Conformal Meta-learners for Predictive Inference of Individual Treatment Effects

SeongSik Choi

February 26, 2025

Seoul National University

Introduction

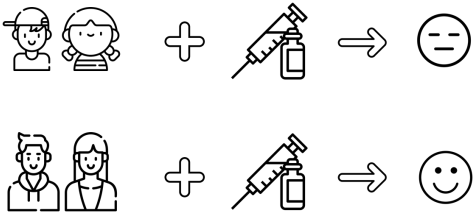
What is the goal?

Consider following example:

X : 나이, $T \in \{0, 1\}$: 백신 접종 여부, Y : 감염 취약도

$\rightarrow Y^t$: 백신 접종 여부 t일 때의 감염 취약도.

나이(X)에 따른 백신 접종(T)에 따른 감염 취약도(Y)의 감소 효과를 추정하자.



기존 causal inference 방법론을 통해서는, 특정 나이에서 백신 접종이 감염 취약도에 미치는 평균 인과 효과를 estimate할 수 있었다.

가령 $X = 30$세인 사람이 백신을 접종할 경우 감염 취약도는 0.02만큼 감소($\mathbb{E}(Y^1 - Y^0|X = 30) = -0.02$)할 것으로 기대될 수 있으나, uncertainty 정보는 알지 못함.

본 논문을 통해서는, 특정 나이에서 백신 접종이 감염 취약도에 미치는 개별 인과 효과에 대한 predictive interval을 구할 수 있다.

가령, $X = 30$세인 사람이 백신을 접종할 경우, 감염 취약도는 유의수준 α에서 $0.01 \sim 0.03$ 정도가 감소할 것으로 기대될 수 있음. ($\mathbb{P}(Y^1 - Y^0 \in (-0.03, -0.01)|X = 30) \geqslant 1 - \alpha$) (Warning : Technically untrue.)

What is the goal of individual treatment effects?

Individual treatment effect(ITE) : $Y_i^1 - Y_i^0$

Goal of existing methods : Estimate the conditional expectation of treatment effect τ such that

$$\tau(x) = \mathbb{E}[Y^1 - Y^0 | X = x]$$

Goal of proposed method : Construct a predictive interval of ITE $\widehat{C}_\alpha$ such that

$$\mathbb{P}[Y_{n+1}^1 - Y_{n+1}^0 \in \widehat{C}_\alpha(X_{n+1})] \geqslant 1 - \alpha \tag{1}$$

for a predetermined target coverage of $1 - \alpha$, with $\alpha \in (0, 1)$, where the probability in (1) accounts for the randomness of the $\widehat{C}_\alpha$ and the test point $(X_{n+1}, Y_{n+1}^1 - Y_{n+1}^0)$.

Conformal prediction of ITE(Oracle)

How to construct a conformal interval on oracle data?

Spoiler : we can construct an oracle predictive interval as following form:

$$\widehat{C}^*_\alpha(x) = \left[\hat{\tau}(x) - Q^*_{\mathcal{V},1-\alpha}, \hat{\tau}(x) + Q^*_{\mathcal{V},1-\alpha}\right]$$

우리가 상상속의 oracle data $\mathcal{D}^* = \{(X_i, Y_i^1 - Y_i^0)\}_{i=1}^n$ 를 얻었다고 가정하자.

$\mathcal{D}^*$ 을 training set $\mathcal{D}^*_t$ 와 calibration set $\mathcal{D}^*_c$ 로 나누자.

$\mathcal{D}^*_t$ 를 이용해서는 ML model $\hat{\tau}$ 을 학습하고,

$\mathcal{D}^*_c$ 를 이용해서는 interval width $Q^*_{\mathcal{V},1-\alpha}$ 를 결정할 것이다.

이를 위한 procedure는 다음 페이지와 같다.

Conformal prediction of ITE (Oracle)

How to construct a conformal interval on oracle data?

1. conformity score $V(\cdot)$ is evaluated for all samples in $\mathcal{D}_c^*$ as follows:

$$V_i^*(\hat{\tau}) = V(X_i, Y_i^1 - Y_i^0; \hat{\tau}), ^\forall i \in \mathcal{D}_c^* \qquad (2)$$

   A common choice of $V(\cdot)$ is absolute residual $V(x, y; \tau) = |y - \tau(x)|$.

