

Conformal Meta-learners for Predictive Inference of Individual Treatment Effects

Ahmed M. Alaa
UC Berkeley and UCSF
amalaa@berkeley.edu

Zaid Ahmad
UC Berkeley
zaidahmad@berkeley.edu

Mark van der Laan
UC Berkeley
laan@berkeley.edu



Berkeley
UNIVERSITY OF CALIFORNIA



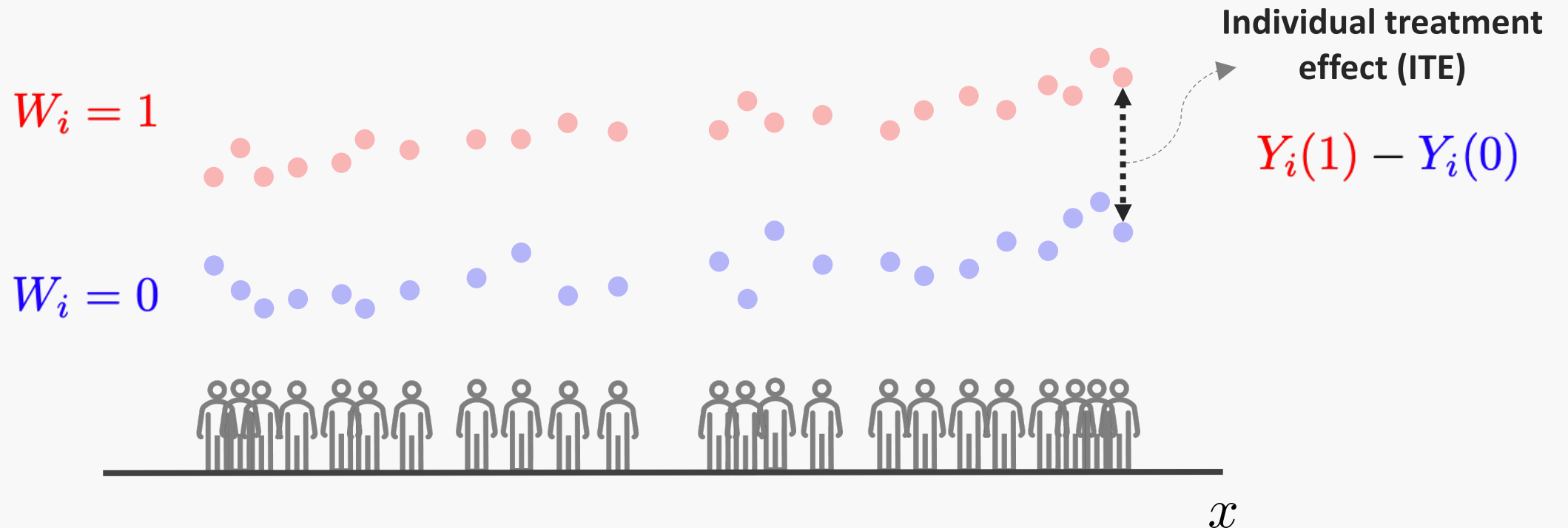
UCSF

University of California
San Francisco

Setup: Potential Outcomes Framework

(Neyman 1923; D. Rubin 1974)

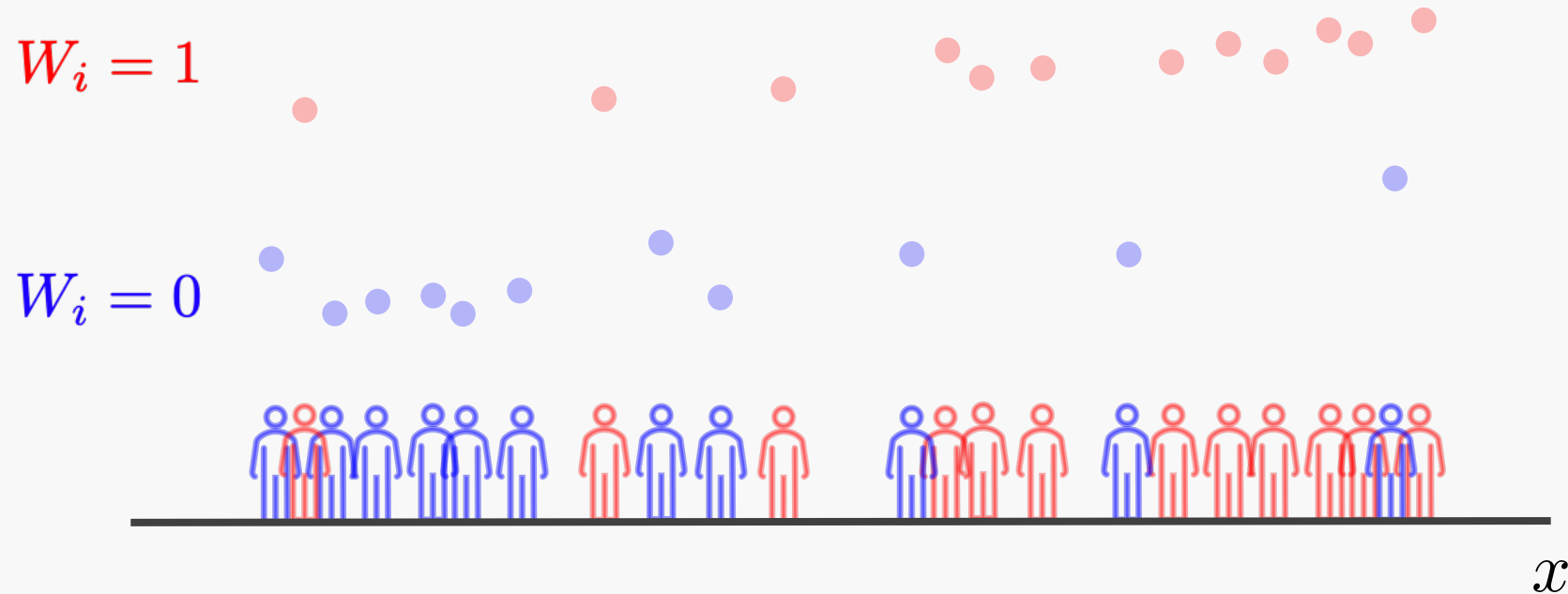
- Binary treatment $W_i \in \{0, 1\} \rightarrow$ two potential outcomes: $Y_i(1)$ and $Y_i(0)$



Setup: Potential Outcomes Framework

- The fundamental problem of causal inference: Counterfactuals are not observed!

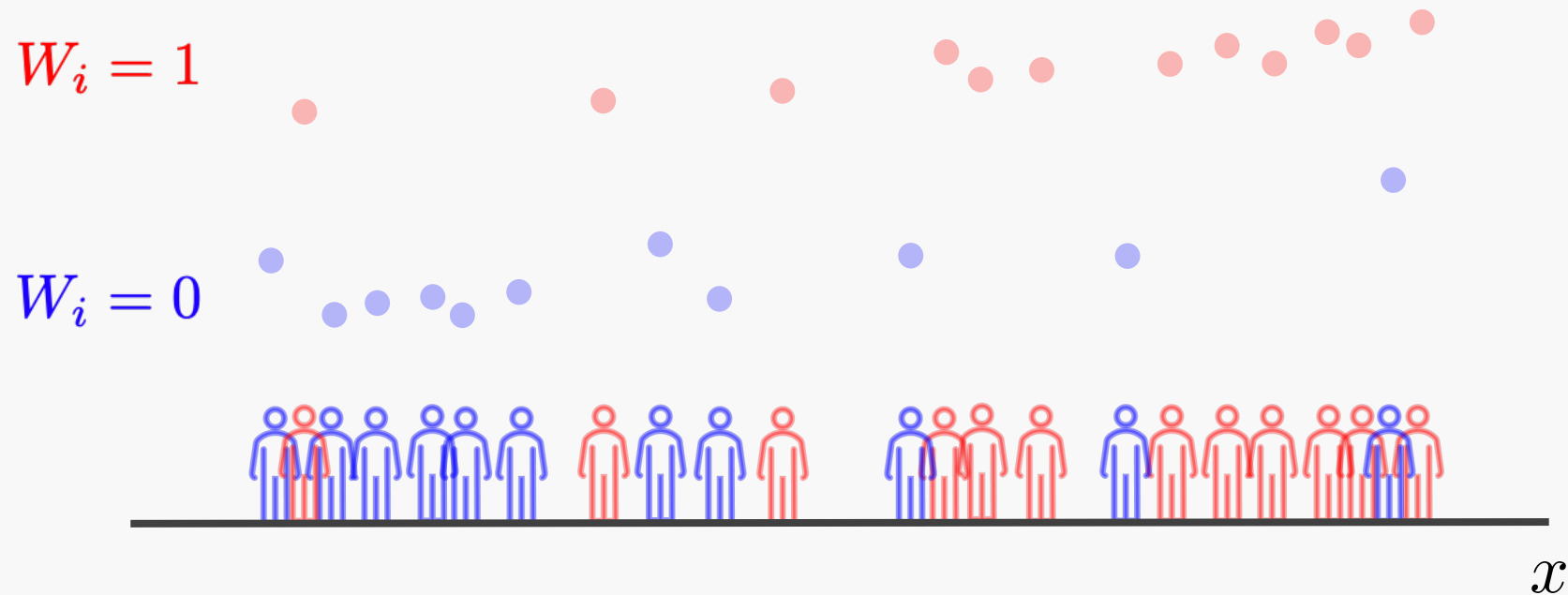
$$Y_i = W_i Y_i(1) + (1 - W_i) Y_i(0)$$



Setup: Potential Outcomes Framework

- **Treatments not randomly assigned:** Propensity score $\rightarrow \pi(x) = P(W = 1|X = x)$

