Inference on a Distribution Function from Ranked Set Samples

Prof. Dr. Lutz Dümbgen
(Universität Bern)

November 6, 2013

Consider independent observations $(X_1, R_1), (X_2, R_2), \ldots, (X_n, R_n)$ with random or fixed ranks R_i in $\{1, 2, ..., k\}$, while conditional on $R_i = r$, the random variable X_i has the same distribution as the r-th order statistic within a random sample of size k from an unknown continuous distribution function F. Such observation schemes are utilized in situations in which ranking observations is much easier than obtaining their precise values. Two wellknown special cases are ranked set sampling (McIntyre 1952) with $k = n$ and $R_i = i$, and judgement post-stratification (MacEachern et al. 2004) with R_i being uniformly distributed on $\{1, 2, ..., k\}$.

Within a rather general setting we analyze and compare the asymptotic distribution of three different estimators of the distribution function F. The asymptotics are for fixed number k and n tending to infinity. The three estimators under consideration are the stratified estimator of Stokes and Sager (1988), a new moment-based estimator, and the nonparametric maximum-likelihood estimator of Kvam and Samaniego (1994). Our results on the latter estimator generalize and refine the analysis of Huang (1997). It turns out that it is asymptotically more efficient than the former two estimators. The efficiency gain over the new estimator is typically rather small, whereas the inefficiency of the stratified estimator may be quite severe in imbalanced settings.

Finally we describe briefly pointwise and simultaneous confidence intervals for the distribution function F with guaranteed coverage probability for finite sample sizes. This may be accomplished by considering the conditional distribution of the data, given the ranks $R_1, R_2, \ldots, R_n$. In particular, pointwise confidence intervals for F may be obtained by adapting a method of Terpstra and Miller (2006).

This is joint work with Ehsan Zamanzade.