2. We then compute a quantile of the empirical distribution of conformity scores

$$Q^*_{\mathcal{V},1-\alpha} := (1-\alpha)\left(1 + 1/\left|\mathcal{D}_c^*\right|\right)\text{-th quantile of } \mathcal{V}^*(\hat{\tau}) \qquad (3)$$

   where $\mathcal{V}^*(\hat{\tau}) = \{V_i^*(\hat{\tau}) : i \in \mathcal{D}_c^*\}$.

3. We can construct the oracle predictive interval at a new point $X_{n+1} = x$ as follows:

$$\widehat{C}_\alpha^*(x) = \left[\hat{\tau}(x) - Q^*_{\mathcal{V},1-\alpha}, \hat{\tau}(x) + Q^*_{\mathcal{V},1-\alpha}\right] \qquad (4)$$

Conformal prediction of ITE (Oracle)

Remarks

- Since the oracle problem is a standard regression, the oracle procedure is marginally valid - i.e., it satisfies the guarantee in (1),

$$\mathbb{P}[Y_{n+1}^1 - Y_{n+1}^0 \in \hat{C}_\alpha^*(X_{n+1})] \geqslant 1 - \alpha$$

- However, oracle CP is infeasible since we cannot observe $Y_i^1 - Y_i^0$, hence we need an alternative procedure that operates on the observed data $\mathcal{D} = \{Z_i = (X_i, T_i, Y_i)\}_{i=1}^n$.

Conformal prediction of ITE (pseudo-intervals)

How to construct a conformal interval on real data?

IDEA : We can calculate pseudo-outcomes $\tilde{Y}_\varphi$ and use it instead of $Y_i^1 - Y_i^0$ to construct a conformal interval.

For nuisance parameter $\varphi = (\pi, \mu_0, \mu_1)$ and observable data $Z = (X, T, Y)$, we can construct a proper pseudo-outcome $\tilde{Y}_\varphi = f(Z; \varphi)$ by choosing proper f such that

$$\mathbb{E}[\tilde{Y}_\varphi | X = x] = \mathbb{E}[Y^1 - Y^0 | X = x] = \tau(x) \tag{5}$$

where $\pi(x) = \mathbb{P}(T = 1 | X = x), \mu_t(x) = \mathbb{E}(Y^t | X = x), ^\forall t \in \{0, 1\}$

Options for f will be presented in the next page.

Using $\tilde{Y}_{\hat{\varphi},i} = f(Z_i, \hat{\varphi})$ instead of $Y_i^1 - Y_i^0$, we can construct $\widehat{C}_{\hat{\varphi},\alpha}$ in a way similar to the oracle procedure of constructing $\widehat{C}^*_\alpha$.

Conformal prediction of ITE (pseudo-intervals)

How to construct a conformal interval on real data?(in detail)

Example of pseudo-outcome :

By choosing f in the table below, the condition in (5) can be satisfied.

IPW-learner [27] : $\widetilde{Y}_\varphi = f(Z;\varphi) = \frac{T-\pi(X)}{\pi(X)(1-\pi(X))}Y$

X-learner [5] : $\widetilde{Y}_\varphi = f(Z;\varphi) = T\left(Y-\mu_0(X)\right) + (1-T)\left(\mu_1(X)-Y\right)$

DR-learner [20] : $\widetilde{Y}_\varphi = f(Z;\varphi) = \frac{T-\pi(X)}{\pi(X)(1-\pi(X))}\left(Y-\mu_T(X)\right) + \mu_1(X) - \mu_0(X)$

Consider $\widehat{\varphi} = (\pi, \widehat{\mu}_0, \widehat{\mu}_1)$ which is an estimate of φ and use $\widetilde{Y}_{\widehat{\varphi},i}$ instead of $Y_i^1 - Y_i^0$. (They assume π is known).

Conformal prediction of ITE (pseudo-intervals)

How to construct a conformal interval on real data?(in detail)

Given a dataset $\mathcal{D} = \{Z_i = (X_i, T_i, Y_i)\}_i$, we create three mutually-exclusive subsets: $\mathcal{D}_\varphi, \mathcal{D}_t$ and $\mathcal{D}_c$.

$\mathcal{D}_\varphi$: used to estimate the nuisance parameters φ.(obtain $\hat{\varphi}$)

$\mathcal{D}_t$: used to train a CATE model τ. (obtain $\hat{\tau}$)

$\mathcal{D}_c$: used to compute conformity scores for $\hat{\tau}$. (obtain $V_{\varphi,i} \Rightarrow \widehat{C}_\varphi(x)$)

Conformal prediction of ITE (pseudo-intervals)

How to construct a conformal interval on real data?(in detail)

Next, the estimates $\widehat{\varphi} = (\pi, \widehat{\mu}_0, \widehat{\mu}_1)$ are used to transform $\{Z_i = (X_i, T_i, Y_i) : i \in \mathcal{D}_t\}$ into covariate/pseudo-outcome pairs $\left\{\left(X_i, \widetilde{Y}_{\varphi,i}\right) : i \in \mathcal{D}_t\right\}$ which are used to train a CATE model $\widehat{\tau}$.

