

EFFICIENT AFT MODEL ESTIMATION WITH AUXILIARY AGGREGATE INFORMATION

Huijuan Ma¹, Manli Cheng¹, Yukun Liu^{*1}, Donglin Zeng² and Yong Zhou¹

¹*East China Normal University and* ²*University of Michigan*

Abstract: With the rapidly increasing availability of aggregate data in the public domain, there has been a growing interest in synthesizing information from individual-level data and aggregate data. This article studies the maximum full likelihood estimation method to integrate the auxiliary information in the estimation of the accelerated failure time model. To overcome the computational challenges in maximizing full likelihood, we propose a novel one-step estimator, where the maximum conditional likelihood estimator without combining any auxiliary information is chosen as an initial estimator. We establish the consistency and asymptotic normality of the proposed one-step estimator and show that it is more efficient than the initial estimator. The asymptotic variance of the proposed one-step estimator has a closed form and is easily estimated by the plug-in rule. Simulation studies show that the proposed one-step estimator yields an efficiency gain over existing approaches. The proposed methodology is illustrated with an analysis of a chemotherapy study for Stage III colon cancer.

Key words and phrases: Accelerated failure time model, empirical likelihood, information synthesis.

1. Introduction

The accelerated failure time (AFT) model is an important and attractive alternative to the Cox proportional hazards model (Cox, 1972) in regression analysis of censored data. It directly relates the log failure time to covariates and has a straightforward physical interpretation. Sometimes it can provide a more accurate or more concise summarization of data than the Cox model (Zeng and Lin, 2007). There has been a vast literature on the semiparametric analysis of the accelerated failure time model. Well-known methods include rank-based approaches (Prentice, 1978; Tsiatis, 1990; Wei, Ying and Lin, 1990; Ying, 1993; Yang, 1997; Jin et al., 2003; Zhou, 2005), least squared approaches (Buckley and James, 1979; Ritov, 1990; Lai and Ying, 1991; Jin, Lin and Ying, 2006), and semiparametric efficient approaches (Zeng and Lin, 2007; Ding and Nan, 2011; Lin and Chen, 2013).

In medical research and precision medicine, comprehensive individual-level data are recognized as the best source of evidence to produce desirable estimates

*Corresponding author. E-mail: ykliu@sfs.ecnu.edu.cn