

Identifiabilité de modèles tensoriels couplés pour l'estimation de lois de probabilité discrète

Philippe FLORES, Konstantin USEVICH, David BRIE

Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France

prenom.nom@univ-lorraine.fr

Résumé – Dans cet article, nous proposons une approche pour fournir des bornes d'identifiabilité générique de modèles tensoriels couplés pour l'estimation de loi de probabilité discrète. Cette approche permet notamment d'étudier l'influence de différentes stratégies de couplage (total, partiel) sur l'identifiabilité du modèle couplés.

Abstract – In this article, we propose an approach to determine upper bounds for generic identifiability of coupled tensor model for estimating discrete probability mass functions. This allows us to study the influence of different coupling strategies on identifiability of the coupled model.

1 Introduction

L'estimation de lois de probabilité discrète à partir d'observations est un problème fondamental de l'apprentissage statistique utilisé dans de nombreuses applications telles que les systèmes de recommandation [1] ou l'analyse de données biologiques [2]. Ce problème se heurte à la malédiction de la dimension : le nombre d'inconnues augmente de manière exponentielle avec le nombre de variables M . Nous suivons l'approche de [1] qui est basé sur un modèle bayésien naïf (MBN) de la loi de probabilité. Il peut s'interpréter comme une décomposition CP (Canonique Polyadique) d'ordre M construite à partir des marginales 3D. Cette approche contourne la malédiction de la dimension car la complexité du MBN est linéaire par rapport à M et nécessite une quantité de données en $O(M^3)$ pour le construire. Pour réduire encore les coûts de calcul, nous proposons dans [2] d'étendre l'approche [1] aux cas de couplages partiels. Cet article a pour but d'étudier l'influence de stratégies de couplage sur l'identifiabilité du modèle CP.

L'organisation de cette communication est la suivante : la section 2 introduit le couplage partiel pour l'estimation de lois de probabilité et présente quelques stratégies de couplage sous la forme de 3-graphes [3]. La section 3 étudie l'identifiabilité générique de modèles couplés par analyse du rang de la matrice jacobienne associée. Enfin, la section 4 fournit des résultats d'identifiabilité pour les stratégies de couplage considérées.

2 Modèle tensoriel de PMF

2.1 Malédiction de la dimension et MBN

Soit $\mathbf{x} = (X^{(1)}, \dots, X^{(M)})$ un vecteur de M variables aléatoires discrètes prenant des valeurs de 1 à I . Notre but est d'estimer la loi de probabilité (PMF) de \mathbf{x} , un tenseur d'ordre M noté \mathcal{H} , à partir de la matrice d'observation \mathbf{X} contenant N réalisations \mathbf{x}_n . L'approche naïve pour estimer \mathcal{H} est de compter le nombre d'occurrences de chaque M -uplets (i_1, \dots, i_M) :

$$\tilde{\mathcal{H}}_{i_1 \dots i_M} = \frac{1}{N} \text{Card} \left\{ n \in \llbracket 1, N \rrbracket \mid \mathbf{x}_n = (i_1, \dots, i_M) \right\} \quad (1)$$

Cette approche requiert un nombre d'échantillons qui croît exponentiellement avec M : c'est la malédiction de la dimension. Pour pallier ce problème, [1] propose de modéliser \mathcal{H} comme un modèle bayésien naïf (MBN) dont la complexité est linéaire $(MI + 1)R \ll I^M$. Le MBN introduit une variable latente L telle que \mathbf{x} est conditionnellement indépendant par rapport à L :

$$\mathcal{H}_{i_1 \dots i_M} = \sum_{r=1}^R \Pr(L = r) \prod_{m=1}^M \Pr(X^{(m)} = i_m \mid L = r), \quad (2)$$

et est interprété comme une décomposition CP [4] de rang R :

$$\mathcal{H} = \llbracket \boldsymbol{\lambda}; \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(M)} \rrbracket, \quad (3)$$

voir [4] pour plus de détails. \mathcal{H} étant un tenseur de probabilité, les matrices facteurs $\mathbf{A}^{(m)} \in \mathbb{R}^{I \times R}$ et $\boldsymbol{\lambda} \in \mathbb{R}^R$ doivent satisfaire les contraintes de *simplex* : la non-négativité $\boldsymbol{\lambda} \geq 0$, $\mathbf{A}^{(m)} \geq 0$, et la contrainte *sum-to-one* : $\mathbb{1}^\top \boldsymbol{\lambda} = 1$, $\mathbb{1}^\top \mathbf{A}^{(m)} = \mathbb{1}^\top$.