Finally, we compute conformity scores for $\widehat{\tau}$ on pseudo-outcomes, i.e.,

$$V_{\varphi,i}(\widehat{\tau}) := V\left(X_i, \widetilde{Y}_{\varphi,i}; \widehat{\tau}\right), \forall i \in \mathcal{D}_c,$$

$$Q_{\mathcal{V}_\varphi,1-\alpha} := (1-\alpha)\left(1 + 1/\left|\mathcal{D}_c\right|\right)\text{-th quantile of } \mathcal{V}(\widehat{\tau})$$

For a target coverage of $1-\alpha$, we construct a predictive interval at a new point $X_{n+1} = x$ as follows:

$$\widehat{C}_\varphi(x) = \left[\widehat{\tau}(x) - Q_{\mathcal{V}_\varphi,1-\alpha}, \widehat{\tau}(x) + Q_{\mathcal{V}_\varphi,1-\alpha}\right]$$

where $\mathcal{V}_\varphi = \{V_{\varphi,k}(\widehat{\tau}) : k \in \mathcal{D}_c\}$. We call $\widehat{C}_\varphi(x)$ a pseudo-interval.

Does psuedo-interval satisfy (1)?

In the following Theorem, we provide sufficient conditions for the validity of conformal meta-learners in terms of the stochastic orders of their conformity scores.

Theorem 1. If $(X_i, T_i, Y_i(0), Y_i(1))$, $i = 1, \ldots, n+1$ are exchangeable, then $\exists \alpha^* \in (0,1)$ such that the pseudo-interval $\widehat{C}_\varphi(X_{n+1})$ constructed using the dataset $\mathcal{D} = \{(X_i, T_i, Y_i)\}_{i=1}^n$ satisfies

$$\mathbb{P}\left(Y_{n+1}(1) - Y_{n+1}(0) \in \widehat{C}_\varphi(X_{n+1})\right) \geqslant 1 - \alpha, \forall \alpha \in \left(0, \alpha^*\right)$$

if at least one of the following **stochastic ordering conditions** hold:

Validity of Psuedo-interval

Does psuedo-interval satisfy (1)?

Stochastic ordering conditions

Let F_{V_φ} and F_{V*} be the corresponding cumulative distribution functions (CDFs) of variable V_φ and V^* respectively.

1. V_φ has first-order stochastic dominance (FOSD) on V^* ($V_\varphi \underset{(1)}{\succeq} V^*$) $\Leftrightarrow F_{V_\varphi}(x) \leqslant F_{V*}(x), \forall x$, with strict inequality for some x.
2. V_φ has second-order stochastic dominance (SOSD) over V^* ($V_\varphi \underset{(2)}{\preceq} V^*$) $\Leftrightarrow \int_{-\infty}^{x} [F_{V_\varphi}(t) - F_{V*}(t)]dt \geqslant 0, \forall x$, with strict inequality for some x.
3. V_φ has monotone convex dominance (MCX) over V^* ($V_\varphi \underset{mcx}{\succeq} V^*$) $\Leftrightarrow \mathbb{E}_{X\sim F_{V_\varphi}}[u(X)] \geqslant \mathbb{E}_{X\sim F_{V*}}[u(X)]$ for all non-decreasing convex functions $u : \mathbb{R} \to \mathbb{R}$.

(Under condition i, we have $\alpha^* = 1$)

Validity of Psuedo-interval

Do V_φ and V^* satisfy the stochastic ordering conditions?

Theorem 2. Let $V_\varphi(\hat{\tau}) = \left|\hat{\tau}(X) - \widetilde{Y}_\varphi\right|$ and assume that the propensity score function $\pi : \mathcal{X} \to [0, 1]$ is known.

Then, the following holds:

1. For the X-learner, V_φ and V^* do not admit to a model- and distribution-free stochastic order.
2. The IPW- and the DR-learners satisfy $V_\varphi \underset{mcx}{\succeq} V^*$. (for any distribution $P(X, T, Y(0), Y(1))$, CATE estimate $\hat{\tau}$, and nuisance estimate $\hat{\varphi}$!)