Targeted Maximum Likelihood Learning:
Application to Airpollution

Mark van der Laan
University of California, Berkeley

ABSTRACT: We are concerned with estimating the causal effect of a decrease in ozone in the LA basin on hospital discharges due to asthma for ages birth to 19 years based on data consisting of quarterly measurements of ozone, asthma hospital discharges and other covariates from 1983 through 2000 inclusively over 195 geographical units. The immediate effect of ozone during a given quarter on asthma hospital discharges during that same quarter, adjusted for a user supplied set of covariates/confounders W, is modeled with a W-adjusted variable importance parameter defined in van der Laan (2005). By assuming a model only for the terms involving ozone and leaving the remainder of the regression unspecified we obtain a semi-parametric regression model in which the parametric component represents the adjusted effect of ozone of interest. We estimate the parametric model parameter with a new targeted maximum likelihood procedure which maps an initial maximum likelihood density estimator into a solution of the efficient influence curve estimating equation with an increased likelihood (and thereby less bias for the parametric component). This is done iteratively, by creating a hardest parametric sub-model with parameter epsilon through the given density estimator with score equal to the efficient influence curve at the density estimator of the parametric component of the semi-parametric regression model. This epsilon is then estimated with the maximum likelihood estimator, and finally a new density estimator is defined as the corresponding update of the original density estimator. This general targeted Maximum Likelihood Estimation procedure is an improvement over the maximum likelihood fit of the semi-parametric regression which suffers from bias for the parametric part due to over-smoothing/bias in the nonparametric part of the semi-parametric regression model. We apply this approach to obtain a W-adjusted variable importance estimate for user supplied subsets W (including singletons) with ozone, one at a time. We then select those subsets of covariates with significant associations using re-sampling based multiple testing adjustments.