PhD Position in Statistics

TITLE: Statistical Inference for Markov and semi-Markov Processes based on Divergence Measures

LABORATORY: Laboratory of Mathematics Raphaël Salem (LMRS) UMR 6085, University of Rouen Normandy, Avenue de l'Université, BP.12, 76801 Saint Étienne du Rouvray, France

PHD SUPERVISOR: Vlad Stefan BARBU (MCF-HDR), LMRS, +33/(0)2/32/95/52/31, barbu@univ-rouen.fr, http://lmrs.univ-rouen.fr/Persopage/Barbu/

SCHOLARSHIP: Research PhD scholarship from the French Ministry of Scientific Research; approx. 1700 €/month gross (approx. 1400 €/month net) for 3 years starting from September 2018

HOW TO APPLY: Please send your application file (including CV + grade transcripts of Master and undergraduate years + cover letter + at least one recommendation letter) to barbu@univ-rouen.fr

REQUIRED PROFILE: The successful candidate should have a master degree in Mathematics with a major in Statistics or related field. Strong skills in stochastic processes, Markov (related) processes, asymptotic statistical methods, parametric and nonparametric statistics are needed. Basics in statistical programming (in R and/or Matlab and/or Python) will be appreciated. Previous experience in statistical applications is also welcome.

ABSTRACT: The research topic that we propose for this thesis concerns statistical aspects for stochastic processes (estimation, tests, model selection), based on tools that come from information theory (divergence measures, entropies). This is a field of modern statistics with high potential for future development and perspectives. To be

more specific, the purpose of this research project is to propose new divergence measures for Markov and semi-Markov processes, to develop associated statistical inference techniques and to undertake some applications of these theoretical results.

KEYWORDS: mathematical statistics, entropy, divergence measures, statistical estimation, hypothesis testing, model selection, applications (reliability, survival analysis), information theory

SUBJECT: Since the introduction of the notion of entropy as a measure of information for a probability distribution (Shannon, 1948), this notion has been quickly not only generalized to other entropy measures but also extended to measure the mutual information concerning two distributions and a plethora of the so-called divergence measures was introduced. Such measures of divergence are used as indices of similarity or dissimilarity between populations, measuring the distance or the discrepancy between two distributions. Classical examples are the well-known Chi-square statistics, Kullback-Leibler divergence, Hellinger distance, Freeman-Tukey's F^2 statistics, etc. Equally important are the limiting versions of divergence measures, known as divergence rates. Besides the numerous limiting properties of these special divergences, such rates can be used in statistical inference in exactly the same manner as the typical (non-limiting) divergence measures. For a comprehensive review of various properties of the rates we refer to Gray (1990). It should also be noted that such measures are used for the construction of model selection criteria (Akaike, 1973; Konishi and Kitagawa, 2008; Mattheou et al., 2009).

Much less attention is paid to divergence measures and entropies for stochastic processes (Rached et al. 2001, 2004). Recent works have investigated generalized divergence measures for Markov chains (Ciuperca et al., 2011; Barbu et al., 2017b, 2018b), by considering generalizations of Alpha divergence measure (Amari and Nagaoka, 2000) and Beta divergence measures (Basu et al., 1998). That work was continued in Barbu et al. (2018a) by the study of the corresponding weighted generalized divergence measures and the associated rates (following the line of weighted divergences, cf., e.g., Belis and Guiasu, 1968; Taneja and Tuteja, 1984).

One interest of this PhD project is to extend these results to semi-Markov processes (cf., e.g., Barbu and Limnios, 2006; 2008), that generalize Markov processes, in the sense that the sojourn time in a state is arbitrarily distributed, as opposed to the Markov case, where it follows an Exponential or Geometric distribution. For this reason, the semi-Markov processes are more complex and interesting theoretical objects, as well as more adapted to practical applications.

A second direction is to propose new divergence measures for stochastic processes, with a special attention paid to Markov and semi-Markov processes. The computation of divergence rates as measures of discrimination between two models will also be addressed.

Undertaking statistical inference, in general (Pardo, 2006), or testing statistical hypothesis, in particular, based on divergence measures or associated rates is one of the interests of these types of study (cf. Cressie and Read, 1984; Lee et al., 2011; Lee, 2013; Mattheou and Karagrigoriou, 2010; Read and Cressie, 1988; Vonta et al., 2012; Zografos et al., 1990). Applying the divergence measures to construct statistical tests is a natural research direction of this thesis.

Depending on the interests of the candidate, potential applications of the theoretical results developed during this doctoral work could be in survival analysis, reliability theory and engineering. Developing an associated R package (as Barbu et al., 2017a for discrete-time semi-Markov modeling and estimation) could be also of interest.

REFERENCES

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(2), 1-8 (to appear), 2018

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. Csiszár, 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 (Phi, a)-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