

Description du projet de thèse :

-- Inférence statistique pour des processus markoviens et semi-markoviens basée sur des mesures de divergence --

Direction : Vlad Barbu

OBJECTIFS

Le sujet de recherche que nous proposons concerne des aspects statistiques pour des processus stochastiques (estimation, tests, sélection de modèle), en se basant principalement sur des outils issus de la théorie de l'information, comme les mesures de divergence et l'entropie. Il s'agit d'un domaine de la statistique moderne avec un important potentiel pour des perspectives et des développements futurs. Les objectifs de ce projet de recherche sont de proposer de nouvelles mesures de divergence pour des processus markoviens ou semi-markoviens, d'élaborer des techniques d'inférence statistique associées et de développer quelques applications de ces résultats théoriques.

CONTEXTE

Depuis l'introduction de la notion d'entropie comme mesure de l'information d'une loi de probabilité (Shannon, 1948), cette notion a été non seulement généralisée à d'autres mesures d'entropie, mais elle a été aussi étendue afin de mesurer l'information mutuelle entre deux lois de probabilité, obtenant ce que l'on appelle une mesure de divergence. Ainsi, un nombre très important de mesures de divergence a été proposé dans la littérature. De telles mesures de divergence sont utilisées comme indices de similarité ou dissimilarité entre deux populations ou distributions. Parmi les exemples classiques nous pouvons citer la statistique du Chi-deux, la divergence de Kullback-Leibler, la distance de Hellinger, la statistique F2 de Freeman-Tukey. Il faut souligner que les versions limite des mesures de divergence, connues sous le nom de taux de divergence, sont également importantes. À part les nombreuses propriétés des mesures de divergence, ces taux de divergence peuvent être utilisés dans l'inférence statistique exactement de la même manière que les mesures classiques (non-limite). Pour une présentation détaillée des diverses propriétés des taux de divergence, nous nous référons à Gray (1990). Un point important à souligner est que ces divergences peuvent être utilisées afin de construire des critères de sélection de modèle (Akaike, 1973; Konishi et Kitagawa, 2008; Mattheou et al., 2009).

Il faut noter que la plupart de ces développements pour les divergences et entropies ont lieu dans un cadre classique et assez restrictif de la statistique, c'est-à-dire pour des données indépendantes (et, en général, aussi identiquement distribuées). Beaucoup moins d'attention est portée sur les divergences et entropies dans des contextes où la dépendance est prise en compte, par le biais, le plus souvent, de processus stochastiques (Rached et al. 2001, 2004). Des travaux récents ont investigué des mesures généralisées de divergence pour des chaînes de Markov (Ciuperca et al., 2011; Barbu et al., 2017b, 2018b), en considérant des généralisations de la mesure Alpha-divergence (Amari et Nagaoka, 2000) et de la mesure Beta-divergence (Basu et al., 1998). Ce travail a été continué dans Barbu et al. (2018a) en étudiant les mesures de divergence généralisées pondérées correspondantes et les taux associés (cf., e.g., Belis et Guiasu, 1968; Taneja et Tuteja, 1984).

DIRECTIONS DE TRAVAIL

Une première direction de travail pour ce projet doctoral est d'étendre ces résultats aux processus semi-markoviens (cf., e.g., Barbu et Limnios, 2006; 2008), qui généralisent les processus de Markov, dans le sens que le temps de séjour dans un état peut suivre un loi de probabilité arbitraire, contrairement au cas markovien, où le temps de séjour suit une loi exponentielle ou géométrique (si le processus est à temps continu, respectivement à temps discret). Pour cette raison, les processus semi-markoviens représentent, d'un point de vue théorique, des objets plus complexes que les processus de Markov et donc très intéressants à investiguer. D'autre part, d'un point de vue appliqué, les processus semi-markoviens sont plus adaptés comme outils de modélisation pour des problèmes concrets d'application.

Une seconde direction de travail est de proposer de nouvelles mesures de divergence pour des processus stochastiques, avec une attention particulière aux processus markoviens et semi-markoviens. Le calcul de taux de divergence comme mesure de dissimilarité entre deux modèles sera aussi abordé.

Réaliser de l'inférence statistique, en général (Pardo, 2006), ou des tests d'hypothèses, en particulier, en se basant sur des mesures de divergence ou sur des taux associés est un des objectifs de ce type d'études (cf. Cressie et Read, 1984; Lee et al., 2011; Lee, 2013; Mattheou et Karagrigoriou, 2010; Read et Cressie, 1988; Vonta et al., 2012; Zografos et al., 1990).

Utiliser des mesures de divergence afin de construire des tests statistiques est une autre direction naturelle de ce travail de thèse. Ceci peut être réalisé, par exemple, dans la suite du travail développé en Barbu et al. (2019) pour construire des tests d'ajustement basés sur les entropies de Tsallis et Kaniadakis.

Des applications intéressantes pour les résultats théoriques développés pendant ce travail doctoral concernent l'analyse de survie (dans le domaine biomédical ou actuariat) et la fiabilité des systèmes (en ingénierie). Développer un/des package(s) R ou Python pour implémenter les programmes élaborés pendant la thèse (comme Barbu et al., 2017a pour la modélisation et l'estimation des modèles semi-markoviens à temps discret) peut être également d'intérêt.