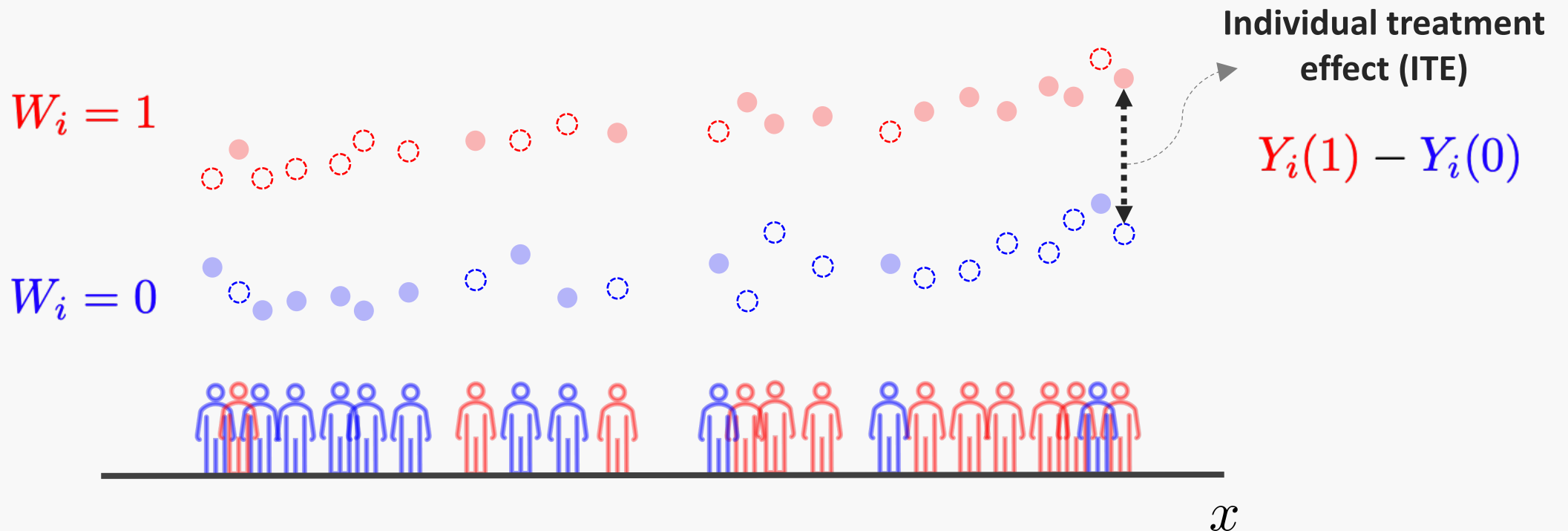
$$Y_i = W_i Y_i(1) + (1 - W_i) Y_i(0)$$



Problem: Valid Predictive Inference on ITEs

- **Predictive Inference:** Construct predictive intervals $\hat{C}(X_{n+1})$ that cover ITEs

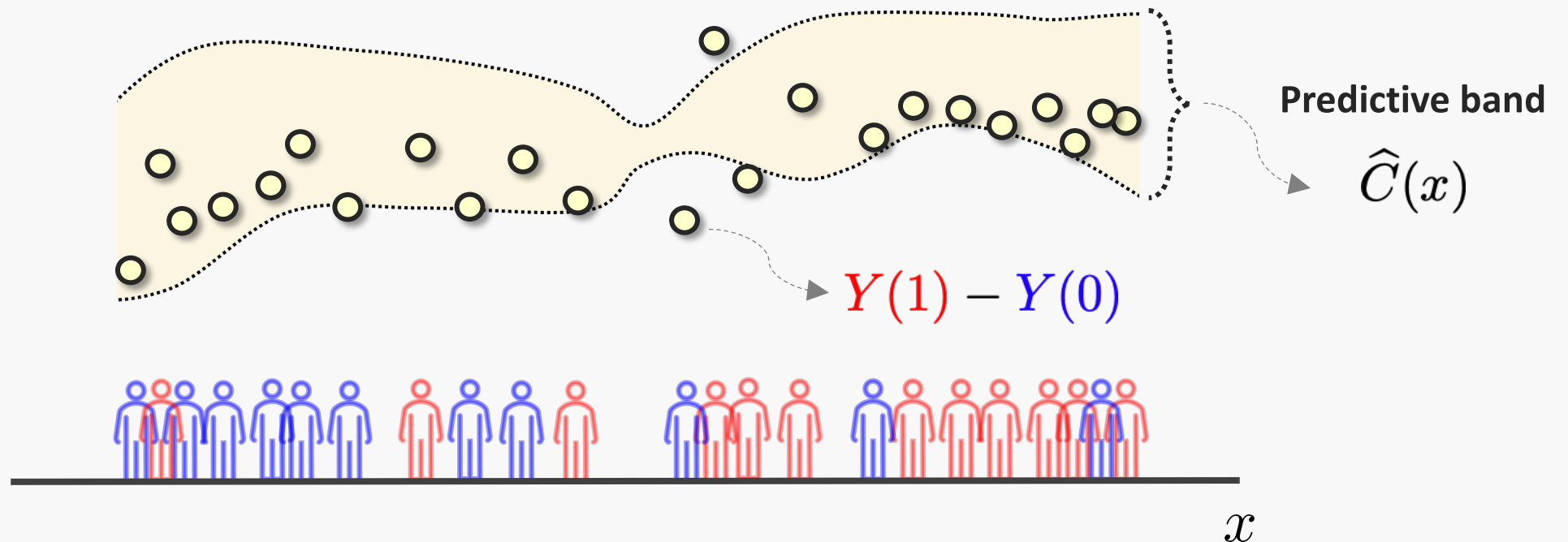
$$P(Y_{n+1}(1) - Y_{n+1}(0) \in \hat{C}(X_{n+1})) \geq 1 - \alpha$$



Problem: Valid Predictive Inference on ITEs

- **Predictive Inference:** Construct predictive intervals $\hat{C}(X_{n+1})$ that cover ITEs

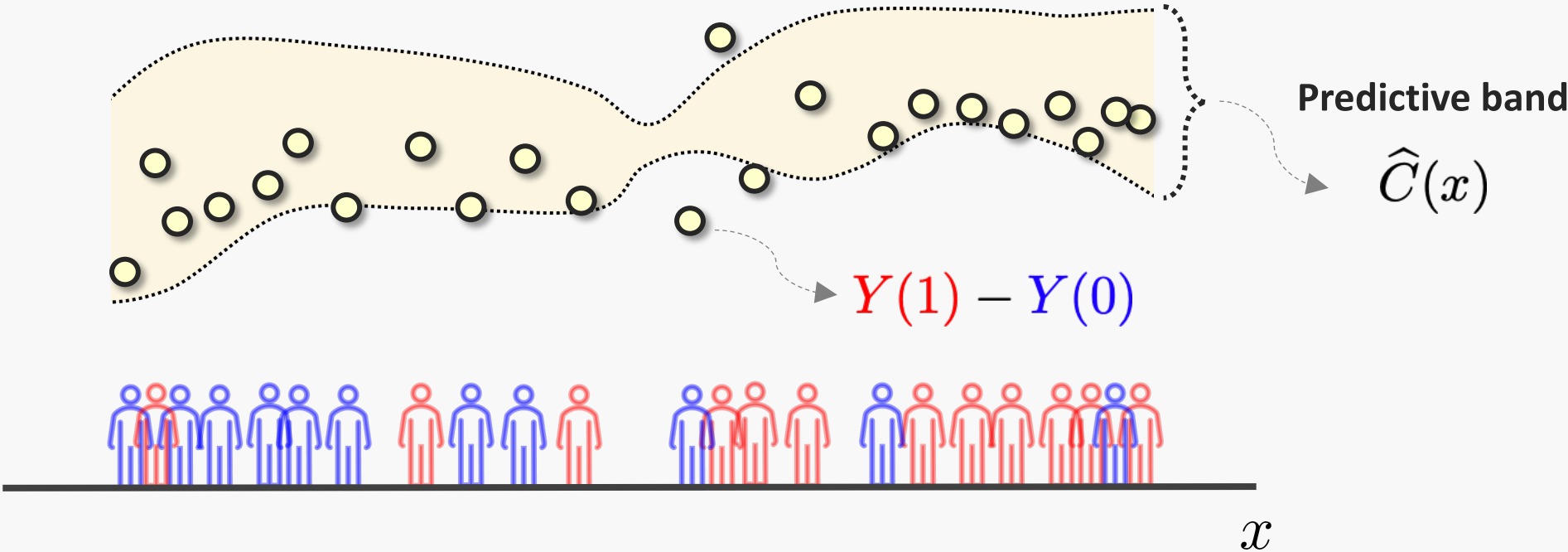
$$P(Y_{n+1}(1) - Y_{n+1}(0) \in \hat{C}(X_{n+1})) \geq 1 - \alpha$$



