Report on Marek Omelka's Habilitation Thesis

Nonparametric Estimation of Copulas, Conditional Copulas, and Conditional Distribution Functions

Marc Hallin, August 31, 2017

Marek Omelka defended his PhD thesis in 2006 from Charles University in Prague. Since then, he has developed a high quality research activity, publishing some 25 papers in top tier journals, among which the *Annals of Statistics*, the *Journal of the Royal Statistical Society Series B*, the *Journal of Multivariate Analysis*, the *Annals of the Institute of Statistical Mathematics*, the *Scandinavian Journal of Statistics*.

Omelka's habilitation thesis essentially summarizes the contributions made in five articles, published between 2009 and 2015 in the *Annals of Statistics*, the *Journal of the Royal Statistical Society Series B*, *Computational Statistics & Data Analysis*, and the *Scandinavian Journal of Statistics*, in collaboration with Irene Gijbels (KULeuven) and Noël Veraverbeke (University of Hasselt). That set of five articles constitutes an in-depth study of estimation problems for copulas, and a remarkable contribution to the literature on nonparametric estimation methods.

Copulas have become increasingly popular in the statistical literature and in such fields of application as quantitative finance, risk modeling or machine learning. The copula transformation (obtained via componentwise univariate probability-integral transformations) indeed nicely disentangles the dependence and marginal features of multivariate distributions, hence constitutes a perfect tool for the analysis of complex dependencies.

A copula is a multivariate probability distribution over the unit cube, with uniform marginals. Estimating a copula thus reduces to the problem of estimating a multivariate distribution function under marginal uniformity constraints. It is a well known fact that smoothing (stepwise) empirical distribution functions in general improves their performances as estimators of continuous population distribution functions, and an abundant literature has been devoted to this smoothing problem, mostly in the univariate case. Unfortunately, standard nonparametric methods, such as local linear kernels, when applied to the estimation of a copula, do not work well—due, essentially, to the special role of the corners and boundaries of the unit cube. The objective in Omelka's thesis is to palliate the deficiency of standard methods by proposing copula-specific alternatives in various conditional and unconditional contexts.

Omelka et al. (2009) introduce (Chapter 2) a clever strategy based on bandwidth shrinking and a transformation technique. Shrinking the bandwidth in the vicinity of the unit cube boundaries considerably reduces, for classical copula families, the order of magnitude of the bias and variance of nonparametric estimators. The transformation idea, which fully exploits the marginal invariance properties of empirical copulas, is particularly ingenious, and somewhat similar to the use of scores in rank-based inference[1].

[1]The results would not be affected if difstinct score functions Φ_j^{-1} were considered for distinct marginals. An investigation of the impact of those score functions on asymptotic performances would be welcome. Are data-

Observations in practice seldom are isolated, though, and usually come in association with covariates. Those covariates potentially carry crucial information, and cannot be left out in the analysis. The subsequent three chapters accordingly are devoted to copula estimation problems in the presence of covariates: nonparametric estimation of conditional copulas then is the objective, with challenging additional difficulties. This objective is of essential importance for applications and, to the best of my knowledge, apart perhaps for some recent Bayesian attempts, the results by Omelka and his coauthors are the only ones available in the literature. Chapter 3 considers the general case; Chapter 4 deals with particular cases where simplifying assumptions can be made and exploited, while Chapter 5 somewhat broadens the scope to the estimation of conditional distribution functions.

A superficial approach of the problem could lead to the hasty conclusion that the problem for conditional copulas boils down to considering, in the unconditional approach, an additional smoothing over the space of covariates. This, however, is quite misleading, as it overlooks the (very likely) impact of covariates on conditional marginals. Gijbels, Veraverbeke and Omelka (2011) therefore suggest a combination of marginal and global smoothing (still, over the space of covariates), and establish the asymptotic properties of the resulting empirical processes under appropriate regularity conditions on the impact of the regressors. Deriving those results is particularly delicate, and constitutes a remarkable and beautiful piece of work in nonparametric asymptotics.

Reading this thesis was quite inspiring and pleasant, and I hope the author will proceed with further exploiting his investigation. A fundamental question indeed is: why should we consider (nonparametric) estimation of conditional copulas? And the answers of course are many, although the copula itself is seldom the ultimate object of interest. Some answers are given attention in Chapter 3, where the dependence on the covariates of some measures of conditional association is investigated. This is but a first step, and it would be interesting, for instance, to look further into (possibly, nonlinear) conditional principal directions, or conditional depth and depth contours; into tests for constant or monotone (increasing or decreasing) conditional dependence; into tests for homogeneity of copula structures (unspecified conditional marginals); into time-series extensions; ... Marek Omelka' s thesis in that respect is opening the door to a rich and diversified spectrum of perspectives—which is the sure sign of a good dissertation.

Summing up, this thesis and its author's subsequent publications not only demonstrate that Marek Omelka, beyond any doubt, possesses the scientific skills usually required for habilitation and promotion to (Associate) Professor, but also announce a productive, bright, and hopefully successful scholarly and academic career.

I very warmly recommend Marek Omelka to the Committee.

Marc Hallin

ECARES and Department of Mathematics Université libre de Bruxelles

Royal Academy of Belgium

-driven scores allowed (e.g., on the model of the score selection ideas proposed, in the context of rank-based inference for regression, by Dodge and Jurečková (*Adaptive Regression*, Springer-Verlag 2000))?