Inference on Local Variable Importance Measures for Heterogeneous Treatment Effects

Paweł Morzywołek
University of Copenhagen

Joint work with Alex Luedtke and Peter Gilbert

MLinPL Conference — Warsaw, October 17$^{\text{th}}$, 2025

UNIVERSITY OF COPENHAGEN

Introduction

- In high-risk domains, e.g. medicine, **decision makers may hesitate to rely on the black-box decision support systems** without understanding the **rationale behind the recommendations**.

- **Goal:** Developing an inferential framework to assess **variable importance** for **heterogeneous treatment effects**.

- The **variable importance measures** we consider are **local** in that they may differ across individuals, while the **inference** is **global** in that it tests whether a given variable is important for any individual.

Introduction

- In high-risk domains, e.g. medicine, **decision makers may hesitate to rely on the black-box decision support systems** without understanding the **rationale behind the recommendations**.

- **Goal:** Developing an inferential framework to assess **variable importance** for **heterogeneous treatment effects**.

- The **variable importance measures** we consider are **local** in that they may differ across individuals, while the **inference** is **global** in that it tests whether a given variable is important for any individual.

Introduction

- In high-risk domains, e.g. medicine, **decision makers may hesitate to rely on the black-box decision support systems** without understanding the **rationale behind the recommendations**.

- **Goal:** Developing an inferential framework to assess **variable importance** for **heterogeneous treatment effects**.

- The **variable importance measures** we consider are **local** in that they may differ across individuals, while the **inference** is **global** in that it tests whether a given variable is important for any individual.

Introduction

- In high-risk domains, e.g. medicine, **decision makers may hesitate to rely on the black-box decision support systems** without understanding the **rationale behind the recommendations**.

- **Goal:** Developing an inferential framework to assess **variable importance** for **heterogeneous treatment effects**.

- The **variable importance measures** we consider are **local** in that they may differ across individuals, while the **inference** is **global** in that it tests whether a given variable is important for any individual.

Problem Setting

- Let $X := (X_1, \ldots, X_d)$ **be sufficient for confounding adjustment** of a **binary treatment** A on **outcome** Y.

- The **conditional average treatment effect (CATE)** is

$$\mathbb{E}(Y^1 - Y^0 | X_j = x_j : j \in S) \tag{1}$$

  where Y^a is the **potential outcome** under treatment $A = a$, for $a = 0, 1$, and $S \subseteq [d] := \{1, \ldots, d\}$ is a subset of indices.

- Under **causal assumptions** it can be identified as

$$\nu(P, S)(x) := \mu_1(P, S)(x) - \mu_0(P, S)(x) \tag{2}$$

  where $\mu_a(P, S)(x) := \mathbb{E}_P\{\mathbb{E}_P(Y | A = a, X = x) | X_j = x_j : j \in S\}$.

Problem Setting

- Let $X := (X_1, \ldots, X_d)$ **be sufficient for confounding adjustment** of a **binary treatment** A on **outcome** Y.
- The **conditional average treatment effect (CATE)** is

$$\mathbb{E}(Y^1 - Y^0 | X_j = x_j : j \in S) \tag{1}$$

where Y^a is the **potential outcome** under treatment $A = a$, for $a = 0, 1$, and $S \subseteq [d] := \{1, \ldots, d\}$ is a subset of indices.

- Under **causal assumptions** it can be identified as

$$\nu(P, S)(x) := \mu_1(P, S)(x) - \mu_0(P, S)(x) \tag{2}$$

where $\mu_a(P, S)(x) := \mathbb{E}_P\{\mathbb{E}_P(Y|A = a, X = x)|X_j = x_j : j \in S\}$.

Problem Setting

- Let $X := (X_1, \ldots, X_d)$ **be sufficient for confounding adjustment** of a **binary treatment** A on **outcome** Y.

- The **conditional average treatment effect (CATE)** is

$$\mathbb{E}(Y^1 - Y^0 | X_j = x_j : j \in S) \tag{1}$$

  where Y^a is the **potential outcome** under treatment $A = a$, for $a = 0, 1$, and $S \subseteq [d] := \{1, \ldots, d\}$ is a subset of indices.

