Efficient Approximate Predictive Inference Under Feedback Covariate Shift with Influence Functions

Drew Prinster  
Suchi Saria  
Anqi Liu  

Department of Computer Science, Johns Hopkins University, Baltimore, MD, United States

Editor: Harris Papadopoulos, Khuong An Nguyen, Henrik Boström and Lars Carlsson

Abstract

We propose JAWA-FCS, which uses higher-order influence functions to approximate predictive intervals of the (previous) jackknife+ weighted for feedback covariate shift for further computational efficiency (no retraining). We evaluate our method on protein design tasks.

Keywords: Jackknife+, feedback covariate shift, influence functions, conformal prediction, biomolecular design.

1. Introduction

Predictive uncertainty intervals for feedback covariate shift Many decision-making scenarios require uncertainty quantification in the presence of feedback loops. A prominent example is biomolecular design: consider a protein engineer who fits a machine-learning regressor $\hat{\mu}$ on a limited dataset of protein sequences $\{X_1, ..., X_n\}$ and experimentally-labeled “fitness” values $\{Y_1, ..., Y_n\}$ (e.g., fluorescence or medicinal efficacy) to then propose a novel protein sequence $X_{\text{test}}$ with high predicted fitness value $\hat{\mu}(X_{\text{test}})$. The use of the trained $\hat{\mu}$ to propose the design sequence $X_{\text{test}}$ induces a feedback covariate shift (FCS) dependency between the training and test data (Fannjiang et al., 2022), which violates common i.i.d. and exchangeability assumptions. Still, given the demands of developing and experimentally measuring novel biomolecules, rigorous uncertainty quantification remains essential for understanding whether $\hat{\mu}(X_{\text{test}})$ is reliable relative to the true label $Y_{\text{test}}$. This work is motivated by approximating predictive uncertainty intervals under FCS when resource constraints of low data-availability and computational budget demand efficient approaches.

Conformal and jackknife+ predictive inference Conformal prediction (CP) (Vovk et al., 2005) is a framework for converting machine-learning predictions into predictive intervals (or prediction sets, more generally) with finite-sample distribution-free coverage (see Angelopoulos and Bates (2021) for a gentle introduction). Full (or transductive) CP is the CP variant with most efficient use of labeled data, which usually results in more precise and informative intervals, but full CP is mainly limited by its notorious computational demands of extensive retraining. On the other hand, split (or inductive) CP (Papadopoulos et al., 2002; Papadopoulos, 2008) is a computationally efficient alternative to full CP that avoids retraining, but split CP suffers from reduced data efficiency due to requiring a holdout set not used for training, which often degrades prediction accuracy. In between the computational-statistical tradeoff poles of full and split CP are cross CP (Vovk, 2015; Vovk et al., 2018) and jackknife+ methods (Barber et al., 2021). Recent works have extended
full and split CP to standard (Tibshirani et al., 2019) and feedback (Fannjiang et al., 2022) covariate shift, as well as the jackknife+ to both these settings (Prinster et al., 2022, 2023).

**Computationally efficient approximation with influence functions** Influence functions (IFs) (Cook, 1979) estimate how model parameters would change if a particular point were removed from training via a Taylor-series approximation, which itself avoids retraining at the main cost of computing the inverse Hessian. IFs have recently become popular in machine learning (Koh and Liang, 2017) and have been proposed for computationally efficient approximation of full CP (Abad et al., 2022) in classification. In regression, higher-order IFs have also been proposed to approximate the leave-one-out (LOO) parameters required by the jackknife+ (Alaa and Van Der Schaar, 2020) and by JAW, the JAckknife+ Weighted for standard covariate shift (Prinster et al., 2022). Giordano et al. (2019) give regularity conditions for consistency of higher-order IF LOO parameter estimation.

**Current work** In the feedback covariate shift setting, we use higher-order influence functions to further improve the computational efficiency of the JAW-FCS method of Prinster et al. (2023) with an approximation that avoids any retraining. Accordingly, we call the current method JAckknife+ Weighted Approximation for FCS (JAWA-FCS). In particular, our JAWA-FCS implementation uses the memory-efficient algorithm for higher-order IFs LOO parameter estimation from Giordano et al. (2019) to approximate the LOO model parameters required by JAW-FCS (Prinster et al., 2023). JAWA-FCS is distinct from the JAWA for standard covariate shift (Prinster et al., 2022) due to the distinct FCS likelihood-ratio weights that require LOO parameter estimation (see Prinster et al. (2023) for weights).

### 2. Experimental Results

As in Fannjiang et al. (2022) and Prinster et al. (2023) we conduct experiments on red and blue fluorescence protein datasets from Poelwijk et al. (2019). We use a small neural network regressor with tanh activation trained on 192 samples, and we approximate LOO parameters with 3rd-order IFs. JAWA-FCS computation time (with hyperparameter search for Hessian damping) was less than 3 minutes, versus about 1 hour 24 minutes for JAW-FCS. Figure 1 compares methods that avoid retraining: only JAWA-FCS and weighted split CP (Tibshirani et al., 2019) maintain coverage at the target level $(1 - \alpha = 0.9)$, though JAWA-FCS does so with smaller (more informative) predictive intervals and higher mean predicted fitness (arrows point to desired metrics). Future work could explore if these results scale to larger predictor models and how IF approximation error impacts coverage guarantees.

Figure 1: Coverage, width, and predicted fitness results for increasing shift $\lambda$. Code, details, and additional experiments at https://github.com/drewprinster/jaws-x
References