2.2 Couplage tensoriel de marginales 3D

Comme l'estimation de $\tilde{\mathcal{H}}$ nécessite une quantité de données trop importante, il est impossible d'estimer la CPD de \mathcal{H} à partir de $\tilde{\mathcal{H}}$. L'idée proposée dans [1] est de coupler les marginales 3D puisque celles-ci sont estimables pour des valeurs de N relativement faibles et permettent d'utiliser les propriétés d'unicité des tenseurs. Pour estimer le MBN, notre approche est de résoudre le problème d'optimisation :

$$\widehat{\boldsymbol{\lambda}}, \widehat{\mathbf{A}}^{(1)}, \dots, \widehat{\mathbf{A}}^{(M)} \quad (4)$$

$$= \underset{\boldsymbol{\lambda}, \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(M)}}{\operatorname{argmin}} \sum_{\{j,k,\ell\} \in \mathcal{T}} \left\| \tilde{\mathcal{H}}^{(jk\ell)} - \llbracket \boldsymbol{\lambda}; \mathbf{A}^{(j)}, \mathbf{A}^{(k)}, \mathbf{A}^{(\ell)} \rrbracket \right\|_F^2$$

$$\text{s.c. } \boldsymbol{\lambda} \geq 0, \mathbf{A}^{(m)} \geq 0, \mathbb{1}^\top \boldsymbol{\lambda} = 1, \mathbb{1}^\top \mathbf{A}^{(m)} = \mathbb{1}^\top$$

appelé factorisation tensorielle couplée où \mathcal{T} est un ensemble de triplets de variables. Ce problème peut être résolu grâce à une procédure d'optimisation alternée couplée [1] utilisant par exemple la méthode ADMM (*Alternating Direction Method of Multipliers*) ou en adaptant une approche de type *Expectation Maximization* [5].

2.3 Stratégies de couplage

Dans [1], \mathcal{T} est l'ensemble de tous les triplets possibles. La nouveauté de notre approche est de considérer un sous-ensemble de triplets afin de réduire les temps de calcul sans perdre significativement en performance d'estimation [2].

Dans ce cas, un couplage partiel \mathcal{T} est interprété comme un 3-graphe où M nœuds sont liés par $T = \text{Card}(\mathcal{T})$ hyperarêtes, chacune contenant 3 nœuds. On définit la matrice d'incidence $\mathbf{V} \in \mathbb{R}^{T \times M}$ et la séquence de degrés $\mathbf{d} \in \mathbb{R}^M$ de \mathcal{T} :

$$v_{tm} = \mathcal{I}_{\tau_t}(m), \quad d_m = \text{Card} \{ \tau \in \mathcal{T} | m \in \tau \}$$

où \mathcal{I}_{τ} est la fonction indicatrice de τ . Par exemple pour $M = 4$ et $\mathcal{T} = \{ \{1, 3, 4\}, \{2, 3, 4\} \}$:

$$\mathbf{V} = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{d} = [1 \quad 1 \quad 2 \quad 2]. \quad (5)$$

Pour garantir la résolution de (4), le 3-graphe lié à \mathcal{T} doit être connexe afin que toutes les variables soient liées entre elles.

Couplage Total : Ce couplage, présenté dans [1], contient tous les triplets donc $\mathcal{T} = \{ \{j, k, \ell\} \subset \llbracket 1, M \rrbracket | j < k < \ell \}$ et a une séquence de degré constante égale à :

$$\mathbf{d} = \left[\binom{M-1}{2} \quad \dots \quad \binom{M-1}{2} \right]. \quad (6)$$

Pour ce type de couplage, $T = \binom{M}{3}$ augmente de manière cubique avec M ce qui devient ingérable en pratique. La malédiction de la dimension a été déplacé à T et le couplage partiel permet de diminuer la complexité de notre problème.

