2, \dots, i\}$ (c'est l'hypothèse du 2)) B_i est *complet*.

(b) Je dis maintenant que toute clique maximale de G est un élément de la suite (B_1, \dots, B_n) . Nous le montrons par récurrence sur n . Supposons que cela

soit vrai pour tout graphe triangulé d'ordre $< n$ muni d'un ordre parfait sur les sommets. Montrons le pour n . D'abord, B_n est une clique maximale, par définition. Ensuite, toute clique maximale du graphe induit sur $\{1, \dots, n-1\}$ apparaît dans la suite (B_1, \dots, B_{n-1}) , d'après l'hypothèse de récurrence. Comme les seules cliques maximales de G sont celles du graphe induit sur $\{1, \dots, n-1\}$ qui ne soient pas des sous ensembles de B_n , plus B_n lui même, la récurrence est étendue.

(c) Nous allons fabriquer une énumération parfaite (C_1, \dots, C_k) des cliques maximales de G à partir de la suite (B_1, \dots, B_n) en supprimant certains éléments et en réordonnant le reste en procédant ainsi

1. Si $i > j$ et $B_i \subset B_j$ on supprime B_i .
2. Après l'épuisement de la procédure 1), soit i fixé. Si j_0 est le plus petit des j tels que $i < j$ et $B_i \subset B_j$ on supprime B_i et on met B_{j_0} à sa place.

Puisque les B_j étaient tous complets, il est clair qu'en appliquant cette procédure il ne reste à la fin que des cliques maximales, et qu'elles sont toutes là, d'après le (b). On a donc ordonné les cliques maximales en une suite (C_1, \dots, C_k)

(d) La tâche qui nous reste est de montrer que cet ordre (C_1, \dots, C_k) satisfait 3). Pour cela introduisons provisoirement une définition : étant donné un graphe $G = (V, E)$, une suite $D = (D_1, D_2, \dots, D_m)$ de parties de V (pas nécessairement distinctes) est dite parfaite si en notant $H_q = D_1 \cup \dots \cup D_q$ alors pour tout $q = 2, \dots, m$ on a $S_q = D_q \cap H_{q-1}$ complet non vide et il existe $p \leq q-1$ tel que $S_q \subset D_p$. Un exemple fondamental de suite parfaite est la suite (B_1, \dots, B_n) définie au a). Cela se voit ainsi : les B_q sont complets et donc les

$$S_q = (B_1 \cup \dots \cup B_{q-1}) \cap B_q = \{1, \dots, q-1\} \cap B_q$$

sont complets. Pour $q > 1$ fixé, soit p le plus grand élément de S_q . Alors $S_q \subset B_p$: en effet si i et j sont dans S_q alors $i \sim q$, $j \sim q$, $p \sim q$ donc $i \sim p$ et donc $i \sim j$.

(e) Observons la chose suivante : supposons que D soit une suite parfaite de longueur m telle qu'il existe $i \neq j$ avec $D_i \subset D_j$. On fabrique alors une nouvelle suite D' de longueur $m-1$ ainsi

1. Si $i > j$ et $D_i \subset D_j$ on supprime D_i .
2. S'il n'existe pas de couple $i > j$ tel que 1) soit vrai, soit j_0 le plus petit des j tels que $i < j$ et $D_i \subset D_j$. Alors on supprime B_i et on met B_{j_0} à sa place.

Le point important est que D' est aussi parfaite. C'est clair dans le premier cas $i > j$. Dans le second cas, observons que nécessairement $S_{j_0} = D_i$. En effet

$$S_{j_0} = D_{j_0} \cap H_{j_0-1} \supset D_{j_0} \cap D_i = D_i.$$

Cependant, comme D est une suite parfaite, il existe $p < j_0$ tel que $S_{j_0} \subset D_p$. Cela entraîne donc $D_i \subset D_p$. Comme le 1) était exclu, on a donc $i \leq p$. Comme j_0 était minimal, on a donc finalement $S_{j_0} = D_i$.

On en déduit que si $S'_i = H_{i-1} \cap D_{j_0}$ alors, puisque $S'_i \subset S_{j_0} = D_i$ on a donc $S'_i = S_i$ et S'_i est donc complet non vide. Il existait par hypothèse $p < i$ tel que $S_i = S'_i \subset D_p$. Ensuite, si $i < q < j_0$ on a

$$S'_q = D_q \cap (H_{q-1} \cup D_{j_0} \cup D_{i+1} \cup \dots \cup D_{q-1}) = D_q \cap (H_{q-1} \cup D_{j_0}) = S_q \cup (D_q \cap (D_{j_0} \setminus D_i)).$$

Cependant, puisque $D_q \cap D_{j_0} \subset S_i = D_{j_0}$ l'ensemble $D_q \cap (D_{j_0} \setminus D_i)$ est donc vide et $S'_q = S_q$ si $i < q < j_0$. Donc S'_q est complet et est inclus dans un D_p pour au moins un $p < q$. Si par hasard ce $p = i$, comme le nouveau en i est $D_{j_0} \supset D_i$ on a $S'_q \subset D_{j_0}$ a fortiori. Quant aux termes de la suite D' d'indice $q \geq j_0$ il est clair qu'ils satisfont $S'_q = S_{q+1}$ et le résultat s'ensuit.

Pour terminer il suffit d'appliquer ce passage de D vers D' de façon répétée en partant de la suite (B_1, \dots, B_n) définie au a) pour arriver à la suite (C_1, \dots, C_k) définie au c). La suite (B_1, \dots, B_n) était parfaite comme on l'a vu au d), la suite (C_1, \dots, C_k) l'est donc aussi.

Exercice 7.1 On décrit un graphe à 7 sommets $\{1, 2, \dots, 7\}$ par ses 4 cliques

$$C_1 = \{4, 3, 2, 1\}, C_2 = \{5, 4, 3, 1\}, C_3 = \{6, 4, 3, 2\}, C_4 = \{7, 6, 2\}.$$

Dessiner ce graphe, dire s'il est triangulé. Si oui, dire si l'ordre des sommets $(1, 2, \dots, 7)$ est parfait et si l'ordre des cliques (C_1, C_2, C_3, C_4) est parfait.

Exercice 7.2 On considère le graphe triangulé de sommets $\{1, 2, \dots, 7\}$ et dont les arêtes sont 12,15,17,23,24,25,27,34,36,37,45,47,67. Donner d'abord une énumération parfaite des *sommets* (v_1, \dots, v_7) , formez ensuite la liste des ensembles B_i décrits dans la démonstration $2 \Rightarrow 3$ du Théorème 7.2 et utilisez l'algorithme décrit là pour construire un ordre parfait des *cliques* du graphe.

8 Séparateurs minimaux et arbres de décomposition

Donnons dès maintenant une définition essentielle :

Définition 8.1. Soit $G = (V, E)$ un graphe triangulé connexe. On dit que $S \subset V$ est un *séparateur minimal* de G si il existe deux éléments a et b de V non adjacents tels que S soit un séparateur minimal de a et b . On appelle *multiplicité* du séparateur minimal S l'entier $\nu(S)$ tel que $1 + \nu(S)$ soit égal au nombre de composantes connexes U de $V \setminus S$ telles que de plus S ne soit pas une clique maximale du graphe induit sur $U \cup S$.

Remarques. 1) Etant donnés les sommets non adjacents a et b , il y a en général plusieurs séparateurs minimaux entre a et b : prendre par exemple le graphe à 4 sommets 1, 2, 3, 4 d'arêtes $\{1, 2\}, \{2, 3\}, \{3, 4\}$: alors $\{2\}$ et $\{3\}$ sont deux séparateurs minimaux de $a = 1$ et $b = 4$.

2) On a vu au théorème 6.1 que un tel séparateur minimal de G est nécessairement complet. Nous allons voir dans cette section que ces séparateurs minimaux de G sont en fait les S_j qui apparaissent dans le Théorème 7.2 quand on considère un ordre parfait entre les cliques maximales. Ils jouent un rôle essentiel avec les cliques maximales dans l'analyse et le maniement des graphes triangulés.

3) Il n'y a pas de séparateur minimal dans un graphe complet.

4) Comme nous l'avons fait en introduisant le terme "cliques maximales", nous éliminons toute ambiguïté par rapport à une littérature très diverse en parlant de "séparateurs minimaux". Il faut savoir que les statisticiens, quand le graphe est triangulé, parlent de séparateur tout court, de clique tout court et de graphe décomposable là où nous parlons de séparateur minimal, de clique maximale et de graphe triangulé.

5) La notion de multiplicité est assez subtile. Considérons par exemple le graphe de sommets $\{1, 2, 3, 4, 5\}$ défini par ses trois cliques maximales $\{1, 2, 3\}$, $\{1, 2, 4\}$, $\{2, 5\}$. On voit facilement qu'il est triangulé et qu'il a deux séparateurs minimaux $\{1, 2\}$ et $\{2\}$. Si on retire $S = \{1, 2\}$ il reste les trois composantes connexes $\{3\}$, $\{4\}$, $\{5\}$. Il est clair que S n'est pas une clique maximale de $\{1, 2, 3\}$ ni de $\{1, 2, 4\}$, mais que c'est une clique maximale de $\{1, 2, 5\}$: la multiplicité de $S = \{1, 2\}$ est donc 1. On voit plus facilement que la multiplicité de $S = \{2\}$ est 1. Un bon moyen de vérifier qu'on ne fait pas d'erreur dans le calcul des multiplicités est le fait que la somme de toutes les multiplicités des séparateurs minimaux doit être égale au nombre de cliques maximales moins un (Corollaire 8.4 ci dessous).

Théorème 8.1. Soit G un graphe triangulé connexe non complet et (C_1, \dots, C_k) un ordre parfait entre les cliques maximales. On note $H_q = C_1 \cup C_2 \cup \dots \cup C_q$ et pour tout $q = 2, \dots, k$ $S_q = C_q \cap H_{q-1}$. Alors pour $q \geq 2$ le graphe induit G_{H_q} est connexe et $(H_{q-1} \setminus S_q, S_q, C_q \setminus S_q)$ est une décomposition de G_{H_q} au sens de la section 6, c'est à dire que c'est une partition propre de H_q et que S_q est un séparateur complet de $(H_{q-1} \setminus S_q, C_q \setminus S_q)$. De plus si $a \in H_{q-1} \setminus S_q$ et $b \in C_q \setminus S_q$ alors S_q est un séparateur minimal de a et b .

Démonstration. Nous montrons d'abord par récurrence sur $k - q$ que G_{H_q} est connexe et que S_q est non vide. C'est vrai pour $q = k$ puisque nous avons supposé G connexe et que cela entraîne trivialement que S_k est non vide. Si c'est vrai pour $k, k - 1, \dots, q + 1$, supposons que G_{H_q} ne soit pas connexe. L'hypothèse de récurrence dit que S_{q+1} est non vide; soit A la composante connexe de H_q qui contient S_q et B son complémentaire dans H_q . Alors $A_1 = A \cup C_q$ et B forment une partition de H_{q+1} . Mais puisque cet ensemble est connexe par l'hypothèse de récurrence, il existe a dans A_1 et b dans B qui sont reliés par un chemin dans $G_{H_{q+1}}$. Comme C_{q+1} est complet, tout chemin empruntant une partie de $C_{q+1} \setminus S_{q+1} = C_{q+1} \setminus A$ peut être remplacé par un chemin dans A , puisque S_{q+1} n'est pas vide. Donc G_{H_q} est connexe. Cela

entraîne que S_q n'est pas vide, car sinon C_q serait une composante connexe de G_{H_q} . La récurrence est étendue.

Montrons ensuite qu'on a une décomposition. Pour q fixé, $C_q \setminus S_q$ n'est pas vide : sinon $S_q \subset C_q$, et comme il existe $p \leq q-1$ tel que $S_q \subset C_p$ alors $C_q \subset C_p$ ce qui contredit la définition de clique maximale. $H_{q-1} \setminus S_q$ n'est pas vide non plus : sinon $C_1 \cup C_2 \cup \dots \cup C_{q-1} \subset C_q$ ce qui contredit la définition de clique maximale. Enfin on a déjà montré par récurrence que S_q n'est pas vide. Il est bien complet, comme partie de l'ensemble complet C_q . Finalement, le Lemme 7.3 appliqué à G_{H_q} montre que S_q est un séparateur de $(H_{q-1} \setminus S_q, C_q \setminus S_q)$.

Montrons enfin que S_q est un séparateur minimal. Soit $a \in H_{q-1} \setminus S_q$ et $b \in C_q \setminus S_q$. On vient de voir par l'emploi du Lemme 7.3 que S_q sépare a et b . Il existe une partie $U \subset S_q$ qui soit un séparateur minimal de a et b . Comme $S_q \subset C_p$ pour un certain $p \leq q-1$ il existe un point $v \in C_p \setminus S_q$ et un chemin de a à v situé dans le graphe connexe $G_{H_{q-1}}$. Si S_{q-1} n'était pas un séparateur minimal, il existerait un $w \in S_q \setminus U$. Comme C_q est complet on a $w \sim b$. De plus w et v sont dans la clique C_p et donc $w \sim v$. Donc, avec le chemin de a à v et la suite v, w, b on a fabriqué un chemin de a à b qui ne passe pas par U et donc U n'est pas un séparateur. Le théorème est complètement montré.

Théorème 8.2. Soit G un graphe triangulé connexe non complet et (C_1, \dots, C_k) un ordre parfait entre les cliques maximales. Soit S un ensemble de sommets. Alors S est un séparateur minimal si et seulement si il existe q dans $\{2, \dots, k\}$ tel que $S = S_q$. En particulier, un séparateur minimal n'est jamais égal à une clique maximale.

Démonstration. On a vu au théorème précédent que S_q est bien un séparateur minimal. Inversement, soit S un séparateur minimal de a et b . Alors a et b sont non adjacents et ne peuvent être dans une même clique maximale. Soit $q(b) = \min\{q; b \in C_q\}$, $q(a) = \min\{q; a \in C_q\}$, supposons sans perte de généralité $q(a) < q(b)$ et montrons que $S = S_j$, pour $2, \dots, q(b)$. D'abord, $b \notin S_{q(b)}$, car sinon il existe un $p \leq q(b) - 1$ tel que $b \in S_{q(b)} \subset C_p$ ce qui contredit la définition de $q(b)$. Donc $b \in R_{q(b)}$ et, selon le Théorème 8.1 précédent, $S_{q(b)}$ est un séparateur minimal de a et b dans $G_{q(b)}$, le graphe induit par $H_{q(b)}$.

Montrons maintenant que l'on a soit $S = S_{q(b)}$, soit $S \cap S_{q(b)} = \emptyset$. On remarque tout d'abord que S étant complet doit être inclus dans une clique C_s , $s = 2, \dots, q$. Si s était tel que $s \geq q(b)$, cela impliquerait que tout chemin entre a et b a une portion qui n'est pas dans $G_{q(b)}$ et cela contredirait le fait que $S_{q(b)}$ est un séparateur minimal de a et b dans $G_{q(b)}$. Donc on a $2 \leq s \leq q(b)$ et $S \subset H_{q(b)}$. On remarque ensuite que si on avait $\emptyset \neq S \cap S_{q(b)} \neq S_{q(b)}$, cela impliquerait que tout chemin entre a et b doit passer et par S et par $S_{q(b)}$, et donc par $S \cap S_{q(b)}$, qui serait alors un séparateur entre a et b strictement inclus dans S et $S_{q(b)}$. Ceci serait évidemment une contradiction au fait que S et $S_{q(b)}$ sont des séparateurs minimaux. On a donc soit $S = S_{q(b)}$, soit $S \cap S_{q(b)} = \emptyset$.

Si $S = S_{q(b)}$, le théorème est prouvé. Si $S \cap S_{q(b)} = \emptyset$, considérons un

chemin $a = v_1, \dots, v_{n-1}, v_n = b$. Sans perte de généralité on peut supposer que $v_{n-1} \in S_{q(b)}$, $v_{n-1} \notin R_{q(b)}$ et que S sépare a de v_{n-1} . Comme l'énumération des cliques est parfaite, cela implique que $v_{n-1} \in C_l$, $2 \leq l \leq q(b) - 1$. Soit $q(v_{n-1}) = \min\{l; v_{n-1} \in C_l\}$. Alors, suivant un raisonnement parallèle à celui suivi pour $q(b)$, on voit que $v_{n-1} \in R_{q(v_{n-1})}$. De plus S et $S_{q(v_{n-1})}$ sont tous deux séparateurs minimaux de a et v_{n-1} . On recommence alors le raisonnement précédent en remplaçant $q(b)$ par $q(v_{n-1})$ pour montrer que soit $S = S_{q(v_{n-1})}$ soit $S \cap S_{q(v_{n-1})} = \emptyset$. Comme l'ensemble des cliques est fini, il existe un q dans $\{2, \dots, q(b)\}$ tel que $S = S_q$.

Finalement, un séparateur minimal S n'est jamais égal à une clique maximale. En effet, il suffit de prendre un ordre parfait (C_1, \dots, C_k) de celles ci. On vient de voir qu'alors il existe $q \geq 2$ tel que $S = S_q$. Donc il existe $p \leq q - 1$ tel que $S \subset C_p$. Si S était une clique maximale alors on aurait à la fois $S = C_p$ et $S = C_q$: contradiction. Le théorème est donc montré.

Encore un peu de vocabulaire de théorie des graphes. Un *arbre* est un graphe (fini et non orienté) qui est connexe et sans cycle. Un sommet d'un arbre qui n'a qu'un voisin est un *bout* de l'arbre. Si on distingue dans l'arbre un sommet r appelé *racine*, c'est un arbre à racine, et l'ensemble de ses sommets est alors structuré en ensemble partiellement ordonné par la relation d'ordre définie ainsi : $s \leq s'$ si l'unique chemin d'extrémités r et s' contient s . Si $s \neq r$, s a un unique *parent* $\gamma(s)$ défini comme l'unique voisin de s tel que $\gamma(s) \leq s$. Les autres voisins de s sont alors appelés ses *descendants*. Un arbre à racine dont tout sommet a 0 ou 2 descendants exactement est appelé un arbre *binaire*.

Soit alors $G = (V, E)$ un graphe triangulé. Un *arbre de décomposition* de G est un arbre binaire à racine dont les sommets sont des parties de V avec les contraintes suivantes

1. La racine est V .
2. Si le sommet V_0 est un bout alors V_0 est complet.
3. Si le sommet V_0 n'est pas un bout, ses deux descendants V_1 et V_2 sont tels que

$$(A, S, B) = (V_1 \setminus V_2, V_1 \cap V_2, V_2 \setminus V_1)$$

est une décomposition de G_{V_0} , $S = V_1 \cap V_2$ est un séparateur minimal de G et S n'est une clique maximale ni de G_{V_1} ni de G_{V_2} .

L'existence d'au moins un arbre de décomposition pour tout graphe triangulé se montre facilement à partir de l'existence d'au moins un ordre parfait entre les cliques (voir le Corollaire 8.4 ci dessous). Un arbre de décomposition n'est pas unique, mais le théorème suivant donne des caractéristiques communes à tous les arbres de décomposition d'un même graphe triangulé.

Théorème 8.3. L'ensemble des bouts d'un arbre de décomposition d'un graphe triangulé connexe est égal à l'ensemble $\{C_1, \dots, C_k\}$ des cliques maximales. En particulier tout arbre de décomposition a le même nombre k de

bouts, et le même nombre $2k - 1$ de sommets. Soit ensuite \mathcal{S} l'ensemble des séparateurs minimaux du graphe. Notons les $k - 1$ sommets de l'arbre de décomposition ayant deux descendants par V_2, \dots, V_k , et à tout j associons $S_j \in \mathcal{S}$ défini comme l'intersection des deux descendants de V_j . Alors pour tout S de \mathcal{S} le nombre de $j = 2, \dots, k$ tels que $S_j = S$ est égal à la multiplicité de S . En particulier cet entier ne dépend pas de l'arbre particulier de décomposition considéré.

Corollaire 8.4. Soit (C_1, \dots, C_k) un ordre parfait de l'ensemble \mathcal{C} des cliques maximales d'un graphe triangulé connexe, soit \mathcal{S} l'ensemble des séparateurs minimaux du graphe, soit $H_q = C_1 \cup \dots \cup C_q$ et, pour $q \geq 2$ soit le séparateur minimal $S_q = H_{q-1} \cap C_q$. Alors pour tout S de \mathcal{S} le nombre de $q = 2, \dots, k$ tels que $S_q = S$ est égal à la multiplicité de S . En particulier cet entier ne dépend pas de l'ordre parfait sur \mathcal{C} considéré.

Démonstration du Corollaire 8.4. A l'ordre parfait sur \mathcal{C} on associe l'arbre de décomposition dont les sommets sont $V = H_k, H_{k-1}, \dots, H_1 = C_1, C_2, \dots, C_k$ où H_q a pour descendants H_{q-1} et C_q .

Démonstration du Théorème 8.3. On procède par récurrence sur le nombre k de cliques maximales. C'est trivial pour $k = 1$ et on suppose le résultat vrai pour tous les graphes triangulés connexes ayant au plus $k - 1$ cliques maximales. Soit alors un arbre de décomposition du graphe triangulé connexe $G = (V, E)$ à $k > 1$ cliques maximales. L'ensemble V n'est donc pas complet et a deux descendants V_1 et V_2 . Soit $S = V_1 \cap V_2$ le séparateur minimal correspondant. Les graphes induits G_{V_i} sont triangulés.

Si C est une clique maximale de G_{V_1} montrons que c'est une clique maximale de G (par définition d'un arbre de décomposition, elle n'est pas égale à S). En effet sinon il existerait une clique maximale $C' \supset C$ de G contenant un point v de $V_2 \setminus S$. Comme C' est complet, v est adjacent à tous les points de C . Comme $C \neq S$, il existerait $u \in C \setminus S$ qui est adjacent à v et donc S ne séparerait pas u et v . Donc finalement les cliques maximales de G_{V_i} sont des cliques maximales de G et l'ensemble \mathcal{C}_1 des cliques maximales de G_{V_1} est disjoint de celui de G_{V_2} , noté \mathcal{C}_2 .

Montrons que $\mathcal{C}_1 \cup \mathcal{C}_2 = \mathcal{C}$. Sinon il existerait une clique maximale C contenant un point $a \in V_1 \setminus S$ et un point $b \in V_2 \setminus S$. Alors a et b seraient adjacents et S ne les séparerait pas : contradiction. L'hypothèse de récurrence entraîne déjà que l'ensemble des bouts de l'arbre de décomposition est égal à \mathcal{C} .

Montrons ensuite que si S_1 est un séparateur minimal de G_{V_1} alors c'est un séparateur minimal de G . En effet, soient a et b dans V_1 séparés par S_1 dans G_{V_1} . S'ils n'étaient pas séparés par S_1 dans G , il existerait un chemin de a à b dans G ne passant par S_1 . Ce chemin utiliserait un point $v \in V_2 \setminus S$. Or a et v sont séparés par S , et b et v sont séparés par S . Donc le chemin de a à b passant par v comprend dans cet ordre a, s, v, s', b avec s et s' dans S . Comme

s et s' sont adjacents on fabriquerait ainsi un chemin de a à b entièrement situé dans V_1 et ne touchant pas S_1 , contradiction. Il est ensuite trivial de constater que S_1 est non seulement un séparateur pour G , mais un séparateur minimal.

Envisageons enfin la question des multiplicités des séparateurs minimaux. Soit p_i le nombre de composantes connexes C de $V_i \setminus S$ telles que S ne soit pas une clique maximale du graphe induit sur $C \cup S$.

Si $p_1 = 1$, cela entraîne que S n'est pas un séparateur minimal de G_{V_1} . En effet si S était séparateur minimal de G_{V_1} il existerait a et b dans V_1 non adjacents séparés par S et donc a et b seraient dans deux composantes connexes différentes de $V_1 \setminus S$. De plus comme S est minimal, ????????

Si $p_1 > 1$ alors de même S est un séparateur minimal de $V_1 \setminus S$ (car cette fois il existe a et b dans V_1 dans deux composantes connexes de G_{V_1}). D'après l'hypothèse de récurrence, la multiplicité $p_1 - 1$ du séparateur S de G_{V_1} est égale au nombre de fois où S apparaît comme intersection de deux descendants dans le sous arbre de sommet V_1 dans l'arbre de décomposition initial. En procédant de même avec V_2 on voit le nombre de fois où S apparaît comme intersection de deux descendants dans l'arbre de décomposition initial est $(p_1 - 1) + (p_2 - 1) + 1 = \nu(S)$. La récurrence est donc complètement étendue et le théorème est montré.

Soit alors $G = (V, E)$ un graphe triangulé connexe et soit \mathcal{C} l'ensemble de ses k cliques maximales. Un *arbre de jonction* de G est un arbre dont les sommets sont tous les éléments de \mathcal{C} avec les contraintes suivantes : si C et C' sont distincts dans \mathcal{C} , soit $C = C_0, C_1, \dots, C_s = C'$ l'unique chemin de C à C' . Alors pour tout $j = 0, \dots, s$ on a

$$C \cap C' \subset C_j. \quad (8.24)$$

Théorème 8.5. Soit un arbre de jonction pour un graphe triangulé connexe. Soit S un séparateur minimal du graphe. Alors le nombre d'arêtes $\{C, C'\}$ de l'arbre telles que $S = C \cap C'$ est égal à la multiplicité de S . En particulier, $C \cap C'$ est un séparateur minimal pour toute arête $\{C, C'\}$.

Démonstration. Soit $\{C, C'\}$ une arête de l'arbre de jonction. Si on supprime cette arête, l'arbre est scindé en deux sous arbres T et T' . Notons par V et V' l'union des cliques maximales sommets de T et T' respectivement. Soit $S = C \cap C' \subset V \cap V'$. On voit d'abord facilement que $S = C \cap C' = V \cap V'$. En effet si $v \in C \cap C'$ il existe des cliques maximales C_0 et C'_0 de T et T' respectivement telles que $v \in C_0 \cap C'_0$. L'unique chemin dans l'arbre de jonction reliant C_0 et C'_0 passe par C et C' et d'après la propriété 8.24) des arbres de jonction on a $v \in C \cap C'$.

Nous montrons ensuite qu'alors $S = C \cap C'$ est un séparateur entre $A = V \setminus S$ et $B = V' \setminus S$. Comme S est complet (intersection de deux cliques maximales), c'est dire que (A, S, B) est une décomposition. Soit donc $a \in A$ et $b \in B$ et un chemin $(a = v_0, v_1, \dots, v_s = b)$ reliant a et b dans le graphe

triangulé. Supposons que ce chemin ne touche pas S . alors nous montrons par récurrence sur i que $v_i \in A$: c'est vrai pour $i = 0$ par définition. Si vrai pour v_{i-1} avec $1 \leq i \leq s$ soit C_0 une clique maximale contenant v_{i-1} et v_i . Puisque $v_{i-1} \in A$ alors $C_0 \subset V$, et donc $v_i \in V$. Puisque le chemin ne touche pas S , alors $v_i \in A$ et la récurrence est étendue. Mais si on l'applique à $i = s$ on obtient la contradiction $b \in A$.

Nous vérifions ensuite que S est minimal. Pour cela nous prenons $(a, b) \in (C \setminus S) \times (C' \setminus S)$ et nous montrons que S est un séparateur minimal de a et b . S'il n'en était pas ainsi, il existerait un séparateur U de a et b contenu dans S et un $w \in S \setminus U$. Mais le chemin (a, w, b) évite U : contradiction.

Montrons ensuite que le nombre d'arêtes $\{C, C'\}$ de l'arbre de jonction telles qu'un séparateur minimal donné S satisfasse $S = C \cap C'$ est exactement égal à sa multiplicité $\nu(S)$. On le montre en créant un arbre de décomposition à partir de l'arbre de jonction. Prenons d'abord une arête $\{C, C'\}$ quelconque de l'arbre de jonction et soit T et T' les sous arbres créés comme ci dessus, et V et V' les unions des cliques maximales de T et T' . On a vu que $(V \setminus S, S, V' \setminus S)$ est une décomposition du graphe triangulé. En réitérant le procédé avec T et T' on crée un arbre de décomposition dont les sommets V_1 à deux descendants sont en bijection

$$V_1 \mapsto \phi(V_1) = \{C_{V_1}, C'_{V_1}\}$$

avec les arêtes de l'arbre de jonction et tel que l'intersection des deux descendants de V_1 soit le séparateur minimal S égal à l'intersection des deux cliques maximales $\{C_{V_1}, C'_{V_1}\}$. Le théorème 8.5 montre alors que le nombre de fois où le séparateur minimal S apparaît comme intersection des deux cliques d'une arête de l'arbre de jonction est la multiplicité $\nu(S)$, nombre de composantes connexes du graphe privé de S , moins un.

