

Lecture 3: Tuesday, May 20 – 10:00 AM

Horwitz-Thompson Estimation for Semiparametric Models and Two-phase Stratified Samples, with Application to Case-Cohort Studies

We consider semiparametric models for which solution of Horvitz-Thompson or inverse probability weighted (IPW) likelihood equations with two phase stratified samples leads to root N consistent and asymptotically Gaussian estimators of both Euclidean and nonparametric parameters. For Bernoulli (i.i.d.) sampling, standard theory shows that the Euclidean parameter estimator is asymptotically linear in the IPW influence function. By proving weak convergence of the IPW empirical process, and borrowing results on weighted bootstrap empirical processes, we derive a parallel asymptotic expansion for finite population stratified sampling at phase two. Variances of estimated regression coefficients are the sum of two asymptotically independent terms: (1) the model based variance of coefficients that would be estimated were complete data available for all subjects in the main cohort (phase one sample); and (2) the design based variance of the Horwitz-Thompson estimate of the sum of main cohort influence function contributions using subjects sampled at phase two. Adjustment of standard sampling weights by the sample survey technique of regression calibration, or by estimation in a parametric model for the probability of inclusion in the phase two sample, can sometimes dramatically lower the design based phase two variances. These results are applied to estimation of hazard ratios in stratified case-cohort studies.