

PROPRIÉTÉS DES M -ESTIMATEURS DANS LES HSMM ET APPLICATIONS

S. BOUZEBDA ET N. LIMNIOS

Les chaînes semi-markoviennes cachées (HSMM, Hidden semi-Markov Model) constituent la colonne vertébrale de notre **projet ANR, INCA**.

Les HSMM généralisent les HMM dans le sens que le processus caché est une chaîne semi-markovienne au lieu d'une chaîne de Markov. Les chaînes semi-markoviennes ont l'avantage de pouvoir prendre en compte de lois de temps de séjours, dans différents états, quelconques sur l'ensemble des entiers naturels et non seulement des lois géométriques comme c'est le cas des modèles de Markov cachés (HMM), [Votsi *et al.* \(2019\)](#), [Barbu and Limnios \(2008\)](#), [Limnios and Oprüşan \(2001\)](#). Ainsi nous allons considérer un processus couple (Z, Y) , où Z est une chaîne semi-markovienne non observée et Y un processus observable. Les dépendances des deux suites stochastiques s'expriment par la relation suivante

$$P(Z_{n+1} = j, U_{n+1} = m, Y_{n+1} \in B | Z_n = i, U_n = s, Y_n = a) = P_a(i, s; j, m)R(i, a; B),$$

où i, j sont des états du processus (Z) , m, n, s des entiers naturels et a, B valeur et sous ensemble des valeurs des (Y) , P_a et R sont des noyaux de transition. Ce qui est différent de HSMM classique car les probabilités de transition du processus (Z, U) dépendent de la valeur des (Y) . C'est une généralisation, adoptée par nos partenaires de INRAE-Toulouse, afin de prendre en compte certaines dépendances particulières dans des applications spécifiques. Nous allons donc proposer des méthodes pour estimer les noyaux de transition P_a et R (Cf. processus OD-HMM thèse de Hanna Bacave, également inscrite dans ce même projet). La difficulté de ce problème consiste au fait que le processus (Z) n'est pas observé. Pour obtenir des estimateurs et leurs propriétés, nous allons utiliser une méthode très différente de celles utilisées jusqu'à présent, basée sur la théorie moderne du processus empirique. La recherche de nouvelles méthodes pour établir les propriétés asymptotiques des estimateurs est nécessaire car la complexité des dépendances et la généralité des processus, tels que les chaînes semi-markoviennes, est grande.

Les méthodes de processus empiriques sont des techniques importantes pour évaluer les propriétés asymptotiques des estimateurs basés sur des modèles (semi-)paramétriques, y compris la convergence forte, la convergence en loi et la validité du bootstrap [Pollard \(1984\)](#), [Dudley \(2014\)](#), [van der Vaart \(1998\)](#), [van der Vaart and Wellner \(1996\)](#), [Kosorok \(2008\)](#).

Les outils d'inférence (semi-)paramétrique complètent les méthodes de processus empiriques en évaluant, entre autres, si les estimateurs sont efficaces. Ici, le travail de recherche concerne les M -estimateurs, qui sont des maximiseurs approximatifs de fonctions objectives calculées à partir des données. Rappelons que les estimateurs basés sur la minimisation des fonctions objectives sont aussi trivialement des M -estimateurs après avoir pris le négatif de la fonction objectif, Bickel *et al.* (1993), van de Geer (2000), Kosorok (2008), Bouzebda and Ferfache (2021, 2023); Bouzebda *et al.* (2022a,b) Bouzebda *et al.* (2018); Papamichail *et al.* (2020). Un résultat clé central à la théorie de la M -estimation, “argmax” théorème, qui permet de dériver des limites faibles des M -estimateurs comme l’argmax du processus limite (généralement gaussien). Ce type de résultat n’existe pas dans notre contexte et sera une grande avancée dans le cadre de la théorie markovienne.