9 Graphes triangulés et matrices définies positives

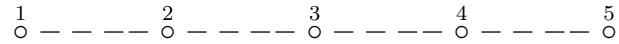
Nous allons donner dans cette section une remarquable caractérisation des graphes triangulés, due à Grone, Johnson, Sá et Wolkovitz. Pour cela, partons d'un graphe $G = (V, E)$, triangulé ou non, avec $V = \{1, \dots, n\}$. On note \tilde{E} l'ensemble des $(i, j) \in V \times V$ tels que ou bien $i = j$ ou bien $\{i, j\} \in E$. Une matrice *incomplète* associée à G est la donnée d'une fonction réelle sur \tilde{E} , soit $A = (a_{ij})_{(i,j) \in \tilde{E}}$. Elle est dite symétrique si $a_{ij} = a_{ji}$. On dit que la matrice $B = (b_{ij})_{i,j \in V}$ *prolonge* A si $b_{ij} = a_{ij}$ pour (i, j) dans \tilde{E} . Nous remarquons d'abord le résultat suivant :

Proposition 9.1. Soit A une matrice symétrique incomplète associée à G et soit \mathcal{A} l'ensemble des matrices symétriques définies positives B qui la prolonge. Alors

1. L'ensemble $\overline{\mathcal{A}}$ des matrices symétriques positives B qui prolongent A est compact.
2. Il existe une et une seule matrice $\hat{A} \in \mathcal{A}$ telle que les coefficients de \hat{A}^{-1} extérieurs à \tilde{E} sont nuls. De plus

$$\det \hat{A} = \max\{\det B; B \in \mathcal{A}\}.$$

Exemple 9.1. Donnons un exemple avant de montrer la proposition. On prend le graphe triangulé



La proposition dit qu'on peut trouver au plus une famille de nombres x et y tels que

$$\hat{A} = \begin{bmatrix} 1 & 1 & x & x & x \\ 1 & 2 & 2 & x & x \\ x & 2 & 3 & 3 & x \\ x & x & 3 & 4 & 4 \\ x & x & x & 4 & 5 \end{bmatrix} = \begin{bmatrix} y & y & 0 & 0 & 0 \\ y & y & y & 0 & 0 \\ 0 & y & y & y & 0 \\ 0 & 0 & y & y & y \\ 0 & 0 & 0 & y & y \end{bmatrix}^{-1}.$$

On verra plus tard au Théorème 9.5

que sur cet exemple on est sûr que \mathcal{A} est non vide et on apprendra à calculer les x et y au Théorème 9.6.

Démonstration. 1) On munit comme d'habitude l'espace des matrices symétriques réelles d'ordre n de la structure euclidienne $\langle B, B' \rangle = \text{tr}(BB')$. On sait que le cône des matrices positives $\overline{\Omega}$ est fermé. Donc $\overline{\mathcal{A}}$ est fermé, comme intersection de $\overline{\Omega}$ et d'une variété affine. Enfin si $B = C^2 \in \overline{\Omega}$ prolonge A alors $\|C\|^2 = \text{tr} B = a_{11} + \dots + a_{nn}$ et donc l'ensemble des $C \in \overline{\Omega}$ tels que C^2 prolonge A est donc borné. Comme $C \mapsto C^2$ est continue, $\overline{\mathcal{A}}$ est donc compact. Montrons que $\overline{\mathcal{A}}$ est la fermeture de \mathcal{A} .

2) \mathcal{A} est l'intersection, supposée non vide, du cône des matrices symétriques définies positives avec un espace affine. De plus la fonction $B \mapsto \log \det B$ est strictement concave et sa différentielle sur l'espace F des matrices symétriques qui prolonge la matrice incomplète nulle est

$$H \mapsto \text{tr} HB^{-1}.$$

Cette fonction atteint donc son maximum sur le compact $\overline{\mathcal{A}}$. Comme elle est nulle sur $\overline{\mathcal{A}} \setminus \mathcal{A}$, elle atteint son maximum sur l'ouvert convexe \mathcal{A} . D'après la stricte convexité c'est atteint en un seul point \hat{A} , qui satisfait à la conclusion de la proposition.

Nous allons maintenant examiner à quelles conditions sur A et G l'ensemble \mathcal{A} ci dessus est non vide. Il est clair que si C est une clique maximale de G

alors la matrice carrée $A_C = (a_{ij})_{i,j \in C}$ est définie positive si \mathcal{A} est non vide. Inversement, supposons que A_C soit définie positive pour toutes les cliques maximales de G . Pouvons nous affirmer que A se prolonge en une matrice symétrique B définie positive? Réponse : non si G n'est pas triangulé (c'est à dire que si G n'est pas triangulé, on peut trouver une matrice incomplète A telle que les A_C soit définies positives mais improlongeable en une définie positive), et oui si G est triangulé. Le premier résultat est le plus facile des deux et dépend de la proposition suivante :

Proposition 9.2. Soit n un entier et $A = (a_{ij})_{1 \leq i, j \leq n}$ une matrice symétrique positive telle que $a_{ij} = 1$ si $|i - j| \leq 1$. Alors $a_{ij} = 1$ pour tous i et j .

Démonstration. C'est trivial pour $n = 2$. Pour $n = 3$ on a $\det A = -(1 - a_{13})^2 \geq 0$ et donc $a_{13} = a_{31} = 1$. Soit alors $n \geq 4$ fixé. Nous allons montrer par récurrence sur k que pour tout $k < n$ et pour tous i, j tels que $|i - j| \leq k$ on a $a_{ij} = 1$. C'est l'hypothèse pour $k = 1$. Si vrai pour $k < n - 1$ soit $j = i + k + 1$. Considérons la matrice extraite de A formée des lignes et colonnes $i, i+k, j$. Elle est positive d'ordre 3. D'après l'hypothèse de récurrence $a_{i,i+k} = a_{i+k,j} = 1$ et la diagonale est 1 : d'après le cas déjà étudié $n = 3$ on a $a_{ij} = 1$ et la récurrence est étendue. En l'appliquant à $k = n - 1$ cela termine la démonstration de la proposition.

Théorème 9.3. Soit $G = (V, E)$ un graphe non triangulé avec $V = \{1, \dots, n\}$. Alors on peut choisir une matrice symétrique incomplète A associée à G telle que A_C soit positive (respectivement : définie positive) pour toute clique maximale C et soit non prolongeable en une matrice symétrique positive (respectivement : définie positive).

Démonstration. Puisque G est non triangulé, il existe donc un cycle sans corde de longueur $p \geq 4$. Sans perte de généralité on suppose que c'est $(1, 2, \dots, p, 1)$. Prenons alors la matrice symétrique incomplète A définie par $a_{i,i+1} = 1$ pour $i = 1, 2, \dots, p - 1$, $a_{1,p} = -1$, $a_{ii} = 1$ pour $i = 1, 2, \dots, n$ et $a_{ij} = 0$ pour le reste de \tilde{E} . Alors A satisfait les hypothèses du théorème dans le cas positif. En effet, pour toute clique C alors $C \cap \{1, 2, \dots, p\}$ a 0, 1 ou 2 éléments. Si c'est 0 ou 1 alors A_C est une matrice identité et est positive. Si c'est 2, alors

$$A_C = \text{diag}\left(\begin{bmatrix} 1 & \epsilon \\ \epsilon & 1 \end{bmatrix}, 1, \dots, 1\right)$$

avec $\epsilon = -1$ si $\{1, p\}$ est inclus dans C et $\epsilon = 1$ dans les autres cas. A_C est donc encore positive dans ce cas.

S'il existait un prolongement positif, alors d'après la Proposition 9.2 on aurait $a_{ij} = 1$ pour tous $1 \leq i, j \leq p$, ce qui est faux pour $i = 1$ et $j = p$. Ensuite, pour traiter le cas défini positif, considérons la même matrice incomplète A et $A_\epsilon = A + \epsilon I_n$ avec $\epsilon > 0$. Alors les $(A_\epsilon)_C$ sont définis positifs. Si pour tout $\epsilon > 0$ A_ϵ était prolongeable en une matrice symétrique définie positive

B_ϵ alors, par compacité (voir Proposition 9.1 partie 1), on pourrait trouver une suite ϵ_n tendant vers 0 telle que B_{ϵ_n} converge vers une matrice B , qui évidemment prolonge A et est symétrique positive. On a vu qu'un tel B ne peut exister et la première partie est donc montrée.

Pour étudier la même question dans le cas où G est triangulé, nous énonçons maintenant une importante propriété des graphes triangulés due à Lueker, Rose et Tarjan (1976).

Proposition 9.4. Soit $G = (V, E)$ et $G' = (V, E')$ deux graphes triangulés sur le même ensemble V de sommets tels que $E \subset E'$. Alors on peut trouver une suite croissante $E = E_0 \subset E_1 \subset \dots \subset E_s = E'$ telle que pour tout $i = 1, \dots, s$ le graphe $G_i = (V, E_i)$ soit triangulé et $E_i \setminus E_{i-1}$ soit réduit à une seule arête.

Démonstration. Voici la démonstration de Lauritzen (1996). On procède par récurrence sur la taille n de V en supposant le résultat vrai pour tous graphes triangulés à $n - 1$ sommets, et on le montre maintenant pour n . Il suffit évidemment de montrer l'existence de E_1 . Comme $G = (V, E)$ n'est pas complet, le lemme de Dirac (Lemme 6.1) garantit l'existence de deux sommets simpliciaux u et v non adjacents. Considérons alors les quatre graphes G_u, G'_u, G_v, G'_v induits par G et G' sur $V \setminus \{u\}$ et $V \setminus \{v\}$. Si $G_u = G'_u$ et $G_v = G'_v$, c'est dire que G et G' ne diffèrent que par l'arête $\{u, v\}$. Sinon, supposons par exemple $G_u \neq G'_u$. L'hypothèse de récurrence (rappelons qu'un graphe induit d'un graphe triangulé est triangulé) entraîne alors l'existence d'une arête $\{w, t\}$ de G'_u non située dans G_u telle que le graphe

$$G_0 = (V \setminus \{u\}, E_{V \setminus \{u\}} \cup \{\{w, t\}\})$$

soit triangulé. Comme u est simplicial dans G l'ensemble S de ses voisins dans G est complet. Par conséquent la partition propre de V suivante

$$V \setminus (S \cup \{u\}), S, \{u\}$$

définit une décomposition du graphe

$$G_1 = (V, E \cup \{\{w, t\}\})$$

qui est donc décomposable, ou triangulé d'après le Théorème 6.3. La proposition est donc montrée.

Théorème 9.5. Soit $G = (V, E)$ un graphe triangulé avec $V = \{1, \dots, n\}$. Alors pour toute matrice symétrique incomplète A associée à G telle que A_C soit positive (respectivement : définie positive) pour toute clique maximale C , il existe une matrice symétrique B qui la prolonge et soit positive (respectivement : définie positive).

L'exemple 9.1 illustre ce théorème : il y a 4 cliques maximales $C_i = \{i, i + 1\}$ avec $i = 1, 2, 3, 4$ et les matrices

$$A_{C_i} = \begin{bmatrix} i & i \\ i & i + 1 \end{bmatrix}$$

sont définies positives. Donc l'ensemble \mathcal{A} de la Proposition 9.1 est non vide.

Démonstration.

Le graphe G est donc triangulé. Dans la Proposition 6.9 nous prenons $G' = (V, E')$ comme le graphe complet. Nous montrons le résultat 2) du théorème par récurrence sur la taille k de $E' \setminus E$. Si $k = 0$ c'est trivial. Supposons le résultat vrai pour $k-1$, donc en particulier pour le graphe G_1 de la Proposition 8.4. Notons par $\{u, v\}$ l'unique arête de G_1 qui n'est pas dans G .

Nous montrons d'abord qu'il y a une et une seule clique maximale du graphe G_1 qui contient l'arête $\{u, v\}$. S'il y en avait 2, disons C et C' , je dis qu'alors $C \cup C'$ serait complet, et donc puisque C est maximal $C = C'$. Soit donc $z \in C$ et $z' \in C'$, et montrons que $\{z, z'\}$ est une arête de G_1 . Si (z, u, z', v) est formé de 4 points distincts, alors ils forment un cycle, car $\{z, u\}$ $\{u, z'\}$ $\{z', v\}$ et $\{v, z\}$ sont des arêtes de G_1 . L'astuce est que ce cycle de G_1 est aussi un cycle de G . Le graphe G étant triangulé, le dit cycle a donc une corde dans G . Ce ne peut être $\{u, v\}$ qui n'est pas dans E . C'est donc $\{z, z'\}$ qui est la corde, et est une arête de G et donc de G_1 . Dans le cas où (z, u, z', v) n'est pas formé de 4 points distincts, la discussion est triviale.

Soit C_0 la clique maximale ci dessus contenant $\{u, v\}$. Sans perte de généralité nous supposons $u = 1, v = 2$ et $C_0 = \{1, 2, \dots, m+2\}$. Nous avons supposé le résultat vrai pour G_1 , c'est à dire que si A_1 est une matrice symétrique incomplète associée à G_1 telle que $(A_1)_C$ est positive (respectivement : définie positive) pour toute clique maximale de G_1 , alors il existe une matrice B symétrique positive (respectivement : définie positive) qui la prolonge. Nous partons ensuite d'une matrice A symétrique incomplète associée à G (telle que A_C est positive ou définie positive pour toute clique maximale de G) et nous devons l'étendre à une matrice A_1 symétrique incomplète associée à G_1 telle que $(A_1)_C$ est positive ou définie positive pour toute clique maximale de G_1 . Cela signifie que, étant donné une matrice positive symétrique C d'ordre m , deux vecteurs colonnes $b^{(1)}$ et $b^{(2)}$ d'ordre m et deux nombres a_{11} et a_{22} tels que les deux matrices par blocs d'ordre $m+1$ suivantes

$$B_1 = \begin{bmatrix} a_{11} & {}^t b^{(1)} \\ b^{(1)} & C \end{bmatrix}, \quad B_2 = \begin{bmatrix} a_{22} & {}^t b^{(2)} \\ b^{(2)} & C \end{bmatrix}$$

soient positives, alors il est possible de trouver un nombre a_{12} tel que la matrice d'ordre $m+2$ suivante

$$B = \begin{bmatrix} a_{11} & a_{12} & {}^t b^{(1)} \\ a_{12} & a_{22} & {}^t b^{(2)} \\ b^{(1)} & b^{(2)} & C \end{bmatrix}$$

soit positive ou définie positive. La matrice incomplète A_1 associée à G_1 ainsi construite aura les propriétés voulues. En effet, si C est une clique maximale de G_1 alors ou bien elle contient $\{u, v\}$ et donc $C = C_0$ et $(A_1)_C = B$. Ou

bien elle ne contient pas $\{u, v\}$ et c'est donc aussi une clique maximale de G et donc $(A_1)_C = A_C$ est positive ou définie positive par hypothèse sur A .

Sans perte de généralité on peut supposer C diagonale (sinon, multiplier B de chaque coté par une matrice $\text{diag}(1, 1, U)$ avec U bien choisi dans le groupe orthogonal d'ordre m . Si $C = \text{diag}(c_1, c_2, \dots, c_m)$, sans perte de généralité on peut supposer les $c_i > 0$ (car si $c_i = 0$, alors $b^{(1)i} = b_i^{(2)} = 0$ puisque B_1 et B_2 sont positives). On écrit pour simplifier B par blocs, avec des notations évidentes, et on l'écrit comme un produit de triangulaires et de diagonale par blocs :

$$B = \begin{bmatrix} \alpha & {}^t\beta \\ \beta & C \end{bmatrix} = \begin{bmatrix} I_2 & {}^t\beta C^{-1} \\ 0 & I_m \end{bmatrix} \begin{bmatrix} \alpha - {}^t\beta C^{-1}\beta & 0 \\ 0 & C \end{bmatrix} \begin{bmatrix} I_2 & 0 \\ C^{-1}\beta & I_m \end{bmatrix}.$$

Il est clair alors que B est positive ou est définie positive en même temps que

$$\alpha - {}^t\beta C^{-1}\beta = \begin{bmatrix} a_{11} - \sum_{i=1}^m \frac{(b_i^{(1)})^2}{c_i} & a_{12} - \sum_{i=1}^m \frac{b_i^{(1)}b_i^{(2)}}{c_i} \\ a_{12} - \sum_{i=1}^m \frac{b_i^{(1)}b_i^{(2)}}{c_i} & a_{22} - \sum_{i=1}^m \frac{(b_i^{(2)})^2}{c_i} \end{bmatrix} \quad (9.25)$$

Rappelons alors une formule classique d'algèbre linéaire qui se démontre par récurrence sur m :

$$\det \begin{bmatrix} a & b_1 & b_2 & \dots & b_m \\ b_1 & c_1 & 0 & \dots & 0 \\ b_2 & 0 & c_2 & \dots & 0 \\ \dots & 0 & \dots & \dots & 0 \\ b_m & 0 & 0 & \dots & c_m \end{bmatrix} = c_1 \dots c_m \left(a - \sum_{i=1}^m \frac{b_i^2}{c_i} \right). \quad (9.26)$$

Cette formule montre que les éléments diagonaux de 10.48) sont ≥ 0 si B_1 et B_2 sont des matrices positives, et qu'ils sont > 0 si B_1 et B_2 sont définies positives. Par conséquent, en choisissant par exemple a_{12} égal à $\sum_{i=1}^m \frac{b_i^{(1)}b_i^{(2)}}{c_i}$ alors 10.48) est bien positive ou définie positive et le théorème 9.5 est montré.

Nous allons maintenant énoncer un résultat beaucoup plus précis que la partie "définie positive" du théorème précédent. Avec les notations de la Proposition 9.1, et dans le cas d'un graphe triangulé, ce théorème donnait les conditions nécessaires et suffisantes sur la matrice incomplète A pour que le cône \mathcal{A} soit non vide (à savoir que les A_C soient définis positifs pour toutes les cliques maximales C). La Proposition 8.1 affirmait alors l'existence d'une matrice unique \hat{A} de \mathcal{A} telle que \hat{A}^{-1} ait ses entrées nulles dans $V \times V \setminus \tilde{E}$. Dans le cas d'un graphe triangulé, le théorème suivant donne en fait une formule explicite pour \hat{A} . Pour énoncer celle ci, adoptons la notation suivante : Si $C \subset V = \{1, 2, \dots, n\}$ et si $M = (m_{i,j})_{i,j \in C}$ est une matrice carrée dont les coefficients sont indexés par C , on notera $(M)_0$ la matrice $(m_{i,j})_{i,j \in V}$ qui prolonge M et est telle que $m_{i,j} = 0$ si i ou j ne sont pas dans C .

Théorème 9.6. Soit $G = (V, E)$ un graphe triangulé connexe, avec $V = \{1, 2, \dots, n\}$. Soit \mathcal{C} et \mathcal{S} les familles des cliques maximales et des séparateurs minimaux, et soit $\nu(S)$ la multiplicité de $S \in \mathcal{S}$ (c'est à dire que $1 + \nu(S)$ est le nombre de composantes connexes U du graphe $G_{V \setminus S}$ telles que S n'est pas une clique maximale de $S \cup U$). Soit ensuite A une matrice symétrique incomplète associée à G telle que A_C soit définie positive pour toute clique maximale. Alors la matrice \hat{A} définie à la Proposition 8.1 est donnée par

$$\hat{A}^{-1} = \sum_{C \in \mathcal{C}} (A_C^{-1})_0 - \sum_{S \in \mathcal{S}} \nu(S) (A_S^{-1})_0 \quad (9.27)$$

$$\det \hat{A} = \prod_{C \in \mathcal{C}} \det A_C \prod_{S \in \mathcal{S}} \det A_S^{-\nu(S)}. \quad (9.28)$$

Exemple 9.2. On prend l'exemple du graphe de l'exemple 9.1 et de la matrice incomplète

$$A = \begin{bmatrix} 1 & 1 & & & \\ 1 & 2 & 2 & & \\ & 2 & 3 & 3 & \\ & & 3 & 4 & 4 \\ & & & 4 & 5 \end{bmatrix}.$$

Les cliques sont $C_q = \{q, q + 1\}$, $q = 1, 2, 3, 4$ et les séparateurs sont $S_q = \{q\}$, $q = 2, 3, 4$, tous de multiplicité 1. On a donc

$$A_{C_q}^{-1} = \begin{bmatrix} 1 + \frac{1}{q} & -1 \\ -1 & 1 \end{bmatrix},$$

et $A_{S_q}^{-1} = 1/q$. Le théorème dit que

$$\begin{aligned} \hat{A}^{-1} &= \begin{bmatrix} 2 & -1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 3/2 & -1 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\ &+ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4/3 & -1 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5/4 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix} \\ &- \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} 2 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}.$$

Exemple 9.3. On prend le graphe à 5 sommets $V = \{0, 1, \dots, 4\}$ avec 4 arêtes $C_q = \{0, q\}$. C'est un arbre, donc les arêtes sont les cliques maximales et le graphe est triangulé. Il n'y qu'un séparateur, qui est $S = \{0\}$. Comme le graphe privé de S a 4 composantes connexes, le séparateur S est de multiplicité 3. Considérons la matrice incomplète A adaptée à ce graphe

$$A = \begin{bmatrix} 2 & 1 & 1 & 1 & 1 \\ 1 & 1 & & & \\ 1 & & 1 & & \\ 1 & & & 1 & \\ 1 & & & & 1 \end{bmatrix}.$$

Alors $A_{C_q}^{-1} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$ et $A_S^{-1} = 1/2$. Donc on a

$$\begin{aligned} \hat{A}^{-1} &= \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\ &+ \begin{bmatrix} 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 2 & \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 2 \end{bmatrix} - 3 \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} \frac{5}{2} & -1 & -1 & -1 & -1 \\ -1 & 2 & 0 & 0 & 0 \\ -1 & 0 & 2 & 0 & 0 \\ -1 & 0 & 0 & 2 & 0 \\ -1 & 0 & 0 & 0 & 2 \end{bmatrix} \end{aligned}$$

Démonstration. Munissons \mathcal{C} d'un ordre parfait (C_1, \dots, C_k) . On sait d'après le Corollaire 8.4 que si $H_q = C_1 \cup \dots \cup C_q$, alors la suite des $S_q = H_{q-1} \cap C_q$ avec $q = 2, \dots, k$ prend ses valeurs dans \mathcal{S} et que $\nu(S)$ est le nombre de q tels que $S = S_q$. Nous devons donc montrer que

$$\hat{A}^{-1} = \sum_{q=1}^k (A_{C_q}^{-1})_0 - \sum_{q=2}^k (A_{S_q}^{-1})_0 \quad (9.29)$$

$$\det \hat{A} = \prod_{q=1}^k \det A_{C_q} \prod_{q=2}^k \det A_{S_q}^{-1}. \quad (9.30)$$

Nous montrons alors le résultat par récurrence sur le nombre k de cliques maximales. Le résultat est trivial pour $k = 1$. Supposons le vrai pour $k - 1$.

On s'appuie alors sur le résultat d'algèbre linéaire :

Proposition 9.7. Soit x une matrice symétrique définie positive écrite par blocs :

$$x = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}.$$

Alors il y a équivalence entre les faits suivants :

1. On a $x_{13} = x_{12}x_{22}^{-1}x_{23}$.
2. On a $(x^{-1})_{13} = 0$.

Dans ces conditions, si on note $x_K = \begin{bmatrix} x_{22} & x_{23} \\ x_{32} & x_{33} \end{bmatrix}$ et $x_M = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$, alors on a les égalités

$$\det x \det x_{22} = \det x_M \det x_K, \quad (9.31)$$

$$x^{-1} = \begin{bmatrix} x_M^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & x_K^{-1} \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & x_{22}^{-1} & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (9.32)$$

Acceptons quelques instants cette proposition, et appliquons la à la décomposition par blocs de $x = \hat{A}$ correspondant à $H_{k-1} \setminus S_k$ pour le bloc 1, à S_k pour le bloc 2 et à $C_k \setminus S_k$ pour le bloc 3. On voit qu'avec les notations de la proposition on a $x_{22} = A_{S_k}$, $x_M = A_{C_k}$ et enfin que

$$x_K = \hat{A}_{H_{k-1}}.$$

On est alors en position d'appliquer l'hypothèse de récurrence à la matrice incomplète B qui est la restriction à H_{k-1} de la matrice incomplète A , puisque $G_{H_{k-1}}$ n'a que $k - 1$ cliques. \hat{B} existe donc et satisfait

$$\hat{B}^{-1} = \sum_{q=1}^{k-1} (A_{C_q}^{-1})_0 - \sum_{q=2}^{k-1} (A_{S_q}^{-1})_0 \quad (9.33)$$

$$\det \hat{B} = \prod_{q=1}^{k-1} \det A_{C_q} \prod_{q=2}^{k-1} \det A_{S_q}^{-1}. \quad (9.34)$$

D'après l'unicité de \hat{B} garantie par la Proposition 9.1, on a $\hat{B} = \hat{A}_{H_{k-1}}$. La Proposition 9.7, 9.33) et 9.34) permettent de passer à 9.27) et 9.28), qui sont donc montrés.

Démonstration de la Proposition 9.7. Elle utilise la formule

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} = \begin{bmatrix} 1 & BC^{-1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} (A - BC^{-1}B^T) & 0 \\ 0 & C \end{bmatrix} \begin{bmatrix} 1 & 0 \\ C^{-1}B^T & 1 \end{bmatrix} \quad (9.35)$$

appliquée aux matrices $A = \begin{bmatrix} x_{11} & x_{13} \\ x_{31} & x_{33} \end{bmatrix}$, $B^T = [x_{21}, x_{23}]$ et $C = [x_{22}]$. Pour éviter des expressions monstrueuses, nous adoptons la notation suivante : pour i, j, k dans $\{1, 2, 3\}$ on note

$$x_{(ij)k} = x_{ij} - x_{ik}x_{kk}^{-1}x_{kj}, \quad (9.36)$$

qu'il n'est pas difficile de mémoriser.

Puisque

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix}^{-1} = \begin{bmatrix} (A - BC^{-1}B^T)^{-1} & X \\ X^T & Y \end{bmatrix}$$

où les X et Y sont des matrices qu'il est inutile de spécifier ici, on obtient

$$\begin{aligned} (A - BC^{-1}B^T)^{-1} &= \left[\begin{bmatrix} x_{11} & x_{13} \\ x_{31} & x_{33} \end{bmatrix} - \begin{bmatrix} x_{12} \\ x_{32} \end{bmatrix} x_{22}^{-1} \begin{bmatrix} x_{21} & x_{23} \end{bmatrix} \right]^{-1} \\ &= \begin{bmatrix} x_{(11)2} & x_{(13)2} \\ x_{(31)2} & x_{(33)2} \end{bmatrix}^{-1} \end{aligned}$$

Cette égalité montre l'équivalence annoncée, puisque l'entrée $(1, 3)$ de x^{-1} est $x_{(13)2} = x_{13} - x_{12}x_{22}^{-1}x_{23}$.

Pour démontrer la fin de la proposition, on utilise l'identité suivante, hideuse mais efficace.

$$x = \begin{bmatrix} 1 & x_{12}x_{22}^{-1} & 0 \\ 0 & 1 & 0 \\ 0 & x_{32}x_{22}^{-1} & 1 \end{bmatrix} \begin{bmatrix} x_{(11)2} & 0 & x_{(13)2} \\ 0 & x_{22} & 0 \\ x_{(31)2} & 0 & x_{(33)2} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ x_{22}^{-1}x_{21} & 1 & x_{22}^{-1}x_{23} \\ 0 & 0 & 1 \end{bmatrix}. \quad (9.37)$$

Elle ne prend pas longtemps à vérifier. Ensuite, si on suppose que $x_{(13)2} = 0$, la matrice centrale du second membre de 9.37) devient diagonale et comme en général

$$\begin{bmatrix} 1 & a & 0 \\ 0 & 1 & 0 \\ 0 & b & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -a & 0 \\ 0 & 1 & 0 \\ 0 & -b & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ a & 1 & b \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -a & 1 & -b \\ 0 & 0 & 1 \end{bmatrix},$$

la matrice x est facile à inverser. On obtient donc l'inverse x^{-1} :

$$\begin{bmatrix} x_{(11)2}^{-1} & -x_{(11)2}^{-1}x_{12}x_{22}^{-1} & 0 \\ -x_{22}^{-1}x_{21}x_{(11)2}^{-1} & x_{22}^{-1}x_{21}x_{(11)2}^{-1}x_{12}x_{22}^{-1} + x_{22}^{-1}x_{23}x_{(33)2}^{-1}x_{32}x_{22}^{-1} + x_{22}^{-1} & -x_{(33)2}^{-1}x_{32}x_{22}^{-1} \\ 0 & -x_{22}^{-1}x_{23}x_{(33)2}^{-1} & x_{33} \end{bmatrix}. \quad (9.38)$$

À ce point, on utilise la formule 9.35) appliquée à la matrice x_K en prenant $A = x_{11}$, $B = x_{12}$ et $C = x_{22}$. Cela nous donne la formule explicite pour x_K^{-1} :

$$x_K^{-1} = \begin{bmatrix} x_{(11)2}^{-1} & -x_{(11)2}^{-1}x_{12}x_{22}^{-1} \\ -x_{22}^{-1}x_{21}x_{(11)2}^{-1} & x_{22}^{-1}x_{21}x_{(11)2}^{-1}x_{12}x_{22}^{-1} + x_{22}^{-1} \end{bmatrix},$$

et une formule similaire pour x_M obtenue en changeant 1 en 3. En comparant avec 9.38), la formule 9.32) est montrée.

Reste à montrer la formule 9.31) sous l'hypothèse $x_{(13)2} = 0$. Dans ce cas la formule 9.37) donne $\det x = \det x_{22} \det x_{(11)2} \det x_{(33)2}$. Cependant la formule 9.35) appliquée à $A = x_{11}$, $B = x_{12}$ et $C = x_{22}$ donne $\det x_K = \det x_{22} \det x_{(11)2}$. De même, $\det x_K = \det x_{22} \det x_{(33)2}$. Donc 9.31) est montré, ce qui achève la démonstration de la Proposition 9.7. A noter que l'exercice 2.3 fournissait une démonstration probabiliste de 9.31).

