Statistica Sinica **37** (2027), 1-21
doi:https://doi.org/10.5705/ss.202024.0013

ESTIMATION OF SUBSIDIARY PERFORMANCE METRICS UNDER OPTIMAL POLICIES

Zhaoqi Li*[1], Houssam Nassif[2] and Alex Luedtke*[1]

[1] *University of Washington and* [2] *Meta*

Abstract: In policy learning, the goal is typically to optimize a primary performance metric, but other subsidiary metrics often also warrant attention. This paper presents two strategies for evaluating these subsidiary metrics under a policy that is optimal for the primary one. The first relies on a novel margin condition that facilitates Wald-type inference. Under this and other regularity conditions, we show that the one-step corrected estimator is efficient. Despite the utility of this margin condition, it places strong restrictions on how the subsidiary metric behaves for nearly-optimal policies, which may not hold in practice. We therefore introduce alternative, two-stage strategies that do not require a margin condition. The first stage constructs a set of candidate policies and the second builds a uniform confidence interval over this set. We provide numerical simulations to evaluate the performance of these methods in different scenarios.

Key words and phrases: Efficient estimator, optimal policies, statistical inference.

1. Introduction

1.1. Literature review

Many fields are interested in learning policies that map from individual-level characteristics to a set of actions. The policies that result in the best possible mean of a subsequent outcome are often referred to as optimal policies (Athey and Wager, 2021). For example, in biomedical sciences the action may take the form of a treatment allocation and the outcome may be disease remission (Ling et al., 2023), whereas in digital marketing the action and outcome may be a recommendation and click-through rate, respectively (Fiez et al., 2024). Various methods have been developed to estimate optimal policies. Examples include Q-learning, which uses regression modeling of the expected outcome to guide policy decisions (Qian and Murphy, 2011), outcome-weighted learning (Zhao et al., 2012), and doubly robust methods (Murphy, 2003; Robins, 2004; Dudík, Langford and Li, 2011; Zhang et al., 2013). Performance guarantees for these methods have been established by several authors (Qian and Murphy, 2011; Zhao et al., 2012; Luedtke and Chambaz, 2020; Athey and Wager, 2021).

*Corresponding author. E-mail: zli9@uw.edu