Couplage Minimal : Une stratégie de couplage minimal possible est de considérer

$$\mathcal{T} = \{ \{1, 2, 3\}, \{3, 4, 5\}, \{5, 6, 7\}, \dots, \{M-2, M-1, M\} \}$$

où deux triplets consécutifs sont reliés par une seule variable. Cette stratégie contient $T = \lfloor M/2 \rfloor$ triplets avec :

$$\mathbf{d} = [1 \quad 1 \quad 2 \quad 1 \quad 2 \quad \dots \quad 2 \quad 1 \quad 2 \quad 1 \quad 1]. \quad (7)$$

Couplage Aléatoire : Pour obtenir une stratégie de couplage aléatoire, $T \in \left[\lfloor \frac{M}{2} \rfloor, \binom{M}{3} \right]$ triplets sont tirés aléatoirement tout en garantissant que \mathcal{T} est connexe. La séquence de degré obtenue est alors non nécessairement constante.

Couplage Équilibré : À T fixé, il est possible de considérer \mathcal{T} avec une séquence de degré \mathbf{d} constante à 1 près. Si on définit δ cette constante et ϵ tel que $3T = \delta M + \epsilon$, on a

$$\mathbf{d} = [\delta \quad \dots \quad \delta \quad \overbrace{\delta+1 \quad \dots \quad \delta+1}^{\epsilon}]. \quad (8)$$

Pour obtenir un couplage équilibré, il est possible d'appliquer des méthodes de génération d'hypergraphes uniformes [3].

Dans la suite, nous étudions l'identifiabilité du MBN (3) en fonction de stratégies de couplage et du nombre de triplets T .

3 Identifiabilité du modèle CP couplé

3.1 Position du problème

Étant donné un ensemble de tenseurs d'ordre 3 et une stratégie de couplage $\{ \mathcal{H}^{(jkl)}, \{j, k, \ell\} \in \mathcal{T} \}$, on dira qu'une décomposition couplée est identifiable si la solution du problème (3) est unique aux indéterminations d'échelle et d'ordre près. Notons que les contraintes simplex permettent de lever l'indétermination d'échelle. On dira que le modèle couplé est identifiable au sens générique si la décomposition est unique pour des matrices facteurs tirées aléatoirement selon une loi de probabilité continue. Nous proposons d'étudier l'influence du choix de la stratégie de couplage sur l'identifiabilité du modèle couplé. À notre connaissance, ce problème n'a pas été traité. Citons cependant [1] qui fournit des conditions suffisantes d'identifiabilité générique. Une analyse des démonstrations montre qu'elle se fondent sur un nombre limité de triplets. Ainsi, dans [1, **Théorème 1**] : si $M \leq I$, alors \mathcal{H} est identifiable si $R \leq I(M-2)$. Ce théorème est prouvé en groupant les $M-2$ dernières variables afin d'analyser l'identifiabilité du modèle $\mathcal{H} = \llbracket \lambda; \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{C} \rrbracket$ où $\mathbf{C} = [\mathbf{A}^{(3)\top} \quad \dots \quad \mathbf{A}^{(M)\top}]^\top$. Cela revient donc à considérer $\mathcal{T} = \{ \{1, 2, \ell\} | \ell \geq 3 \}$ et donc à n'utiliser que $M-2$ triplets sur les $\binom{M}{3}$ disponibles. D'autres partitions conduisent à des résultats différents qui permettent notamment de prendre en compte le cas $M > I$. Notons que les conditions de [1] relâchent les contraintes de simplex et de positivité qui ne sont pas prises en compte. L'étude du rang de la matrice jacobienne permet d'établir des conditions d'identifiabilité générique. Elles relâchent les contraintes de positivité mais prennent en compte les contraintes de simplex ainsi que la stratégie de couplage.

3.2 Paramétrisation et matrice jacobienne

Dans sa forme initiale, la contrainte de simplex définit une paramétrisation implicite. Pour se ramener à une paramétrisation explicite et prendre en compte ces contraintes, on considère la paramétrisation \mathcal{P} et sa matrice jacobienne $\mathcal{J}_{\mathcal{P}}$:

$$\mathcal{P}(\mathbf{a}) = \begin{bmatrix} a_1 \\ \vdots \\ a_{I-1} \\ 1 - \sum_{i=1}^{I-1} a_i \end{bmatrix}, \quad \mathcal{J}_{\mathcal{P}} = \begin{bmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ -1 & \dots & & -1 \end{bmatrix}.$$