Remarque. La formule $x_{13} = x_{12}x_{22}^{-1}x_{23}$ de la Proposition 9.7, contient un algorithme pour calculer directement \hat{A} , alors que le Théorème 9.6 ne donnait que \hat{A}^{-1} et qu'il peut être pénible de calculer l'inverse d'une grande matrice. Cet algorithme n'utilise que des inverses de petites matrices, de la forme A_S où S est un séparateur minimal du graphe triangulé. On part d'un ordre parfait (C_1, \dots, C_k) entre les cliques, et on note (S_2, \dots, S_k) la suite associée de séparateurs minimaux. On applique $x_{13} = x_{12}x_{22}^{-1}x_{23}$ au cas où le bloc 1 est défini par $C_1 \setminus S_2$, le bloc 2 est défini par S_2 et le bloc 3 est défini par $C_2 \setminus S_2$. Ainsi la partie de \hat{A} correspondant à $C_1 \setminus S_2 \times C_2 \setminus S_2$ se trouve t-elle déterminée, au prix de l'inversion de $x_{22} = A_{S_2}$. La matrice \hat{A} est donc maintenant connue sur $H_2 \times H_2$, avec $H_2 = C_1 \cup C_2$. De même on détermine la partie de \hat{A} correspondant à $H_1 \setminus S_3 \times C_3 \setminus S_3$ au prix de l'inversion de $x_{22} = A_{S_3}$, et la matrice \hat{A} est donc maintenant connue sur $H_3 \times H_3$, avec $H_3 = H_2 \cup C_3$. Et ainsi de suite.

Revenons à l'exemple 9.2. L'ordre (C_1, C_2, C_3, C_4) est parfait. Notons $\hat{A} = (a_{ij})_{1 \leq i, j \leq k}$. En appliquant la procédure précédente on obtient successivement

$$\begin{aligned} a_{13} &= 1 \times \frac{1}{2} \times 2 = 1 \\ \begin{bmatrix} a_{14} \\ a_{24} \end{bmatrix} &= \begin{bmatrix} 1 \\ 2 \end{bmatrix} \times \frac{1}{3} \times 3 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \\ \begin{bmatrix} a_{15} \\ a_{25} \\ a_{35} \end{bmatrix} &= \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \times \frac{1}{4} \times 4 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \end{aligned}$$

Exercice 9.1. Voici une matrice incomplète A d'ordre 8. Trouver le graphe triangulé à huit sommets auquel elle est associée, ses cliques maximales, ses séparateurs minimaux et leurs multiplicités respectives. Vérifier que A satisfait les hypothèses du théorème 9.6. Trouver ensuite \hat{A}^{-1} . (Ici, le calcul direct de \hat{A} par l'algorithme

De plus

- Si $y \in \overline{P_G}$ alors y engendre une demi droite extrémale si et seulement si y est de rang 1 et si et seulement si il existe une clique maximale C et une matrice $b = (b_1, \dots, b_n)$ non nulle telle que $b_i = 0$ si $i \notin C$ tels que $y = b^T b$. Tout élément de $\overline{P_G}$ est somme d'au plus n éléments de $\overline{P_G}$ de rang 1.
- Si $x \in \overline{Q_G}$ alors x engendre une demi droite extrémale si et seulement si x est prolongeable en une matrice de rang 1 et si et seulement si pour toute clique maximale C alors x_C est de rang 1. Tout élément de $\overline{Q_G}$ est somme d'au plus n éléments de $\overline{Q_G}$ appartenant à des demi droites extrémales.

Démonstration. Elle se fait en trois parties : les extrémales de $\overline{P_G}$, celles de $\overline{Q_G}$ et la dualité. Il est clair que $\overline{P_G}$ et $\overline{Q_G}$ sont les fermetures de P_G et de Q_G .

LES EXTRÉMALES DE $\overline{P_G}$. Si y dans $\overline{P_G}$ est de rang 1, alors il existe $b^T = (b_1, \dots, b_n)$ non nulle telle que $y = bb^T$. Soit $S = \{i; b_i \neq 0\}$. Alors S est complet, car si il existe i et j dans S tels que $\{i, j\} \notin E$, alors $y_{ij} = b_i b_j = 0$, contradiction. Une clique maximale C contenant S a donc les propriétés voulues. Pour voir que si $y = bb^T$ alors $D = \{\lambda y; \lambda \geq 0\}$ est une génératrice extrémale, on écrit $bb^T = y' + y''$ avec y' et y'' dans $\overline{P_G}$. Pour tout $a^T = (a_1, \dots, a_n)$ tel que $a^T b = a_1 b_1 + \dots + a_n b_n = 0$ on a donc

$$0 = (a^T b)^2 = a^T y' a + a^T y'' a$$

et donc $a^T y' a = a^T y'' a = 0$ à cause de la positivité des matrices y' et y'' . Comme c'est vrai pour tout a orthogonal à b dans \mathbb{R}^n euclidien canonique cela montre que y' et y'' sont dans D .

Inversement soit $y \in \overline{P_G}$ et montrons qu'on peut l'écrire comme somme d'au plus n éléments de rang 1 de $\overline{P_G}$ linéairement indépendants. Ici interviendra le fait que G est triangulé. On procède par récurrence sur n . C'est trivial pour $n = 1$ et on suppose le résultat vrai jusqu'à $n - 1$. Sans perte de généralité, d'après le Théorème 7.2 on suppose que $(1, 2, \dots, n)$ est un ordre parfait entre les sommets. Soit $k = \max\{j; y_{jj} \neq 0\}$. Alors $y_{ij} = 0$ si i ou j est $> k$. Si $k < n$ on applique l'hypothèse de récurrence au graphe induit sur $\{1, \dots, k\}$. Si $k = n$ on écrit

$$y = \begin{bmatrix} a & b \\ b^T & y_{nn} \end{bmatrix} = \begin{bmatrix} I_{n-1} & by_{nn}^{-1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a - by_{nn}^{-1}b^T & 0 \\ 0 & y_{nn} \end{bmatrix} \begin{bmatrix} I_{n-1} & 0 \\ y_{nn}^{-1}b^T & 1 \end{bmatrix}$$

ce qui montre que $a - by_{nn}^{-1}b^T$ est positive d'ordre $n - 1$. Ensuite, comme n est simplicial (le point crucial), si $b^T = (b_1, \dots, b_{n-1})$ alors $b_j \neq 0$ entraîne $j \sim n$ car $y \in \overline{P_G}$. Donc si $b_0^T = (b_1, \dots, b_{n-1}, y_n)$ on a donc $b_0 \neq 0$ et $\frac{1}{y_{nn}} b_0 b_0^T \in \overline{P_G}$. Donc

$$y = \begin{bmatrix} a - by_{nn}^{-1}b^T & 0 \\ 0 & y_{nn} \end{bmatrix} + \frac{1}{y_{nn}} b_0 b_0^T.$$

Il est clair que les deux matrices du second membre sont dans $\overline{P_G}$ et sont linéairement indépendantes. On peut alors appliquer l'hypothèse de récurrence à $a - by_{nn}^{-1}b^T$ et au graphe triangulé induit sur $\{1, \dots, n-1\}$ et le résultat est montré. Il entraîne que les génératrices extrémales ont la forme indiquée.

LES EXTRÉMALES DE $\overline{Q_G}$. Commençons par montrer la dernière partie. C'est trivial dans un sens : si $x = (x_{ij})$ est la restriction d'une matrice de rang 1, alors x_C est certainement de rang 1 pour toute clique maximale. Inversement, nous allons montrer par récurrence sur le nombre k de cliques maximales de G le résultat suivant : si $x \in Q_G$ est telle que x_C est de rang 1 pour toute clique maximale C , alors il existe une matrice symétrique positive de rang 1 dont x est la restriction. Pour $k = 1$, c'est trivial. Supposons le résultat vrai pour $k - 1$ et soit (C_1, \dots, C_k) une énumération parfaite des cliques. Notons pour simplifier $M = C_1 \cup \dots \cup C_{k-1}$, $K = C_k$ et $S = M \cap K$. D'après l'hypothèse de récurrence, x restreint à G_M est la restriction d'une matrice $(a_i a_j)_{i,j \in M}$. D'autre part, par hypothèse x_K est de la forme $(b_i b_j)_{i,j \in K}$ et on a $a_i a_j = b_i b_j$ pour tous $i, j \in S$. Donc $a_i^2 = b_i^2$ si $i \in S$, et il existe donc $\epsilon_i = \pm 1$ tel que $b_i = \epsilon_i a_i$. Donc $\epsilon_i \epsilon_j = 1$ pour tous $i, j \in S$ et donc ϵ_i est égal à la constante ϵ_S sur S . Définissons alors a_j pour $j \in K \setminus S$ par $a_j = \epsilon_S b_j$. On a alors $a_i a_j = b_i b_j$ pour tous $i, j \in K$. C'est évident pour $i, j \in S$ et $i, j \in K \setminus S$. Si $i \in S$ et $j \in K \setminus S$ alors $b_i b_j = \epsilon_i a_i \epsilon_S a_j = a_i a_j$. La matrice incomplète x est bien la restriction de la matrice de rang 1 $(a_i a_j)_{1 \leq i, j \leq n}$ et la récurrence est étendue.

Montrons ensuite que si $x \in \overline{Q_G}$ et si x_C est de rang 1 pour tout C alors x engendre une demi droite extrémale de $\overline{Q_G}$. En effet, si $x = x' + x''$ avec x' et x'' dans $\overline{Q_G}$ alors pour tout C on a $x_C = x'_C + x''_C$. Comme x_C est de rang 1 et que x'_C et x''_C sont positives cela entraîne qu'il existe des nombres λ'_C et λ''_C tels que $x'_C = \lambda'_C x_C$ et $x''_C = \lambda''_C x_C$. Mais comme G est connexe, on voit facilement que $C \mapsto (\lambda'_C, \lambda''_C)$ est constant et donc $x' = \lambda' x$, $x'' = \lambda'' x$. Donc x engendre une demi droite extrémale de $\overline{Q_G}$.

Montrons enfin que si x engendre une demi droite extrémale de $\overline{Q_G}$ alors x est la restriction d'une matrice de rang 1. En effet, d'après le Théorème 9.5, il existe une matrice symétrique positive M dont x est la restriction. Si cette matrice M n'est pas de rang 1, on l'écrit comme somme de q matrices symétriques positives M_j de rang 1, avec $2 \leq q \leq n$, soit $M = M_1 + \dots + M_q$. On note x_j la restriction de M_j . Il est clair que $x_j \neq 0$ car M_j a au moins un élément diagonal non nul, et donc $x = x_1 + \dots + x_q$ n'engendre pas une demi droite extrémale.

LA DUALITÉ. Les éléments de P_G et Q_G s'identifient naturellement à des éléments de $F = \mathbb{R}^{V \cup E}$. Il est connu que l'ensemble des matrices symétriques définies positives d'ordre n forment un cône convexe ouvert de l'espace $\mathbb{R}^{n(n+1)/2}$, dont F est un sous espace. Donc P_G est un ouvert de F . Sa fermeture $\overline{P_G}$ dans F est le cône convexe fermé de la Proposition 7.4. Notons par P_G^* le second membre de 10.39).

Montrons $Q_G \supset P_G^*$. Soit $x \in P_G^*$. Alors en tant qu'élément de F , x est identifié à une matrice symétrique incomplète gouvernée par E . Si C est une clique maximale, montrons que x_C est définie positive. Soit $b^T = (b_1, \dots, b_n)$ un vecteur ligne quelconque tel que $b_i = 0$ si $i \notin C$. Notons b_C sa restriction à C . Alors $y = b^T b$ est dans $\overline{P_G}$ et puisque $x \in P_G^*$ on a

$$0 < \text{tr}(xy) = b_C^T x_C b_C$$

et x_C est bien définie positive, puisque c'est vrai pour tout b_C . Donc $x \in Q_G$.

Montrons $Q_G \subset P_G^*$. Soit $x \in Q_G$. Pour toute clique maximale C et tout $b^T = (b_1, \dots, b_n)$ vecteur ligne quelconque tel que $b_i = 0$ si $i \notin C$ on vient de voir que si $y = b^T b$ on a $\text{tr} xy > 0$. Or on a vu à la Proposition 7.4 que tout y de $\overline{P_G}$ non nul est somme d'un nombre fini non nul de $y_i \in \overline{P_G}$ de rang 1, donc de la forme $y_i = b_i^T b_i$ où les éléments de la matrice colonne b_i sont nuls en dehors d'une clique maximale C_i . Il est clair que $\text{tr} xy = \sum_i (b_i^T)_{C_i} x_{C_i} (b_i)_{C_i} > 0$ pour un tel y et donc $x \in P_G^*$, ce qui achève la démonstration du Théorème 10.1.

Pour terminer ces remarques sur les cônes, il peut être utile de mentionner qu'en général, les cônes convexes Q_G et P_G ne sont pas des cônes homogènes, c'est à dire des cônes tels que leurs groupes d'automorphismes linéaires agiraient transitivement sur le cône. Le plus simple des contre exemples est donné par le graphe A_4 suivant



Pour vérifier que Q_{A_4} n'est pas homogène, caractérisons d'abord son groupe d'automorphismes. Nous notons par $G(Q_{A_4})$ le groupe des automorphismes linéaires de Q_{A_4} , c'est à dire l'ensemble des g automorphismes de F , l'espace des matrices incomplètes associées à A_4 tels que de plus $g(Q_{A_4}) = Q_{A_4}$. Un raisonnement simple montre alors que $g(\overline{Q_{A_4}}) = \overline{Q_{A_4}}$. Paramétrons les éléments de F ainsi

$$M(x; y) = \begin{bmatrix} x_1 & y_1 & & & \\ y_1 & x_2 & y_2 & & \\ & y_2 & x_3 & y_3 & \\ & & y_3 & x_4 & \end{bmatrix}.$$

Soit G le sous groupe de $G(Q_{A_4})$ formé des g tels que il existe des nombres a_1, a_2, a_3, a_4 tous non nuls, et deux nombres b et c tels que

$$g(M(x; y)) = \begin{bmatrix} a_1^2 x_1 + 2a_1 b y_1 + b^2 x_2 & a_2(a_1 y_1 + b x_2) & & & \\ a_2(a_1 y_1 + b x_2) & a_2^2 x_2 & a_2 a_3 y_2 & & \\ & a_2 a_3 y_2 & a_3^2 x_3 & a_3(a_4 y_3 + c x_3) & \\ & & a_3(a_4 y_3 + c x_3) & a_4^2 x_4 + 2a_4 c y_1 + c^2 x_3 & \end{bmatrix}. \tag{10.40}$$

Nous allons voir que

- G est bien un sous groupe de $G(Q_{A_4})$;

- G n'agit pas transitivement sur Q_{A_4} ;
- L'indice de G dans $G(Q_{A_4})$ est 2;
- $G(Q_{A_4})$ n'agit pas transitivement sur Q_{A_4} .

Le premier point se vérifie facilement en contrôlant la définie positivité de trois matrices d'ordre 2 par calcul de leurs déterminants respectifs. Pour le second point on raisonne par l'absurde en supposant que G agit transitivement. On le fait agir sur $M(1, 1, 1, 1; 0, 0, 0)$ pour obtenir

$$g(M(1, 1, 1, 1; 0, 0, 0)) = \begin{bmatrix} a_1^2 + b^2 & a_2b & & & \\ & a_2b & a_2^2 & 0 & \\ & & 0 & a_3^2 & a_3c \\ & & & a_3c & a_4^2 + c^2 \end{bmatrix}.$$

On voit que les $M(x; y)$ tels que $y_2 \neq 0$ ne peuvent être dans $G(M(1, 1, 1, 1; 0, 0, 0))$ et la contradiction est acquise.

Le troisième point est plus délicat. Commençons par expliquer comment on était conduit à 10.40). Pour $i = 1, 2, 3$ les inéquations $x_i x_{i+1} - y_i^2 \geq 0$, $x_i \geq 0$ et $x_{i+1} \geq 0$, avec toutes les autres variables x_j et y_j nulles sont celles d'une face F_i de dimension 3 du cône convexe fermé $\overline{Q_{A_4}}$. Montrons que $g \in G$ si et seulement si $g \in G(Q_{A_4})$ et $g(F_i) = F_i$ pour $i = 1, 2, 3$. Le "seulement si" est clair. Pour voir la partie "si", nous admettons le résultat suivant : si P_r est le cône des matrices symétriques définies positives réelles d'ordre r et si g est un automorphisme de l'espace des matrices symétriques réelles d'ordre r , alors $g(P_r) = P_r$ si et seulement si il existe une matrice réelle inversible A d'ordre r telle que $g(x) = Ax A^T$ pour tout $x \in P_r$ (C'est un amusant exercice d'algèbre linéaire : pour révéler a_{11} faire agir g sur E_{11} , pour révéler a_{12} faire agir g sur $E_{11} + E_{12} + E_{21} + E_{22}$, où E_{ij} est la matrice avec un 1 en (i, j) et des zéros ailleurs).

Appliquons alors ce principe à $r = 2$. Puisque $g(F_i) = F_i$ il existe des matrices $A_i = \begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix}$ telles que la restriction de $g \in G(Q_{A_4})$ à F_i soit de la forme

$$A_i \begin{bmatrix} x_i & y_i \\ y_i & x_{i+1} \end{bmatrix} A_i^T.$$

En écrivant les conditions de compatibilité pour les entrées x_2 et x_3 pour $g(M(x, y))$ on voit qu'il faut que $c_1 = c_2 = b_2 = b_3 = 0$ et que $d_i = a_{i+1}$, ce qui conduit à 10.40). Ensuite, G est au moins d'indice 2, puisque la symétrie S

$$(x, y) \mapsto (x_4, x_3, x_2, x_1; y_3, y_2, y_1)$$

est clairement un élément de $G(Q_{A_4})$ qui n'est pas dans G . Nous montrons maintenant que $G(Q_{A_4}) = G \cup SG$. Cela entraîne le quatrième point, puisque que les $M(x; y)$ tels que $y_2 \neq 0$ ne peuvent pas non plus être dans $SG(M(1, 1, 1, 1; 0, 0, 0))$.

Pour conclure, on remarque que le fait que Q_{A_4} ne soit pas homogène entraîne que P_{A_4} ne l'est pas non plus. En effet ces deux cônes sont duaux.

??

Le passage précédent est à revoir. Il faut en fait montrer que les éléments du groupe des automorphismes de P_G pour un graphe triangulé quelconque sont de la forme $x \mapsto axa^*$. Les ordres parfaits des cliques pourraient être liés aux représentations de ce groupe.

??

Nous calculons pour commencer une intégrale auxiliaire où les graphes triangulés n'ont rien à faire :

Proposition 10.2. Soit E et F deux espaces euclidiens de dimensions n et m , et soit a et b des endorphismes symétriques définis positifs de E et F respectivement. On munit $L = L(E, F)$ de la structure euclidienne $\langle x, y \rangle = \text{tr}(x^*y)$, où $x^* \in L(F, E)$ est l'adjoint de x , et on munit L de la mesure de Lebesgue dx associée à cette structure euclidienne. Alors

1. L'application $x \mapsto bxa$ est un endomorphisme symétrique défini positif de L , de déterminant $D = \det a^m \det b^n$.
2. Pour tout $y \in L$ on a

$$\int_L e^{-\text{tr}(x^*bxa)+2\text{tr}(x^*y)} dx = \frac{\pi^{mn/2}}{D^{1/2}} e^{\text{tr}(y^*b^{-1}ya^{-1})}. \tag{10.41}$$

3. La fonction sur L définie par $x \mapsto e^{-\text{tr}(x^*bxa)+2\text{tr}(x^*y)}$ est proportionnelle à la densité d'une loi gaussienne sur L de covariance $x \mapsto \frac{1}{2}b^{-1}xa^{-1}$ et de moyenne $b^{-1}ya^{-1}$.

Démonstration. Puisque $(y^*bxa)^* = ax^*by$ il est clair que $\langle y, bxa \rangle = \langle x, bya \rangle$ et donc que $x \mapsto bxa$ est un endomorphisme symétrique de L . Soit e et f des bases orthonormées de E et F qui diagonalisent a et b . Notons $[a]_e^e = \text{diag}(\alpha_1, \dots, \alpha_n)$ et $[b]_f^f = \text{diag}(\beta_1, \dots, \beta_m)$.

Si $x \in L$ a pour matrice représentative $[x]_e^f = (x_{ij})$ alors pour $x \neq 0$ on a

$$\text{tr}(x^*bxa) = \sum_{i=1}^m \sum_{j=1}^n \alpha_j \beta_i x_{ij}^2 > 0, \tag{10.42}$$

ce qui montre la positivité de $x \mapsto bxa$. Il est clair que ses vecteurs propres sont les $f_i \otimes e_j$, l'élément de L défini par $v \mapsto f_i \langle e_j, v \rangle$, et qui est associé à la valeur propre $\alpha_j \beta_i$. Donc son déterminant est

$$D = \prod_{ij} (\alpha_j \beta_i) = \left(\prod_{i=1}^m \beta_i \right)^n \left(\prod_{j=1}^n \alpha_j \right)^m = (\det a)^m (\det b)^n.$$

Ensuite, pour exploiter 10.42) on écrit

$$\int_{-\infty}^{\infty} e^{-\alpha_j \beta_i x_{ij}^2 + 2x_{ij} y_{ij}} dx_{ij} = \frac{\pi^{1/2}}{(\alpha_j \beta_i)^{1/2}} e^{\frac{y_{ij}^2}{\alpha_j \beta_i}}.$$

En utilisant $\sum_{i=1}^m \sum_{j=1}^n \frac{y_{ij}^2}{\alpha_j \beta_i} = \text{tr}(y^* b^{-1} y a^{-1})$ ainsi que 10.42) on a le résultat 10.41). La troisième partie est une conséquence immédiate de 10.41).

Citons ensuite l'intégrale de Wishart :

Proposition 10.3. Si P_r est le cône des matrices symétriques d'ordre r définies positives et si $p > (r-1)/2$ alors pour $\theta \in P_r$ on a

$$\int_{P_r} \exp -\text{tr}(\theta x^{-1}) (\det x)^{-p-\frac{r+1}{2}} dx = \int_{P_r} \exp -\text{tr}(\theta y) (\det y)^{p-\frac{r+1}{2}} dy = \frac{\Gamma_r(p)}{(\det \theta)^p}. \quad (10.43)$$

où dy est la mesure de Lebesgue¹ sur P_r définie par $dy = \prod_{1 \leq i \leq j \leq r} dy_{ij}$ et où

$$\Gamma_r(p) = \pi^{\frac{r(r-1)}{4}} \prod_{j=1}^r \Gamma(p - \frac{j-1}{2}). \quad (10.44)$$

Remarque. Si $p > (r-1)/2$, $\theta \in P_r$ et $\sigma \in P_r$ les probabilités sur P_r

$$\begin{aligned} & \frac{(\det \theta)^p}{\Gamma_r(p)} \exp -\text{tr}(\theta x^{-1}) (\det x)^{-p-\frac{r+1}{2}} \mathbf{1}_{P_r}(x) dx, \\ & \frac{1}{(\det \sigma)^p \Gamma_r(p)} \exp -\text{tr}(\sigma^{-1} y) (\det y)^{p-\frac{r+1}{2}} \mathbf{1}_{P_r}(y) dy \end{aligned} \quad (10.45)$$

sont respectivement appelées loi inverse Wishart ordinaire de paramètres p et θ , et loi de Wishart ordinaire de paramètres p et σ . Il est évident que si $X = Y^{-1}$ alors Y suit une loi de Wishart ordinaire de paramètres (p, σ) si et seulement si X suit une loi inverse Wishart ordinaire de paramètres (p, θ) avec $\theta = \sigma^{-1}$. Nous aurons à considérer leur généralisation associée à un graphe triangulé. La loi de Wishart ordinaire a des généralisations dans d'autres directions, la plus simple consistant à prendre des valeurs de $p \leq (r-1)/2$. Ce sont les lois de Wishart singulières. Toutefois l'analogue des Wishart singulières associé à un graphe triangulé est *terra incognita*.

Démonstration. La première égalité de 10.43) découle du changement de variable $y = x^{-1}$ dont la différentielle est $h \mapsto -x^{-1} h x^{-1}$, un endomorphisme de l'espace des matrices réelles symétriques d'ordre r . En imitant la démonstration de la partie 1 de la Proposition 10.2, on voit que le déterminant de cet endomorphisme est $(-1)^r (\det x)^{-r-1}$, ce qui permet de calculer le jacobien de $x \mapsto x^{-1}$ et de montrer cette première égalité. Bien que la démonstration de Bartlett de la deuxième égalité de 10.43) figure dans tous les ouvrages d'analyse multivariée, nous allons en donner une autre en guise de préparation aux méthodes de cette section. On procède par récurrence sur r . Pour $r = 1$, c'est la définition de la fonction gamma. Supposons le résultat vrai pour tous les

¹C'est l'usage des statisticiens que nous conservons ici plutôt que de prendre la mesure de Lebesgue canonique pour la structure euclidienne engendrée par $\text{tr}(yy^T)$.

entiers $< r$ et soit m et n des entiers > 0 tels que $m + n = r$. Ecrivons alors y et θ dans P_r ainsi :

$$y = \begin{bmatrix} y_1 & y_{12} \\ y_{21} & y_2 \end{bmatrix}, \quad \theta = \begin{bmatrix} \theta_1 & \theta_{12} \\ \theta_{21} & \theta_2 \end{bmatrix}$$

avec $y_1 \in P_m$ et $y_2 \in P_n$, et notons $y'_1 = y_1 - y_{12}y_2^{-1}y_{21}$. L'égalité habituelle

$$y = \begin{bmatrix} 1 & y_{12}y_2^{-1} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y'_1 & 0 \\ 0 & y_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ y_2^{-1}y_{21} & 1 \end{bmatrix},$$

montre que $y \in P_r$ si et seulement si $y'_1 \in P_m$ et $y_2 \in P_n$. On a alors

$$\begin{aligned} & \int_{P_r} \exp -\text{tr}(\theta y) (\det y)^{p-\frac{r+1}{2}} dy \\ & \stackrel{(1)}{=} \int_{P_m \times P_n \times R^{nm}} e^{-\text{tr}(\theta_1 y'_1) - \text{tr}(\theta_2 y_2) + \text{tr}(\theta_1 y_{12} y_2^{-1} y_{21}) - 2\text{tr}(\theta_{21} y_{12})} \\ & \quad \times (\det y'_1)^{p-\frac{1}{2}(r+1)} (\det y_2)^{p-\frac{1}{2}(r+1)} dy'_1 dy_{12} dy_2 \\ & \stackrel{(2)}{=} \pi^{mn/2} (\det \theta_1)^{-n/2} \int_{P_m \times P_n} e^{-\text{tr}(\theta_1 y'_1) - \text{tr}(y_2(\theta_2 - \theta_{21}\theta_1^{-1}\theta_{12}))} \\ & \quad \times (\det y'_1)^{p-\frac{1}{2}(r+1)} (\det y_2)^{p-\frac{1}{2}(n+1)} dy'_1 dy_{12} dy_2 \\ & \stackrel{(3)}{=} \pi^{mn/2} (\det \theta_1)^{-n/2} \frac{\Gamma_m(p - \frac{n}{2})}{(\det \theta_1)^{p-\frac{n}{2}}} \frac{\Gamma_n(p)}{(\det(\theta_2 - \theta_{21}\theta_1^{-1}\theta_{12}))^p} \\ & \stackrel{(4)}{=} \pi^{mn/2} \frac{\Gamma_m(p - \frac{n}{2}) \Gamma_n(p)}{(\det(\theta))^p} \stackrel{(5)}{=} \frac{\Gamma_r(p)}{(\det \theta)^p}. \end{aligned}$$

Dans cette suite d'égalités, (1) vient du changement de variable $y \mapsto (y'_1, y_2, y_{12})$ de jacobien 1, l'égalité (2) vient de l'intégrale par rapport à dy_{12} calculée en appliquant la Proposition 10.3 à $(a, b, y) = (\theta_1, y_2^{-1}, \theta_{12})$. L'égalité (3) vient de l'hypothèse de récurrence, (4) de $\det \theta = (\det \theta_1)(\det(\theta_2 - \theta_{21}\theta_1^{-1}\theta_{12}))$ et (5) de la définition 10.44) de la fonction $\Gamma_r(p)$. La puissance de π est la bonne, puisque si $r = m + n$ on a

$$\frac{m(m-1)}{4} + \frac{n(n-1)}{4} + \frac{mn}{2} = \frac{r(r-1)}{4}. \quad (10.46)$$

La Proposition 10.3 est donc montrée.

Nous allons calculer dans la suite quelques intégrales liées à un graphe triangulé connexe $G = (V, E)$, avec $V = \{1, \dots, n\}$. Soit P_G et Q_G les cônes convexes de la Proposition 10.1. On s'astreint toujours à noter par x les éléments de Q_G et par y ceux de P_G . Soit \mathcal{C} et \mathcal{S} les ensembles de cliques maximales et de séparateurs minimaux, et soit $\nu(S)$ la multiplicité de $S \in \mathcal{S}$. Soit enfin α et β des fonctions réelles sur \mathcal{C} et \mathcal{S} respectivement. Adoptons la notation suivante pour $x \in Q_G$:

$$H_G(\alpha, \beta; x) = \frac{\prod_{C \in \mathcal{C}} (\det x_C)^{\alpha(C)}}{\prod_{S \in \mathcal{S}} (\det x_S)^{\nu(S)\beta(S)}} \quad (10.47)$$

Bien que la définition de H_G ne dépende pas d'un ordre mis entre les cliques et entre les séparateurs, les propriétés de H_G seront fortement reliées aux ordres parfaits possibles sur \mathcal{C} . Pour cette raison, il va nous falloir envisager l'ensemble \mathcal{P}_G des ordres parfaits sur la famille \mathcal{C} des cliques maximales du graphe triangulé G .