- Under **causal assumptions** it can be identified as

$$\nu(P, S)(x) := \mu_1(P, S)(x) - \mu_0(P, S)(x) \tag{2}$$

  where $\mu_a(P, S)(x) := \mathbb{E}_P\{\mathbb{E}_P(Y | A = a, X = x) | X_j = x_j : j \in S\}$.

Problem Setting

- Let $X := (X_1, \ldots, X_d)$ **be sufficient for confounding adjustment** of a **binary treatment** A on **outcome** Y.
- The **conditional average treatment effect (CATE)** is

$$\mathbb{E}(Y^1 - Y^0 | X_j = x_j : j \in S) \tag{1}$$

where Y^a is the **potential outcome** under treatment $A = a$, for $a = 0, 1$, and $S \subseteq [d] := \{1, \ldots, d\}$ is a subset of indices.

- Under **causal assumptions** it can be identified as

$$\nu(P, S)(x) := \mu_1(P, S)(x) - \mu_0(P, S)(x) \tag{2}$$

where $\mu_a(P, S)(x) := \mathbb{E}_P\{\mathbb{E}_P(Y|A = a, X = x)|X_j = x_j : j \in S\}$.

How to quantify the variable importance for CATE?

Leave-Out-Covariates (LOCO) variable importance measure

Leave-out-covariates (LOCO) variable importance measure for CATE:

$$\gamma_i(P, S)(x) := \nu(P, S \cup \{i\})(x_{S \cup \{i\}}) - \nu(P, S)(x_S). \tag{3}$$

- Let $S = [d] \backslash \{i\}$, then $\gamma_i(P, S)$ is **leave-one-out (LOO)** variable importance measure.
- Let $S = \emptyset$, then $\gamma_i(P, S)$ is **keep-one-in (KOI)** variable importance measure.

Leave-Out-Covariates (LOCO) variable importance measure

Leave-out-covariates (LOCO) variable importance measure for CATE:

$$\gamma_i(P, S)(x) := \nu(P, S \cup \{i\})(x_{S \cup \{i\}}) - \nu(P, S)(x_S). \tag{3}$$

- Let $S = [d] \backslash \{i\}$, then $\gamma_i(P, S)$ is **leave-one-out (LOO)** variable importance measure.
- Let $S = \emptyset$, then $\gamma_i(P, S)$ is **keep-one-in (KOI)** variable importance measure.

Leave-Out-Covariates (LOCO) variable importance measure

Leave-out-covariates (LOCO) variable importance measure for CATE:

$$\gamma_i(P, S)(x) := \nu(P, S \cup \{i\})(x_{S \cup \{i\}}) - \nu(P, S)(x_S). \tag{3}$$

- Let $S = [d] \backslash \{i\}$, then $\gamma_i(P, S)$ is **leave-one-out (LOO)** variable importance measure.
- Let $S = \emptyset$, then $\gamma_i(P, S)$ is **keep-one-in (KOI)** variable importance measure.

Wald-type test of null hypothesis of zero-importance

Test the **null hypothesis of zero-importance of i^th feature**

$$H_0 \colon \gamma_i(P_0, S) = 0 \quad P\text{-a.s.}$$

against the alternative

$$H_A \colon \text{Not } H_0.$$

- **Proposal:**

  Consider test statistic $\|\gamma_i(P, S)\|^2_{L^2(P)}$

- **Problem:**

  Under H_0 the EIF of $\|\gamma_i(P, S)\|^2_{L^2(P)}$ is 0

  $\Downarrow$

  Higher-order parameter expansion needed to develop valid tests.

- **Solution:**

  Reproducing Kernel Hilbert Space (RKHS) embedding

Wald-type test of null hypothesis of zero-importance

Test the **null hypothesis of zero-importance of i$^{\text{th}}$ feature**

$$H_0 \colon \gamma_i(P_0, S) = 0 \quad P\text{-a.s.}$$

against the alternative

$$H_A \colon \text{Not } H_0.$$

- **Proposal:**

  Consider test statistic $\|\gamma_i(P, S)\|^2_{L^2(P)}$

- **Problem:**

  Under H_0 the EIF of $\|\gamma_i(P, S)\|^2_{L^2(P)}$ is 0

  $\Downarrow$

  Higher-order parameter expansion needed to develop valid tests.

