

How good is the semiparametric-efficient estimator if the model is nearly correct?

Thomas Lumley¹

¹ Department of Biostatistics, University of Washington, USA

Inference in the presence of subsampled, missing, or otherwise incomplete data is described in great generality by [1]. They present two classes of estimators. The first is the augmented inverse-probability weighted (AIPW) estimators. These are consistent for the same parameter that would be estimated with complete data, whether or not the complete-data model is correctly specified. These are equivalent to the generalized raking estimators [2] from the survey literature. The second class contains the semiparametric efficient estimator. This is typically more efficient than any AIPW estimator, but is typically not consistent for the same parameter that would be estimated in complete data unless the complete-data model is correctly specified. When the complete-data model is grossly misspecified, ordinary care in data analysis will detect the problem. As a result the impact of model misspecification has been studied only in a few special cases (eg [3]). However, the difference between the efficient estimator and an AIPW estimator is asymptotically small, so it is not clear if the efficiency advantage of the efficient estimator is maintained under even minor model misspecification. I discuss the case when the complete-data model is misspecified, but by a small enough amount that no model misspecification test can reliably detect it. This 'nearly-true' model is represented asymptotically by the concept of mutually contiguous series of models. I show that the efficiency advantage of the efficient estimator is not in fact maintained under contiguous model misspecification. I discuss applications where this may be relevant in epidemiology and in genetic association studies.

References

- [1] Robins JM, Rotnitzky A, Zhao LP (1994) Estimation of Regression Coefficients When Some Regressors Are Not Always Observed. *J Amer Stat Assoc.* 89: 846–866.
- [2] Deville J-C, Sarndal C-E, Sautory O (1993) Generalized Raking Procedures in Survey Sampling. *JASA* 88:1013-1020
- [3] Scott, AJ. and Wild, CJ., "On the robustness of weighted methods for fitting models to case-control data", *Journal of the Royal Statistical Society, B*, 64, 207-219, 2002.