Cette paramétrisation d'un facteur supprime un degré de liberté car le rang de $\mathcal{J}_{\mathcal{P}}$ est $I-1$. En définissant les facteurs tronqués $\underline{\mathbf{a}}_r^{(m)} = (\mathbf{a}_{i,r}^{(m)})_{i=1, r=1}^{I-1, R}$, l'estimation des paramètres

$$\boldsymbol{\theta} = \text{vec}(\lambda_1, \underline{\mathbf{a}}_1^{(1)}, \dots, \underline{\mathbf{a}}_1^{(M)}, \dots, \lambda_R, \underline{\mathbf{a}}_R^{(1)}, \dots, \underline{\mathbf{a}}_R^{(M)})$$

est définie par la fonction de paramétrisation μ :

$$\mu : \begin{cases} \mathbb{R}^{(M(I-1)+1)R} & \longrightarrow & \mathbb{R}^{TI^3} \\ \boldsymbol{\theta} & \longmapsto & \text{vec}(\mathcal{H}^{(\tau_1)}, \dots, \mathcal{H}^{(\tau_T)}) \end{cases}, \quad (10)$$

$$\mathcal{J}_\mu(\theta_r) = \underbrace{\begin{bmatrix} \mathbf{a}_r^{(4)} \otimes \mathbf{a}_r^{(3)} \otimes \mathbf{a}_r^{(1)} \\ \mathbf{a}_r^{(4)} \otimes \mathbf{a}_r^{(3)} \otimes \mathbf{a}_r^{(2)} \end{bmatrix}}_{\text{Poids } \lambda_r} \underbrace{\begin{bmatrix} \lambda_r(\mathbf{a}_r^{(4)} \otimes \mathbf{a}_r^{(3)} \otimes \mathcal{J}_P) & \mathbf{0}_{I^3 \times I} \\ \mathbf{0}_{I^3 \times I} & \lambda_r(\mathbf{a}_r^{(4)} \otimes \mathbf{a}_r^{(3)} \otimes \mathcal{J}_P) \end{bmatrix}}_{\text{Facteur } \mathbf{a}_r^{(1)}} \underbrace{\begin{bmatrix} \mathbf{0}_{I^3 \times I} & \lambda_r(\mathbf{a}_r^{(4)} \otimes \mathcal{J}_P \otimes \mathbf{a}_r^{(1)}) \\ \lambda_r(\mathbf{a}_r^{(4)} \otimes \mathbf{a}_r^{(3)} \otimes \mathcal{J}_P) & \lambda_r(\mathbf{a}_r^{(4)} \otimes \mathcal{J}_P \otimes \mathbf{a}_r^{(2)}) \end{bmatrix}}_{\text{Facteur } \mathbf{a}_r^{(2)}} \underbrace{\begin{bmatrix} \lambda_r(\mathbf{a}_r^{(4)} \otimes \mathcal{J}_P \otimes \mathbf{a}_r^{(1)}) & \lambda_r(\mathcal{J}_P \otimes \mathbf{a}_r^{(3)} \otimes \mathbf{a}_r^{(1)}) \\ \lambda_r(\mathbf{a}_r^{(4)} \otimes \mathcal{J}_P \otimes \mathbf{a}_r^{(2)}) & \lambda_r(\mathcal{J}_P \otimes \mathbf{a}_r^{(3)} \otimes \mathbf{a}_r^{(2)}) \end{bmatrix}}_{\text{Facteur } \mathbf{a}_r^{(3)}} \underbrace{\begin{bmatrix} \lambda_r(\mathcal{J}_P \otimes \mathbf{a}_r^{(3)} \otimes \mathbf{a}_r^{(1)}) \\ \lambda_r(\mathcal{J}_P \otimes \mathbf{a}_r^{(3)} \otimes \mathbf{a}_r^{(2)}) \end{bmatrix}}_{\text{Facteur } \mathbf{a}_r^{(4)}} \quad (9)$$

où la marginale 3D $\mathcal{H}^{(\tau)}$ est paramétrée par :

$$\mathcal{H}^{(\tau)} = \left[\boldsymbol{\lambda}; \mathcal{P}(\mathbf{A}^{(j_\tau)}), \mathcal{P}(\mathbf{A}^{(k_\tau)}), \mathcal{P}(\mathbf{A}^{(\ell_\tau)}) \right]. \quad (11)$$