- **Solution:**

  Reproducing Kernel Hilbert Space (RKHS) embedding

Wald-type test of null hypothesis of zero-importance

Test the **null hypothesis of zero-importance of i$^{\text{th}}$ feature**

$$H_0 \colon \gamma_i(P_0, S) = 0 \quad P\text{-a.s.}$$

against the alternative

$$H_A \colon \text{Not } \; H_0.$$

- **Proposal:**

  Consider test statistic $\|\gamma_i(P,S)\|^2_{L^2(P)}$

- **Problem:**

  Under H_0 the EIF of $\|\gamma_i(P,S)\|^2_{L^2(P)}$ is 0

  $\Downarrow$

  Higher-order parameter expansion needed to develop valid tests.

- **Solution:**

  Reproducing Kernel Hilbert Space (RKHS) embedding

Wald-type test of null hypothesis of zero-importance

Test the **null hypothesis of zero-importance of i$^{\text{th}}$ feature**

$$H_0 : \gamma_i(P_0, S) = 0 \quad P\text{-a.s.}$$

against the alternative

$$H_A : \text{Not } \; H_0.$$

- **Proposal:**

  Consider test statistic $\|\gamma_i(P, S)\|^2_{L^2(P)}$

- **Problem:**

  Under H_0 the EIF of $\|\gamma_i(P, S)\|^2_{L^2(P)}$ is 0

  $\Downarrow$

  Higher-order parameter expansion needed to develop valid tests.

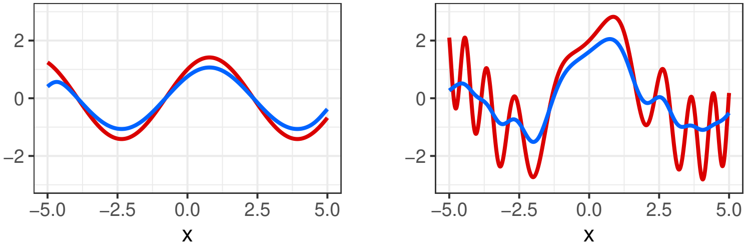
- **Solution:**

  Reproducing Kernel Hilbert Space (RKHS) embedding

RKHS embedding

For a symmetric, continuous positive semi-definite kernel function $\mathcal{K}\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, we can define

$$f(\cdot) \mapsto \int \mathcal{K}(\cdot, x')f(x')P_X(dx') \tag{4}$$

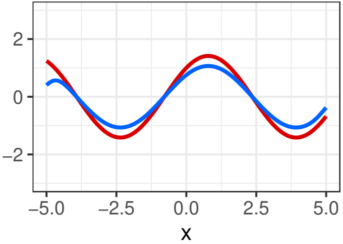


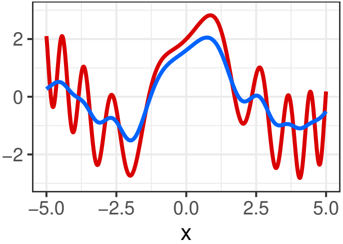
RKHS embedding is injective for an appropriate choice of the kernel.

RKHS embedding

For a symmetric, continuous positive semi-definite kernel function $\mathcal{K}\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, we can define

$$f(\cdot) \mapsto \int \mathcal{K}(\cdot, x')f(x')P_X(dx') \tag{4}$$





RKHS embedding is injective for an appropriate choice of the kernel.

RKHS embedding of the LOCO variable importance measure

RKHS embedding of the LOCO variable importance measure:

$$\gamma_i^{\mathcal{K}}(P, S)(x) = \int_{\mathcal{X}} \mathcal{K}(x, x')\gamma_i(P, S)(x')P_X(dx'). \tag{5}$$

RKHS embedding of **CATE** is pathwise differentiable and has EIF

$\Downarrow$

RKHS embedding of the **LOCO variable importance measure**
is pathwise differentiable and has EIF

RKHS embedding of the LOCO variable importance measure

RKHS embedding of the LOCO variable importance measure:

$$\gamma_i^{\mathcal{K}}(P,S)(x) = \int_{\mathcal{X}} \mathcal{K}(x,x')\gamma_i(P,S)(x')P_X(dx'). \tag{5}$$

