Generalized Multilevel Function-on-Scalar Regression and Principal Component Analysis

Jeff Goldsmith*
Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, United States, jeff.goldsmith@columbia.edu

Vadim Zipunnikov
Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, United States, vzipunni@jhsph.edu

Jennifer Schrack
Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, United States, jschrac1@jhu.edu

This manuscript considers regression models for generalized, multilevel functional responses: functions are *generalized* in that they follow an exponential family distribution and *multilevel* in that they are clustered within groups or subjects. This data structure is increasingly common across scientific domains and is exemplified by our motivating example, in which binary curves indicating physical activity or inactivity are observed for nearly six hundred subjects over five days. We use a generalized linear model to incorporate scalar covariates into the mean structure, and decompose subject-specific and subject-day-specific deviations using multilevel functional principal components analysis. Fixed effect coefficient functions and principal component basis functions are estimated using penalized splines; model parameters are estimated in a Bayesian framework. Simulations designed to mimic the application have good estimation and inferential properties with reasonable computation times for moderate datasets, in both cross-sectional and multilevel scenarios; code is publicly available. In the application we identify effects of age and BMI on the time-specific change in probability of being active over a twenty-four hour period; in addition, the principal components analysis identifies the patterns of activity that distinguish subjects and days within subjects.

Keywords: Accelerometry, Bayesian Inference, Generalized Functional Data, Hamiltonian Monte Carlo, Penalized Splines.