Examinons maintenant les propriétés de H_G . D'abord, la Proposition 9.7 implique

$$H_G(p, p; x) = (\det \hat{x})^p, \quad H_G(-p, -p; \varphi(y)) = (\det \hat{y})^p$$

(Rappelons qu'on note toujours $\varphi : P_G \rightarrow G_G$ la fonction réciproque $x = \varphi(y)$ de $x \mapsto \hat{x}^{-1}$). Supposons maintenant que $P = (C_1, \dots, C_k)$ soit un ordre parfait entre les cliques (et notons $\alpha(C_q) = \alpha_q$ et $\beta(S_q) = \beta_q$). Pour simplifier les notations, nous écrivons $M = C_1 \cup \dots \cup C_{k-1}$ plutôt que H_{k-1} , $K = C_k$. Donc $S_k = M \cap K$. L'ordre parfait (C_1, \dots, C_{k-1}) induit sur les cliques maximales de M est noté P_M .

Si $x \in Q_G$ notons $x_1 = (x)_{M \setminus K}$, $x_3 = (x)_{K \setminus M}$ et $x_2 = (x)_{M \cap K}$. On écrit donc

$$x = \begin{bmatrix} x_1 & x_{12} & & \\ x_{21} & x_2 & x_{23} & \\ & x_{32} & x_3 & \end{bmatrix}. \quad (10.48)$$

Par conséquent

$$x_M = \begin{bmatrix} x_1 & x_{12} \\ x_{21} & x_2 \end{bmatrix}, \quad x_K = \begin{bmatrix} x_2 & x_{23} \\ x_{32} & x_3 \end{bmatrix}.$$

Notez que x_M est une matrice incomplète du cône Q_{G_M} , alors que x_K est une matrice symétrique définie positive ordinaire.

Avec ces notations, on voit donc que

$$H_G(\alpha, \beta; x) = H_{G_M}(\alpha_M, \beta_M; x_M) (\det x_K)^{\alpha_k} (\det x_2)^{-\beta_k},$$

en convenant que α_M et β_M sont les restrictions de α et β aux cliques et séparateurs de G_M . En fait, puisque suivant une formule cent fois utilisée on a $\det x_K = \det x_2 \det(x_3 - x_{32}x_2^{-1}x_{23})$, il est encore plus utile d'écrire

$$H_G(\alpha, \beta; x) = H_{G_M}(\alpha_M, \beta_M; x_M) (\det(x_3 - x_{32}x_2^{-1}x_{23}))^{\alpha_k} (\det x_2)^{\alpha_k - \beta_k}. \quad (10.49)$$

Parmi les fonctions α et β on aura à utiliser les fonctions $-(c+1)/2$ et $-(s+1)/2$ qui signifient $\alpha(C) = -(|C|+1)/2$ et $\beta(S) = -(|S|+1)/2$. En cas d'ordre parfait, c_q et s_q sont les tailles de C_q et S_q .

Nous allons maintenant considérer deux remarquables intégrales, qui utilisent la bijection $x \mapsto \hat{x}^{-1}$ entre Q_G et P_G du Théorème 9.6, et définies pour $\theta \in Q_G$. Pour simplifier leur présentation, nous considérons la mesure sur Q_G suivante² :

²La mesure $\nu_G(dy)$ sur P_G , image de $\mu_G(dx)$ par $x \mapsto y = \hat{x}^{-1}$ sera de même considérée au Corollaire 11.3.

$$\mu_G(dx) = H_G\left(-\frac{1}{2}(c+1), -\frac{1}{2}(s+1); x\right) \mathbf{1}_{Q_G} dx \quad (10.50)$$

où $dx = \prod_{i \leq j; (i,j) \in \tilde{E}} dx_{ij}$ est la mesure de Lebesgue sur l'espace des matrices incomplètes associées à G . Notons que si x est écrit comme dans 10.48) on a le droit d'écrire

$$dx = dx_M dx_{23} dx_3$$

avec les normalisations pour les mesures de Lebesgue faites aux Propositions 10.2 et 10.3. Cette précaution a son importance pour la signification et la valeur des intégrales suivantes :

$$I_1(\alpha, \beta, \theta) = \int_{Q_G} e^{-\text{tr}(x\hat{\theta}^{-1})} H_G(\alpha, \beta, x) \mu_G(dx) \quad (10.51)$$

$$I_2(\alpha, \beta, \theta) = \int_{Q_G} e^{-\text{tr}(\theta\hat{x}^{-1})} H_G(\alpha, \beta, x) \mu_G(dx) \quad (10.52)$$

où α et β sont choisis pour qu'elles convergent. Nous allons nous intéresser (Théorèmes 10.4 et 10.5) aux cas où ces intégrales sont proportionnelles à $H_G(\alpha, \beta; \theta)$. En d'autres termes, si on considère les deux opérateurs K_1 et K_2 sur l'espace de Hilbert $L^2(\mu_G)$ définis par

$$K_1(f)(\theta) = \int_{Q_G} e^{-\text{tr}(x\hat{\theta}^{-1})} f(x) \mu_G(dx), \quad K_2(f)(\theta) = \int_{Q_G} e^{-\text{tr}(\theta\hat{x}^{-1})} f(x) \mu_G(dx),$$

nous allons étudier certaines valeurs de (α, β) telles que $H_G(\alpha, \beta; \cdot)$ soit une fonction propre pour K_1 ou K_2 . La Proposition 10.3 donnait une illustration de ce phénomène dans le cas $k = 1$ en montrant qu'alors $y \mapsto (\det y)^p$ et $x \mapsto (\det x)^{-p}$ sont des fonctions propres de K_1 et K_2 pour la même valeur propre $\Gamma_r(p)$, avec $p > (r-1)/2$.

Pour $k \geq 2$ la situation est plus compliquée. En fait, à chaque ordre parfait $P \in \mathcal{P}_G$ nous allons associer deux ouverts A_P et B_P de \mathbb{R}^{k+1} formés de (α, β) tels que $H_G(\alpha, \beta; \cdot)$ soit vecteur propre pour K_1 ou K_2 respectivement, et nous allons calculer les valeurs propres $\Gamma_1(\alpha, \beta)$ et $\Gamma_2(\alpha, \beta)$ correspondantes. Nous allons exclure le cas $k = 1$ (c'est à dire supposer G non complet), d'abord parce que il est traité par la Proposition 10.3, ensuite parce qu'il est radicalement différent du cas $k \geq 2$, puisque la variété des vecteurs propres passe de la dimension 1 à la dimension $k+1$.

Théorème 10.4. Soit G un graphe triangulé connexe non complet, soit $P = (C_1, \dots, C_k)$ un ordre parfait de \mathcal{C} et soit (S_2, \dots, S_k) la suite des séparateurs minimaux associée. On note $J(P, S)$ l'ensemble des $j = 2, \dots, k$ tels que $S_j = S$. Soit A_P l'ensemble des (α, β) , fonctions sur \mathcal{C} et \mathcal{S} , telles que pour tout S différent de S_2 on ait

$$\nu(S)\beta(S) = \sum_{j \in J(P, S)} \alpha_j,$$

telles que $\alpha(C) > \frac{|C|-1}{2}$ pour tout $C \in \mathcal{C}$, et telles que

$$\alpha_1 + \sum_{j \in J(P, S_2)} \alpha_j > \nu(S_2)\beta_2 + \frac{s_2 - 1}{2}.$$

Alors pour $\theta \in Q_G$ et pour $(\alpha, \beta) \in A_P$ l'intégrale $I_1(\alpha, \beta; \theta)$ converge et la fonction $H_G(\alpha, \beta; \cdot)$ est une fonction propre de K_1 pour la valeur propre

$$\begin{aligned} \Gamma_1(\alpha, \beta) &= \pi^{\frac{1}{2}((c_1 - s_2)s_2 + \sum_{j=2}^k (c_j - s_j)s_j)} \Gamma_{s_2} \left(-\nu(S_2)\beta_2 + \alpha_1 + \sum_{j \in J(P, S_2)} \alpha_j \right) \\ &\quad \Gamma_{c_1 - s_2} \left(\alpha_1 - \frac{s_2}{2} \right) \prod_{j=2}^k \Gamma_{c_j - s_j} \left(\alpha_j - \frac{s_j}{2} \right) \\ &= \Gamma_{s_2} \left(-\nu(S_2)\beta_2 + \alpha_1 + \sum_{j \in J(P, S_2)} \alpha_j \right) \frac{\Gamma_{c_1}(\alpha_1)}{\Gamma_{s_2}(\alpha_1)} \prod_{q=2}^k \frac{\Gamma_{c_q}(\alpha_q)}{\Gamma_{s_q}(\alpha_q)} \end{aligned}$$

Remarques. En utilisant la fonction réciproque $\theta = \varphi(y)$ de $y = \hat{\theta}^{-1}$ on reformule le théorème en disant que pour $y \in P_G$ et pour $(\alpha, \beta) \in A_P$ on a

$$\int_{Q_G} e^{-\text{tr } xy} H(\alpha, \beta; x) \mu_G(dx) = \Gamma_1(\alpha, \beta) H_G(\alpha, \beta; \varphi(y)). \quad (10.53)$$

Il peut être plus agréable d'introduire les quantités

$$\begin{aligned} \gamma(P, S) &= -\nu(S)\beta(S) + \sum_{j \in J(P, S)} \alpha_j \text{ pour } S \neq S_2 \\ \gamma(P, S_2) &= -\nu(S_2)\beta_2 + \alpha_1 + \sum_{j \in J(P, S)} \alpha_j \end{aligned}$$

La description de A_P se fait alors par $\alpha_j > (c_j - 1)/2$ pour tout $j = 1, \dots, k$, $\gamma(P, S) = 0$ si $S \neq 2$ et $\gamma(P, S_2) > (s_2 - 1)/2$.

Insistons sur le fait que par définition la fonction Γ_1 dépend apparemment de l'ordre parfait ambiant P . Toutefois la formule 10.53 montre que si $(\alpha, \beta) \in A_P \cap A_{P'}$ alors les définitions P et P' pour $\Gamma_1(\alpha, \beta)$ coïncident. Donc Γ_1 est une fonction bien définie sur $\cup_{P \in \mathcal{P}_G} A_P$. Notons aussi que les deux expressions données de Γ_1 coïncident bien : la formule 10.46 appliquée à $m = c_q - s_q$ et $n = s_q$ explique la disparition de la puissance de π . En particulier, si les α_q et β_q sont tous égaux à un même nombre $p > \max_q \frac{c_q - 1}{2}$, le nombre

$$\Gamma_1(p, p) = \frac{\prod_{q=1}^k \Gamma_{c_q}(p)}{\prod_{q=2}^k \Gamma_{s_q}(p)} \quad (10.54)$$

ne dépend pas de l'ordre parfait P . Nous utiliserons cela à la section 13.

Ce théorème est assez étrange pour que nous l'illustrions d'un exemple.

Exemple 10.1. Considérons G triangulé à 6 sommets dont les cliques maximales sont respectivement

$$A = \{1, 2, 3\}, B = \{1, 2, 4\}, C = \{1, 2, 5\}, D = \{2, 6\}.$$

On voit que les séparateurs minimaux sont $U = \{1, 2\}$ et $V = \{2\}$ avec $\nu(U) = 2$ et $\nu(V) = 1$. On voit aussi que les 24 ordres possibles des 4 cliques sont tous parfaits. Comme A, B, C jouent des rôles symétriques nous n'allons considérer que 4 des ordres parfaits, à savoir

$$P_1 = (A, B, C, D), P_2 = (A, B, D, C), P_3 = (A, D, B, C), P_4 = (D, A, B, C)$$

On écrit alors pour chacun l'ordre correspondant des séparateurs et une description des ensembles A_{P_i} (à laquelle il faut ajouter $\alpha(A), \alpha(B), \alpha(C) > 1$ et $\alpha(D) > 1/2$)

$$\begin{array}{llll} A & B & C & D \\ & U & U & V \end{array} \quad \alpha(D) = \beta(V) \quad \alpha(A) + \alpha(B) + \alpha(C) > 2\beta(U) + 1$$

$$\begin{array}{llll} A & B & D & C \\ & U & V & U \end{array} \quad \alpha(D) = \beta(V) \quad \alpha(A) + \alpha(B) + \alpha(C) > 2\beta(U) + 1$$

$$\begin{array}{llll} A & D & B & C \\ & V & U & U \end{array} \quad \alpha(B) + \alpha(C) = 2\beta(U) \quad \alpha(A) + \alpha(D) > \beta(V) + \frac{1}{2}$$

$$\begin{array}{llll} D & A & B & C \\ & V & U & U \end{array} \quad \alpha(B) + \alpha(C) = 2\beta(U) \quad \alpha(D) + \alpha(A) > \beta(V) + \frac{1}{2}$$

Démonstration du Théorème 10.4. Celle ci se fait en plusieurs étapes. Pour simplifier les notations nous écrivons toujours $M = C_1 \cup \dots \cup C_{k-1}$, $K = C_k$ et donc $M \cap K = S_k$. On paramètre alors les éléments de Q_G comme expliqué en 10.48.

ÉTAPE 1. Si x_2 est fixé défini positif, considérons l'intégrale

$$F(x_2) = \int e^{-\text{tr}(x_K \theta_K^{-1})} (\det x_K)^{\alpha_k - (c_k + 1)/2} dx_{23} dx_3$$

où le domaine d'intégration est l'ensemble des (x_{23}, x_3) tels que $x'_3 = x_3 - x_{32} x_2^{-1} x_{23}$ soit défini positif (la notation x'_3 est plus commode ici que la notation $x_{(33)2}$ de la démonstration de la Proposition 9.7).

Calculons alors $F(x_2)$. Pour quelques instants, nous écrivons $\theta_K^{-1} = \begin{bmatrix} t_2 & t_{23} \\ t_{32} & t_3 \end{bmatrix}$.

On fait le changement de variable $(x_{23}, x_3) \mapsto (x_{23}, x'_3)$ et donc $dx_{23} dx_3 = dx_{23} dx'_3$. Le domaine d'intégration est alors un produit d'ensembles. De plus

$$\begin{aligned} e^{-\text{tr}(x_K \theta_K^{-1})} (\det x_K)^{\alpha_k - (c_k + 1)/2} &= e^{-\text{tr}(x_2 t_2)} (\det x_2)^{\alpha_k - (c_k + 1)/2} \\ &\quad \times e^{-2\text{tr} t_{32} x_{23} - \text{tr}(x_{32} x_2^{-1} x_{23} t_3)} \\ &\quad \times e^{-\text{tr}(x'_3 t_3)} (\det x'_3)^{\alpha_k - (c_k + 1)/2} \end{aligned}$$

donc $F(x_2)$ est un produit de deux intégrales. La première se calcule à l'aide de la Proposition 10.2 appliquée à $a = t_3$ $b = x_2^{-1}$, à $y = t_{23}$ et la seconde à l'aide de la Proposition 10.3 appliquée à $r = c_k - s_k$ à $p = \alpha_k - \frac{1}{2}(c_k + 1) + \frac{1}{2}(c_k - s_k + 1)$ et en remplaçant θ par t_3 . On obtient ainsi

$$\begin{aligned}
F(x_2) &= C e^{-\text{tr}(x_2 t_2)} (\det x_2)^{\alpha_k - \frac{1}{2}(c_k + 1)} \\
&\quad \times e^{\text{tr}(t_{32} x_2 t_{23} t_3^{-1})} (\det x_2)^{\frac{1}{2}(c_k - s_k)} (\det t_3)^{\frac{s_k}{2}} \\
&\quad \times (\det t_3)^{-\alpha_k + \frac{1}{2}(c_k + 1) - \frac{1}{2}(c_k - s_k + 1)} \\
&= C e^{-\text{tr}(x_2 (t_2 - t_{23} t_3^{-1} t_{32}))} (\det x_2)^{\alpha_k - \frac{1}{2}(s_k + 1)} (\det t_3)^{-\alpha_k} \\
&= C e^{-\text{tr}(x_2 \theta_2^{-1})} (\det x_2)^{\alpha_k - \frac{1}{2}(s_k + 1)} (\det \theta_3 - \theta_{32} \theta_2^{-1} \theta_{23})^{\alpha_k} \\
&= C e^{-\text{tr}(x_2 \theta_2^{-1})} (\det x_2)^{\alpha_k - \frac{1}{2}(s_k + 1)} (\det \theta_K)^{\alpha_k} (\det \theta_2)^{-\alpha_k}
\end{aligned}$$

où

$$C = \pi^{\frac{1}{2} s_k (c_k - s_k)} \Gamma_{c_k - s_k}(\alpha_k - \frac{s_k}{2}) = \frac{\Gamma_{c_k}(\alpha_k)}{\Gamma_{s_k}(\alpha_k)}. \quad (10.55)$$

Notons que l'intégrale $F(x_2)$ converge si et seulement si $\alpha_k > \frac{c_k - 1}{2}$.

ÉTAPE 2. Le lien entre I_1 et $F(x_2)$ est fourni par

$$I_1(\alpha, \beta; \theta) = \int_{Q_{G_M}} e^{-\text{tr}(x_M \theta_M^{-1}) + \text{tr}(x_2 \theta_2^{-1})} H_{G_M}(\alpha_M, \beta_M; x_M) (\det x_2)^{-\beta_k + \frac{1}{2}(s_k + 1)} F(x_2) \mu_{G_M}(dx_M).$$

Donc d'après la première étape :

$$\begin{aligned}
I_1(\alpha, \beta; \theta) &= C (\det \theta_K)^{\alpha_k} (\det \theta_2)^{-\alpha_k} \\
&\times \int_{Q_{G_M}} e^{-\text{tr}(x_M \theta_M^{-1})} H_{G_M}(\alpha_M, \beta_M; x_M) (\det x_2)^{\alpha_k - \beta_k} \mu_{G_M}(dx_M). \quad (10.56)
\end{aligned}$$

ÉTAPE 3. Nous montrons dans cette étape le résultat du théorème dans le cas $k = 2$. On suppose donc que M est aussi une clique maximale. Pour quelques instants, nous écrivons $\theta_M^{-1} = \begin{bmatrix} s_1 & s_{12} \\ s_{21} & s_2 \end{bmatrix}$. Alors 10.56 devient

$$(\det \theta_K)^{\alpha_k} (\det \theta_2)^{-\alpha_k} C \int_{Q_{G_M}} e^{-\text{tr}(x_1 s_1) - 2\text{tr}(x_{21} s_{12}) - \text{tr} s_2 x_2} (\det x_2)^{\alpha_2 - \beta_2} (\det x_M)^{\alpha_1 - \frac{c_1 + 1}{2}} dx_M.$$

En posant $x'_1 = x_1 - x_{12} x_2^{-1} x_{21}$, en exploitant $\det x_M = \det x_2 \det x'_1$ et en faisant le changement de variable $(x_{12}, x_1) \mapsto (x_{12}, x'_1)$ des calculs entièrement

analogues à ceux qui ont donné l'expression explicite de $F(x_2)$ plus haut fournissent le résultat voulu, avec la valeur propre

$$\begin{aligned} \Gamma_1(\alpha_1, \alpha_2, \beta_2) &= \pi^{\frac{s_2}{2}(c_1+c_2-2s_2)} \Gamma_{c_1-s_2}(\alpha_1 - \frac{s_2}{2}) \Gamma_{c_2-s_2}(\alpha_2 - \frac{s_2}{2}) \Gamma_{s_2}(\alpha_1 + \alpha_2 - \beta_2) \\ &= \frac{\Gamma_{s_2}(\alpha_1 + \alpha_2 - \beta_2) \Gamma_{c_1}(\alpha_1) \Gamma_{c_2}(\alpha_2)}{\Gamma_{s_2}(\alpha_1) \Gamma_{s_2}(\alpha_2)}. \end{aligned}$$

Notons que l'intégrale converge si et seulement si $\alpha_i > \frac{c_i-1}{2}$ et $\alpha_1 + \alpha_2 - \beta_2 > \frac{s_2-1}{2}$. Le cas $k = 2$ est donc terminé.

ÉTAPE 4. Nous montrons enfin le résultat général par récurrence sur k . C'est vrai pour $k = 2$ par la troisième étape, et supposons le résultat vrai pour $k-1$. On distingue alors les trois cas $\nu(S_k) = 1$; $\nu(S_k) > 1$ et $S_k \neq S_2$; $\nu(S_k) > 1$ et $S_k = S_2$.

Le cas $\nu(S_k) = 1$ est facile : puisque que $k > 2$, on a donc $S_k \neq S_2$. L'hypothèse entraîne $\alpha_k = \beta_k$ et l'intégrale 10.56 ramène au cas $k-1$ et l'hypothèse de récurrence permet de conclure.

Si $\nu(S_k) > 1$ et $S_k \neq S_2$ l'intégrale 10.56 montre que

$$I_1(\alpha, \beta; \theta) = C \frac{(\det \theta_K)^{\alpha_k}}{(\det \theta_2)^{\alpha_k}} I_{G_M}(\alpha_M, \beta'; \theta) \quad (10.57)$$

avec $\beta'(S) = \beta(S)$ si $S \neq S_k$ et $(\nu(S_k) - 1)\beta'(S_k) = \beta_k - \alpha_k$. Comme par hypothèse $\nu(S_k)\beta_k = \sum_{j \in J(P, S_k)} \alpha_j$ on a donc

$$(\nu(S) - 1)\beta'(S) = \sum_{j \in J(P_M, S)} \alpha_j$$

pour l'ordre parfait $P_M = (C_1, \dots, C_{k-1})$ du graphe G_M . La récurrence est donc étendue.

Si $\nu(S_k) > 1$ et $S_k = S_2$ la formule 10.57 est encore valable, mais comme $S_k = S_2$ on n'a pas à se soucier de contrainte du type $\nu(S_k)\beta_k = \sum_{j \in J(P, S_k)} \alpha_j$. Le Théorème 10.4 est montré.

Nous faisons maintenant la même chose avec l'intégrale de deuxième espèce $I_2(\alpha, \beta; \theta)$.

Théorème 10.5. Soit G un graphe triangulé connexe non complet, soit $P = (C_1, \dots, C_k)$ un ordre parfait de \mathcal{C} et soit (S_2, \dots, S_k) la suite des séparateurs minimaux associée. Si $S \in \mathcal{S}$ on note $J(P, S)$ l'ensemble des $j = 2, \dots, k$ tels que $S_j = S$. Soit B_P l'ensemble des (α, β) , fonctions sur \mathcal{C} et \mathcal{S} , telles que pour tout S différent de S_2 on ait

$$\nu(S)(\beta(S) + \frac{1}{2}(|S| + 1)) = \sum_{j \in J(P, S)} (\alpha_j + \frac{1}{2}(c_j + 1)),$$

telles que $-\alpha_q > \frac{1}{2}(c_q - s_q - 1)$ pour tout $q = 2, \dots, k$ ainsi que $-\alpha_1 > \frac{1}{2}(c_1 - s_1 - 1)$, et telles que

$$-\alpha_1 - \frac{1}{2}(c_1 + 1) + \sum_{j \in J(P, S_2)} (-\alpha_j - \frac{1}{2}(c_j + 1)) + \nu(S_2)(\beta_2 + \frac{1}{2}(s_2 + 1)) > 0.$$

Alors pour $\theta \in Q_G$ et pour $(\alpha, \beta) \in B_P$ l'intégrale $I_2(\alpha, \beta; \theta)$ converge et la fonction $H_G(\alpha, \beta; \cdot)$ est une fonction propre de K_2 pour la valeur propre

$$\begin{aligned} \Gamma_2(\alpha, \beta) &= \pi^{\frac{1}{2}((c_1 - s_2)s_2 + \sum_{j=2}^k (c_j - s_j)s_j)} \\ &\quad \Gamma_{s_2}(\nu(S_2)(\beta_2 + \frac{1}{2}(s_2 + 1)) + \frac{s_2 + 1}{2} - \alpha_1 - \frac{1}{2}(c_1 + 1) + \sum_{j \in J(P, S_2)} (-\alpha_j - \frac{1}{2}(c_j + 1))) \\ &\quad \Gamma_{c_1 - s_2}(-\alpha_1) \prod_{j=2}^k \Gamma_{c_j - s_j}(-\alpha_j) \end{aligned}$$

Remarques. On reformule le théorème en disant que pour $x \in Q_G$ et pour $(\alpha, \beta) \in B_P$ on a

$$\int_{P_G} e^{-\text{tr } xy} H(\alpha, \beta; \varphi(y)) \nu_G(dy) = \Gamma_2(\alpha, \beta) H_G(\alpha, \beta; x), \quad (10.58)$$

où $\nu_G(dy)$ est l'image de $\mu_G(dx)$ par $x \mapsto y = \hat{x}^{-1}$ et est calculée au Corollaire 11.3 ci dessous. Par souci de symétrie avec 10.53 on a remplacé θ par x dans 10.58. Comme au Théorème 10.4, on observe que Γ_2 est une fonction bien définie sur $\cup_{P \in \mathcal{P}_G} B_P$.

Il peut être plus agréable d'introduire les quantités

$$\begin{aligned} \tilde{\alpha}_j &= -\alpha_j - \frac{1}{2}(c_j + 1) \\ \tilde{\beta}_j &= -\beta_j - \frac{1}{2}(s_j + 1) \\ \tilde{\gamma}(P, S) &= -\nu(S)\tilde{\beta}(S) + \sum_{j \in J(P, S)} \tilde{\alpha}_j \text{ pour } S \neq S_2 \\ \tilde{\gamma}(P, S_2) &= -\nu(S_2)\tilde{\beta}_2 + \tilde{\alpha}_1 + \sum_{j \in J(P, S)} \tilde{\alpha}_j \end{aligned}$$

La description de B_P se fait alors par $\tilde{\alpha}_j > -1 - s_j/2$ pour tout $j = 2, \dots, k$, $\tilde{\alpha}_1 > -1 - s_2/2$, $\tilde{\gamma}(P, S) = 0$ si $S \neq 2$ et $\tilde{\gamma}(P, S_2) > 0$. Notez que la symétrie $(\alpha, \beta) \mapsto (\tilde{\alpha}, \tilde{\beta})$ ne transforme pas B_P en A_P et que $\Gamma_1(\tilde{\alpha}, \tilde{\beta}) \neq \Gamma_2(\alpha, \beta)$. On peut trouver que $\Gamma_2(\alpha, \beta)$ est moins intimidant quand il est écrit

$$\Gamma_2(\alpha, \beta) = \Gamma_{s_2}(\frac{1}{2}(s_2 + 1) + \tilde{\gamma}(P, S_2)) \pi^{\frac{1}{2}((c_1 - s_2)s_2)} \Gamma_{c_1 - s_2}(-\alpha_1) \prod_{j=2}^k \pi^{\frac{1}{2}(c_j - s_j)s_j} \Gamma_{c_j - s_j}(-\alpha_j)$$

Utiliser comme au Théorème 10.4 que $\pi^{\frac{1}{2}((c-s)s)}\Gamma_{c-s}(p-\frac{s}{2}) = \Gamma_c(p)/\Gamma_s(p)$ n'apporte pas grand simplification.

Finalement, le cas où il existe $p > 0$ tel que pour tous j on a $\tilde{\alpha}_j = \tilde{\beta}_j = p$ va s'avérer important à la section 13. Dans ce cas, en convenant $s_1 = s_2$

$$\Gamma_2(\alpha, \beta) = \Gamma_{s_2}(p + \frac{1}{2}(s_2 + 1)) \prod_{j=1}^k \pi^{\frac{1}{2}(c_j - s_j)s_j} \Gamma_{c_j - s_j}(p + \frac{1}{2}(c_j + 1)) \quad (10.59)$$

Démonstration du Théorème 10.5. Parallèlement au théorème précédent nous le montrons en 4 étapes. On paramètre toujours les éléments de Q_G comme en 10.48.

ÉTAPE 1. Si x_2 est fixé défini positif, considérons l'intégrale

$$G(x_2) = \int e^{-\text{tr}(\theta_K x_K^{-1})} (\det x_K)^{\alpha_k - (c_k + 1)/2} dx_{23} dx_3$$

où, comme au Théorème 10.4, le domaine d'intégration est l'ensemble des (x_{23}, x_3) tels que $x'_3 = x_3 - x_{32}x_2^{-1}x_{23}$ soit défini positif. On posera aussi $x'_{32} = x_{32}x_2^{-1}$.