RKHS embedding of **CATE** is pathwise differentiable and has EIF

$\Downarrow$

RKHS embedding of the **LOCO variable importance measure**
is pathwise differentiable and has EIF

One-step estimator and weak convergence result

The **one-step estimator** of the RKHS embedding of the LOCO:

$$\underbrace{\hat{\gamma}^{\mathcal{K}}_{i,S}(\cdot)}_{\text{One-step estimator}} := \underbrace{\gamma^{\mathcal{K}}_{i}(\hat{P}_n, S)(\cdot)}_{\text{Plug-in estimator}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n} \phi^{S}_{n}(\cdot)}_{\text{Bias correction}} \tag{6}$$

We have the following **weak convergence** result:

$$n^{1/2}\left[\hat{\gamma}^{\mathcal{K}}_{i,S} - \gamma^{\mathcal{K}}_{i}(P_0, S)\right] \rightsquigarrow \mathbb{H}_S, \tag{7}$$

where $\mathbb{H}_S$ is a Hilbert-valued Gaussian random variable, s.t. for each $h \in \mathcal{H}$, $\langle \mathbb{H}_S, h\rangle_{\mathcal{H}} \sim \mathcal{N}\left(0, E_0\left[\langle \phi^S_0(Y, A, X), h\rangle^2_{\mathcal{H}}\right]\right)$.

One-step estimator and weak convergence result

The **one-step estimator** of the RKHS embedding of the LOCO:

$$\underbrace{\hat{\gamma}^{\mathcal{K}}_{i,S}(\cdot)}_{\text{One-step estimator}} := \underbrace{\gamma^{\mathcal{K}}_{i}(\hat{P}_n, S)(\cdot)}_{\text{Plug-in estimator}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n}\phi^{S}_{n}(\cdot)}_{\text{Bias correction}} \tag{6}$$

We have the following **weak convergence** result:

$$n^{1/2}\left[\hat{\gamma}^{\mathcal{K}}_{i,S} - \gamma^{\mathcal{K}}_{i}(P_0, S)\right] \rightsquigarrow \mathbb{H}_S, \tag{7}$$

where $\mathbb{H}_S$ is a Hilbert-valued Gaussian random variable, s.t. for each $h \in \mathcal{H}$, $\langle \mathbb{H}_S, h\rangle_{\mathcal{H}} \sim \mathcal{N}\left(0, E_0\left[\langle \phi^{S}_{0}(Y, A, X), h\rangle^2_{\mathcal{H}}\right]\right)$.

Wald-type test for RKHS embedding

- Equivalent test of zero-importance

$$H_0\colon \left\|\gamma_i^{\mathcal{K}}(P_0, S)\right\|_{\mathcal{H}} = 0$$

against the alternative

$$H_A\colon \left\|\gamma_i^{\mathcal{K}}(P_0, S)\right\|_{\mathcal{H}} \neq 0.$$

- One rejects the null hypothesis, if

$$\left\|\hat{\gamma}_{i,S}^{\mathcal{K}}\right\|_{\mathcal{H}} > \sqrt{\frac{\widehat{\xi}}{n}},$$

where $\widehat{\xi}$ is $(1-\alpha)$-quantile of $\|\mathbb{H}_S\|_{\mathcal{H}}^2$ obtained via bootstrap.

- This test asymptotically controls type I error at level α.
- It has asymptotically non-zero power against local alternatives.

Wald-type test for RKHS embedding

- Equivalent test of zero-importance

$$H_0\colon \left\|\gamma_i^{\mathcal{K}}(P_0, S)\right\|_{\mathcal{H}} = 0$$

against the alternative

$$H_A\colon \left\|\gamma_i^{\mathcal{K}}(P_0, S)\right\|_{\mathcal{H}} \neq 0.$$

- One rejects the null hypothesis, if

$$\left\|\hat{\gamma}_{i,S}^{\mathcal{K}}\right\|_{\mathcal{H}} > \sqrt{\frac{\widehat{\xi}}{n}},$$

where $\widehat{\xi}$ is $(1-\alpha)$-quantile of $\|\mathbb{H}_S\|_{\mathcal{H}}^2$ obtained via bootstrap.

- This test asymptotically controls type I error at level α.
- It has asymptotically non-zero power against local alternatives.