Proposed Method: Conformal Meta-Learners

Concept 1:
Pseudo-outcome Regression

Concept 2:
Conformal Prediction



Concept 1: Pseudo-outcome Regression (Meta-learners)

(van der Laan 2006; E. Kennedy 2020 and others)

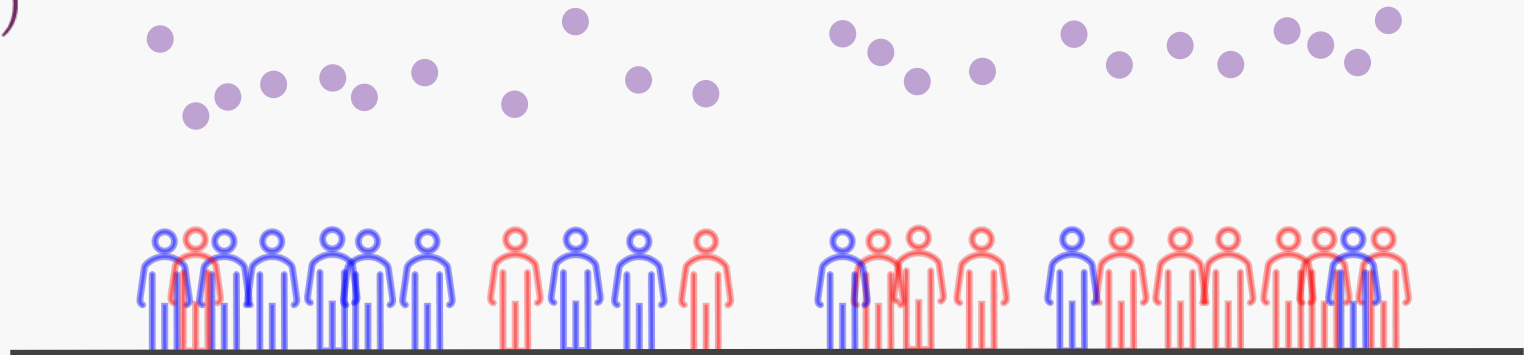
- **Pseudo-outcomes:** Transformations of (X, W, Y) that preserve conditional effects

$$E[\phi(X, W, Y) | X = x] = E[Y(1) - Y(0) | X = x]$$

Inverse Propensity
Weighting

$$\phi_{\text{IPW}} = \frac{W - \pi(X)}{\pi(X)(1 - \pi(X))} Y$$

$\phi(X, W, Y)$



x

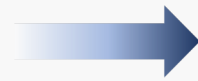
Concept 1: Pseudo-outcome Regression (Meta-learners)

(van der Laan 2006; E. Kennedy 2020 and others)

- **Pseudo-outcomes:** Transformations of (X, W, Y) that preserve conditional effects

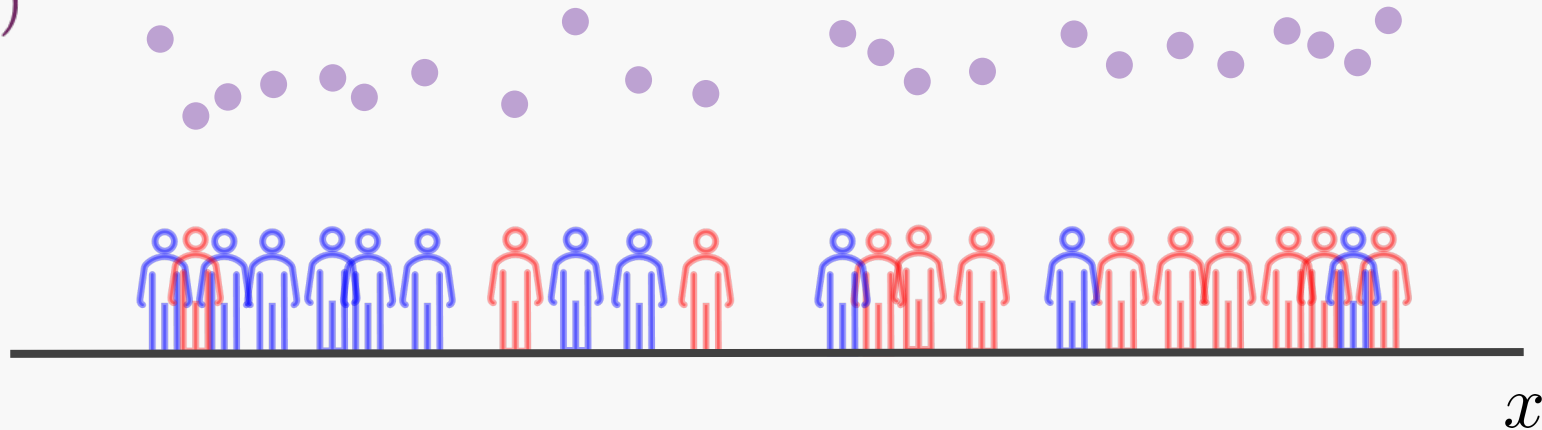
$$E[\phi(X, W, Y) | X = x] = E[Y(1) - Y(0) | X = x]$$

$$\tau = E[Y(1) - Y(0)]$$



$$\hat{\tau} = \frac{1}{n} \sum_i \phi(X_i, W_i, Y_i)$$

$\phi(X, W, Y)$

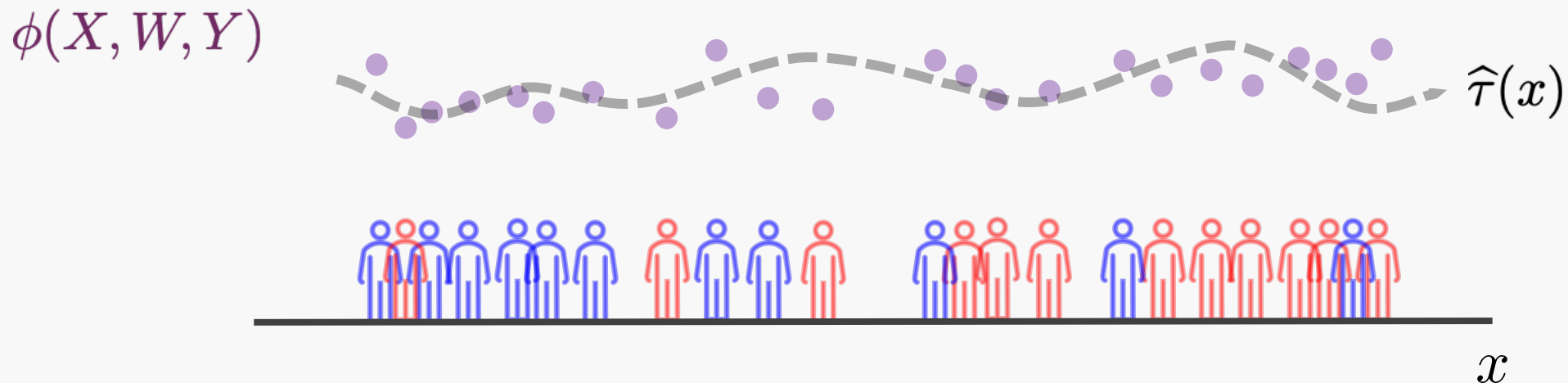


Concept 1: Pseudo-outcome Regression (Meta-learners)

- **Pseudo-outcomes:** Transformations of (X, W, Y) that preserve conditional effects

$$E[\phi(X, W, Y) | X = x] = E[Y(1) - Y(0) | X = x]$$

$$\tau(x) = E[Y(1) - Y(0) | X = x] \longrightarrow \hat{\tau}(x) : \text{Regress } \phi(X, W, Y) \text{ on } X$$



Concept 2: Conformal Prediction

- **Conformal Prediction:** A general approach for **post-hoc** predictive inference (V. Vovk 2012)

Finite-sample validity

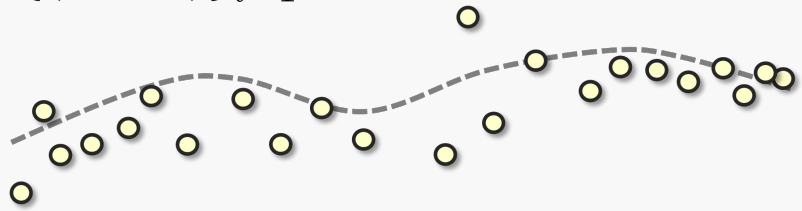
Model-free

Distribution-free

Any fitted ML model
(point predictions)