Calculons alors $G(x_2)$. Nous écrivons

$$x_K = \begin{bmatrix} x_2 & x_{23} \\ x_{32} & x_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ x'_{32} & 1 \end{bmatrix} \begin{bmatrix} x_2 & 0 \\ 0 & x'_3 \end{bmatrix} \begin{bmatrix} 1 & x'_{23} \\ 0 & 1 \end{bmatrix},$$

et donc

$$\begin{aligned} x_K^{-1} &= \begin{bmatrix} 1 & -x'_{23} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_2^{-1} & 0 \\ 0 & (x'_3)^{-1} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -x'_{32} & 1 \end{bmatrix} \\ &= \begin{bmatrix} x_2^{-1} + x'_{23}(x'_3)^{-1}x'_{32} & -x'_{23}(x'_3)^{-1} \\ -(x'_3)^{-1}x'_{32} & (x'_3)^{-1} \end{bmatrix}. \end{aligned} \quad (10.60)$$

On fait le changement de variable $(x_{23}, x_3) \mapsto (x'_{23}, x'_3)$. En appliquant la Proposition 10.2 partie 1, on voit que

$$dx_{23} dx_3 = (\det x_2)^{c_k - s_k} dx'_{23} dx'_3.$$

Le domaine d'intégration est alors un produit d'ensembles. De plus

$$\begin{aligned} e^{-\text{tr} \theta_K x_K^{-1}} (\det x_K)^{\alpha_k - (c_k + 1)/2} &= e^{-\text{tr}(\theta_2 x_2^{-1})} (\det x_2)^{\alpha_k - (c_k + 1)/2} \\ &\quad \times e^{2\text{tr}(\theta_{32} x'_{23} (x'_3)^{-1}) - \text{tr}(\theta_2 x'_{23} (x'_3)^{-1} x'_{32})} \\ &\quad \times e^{-\text{tr}(\theta_3 (x'_3)^{-1})} (\det x'_3)^{\alpha_k - (c_k + 1)/2}. \end{aligned}$$

À la différence du calcul de $F(x_2)$ dans l'intégrale de première espèce du Théorème 10.4, $G(x_2)$ n'est pas l'intégrale d'un produit de deux fonctions de x'_{23}

seul et de x'_3 seul. Mais on commence par intégrer par rapport à x'_{23} à l'aide de la Proposition 10.2 appliquée à $a = \theta_2$, $b = (x'_3)^{-1}$, $y = \theta_{23}(x'_3)^{-1}$. On intègre ensuite par rapport à x'_3 à l'aide de la Proposition 10.3 appliquée à $r = c_k - s_k$, et à $-p = \alpha_k$ et en remplaçant θ par $\theta_3 - \theta_{32}\theta_2^{-1}\theta_{23}$. On obtient ainsi

$$G(x_2) = Ce^{-\text{tr}(\theta_2 x_2^{-1})} (\det x_2)^{\alpha_k - \frac{1}{2}(c_k+1) + c_k - s_k} \\ \times (\det \theta_2)^{-\frac{1}{2}(c_k - s_k)} (\det(\theta_3 - \theta_{32}\theta_2^{-1}\theta_{23}))^{\alpha_k}$$

où C est une fonction de α et β . Exploitant le fait que $(\det \theta_2)(\det(\theta_3 - \theta_{32}\theta_2^{-1}\theta_{23})) = \det \theta_K$, on obtient

$$G(x_2) = Ce^{-\text{tr}(\theta_2 x_2^{-1})} (\det x_2)^{\alpha_k - \frac{1}{2}(c_k+1) + c_k - s_k} (\det \theta_K)^{\alpha_k} (\det \theta_2)^{-\alpha_k - \frac{1}{2}(c_k - s_k)},$$

avec $C = \pi^{\frac{1}{2}s_k(c_k - s_k)} \Gamma_{c_k - s_k}(-\alpha_k)$. L'intégrale $G(x_2)$ converge si et seulement si $-\alpha_k > \frac{c_k - s_k - 1}{2}$.

ÉTAPE 2. On peut écrire en appliquant le Théorème 9.6 et l'identité 10.49 :

$$I_2(\alpha, \beta; \theta) = \int_{Q_{G_M}} e^{-\text{tr}(\theta_M x_M^{-1}) + \text{tr}(\theta_2 x_2^{-1})} H_{G_M}(\alpha_M, \beta_M; x_M) (\det x_2)^{-\beta_k + \frac{1}{2}(s_k+1)} G(x_2) \mu_{G_M}(dx_M).$$

Nous transportons alors le résultat pour $G(x_2)$ dans $I_2(\alpha, \beta; \theta)$. On obtient

$$I_2(\alpha, \beta; \theta) = C (\det \theta_K)^{\alpha_k} (\det \theta_2)^{-\alpha_k + \frac{1}{2}(c_k - s_k)} \\ \int_{Q_{G_M}} e^{-\text{tr}(\theta_M x_M^{-1})} H_{G_M}(\alpha_M, \beta_M; x_M) (\det x_2)^{\alpha_k - \beta_k - \frac{1}{2}(c_k - s_k)} \mu_{G_M}(dx_M). \quad (10.61)$$

ÉTAPE 3. On montre le résultat pour $k = 2$. On suppose donc que M est aussi une clique maximale. On posera $x'_1 = x_1 - x_{12}x_2^{-1}x_{21}$ et $x'_{21} = x_2^{-1}x_{21}$. Comme on l'a fait pour x_K on écrit maintenant

$$x_M^{-1} = \begin{bmatrix} (x'_1)^{-1} & -(x'_1)^{-1}x'_{12} \\ -x'_{21}(x'_1)^{-1} & x_2^{-1} + x'_{21}(x'_1)^{-1}x'_{12} \end{bmatrix}.$$

On fait alors le changement de variables $(x_1, x_{12}, x_2) \mapsto (x'_1, x'_{12}, x_2)$ dans 10.61, qui devient une intégrale triple qu'on calcule en intégrant d'abord par rapport à x'_{12} , puis x'_1 puis x_2 . On a par la Proposition 10.2 partie 1 que $dx_1 dx_{12} dx_2 = (\det x_2)^{c_1 - s_2} dx'_1 dx'_{12} dx_2$.

L'intégrale par rapport à x'_{12} se calcule à l'aide de la Proposition 10.2 appliquée à $a = (x'_1)^{-1}$, $b = \theta_2$ et $y = (x'_1)^{-1}\theta_{12}$.

$$\begin{aligned}
I_2(\alpha, \beta; \theta) &= C(\det \theta_K)^{\alpha_k} (\det \theta_2)^{-\alpha_2 - \frac{c_2}{2}} \\
&\int \int e^{-\text{tr}((x'_1)^{-1}(\theta_1 - \theta_{12}\theta_2^{-1}\theta_{21}) - \text{tr} \theta_2 x_2^{-1})} \\
&(\det x'_1)^{\alpha_1 - \frac{1}{2}(c_1 + s_2 + 1)} (\det x_2)^{\alpha_1 + \alpha_2 - \beta_2 + \frac{1}{2}(c_1 + c_2 - 3s_2 - 1)} dx_2 dx'_1.
\end{aligned}$$

On applique ensuite la Proposition 10.3 pour faire l'intégration suivant dx'_1 et dx_2 . Les calculs fournissent le résultat voulu, à savoir

$$I_2(\alpha_1, \alpha_2, \beta_2; \theta) = \Gamma_2(\alpha_1, \alpha_2, \beta_2) \frac{(\det \theta_M)^{\alpha_1} (\det \theta_K)^{\alpha_2}}{(\det \theta_S)^{\beta_2}},$$

avec

$$\begin{aligned}
\Gamma_2(\alpha_1, \alpha_2, \beta_2) &= \pi^{\frac{s_2}{2}(c_1 + c_2 - 2s_2)} \\
&\Gamma_{c_1 - s_2}(-\alpha_1) \Gamma_{c_2 - s_2}(-\alpha_2) \Gamma_{s_2}(-\alpha_1 - \alpha_2 + \beta_2 - \frac{1}{2}(c_1 + c_2 - 2s_2)).
\end{aligned}$$

Il y a convergence si $-\alpha_i > \frac{1}{2}(c_i - s_2 - 1)$ pour $i = 1, 2$ et $-\alpha_1 - \alpha_2 + \beta_2 - \frac{1}{2}(c_1 + c_2 - 2s_2) > \frac{1}{2}(s_2 - 1)$.

ÉTAPE 4. On montre enfin le résultat général en procédant par récurrence sur $k \geq 2$. C'est entièrement analogue au Théorème 4 et nous ne le faisons pas.

11 Lois de Wishart pour un graphe triangulé

Dans cette section, nous introduisons certaines familles exponentielles sur les cônes P_G et Q_G , nous examinons leurs objets classiques et des propriétés de composition (Proposition 11.4). Les cas particuliers des G Wishart s'appliquant aux les modèles graphiques gaussiens seront détaillés dans la section 13. Cette section 11 contient aussi un intéressant point technique : le calcul du jacobien de $x \mapsto \hat{x}^{-1}$ entre Q_G et P_G (Proposition 11.2).

Définition 11.1. Soit G un graphe triangulé, P un ordre parfait de ses cliques maximales et (α, β) dans l'ensemble A_P défini au Théorème 10.4. Soit enfin $\sigma \in Q_G$. On appelle loi de *Wishart de première espèce*, de paramètre de forme (α, β) et de paramètre d'échelle σ la loi de probabilité sur Q_G suivante

$$\gamma_{\alpha, \beta; \sigma}(dx) = e^{-\text{tr}(x\hat{\sigma}^{-1})} \frac{H_G(\alpha, \beta; x)}{\Gamma_1(\alpha, \beta) H_G(\alpha, \beta; \sigma)} \mu_G(dx). \quad (11.62)$$

On voit que pour (α, β) fixé dans A_P la famille $\{\gamma_{\alpha, \beta; \sigma}; \sigma \in Q_G\}$ n'est autre que la famille exponentielle naturelle (FEN) engendrée par la mesure sur Q_G égale à

$$\mu(dx) = \frac{H_G(\alpha, \beta; x)}{\Gamma_1(\alpha, \beta)} \mu_G(dx)$$

dont la transformée de Laplace est définie pour $-\theta \in P_G$ par

$$L_\mu(\theta) = \int_{Q_G} e^{\text{tr}(x\theta)} \mu(dx) = H_G(\alpha, \beta; -\varphi(\theta)) :$$

c'est une reformulation du Théorème 10.4. Ceci entraîne que la transformée de Laplace de $\gamma_{\alpha, \beta; \sigma}$ est définie pour $-\theta + \hat{\sigma}^{-1} \in P_G$ par

$$\frac{H_G(\alpha, \beta; \varphi(\hat{\sigma}^{-1} - \theta))}{H_G(\alpha, \beta; \sigma)}.$$

On voit déjà que si $(\alpha, \beta) \in A_P$ et $(\alpha', \beta') \in A_P$, alors comme A_P est fermé pour l'addition on a la convolution

$$\gamma_{\alpha, \beta; \sigma} * \gamma_{\alpha', \beta'; \sigma} = \gamma_{\alpha + \alpha', \beta + \beta'; \sigma}.$$

Pour dire quelque chose des objets classiques de cette famille exponentielle (moyenne en fonction de θ , inverse ψ_μ , fonction variance etc) qui ne sont pas toujours explicitables nous avons besoin d'une proposition.

Proposition 11.1. Soit P_r le cône des matrices symétriques réelles (r, r) inversibles et $-\theta \in P_r$. Soit $C \subset \{1, \dots, r\}$ et notons $\sigma_C(\theta) = ((-\theta)^{-1})_C$. Si on écrit θ par blocs correspondant à C et son complémentaire

$$\theta = \begin{bmatrix} \theta_1 & \theta_{12} \\ \theta_{21} & \theta_2 \end{bmatrix},$$

on écrit $\theta'_{12} = \theta_{12}\theta_2^{-1}$. Avec ces notations, alors la différentielle de $\theta \mapsto \sigma_C(\theta)$ est

$$h = \begin{bmatrix} h_1 & h_{12} \\ h_{21} & h_2 \end{bmatrix} \mapsto \begin{bmatrix} \sigma_C & -\sigma_C \theta'_{12} \\ \theta'_{21} \sigma_C & \theta_2 \end{bmatrix} \begin{bmatrix} h_1 & h_{12} \\ h_{21} & h_2 \end{bmatrix} \begin{bmatrix} \sigma_C \\ \theta'_{21} \sigma_C \end{bmatrix}.$$

De plus, la différentielle de $\theta \mapsto \kappa_C(\theta) = -\log \det \sigma_C(\theta)$ est

$$h \mapsto \text{tr} \begin{bmatrix} h_1 & h_{12} \\ h_{21} & h_2 \end{bmatrix} \begin{bmatrix} \sigma_C & -\sigma_C \theta'_{12} \\ -\theta'_{21} \sigma_C & \theta_2 \end{bmatrix}.$$

Démonstration. Standard.

Nous appliquons cette proposition en restreignant θ à P_G et en prenant C comme clique maximale ou séparateur minimal. Puisque la fonction de cumulants est

$$\log L_\mu(\theta) = k_\mu(\theta) = \sum_{C \in \mathcal{C}} \alpha(C) \kappa_C(\theta) - \sum_{S \in \mathcal{S}} \nu(S) \beta(S) \kappa_S(\theta), \quad (11.63)$$

alors sa différentielle est en principe calculable. Nous laissons le lecteur le faire dans le cas où $k = 2$.

Toutefois, il faut noter que dans le cas $\alpha = \beta = p$ on a $k'_\mu(\theta) = p\sigma$ donc $\sigma = m/p$. Ce ci se montre d'une manière beaucoup plus simple. En effet,

puisque d'après la Proposition 9.7 on a $H_G(p, p, x) = (\det \hat{x})^p$ alors $L_\mu(\theta) = (\det(-\theta))^{-p}$, $k_\mu(\theta) = -p \log \det(-\theta)$, $k'_\mu(\theta)(h) = p \operatorname{tr}(h(-\theta)^{-1})$, avec h dans l'espace F des matrices incomplètes gouvernées par G . Comme $-\theta = \hat{\sigma}^{-1}$ on a bien $k'_\mu(\theta) = p\sigma$.

Définition 11.2. Soit G un graphe triangulé, P un ordre parfait de ses cliques maximales et (α, β) dans l'ensemble B_P défini au Théorème 10.5. Soit enfin $\theta \in Q_G$. On appelle loi de Wishart inverse, ou loi *inverse-Wishart*, de paramètres de forme et d'échelle (α, β) et θ la loi de probabilité sur Q_G suivante

$$IW_{\alpha, \beta; \theta}(dx) = e^{-\operatorname{tr}(\theta \hat{x}^{-1})} \frac{H_G(\alpha, \beta; x)}{\Gamma_2(\alpha, \beta) H_G(\alpha, \beta; \theta)} \mu_G(dx). \quad (11.64)$$

On appelle loi de *Wishart de deuxième espèce* de paramètres de forme et d'échelle (α, β) et θ la loi de probabilité $W_{\alpha, \beta; \sigma}(dy)$ sur P_G image de la loi $IW_{\alpha, \beta; \theta}(dx)$ par l'application $x \mapsto y = \hat{x}^{-1}$.

Le fait que $IW_{\alpha, \beta; \theta}$ soit une probabilité découle du Théorème 10.5. La famille $\{IW_{\alpha, \beta; \theta}; \theta \in P_G\}$ est une famille exponentielle *générale* engendrée par la mesure $\frac{H_G(\alpha, \beta; x)}{\Gamma_2(\alpha, \beta)} \mu_G(dx)$ et la fonction $t(x) = \hat{x}^{-1}$. La famille $\{W_{\alpha, \beta; \theta}, \theta \in P_G\}$ est sa famille exponentielle naturelle associée. Pour calculer la densité de la loi $W_{\alpha, \beta; \theta}(dy)$ nous montrons la proposition suivante. Elle est due à Roverato (2002), avec une démonstration différente. On rappelle que $\varphi(y) = x$ est défini par $y = \hat{x}^{-1}$.

Proposition 11.2. La valeur absolue du jacobien de l'application $x \mapsto y = \hat{x}^{-1}$ est $H_G(-c-1, s-1; x)$. La valeur absolue du jacobien de $y \mapsto \varphi(y)$ est $H_G(c+1, s+1; \varphi(y))$.

Démonstration. On procède par récurrence sur le nombre k de cliques maximales. Si $k = 1$, c'est un résultat classique rappelé à la Proposition 10.3 que la différentielle de $x \mapsto x^{-1}$ est l'endomorphisme de l'espace des matrices symétriques (n, n) défini par $h \mapsto -x^{-1}hx^{-1}$, dont le déterminant est $(-1)^n(\det x)^{-n-1}$ (ce calcul du déterminant se fait à la manière de la Proposition 9.2). Pour $k = 1$ donc, le jacobien a bien la forme indiquée. Supposons alors la proposition montrée pour $k-1$. Soit (C_1, \dots, C_k) un ordre parfait entre les cliques maximales et notons comme d'habitude pour simplifier $M = C_1 \cup \dots \cup C_{k-1}$, $K = C_k$ et $S = M \cap K$. Si $x \in Q_G$ nous adoptons la notation 10.48).

On écrit \hat{x}^{-1} sous la forme

$$\begin{aligned} y = \hat{x}^{-1} &= (x_{C_k}^{-1})_0 - (x_{S_k}^{-1})_0 + \sum_{q=1}^{k-1} (x_{C_q}^{-1})_0 - \sum_{q=2}^{k-1} (x_{S_q}^{-1})_0 \\ &= (x_K^{-1})_0 - (x_2^{-1})_0 + ((\widehat{x_M})^{-1})_0 \\ &= \begin{bmatrix} y_1 & y_{12} & 0 \\ y_{21} & y_2 & y_{23} \\ 0 & y_{32} & y_3 \end{bmatrix}. \end{aligned} \quad (11.65)$$

Ensuite on introduit les nouvelles variables $x'_3 = x_3 - x_{32}x_2^{-1}x_{23}$ et $x'_{32} = x_{32}x_2^{-1}$ (avec $x'_{23} = x_2^{-1}x_{23}$ naturellement). En effet, $(x_K^{-1})_0 - (x_2^{-1})_0$ ne dépend que de (x'_3, x'_{32}) car comme on l'a vu à 10.60

$$\begin{bmatrix} x_2 & x_{23} \\ x_{32} & x_3 \end{bmatrix}^{-1} - \begin{bmatrix} x_2^{-1} & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} x'_{23}(x'_3)^{-1}x'_{32} & -x'_{23}(x'_3)^{-1} \\ -(x'_3)^{-1}x'_{32} & (x'_3)^{-1} \end{bmatrix}. \quad (11.66)$$

Notons aussi que $\frac{\det x_K}{\det x_2} = \det x'_3$.

D'après la Proposition 10.2 l'endomorphisme de $L(\mathbb{R}^S, \mathbb{R}^{K \setminus S})$ défini par $x_{23} \mapsto x'_{23} = (x_2)^{-1}x_{23}$ a pour déterminant $(\det x_2)^{|S|-|K|}$. Il est à peu près clair que l'étude du jacobien de $(x_M, x_{23}, x_3) \mapsto (x_M, x'_{23}, x'_3)$ fournit

$$dx = dx_M dx_{23} dx_3 = (\det x_2)^{-|S|+|K|} dx_M dx'_{23} dx'_3. \quad (11.67)$$

On fait maintenant le changement de variable

$$y_M = (\widehat{x_M})^{-1}, \quad y_{23} = -x'_{23}(x'_3)^{-1}, \quad y_3 = (x'_3)^{-1}.$$

Alors on a $dy_3 = (\det x'_3)^{-|K|+|S|-1} dx'_3$ d'après la formule rappelée au début de la démonstration pour $k = 1$. Ensuite, d'après la Proposition 10.2 l'endomorphisme de $L(\mathbb{R}^{K \setminus S}, \mathbb{R}^S)$, défini par $x'_{23} \mapsto x'_{23}(x'_3)^{-1}$ a pour déterminant $(\det x'_3)^{-|S|}$ et donc

$$dy_{23} dy_3 = (\det x'_3)^{-|K|-1} dx'_{23} dx'_3.$$

Comme d'après l'hypothèse de récurrence on avait

$$dy_M = H_{G_M}(-c_M - 1, -s_M - 1; x_M) dx_M,$$

on obtient

$$\begin{aligned} dy &= dy_M dy_{23} dy_3 = H_{G_M}(-c_M - 1, -s_M - 1; x_M) \frac{(\det x_2)^{|S|+1}}{(\det x_K)^{|K|+1}} dx_M dx_{23} dx_3 \\ &= H_G(-c - 1, -s - 1; x) dx, \end{aligned}$$

et la récurrence est étendue. Le jacobien de $y \mapsto \varphi(y)$ s'en déduit immédiatement et la Proposition 11.2 est montrée.

Corollaire 11.3. Soit $\nu_G(dy)$ l'image de $\mu_G(dx)$ par $x \mapsto y = \hat{x}^{-1}$. Alors $\nu_G(dy) = H_G(\frac{1}{2}(c+1), \frac{1}{2}(s+1); \varphi(y)) \mathbf{1}_{P_G}(y) dy$ et

$$W_{\alpha, \beta; \theta}(dy) = e^{-\text{tr}(\theta y)} \frac{H_G(\alpha, \beta; \varphi(y))}{\Gamma_2(\alpha, \beta) H_G(\alpha, \beta; \theta)} \nu_G(dy) \quad (11.68)$$

Démonstration. Immédiate.

Par conséquent, la FEN $\{W_{\alpha,\beta;\theta}, \theta \in P_G\}$ est engendrée par la mesure sur P_G

$$\nu(dy) = \frac{H_G(\alpha, \beta; \varphi(y))}{\Gamma_2(\alpha, \beta)} \nu_G(dy)$$

dont la transformée de Laplace est définie pour $-\theta \in Q_G$ par $H(\alpha, \beta; -\theta)$.

On peut chercher à déterminer ses objets classiques. Auparavant, nous faisons une remarque utile :

Proposition 11.4. Soit P un ordre parfait des cliques maximales du graphe triangulé connexe non complet G . Soit $(\alpha, \beta) \in A_P$ et $(\alpha', \beta') \in B_P$ comme aux Théorèmes 10.4 et 10.5. Considérons la loi jointe de (X, Σ) sur $Q_G \times Q_G$ définie par $\gamma_{\alpha,\beta;\sigma}(dx)IW_{\alpha',\beta';\theta}(d\sigma)$. Alors la loi conditionnelle de Σ sachant $X = x$ est $IW_{\alpha'-\alpha, \beta'-\beta; \theta+x}$.

Remarques. Cette proposition dit donc que si on rend aléatoire le paramètre d'échelle σ de la loi de Wishart de première espèce $\gamma_{\alpha,\beta;\sigma}(dx)$ en lui donnant une loi inverse-Wishart comme loi *a priori* $IW_{\alpha',\beta';\theta}(d\sigma)$ dont les paramètres de forme α', β' sont adaptés au *même ordre parfait* P que α, β , alors la loi *a posteriori* est encore une loi inverse Wishart, dont les paramètres de forme $\alpha' - \alpha, \beta' - \beta$ sont encore dans B_P . Autrement dit, si $(\alpha, \beta) \in A_P$ alors la famille

$$\{IW_{\alpha',\beta';\theta}, \theta \in P_G, (\alpha', \beta') \in B_P\}$$

est une famille conjuguée au sens bayésien de la FEN $\{\gamma_{\alpha,\beta;\sigma} \in Q_G\}$. C'est plus grand que la famille conjuguée classique (aussi dite conjuguée standard, conjuguée naturelle ou encore dite conjuguée au sens de Diaconis et Ylvisaker) qui se réduirait à

$$\{IW_{p-\alpha, p-\beta; \theta}; \theta \in P_G, p < 0\}.$$

Démonstration de la Proposition 11.4. On écrit $\gamma_{\alpha,\beta;\sigma}(dx)IW_{\alpha',\beta';\theta}(d\sigma) =$

$$\frac{\Gamma_2(\alpha' - \alpha, \beta' - \beta)}{\Gamma_1(\alpha, \beta)\Gamma_2(\alpha', \beta')} H(\alpha, \beta; x) H(\alpha' - \alpha, \beta' - \beta; \theta + x) \mu_G(dx) IW_{\alpha' - \alpha, \beta' - \beta; \theta + x}(d\sigma).$$

12 Le graphe de la marguerite, le graphe A_n

Dans cette section nous illustrons les résultats des trois sections précédentes pour deux graphes particuliers. Leur utilité statistique est modeste, mais ils ont assez de régularité pour produire des formules explicites inspiratrices de généralisations (ainsi les auteurs sont ils passés des formules d'Andersson-Vinberg ci dessus (Théorème 12.1) aux Théorèmes 10.4 et 10.5.

Le GRAPHE DE LA MARGUERITE à $k+1$ sommets a pour sommets $\{0, 1, \dots, k\}$ et pour uniques arêtes les $C_q = \{0, q\}$ avec $q = 1, \dots, k$. C'est donc un arbre, les arêtes sont les cliques maximales et c'est un graphe triangulé. L'unique

séparateur minimal est $\{0\}$ et il est de multiplicité $k - 1$. Notons par F l'espace des matrices symétriques d'ordre $k + 1$ de la forme

$$y \begin{bmatrix} y_0 & y_{01} & \cdots & y_{0k} \\ y_{01} & y_1 & \cdots & 0 \\ \cdots & 0 & \cdots & 0 \\ y_{0k} & 0 & \cdots & y_k \end{bmatrix}. \quad (12.69)$$

Alors si $y_1 \dots y_k \neq 0$ c'est un exercice classique que de vérifier par récurrence sur k que

$$\det y = y_1 \dots y_k \left(y_0 - \sum_{q=1}^k \frac{y_{0q}^2}{y_q} \right).$$

A partir de là on montre facilement, à l'aide de la caractérisation des matrices définies positives par la positivité des déterminants principaux, que si P_G est l'ensemble des $y \in F$ qui sont définis positifs, alors

$$y \in P_G \Leftrightarrow y_q > 0 \quad \forall q = 1, \dots, k; \quad y_0 - \sum_{q=1}^k \frac{y_{0q}^2}{y_q} > 0.$$

On peut ajouter que le cône $\overline{P_G}$ des éléments positifs de F est caractérisé par

$$y \in \overline{P_G} \Leftrightarrow y_q \geq 0 \quad \forall q = 1, \dots, k; \quad y_0 - \sum_{q=1}^k \frac{y_{0q}^2}{y_q} \geq 0.$$

D'après le Théorème 10.1, les demi droites extrémales de $\overline{P_G}$ sont engendrées par les y de la forme suivante : ou bien $y_0 > 0$ et tous les autres y_q et y_{0q} sont nuls ; ou bien il existe j tel que $y_0 y_j = y_{0j}^2$ avec $y_0 > 0$ et tous les autres y_q et y_{0q} sont nuls.

De même le cône Q_G des matrices incomplètes est l'ensemble des

$$x = \begin{bmatrix} x_0 & x_{01} & \cdots & x_{0k} \\ x_{01} & x_1 & \cdots & \\ \cdots & & \cdots & \\ x_{0k} & & \cdots & x_k \end{bmatrix} \quad (12.70)$$

tels que $x_0 x_q - x_{0q}^2 > 0$ pour tous $q = 1, \dots, k$ et $x_0 > 0$. Le cône $\overline{Q_G}$ est caractérisé par $x_0 x_q - x_{0q}^2 \geq 0$ et $x_q \geq 0$ pour tous $q = 1, \dots, k$, et $x_0 \geq 0$. D'après le Théorème 10.1, les demi droites extrémales de $\overline{Q_G}$ sont engendrées par les x de la forme suivante : ou bien $x_0 > 0$ et tous les autres x_q et x_{0q} sont nuls ; ou bien $x_0 x_q = x_{0q}^2$ tous $q = 1, \dots, k$ et $x_0 > 0$. Les résultats précédents concernant les cônes P_G et Q_G du graphe de la marguerite sont en appendice de la thèse de M. Casalis (1990), où il est aussi montré que P_G et Q_G sont homogènes. Maintenant

- Pour x défini par 12.70) dans Q_G , nous calculons \hat{x}^{-1} , \hat{x} et $\det \hat{x}$.
- Pour y défini par 12.69) dans P_G , nous calculons $\varphi(y)$ et $\widehat{\varphi}(y)$.

- Nous calculons les fonctions $H_G(\alpha, \beta; x)$ et $H_G(\alpha, \beta; \varphi(y))$ pour $x \in Q_G$ et $y \in P_G$ ainsi que les mesures $\mu_G(dx)$ et $\nu_G(dy)$.
- Nous calculons les ensembles A_P et B_P et les fonctions $\Gamma_1(\alpha, \beta)$ et $\Gamma_2(\alpha, \beta)$.

Comme $x_{C_q} = \begin{bmatrix} x_0 & x_{0q} \\ x_{0q} & x_q \end{bmatrix}$ alors

$$x_{C_q}^{-1} = \frac{1}{x_0x_q - x_{0q}^2} \begin{bmatrix} x_q & -x_{0q} \\ -x_{0q} & x_0 \end{bmatrix}.$$

Comme le séparateur $\{0\}$ a pour multiplicité $k-1$ on obtient par le Théorème 10.6

$$\hat{x}^{-1} = \begin{bmatrix} \frac{-k+1}{x_0} + \sum_{i=1}^k \frac{x_i}{x_0x_i - x_{0i}^2} & \frac{-x_{01}}{x_0x_1 - x_{01}^2} & \cdots & \frac{-x_{0k}}{x_0x_k - x_{0k}^2} \\ \frac{-x_{01}}{x_0x_1 - x_{01}^2} & \frac{x_0}{x_0x_1 - x_{01}^2} & \cdots & 0 \\ \cdots & 0 & \cdots & 0 \\ \frac{-x_{0k}}{x_0x_k - x_{0k}^2} & 0 & \cdots & \frac{x_0}{x_0x_k - x_{0k}^2} \end{bmatrix}.$$

Pour calculer $\hat{x} = (x_{ij})_{0 \leq i, j \leq k}$ le plus simple est d'utiliser la remarque de la fin de la section 9. L'ordre (C_1, \dots, C_k) est évidemment un ordre parfait puisqu'il n'y a qu'un séparateur. Donc $x_{12} = x_{01}x_{02}/x_0$, en général $x_{ij} = x_{0i}x_{0j}/x_0$, et donc :

$$\hat{x} = \begin{bmatrix} x_0 & x_{01} & \cdots & x_{0k} \\ x_{01} & x_1 & \cdots & \frac{x_{01}x_{0k}}{x_0} \\ \cdots & \cdots & \cdots & \cdots \\ x_{0k} & \frac{x_{0k}x_{01}}{x_0} & \cdots & x_k \end{bmatrix}$$

Quant à $\det \hat{x}$ il est donné par la Proposition 9.7 et est égal à

$$\det \hat{x} = H_G(1, 1; x) = \frac{1}{x_0^{k-1}} \prod_{q=1}^k (x_0x_q - x_{0q}^2).$$

Pour calculer $x = \varphi(y)$, il nous faut bravement résoudre les équations en x :

$$\begin{aligned} y_0 &= \frac{-k+1}{x_0} + \sum_{q=1}^k \frac{x_q}{x_0x_q - x_{0q}^2}, \\ y_{0q} &= \frac{-x_{0q}}{x_0x_q - x_{0q}^2}, \quad q = 1, \dots, k, \\ y_q &= \frac{x_0}{x_0x_q - x_{0q}^2}, \quad q = 1, \dots, k. \end{aligned}$$

L'astuce est simplement de calculer $y_0 - \sum_{q=1}^k \frac{y_{0q}^2}{y_q}$ et de constater que c'est $1/x_0$. Ensuite, on écrit $x_{0q} = -y_{0q}x_0/y_q$ puis, comme $x_0x_q - x_{0q}^2 = x_0/y_q$, on écrit

$$x_q = \frac{1}{y_q} + \frac{x_{0q}^2}{x_0} = \frac{1}{y_q} + \frac{y_{0q}^2 x_0}{y_q^2} = \frac{x_0}{y_q} \left(y_0 - \sum_{i \neq q} \frac{y_{0i}^2}{y_i} \right).$$

$x = \varphi(y)$ est donc calculé. Le coefficient x_{ij} de \hat{x} est alors $x_{ij} = x_0 \frac{y_{0i} y_{0j}}{y_i y_j}$.