Wald-type test for RKHS embedding

- Equivalent test of zero-importance

$$H_0\colon \left\|\gamma_i^{\mathcal{K}}(P_0, S)\right\|_{\mathcal{H}} = 0$$

  against the alternative

$$H_A\colon \left\|\gamma_i^{\mathcal{K}}(P_0, S)\right\|_{\mathcal{H}} \neq 0.$$

- One rejects the null hypothesis, if

$$\left\|\hat{\gamma}_{i,S}^{\mathcal{K}}\right\|_{\mathcal{H}} > \sqrt{\frac{\widehat{\xi}}{n}},$$

  where $\widehat{\xi}$ is $(1-\alpha)$-quantile of $\|\mathbb{H}_S\|_{\mathcal{H}}^2$ obtained via bootstrap.

- This test asymptotically controls type I error at level α.
- It has asymptotically non-zero power against local alternatives.

Wald-type test for RKHS embedding

- Equivalent test of zero-importance

$$H_0\colon \left\|\gamma_i^{\mathcal{K}}(P_0, S)\right\|_{\mathcal{H}} = 0$$

against the alternative

$$H_A\colon \left\|\gamma_i^{\mathcal{K}}(P_0, S)\right\|_{\mathcal{H}} \neq 0.$$

- One rejects the null hypothesis, if

$$\left\|\hat{\gamma}_{i,S}^{\mathcal{K}}\right\|_{\mathcal{H}} > \sqrt{\frac{\widehat{\xi}}{n}},$$

where $\widehat{\xi}$ is $(1-\alpha)$-quantile of $\|\mathbb{H}_S\|_{\mathcal{H}}^2$ obtained via bootstrap.

- This test asymptotically controls type I error at level α.
- It has asymptotically non-zero power against local alternatives.

Wald-type test for RKHS embedding

- Equivalent test of zero-importance

$$H_0\colon \left\| \gamma_i^{\mathcal{K}}(P_0, S) \right\|_{\mathcal{H}} = 0$$

against the alternative

$$H_A\colon \left\| \gamma_i^{\mathcal{K}}(P_0, S) \right\|_{\mathcal{H}} \neq 0.$$

- One rejects the null hypothesis, if

$$\left\| \hat{\gamma}_{i,S}^{\mathcal{K}} \right\|_{\mathcal{H}} > \sqrt{\frac{\widehat{\xi}}{n}},$$

where $\widehat{\xi}$ is $(1-\alpha)$-quantile of $\|\mathbb{H}_S\|_{\mathcal{H}}^2$ obtained via bootstrap.

- This test asymptotically controls type I error at level α.
- It has asymptotically non-zero power against local alternatives.

Limitations of the LOCO

- If X_i and X_j, for $j \in S$, are **highly correlated**, then $\gamma_i(P_0, S)$ will typically be close to 0, potentially underestimating the importance of X_i.

- LOCO does not capture **interaction effects** well. A variable might appear unimportant alone but be crucial in interaction with another feature.

- The definition of LOCO relies on an **arbitrary choice** of the subset S.

Limitations of the LOCO

- If X_i and X_j, for $j \in S$, are **highly correlated**, then $\gamma_i\,(P_0, S)$ will typically be close to 0, potentially underestimating the importance of X_i.

- LOCO does not capture **interaction effects** well. A variable might appear unimportant alone but be crucial in interaction with another feature.

- The definition of LOCO relies on an **arbitrary choice** of the subset S.

Limitations of the LOCO

- If X_i and X_j, for $j \in S$, are **highly correlated**, then $\gamma_i(P_0, S)$ will typically be close to 0, potentially underestimating the importance of X_i.
- LOCO does not capture **interaction effects** well. A variable might appear unimportant alone but be crucial in interaction with another feature.
- The definition of LOCO relies on an **arbitrary choice** of the subset S.

Cooperative game theory

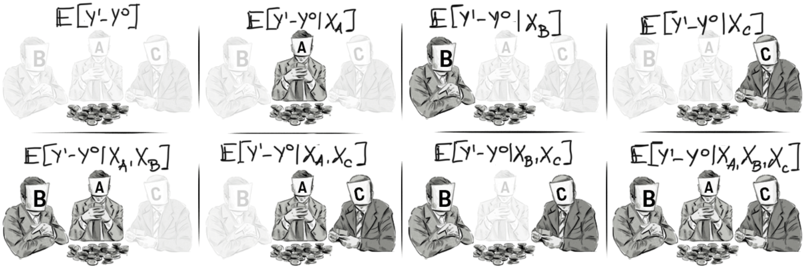
Cooperative game theory aims to distribute the payout among a group of cooperative players in a fair manner.

