

Estimation de modèles markoviens discrets dans un cadre industriel fiabiliste à données manquantes

Alberto Pasanisi, Shuai Fu, Nicolas Bousquet

► **To cite this version:**

Alberto Pasanisi, Shuai Fu, Nicolas Bousquet. Estimation de modèles markoviens discrets dans un cadre industriel fiabiliste à données manquantes. 42èmes Journées de Statistique, 2010, Marseille, France, France. 2010. <inria-00494688>

HAL Id: inria-00494688

<https://hal.inria.fr/inria-00494688>

Submitted on 24 Jun 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ESTIMATION DE MODÈLES MARKOVIENS DISCRETS DANS UN CADRE INDUSTRIEL FIABILISTE À DONNÉES MANQUANTES

Alberto Pasanisi¹, Shuai Fu^{1,2} & Nicolas Bousquet¹

¹*EDF R&D, Dépt. Management des Risques Industriels, 6 quai Watier, 78401 Chatou*

²*Département de Mathématiques, Université Paris-Sud 11, 91405 Orsay*

Résumé. Les modèles markoviens sont particulièrement utiles pour décrire des systèmes qui, au long de leur vie, passent à travers différents états. Les paramètres de ces modèles sont les probabilités de transition entre les états. Normalement, les données disponibles pour l'inférence statistique sont des séquences temporelles d'états pour un nombre donné d'individus. Quand les séquences sont incomplètes, l'estimation de la matrice de transition n'est pas triviale et demande l'utilisation de techniques plus avancées. Dans cette communication, nous nous focalisons sur l'estimation bayésienne des probabilités de transition. Premièrement, nous présentons différentes méthodes MCMC, en fonction de la structure des données manquantes. Ensuite, nous proposons une manière d'accélérer les calculs MCMC en tenant compte de la dépendance entre les lignes de la matrice de transition. Finalement, nous montrons les résultats d'essais simulés menés sur des matrices typiquement rencontrées dans les problèmes de fiabilité industrielle.

Summary. Markov models are very useful to describe systems that pass, during their life, through different states. The parameters of such a model are transition probabilities between states. Normally, data for statistical inference consist in time sequences of the states for a given number of individuals. When time sequences are incomplete the problem is not trivial and requires more advanced statistical techniques. In this communication we will focus on Bayesian estimation of transition probabilities. First, depending on missed data structure, we present different MCMC calculation methods. Then, we show how to accelerate MCMC calculations, by taking into account the dependence between the lines of the transition matrix. Finally, we show some results condensing trials led on matrices typically encountered in industrial reliability issues.

Mots-clés. Matrice de transition markovienne; données manquantes; algorithme de Gibbs; algorithme de Metropolis-Hastings; copule gaussienne.

1 Cadre de travail

La fiabilité de nombreux systèmes et composants industriels (résumés par la notation Σ) peut être suivie au moyen de mesures \mathbf{z} d'avancement de la défaillance (par exemple des

classes progressives de tailles de fissures), qui peuvent correspondre à un nombre fini r de situations nommées *états*. En supposant que pour une échelle de temps de fonctionnement réaliste, le comportement de Σ (incluant possiblement des réparations) n'est dépendant que de son état au pas de temps précédent, et qu'il suit donc un processus markovien stationnaire du premier ordre, l'objectif du fiabiliste est d'estimer la matrice de transition :

$$\boldsymbol{\psi} = \begin{pmatrix} \boldsymbol{\psi}_1 \\ \boldsymbol{\psi}_2 \\ \vdots \\ \boldsymbol{\psi}_r \end{pmatrix} = \begin{pmatrix} \psi_{1,1} & \dots & \psi_{1,r} \\ \psi_{2,1} & \dots & \psi_{2,r} \\ \vdots & \dots & \vdots \\ \psi_{r,1} & \dots & \psi_{r,r} \end{pmatrix}$$

où $\psi_{i,j}$ correspond à la probabilité pour Σ de passer d'un état S_i à un autre S_j . Ce faisant, le fiabiliste évite de faire des hypothèses paramétriques sur les modèles de dégradation (et éventuellement de réparation) sous-jacents.