Pour calculer les prochaines quantités il va être commode d'utiliser l'abréviation (en accord avec les remarques suivant le Théorème 10.4)

$$\gamma(\alpha, \beta) = -(k-1)\beta + \sum_{q=1}^k \alpha_q.$$

On écrit d'abord

$$H_G(\alpha, \beta; x) = x_0^{-(k-1)\beta} \prod_{q=1}^k (x_0 x_q - x_{0q}^2)^{\alpha_q}.$$

Le calcul fait de $\varphi(y)$ permet alors d'écrire

$$H_G(\alpha, \beta; \varphi(y)) = \left(y_0 - \sum_{q=1}^k \frac{y_{0q}^2}{y_q} \right)^{-\gamma(\alpha, \beta)} \prod_{q=1}^k y_q^{-\alpha_q}$$

(On rappelle que si $x = \varphi(y)$ alors $x_0 x_q - x_{0q}^2 = x_0 / y_q$). Quant aux mesures de référence ce sont

$$\mu_G(dx) = x_0^{k-1} \prod_{q=1}^k (x_0 x_q - x_{0q}^2)^{-3/2} \mathbf{1}_{Q_G}(x) dx \quad (12.71)$$

$$\nu_G(dy) = \left(y_0 - \sum_{q=1}^k \frac{y_{0q}^2}{y_q} \right)^{1+\frac{k}{2}} \prod_{q=1}^k y_q^{3/2} \mathbf{1}_{P_G}(y) dy \quad (12.72)$$

Pour le graphe de la marguerite tous les ordres de cliques maximales sont parfaits, et comme il n'y a qu'un séparateur, tous les ensembles A_P coïncident. Il en va de même pour les B_P . On obtient, d'après les Théorèmes 10.4 et 10.5 :

$$A_P = \{(\alpha_1, \dots, \alpha_k, \beta); \gamma(\alpha, \beta) > 0 \text{ et } \alpha_q > 1/2 \forall q = 1, \dots, k\},$$

$$B_P = \{(\alpha_1, \dots, \alpha_k, \beta); \gamma(\alpha, \beta) < \frac{k}{2} \text{ et } \alpha_q < 0 \forall q = 1, \dots, k\}$$

$$\Gamma_1(\alpha, \beta) = \pi^{k/2} \Gamma(\gamma(\alpha, \beta)) \prod_{q=1}^k \Gamma(\alpha_q - \frac{1}{2})$$

$$\Gamma_2(\alpha, \beta) = \pi^{k/2} \Gamma(-\gamma(\alpha, \beta) + \frac{k}{2}) \prod_{q=1}^k \Gamma(-\alpha_q).$$

L'application des Théorèmes 10.4 et 10.5 au graphe de la marguerite fournit alors les deux remarquables FORMULES D'ANDERSSON-VINBERG suivantes

Théorème 12.1. Avec $x \in Q_G$ et $y \in P_G$ où G est le graphe de la marguerite à $k+1$ sommets, on a pour $(\alpha, \beta) \in A_P$

$$\int_{Q_G} e^{-\text{tr } xy} H(\alpha, \beta; x) \mu_G(dx) = \Gamma_1(\alpha, \beta) H_G(\alpha, \beta; \varphi(y)),$$

et pour $(\alpha, \beta) \in B_P$ on a

$$\int_{P_G} e^{-\text{tr } xy} H(\alpha, \beta; \varphi(y)) \nu_G(dy) = \Gamma_2(\alpha, \beta) H_G(\alpha, \beta; x).$$

Démonstration. Il y en a trois possibles : soit appliquer les Théorèmes 10.4 et 10.5, soit faire le calcul direct (pour la première, faire l'intégration d'abord par rapport aux x_q , puis par rapport aux x_{0q} , puis par rapport à x_0 ; pour la deuxième, utiliser l'ordre y_0, y_{0q}, y_q), soit utiliser la méthode d'Andersson (1999), que nous n'allons pas détailler ici, et qui est basée sur le fait que dans le cas du graphe de la marguerite, alors les cônes Q_G et P_G sont homogènes et que la structure peut être considérée à la lumière des algèbres de Vinberg (voir Andersson et Vojnar (2001) pour un exposé récent sur le sujet). Dans ce contexte, les fonctions $H(\alpha, \beta; \cdot)$ sont des produits de caractères réels du groupe des automorphismes de Q_G ou de P_G . Rappelons que pour un graphe triangulé, Q_G et P_G ne sont pas en général homogènes (exemple $G = A_4$). Les deux approches "graphes triangulés" et "algèbres de Vinberg" fournissent les mêmes formules pour la marguerite, mais l'enveloppe commune des deux théories, dont aucune ne recouvre l'autre, reste donc à découvrir.

CALCUL DE LA FONCTION VARIANCE POUR UNE WISHART DE PREMIÈRE ESPÈCE POUR LA MARGUERITE. On sait qu'alors la famille exponentielle est engendrée par une mesure μ dont la transformée de Laplace version physicienne est $S_\mu(y) = \int_{Q_G} e^{-\text{tr}(yx)} \mu(dx) = H_G(\alpha, \beta, \varphi(y))$. Attention, on est dans le cas dit non classique, car

$$xQy^T = \text{tr}(xy) = \sum_{q=0}^k x_q y_q + 2 \sum_{q=1}^k x_{0q} y_{0q}.$$

Notons pour simplifier $\gamma = -(k-1)\beta + \sum_{q=1}^k \alpha_q$ et $\frac{1}{x_0} = y_0 - \sum_{q=1}^k \frac{y_{0q}^2}{y_q}$, alors

$$s_\mu(y) = -\gamma \log x_0 + \sum_{q=1}^k \alpha_q \log y_q.$$

Compte tenu de la méthode décrite dans la section 4, nous avons à calculer

- $s'_\mu(y) = m = Q^{-1}(\frac{\partial}{\partial y} s_\mu(y))$.
- La fonction réciproque $y = \rho_\mu(m)$.
- La matrice hessienne $H(y)$ des dérivées secondes de $s_\mu(y)$.
- Et enfin la fonction variance $V(m) = -Q^{-1}H(\rho_\mu(m))Q^{-1}$.

On obtient d'abord

$$\begin{array}{lll} \frac{\partial}{\partial y_0} x_0 = -x_0^2 & \frac{\partial}{\partial y_q} x_0 = -x_0^2 \frac{y_{0q}^2}{y_q^2} & \frac{\partial}{\partial y_{0q}} x_0 = 2x_0^2 \frac{y_{0q}}{y_q} \\ \frac{\partial}{\partial y_0} \log x_0 = -x_0 & \frac{\partial}{\partial y_q} \log x_0 = -x_0 \frac{y_{0q}^2}{y_q^2} & \frac{\partial}{\partial y_{0q}} \log x_0 = 2x_0 \frac{y_{0q}}{y_q} \\ \frac{\partial}{\partial y_0} s_\mu(y) = \gamma x_0 = m_0 & \frac{\partial}{\partial y_q} s_\mu(y) = \gamma x_0 \frac{y_{0q}^2}{y_q^2} + \frac{\alpha_q}{y_q} = m_q & \frac{\partial}{\partial y_{0q}} s_\mu(y) = -2\gamma x_0 \frac{y_{0q}}{y_q} = 2m_{0q}. \end{array}$$

Introduisons alors les quantités $\Delta_q = m_q m_0 - m_{0q}^2$. On les utilise pour simplifier la présentation de $y = \rho_\mu(m)$: après calculs on obtient

$$y_0 = \frac{1}{m_0} \left(\gamma + \sum_{q=1}^k \alpha_q \frac{m_{0q}^2}{\Delta_q} \right), \quad y_q = \alpha_q \frac{m_0}{\Delta_q}, \quad y_{0q} = -\alpha_q \frac{m_{0q}}{\Delta_q}.$$

Il est ensuite assez fastidieux de décrire le calcul de $H(\rho_\mu(m))$. Nous allons nous contenter de donner le résultat pour $V(m)$. Et encore, pour simplifier, nous prenons $k = 2$. Mais ceci n'est pas vraiment une restriction. Pour passer à k quelconque, il suffit de remplacer (q, q') par $(1, 2)$ si $q \neq q'$ et par $(1, 1)$ si $q = q'$. On obtient que $V_F(m)$ égale

$$\left[\begin{array}{cc|cc|cc} \frac{m_0^2}{\gamma} & \frac{m_{01}^2}{\gamma m_0^2} & \frac{m_{02}^2}{\gamma m_0^2} & \frac{m_0 m_{01}}{\gamma m_0} & \frac{m_0 m_{02}}{\gamma m_0} & \\ \frac{m_{01}^2}{\gamma} & \frac{m_{01}^4}{\gamma m_0^2} + \frac{2m_{01}^2 \Delta_1}{\alpha_1 m_0^2} + \frac{\Delta_1^2}{\alpha_1 m_0^2} & \frac{m_{01}^2 m_{02}^2}{\gamma m_0^2} & \frac{m_{01}^3}{\gamma m_0} + \frac{m_{01} \Delta_1}{\alpha_1 m_0} & \frac{m_{01}^2 m_{02}}{\gamma m_0} & \\ \frac{m_{02}^2}{\gamma} & \frac{m_{01}^2 m_{02}^2}{\gamma m_0^2} & \frac{m_{02}^4}{\gamma m_0^2} + \frac{2m_{02}^2 \Delta_2}{\alpha_2 m_0^2} + \frac{\Delta_2^2}{\alpha_2 m_0^2} & \frac{m_{01} m_{02}^2}{\gamma m_0} & \frac{m_{02}^3}{\gamma m_0} + \frac{m_{02} \Delta_2}{\alpha_2 m_0} & \\ \frac{m_0 m_{01}}{\gamma} & \frac{m_{01}^3}{\gamma m_0} + \frac{m_{01} \Delta_1}{\alpha_1 m_0} & \frac{m_{01} m_{02}^2}{\gamma m_0} & \frac{m_{01}^2}{\gamma} + \frac{\Delta_1}{2\alpha_1} & \frac{m_0 m_{01}}{\gamma} & \\ \frac{m_0 m_{02}}{\gamma} & \frac{m_{01}^2 m_{02}}{\gamma m_0} & \frac{m_{02}^3}{\gamma m_0} + \frac{m_{02} \Delta_2}{\alpha_2 m_0} & \frac{m_0 m_{01}}{\gamma} & \frac{m_{02}^2}{\gamma} + \frac{\Delta_2}{2\alpha_2} & \end{array} \right]$$

LE GRAPHE A_{k+1} À $k + 1$ SOMMETS.

Il a pour sommets

$\{1, \dots, k + 1\}$ et pour uniques arêtes les $C_q = \{q, q + 1\}$ avec $q = 1, \dots, k$.

C'est donc un arbre, les arêtes sont les cliques maximales et c'est un graphe triangulé. Les séparateurs sont les $S_q = \{q\}$ avec $q = 2, \dots, k$ et sont tous de multiplicité 1. Notons par F l'espace des matrices symétriques d'ordre $k + 1$ qui sont tridiagonales, c'est à dire donc les matrices $y = (y_{ij})_{1 \leq i, j \leq k+1}$ telles que $y_{ij} = 0$ si $|i - j| > 1$ et $y_{ij} = y_{ji}$. Pour simplifier, on note comme d'habitude $y_{qq} = y_q$. Le cône P_G est donc l'ensemble des $y \in F$ qui sont définis positifs, le cône Q_G est l'ensemble des matrices symétriques incomplètes $x = (x_{ij})_{(i,j) \in \bar{E}}$ telles que $x_q > 0$ pour tout $q = 1, \dots, k + 1$ et $\Delta_q = x_q x_{q+1} - x_{q,q+1}^2 > 0$.

La problématique est la même que pour la marguerite, il faut calculer $y = \hat{x}^{-1}$, \hat{x} , $\varphi(y)$, $H_G(\alpha, \beta; x)$, $H_G(\alpha, \beta; \varphi(y))$, μ_G , ν_G , A_P , B_P , $\Gamma_1(\alpha, \beta)$ et $\Gamma_2(\alpha, \beta)$. La réalisation de la totalité de ce programme est toutefois subordonnée à la solution d'un problème difficile, le calcul de $\varphi(y)$. Nous commençons donc par les choses faisables. On a donc d'après le Théorème 9.6, si $y = \hat{x}^{-1}$

$$y_1 = \frac{x_2}{\Delta_1} \quad y_q = \frac{x_{q-1}}{\Delta_{q-1}} + \frac{x_{q+1}}{\Delta_q} \quad y_{q,q+1} = \frac{-x_{q,q+1}}{\Delta_q} \quad q = 2, \dots, k, \quad y_{k+1} = \frac{x_k}{\Delta_k}.$$

La méthode de la remarque qui termine la section 9 permet de calculer $\hat{x} = (x_{ij})$ et on obtient pour $1 \leq i < j \leq k + 1$

$$x_{ij} = \frac{\prod_{q=i}^{j-1} x_{q,q+1}}{\prod_{q=i+1}^{j-1} x_q}.$$

Ensuite on a

$$H_G(\alpha, \beta; x) = \frac{\prod_{q=1}^k \Delta_q^{\alpha_q}}{\prod_{q=2}^k x_q^{\beta_q}}, \quad \mu_G(dx) = \frac{\prod_{q=2}^k x_q}{\prod_{q=1}^k \Delta_q^{3/2}} \mathbf{1}_{Q_G}(x) dx.$$

Quant aux ensembles A_P et B_P pour l'ordre parfait $P = (C_1, \dots, C_k)$ ce sont

$$\begin{aligned} A_P &= \{(\alpha, \beta); \alpha_1 + \alpha_2 > \beta_2, \alpha_q > 1/2 \ q = 1, \dots, k, \alpha_q = \beta_q, \ q = 3, \dots, k\}, \\ B_P &= \{(\alpha, \beta); \alpha_1 + \alpha_2 - \beta_2 < 2, \alpha_q < 0 \ q = 1, \dots, k, \alpha_q = \beta_q + \frac{1}{2}, \ q = 3, \dots, k\}. \end{aligned}$$

Enfin voici les fonctions gamma

$$\begin{aligned} \Gamma_1(\alpha, \beta) &= \pi^{\frac{k+1}{2}} \Gamma(\alpha_1 + \alpha_2 - \beta_2) \prod_{q=1}^k \Gamma(\alpha_q - \frac{1}{2}) \\ \Gamma_2(\alpha, \beta) &= \pi^{\frac{k+1}{2}} \Gamma(-\alpha_1 - \alpha_2 + \beta_2 + \frac{3}{2}) \prod_{q=1}^k \Gamma(-\alpha_q). \end{aligned}$$

Nous nous embarquons maintenant pour le calcul de $\varphi(y)$. Les matrices $y \in P_G$, c'est à dire les matrices tridiagonales et définies positives ont été fort étudiées dans la littérature à propos des fractions continues analytiques (voir par exemple Wall (1948) page 46 et suivantes). On va d'abord calculer $\varphi(y)$ dans le cas où la diagonale de y est formée de 1. L'idée inspirée de Wall (et de ses prédécesseurs Pringsheim et Van Vleck) est alors de paramétrer les $y_{q,q+1}$ par des nombres h_1, \dots, h_{k+1} tels que

$$y_{q,q+1}^2 = h_q(1 - h_{q+1}).$$

(En fait ces auteurs écrivent plutôt $y_{q,q+1}^2 = (1 - g_{q-1})g_q$). Avant de donner les détails voici une proposition technique :

Proposition 12.2. Soit h_1, \dots, h_{k+1} des nombres de $[0, 1]$ soit $\epsilon = (\epsilon_1, \dots, \epsilon_k) \in \{-1, 1\}^k$ et soit la matrice $M_\epsilon(h_1, \dots, h_{k+1})$ symétrique d'ordre $k+1$:

$$\begin{bmatrix} 1 & \epsilon_1(h_1(1-h_2))^{1/2} & 0 & \dots & 0 & 0 \\ \epsilon_1(h_1(1-h_2))^{1/2} & 1 & \epsilon_2(h_2(1-h_3))^{1/2} & \dots & 0 & 0 \\ 0 & \epsilon_2(h_2(1-h_3))^{1/2} & 1 \dots & 0 & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \epsilon_k(h_k(1-h_{k+1}))^{1/2} & 1 \end{bmatrix}. \quad (12.73)$$

Alors elle est positive et son déterminant $D(h_1, \dots, h_{k+1})$ ne dépend pas de ϵ . De plus, si $0 \leq h_j < 1$ pour tout $j = 1, \dots, k+1$ et si $p_j = h_j/(1-h_j)$ alors la matrice est définie positive et

$$\frac{D(h_1, \dots, h_{k+1})}{(1-h_1)(1-h_2)\dots(1-h_{k+1})} = \sum_{j=0}^{k+1} \prod_{i=j+1}^{k+1} p_i. \quad (12.74)$$

Finalement, si de plus $h_{k+1} = 0$ notons par x_q et $x_{q,q+1}$ les entrées (q, q) et $(q, q+1)$ de $(M_\epsilon(h_1, \dots, h_{k+1}))^{-1}$. Alors

$$D(h_1, \dots, h_k, 0) = (1 - h_1) \dots (1 - h_k) \quad (12.75)$$

$$\begin{aligned} x_q &= \frac{D(h_1, \dots, h_{q-1})}{(1 - h_1)(1 - h_2) \dots (1 - h_q)} \\ &= \frac{1}{1 - h_q} [1 + p_{q-1} + p_{q-1}p_{q-2} + \dots + p_{q-1}p_{q-2} \dots p_1] \end{aligned} \quad (12.76)$$

$$\begin{aligned} x_{q,q+1} &= -\frac{\epsilon_q(h_q(1 - h_{q+1}))^{1/2}}{(1 - h_q)(1 - h_{q+1})} [1 + p_{q-1} + p_{q-1}p_{q-2} + \dots + p_{q-1}p_{q-2} \dots p_1] \\ &= -\frac{\epsilon_q(h_q(1 - h_{q+1}))^{1/2}}{1 - h_{q+1}} x_q \end{aligned} \quad (12.77)$$

$$x_q x_{q+1} - x_{q,q+1}^2 = \frac{D(h_1, \dots, h_{q-1})D(h_1, \dots, h_{q+1})}{(1 - h_1)^2 \dots (1 - h_q)^2 (1 - h_{q+1})} \quad (12.78)$$

Démonstration. En développant le déterminant par rapport à la dernière ligne on a immédiatement la relation de récurrence

$$D(h_1, \dots, h_{k+1}) = D(h_1, \dots, h_k) - h_k(1 - h_{k+1})D(h_1, \dots, h_{k-1}). \quad (12.79)$$

Comme $D(h_1) = 1$ et $D(h_1, h_2) = 1 - h_1 + h_1 h_2$ l'indépendance vis à vis de ϵ est claire. Supposons maintenant $0 \leq h_j < 1$ pour tout $j = 1, \dots, k+1$. Alors 12.79 permet facilement de montrer par récurrence la formule 12.74. En particulier dans ce cas 12.74 montre $D(h_1, \dots, h_{k+1}) > 0$. Comme c'est vrai pour tout k , le critère classique des déterminants principaux montre que $M_\epsilon(h_1, \dots, h_{k+1})$ est définie positive. Par densité on voit que $M_\epsilon(h_1, \dots, h_{k+1})$ est positive si certains h_j sont égaux à 1.

Supposons enfin $h_{k+1} = 0$. Alors 12.79 montre par récurrence la formule 12.75. Ensuite, pour calculer x_q on prend le mineur de l'entrée (q, q) de la matrice $M_\epsilon(h_1, \dots, h_{k+1})$ et on voit qu'il est égal à $D(h_1, \dots, h_{q-1})D(h_{q+1}, \dots, h_{k+1})$. Le nombre x_q est le quotient de ce nombre par $D(h_1, \dots, h_{k+1})$. Maintenant, nous avons supposé $h_{k+1} = 0$ et donc d'après 12.75 on a

$$D(h_1, \dots, h_{k+1})/D(h_{j+1}, \dots, h_{k+1}) = (1 - h_1) \dots (1 - h_j).$$

Mettant ensemble cette dernière formule et 12.74 on a bien 12.76. On procède de même par quotient de mineur et déterminant pour avoir 12.77 puis 12.78 et la proposition est complètement montrée.

Cette proposition calculait essentiellement $x = \varphi(y)$ dans le cas particulier où les $y_q = 1$ pour tout q . En effet pour un tel $y \in P_G$ notons plutôt $z_{q,q+1} = y_{q,q+1}$. Supposons de plus que $z_{q,q+1} \neq 0$ pour tout $q = 1, \dots, k$. On définit ϵ_q

comme le signe de $z_{q,q+1}$ et on peut trouver des $(h_1, \dots, h_k) \in (0, 1)$ tels que $z_{q,q+1}^2 = h_q(1 - h_{q+1})$ avec $h_{k+1} = 0$. Il suffit pour cela de prendre

$$h_k = z_{k,k+1}^2, \quad h_{k-1} = \frac{z_{k-1,k}^2}{1 - z_{k,k+1}^2}, \quad h_{k-1} = \frac{z_{k-2,k-1}^2}{1 - \frac{z_{k-1,k}^2}{1 - z_{k,k+1}^2}} \dots$$

Nous renvoyons à Wall (1948) pages 67 -69 pour la vérification du fait que les h_q sont effectivement dans $(0, 1)$. Quand on remplace ces h_q par leur expression en fonction des $z_{q,q+1}$ dans 12.76 et 12.77 on a l'expression explicite de $\varphi(y)$ voulue (toujours sous l'hypothèse où les $y_q = 1$).

Reste à passer au cas où les y_q sont > 0 quelconques. Il suffit alors d'observer qu'alors si $D = \text{diag}((y_1)^{1/2}, \dots, (y_{q+1})^{1/2})$ alors le passage de y à $D^{-1}yD^{-1}$ ramène au cas précédent avec

$$z_{q,q+1} = \frac{y_{q,q+1}}{(y_q y_{q+1})^{1/2}}.$$

On calcule les h_q en fonction de ces $z_{q,q+1}$ et on a

$$x_q = \frac{1}{y_q} \times \frac{1}{1 - h_q} [1 + p_{q-1} + p_{q-1}p_{q-2} + \dots + p_{q-1}p_{q-2} \dots p_1] \quad (12.80)$$

$$x_{q,q+1} = -\frac{1}{(y_q y_{q+1})^{1/2}} \times \epsilon_q (h_q(1 - h_{q+1}))^{1/2} \frac{D(h_1, \dots, h_{q-1})}{(1 - h_1) \dots (1 - h_{q+1})} \quad (12.81)$$

$$\Delta_q = x_q x_{q+1} - x_{q,q+1}^2 = \frac{1}{y_q y_{q+1}} \times \frac{D(h_1, \dots, h_{q-1}) D(h_1, \dots, h_{q+1})}{(1 - h_1)^2 \dots (1 - h_q)^2 (1 - h_{q+1})} \quad (12.82)$$

Nous avons ignoré le cas où certains $y_{q,q+1} = 0$. Dans ce cas, la matrice y est formée d'un certain nombre de blocs tridiagonaux auxquels on applique la méthode précédente.

13 Les Lois G -Wishart et leur application à l'estimation.

Dans cette section, nous allons réexaminer les lois de Wishart et inverse Wishart associées à un graphe triangulé G introduites à la section 11 pour les considérer dans le cas particulier où (α, β) est une constante p (pour la première espèce) et le cas particulier où $(\alpha + \frac{c+1}{2}, \beta + \frac{s+1}{2})$ est une constante $-p < 0$ (pour la deuxième espèce). Ces cas particuliers ont des propriétés statistiques fort intéressantes.

Définition 13.1. Soit $G = (V, E)$ un graphe triangulé connexe sur $V = \{1, \dots, n\}$, soit $P = (C_1, \dots, C_k)$ un ordre parfait des cliques maximales et (S_2, \dots, S_k) les séparateurs correspondants, soit c_q et s_q les tailles de C_q et

S_q , soit $p > \max_q \frac{c_q-1}{2}$ et soit $\sigma \in Q_G$. Alors la loi G -Wishart de première espèce sur le cône de matrices incomplètes Q_G , de paramètres p et σ est la loi $\gamma_{p,p;\sigma}(dx)$

PROPRIÉTÉS DE LA G WISHART, 1ÈRE ESPÈCE On a donc

$$\gamma_{p,p;\sigma}(dx) = e^{-\text{tr}(x\hat{\sigma}^{-1})} \frac{(\det \hat{x})^p}{\Gamma_1(p,p)(\det \hat{\sigma})^p} \mu_G(dx) \quad (13.83)$$

$$= e^{-\text{tr}(x\hat{\sigma}^{-1})} \frac{\prod_{q=1}^k (\det x_{C_q})^{p-\frac{c_q+1}{2}}}{\prod_{q=2}^k (\det x_{S_q})^{p-\frac{s_q+1}{2}}} \frac{\prod_{q=2}^k \Gamma_{s_q}(p)}{(\det \hat{\sigma})^p \prod_{q=1}^k \Gamma_{s_q}(p)} \mathbf{1}_{Q_G}(x) \quad (13.84)$$

Les présentations 13.83 et 13.84 de $\gamma_{p,p;\sigma}(dx)$ résultent de l'aspect particulier que prennent les fonctions $H_G(\alpha, \beta; x)$ et $\Gamma_1(\alpha, \beta)$ quand α et β sont constants sur les cliques et les séparateurs. On peut même remarquer que si pour une partie complète S de taille s de G on note

$$f_S(x_S) = e^{-\text{tr}(x_S \sigma_S^{-1})} \left(\frac{\det x_S}{\det \sigma_S} \right)^{p-\frac{s+1}{2}} \frac{1}{\Gamma_s(p)}$$

qui est après tout la densité d'une loi de Wishart ordinaire sur le cône P_s alors on a l'étonnante expression

$$\gamma_{p,p;\sigma}(dx) = \frac{\prod_{q=1}^k f_{C_q}(x_{C_q})}{\prod_{q=2}^k f_{S_q}(x_{S_q})} \mathbf{1}_{Q_G}(x) dx. \quad (13.85)$$

La transformée de Laplace de la G Wishart de première espèce $\gamma_{p,p;\sigma}$ pour $\hat{\sigma}^{-1} - \theta \in P_G$ et d'après la section 11 est $H_G(p, p, \varphi(\hat{\sigma}^{-1} - \theta)) / H_G(p, p, \sigma)$ ou

$$L_{\gamma_{p,p;\sigma}}(\theta) = \frac{(\det \hat{\sigma})^p}{(\det(\hat{\sigma}^{-1} - \theta))^p} = \frac{1}{(\det(I_n - \theta \hat{\sigma}))^p} \quad (13.86)$$

La loi G Wishart de première espèce est la loi de l'estimateur de la projection de la covariance inconnue dans le cas d'un modèle triangulé gaussien. Plus précisément

Théorème 13.2. Soit $G = (V, E)$ un graphe triangulé connexe sur $V = \{1, \dots, n\}$, soit F l'espace des matrices symétriques incomplètes gouvernées par G . Soit N un entier supérieur ou égal à la taille de la plus grande clique maximale. Soit Z_1, \dots, Z_N des variables aléatoires de \mathbb{R}^n gaussiennes indépendantes centrées de variance inconnue Σ telle que $\Sigma^{-1} \in P_G$. Alors l'estimateur du maximum de vraisemblance de Σ est \hat{X} où X est l'élément de Q_G obtenu par projection sur F de

$$\frac{1}{N} \sum_{j=1}^N Z_j Z_j^T.$$

De plus, si Σ_0 est la vraie valeur de Σ , avec $\Sigma_0^{-1} \in P_G$, la loi de $X/2$ est alors $\gamma_{N/2, N/2; \sigma}$, où σ est la projection sur F de $2\Sigma_0$.

Nous allons maintenant enoncer les proprietes d'indépendance de l'hyper Wishart (alias G-Wishart de premiere espece).