Conformal prediction

$\{(X_i, Y_i)\}_{i=1}^n$



x

Concept 2: Conformal Prediction

- **Conformal Prediction:** A general approach for **post-hoc** predictive inference (V. Vovk 2012)

Finite-sample validity

Model-free

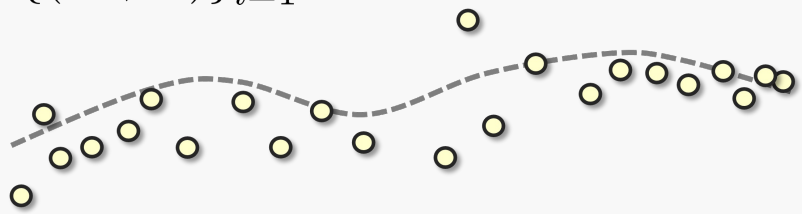
Distribution-free

Any fitted ML model
(point predictions)

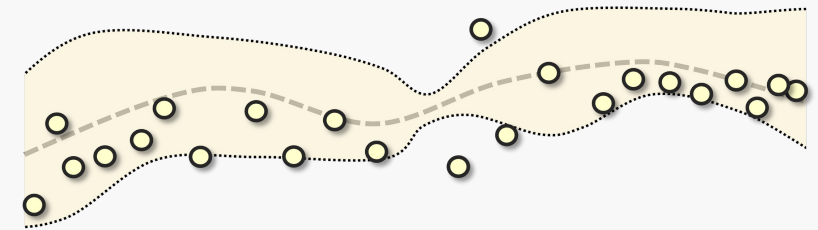
Conformal prediction

Valid uncertainty
intervals

$$\{(X_i, Y_i)\}_{i=1}^n$$



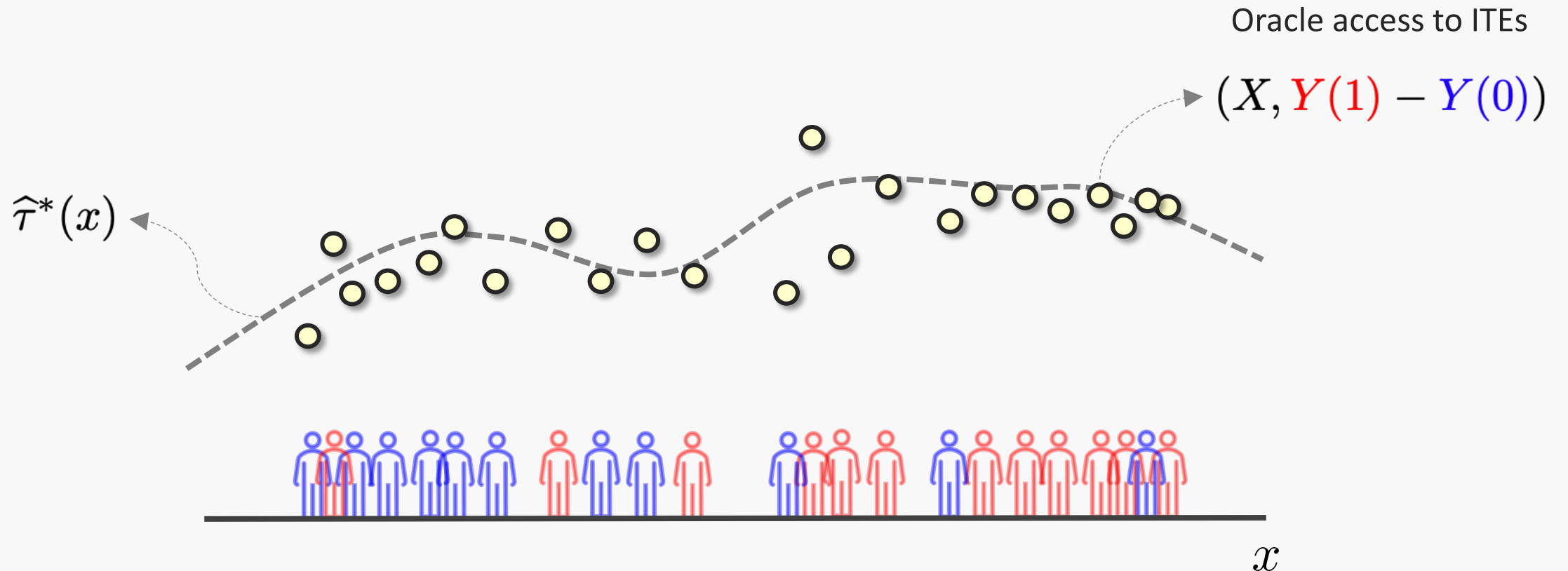
x



x

Concept 2: Conformal Prediction

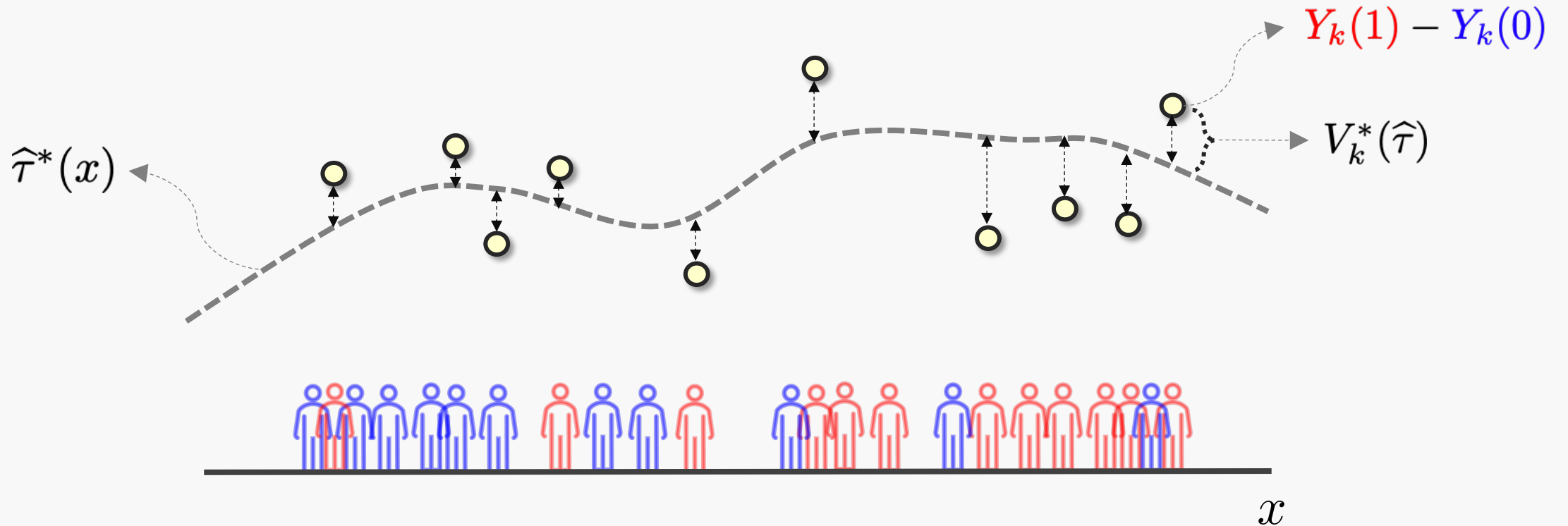
- **Step 1:** Train a machine learning model $\hat{\tau}^*(x)$ using $\{(X_i, Y_i(1) - Y_i(0))\}_i$