Le choix d'un cadre d'estimation bayésien (Lee et al. 1968), où la loi *a priori* de $\boldsymbol{\psi}$ est définie comme le produit de r lois (indépendantes) de Dirichlet :

$$\boldsymbol{\psi}_i \sim \mathcal{D}_{ir}(\gamma_{i1}, \dots, \gamma_{ir}), \quad (1)$$

permet au statisticien de s'affranchir du problème d'estimation de $\boldsymbol{\psi}$ sous les contraintes $0 \leq \psi_{i,j} \leq 1$ et $\sum_{j=1}^r \psi_{i,j} = 1$, et facilite la prise de décision prédictive vis-à-vis d'un cadre fréquentiste classique (cf. Jones 2005), en tenant éventuellement compte de l'information supplémentaire apportée par des experts industriels. Typiquement, on souhaite connaître la distribution *a posteriori* de la probabilité inconditionnelle d'être dans les états 1 à r au temps t :

$$\mathbf{p}(t) = (p_1(t), \dots, p_r(t)) = \mathbf{p}(0) \cdot \boldsymbol{\psi}^t.$$

Le cadre idéal de travail est le suivant : étant donnés m matériels identiques Σ vérifiés à des instants $1, \dots, T$, la matrice complète d'observations iid. $\mathbf{z} = (z_{k,t})_{k=1, \dots, m, t=1, \dots, T}$ permet de définir les nombres $\omega_{i,j} = \sum_{t=1}^T \sum_{k=1}^m \mathbf{1}(z_{k,t-1} = S_i, z_{k,t} = S_j)$ de transitions en un pas de temps d'un état S_i à un autre S_j . Le produit des vecteurs $\boldsymbol{\omega}_i = (\omega_{i,j})_{j=1, \dots, r}$ suit alors la loi produit de multinomiales $\otimes_{i=1}^r \mathcal{M}(\sum_{j=1}^r \omega_{i,j}; \psi_{i,1}, \dots, \psi_{i,r})$. Dans ce cas, la loi *a priori* $\pi(\boldsymbol{\psi}_i)$ est conjuguée, donc la loi *a posteriori* $\pi(\boldsymbol{\psi}|\mathbf{z})$ est un produit explicite de lois de Dirichlet.

2 Problème à données manquantes

Cette inférence directe se révèle impossible dans les cas, fréquents dans l'industrie (Pasani et Parent 2004), où les retours d'expérience contiennent des données manquantes. Celles-ci peuvent être de deux types :

1. certains $z_{j,t}$ ne sont pas observés, par exemple parce que la visite des installations à un instant t ne permet de vérifier qu'un petit nombre d'exemplaires de Σ sur les m en fonctionnement ;
2. les données \mathbf{z} sont résumées sous la forme de données *agrégées* $\mathbf{n} = (n_i(t))_{i=1,\dots,r,t=1,\dots,T}$, où :

$$n_i(t) = \sum_{k=1}^m \mathbb{1}_{(t=t_k, z(k,t_k)=S_i)},$$

correspondant aux nombres de matériels Σ dans un état i à un instant donné ; par exemple parce que la conservation des trajectoires individuelles des matériels n'a pas été jugée particulièrement intéressante au moment des visites des installations.

Le cas des séquences partiellement observées peut être classiquement réglé par l'utilisation d'un algorithme de Gibbs couplé à une méthode d'*augmentation de données* (Robert et Casella 2004) : on considère que les données manquantes jouent le rôle de paramètres additionnels que l'on peut simuler à chaque étape de l'algorithme (cf. l'estimation du modèle d'Arnason-Schwarz, Marin et Robert 2007). Une fois les séquences reconstituées, on retombe dans le cadre d'inférence idéal (*l'a priori* (1) est conjugué) et la simulation *a posteriori* conditionnelle est immédiate.

En présence de données agrégées, l'algorithme de Metropolis-Hastings (MH) est une façon standard de réaliser l'estimation bayésienne de $\boldsymbol{\psi}$. Dans ce cas, la vraisemblance s'écrit comme un produit de T multinomiales de paramètres $\mathbf{p}(\mathbf{t})$ et $\sum_{i=1}^r n_i(t)$ et *l'a priori* (1) n'est plus conjugué. La difficulté traditionnelle est le choix de la loi instrumentale $J_h(\cdot|\boldsymbol{\psi}^{[h-1]})$ à l'étape h de cet algorithme.

Formellement, on montre que le cas fréquent où on ne fait qu'une seule observation d'un matériel Σ sur la période $t = [1, T]$ peut alternativement être considéré comme un problème de séquences partiellement observées ou comme un problème de données agrégées. Les deux méthodes d'estimation sont donc possibles. Cependant, l'algorithme de Gibbs, de par la nécessité de simuler $T - 1$ données manquantes pour chacune des m trajectoires, menace de se révéler trop coûteux en temps CPU. Ainsi, dans de nombreux cas industriels, le problème posé par une estimation rapide de $\boldsymbol{\psi}$ (en particulier pour des dimensions r élevées, entre 5 et 7 dans la pratique industrielle) revient à celui de l'accélération des algorithmes MH, inhérent au choix de J_h .