Proposition 13.3. Soit G un graphe triangulé et soit $\mathcal{C} = \{C_1, \dots, C_k\}$ un ordre parfait des cliques maximales. Soit $\mathcal{S} = \{S_2, \dots, S_k\}$ l'ensemble correspondants des séparateurs minimaux. Soit $X \in Q_G$ une variable aléatoire qui suit une loi de densité

$$\mathcal{HW}(x; \sigma, N) = \frac{\prod_{i=1}^k w_{c_i}(x_{C_i}; \sigma_{C_i}, N)}{\prod_{i=2}^k w_{s_i}(x_{S_i}; \sigma_{S_i}, N)}, \quad (13.87)$$

où pour tout sous ensemble complet $C \subset V$ de cardinalité c

$$w_c(x_C; \sigma_C, N) = \frac{(\det x_C)^{\frac{N}{2} - \frac{c+1}{2}}}{\Gamma_c(\frac{N}{2})(\det \sigma_C)^{\frac{N}{2}}} \exp\langle x_C, \sigma_C^{-1} \rangle. \quad (13.88)$$

Alors

1. $X_{R_k \cdot S_k} \perp (X_{R_k, S_k} X_{S_k}^{-1}, X_{H_{k-1}})$,
2. $X_{R_k \cdot S_k} \sim w_{r_k}(\frac{N-s_k}{2}, \sigma_{2,1})$
3. $X_{R_k, S_k} X_{S_k}^{-1} | X_{H_{k-1}} \sim N(\sigma_{R_k, S_k} \sigma_{S_k}^{-1}, B_k)$

où B_k est l'opérateur bilinéaire

$$(u, v) \in \mathbb{R}^{R_k, S_k} \otimes \mathbb{R}^{S_k, R_k} \mapsto u X_{S_k} v \sigma_{R_k \cdot S_k}^{-1}.$$

Démonstration Comme les $\sigma_{C_j}, j = 1 \dots, k$ sont définies positives et que le graphe G est decomposable, il en résulte que $\hat{\sigma}^{-1}$ est dans P_G . Nous avons donc

$$\hat{\sigma}^{-1} = \sum_{i=1}^k \sigma_{C_i}^{-1} - \sum_{i=2}^k \sigma_{S_i}^{-1}.$$

et

$$\frac{\prod_{i=1}^k \exp -\frac{1}{2} \langle x_{C_i}, \sigma_{C_i}^{-1} \rangle}{\prod_{i=2}^k \exp -\frac{1}{2} \langle x_{S_i}, \sigma_{S_i}^{-1} \rangle} = \exp -\frac{1}{2} \left(\langle x_{H_{k-1}}, \sigma_{H_{k-1}}^{-1} \rangle + \langle x_{C_k}, \sigma_{C_k}^{-1} \rangle - \langle x_{S_k}, \sigma_{S_k}^{-1} \rangle \right). \quad (13.89)$$

Nous decomposons maintenant les matrices x_{C_k} and σ_{C_k} en blocks correspondant a S_k and $R_k = C_k \setminus S_k$ et que nous notons respectivement "1" and "2" pour alléger les notations. Nous utilisons la decomposition de Cholesky pour chacune de ces matrices et nous avons donc

$$x = \begin{pmatrix} x_1 & x_{12} \\ x_{21} & x_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ x_{21} x_1^{-1} & 1 \end{pmatrix} \begin{pmatrix} x_1 & 0 \\ 0 & x_{2,1} \end{pmatrix} \begin{pmatrix} 1 & x_1^{-1} x_{12} \\ 0 & 1 \end{pmatrix},$$

et nous avons une decomposition similaire pour Σ . De simples multiplications de matrices donnent alors

$$\langle x_{C_k}, \sigma_{C_k}^{-1} \rangle - \langle x_{S_k}, \sigma_{S_k}^{-1} \rangle = \langle x_{2,1}, \sigma_{2,1}^{-1} \rangle + \langle (x_{21} x_1^{-1} - \sigma_{21} \sigma_1^{-1}), x_1 (x_1^{-1} x_{12} - \sigma_1^{-1} \sigma_{12}) \sigma_{2,1}^{-1} \rangle \quad (13.90)$$

Nous decomposons aussi les determinants de la facon suivante

$$\begin{aligned}
& \frac{\prod_{i=1}^k (\det x_{C_i})^{\frac{N}{2} - \frac{c_i+1}{2}}}{\prod_{i=2}^k (\det x_{S_i})^{\frac{N}{2} - \frac{s_i+1}{2}}} \\
&= \frac{\prod_{i=1}^{k-1} (\det x_{C_i})^{\frac{N}{2} - \frac{c_i+1}{2}} (\det x_{C_k})^{\frac{N}{2} - \frac{c_k+1}{2}}}{\prod_{i=2}^{k-1} (\det x_{S_i})^{\frac{N}{2} - \frac{s_i+1}{2}} (\det x_{S_k})^{\frac{N}{2} - \frac{c_k+1}{2}}} \\
&= \frac{\prod_{i=1}^{k-1} (\det x_{C_i})^{\frac{N}{2} - \frac{c_i+1}{2}}}{\prod_{i=2}^{k-1} (\det x_{S_i})^{\frac{N}{2} - \frac{s_i+1}{2}}} (\det x_{2,1})^{\frac{N-s_k}{2} - \frac{r_k+1}{2}} (\det x_1)^{-\frac{c_k-s_k}{2}} \quad (13.91)
\end{aligned}$$

Nous decomposons maintenant la constante de normalisation.

$$\begin{aligned}
& \frac{\prod_{i=1}^k \Gamma_{c_i}(\frac{N}{2})(\det \sigma_{C_i})^{\frac{N}{2}}}{\prod_{i=2}^k \Gamma_{s_i}(\frac{N}{2})(\det \sigma_{S_i})^{\frac{N}{2}}} \\
&= \frac{\prod_{i=1}^{k-1} \Gamma_{c_i}(\frac{N}{2})(\det \sigma_{C_i})^{\frac{N}{2}}}{\prod_{i=2}^{k-1} \Gamma_{s_i}(\frac{N}{2})(\det \sigma_{S_i})^{\frac{N}{2}}} \frac{1}{\pi^{\frac{r_k s_k N}{2}} \Gamma_{c_k}(\frac{N-s_k}{2}) \sigma_{2,1}^{\frac{N-s_k}{2}} \sigma_{2,1}^{\frac{s_k}{2}}} \quad (13.92)
\end{aligned}$$

Faisons maintenant le changement de variable

$$x \in Q_G \mapsto (x_{H_{k-1}}, x_{2,1} x_1^{-1}, x_1)$$

dont le Jacobien est $(\det x_1)^{c_k - s_k}$. En utilisant 13.89, 13.90, 13.92, 13.91 et le Jacobien ci dessus, nous obtenons

$$\begin{aligned}
\mathcal{HW}(x; \sigma, N) dx &= \frac{\prod_{i=1}^k w_{c_i}(x_{C_i}; \sigma_{C_i}, N)}{\prod_{i=2}^k w_{s_i}(x_{S_i}; \sigma_{S_i}, N)} \\
&\times \mathcal{HW}(x_{H_{k-1}}; \sigma_{H_{k-1}}, N) \frac{(\det x_{2,1})^{\frac{N-s_k}{2} - \frac{r_k+1}{2}}}{\Gamma_{c_k}(\frac{N-s_k}{2}) \sigma_{2,1}^{\frac{N-s_k}{2}}} \exp - \langle x_{2,1}, \sigma_{2,1}^{-1} \rangle \\
&\times \frac{(\det x_1)^{\frac{c_k-s_k}{2}}}{\pi^{\frac{r_k s_k N}{2}} \sigma_{2,1}^{\frac{s_k}{2}}} \exp - \langle (x_{2,1} x_1^{-1} - \sigma_{2,1} \sigma_1^{-1}), x_1 (x_1^{-1} x_{12} - \sigma_1^{-1} \sigma_{12}) \sigma_{2,1}^{-1} \rangle \\
&\times dx_{H_{k-1}} dx_1 dx_{2,1} . \quad (13.93)
\end{aligned}$$

Nous voyons donc clairement que $x_{2,1} = x_{R_k \cdot S_k}$ et $x_{2,1} x_1^{-1} = x_{R_k S_k} x_{S_k}^{-1}$ etant donne $x_{H_{k-1}}$ suivent bien les distributions indiquees dans l'enonce et que $x_{R_k \cdot S_k}$ est independant de $(x_{R_k S_k} x_{S_k}^{-1}, x_{H_{k-1}})$. Nous pouvons evidemment repeter pour n'importe quel $j = 2, \dots, k-1$ l'argument que nous avons donne pour k et il s'en suit que pour tout $j = 2, \dots, k$, nous avons que $x_{R_j \cdot S_j}$ est independant de $(x_{R_j S_j} x_{S_j}^{-1}, x_{H_{j-1}})$.

Proposition 13.4. Soit $P = (C_1, \dots, C_k)$ un ordre parfait des cliques maximales du graphe triangulé G , soit $p > 0$ et soit $(\alpha, \beta) = (-p - \frac{c+1}{2}, -\delta - \frac{s+1}{2})$.

Soit $\theta \in Q_G$ et soit X une variable aléatoire sur Q_G de loi inverse Wishart $IW_{\alpha, \beta; \theta}$. On note par X_2 la restriction de X au séparateur $S = S_k$, on note par

$$X_K = \begin{bmatrix} X_2 & X_{23} \\ X_{32} & X_3 \end{bmatrix}, \quad X_M = \begin{bmatrix} X_1 & X_{12} \\ X_{21} & X_2 \end{bmatrix}$$

les restrictions respectives de X à $K = C_k$ et $M = C_1 \cup \dots \cup C_{k-1}$, on note $X'_3 = X_3 - X_{32}X_2^{-1}X_{23}$ et $X'_{32} = X_{32}X_2^{-1}$. Alors X_M et (X'_{32}, X'_3) sont indépendantes. De plus, X_M suit une loi inverse Wishart de paramètres $(\alpha_M, \beta_M; \theta_M)$. En convenant $\theta'_3 = \theta_3 - \theta_{32}\theta_2^{-1}\theta_{23}$ on a que X'_3 suit la loi inverse Wishart ordinaire sur le cône $P_{c_k - s_k}$ des matrices symétriques définies positives d'ordre $|K| - |S| = c_k - s_k$

$$\frac{1}{\Gamma_{c_k - s_k}(p + \frac{c_k + 1}{2})} (\det x'_3)^{-p - \frac{c_k + 1}{2} - \frac{c_k - s_k + 1}{2}} e^{-\text{tr}(\theta'_3(x'_3)^{-1})} \mathbf{1}_{P_{c_k - s_k}}(x'_3) dx'_3.$$

Enfin, la loi de X'_{32} conditionnée par X'_3 est une loi normale sur $L = L(\mathbb{R}^S, \mathbb{R}^{K \setminus S})$ de covariance $x'_{23} \mapsto \frac{1}{2}\theta_2^{-1}x'_{23}$ et de moyenne $\theta_2^{-1}\theta_{23}(X'_3)^{-1}$.

Démonstration. Elle est très proche de celle de la Proposition 11.2, puisqu'il suffit de calculer la loi de (X_M, X'_{32}, X'_3) . Nous conservons à dessein la notation (α, β) pour que le lecteur réalise exactement à quel moment l'hypothèse $(-p - \frac{c+1}{2}, -p - \frac{s+1}{2}) = (\alpha, \beta)$ devient nécessaire. La loi de (X_M, X'_{32}, X'_3) est, avec une constante multiplicative K

$$\begin{aligned} & K e^{-\text{tr}(\theta_M(x_M)^{-1})} H_{G_M}(\alpha_M, \beta_M; x_M) \mu_{G_M}(dx_M) \\ & e^{-\text{tr}(\theta_2 x'_{23}(x'_3)^{-1} x'_{32} - \text{tr}(\theta_3(x'_3)^{-1}) + 2\text{tr}(\theta_{32} x'_{23}(x'_3)^{-1})} \\ & (\det x'_3)^{\alpha_k - \frac{c_k + 1}{2}} (\det x_2)^{\alpha_k - \beta_k + \frac{c_k - s_k}{2}} dx'_{23} dx'_3. \end{aligned} \quad (13.94)$$

Pour écrire cela nous avons utilisé les formules 10.49, 11.64, 11.65, 11.66 et 11.67. Ensuite, le choix particulier de (α, β) permet de tuer l'exposant de $\det x_2$ et d'obtenir l'indépendance de X_M avec (X'_{32}, X'_3) . Pour trouver la loi marginale de X'_3 il suffit d'intégrer par rapport à x'_{23} sur tout $\mathbb{R}^{|K| - |S|}$ la densité de la loi jointe de (X'_{32}, X'_3) (proportionnelle au produit des deux dernières lignes de 13.94) en s'aidant de la Proposition 10.2, et nous sautons ce calcul fastidieux. Les constantes d'intégration sont données par la Proposition 10.3. On en déduit alors facilement que la loi conditionnelle de X'_{23} sachant X'_3 est une loi gaussienne, dont les paramètres se calculent à l'aide de la troisième partie de la Proposition 10.2.

Corollaire 13.5. Sous les hypothèses de la Proposition 13.4, pour $q = 2, \dots, k$ notons $R_q = C_q \setminus S_q$. Si A et B sont des ensembles complets de sommets on note X_A et $X_{A,B}$ les restrictions de X à $A \times A$ et à $A \times B$. Notons aussi

$$X'_{R_q} = X_{R_q} - X_{R_q, S_q} X_{S_q}^{-1} X_{S_q, R_q}, \quad X'_{R_q, S_q} = X_{R_q, S_q} X_{S_q}^{-1}.$$

Alors les variables aléatoires

$$X_{C_1}, (X'_{R_2}, X'_{R_2, S_2}), (X'_{R_3}, X'_{R_3, S_3}), \dots, (X'_{R_k}, X'_{R_k, S_k})$$

sont indépendantes. De plus en convenant $\theta'_{R_q} = \theta_{R_q} - \theta_{R_q, S_q} \theta_{S_q}^{-1} \theta_{S_q, R_q}$ alors X'_{R_q} suit une loi inverse Wishart ordinaire de paramètres θ'_{R_q} et $c_q - s_q$ et X_{C_q} suit une loi inverse Wishart ordinaire de paramètres θ_{C_q} et c_q . Enfin, conditionnée par X'_{R_q} , la loi de X'_{S_q, R_q} est gaussienne de covariance $x_{S_q, R_q} \mapsto \frac{1}{2} \theta_{S_q}^{-1} x_{S_q, R_q}$ et de moyenne $\theta_{S_q}^{-1} \theta_{S_q, R_q} (X'_{R_q})^{-1}$.

Démonstration. Par récurrence sur k à partir de la Proposition 13.4. Remarquons que cette récurrence part de $k = 1$ et fournit que X_{C_1} suit une loi inverse Wishart ordinaire de paramètres θ_{C_1} et c_1 . Pour obtenir que X_{C_q} suit une loi inverse Wishart ordinaire de paramètres θ_{C_q} et c_q , il faut se rappeler un détail concernant les ordres parfaits sur \mathcal{C} : n'importe quel élément de \mathcal{C} peut être pris comme élément initial d'un ordre parfait (voir fin du Théorème 7.2).

.....

14 Estimation pour un modèle discret triangulé.

Rappelons qu'un modèle graphique discret est la donnée d'un graphe (V, E) (fini, et non orienté), et d'une famille d'ensembles finis $(A_v)_{v \in V}$. On note par \mathcal{C} l'ensemble des cliques maximales de ce graphe : c'est toujours une famille de Sperner. Le modèle est alors l'ensemble des probabilités $p = (p(a))_{a \in A}$ sur $A = \times_{v \in V} A_v$ telles que pour tout $a \in A$ on ait $p(a) > 0$ et telles que $\log p$ est une combinaison linéaire de fonctions f_C indexées par \mathcal{C} où $a = (a_v)_{v \in V} \mapsto f_C(a)$ ne dépend que des a_v tels que $v \in C$. Il s'agit d'une famille exponentielle *générale*. Comme on le sait, l'estimation de p par maximum de vraisemblance est standard dans le cas d'un modèle saturé, c'est à dire si le graphe était complet. Elle n'est pas résoluble explicitement en général, mais nous aurons une solution explicite si le graphe est triangulé.

Commençons par réviser l'estimation des paramètres dans les cas d'une loi binomiale ou multinomiale sous l'angle des familles exponentielles.

Loi binomiale. Soit

$$B(N, p) = \sum_{k=0}^N C_N^k p^k (1-p)^{N-k} \delta_k$$

la loi binomiale où N est fixé et où p est inconnu. Si on pose $e^\theta = \frac{p}{1-p}$ on en déduit que $p = \frac{e^\theta}{1+e^\theta}$, que $1-p = \frac{1}{1+e^\theta}$ et donc que

$$C_N^k p^k (1-p)^{N-k} = e^{k\theta - N \log(1+e^\theta)}.$$

Autrement dit, le modèle $(B(N, p); p \in]0, 1[)$ est une famille exponentielle naturelle engendrée par la mesure $\mu = \sum_{k=0}^N C_N^k \delta_k$ dont la fonction de cumulants est $k_\mu(\theta) = N \log(1+e^\theta)$ et où $\Theta(\mu) = \mathbb{R}$. Si k est l'observation et si $0 < k < N$, le maximum de vraisemblance est atteint pour $\hat{p} = \frac{e^{\hat{\theta}}}{1+e^{\hat{\theta}}} = \frac{k}{N}$.

Loi multinomiale. Soit A un entier > 0 . On se donne d'abord une probabilité $p = (p_0, \dots, p_A)$ sur $\{0, \dots, A\}$. Soit (e_0, \dots, e_A) la base canonique de \mathbb{R}^{1+A} . Soit X une variable aléatoire de \mathbb{R}^{1+A} telle que $\Pr(X = e_a) = p_a$. Soit X_1, \dots, X_N des v.a. indépendantes et de même loi que X et soit $S_N = X_1 + \dots + X_N$. Si $S_N = (s_a)_{a=0}^A$, avec $s_0 + \dots + s_A = N$ et $s_a \in \mathbb{N}$, c'est dire que en N expériences, on a obtenu s_a fois le résultat a . S_N suit alors la loi multinomiale

$$\Pr(S_N = (s_a)_{a=0}^A) = \frac{N!}{\prod_{a=0}^A s_a!} \prod_{a=0}^A p_a^{s_a}.$$

Le vocabulaire traditionnel recèle un piège, car la loi binomiale n'est pas le cas particulier de la loi multinomiale correspondant à $A = 1$, du moins pas tout à fait : si $A = 1$ la loi multinomiale est concentrée sur un segment de \mathbb{R}^2 et la loi binomiale sera sa projection sur \mathbb{R} . L'inconvénient de la loi multinomiale telle que nous l'avons définie est qu'elle est concentrée sur l'hyperplan affine de \mathbb{R}^{1+A} défini par $s_0 + \dots + s_A = N$ ce qui gêne l'application des théorèmes standard sur les familles exponentielles naturelles. Pour pallier cet inconvénient, on fait la projection π sur \mathbb{R}^A définie par $\pi : (s_a)_{a=0}^A \mapsto (s_a)_{a=1}^A$. Aucune information n'est perdue, mais on a quand même dissymétrisé les observations en faisant jouer un rôle particulier à celle qui est numérotée par 0. La loi de $\pi(S_N)$ est alors

$$\Pr(\pi(S_N) = (s_a)_{a=1}^A) = \frac{N!}{\prod_{a=0}^A s_a!} p_1^{s_1} \dots p_A^{s_A} (1 - p_1 - \dots - p_A)^{N - s_1 - \dots - s_N}.$$

Posons alors $e^{\theta_j} = \frac{p_j}{1 - p_1 - \dots - p_A}$ pour $j = 1, \dots, A$. Alors

$$p_j = \frac{e^{\theta_j}}{1 + e^{\theta_1} + \dots + e^{\theta_A}}, \tag{14.95}$$

$$1 - p_1 - \dots - p_A = \frac{1}{1 + e^{\theta_1} + \dots + e^{\theta_A}}.$$

La loi de $\pi(S_N)$ prend alors la forme d'une famille exponentielle naturelle, ce qui était peu apparent avant :

$$\Pr(\pi(S_N) = (s_a)_{a=1}^A) = \frac{N!}{\prod_{a=0}^A s_a!} e^{\theta_1 s_1 + \dots + \theta_A s_A - N \log(1 + e^{\theta_1} + \dots + e^{\theta_A})}.$$

Nous laissons au lecteur le soin de décrire une mesure génératrice μ correspondant à la fonction de cumulants $k_\mu(\theta) = N \log(1 + e^{\theta_1} + \dots + e^{\theta_A})$ qui est définie sur $\Theta(\mu) = \mathbb{R}^A$. Si maintenant N est fixé et si p est inconnu, on

applique facilement la méthode du maximum de vraisemblance pour estimer les θ et p : si $(s_a)_{a=1}^A$ est l'observation, alors le maximum de vraisemblance est atteint pour

$$\hat{p}_a = \frac{e^{\hat{\theta}_a}}{1 + e^{\hat{\theta}_1} + \dots + e^{\hat{\theta}_A}} = \frac{s_a}{N}.$$

Modèle de Poisson et modèle multinomial. Il est parfois commode pour réaliser une loi multinomiale d'utiliser la proposition classique, qu'il est facile de montrer :

Proposition 14.1. Soit $\lambda(a) > 0$ pour $a = 0, \dots, A$. On note $\lambda = \sum_{a=0}^A \lambda(a)$ et $p_a = \lambda(a)/\lambda$. Soit $(N_a)_{a=0}^A$ une suite de variables aléatoires indépendantes de Poisson de moyennes respectives $\lambda(a)$ et soit $N = \sum_{a=0}^A N_a$. Alors la loi conditionnelle du vecteur $(N_a)_{a=0}^A$ sachant $N = n$ est une loi multinomiale de paramètres n et $p = (p_a)_{a=0}^A$.

Theorème 14.2. Soit A un ensemble fini et la mesure μ sur \mathbb{R}^A définie par

$$\mu(dx) = \sum_{x \in \mathbb{N}^A} \frac{1}{\prod_{a \in A} x_a!} \delta_x.$$

Soit $F = F(\mu)$ la famille exponentielle naturelle engendrée des lois de Poisson. Soit H un sous espace affine de \mathbb{R}^A . Alors le maximum de vraisemblance pour la famille restreinte

$$F_H = \{P(\theta, \mu); \theta \in H\}$$

existe et est unique pour une observation $(s_a)_{a \in A}$ telle que $s_a > 0$ pour tout a de A .

Démonstration L'unicité découle de la théorie faite à la section 5, Proposition 5.1. Pour l'existence, il est évident que $k_\mu(\theta) = \sum_a e^{\theta_a}$. Il s'agit donc de montrer que la fonction suivante

$$\theta \mapsto F(\theta) = \sum_a (s_a \theta_a - e^{\theta_a}),$$

restreinte à la variété affine H , y atteint son maximum. Pour cela, si $s > 0$, on trace d'abord le graphe de la fonction sur \mathbb{R} définie par $f_s(x) = sx - e^x$ et on constate qu'elle tend vers $-\infty$ si $|x| \rightarrow \infty$. On en déduit que la fonction continue

$$F(\theta) = \sum_a f_{s_a}(\theta_a)$$

tend vers $-\infty$ si $\|\theta\| = \sum_a |\theta_a| \rightarrow \infty$. Il en est de même quand on la restreint à H et elle atteint donc son maximum sur H et le théorème est montré.

Analysons en les conséquences à la lumière de la section 5. Pour simplifier, supposons que l'espace affine H ci dessus soit un sous espace vectoriel. Notons par m_a la moyenne inconnue de la variable de Poisson X_a . Comme la famille

exponentielle des (X_a) est gouvernée par la fonction de cumulant $k_\mu(\theta) = \sum_{a \in A} e^{\theta_a}$, de différentielle

$$k'_\mu(\theta) = (e^{\theta_a})_{a \in A} = (m_a)_{a \in A},$$

on voit que si t est la projection de \mathbb{R}^A sur H et si $q = t(s)$, où $s = (s_a)_{a \in A}$ est l'observation, alors l'égalité $k'_\mu(\theta) = q + u_q$ de 5.17) entraîne que $t(m) = t(s)$, ce qu'on résume par :

Pour le modèle de Poisson restreint au sous espace H , *Projection sur H de l'estimé des moyennes = Projection sur H du vecteur des observations.*

Attirons maintenant l'attention sur le fait que dans le modèle multinomial surveillé sur un sous espace affine des θ (baptisé *modèle log affine* dans la littérature) la même formule a lieu si on change un peu le paramètre inconnu $p = (p_a)_{a \in A}$ pour $m_a = Np_a$. En effet, on a vu que la famille exponentielle des (X_a) est gouvernée par la fonction de cumulant $k_\mu(\theta) = N \log(1 + \sum_{a \in A} e^{\theta_a})$, de différentielle

$$k'_\mu(\theta) = \frac{N}{1 + \sum_{a \in A} e^{\theta_a}} (e^{\theta_a})_{a \in A} = (m_a)_{a \in A} = (m_a)_{a \in A}.$$

Si $q = t(s)$, où $s = (s_a)_{a \in A}$ est l'observation, alors l'égalité $k'_\mu(\theta) = q + u_q$ de 5.17) entraîne que $t(s) = t(m)$ d'où :

Pour le modèle multinomial logaffine, N fois *Projection sur H de l'estimé des p = Projection sur H de l'estimé des m = Projection sur H du vecteur des observations.*

Modèle hiérarchique. On se rappelle qu'un modèle hiérarchique est gouverné par un produit d'ensembles finis $A = \times_{v \in V} A_v$ et par une famille de Sperner $\{V_1, \dots, V_k\}$, c'est à dire des parties non vides de V telles que $V_i \subset V_j$ entraîne $i = j$. Notons par M l'espace vectoriel $M = E_{V_1} + \dots + E_{V_k}$ défini à la proposition 3.1, 3.10). Le modèle hiérarchique est l'ensemble des probabilités p sur A telles que $\log p \in M$ (voir 3.8). C'est donc un sous modèle du modèle multinomial. On a vu que celui ci est un modèle exponentiel, pourvu qu'on y distingue un point 0 particulier. On a vu aussi que les modèles exponentiels, si on les surveille sur une variété affine, ou mieux, un sous espace vectoriel, peuvent parfois encore se prêter au calcul du maximum de vraisemblance.

Il y a toutefois un point à éclaircir : le lien entre le point particulier et la variété affine. Dans l'ensemble A fixons un point noté 0, et soit $A^* = A \setminus \{0\}$. Notons par H le sous espace vectoriel de M formé par les f tels que $f(0) = 0$. On paramètre par les $(\theta(a))_{a \in A^*}$ et il est commode de convenir $\theta(0) = 0$. Alors 14.95) s'écrit pour tout $a \in A^*$

$$p(a) = \frac{e^{\theta(a)}}{\sum_{b \in A} e^{\theta(b)}},$$

$$\theta(a) = \log p(a) - \log(1 - \sum_{b \in A^*} p(b)) = \log p(a) - \log p(0).$$

Il est donc clair que $\log p \in M \Leftrightarrow (\theta(a))_{a \in A} \in H$.

Compte tenu de la théorie développée à la section 5, l'estimation de p ou de θ après N observations se fait ainsi :

- Nous observons le tableau de contingence $(s(a))_{a \in A}$ ou $(s(a))_{a \in A^*}$.
- Nous le projetons orthogonalement sur H , obtenant ainsi un point $q = (q(a))_{a \in A}$ (rappel $q(0) = 0$).
- Nous recherchons $u(q) \in H^\perp$ tel que $k'_\mu(\theta) = q + u(q)$ ait une solution dans H , c'est à dire telle que pour tout $a \in A^*$ on ait

$$\frac{N}{\sum_{b \in A} e^{\theta(b)}} e^{\theta(a)} = q(a) + u(q)(a). \quad (14.96)$$

Je laisse de côté la question de l'existence et du domaine des moyennes de la famille exponentielle associée concentrée sur H . La difficulté est dans le troisième point du programme ci dessus. Le cas du modèle hiérarchique, ou même du modèle graphique général relève de l'analyse numérique (algorithme de Deming et Stefan : voir Lauritzen (1996)). Seul le cas triangulé est susceptible d'une solution explicite.

Pour cela, nous démontrons un résultat important sur les modèles hiérarchiques . Il exige encore quelques définitions.

Définitions. Soit V un ensemble fini et \mathcal{X} et \mathcal{Y} des familles de Sperner de V .

- Si \mathcal{A} est une famille de parties de A alors $\text{Sperner}(\mathcal{A})$ est la famille de ses éléments maximaux.
- $\mathcal{X} \vee \mathcal{Y}$ est Sperner $(\mathcal{X} \cup \mathcal{Y})$.
- $\mathcal{X} \wedge \mathcal{Y}$ est Sperner $\{X \cap Y; X \in \mathcal{X}, Y \in \mathcal{Y}\}$.
- On dit que \mathcal{X} et \mathcal{Y} sont *en somme directe* si $\mathcal{X} \wedge \mathcal{Y}$ est réduit à un seul élément.

Proposition 14.3 Si les familles de Sperner \mathcal{X} et \mathcal{Y} sont en somme directe, alors

$$\mathcal{X} \wedge \mathcal{Y} = \{(\cup_{X \in \mathcal{X}} X) \cap (\cup_{Y \in \mathcal{Y}} Y)\}.$$

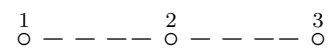
Démonstration : exercice.