Concept 2: Conformal Prediction

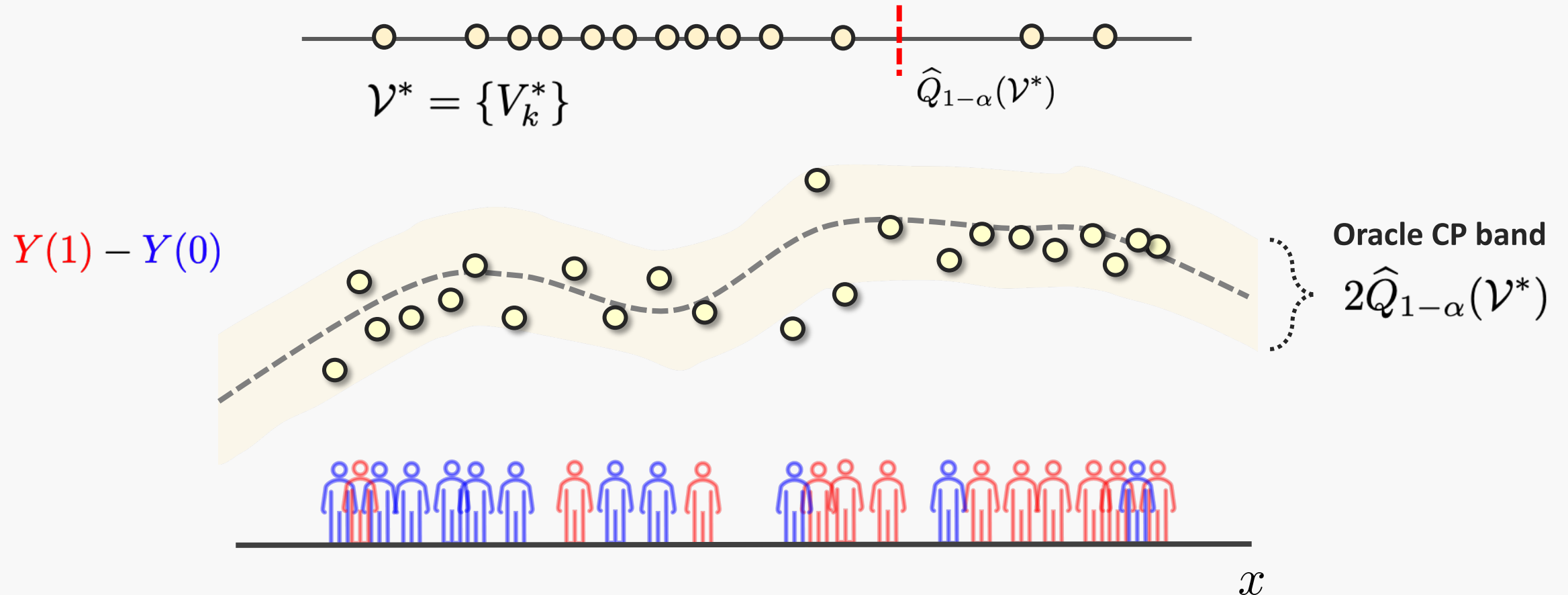
- **Step 2:** Evaluate **conformity scores** on a held-out calibration set

$$V_k^*(\hat{\tau}) = V(\hat{\tau}(X_k), Y_k(1) - Y_k(0))$$



Concept 2: Conformal Prediction

- **Step 3:** Construct a predictive interval using the empirical quantile of conformity scores

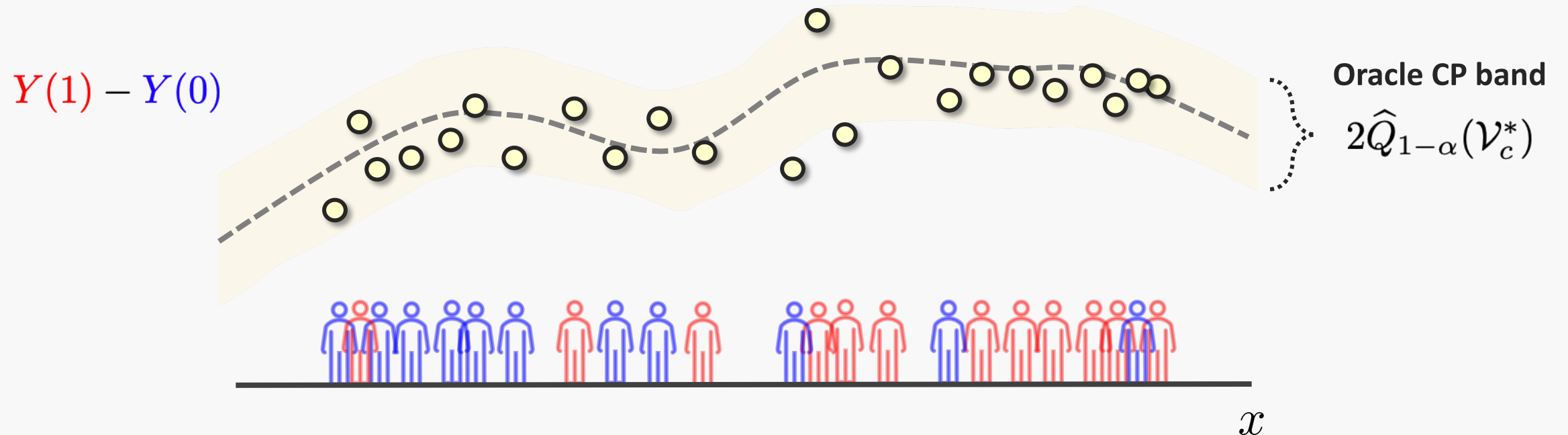


Concept 2: Conformal Prediction

- **Step 3:** Construct a predictive interval using the empirical quantile of conformity scores

