Semiparametrically Efficient One-Step R-Estimation

Marc Hallin[1], and Davy Paindaveine[1]

[1] Mathematics Department, ISRO, and ECARES, Université Libre de Bruxelles, Bruxelles, Belgium

Abstract

Despite a long history, R-estimation methods, unlike rank tests, never made their way to applications. And, even among the experts of rank-based methods, a pretty widespread opinion is that "ranks are fine for testing but not for estimation".

The reasons for this lack of symmetry between estimation and testing are twofold. Practical reasons first: unlike rank test statistics, R-estimators in general are not given under explicit closed forms, but follow from unpleasant optimization procedures, involving discrete-valued objective functions. More fundamental reasons, too: consistency and asymptotic normality proofs are rather elaborate, and restricted to some traditional cases. And, asymptotic variances of R-estimators typically depend on the unknown underlying density. Such variances cannot be computed exactly, and cannot be estimated easily. As a result, R-estimators, contrary to rank tests are seldom considered in practice.

Statistical decision theory however suggests that the advantages of rank-based methods depend on the local properties of the model under study, not on the specific inference problem under consideration. From the point of view of the asymptotic theory of statistical experiments, once a type of scores (Wilcoxon, van der Waerden, Laplace, ... ; in general, this choice is associated with some "target density" f) has been chosen, the rank transformation simply consists in mapping the original sequence of statistical experiments $\mathcal{E}^{(n)}$, say, onto another sequence $\mathcal{E}^{(n)}_{f,g}$, where g is the actual unknown density of the observations. The local properties of the resulting model are considered attractive from the point of view of hypothesis testing (distribution-freeness and invariance, local powers). These properties belong to the corresponding local Gaussian shift experiments, hence are fully characterized by $\mathcal{E}^{(n)}_{f,g}$'s information matrices. Typically, the performance of optimal estimators in such models is measured by a covariance matrix which is the inverse of the information matrix characterizing the noncentrality parameter, under local alternatives, of the chi-square distributions of optimal test statistic. If these information matrices are attractive from the point of view of hypothesis testing, *they should be equally attractive from the point of view of point estimation.* Actually, it has been shown by Hallin and Werker (2002) that, under very general assumptions, these matrices coincide with the semiparametrically efficient information matrices.

Now, the practical problems related with the implementation of R-estimation remain. In a sense, they are the same as for the implementation of most M-estimators, including the maximum likelihood ones: as a rule, no explicit form is provided, and the estimator results from the minimization (maximization) of some rank-based objective function. In R-estimation, however, the form of objective functions, which are intrinsically piecewise constant, creates some additional trouble. To the best of our knowledge, the problems resulting from this discrete nature of rank-based objective functions have been solved for location and linear regression models only (the seminal paper in this direction is Jurečková 1971). An unsuccessful attempt has been made in the context of linear (ARMA) time-series models (Allal et al. 2001), who explain why the classical rank-based objective function approach fails in that case.

The usual way to escape numerical optimization, and to provide closed form versions of asymptotically optimal estimators is (in the context of LAN experiments) Le Cam's *one-step* method. Implementation of this method in the present context however runs into the same difficulties as the estimation of asymptotic covariance matrices of R-estimators. We are showing how a very intuitive local maximum likelihood argument allows for a simple and feasible solution.

Applications include classical R-estimation procedures but also less traditional ones, such as those involving serial rank statistics, rank-and-sign statistics, or multivariate, hyperplane-based signed rank statistics.