Théorème 14.4 Soit deux modèles hiérarchiques sur $A = \times_{v \in V} A_v$ gouvernés par les familles de Sperner sur V notés \mathcal{X} et \mathcal{Y} . On suppose que ces familles de Sperner sont en somme directe. On note par $s = (s_a)_{a \in A}$ le vecteur des N observations et $m_a = Np_a$ où p est la probabilité inconnue sur A . On note par \hat{m} , \hat{m}_1 et \hat{m}_2 les estimés respectifs de m suivant les modèles hiérarchiques gouvernés par $\mathcal{X} \vee \mathcal{Y}$, \mathcal{X} et \mathcal{Y} respectivement. On note $V_1 = \cup_{X \in \mathcal{X}} X$, $V_2 = \cup_{Y \in \mathcal{Y}} Y$ et $V_3 = V_1 \cap V_2$. Alors

$$\hat{m}_a = \frac{\hat{m}_1(a_{V_1}) \hat{m}_2(a_{V_2})}{s(a_{V_3})}.$$

(notation : si $C \subset V$ et $a \in A$, $s(a_C) = \sum\{s_b; b_v = a_v \forall v \in C\}$).

Modèle graphique discret triangulé. Nous allons d'abord examiner le graphe suivant



Appendice 1. Examen du cours de DEA "Modèles graphiques, 8 février 2001

Exercice A. Soient $X_1, Y_1, \dots, X_N, Y_N$ des variables aléatoires indépendantes, de lois respectives pour les X et les Y :

$$ae^{-ax} \mathbf{1}_{]0, \infty[}(x) dx, \quad be^{-by} \mathbf{1}_{]0, \infty[}(y) dy.$$

On suppose que a et b sont des paramètres inconnus positifs *constraints* par $a + b = 1$, et on dispose pour les estimer des observations $X_1, Y_1, \dots, X_N, Y_N$. On posera pour simplifier $\bar{X} = \frac{1}{N}(X_1 + \dots + X_N)$, $\bar{Y} = \frac{1}{N}(Y_1 + \dots + Y_N)$ et $\bar{Z} = \frac{1}{2}(\bar{X} - \bar{Y})$.

1. Calculer le maximum de vraisemblance de a en utilisant simplement sa définition (on obtiendra une équation du second degré dont on choisira la racine convenable).
2. On reprend le problème précédent en posant $a = \frac{1}{2}(1 + \theta)$ et $b = \frac{1}{2}(1 - \theta)$. Interpréter le problème à l'aide d'une famille exponentielle naturelle sur \mathbb{R}^2 euclidien, où le paramètre est restreint au sous espace $H = \{(\theta_1, \theta_2) \in \mathbb{R}^2; \theta_1 + \theta_2 = 0\}$. Précisez les objets classiques de la famille exponentielle générale induite (soit Ω, ν, t, E, μ).
3. On identifie la famille exponentielle naturelle associée sur H à une famille sur \mathbb{R} (par $(x, y) \mapsto x$ si $x + y = 0$). Précisez ses objets classiques : $\mu, \Theta(\mu), k_\mu, M_F, \psi_\mu$. A l'aide de ces résultats et du cours, retrouver les résultats de 1).

Exercice B. (Il est préférable d'étudier l'exercice A avant) Soient $X_1, Y_1, Z_1, \dots, X_N, Y_N, Z_N$ des variables aléatoires indépendantes, de lois respectives :

$$ae^{-ax} \mathbf{1}_{]0, \infty[}(x) dx, \quad be^{-by} \mathbf{1}_{]0, \infty[}(y) dy, \quad ce^{-cz} \mathbf{1}_{]0, \infty[}(z) dz.$$

On suppose que a, b et c sont des paramètres inconnus positifs *constraints* par $a + b + c = 1$, qu'on veut estimer par la méthode du maximum de vraisemblance. Pour cela on pose $a = \frac{1}{3}(1 + \theta_1)$, $b = \frac{1}{3}(1 + \theta_2)$ et $c = \frac{1}{3}(1 + \theta_3)$ et on introduit le plan vectoriel

$$H = \{(\theta_1, \theta_2, \theta_3) \in \mathbb{R}^3; \theta_1 + \theta_2 + \theta_3 = 0\}$$

de l'espace euclidien canonique \mathbb{R}^3 .

1. Donner un vecteur \vec{n} orthogonal à H qui soit de norme 1 et en déduire la projection orthogonale sur H d'un vecteur $\vec{v} = (x, y, z)$, soit $\vec{v} - \langle \vec{v}, \vec{n} \rangle \vec{n}$.
2. Si $\vec{q} = (q_1, q_2, q_3) \in H$, montrer que l'ensemble des $u \in \mathbb{R}$ tels qu'on puisse trouver $(\theta_1, \theta_2, \theta_3) \in H$ avec

$$(q_1 + u, q_2 + u, q_3 + u) = \left(\frac{1}{1 - \theta_1}, \frac{1}{1 - \theta_2}, \frac{1}{1 - \theta_3} \right)$$

est l'ensemble des solutions d'une équation de degré 3 qu'on ne cherchera pas à résoudre explicitement.

3. A l'aide du cours et des questions qui précèdent, expliquer comment procéder à l'estimation des paramètres (a, b, c) (plus les détails sur les objets classiques des familles exponentielles impliquées sont fournis, meilleure est la note donnée à cette question).

Exercice C. Soit le graphe :



Est-il triangulé? Quelles sont ses cliques maximales, ses séparateurs et leurs multiplicités? Trouver un arbre binaire de décomposition. On considère ensuite un modèle graphique discret (multinomial) sur $A = A_0 \times A_1 \times A_2 \times A_3$, avec $A_j = \{0, 1\}$, gouverné par le graphe précédent. On fait 50 expériences qui donnent les observations suivantes $(s(a))_{a \in A}$, avec $s(a) = s(a_0, a_1, a_2, a_3)$ et $a_j = 0$ ou 1, résumées en 4 tableaux :

Si $a_0 = a_3 = 0$:	$\begin{array}{cc cc} a_2 \setminus a_1 & 0 & 1 & \\ \hline 0 & 3 & 2 & \\ 1 & 2 & 1 & \end{array}$, si $a_0 = 1, a_3 = 0$:	$\begin{array}{cc cc} a_2 \setminus a_1 & 0 & 1 & \\ \hline 0 & 4 & 5 & \\ 1 & 9 & 2 & \end{array}$
si $a_0 = 0, a_3 = 1$:	$\begin{array}{cc cc} a_2 \setminus a_1 & 0 & 1 & \\ \hline 0 & 4 & 2 & \\ 1 & 2 & 4 & \end{array}$, si $a_0 = a_3 = 1$:	$\begin{array}{cc cc} a_2 \setminus a_1 & 0 & 1 & \\ \hline 0 & 6 & 1 & \\ 1 & 2 & 1 & \end{array}$

Estimer par maximum de vraisemblance la valeur de la probabilité inconnue $p(a)$ sur A pour les 3 valeurs suivantes de a : 0000, 0110 et 1111.

Exercice D. (Il est préférable d'étudier l'exercice C avant) Soit (X_0, X_1, X_2, X_3) une variable aléatoire gaussienne de \mathbb{R}^4 , possédant une densité, centrée, de matrice de covariance inconnue Σ . On suppose que

- X_0 et X_3 sont conditionnellement indépendantes sachant (X_1, X_2) ;
- X_0 et X_2 sont conditionnellement indépendantes sachant (X_1, X_3) ;
- X_1 et X_3 sont conditionnellement indépendantes sachant (X_0, X_2) .

On fait 10 observations qui donnent la covariance empirique suivante

$$S_{10} = \frac{1}{10} \sum_{j=1}^{10} X^j (X^j)^T = \begin{bmatrix} 5,2 & -0,4 & 1,3 & 1,2 \\ -0,4 & 6 & -0,6 & 1,4 \\ 1,3 & -0,6 & 4,2 & 0,8 \\ 1,2 & 1,4 & 0,8 & 3,2 \end{bmatrix}.$$

Expliquer comment calculer l'estimateur du maximum de vraisemblance de Σ .

Exercice E. Soit $a > 0$ fixé. On considère la fonction sur $] - 1, 1[$ définie par $f(z) = e^{a \arcsin z}$. Montrer que $(1 - z^2)f''(z) - zf'(z) - a^2f(z) = 0$ et en déduire que les nombres $p_k(a)$ définis par

$$f(z) = \sum_{k=0}^{\infty} \frac{p_k(a)}{k!} z^k$$

sont positifs. On considère alors la mesure sur \mathbb{R} définie par

$$\mu(dx) = \sum_{k=0}^{\infty} \frac{p_k(a)}{k!} \delta_k(dx).$$

Calculer les objets classiques de la famille exponentielle naturelle $F(\mu)$.

Appendice 3 : Droites extrémales de certains cones convexes

On note par S l'ensemble des matrices symétriques réelles d'ordre 4, et par Z l'ensemble des éléments de S de la forme

$$\begin{bmatrix} x_1 & a & 0 & d \\ a & x_2 & b & 0 \\ 0 & b & x_3 & c \\ d & 0 & c & x_4 \end{bmatrix}.$$

On note par I l'ensemble des matrices symétriques incomplètes de la forme

$$\begin{bmatrix} x_1 & a & & d \\ a & x_2 & b & \\ & b & x_3 & c \\ d & & c & x_4 \end{bmatrix}.$$

On note par p la projection canonique de S sur I . On note par P_4 le cône convexe ouvert des éléments de S qui sont définis positifs. On note $P = P_4 \cap Z$ et $Q = p(P_4)$. On note par $\overline{P_4}$ le cône convexe fermé des éléments de S qui sont positifs. On note $\overline{P} = \overline{P_4} \cap Z$ et $\overline{Q} = p(\overline{P_4})$. Nous voulons trouver les droites extrémales des deux cônes fermés \overline{P} et \overline{Q} .

Il est utile pour cela de considérer les sous ensembles P_1 et Q_1 de \overline{P} et \overline{Q} de diagonale $(1, 1, 1, 1)$ en les paramétrant avec 4 nombres $\alpha, \beta, \gamma, \delta \geq 0$ et 4 signes $\epsilon = (\epsilon_a, \dots, \epsilon_d)$ de la façon suivante

$$x = x(\epsilon; \alpha, \beta, \gamma, \delta) = \begin{bmatrix} 1 & \epsilon_a \sqrt{\alpha} & & \epsilon_d \sqrt{\delta} \\ \epsilon_a \sqrt{\alpha} & 1 & \epsilon_b \sqrt{\beta} & \\ & \epsilon_b \sqrt{\beta} & 1 & \epsilon_c \sqrt{\gamma} \\ \epsilon_d \sqrt{\delta} & & \epsilon_c \sqrt{\gamma} & 1 \end{bmatrix}, \quad (14.97)$$

$$y = y(\epsilon; \alpha, \beta, \gamma, \delta) = \begin{bmatrix} 1 & \epsilon_a \sqrt{\alpha} & 0 & \epsilon_d \sqrt{\delta} \\ \epsilon_a \sqrt{\alpha} & 1 & \epsilon_b \sqrt{\beta} & 0 \\ 0 & \epsilon_b \sqrt{\beta} & 1 & \epsilon_c \sqrt{\gamma} \\ \epsilon_d \sqrt{\delta} & 0 & \epsilon_c \sqrt{\gamma} & 1 \end{bmatrix}. \quad (14.98)$$

On utilisera souvent la notation $\alpha' = 1 - \alpha, \dots, \delta' = 1 - \delta$. On peut remarquer que

$$1 - \alpha - \beta - \gamma - \delta + \alpha\gamma + \beta\delta = \alpha'\gamma' + \beta'\delta' - 1.$$

Avec ces notations on voit facilement avec les déterminants que $y \in P_1$ si et seulement si

$$\alpha + \beta \leq 1, \beta + \gamma \leq 1, \gamma + \delta \leq 1, \delta + \alpha \leq 1 \quad (14.99)$$

$$\alpha'\gamma' + \beta'\delta' - 1 \geq 2\epsilon_a \epsilon_b \epsilon_c \epsilon_d \sqrt{\alpha\beta\gamma\delta}. \quad (14.100)$$

Les inégalités de 14.99 s'obtiennent avec les déterminants d'ordre 3, celles de 14.100 avec le déterminant de y . On considèrera en particulier pour $0 < \alpha < 1$ l'élément suivant de P_1 :

$$y_\alpha = y(1, 1, 1, -1; \alpha, \alpha', \alpha, \alpha') = \begin{bmatrix} 1 & \sqrt{\alpha} & 0 & -\sqrt{\alpha'} \\ \sqrt{\alpha} & 1 & \sqrt{\alpha'} & 0 \\ 0 & \sqrt{\alpha'} & 1 & \sqrt{\alpha} \\ -\sqrt{\alpha'} & 0 & \sqrt{\alpha} & 1 \end{bmatrix}. \quad (14.101)$$

Le lecteur peut vérifier que pour y_α , les 5 inégalités 14.99 et 14.100 sont des égalités, et donc que y_α est de rang 2.

Fiedler a démontré que $x \in Q_1$ si et seulement si $0 \leq \alpha, \dots, \delta \leq 1$ et si

$$\alpha + \beta + \gamma + \delta \leq 2(1 + \sqrt{\alpha\beta\gamma\delta} + \sqrt{\alpha'\beta'\gamma'\delta'}).$$

Cette dernière condition pourrait être réécrite

$$\alpha\gamma + \beta\delta - \alpha'\gamma' - \beta'\delta' \leq 2\sqrt{\alpha\beta\gamma\delta} + 2\sqrt{\alpha'\beta'\gamma'\delta'}.$$

Il est intéressant de rechercher les génératrices extrémales des cônes \overline{P} et \overline{Q} , car cela permettrait sans doute de montrer que P et Q sont des cônes duaux, et donc de clarifier la question du domaine d'existence des transformées de Laplace des lois de Wishart liées au modèle graphique non triangulé du carré.

Lemme 1. Soit E un espace euclidien et soit $\overline{P}(E)$ le cône fermé de ses endomorphismes symétriques positifs. Soit $C \subset \overline{P}(E)$ un cône convexe fermé et $y \in C$ de rang 1. Alors y est sur une génératrice extrémale de C .

Démonstration du lemme. Soit y' et y'' dans C tels que $y = y' + y''$. Soit e une base de diagonalisation de y . Alors il existe $\lambda > 0$ tel que

$$\text{diag}(\lambda, 0, \dots, 0) = [y]_e^e = [y']_e^e + [y'']_e^e.$$

Puisque $[y']_e^e$ et $[y'']_e^e$ sont des matrices symétriques positives cela entraîne qu'elles sont proportionnelles à $[y]_e^e$ ce qui entraîne le résultat.

Théorème 2, conjecturé. Les génératrices extrémales de \overline{P} sont les matrices de rang 1 et les matrices proportionnelles aux éléments de P_1 de la forme $y(\epsilon; \alpha, \alpha', \alpha, \alpha')$ avec $0 < \alpha < 1$ et $\epsilon_a \epsilon_b \epsilon_c \epsilon_d = -1$.

Démonstration. Ce résultat est plausible. Voici plusieurs étapes de sa démonstration.

1) Montrons que y_α est sur une génératrice extrémale. Pour cela on introduit

$$A_\alpha = \begin{bmatrix} 1 & \sqrt{\alpha} \\ \sqrt{\alpha} & 1 \end{bmatrix}, \quad B_\alpha = \begin{bmatrix} 0 & -\sqrt{\alpha'} \\ \sqrt{\alpha'} & 0 \end{bmatrix}$$

ce qui fait que $y_\alpha = \begin{bmatrix} A_\alpha & B_\alpha \\ B_\alpha^T & A_\alpha \end{bmatrix}$. Une remarque essentielle est que $B_\alpha A_\alpha^{-1} B_\alpha = A_\alpha$ ce qui fait qu'en posant $U = \begin{bmatrix} I_2 & A_\alpha^{-1} B_\alpha \\ 0 & I_2 \end{bmatrix}$ on a $y_\alpha = U^T \begin{bmatrix} A_\alpha & 0 \\ 0 & 0 \end{bmatrix} U$. Attention, U n'est nullement orthogonale, et les valeurs propres de $\begin{bmatrix} A_\alpha & 0 \\ 0 & 0 \end{bmatrix}$, soit $(1 + \sqrt{\alpha}, 1 - \sqrt{\alpha}, 0, 0)$, ne sont nullement celles de y_α , soit $(2, 2, 0, 0)$.

Supposons maintenant que $y_\alpha = y + y'$ avec y et y' dans \overline{P} . Pour montrer que y_α est sur une génératrice extrémale, il nous faut montrer que y et y' sont proportionnels à y_α . Pour cela, on écrit y et y' par blocs :

$$y = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}, \quad y' = \begin{bmatrix} A' & B' \\ (B')^T & C' \end{bmatrix}$$

avec les notations

$$A = \begin{bmatrix} x_1 & a\sqrt{\alpha} \\ a\sqrt{\alpha} & x_2 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & -d\sqrt{\alpha'} \\ b\sqrt{\alpha'} & 0 \end{bmatrix}, \quad C = \begin{bmatrix} x_3 & c\sqrt{\alpha} \\ c\sqrt{\alpha} & x_4 \end{bmatrix},$$

et une notation analogue pour y' . L'astuce est alors de multiplier l'égalité $y_\alpha = y + y'$ de chaque coté par U^{-1} et $(U^{-1})^T$. On obtient ainsi

$$\begin{bmatrix} A_\alpha & 0 \\ 0 & 0 \end{bmatrix} = (U^{-1})^T \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} U^{-1} + (U^{-1})^T \begin{bmatrix} A' & B' \\ (B')^T & C' \end{bmatrix} U^{-1}.$$

On écrit alors

$$(U^{-1})^T \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} U^{-1} = \begin{bmatrix} A & B_1 \\ B_1^T & C_1 \end{bmatrix}, \quad (U^{-1})^T \begin{bmatrix} A' & B' \\ (B')^T & C' \end{bmatrix} U^{-1} = \begin{bmatrix} A' & B'_1 \\ (B'_1)^T & C'_1 \end{bmatrix}$$

avec $B_1 = B - AA_\alpha^{-1}B_\alpha$ et $C_1 = C - AA_\alpha^{-1}B_\alpha - (AA_\alpha^{-1}B_\alpha)^T + B_\alpha^T A_\alpha^{-1} AA_\alpha^{-1} B_\alpha$ avec des expressions analogues pour y' . Puisque les matrices $\begin{bmatrix} A & B_1 \\ B_1^T & C_1 \end{bmatrix}$ et

$\begin{bmatrix} A' & B'_1 \\ (B'_1)^T & C'_1 \end{bmatrix}$ sont positives et de somme $\begin{bmatrix} A_\alpha & 0 \\ 0 & 0 \end{bmatrix}$ cela entraîne que $C_1 = C'_1 = 0$, puis que $B_1 = B'_1 = 0$. En écrivant les équations correspondantes à ces 4 égalités matricielles, on obtient que $y = ay_\alpha$ et $y' = a'y_\alpha$ avec $0 \leq a = 1 - a' \leq 1$, ce qui montre l'extrémalité.

2) Montrons que $y(\epsilon; \alpha, \alpha', \alpha, \alpha')$ avec $0 < \alpha < 1$ et $\epsilon_a \epsilon_b \epsilon_c \epsilon_d = -1$ est sur une génératrice extrémale. On pourrait imiter la démonstration du 1), mais il est plus élégant d'observer que si $D = \text{diag}(\eta_1, \eta_2, \eta_3, \eta_4)$ avec $\eta_j = \pm 1$ alors l'application $y \mapsto DyD$ est un automorphisme du cône \overline{P} . Cela entraîne que DyD est sur une génératrice extrémale si y l'est. Comme $\epsilon_a \epsilon_b \epsilon_c \epsilon_d = -1$ par hypothèse et que $(\eta_1 \eta_2 \eta_3 \eta_4)^2 = 1$ il existe D tel que $y(\epsilon; \alpha, \alpha', \alpha, \alpha') = Dy_\alpha D$ et l'extrémalité de $y(\epsilon; \alpha, \alpha', \alpha, \alpha')$ est montrée.

3) Les matrices de rang 1 situées dans \overline{P} sont extrémales d'après le Lemme 1.

4) Examinons quelles matrices y de \overline{P} sont sommes de matrices de rang 1. Bien entendu, à cause du 1) et 2) on sait que tous les éléments de \overline{P} ne sont pas de ce type.

On sait que toute matrice symétrique positive est somme de matrices symétriques positives de rang 1. C'est vrai en particulier pour les matrices d'ordre 2. Il s'agit donc d'étudier quels y dans \overline{P} sont telles qu'il existe 4 matrices A, B, C, D dans \overline{P} telles que $y = A + B + C + D$ et telles que A, B, C, D soient de la forme

$$A = \begin{bmatrix} a_1 & a_{12} & 0 & 0 \\ a_{12} & a_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \dots, D = \begin{bmatrix} d_1 & 0 & 0 & d_{41} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ d_{41} & 0 & 0 & d_4 \end{bmatrix}.$$

Si (x_1, x_2, x_3, x_4) est la diagonale de y , alors y est somme de matrices de rang 1 dans le cas particulier où cette suite a un ou plusieurs zéros. À cause des symétries, il suffit de discuter les cas où les seuls zéros sont les suivants

1. Si tous les x_j sont nuls, alors $y = 0$ et c'est trivial.
2. Si $x_2 = x_3 = x_4 = 0$ c'est trivial avec $B = C = D = 0$ et $A = \text{diag}(x_1, 0, 0, 0)$.
3. Si $x_3 = x_4 = 0$, prendre $A = y$ et $B = C = D = O$.
4. Si $x_2 = x_4 = 0$ prendre $A = \text{diag}(x_1, 0, 0, 0)$ et $B = \text{diag}(0, 0, x_3, 0)$ et $C = D = 0$.
5. Si $x_4 = 0$ on prend $D = 0$ et l'existence de A, B, C découle d'un théorème du cours sur les graphes triangulés.

Donc sans perte de généralité on suppose maintenant $x_j > 0$. En considérant

$$\Delta = \text{diag}(x_1^{-1/2}, x_2^{-1/2}, x_3^{-1/2}, x_4^{-1/2})$$

la matrice $\Delta y \Delta$ est dans P_1 et il suffit de trouver A, B, C, D avec l'hypothèse supplémentaire que $y \in P_1$. On adopte la notation 14.97. Une discussion du même genre que la précédente sur les zéros de la suite (α, \dots, δ) permet de se ramener sans perte de généralité au cas difficile où ces 4 nombres sont > 0 . Les 4 matrices A, \dots, D doivent donc satisfaire aux équations

$$d_1 + a_1 = a_2 + b_2 = b_3 + c_3 = c_4 + d_4 = 1, \quad a_{12} = \epsilon_a \sqrt{\alpha}, \dots, d_{41} = \epsilon_d \sqrt{\delta}.$$

y est somme de matrices de rang 1 si on peut trouver des solutions à ce système linéaire telles que de plus les a_1, \dots, d_4 soient ≥ 0 et tels que $a_{12}^2 \leq a_1 a_2, \dots, d_{41}^2 \leq d_1 d_4$ lorsque les α, \dots, δ et les ϵ satisfont 14.99 et 14.100.

Lemme 3. Les éléments de P_1 tels que $\alpha > 0, \dots, \delta > 0$ et qui sont sommes de matrices de rang 1 sont ceux qui satisfont une des 4 inégalités

$$\delta \leq \left(1 - \frac{\alpha}{\beta'}\right) \left(1 - \frac{\gamma}{\beta'}\right) \quad (14.102)$$

et les 3 autres analogues obtenues par symétrie.

Démonstration. Supposons que $y = y(\epsilon; \alpha, \dots, \delta) \in P_1$ satisfasse 14.102. Je dis qu'alors on peut trouver $d \in]\delta, 1[$ tel que

$$y' = \begin{bmatrix} 1-d & \epsilon_a \sqrt{\alpha} & 0 & 0 \\ \epsilon_a \sqrt{\alpha} & 1 & \epsilon_b \sqrt{\beta} & 0 \\ 0 & \epsilon_b \sqrt{\beta} & 1 & \epsilon_c \sqrt{\gamma} \\ 0 & 0 & \epsilon_c \sqrt{\gamma} & 1 - \frac{\delta}{d} \end{bmatrix}$$

soit une matrice positive. Pour cela en multipliant y' par deux matrice diagonale convenable on se ramene à l'examen de

$$y(\epsilon; \frac{\alpha}{1-d}, \beta, \frac{\gamma}{1-\frac{\delta}{d}})$$

qui satisfait les conditions 14.99 et 14.100 si

Inversement, supposons que $y \in P_1$ soit somme de matrices de P_1 de rang 1. Donc $y = A + B + C + D$ avec les notations ci dessus. Comme on a supposé que α, \dots, δ sont tous non nuls, il existe donc un des A, B, C, D qui est non diagonal. Supposons que ce soit

5) Au vu de la discussion du 4) qui ramène au cas de la diagonale sans zéros, pour montrer le théorème, il suffirait de montrer que tout élément de P_1 est combinaison convexe des $y(\epsilon; \alpha, \alpha', \alpha, \alpha')$ et des huit matrices symétriques de la forme

$$\begin{bmatrix} 1 & \pm 1 & 0 & 0 \\ \pm 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & \pm 1 \\ 0 & 0 & \pm 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 & \pm 1 \\ 0 & 1 & \pm 1 & 0 \\ 0 & \pm 1 & 1 & 0 \\ \pm 1 & 0 & 0 & 1 \end{bmatrix}. \tag{14.103}$$

Autrement dit, il suffit de résoudre le problème auxiliaire de la recherche des points extrémaux du convexe compact P_1 (ou si on veut, des génératrices extrémales du cône convexe ferme des éléments de \overline{P} de diagonale constante). Le fait que y_α soit sur une génératrice extrémale de \overline{P} entraîne que c'est un point extrémal de P_1 . Les matrices de 14.103 ne sont évidemment pas sur des génératrices extrémales de \overline{P} puisque somme de deux matrices de rang 1. Par conséquent le fait qu'elles soient néanmoins des points extrémaux de P_1 demande démonstration. Contentons nous de montrer que

$$y(1, 1, 1, 1; 1, 0, 1, 0) = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

est un point extrémal de P_1 . Les 7 autres s'en déduisent par application d'un automorphisme de P_1 convenable. Supposons donc que $y(1, 1, 1, 1; 1, 0, 1, 0) =$

$y + y'$ avec y et y' positivement proportionnels à des éléments de P_1 . Ecrivons alors

$$y = \begin{bmatrix} x & a & 0 & d \\ a & x & b & 0 \\ 0 & b & x & c \\ d & 0 & c & x \end{bmatrix}, \quad y' = \begin{bmatrix} x' & a' & 0 & -d \\ a' & x' & b & 0 \\ 0 & -b & x' & c' \\ -d & 0 & c' & x' \end{bmatrix}$$

avec $x = 1 - x'$, $a = 1 - a'$ et $c = 1 - c'$. L'égalité $y(1, 1, 1, 1; 1, 0, 1, 0) = y + y'$ entraîne en particulier

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} x & a \\ a & x \end{bmatrix} + \begin{bmatrix} x' & a' \\ a' & x' \end{bmatrix} = \begin{bmatrix} x & c \\ c & x \end{bmatrix} + \begin{bmatrix} x' & c' \\ c' & x' \end{bmatrix}. \quad (14.104)$$

On a vu au lemme 1 que $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ est sur une génératrice extrémale de $\overline{P_2}$.

Comme les 4 autres matrices de 14.104 sont dans $\overline{P_2}$ elles lui sont proportionnelles et donc $a = c = x$. Sans perte de généralité à cause de la symétrie entre y et y' on peut supposer $a > 0$. Pour voir que $b = 0$ on écrit que le déterminant de

$$\begin{bmatrix} a & a & 0 \\ a & a & b \\ 0 & b & a \end{bmatrix}$$

doit être ≥ 0 . Ce déterminant est $-ab^2$ et donc $b = 0$. On procède de même pour montrer que $d = 0$.

15 Bibliographie.

- EDWARDS, D. (1995) *Introduction to Graphical Modelling*, Springer, New York.
- DAWID, A.P., LAURITZEN, S.L. (1993) Hyper Markov laws in the statistical analysis of decomposable graphical models, *Ann. Statist.* **21**, 1272-1317.
- DARROCH J.N., LAURITZEN, S.L., SPEED T.P. (1980) Markov fields and log-linear interaction models for contingency tables, *Ann. Statist.* **8**, 522-539.
- GOLUBIC, C. (1981) *Algorithmic graph theory and perfect graphs*, Academic Press, New York.
- GRONE, R., JOHNSON, C.R., SÁ, E.M, WOLKOWICZ, H. (1984) Positive Definite Completion of Partial Hermitian Matrices, *Linear Algebra and Its Applications* **58**, 109-124.
- LAURITZEN, S., SPEED, T.P., VIVAYAN, K. (1984) Decomposable graphs and hypergraphs, *J. Austral. Math. Soc. (Series A)* **36**, 12-29.
- LAURITZEN, S. (1996) *Graphical models*, Clarendon Press, Oxford.
- LEIMER, H.-G.. (1993) Optimal decomposition by clique separators *Discrete Math.* **113**, 99-123.
- PARTER, S. (1961) The use of linear graphs in Gauss elimination *SIAM Review* **3**, 119-130.

- ROSE, D.J. (1970) Triangulated graphs and the elimination process *J. Math. Anal. and Appl.* **32**, 597-609.
- ROSE, D.J., TARJAN, R.E., LUEKER G.S. (1976) Algorithmic aspects of vertex elimination on graphs, *SIAM J. Comput.* **5**, 266-283.
- SPEED, T.P., KIIVERI, H.T. (1986) Gaussian Markov distributions over finite graphs, *Annals of Statist.* **14**, 138-150.
- TARJAN, R.E. (1985) Decomposition by clique separators *Discrete Math.* **55**, 221-232.
- VOROBEV N.N. (1962) Markov measures and Markov extensions *Theor. Prob. and Applications.* ??, 420-429.