Source: https://bernard-mlab.com/post/mta-sharpley-value/

Shapley values

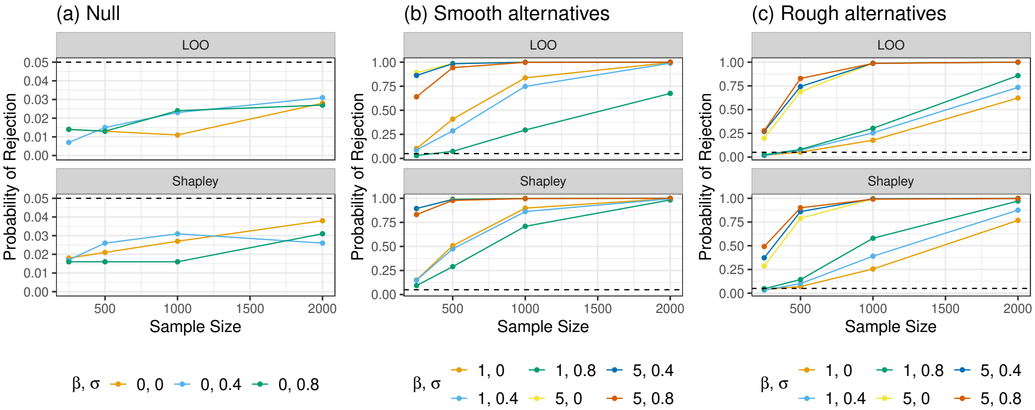
Shapley value of i^{th} feature for CATE:

$$\zeta_i(P)(x) = \frac{1}{d} \sum_{S \subseteq [d] \setminus \{i\}} \binom{d-1}{|S|}^{-1} \left\{ \nu(P; S \cup \{i\})(x_{S \cup \{i\}}) - \nu(P; S)(x_S) \right\} \tag{8}$$



Source: https://bernard-mlab.com/post/mta-sharpley-value/

Simulations



Discussion

- We present an inferential framework for evaluating **variable importance** for **heterogeneous treatment effects**.
- We provide **efficient estimator** of this measure together **with corresponding confidence bands**, and introduce a **Wald-type test** to assess the **null hypothesis of zero-importance**.
- Our approach focuses on **local variable importance measures** that vary among individuals, coupled with **global inference** that assesses the overall significance of a variable across all individuals.

Discussion

- We present an inferential framework for evaluating **variable importance** for **heterogeneous treatment effects**.
- We provide **efficient estimator** of this measure together **with corresponding confidence bands**, and introduce a **Wald-type test** to assess the **null hypothesis of zero-importance**.
- Our approach focuses on **local variable importance measures** that vary among individuals, coupled with **global inference** that assesses the overall significance of a variable across all individuals.

Discussion

- We present an inferential framework for evaluating **variable importance** for **heterogeneous treatment effects**.
- We provide **efficient estimator** of this measure together **with corresponding confidence bands**, and introduce a **Wald-type test** to assess the **null hypothesis of zero-importance**.
- Our approach focuses on **local variable importance measures** that vary among individuals, coupled with **global inference** that assesses the overall significance of a variable across all individuals.

Discussion

- We present an inferential framework for evaluating **variable importance** for **heterogeneous treatment effects**.
- We provide **efficient estimator** of this measure together **with corresponding confidence bands**, and introduce a **Wald-type test** to assess the **null hypothesis of zero-importance**.
- Our approach focuses on **local variable importance measures** that vary among individuals, coupled with **global inference** that assesses the overall significance of a variable across all individuals.

Selected References

[1] Luedtke, A., Chung, I., One-step estimation of differentiable Hilbert-valued parameters. Annals of Statistics 2024.

[2] Shapley, L., A value for n-person games. Contributions to the Theory of Games (1953).