$$P(Y_{n+1}(1) - Y_{n+1}(0) \in \hat{C}^*(X_{n+1})) \geq 1 - \alpha$$

If calibration and test data are exchangeable



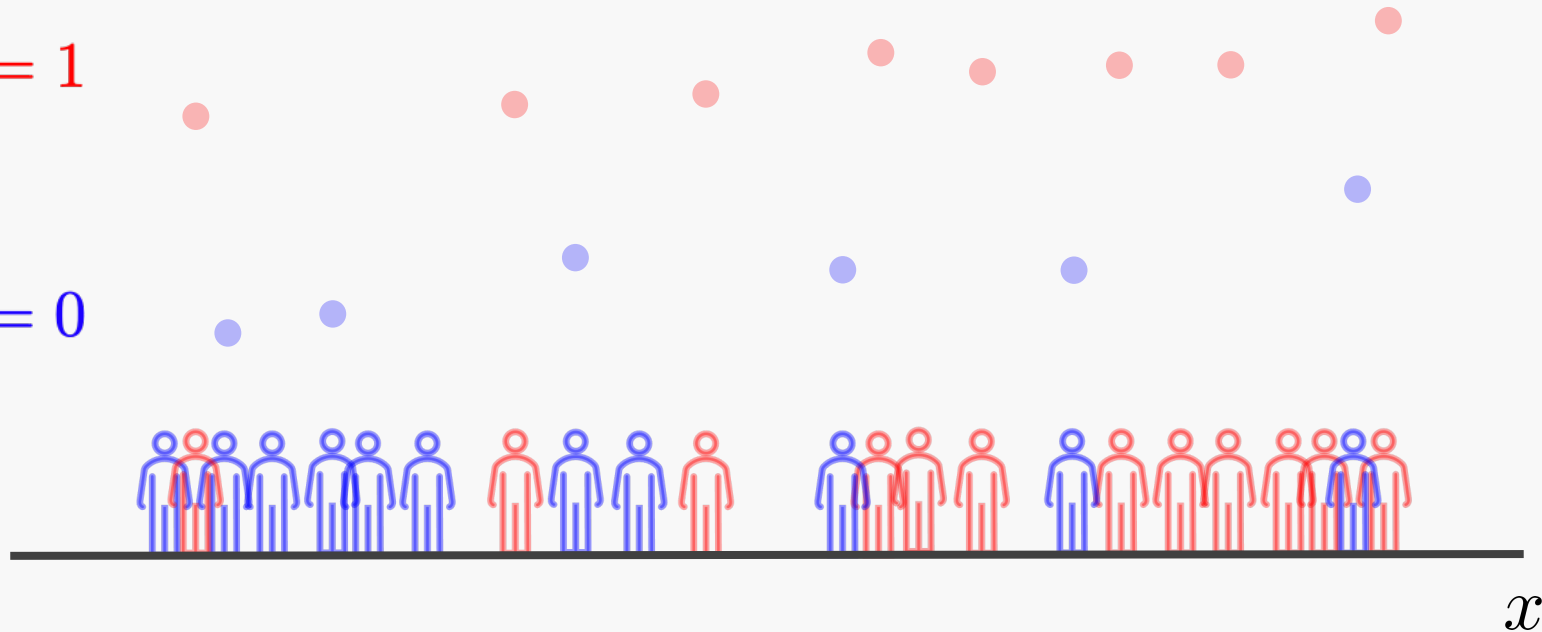
Method: Conformal Meta-learners

Concept 1:
Pseudo-outcome Regression

Concept 2:
Conformal Prediction

$$W_i = 1$$

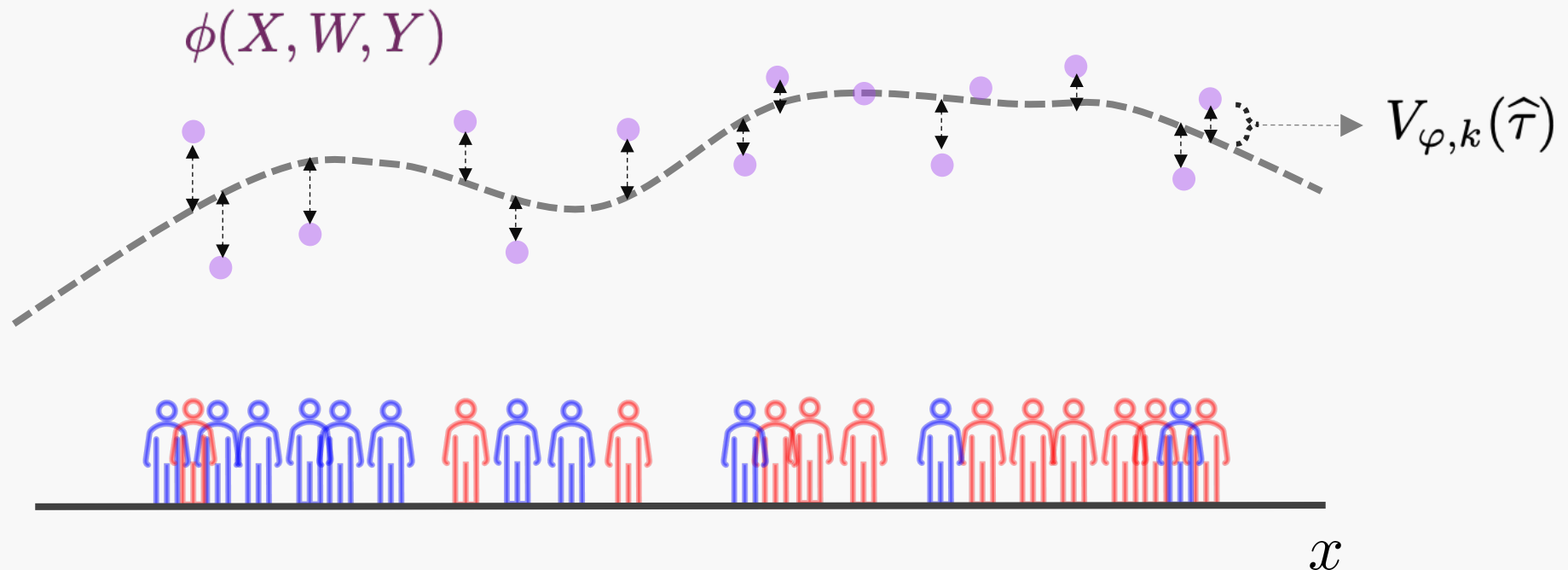
$$W_i = 0$$



Method: Conformal Meta-learners

- **Key Idea:** Apply CP to pseudo-outcomes instead of unobserved ITEs!

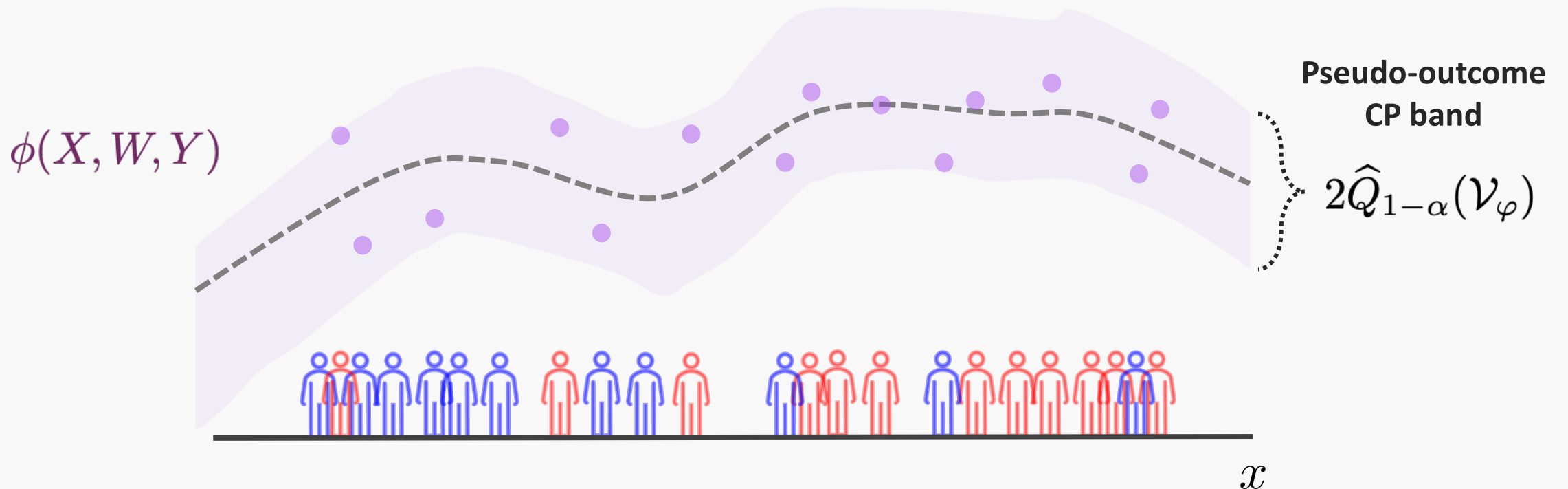
$$V_{\varphi,k}(\hat{\tau}) = V(\hat{\tau}(X_k), \phi(X_k, W_k, Y_k))$$



Method: Conformal Meta-learners

- **Key Idea:** Apply CP to pseudo-outcomes instead of unobserved ITEs.

$$P(\phi(X_{n+1}, W_{n+1}, Y_{n+1}) \in \hat{C}_\varphi(X_{n+1})) \geq 1 - \alpha$$



Method: Validity of Meta-learners via Stochastic Ordering Theory

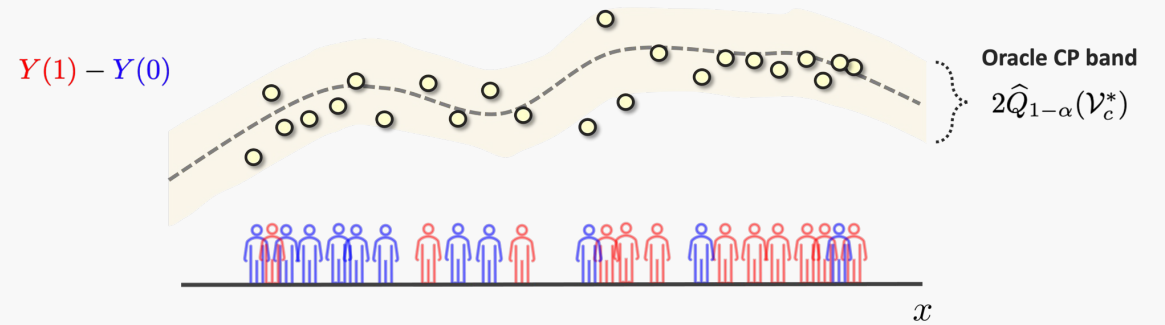
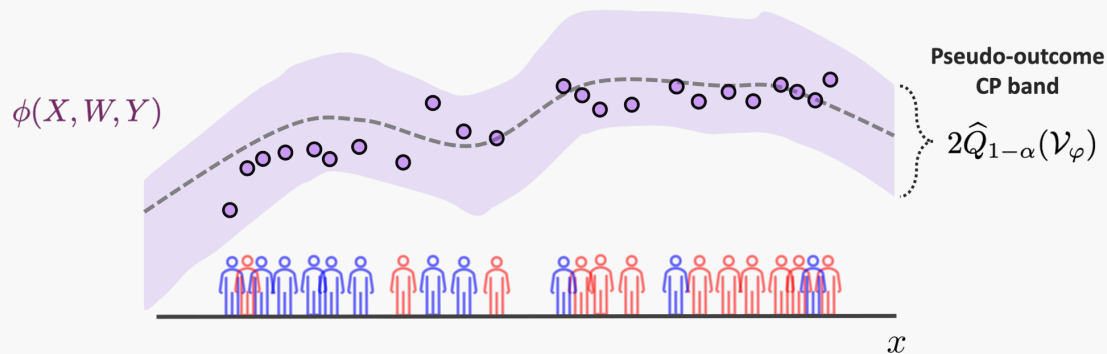
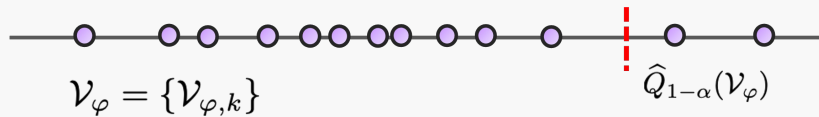
- Under what conditions are predictive intervals for pseudo-outcomes valid for ITEs?