Pour simplifier l'analyse, la contrainte de sum-to-one sur $\boldsymbol{\lambda}$ est relâcher car elle n'affecte pas l'identifiabilité du modèle. Selon [6], la matrice jacobienne de la paramétrisation est de rang plein pour un paramètre générique si et seulement si le modèle est recouvrable, *i.e.* il existe un nombre fini de solutions. De plus, garantir la recouvrabilité est une notion beaucoup plus forte qu'une condition nécessaire d'identifiabilité. La notion de rang plein de matrice jacobienne est fréquemment utilisée pour étudier l'identifiabilité des décompositions tensorielles. Dans certains scénarios, cette condition est équivalente à l'inversibilité de la matrice d'information de Fisher [7] ce qui fournit des garanties sur l'identifiabilité locale du modèle [8]. On donne en section 3.3 la structure de la matrice jacobienne associée à la paramétrisation (10).

3.3 Structure de la matrice jacobienne

La matrice jacobienne de μ noté $\mathcal{J}_\mu(\boldsymbol{\theta})$ est de taille $TI^3 \times (M(I-1)+1)R$. Comme le MBN est additif par rapport aux termes de rang 1, $\mathcal{J}_\mu(\boldsymbol{\theta})$ est une matrice à R blocs colonnes :

$$\mathcal{J}_\mu(\boldsymbol{\theta}) = \begin{bmatrix} \mathcal{J}_\mu(\boldsymbol{\theta}_1) & \cdots & \mathcal{J}_\mu(\boldsymbol{\theta}_R) \end{bmatrix} \quad (12)$$

Ainsi, chaque bloc $\mathcal{J}_\mu(\boldsymbol{\theta}_r)$ représente la matrice jacobienne de la paramétrisation du terme de rang 1 $\boldsymbol{\theta}_r$. De plus, $\mathcal{J}_\mu(\boldsymbol{\theta})$ est décomposable en T blocs lignes :

$$\mathcal{J}_\mu(\boldsymbol{\theta}) = \begin{bmatrix} \mathcal{J}_\mu^{(\tau_1)}(\boldsymbol{\theta}) \\ \vdots \\ \mathcal{J}_\mu^{(\tau_T)}(\boldsymbol{\theta}) \end{bmatrix} \quad (13)$$

Pour chaque triplet $\tau \in \mathcal{T}$, $\mathcal{J}_\mu^{(\tau)}(\boldsymbol{\theta})$ représente la matrice jacobienne de la paramétrisation de $\mathcal{H}^{(j_\tau k_\tau \ell_\tau)}$ (11).

En exemple, si on considère le cas présenté en équation (5), la matrice jacobienne du r -ième terme de rang 1 $\boldsymbol{\theta}_r$ est présenté en équation (9). À noter que la structure de $\mathcal{J}_\mu(\boldsymbol{\theta})$ est étroitement liée à la matrice d'incidence \mathbf{V} . En effet, si la première colonne de chaque $\mathcal{J}_\mu(\boldsymbol{\theta}_r)$ est allouée à λ_r , les autres blocs de cette matrice reproduisent la structure de la matrice d'incidence \mathbf{V} .

4 Bornes d'identifiabilité

L'ensemble des rangs pour lesquels la décomposition (3) est identifiable est obtenu grâce à la procédure itérative de l'algorithme 1, inspirée de [9] pour la détermination du rang générique d'un tenseur. Comme $\mathcal{J}_\mu(\boldsymbol{\theta})$ a une structure à R blocs

colonnes (12), le principe de l'algorithme 1 est de générer des paramètres $\boldsymbol{\theta}_r$ tant que le rang de $\mathcal{J}_\mu(\boldsymbol{\theta})$ augmente de $M(I-1)+1$. Cet algorithme fournit R_{\max} tel que le modèle est recouvrable si $R \leq R_{\max}$ et R_{\max} est une borne d'identifiabilité car le modèle est nécessairement non identifiable si $R > R_{\max}$.