3 Sélection de lois instrumentales pour MH

De façon générale, sélectionner J_h comme un produit de lois de Dirichlet semble le plus naturel pour se libérer des contraintes d'estimation liées à la nature de la matrice de transition $\boldsymbol{\psi}$. On propose trois façons de construire cette loi instrumentale, dont nous

comparons empiriquement au § 4 les propriétés de mélange sur une classe de matrices modélisant typiquement la dégradation des composants industriels. L'adaptation n'est cependant considérée qu'en temps limité afin de préserver la stationnarité des chaînes produites (Roberts et Rosenthal 2007).

Méthode 1. Dans un but simplificateur, on ne cherche d'abord pas à prendre en compte les corrélations entre les vecteurs $\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_r$. A l'étape h , la loi instrumentale de $\boldsymbol{\psi}$ est le produit de r lois de Dirichlet indépendantes du type :

$$\boldsymbol{\psi}_i^{[h]*} | \boldsymbol{\psi}_i^{[h-1]} \sim \mathcal{D}_{ir} \left(d_i \cdot \psi_{i,1}^{[h-1]}, \dots, d_i \cdot \psi_{i,r}^{[h-1]} \right).$$

L'espérance conditionnelle de $J_h(\cdot | \boldsymbol{\psi}^{[h-1]})$ prend la valeur de l'estimation précédente $\boldsymbol{\psi}^{[h-1]}$. La variance conditionnelle :

$$\mathbb{V}(\psi_{i,j}^{[h]*} | \psi_i^{[h-1]}) = \frac{\psi_{i,j}^{[h-1]}(1 - \psi_{i,j}^{[h-1]})}{d_i + 1}$$

est une fonction qui ne dépend que du seul paramètre d_i . Le taux d'acceptation de l'algorithme MH peut donc être réglé en modifiant le paramètre d_i .

Méthode 2. La deuxième méthode cherche à améliorer la prise en compte des corrélations entre vecteurs-lignes de $\boldsymbol{\psi}$ en menant un apprentissage itératif. Après quelques itérations réalisées *via* la méthode 1,

1. on simule le vecteur $\boldsymbol{\theta}^{[h]*}$ composé des éléments diagonaux $(\psi_{i,i}^{[h]*})$, $i = 1, \dots, r$, en utilisant une copule estimée empiriquement à partir des itérations précédentes ; la copule est choisie gaussienne et les marginales sont des lois Bêta :

$$\psi_{ii}^{[h]*} | \psi_{ii}^{[h-1]} \sim \mathcal{Be}(d_i \cdot \psi_{ii}^{[h-1]}, d_i(1 - \psi_{ii}^{[h-1]}));$$

2. une fois simulé l'élément $\psi_{i,i}^{[h]*}$, on simule le reste de la ligne i , à une constante $1 - \psi_{i,i}^{[h]*}$ près, selon la loi de Dirichlet

$$\mathcal{D}_{ir} \left(\frac{\psi_{i,1}^{[h-1]}}{1 - \psi_{i,i}^{[h-1]}} \cdot d_i, \dots, \frac{\psi_{i,i-1}^{[h-1]}}{1 - \psi_{i,i}^{[h-1]}} \cdot d_i, \frac{\psi_{i,i+1}^{[h-1]}}{1 - \psi_{i,i}^{[h-1]}} \cdot d_i, \dots, \frac{\psi_{i,r}^{[h-1]}}{1 - \psi_{i,i}^{[h-1]}} \cdot d_i \right)$$

puis le candidat $\boldsymbol{\psi}_i^{[h]*}$ est reconstruit après une étape de renormalisation.

Méthode 3. La troisième méthode généralise ce processus en simulant aléatoirement la position de l'élément i de $\boldsymbol{\theta}^{[h]*}$ dans la ligne i de $\boldsymbol{\psi}$ (et non plus seulement à la position i de cette même ligne). Dans les deux cas, l'apprentissage doit être mené sur des matrices $\boldsymbol{\psi}^{[h-p]}, \dots, \boldsymbol{\psi}^{[h-2]}$ suffisamment différenciées (avec p suffisamment grand) pour obtenir des estimateurs robustes de la matrice de corrélation de la copule gaussienne.