Conformity scores evaluated on pseudo-outcome

Oracle Conformity scores evaluated on true ITEs

$$V_{\varphi,k}(\hat{\tau}) = V(\hat{\tau}(X_k), \phi(X_k, W_k, Y_k))$$

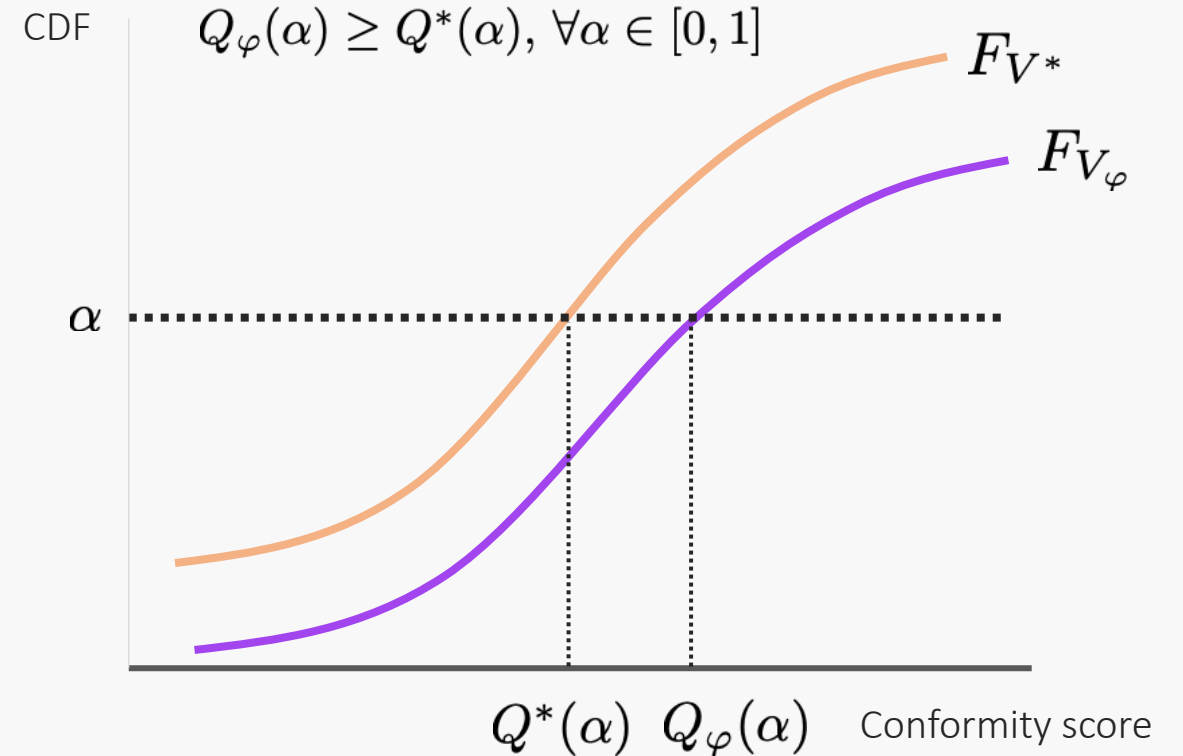
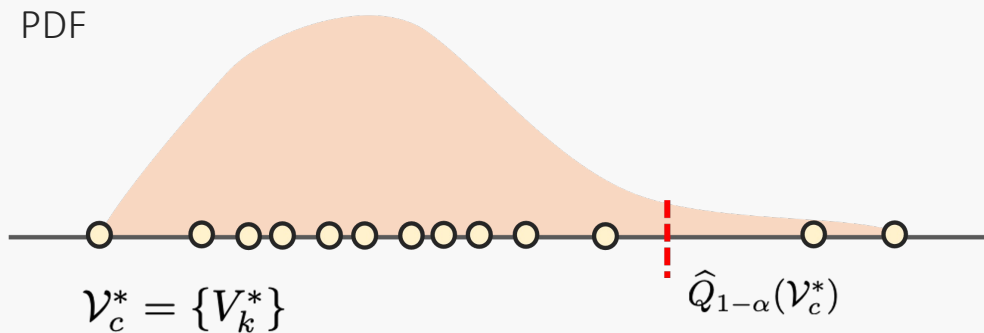
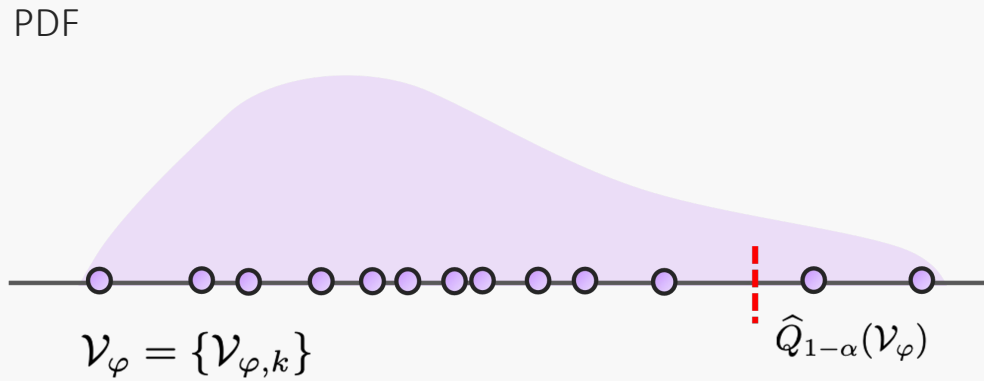
$$V_k^*(\hat{\tau}) = V(\hat{\tau}(X_k), Y_k(1) - Y_k(0))$$



Method: Validity of Meta-learners via Stochastic Ordering Theory

- **Sufficient condition for validity:** First-order stochastic dominance!

$$V_\varphi(\hat{\tau}) \succeq V^*(\hat{\tau})$$



Method: Validity of Meta-learners via Stochastic Ordering Theory

- Unified analysis of validity of meta-learners = stochastic orders of $V_\varphi(\hat{\tau})$ and $V^*(\hat{\tau})$

| Meta-learner | Pseudo-outcome |
|--|---|
| X-learner | $\phi_X = W(Y - \hat{\mu}_0(X)) + (1 - W)(\hat{\mu}_1(X) - Y)$ |
| IPW-learner Inverse propensity weighted | $\phi_{\text{IPW}} = \frac{W - \pi(X)}{\pi(X)(1 - \pi(X))} Y$ |
| DR-learner Doubly-robust learner | $\phi_{\text{DR}} = \frac{W - \pi(X)}{\pi(X)(1 - \pi(X))} (Y - \hat{\mu}_W(X)) + (\hat{\mu}_1(X) - \hat{\mu}_0(X))$ |

Method: Validity of Meta-learners via Stochastic Ordering Theory

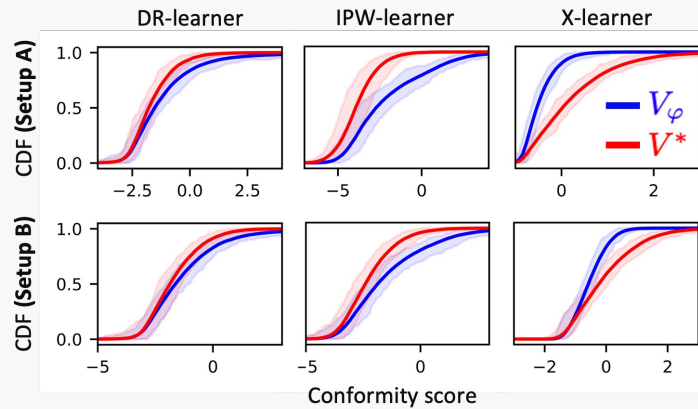
- Commonly-used meta-learners guarantee model-/distribution-free stochastic orders!

| Meta-learner | Stochastic orders of Conformity Scores |
|--|--|
| X-learner | No distribution-free stochastic order! |
| IPW-learner Inverse propensity weighted | $V^* \succeq_{(2)} V_{\text{IPW}}$ |
| DR-learner Doubly-robust learner | $V^* \succeq_{(2)} V_{\text{DR}}$ |

