

SIMULTANEOUS INFERENCE FOR MEAN CURVES OF FUNCTIONAL AND LONGITUDINAL DATA: A UNIFIED THEORY

Zhipeng Lou* and Wei Biao Wu

University of Pittsburgh and The University of Chicago

Abstract: The paper considers simultaneous statistical inference for mean curves of functional and longitudinal data in a unified framework. We establish the asymptotic distribution for the normalized maximum deviations of the local linear estimators from the true mean functions. The asymptotic distribution leads to simultaneous confidence bands with asymptotically correct coverage probabilities. A Gaussian multiplier bootstrap procedure is proposed to obtain the cutoff values and our fully data-driven approach has a good finite sample performance. All the results obtained in the present paper are unified with respect to the sampling schemes.

Key words and phrases: Functional data analysis, Gaussian approximation, Gaussian multiplier bootstrap, local linear smoothing, longitudinal data analysis, simultaneous confidence bands.

1. Introduction

Functional and longitudinal data arise in many disciplines and have received considerable attention in the statistics community over the last two decades. Notable work includes Ramsay and Silverman (2005), Ferraty and Vieu (2006), Wu and Zhang (2006), Ramsay and Silverman (2007), Horváth and Kokoszka (2012), and Hsing and Eubank (2015); see also Cuevas (2014) and Wang, Chiou and Müller (2016) for a comprehensive overview. As a typical setting, one considers independent and identically distributed (i.i.d.) random functions $X_1(t), \dots, X_n(t)$ defined on a compact interval $\mathcal{T}$, with mean function $\mu(t) = \mathbb{E}\{X_1(t)\}$ and covariance function $\gamma(t, s) = \text{Cov}\{X_1(t), X_1(s)\}$. In practice, the entire trajectory of $X_i(t)$ is not observable as data can only be collected discretely over time. Correspondingly, as a more realistic formulation, each process is supposed to be observed with additive measurement errors at $m_i \geq 1$ discrete random time points $t_{ij} \in \mathcal{T}$, $j = 1, \dots, m_i$. The actual observations then follow

$$Y_{ij} = X_i(t_{ij}) + \varepsilon_{ij} = \mu(t_{ij}) + \nu_i(t_{ij}) + \varepsilon_{ij}, \quad (1.1)$$

*Corresponding author. E-mail: ZHL318@pitt.edu