4 Essais sur une classe de matrices de dégradation

Les matrices ψ trigonales supérieures, de diagonale non-nulle et dont le dernier état est absorbant ($\psi_r = (0, \dots, 0, 1)$), peuvent bien représenter les divers états de dégradation de systèmes Σ soumis au mieux à des réparations non-rajeunissantes (*at best as bad as old*). Une règle est construite pour créer de telles matrices en fonction de la dimension r et d'un paramètre p qui correspond à la raison d'un taux de dégradation géométrique :

$$\psi_{i,j} = \begin{cases} \frac{(1-p)}{p^{i-1} - p^r} p^{j-1} & \text{si } j \geq i, \\ 0 & \text{sinon.} \end{cases}$$

Les essais numériques d'estimation sont menés à partir des choix suivants : la dimension r évolue entre 3 et 6 ; la raison p est choisie égale à 0.5. On choisit $m = 1000$ et $T = 15$. À chaque étape des algorithmes, les d_i sont simulés uniformément sur $[100, 2500]$. Enfin, 100 jeux de données sont simulés pour chaque dimension r .

En utilisant la statistique multivariée de Brooks-Gelman (Robert et Casella 2004) et une comparaison entre la matrice ψ_0 de simulation et son estimateur bayésien minimisant la perte quadratique, on peut comparer les trois méthodes par le nombre d'itérations nécessaire pour obtenir la stationnarité des chaînes selon les trois méthodes. Les étendues de chaque distribution de période de chauffe sont représentées sur la Figure 1 en fonction de la dimension r . Une évolution similaire a été constatée lorsque le taux d'acceptation des chaînes a été contrôlé empiriquement. Ces résultats sont en cohérence avec d'autres résultats de simulation montrant que le taux d'acceptation des méthodes 2 et 3 est supérieur à celui de la méthode 1, permettant une meilleure exploration de l'espace paramétrique.

5 Conclusion

Pour une classe de matrices de transition réaliste dans le cadre de la fiabilité industrielle, l'emploi d'une modélisation (même grossière) de la dépendance entre vecteurs de probabilités au sein des lois instrumentales de l'algorithme MH permet d'accélérer la procédure d'estimation *a posteriori* de ψ , au sens où la période de chauffe décroît significativement vis-à-vis du cas où l'on ne prend pas en compte la dépendance. La sélection de la copule via des tests réalisés à des instants réguliers du déroulement des chaînes, permettrait peut-être d'améliorer encore ces résultats. On a cependant noté que le temps CPU d'estimation de la copule sous-jacente peut parfois freiner l'applicabilité de la méthode dans des cas de dimension supérieure à 6.

Remerciements. Ce travail a été partiellement financé par la Direction Générale de la Compétitivité, de l'Industrie et des Services (Ministère de l'Economie, de l'Industrie et

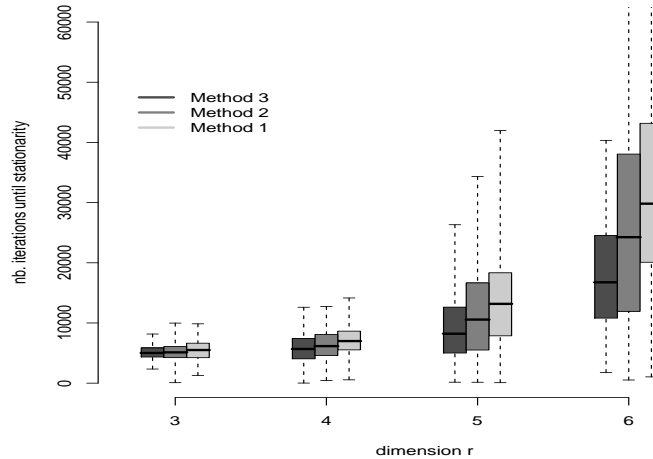


FIGURE 1 – Étendue (domaine à 95%) des périodes de chauffe des trois méthodes selon la dimension r de ψ . Les tirets correspondent aux valeurs médianes.

de l'Emploi) dans le cadre du projet CSDL (Complex Systems Design Lab) du Pôle de Compétitivité System@tic - Paris Région.

Bibliographie

- [1] Jones, M.T. (2005) Estimating Markov transition matrices using proportions data : an application to credit risk. *IMF Working Paper* 05/219. International Monetary Fund.
- [2] Lee, T.C. Judge, G.G. et Zellner, A. (1968) Maximum Likelihood and Bayesian Estimation of Transition Probabilities. *Journal of the American Statistical Association*, 63, 1162-1179.
- [3] Marin, J.M. et Robert, C.P. (2007) *Bayesian Core : A Practical Approach to Computational Bayesian Statistics*. New York : Springer.
- [4] Pasanisi, A. et Parent, E. (2004) Modélisation bayésienne du vieillissement des compteurs d'eau par mélange de classes d'appareils de différents états de dégradation. *Revue de Statistique Appliquée*, 52, 39-65.
- [5] Robert, C.P. et Casella, G. (2004) *Monte Carlo Statistical Methods*. New York : Springer.
- [6] Roberts, G.O. et Rosenthal, J.S. (2007) Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *Journal of Applied Probability*, 44, 458-475.