COLLABORATIONS INTERNATIONALES LIEES AU SUJET DE THESE

Le travail doctoral va se réaliser en collaboration avec les partenaires internationaux du directeur de thèse, experts dans des sujets abordés dans ce projet:

- G. D'Amico, Université "G. D'Annunzio" de Chieti-Pescara, Italie
- A. Karagrigoriou, Université Aegean, Samos, Grèce
- V. Preda, Université de Bucarest et Institut de Statistique Mathématique et Mathématiques Appliquées "Gheorghe Mihoc-Caius Iacob" de l'Académie Roumaine, Bucarest, Roumanie

BIBLIOGRAPHIE

- H. Akaike, Information theory and an extension of the maximum likelihood principle, Proceeding of the Second International Symposium on Information Theory, B.N. Petrov and F. Csaki (eds.), Akademia Kaido, Budapest, 267-281, 1973
- S. M. Ali, S. D. Silvey, A general class of coefficients of divergence of one distribution from another, J. Roy. Statist. Soc. B, 28, 131-142, 1966
- S. Amari, H. Nagaoka, Methods of Information Geometry, Oxford University Press, New York, 2000
- V. S. Barbu, C. Bérard, D. Cellier, M. Sautreuil, N. Vergne, Package R SMM, Simulation and Estimation of Multi-State Discrete-Time Semi-Markov and Markov Models, 2017, <https://cran.r-project.org/web/packages/SMM>
- V. S. Barbu, A. Karagrigoriou, V. Preda, Entropy and divergence rates for Markov chains : I. The Alpha-Gamma and Beta-Gamma case, Proceedings of the Romanian Academy-series A : Mathematics, Physics, Technical Sciences, Information Science, 18(4), 293-301, 2017
- V. S. Barbu, A. Karagrigoriou, V. Preda, Entropy and divergence rates for Markov chains : II. The weighted case, Proceedings of the Romanian Academy-series A : Mathematics, Physics, Technical Sciences, Information Science, 19(1), 3-10, 2018
- V. S. Barbu, A. Karagrigoriou, V. Preda, Entropy and divergence rates for Markov chains : III. The Cressie and Read case and applications, Proceedings of the Romanian Academy-series A : Mathematics, Physics, Technical Sciences, Information Science, 19(3), 413-421, 2018
- V. S. Barbu, A. Karagrigoriou, V. Preda, Maximum Tsallis and Kaniadakis entropy type tests of fit, en cours (phase de rédaction), 2019
- V. S. Barbu, N. Limnios, Empirical estimation for discrete time semi-Markov processes with applications in reliability, Journal of Nonparametric Statistics, 18(7-8), 483-498, 2006
- V. S. Barbu, N. Limnios, Semi-Markov Chains and Hidden Semi-Markov Models toward Applications - Their use in Reliability and DNA Analysis, Lecture Notes in Statistics, vol. 191, Springer, New York, 2008
- A. Basu, I. R. Harris, N. L. Hjort, M. C. Jones, Robust and efficient estimation by minimising a density power divergence, Biometrika, 85, 549-559, 1998
- M. Belis, S. Guiasu, A quantitative-qualitative measure of information in cybernetic systems, IEEE Transactions on Information Theory, 14(4), 593-594, 1968
- G. Ciuperca, V. Girardin, L. Lhote, Computation and estimation of generalized entropy rates for denumerable Markov chains, IEEE Transactions On Information Theory, 57(6), 2011
- N. Cressie, T. R. C. Read, Multinomial goodness-of-fit tests, J. R. Statist. Soc., 5, 440-454, 1984
- I. Csizsár, Eine informationstheoretische ungleichung und ihre anwendung auf den beweis der ergodizität von markoffschen ketten, Publication of the Mathematical Institute of the Hungarian Academy of Sciences, 8, 84-108, 1963
- R. M. Gray, Entropy and Information Theory, New York, Springer-Verlag, 1990
- J. N. Kapur, Measures of Information and Their Applications, Wiley, New Delhi, 1994
- S. Konishi, G. Kitagawa, Information Criteria and Statistical Modeling, Springer, 2008
- S. Lee, I. Vonta, A. Karagrigoriou, A maximum entropy type test of fit, Comput. Statist. Data Anal., 55, 2635-2643, 2011
- S. Lee, A maximum entropy type test of fit: Composite hypothesis case, Comput. Statist. Data Anal., 57, 59-67, 2013

K. Mattheou, S. Lee, A. Karagrigoriou, A model selection criterion based on the BHHJ measure of divergence, *Journal of Statistical Planning and Inference*, 139, 228-235, 2009

K. Mattheou, A. Karagrigoriou, A new family of divergence measures for tests of fit, *Austr. and N. Zealand J. of Statist.*, 52(2), 187-200, 2010

L. Pardo, *Statistical Inference Based on Divergence Measures*, Chapman and Hall, 2006

Z. Rached, F. Alajajai, L. L. Campbell, Rényi's divergence and entropy rates for finite alphabet Markov sources, *IEEE Transc. Inf. Theory*, 47(4), 1553-1561, 2001

Z. Rached, F. Alajajai, L. L. Campbell, Kullback-Leibler divergence rate between Markov sources, *IEEE Transc. Inf. Theory*, 50(5), 917-921, 2004

T. R. C. Read, N. Cressie, *Goodness-of-Fit Statistics for Discrete Multivariate Data*, New York, Springer-Verlag, 1988

C. E. Shannon, A mathematical theory of communication, *Bell System Technical Journal*, 27, 379-423, 1948

H. C. Taneja, R. K. Tuteja, Characterization of a quantitative-qualitative measure of relative information, *Information Sciences*, 33, 217-222, 1984

F. Vonta, K. Mattheou, A. Karagrigoriou, On properties of the (Φ, α) -power divergence family with applications in goodness of fit tests, *Methodol. Comput. Appl. Probab.*, 14, 335-356, 2012

K. Zografos, K. Ferentinos, T. Papaioannou, Phi-divergence statistics: Sampling properties, multinomial goodness of fit and divergence tests, *Comm. in Statist. Theor. Meth.*, 19(5), 1785-1802, 1990