

U-statistics with conditional kernels for incomplete data models

Ao Yuan¹ · Mihai Giurcanu² · George Luta¹ ·
Ming T. Tan¹

Received: 26 March 2014 / Revised: 18 June 2015 / Published online: 23 August 2015
© The Institute of Statistical Mathematics, Tokyo 2015

Abstract For incomplete data models, the classical U-statistic estimator of a functional parameter of the underlying distribution cannot be computed directly since the data are not fully observed. To estimate such a functional parameter, we propose a U-statistic using a substitution estimator of the conditional kernel given the observed data. This kernel estimator is obtained by substituting the non-parametric maximum likelihood estimator for the underlying distribution function in the expression of the conditional kernel. We study the asymptotic properties of the proposed U-statistic for several incomplete data models, and in a simulation study, we assess the finite sample performance of the Mann–Whitney U-statistic with conditional kernel in the current status model. The analysis of a real-world data set illustrates the application of the proposed methods in practice.

Keywords U-statistics · Censored data · Incomplete data models · Non-parametric MLE

✉ Ao Yuan
ay312@georgetown.edu

Mihai Giurcanu
giurcanu@ufl.edu

George Luta
gl77@georgetown.edu

Ming T. Tan
mtt34@georgetown.edu

¹ Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University, 4000 Reservoir Road, NW, Washington, DC 20057, USA

² Department of Statistics, University of Florida, McCarty C, Gainesville, Florida 32611, USA