Statistica Sinica **36** (2026), 1-6
doi:https://doi.org/10.5705/ss.202025.0164

DISCUSSION ON "CAUSAL AND COUNTERFACTUAL VIEWS OF MISSING DATA MODELS"

Zeyi Wang* and Mark J. van der Laan

Oklahoma State University and University of California

1. Introduction

Viewing missingness as longitudinal intervention enabled closed-form efficient estimation for challenging bivariate censoring problems.

Opposite to the common view of causal inference as missing data problems, viewing missingness as (longitudinal) intervention has also been a fruitful strategy, leveraging identification and estimation techniques from causal inference literature. For example, in the challenging bivariate censoring problem, viewing the complex censoring mechanism as longitudinal interventions under sequential randomization assumptions (SRA; it can be shown that SRA $\subset$ MAR for multivariate right-censored data and SRA = MAR for univariate right-censored data) leads to closed-form efficient influence curves which would have not been possible with only the MAR assumptions (Rubin, 1976; van der Laan and Robins, 2003b). For example, if the full data is $X_j(t) = I(T_j \leq t), j = 1, 2$ defined by survival times T_1, T_2 that are subject to censoring times C_1, C_2, then only assuming MAR is significantly weaker than assuming $(C_1, C_2) \perp (X_1, X_2)$, resulting in an efficient influence curve that does not exist in closed form. Even though such SRA based estimators are inefficient if one truly only assumes MAR, they can flexibly incorporate time-dependent covariate information and provide meaningful efficiency improvement (van der Laan and Robins, 2003a). There also exist scenarios—for example, when C_1 and T_2 are dependent, such as when monitoring, defined as interval censoring, is part of the intervention (Carone, Petersen and van der Laan, 2012; van der Laan, 2018)—where SRA is plausible, whereas MAR is violated.

The missing data model is a special case of multivariate/bivariate censoring.

In the paper under discussion, the missing data model defined by full data variables $L_1^{(1)}, \ldots, L_K^{(1)}$, missingness indicators $R_1, \ldots, R_K$, and observed data variables $L_1, \ldots, L_K$ is a special case of multivariate censoring. For $k = 1, \ldots, K$, let $T_k \in \mathbb{R}$, $C_k \in \{-\infty, \infty\}$, and define $\tilde{T}_k = \min\{T_k, C_k\}$ with a non-censoring

*Corresponding author. E-mail: zwang107@gmail.com