

WIN RATIO FOR PARTIALLY ORDERED DATA

Lu Mao

University of Wisconsin-Madison

Abstract: The win ratio, initially developed for time-to-event data, can be extended to any data type equipped with a partial order. We study this extension in both nonparametric inference and semiparametric regression. We begin by formulating the win ratio as an estimand of contrast between two populations with partially ordered responses, showing that it reduces to the familiar odds ratio in the case of binary data. For hypothesis testing, we prove that the empirical two-sample win ratio is consistent against stochastically ordered distributions and efficient against proportional odds alternatives under a total order. In regression, we model the conditional win ratio multiplicatively against covariates, extending logistic regression from binary to partially ordered responses. This model is implied by a generalized continuation-ratio logit model but requires fewer assumptions on the relationship between response levels. To make inference, we construct a class of weighted U -statistic estimating equations and derive pseudo-efficient weights to improve efficiency. Simulation studies demonstrate that the proposed procedures perform well in both testing and regression under finite samples. As illustrations, we analyze bivariate radiologic assessments in a recent liver disease study and subject smoking status in a youth tobacco use study, treating them both as partially ordered outcomes. The proposed methodology is implemented in the R package `poset`, publicly available on GitHub at <https://lmaowisc.github.io/poset> and on the Comprehensive R Archive Network (CRAN).

Key words and phrases: Continuation ratio, logistic regression, odds ratio, ordinal data, stochastic order, U -statistic.

1. Introduction

Partially ordered data, a common variant of ordinal data, frequently arise in medical and sociological studies. Unlike totally ordered tumor grades or Likert scales, partially ordered data are not necessarily pairwise comparable. For instance, in the tumor-node-metastasis cancer staging system (Edge et al., 2010), primary and metastatic tumors are scored on four- and three-level severity scales, respectively. Two patients can be compared in overall severity only if one scores higher on both scales (Lin, Wang and Chappell, 2018). Similar partial orders are found in groups of radiologic ratings or survey responses.

Despite their prevalence, partially ordered data have received only limited methodological attention. In the nonparametric setting, Rosenbaum (1991)

*Corresponding author. E-mail: lmao@biostat.wisc.edu