Smoothing Spline Estimation of Variance Functions

Anna Liu, Tiejun Tong, and Yuedong Wang

This article considers spline smoothing of variance functions. We focus on selection of the smoothing parameters and develop three direct data-driven methods: unbiased risk (UBR), generalized approximate cross-validation (GACV), and generalized maximum likelihood (GML). In addition to guaranteed convergence, simulations show that these direct methods perform better than existing indirect UBR, generalized cross-validation (GCV), and GML methods. The direct UBR and GML methods perform better than the GACV method. An application to array-based comparative genomic hybridization data illustrates the usefulness of the proposed methods.

Key Words: Array-based comparative genomic hybridization; Generalized approximate cross-validation; Generalized maximum likelihood; Heteroscedasticity; Smoothing parameter; Unbiased risk.

1. INTRODUCTION

Modeling local variability in terms of variance functions is an important problem with a wide range of applications. For example, variance function estimation is necessary in finance, quality control, and immunoassay for measuring volatility or risk (Andersen and Lund 1997; Gallant and Tauchen 1997), experimental design (Box 1988), prediction (Carroll 1987; Yao and Tong 1994) and calibration (Raab 1981; Watters, Carroll, and Spiegelman 1987). Variance estimation is especially important for detecting genes with differential expression across experimental conditions based on microarray data (Huang and Pan 2002; Wang and Guo 2004). With a small number of replicated array experiments, the standard estimates of variances are unreliable. Various methods have been proposed to improve estimation of the variances which usually lead to more powerful tests (Huang and Pan 2002; Wang and Guo 2004; Cui et al. 2005). We apply our variance function estimation

Anna Liu is Assistant Professor, Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01003 (E-mail: *anna@math.umass.edu*). Tiejun Tong is Postdoctoral Associate, Department of Epidemiology and Public Health, Yale University, New Haven, CT 06520 (E-mail: *tiejun.tong@yale.edu*). Yuedong Wang is Professor, Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106 (E-mail: *yuedong@pstat.ucsb.edu*). Address for correspondence: Yuedong Wang, Department of Statistics and Applied Probability, University of California, Santa Barbara, California 93106.