Algorithme 1: Borne d'identifiabilité du modèle (11)

Entrée: M, I, \mathcal{T}

Initialisation : $R = 1, D = 0,$

$\boldsymbol{\theta} = \text{vec}(\boldsymbol{\lambda}_1, \mathbf{a}_1^{(1)}, \dots, \mathbf{a}_1^{(M)})$ aléatoire

Tant que $\text{rang}(\mathcal{J}_\mu(\boldsymbol{\theta})) - D = M(I-1) + 1$

$D = \text{rang}(\mathcal{J}_\mu(\boldsymbol{\theta}))$

$\boldsymbol{\theta}_{R+1} = \text{vec}(\boldsymbol{\lambda}_{R+1}, \mathbf{a}_{R+1}^{(1)}, \dots, \mathbf{a}_{R+1}^{(M)})$ aléatoire

$\boldsymbol{\theta} = \text{vec}(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_R, \boldsymbol{\theta}_{R+1}), R = R + 1$

fin

Résultat: $R - 1$

4.1 Cas du couplage total

Conjecture 1 Soit \mathcal{T} la stratégie de couplage total. La borne d'identifiabilité générique du modèle (11) est

$$R_{\max} = \left\lfloor \frac{1 + M(I-1) + \binom{M}{2}(I-1)^2 + \binom{M}{3}I^3}{M(I-1) + 1} \right\rfloor. \quad (14)$$

Cette conjecture est obtenue en comparant le nombre d'observations différentes et le nombre de paramètres $(M(I-1)+1)R$. Puisque la méthode se base sur du couplage de marginales, une partie des TI^3 observations sont redondantes. Si toutes les marginales 3D de \mathcal{H} sont observées, le degré de liberté du modèle (11) est la somme de : 1 pour la marginale de dimension 0; $(I-1)$ pour chacune des M marginales 1D; $(I-1)^2$ pour les $\binom{M}{2}$ marginales 2D; et de même $(I-1)^3$ pour chaque marginales 3D. Pour $M \in \{4, 6, 8\}$ et $I \in \llbracket 3, 15 \rrbracket$, l'algorithme 1 donne des bornes en adéquation avec (14). La figure 1 montre que ces résultats sont supérieurs à ceux de [1], même si [1] présente des conditions suffisantes d'identifiabilité alors que la notre est nécessaire.

4.2 Couplages partiels aléatoires

Pour le couplage partiel, le degré de liberté du modèle dépend du nombre de triplets T et de P le nombre de paires différentes de variables liées entre elles par un triplet.

Proposition 1 Pour \mathcal{T} quelconque, la borne d'identifiabilité générique du modèle (11) est

$$R_{\max} \leq \left\lfloor \frac{1 + M(I-1) + P(I-1)^2 + T(I-1)^3}{M(I-1) + 1} \right\rfloor. \quad (15)$$

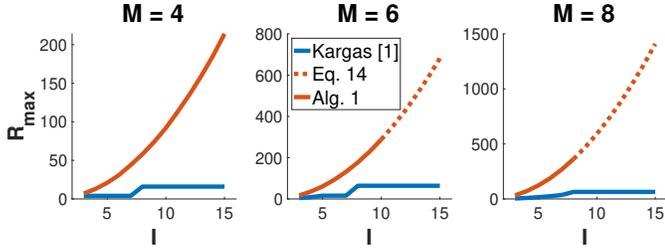


FIGURE 1 – Borne d’identifiabilité en fonction de M et I comparés avec [1]. La conjecture 1 est représentée pour les points non calculés par l’algorithme 1 (temps de calcul trop élevés).

Afin de vérifier cette condition, 1000 stratégies ont été générées pour $M = 8$, $I = 4$ et $T \in [4, 16]$. La figure 2 montre la répartition de la borne d’identifiabilité en fonction de T . À T fixé, plusieurs bornes existent car P n’est pas constant (15). Deuxièmement, une partie non négligeable des stratégies donne une borne plus faible. Le cas $R = 16$ est étudié dans la section suivante.

4.3 Exemple de couplage avec perte de rang

Dans la section précédente, on observe sur la figure 2 des stratégies aléatoires où la borne (15) n’est pas atteinte. Cette situation intervient pour des stratégies de couplage particulières. Par exemple, le cas où la borne est atteinte pour $R = 16$ est obtenu si on considère \mathcal{T} une stratégie de couplage telle que $d_1 = 1$ et $d_2 > 1$, c’est-à-dire si une variable est présente uniquement dans un triplet. Pour ces stratégies de couplage, la matrice $\mathcal{J}_\mu(\theta_r)$ est de la forme

$$\mathcal{J}_\mu(\theta_r) = \begin{bmatrix} \mathbf{B}_r & | & \cdots \\ \mathbf{0} & | & \mathbf{C}_r \end{bmatrix}$$

