BAYESIAN EMPIRICAL LIKELIHOOD IN SMALL AREA ESTIMATION

SANJAY CHAUDHURI

Large scale countrywide sample surveys often contain very few observations from *small areas* like a specific state, a specific county etc. Accurate estimation of such small area level parameters are of major interest in many applications. In order to reduce the variability of such area level parameter estimates, one needs to increase the number of observations by borrowing information from several similar small areas. Bayesian methods provide a natural way to borrow such information. However, virtually all such Bayesian methodologies are based on parametric likelihoods. These parametric assumptions are mostly made for analytical convenience and cannot be justified easily.

Our main objective is to explore an alternative semi-parametric Bayesian approach based on empirical likelihood. We also develop the methodologies required to analyze and justify such procedures.

In Chaudhuri and Ghosh (2011) (Biometrika, Vol 98, number 2, pp 473 - 480) we discuss how to use empirical likelihood in a Bayesian context in small area estimation. We consider the *Fay-Herriot* model to estimate the median income of four person families in 50 US states and the district of Columbia. It turns out that our method produces better estimates than the available ones (Table 1).

Use of empirical likelihood in Bayesian procedures is an emerging area in current statistics. We hope this article will increase interest in such procedures and motivate their application in different areas of statistics.

Estimate	Average relative bias	Average squared relative bias	Average absolute bias	Average Squared deviation
CPS	.0735	.0084	$2,928.82$	$13,811,122.39$
HB	.0363	.0021	$1,457.47$	$3,468,496.61$
L_{E_1}	.0348	.0019	$1,325.25$	$2,638,343.00$
L_{T_1}	.0349	.0019	$1,334.50$	$2,642,001.00$
L_{E_2}	.0348	.0019	$1,326.02$	$2,638,143.00$
L_{T_2}	.0349	.0019	$1,337.38$	$2,644,497.00$

TABLE 1. Comparison of estimates for *Fay-Herriot* data. CPS and HB correspond to the estimates obtained from the Current Population Survey and the hierarchical Bayes procedure by Ghosh, Nangia and Kim (1996) (Journal of the american statistical association, vol 91, pp 1423 - 1431). L_E and L_T respectively denotes empirical and exponentially tilted empirical likelihood.(L_{E_1}, L_{T_1}) and (L_{E_2}, L_{T_2}) correspond to two different sets of priors distributions.