[3] Williamson B., Gilbert P., Carone M., Simon N., Nonparametric variable importance assessment using machine learning techniques. Biometrics 2019.

[4] Williamson B., Gilbert P., Simon N., Carone M., A General Framework for Inference on Algorithm-Agnostic Variable Importance. Journal of the American Statistical Association 2022.

Thank You!

E-mail: pawel.morzywolek@sund.ku.dk

Website: pmorzywolek.github.io

Pacific Institute for the Mathematical Sciences

eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

Appendix

Efficient influence function

The efficient influence function for $\gamma_i^{\mathcal{K}}(P, S)$ is

$$
\begin{aligned}
&\phi_P^S(y, a, x)\left(x'\right) = \\
&\left[E_P\left\{\mathcal{K}_{x'}(X) \mid X_{S\cup\{i\}} = x_{S\cup\{i\}}\right\} - E_P\left\{\mathcal{K}_{x'}(X) \mid X_S = x_S\right\}\right]\left[\frac{2a-1}{g_P(1 \mid x)}\left\{y - \mu_{P,a}(x)\right\} + \mu_{P,1}(x) - \mu_{P,0}(x)\right] \\
&+ \left[\mathcal{K}_{x'}(x) - E_P\left\{\mathcal{K}_{x'}(X) \mid X_{S\cup\{i\}} = x_{S\cup\{i\}}\right\}\right] E_P\left\{\mu_{P,1}(X) - \mu_{P,0}(X) \mid X_{S\cup\{i\}} = x_{S\cup\{i\}}\right\} \\
&- \left[\mathcal{K}_{x'}(x) - E_P\left\{\mathcal{K}_{x'}(X) \mid X_S = x_S\right\}\right] E_P\left\{\mu_{P,1}(X) - \mu_{P,0}(X) \mid X_S = x_S\right\} \\
&- \underbrace{\left(E_P\left[\mathcal{K}\left(x', X\right) E_P\left\{\mu_{P,1}(X) - \mu_{P,0}(X) \mid X_{S\cup\{i\}}\right\}\right] - E_P\left[\mathcal{K}\left(x', X\right) E_P\left\{\mu_{P,1}(X) - \mu_{P,0}(X) \mid X_S\right\}\right]\right)}_{\gamma_i^{\mathcal{K}}(P,S)(x')}
\end{aligned}
$$

Confidence interval

A confidence interval for $\left\| \gamma_i^{\mathcal{K}}(P_0, S) \right\|_{\mathcal{H}}$ is given by

$$\left\| \hat{\gamma}_{i,S}^{\mathcal{K}} \right\|_{\mathcal{H}} \pm \sqrt{\widehat{\xi}/n}.$$

This is justified by the reverse triangle inequality and the fact that the spherical confidence set is asymptotically valid, since

$$\left| \left\| \hat{\gamma}_{i,S}^{\mathcal{K}} \right\|_{\mathcal{H}} - \left\| \gamma_i^{\mathcal{K}}(P_0, S) \right\|_{\mathcal{H}} \right| \leq \left\| \hat{\gamma}_{i,S}^{\mathcal{K}} - \gamma_i^{\mathcal{K}}(P_0, S) \right\|_{\mathcal{H}} \leq \sqrt{\widehat{\xi}/n}$$

with probability tending to $(1 - \alpha)$.
The presented confidence interval is asymptotically non-conservative in the special case where $\gamma_i^{\mathcal{K}}(P_0, S) = 0$, but presumably will otherwise be conservative.

General variable importance estimand

- For a vector of weights $\omega = \{\omega_S : S \subseteq [d]\}$, the general variable importance estimand is defined as follows:

$$\theta_\omega(P)(x) := \sum_{S \subseteq [d]} \omega_S \nu(P; S)(x).$$

- The parameter $\gamma_i(P, S)$ is a special case of $\theta_\omega(P)$ with $\omega_{S \cup \{i\}} = 1$ and $\omega_S = -1$.
- The parameter $\zeta_i(P)$ is a special case of $\theta_\omega(P)$ with $\omega_{S \cup \{i\}} = \frac{1}{d}\binom{d-1}{|S|}^{-1}$ and $\omega_S = -\frac{1}{d}\binom{d-1}{|S|}^{-1}$, for $S \subseteq [d] \setminus \{i\}$.