où $\mathbf{B}_r \in \mathbb{R}^{I^3 \times I}$ représente la matrice jacobienne liée à $\mathbf{a}_r^{(1)}$ et $\mathbf{C}_r \in \mathbb{R}^{(T-1)I^3 \times (M-1)I+1}$ celle liée aux autres facteurs et à λ_r . Comme $\mathcal{J}_\mu(\theta_r)$ a une structure triangulaire, $\mathcal{J}_\mu(\theta_r)$ est de rang colonnes plein si la matrice $\mathbf{B} = [\mathbf{B}_1 \cdots \mathbf{B}_R]$ est de rang colonnes plein. Avec la contrainte *sum-to-one*, \mathbf{B} contient au plus $I^3 - I^2$ lignes indépendantes donc on obtient la condition $R(I - 1) \leq I^3 - I^2$ dans le cas générique.

4.4 Couplages partiels équilibrés

Afin d’éviter les cas engendrant des pertes de rang, il est possible de choisir \mathcal{T} telle que \mathbf{d} soit constante à 1 près : ces stratégies sont dites équilibrées. Pour ces stratégies, on observe que la condition d’identifiabilité est bien celle de l’équation (15). La figure 2 présente la borne d’identifiabilité obtenue par l’algorithme 1. Les stratégies de couplage équilibrées n’engendrent pas de pertes d’identifiabilité et permettent donc se prémunir de situations telles que celle présentée en section 4.3.

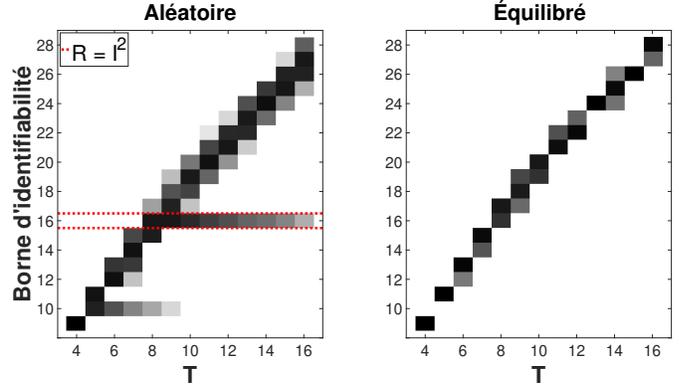


FIGURE 2 – Répartition de la borne d’identifiabilité en fonction du nombre de triplets T pour des stratégies aléatoires et équilibrées avec $M = 8$ et $I = 4$.

5 Conclusion

Nous proposons une approche pour l’étude de l’identifiabilité générique de modèles tensoriels couplés pour l’estimation de loi de probabilité. Des bornes d’identifiabilité générique ont été obtenues à l’aide d’un algorithme étudiant le rang de la matrice jacobienne associée à la paramétrisation de notre modèle. L’influence de différentes stratégies de couplage (total, partiel aléatoire, équilibré) sur l’identifiabilité a été étudiée. Le couplage total fournit la borne la plus grande alors que le couplage aléatoire peut engendrer des cas de pertes de rang. Celles-ci sont expliquées par la structure de la matrice jacobienne mais le couplage équilibré n’admet pas de pertes de rang.

Références

- [1] N. Kargas et al., “Tensors, learning, and “Kolmogorov extension” for finite-alphabet random vectors”, *TSP*, 2018.
- [2] P. Flores et al., “Coupled tensor factorization for flow cytometry data analysis,” *IEEE MLSP*, 2022.
- [3] A. Frosini et al., “On the degree sequences of uniform hypergraphs,” 17th IAPR International Conference, 2013.
- [4] T. Kolda and B. Bader, “Tensor decompositions and applications”, *SIAM Review*, 2009.
- [5] J. K. Chege et al., “Efficient probability mass function estimation from partially observed data,” *Asilomar*, 2022.
- [6] P. Breiding et al., “Algebraic compressed sensing”, *Applied and Computational Harmonic Analysis*, 2023.
- [7] C. Prévost et al., “Constrained Cramér–Rao bounds for reconstruction problems formulated as coupled canonical polyadic decompositions”, *Signal Processing*, 2022.
- [8] C. Jauffret, “Observability and Fisher information matrix in nonlinear regression”, *IEEE TAES*, 2007.
- [9] P. Comon et al., “Generic and typical ranks of multi-way arrays”, *Linear Algebra and its Applications*, 2009.