Correcting Selection Bias via Functional Empirical Bayes

Yingying Fan, Ph.D.
Dean’s Associate Professor in Business Administration
Data Sciences and Operation
University of Southern California
Los Angeles, CA

Abstract:
Selection bias results from the sampling of extreme observations and is a well recognized issue for standard scalar or multivariate data. Numerous approaches have been proposed to address the issue, dating back at least as far as the James-Stein shrinkage estimator. However, the same potential issue arises, albeit with additional complications, for functional data. Given a set of observed functions, one may wish to select for further analysis those which are most extreme according to some metric such as the average, variance, or maximum value of the function. However, given that functions are often noisy realizations of some underlying mean process, these outliers are likely to generate biased estimates of the quantity of interest. In this paper we propose an Empirical Bayes approach, using a variant of Tweedie’s formula, to adjust such functional data to generate approximately unbiased estimates of the true mean functions. Our approach has several advantages. It is non-parametric in nature, but is capable of automatically shrinking back towards a James-Stein type estimator in low signal situations. It is also computationally efficient and possesses desirable theoretical properties. Furthermore, we demonstrate through extensive simulations and a real data analysis that our approach can produce significant improvements in prediction accuracy relative to possible competitors.

Biography:
Yingying Fan is Dean's Associate Professor in Business Administration in Data Sciences and Operations Department of the Marshall School of Business at the University of Southern California, Associate Professor in Departments of Economics and Computer Science at USC, and an Associate Fellow of USC Dornsife Institute for New Economic Thinking (INET). She received her Ph.D. in Operations Research and Financial Engineering from Princeton University in 2007 under the supervision of Jianqing Fan. She was Lecturer in the Department of Statistics at Harvard University from 2007-2008. Her research interests include high-dimensional statistics, big data problems, high-dimensional classification, large-scale inference and false discovery rate control, statistical machine learning, networks, causal inference, nonparametric statistics, financial econometrics and business applications, and deep learning.