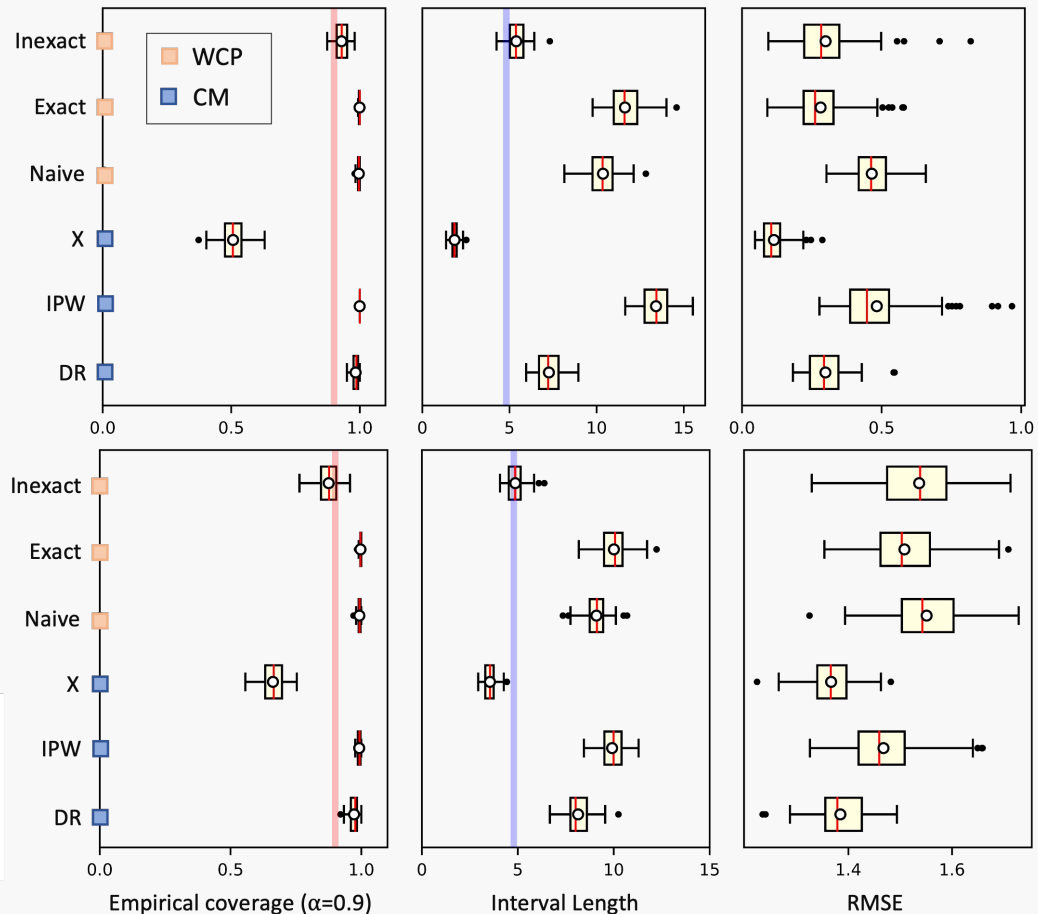
Results and Takeaway

- **TL;DR: Conformal meta-learners = valid predictive inference + accurate point predictions**

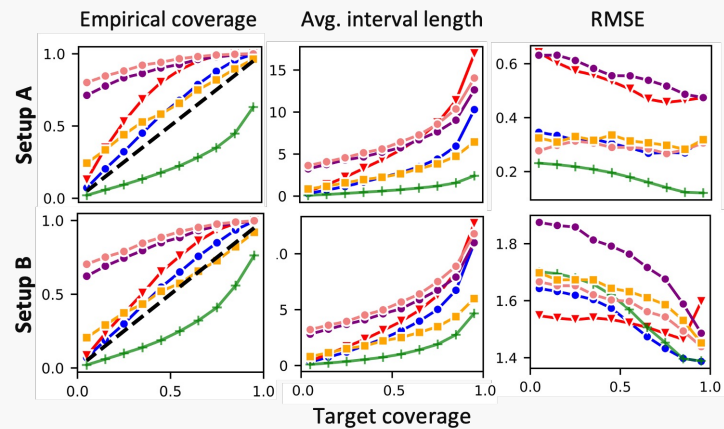
(a) Empirical assessment of stochastic orders



(b) Coverage, efficiency and RMSE for **Setup A** (top) and **Setup B** (bottom)



(c) Performance at different levels of target coverage



Poster Session 2

Tuesday Dec. 12

5:15 pm – 7:15 pm

Conformal Meta-Learners for Predictive Inference of Individual Treatment Effects

Ahmed M. Alaa, Zaid Ahmad and Mark van der Laan

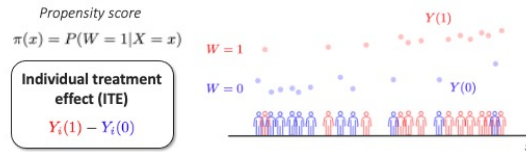


Berkeley
UNIVERSITY OF CALIFORNIA



Problem

- **Potential outcomes (PO) framework:** Each subject with feature X_i has two POs $\rightarrow Y_i(0)$: outcome w/o treatment and $Y_i(1)$: outcome w/ treatment.
- **Observational data:** $\{(X_i, W_i, Y_i)\}_i$, $W_i \in \{0, 1\}$ is a treatment indicator, $Y_i = W_i Y_i(1) + (1 - W_i) Y_i(0)$ is the observed (factual) outcome.



- **Our goal:** Valid predictive inference of ITEs...

$$P(Y_{n+1}(1) - Y_{n+1}(0) \in \hat{C}(X_{n+1})) \geq 1 - \alpha$$

Challenge: We never observe the ITEs in our data!

Related Work: Two Key Ideas

Pseudo-outcome Regression

(van der Laan 2006; E. Kennedy 2020 and others)

- ITEs are not observed \rightarrow Replace w/ proximal **pseudo-outcome** φ

$$E[\varphi(X, W, Y)|X] = E[Y(1) - Y(0)|X]$$

Example: Inverse propensity weighting

- Average treatment effects (ATE):

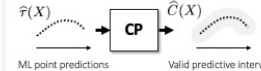
$$\hat{\tau} = \frac{1}{n} \sum_i \varphi(X_i, W_i, Y_i)$$

- **"Meta-learners"** for Conditional average treatment effects (CATE): Train an ML model on $\{(X_i, \varphi_i)\}_i$

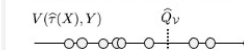
Conformal Prediction

(V. Vovk 2012)

- General purpose method for post-hoc ML-based predictive inference



- How are intervals constructed? Empirical quantiles of **conformity scores** on a held-out calibration set

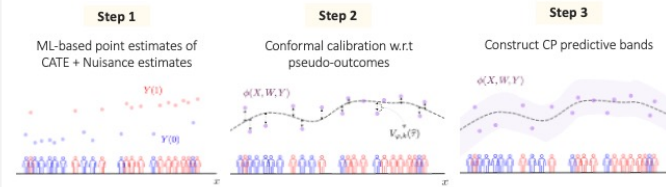


Proposed Method: Conformal Meta-Learners

- **Validity of conformal in regression** \rightarrow Marginal coverage $P(Y_{n+1} \in \hat{C}(X_{n+1})) \geq 1 - \alpha$

Finite-sample Model-free Distribution-free Assumption: Exchangeability of calibration and test data!

- **Key Idea:** Apply conformal prediction w/ pseudo-outcomes to construct intervals for ITEs



- **Example:** Doubly-Robust pseudo-outcomes $\varphi_{DR} = \frac{W - \pi(X)}{\pi(X)(1 - \pi(X))} (Y - \hat{\mu}_W(X)) + (\hat{\beta}_1(X) - \hat{\beta}_0(X))$

Stochastic Ordering Theory of Valid Inference

- **Conformal prediction w/ pseudo-outcomes** $\rightarrow P(\varphi(X_{n+1}, W_{n+1}, Y_{n+1}) \in \hat{C}_\varphi(X_{n+1})) \geq 1 - \alpha$
- Under what conditions will these intervals be valid for ITEs?

Pseudo-outcome conformity scores

$$V_{\varphi,k}(\hat{\tau}) = V(\hat{\tau}(X_k), \varphi_k(X_k, W_k, Y_k))$$

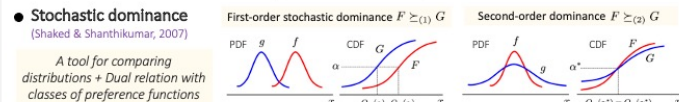


"Oracle" conformity scores

$$V_k^*(\hat{\tau}) = V(\hat{\tau}(X_k), Y_k(1) - Y_k(0))$$



Stochastic order



- **Stochastic dominance** (Shaked & Shanthikumar, 2007)

A tool for comparing distributions + Dual relation with classes of preference functions

Main Results

- **Theorem 1:** Sufficient condition for validity of meta-learners

$$V(\hat{\tau}(X), \varphi(X, W, Y)) \succeq_{(1)} V(\hat{\tau}(X), Y(1) - Y(0))$$

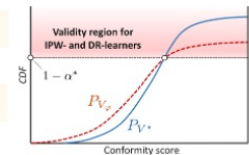
- **Theorem 2:** Stochastic orders of common meta-learners

X-learner (Künzel et al, 2019)

No stochastic order!

IPW- & DR-learner (Kennedy, 2020)

$$V^*(\hat{\tau}) \succeq_{(2)} V_\varphi(\hat{\tau})$$

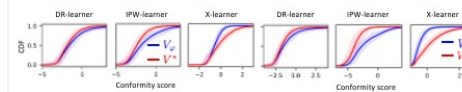


- IPW- & DR-learners: valid for high-probability target coverage
- Stochastic orders are model- and distribution-free!

Experiments

- **Pros:** Accurate point prediction + calibrated predictive intervals!
- **Cons:** Must know $\pi(x)$ + CDF cutoff α^* is unknown.

Experiments on synthetic data Experiments on real data (IHDP)



| IHDP | | | |
|---------|-------------|-------------|-------------|
| | Coverage | Avg. len. | RMSE |
| Naive | 0.89 (0.02) | 18.9 (4.04) | 4.73 (1.00) |
| Exact | 0.99 (0.00) | 29.8 (7.60) | 4.30 (0.97) |
| Inexact | 0.63 (0.04) | 8.49 (1.36) | 4.61 (0.99) |
| X | 0.65 (0.04) | 11.0 (3.04) | 3.34 (0.56) |
| IPW | 0.99 (0.00) | 112 (23.0) | 19.9 (3.44) |
| DR | 0.96 (0.01) | 16.7 (3.30) | 3.32 (0.53) |