Les applications concernent en premier lieu les métapopulations (espèce dont les populations sont séparées géographiquement mais avec interactions) avec populations partiellement observables en agriculture. Les interactions sont traduites par une structure pré-établie et traduite au niveau probabiliste par des processus type Markov caché ou semi-Markov caché (Cf. processus OD-HMM thèse de Hanna Bacave, également inscrite dans ce même projet). L’objectif applicatif est donc de pouvoir décrire par une modélisation précise comment évoluent les adventices dans les différents champs agricoles. Une deuxième application concerne le traitement des données de séisme. Nous disposons de ces données depuis 1840 à nos jours pour la mer Egée. Elles nous permettront par une modélisation plus fine que ce que nous avons effectué jusqu’à présent, de mettre en évidence l’influence des différentiels des tensions terrestres (mesurables) sur l’intensité de séismes. Des applications importantes peuvent être envisagées en biologie, telles que la recherche des structures particulières dans l’ADN, comme les GpC, les séquences d’acides aminés, etc.

REFERENCES

- Barbu, V. S. and Limnios, N. (2008). *Semi-Markov chains and hidden semi-Markov models toward applications*, volume 191 of *Lecture Notes in Statistics*. Springer, New York. Their use in reliability and DNA analysis.
- Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and adaptive estimation for semiparametric models*. Johns Hopkins Series in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD.
- Bouzebda, S. and Ferfache, A. A. (2021). Asymptotic properties of M -estimators based on estimating equations and censored data in semi-parametric models with multiple change points. *J. Math. Anal. Appl.*, **497**(2), Paper No. 124883, 44.
- Bouzebda, S. and Ferfache, A. A. (2023). Asymptotic properties of semiparametric M -estimators with multiple change points. *Phys. A*, **609**, Paper No. 128363.

- Bouzebda, S., Papamichail, C., and Limnios, N. (2018). On a multidimensional general bootstrap for empirical estimator of continuous-time semi-Markov kernels with applications. *J. Nonparametr. Stat.*, **30**(1), 49–86.
- Bouzebda, S., El-hadjali, T., and Ferfache, A. A. (2022a). Central limit theorems for functional z-estimators with functional nuisance parameters. *Comm. Statist. Theory Methods*, pages 1–43.
- Bouzebda, S., Elhattab, I., and Ferfache, A. A. (2022b). General M-Estimator Processes and their m out of n Bootstrap with Functional Nuisance Parameters. *Methodol. Comput. Appl. Probab.*, **24**(4), 2961–3005.
- Dudley, R. M. (2014). *Uniform central limit theorems*, volume 142 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, New York, second edition.
- Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer Series in Statistics. Springer, New York.
- Limnios, N. and Oprüsan, G. (2001). *Semi-Markov processes and reliability*. Statistics for Industry and Technology. Birkhäuser Boston, Inc., Boston, MA.
- Papamichail, C., Bouzebda, S., and Limnios, N. (2020). Regression analysis of stochastic fatigue crack growth model in a martingale difference framework. *J. Stat. Theory Pract.*, **14**(3), Paper No. 44, 43.
- Pollard, D. (1984). *Convergence of stochastic processes*. Springer Series in Statistics. Springer-Verlag, New York.
- van de Geer, S. A. (2000). *Applications of empirical process theory*, volume 6 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.
- van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.
- van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics.
- Votsi, I., Limnios, N., Papadimitriou, E., and Tsaklidis, G. (2019). *Earthquake statistical analysis through multi-state modeling*. Mathematics and Statistics Series. ISTE, London; John Wiley & Sons, Inc., Hoboken, NJ. Statistical methods for earthquakes set. Vol. 2.

(S. Bouzebda et N. Limnios) UNIVERSITÉ DE TECHNOLOGIE DE COMPIÈGNE,, LMAC (LABORATORY OF APPLIED MATHEMATICS OF COMPIÈGNE),, CS 60 319 - 60 203 COMPIÈGNE CEDEX
Email address: Salim.Bouzebda@utc.fr; Nikolaos.Limnios@utc.fr