A unified approach to nonparametric variable importance assessment

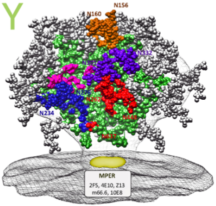
Brian D. Williamson
Marco Carone, Noah Simon
Department of Biostatistics, University of Washington
Funding: NIAID F31AI140836

31 July 2019

Motivation

AMP
ANTIBODY · MEDIATED · PREVENTION
STUDY

W CARDIOLOGY
UNIVERSITY OF WASHINGTON



Motivation

In general: data $O_1, \ldots, O_n \overset{iid}{\sim} P_0$

Motivation

In general: data $O_1, \ldots, O_n \overset{iid}{\sim} P_0$

- $O_i := (X_i, Y_i)$;

Motivation

In general: data $O_1, \ldots, O_n \overset{iid}{\sim} P_0$

- $O_i := (X_i, Y_i)$;
- $X_i \in \mathbb{R}^p$ is a vector of covariates

Motivation

In general: data $O_1, \ldots, O_n \overset{iid}{\sim} P_0$

- $O_i := (X_i, Y_i)$;
- $X_i \in \mathbb{R}^p$ is a vector of covariates
- $Y_i \in \mathbb{R}$ is the outcome of interest

Motivation

In general: data $O_1, \ldots, O_n \overset{iid}{\sim} P_0$

- $O_i := (X_i, Y_i)$;
- $X_i \in \mathbb{R}^p$ is a vector of covariates
- $Y_i \in \mathbb{R}$ is the outcome of interest

Goal: estimate the conditional mean, $E_{P_0}(Y \mid X = x)$

Motivation

In general: data $O_1, \ldots, O_n \overset{iid}{\sim} P_0$

- $O_i := (X_i, Y_i)$;
- $X_i \in \mathbb{R}^p$ is a vector of covariates
- $Y_i \in \mathbb{R}$ is the outcome of interest

Goal: estimate the conditional mean, $E_{P_0}(Y \mid X = x)$

Goal: describe the predictiveness of our estimator

Motivation

In general: data $O_1, \ldots, O_n \overset{iid}{\sim} P_0$

- $O_i := (X_i, Y_i)$;
- $X_i \in \mathbb{R}^p$ is a vector of covariates
- $Y_i \in \mathbb{R}$ is the outcome of interest

Goal: estimate the conditional mean, $E_{P_0}(Y \mid X = x)$

Goal: describe the predictiveness of our estimator

Goal: compare predictiveness of multiple estimators

Linear regression variable importance

Objectives:

1. estimate the conditional mean
2. estimate the importance of X_s, $s \subseteq \{1, \ldots, p\}$

How do we typically do this in linear regression?

Linear regression variable importance

Objectives:

1. estimate the conditional mean
2. estimate the importance of X_s, $s \subseteq \{1, \ldots, p\}$

How do we typically do this in linear regression?

1. Fit a linear regression of Y on $X \rightarrow \hat{\mu}(X)$

Linear regression variable importance

Objectives:

1. estimate the conditional mean
2. estimate the importance of X_s, $s \subseteq \{1, \ldots, p\}$

How do we typically do this in linear regression?

1. Fit a linear regression of Y on $X \to \hat{\mu}(X)$
2. Fit a linear regression of Y on $X_{-s} \to \hat{\mu}_s(X)$

Linear regression variable importance

Objectives:

1. estimate the conditional mean
2. estimate the importance of X_s, $s \subseteq \{1, \dots, p\}$

How do we typically do this in linear regression?

1. Fit a linear regression of Y on $X \rightarrow \hat{\mu}(X)$
2. Fit a linear regression of Y on $X_{-s} \rightarrow \hat{\mu}_s(X)$
3. Compare the fitted values $[\hat{\mu}(X_i), \hat{\mu}_s(X_i)]$

Linear regression variable importance

Objectives:

1. estimate the conditional mean
2. estimate the importance of X_s, $s \subseteq \{1, \dots, p\}$

How do we typically do this in linear regression?

1. Fit a linear regression of Y on $X \to \hat{\mu}(X)$
2. Fit a linear regression of Y on $X_{-s} \to \hat{\mu}_s(X)$
3. Compare the fitted values $[\hat{\mu}(X_i), \hat{\mu}_s(X_i)]$

Many ways to compare fitted values, including:

- ANOVA decomposition
- Difference in R^2

Linear regression variable importance

Difference in R^2:

$$\left[1 - \frac{MSE(\hat{\mu})}{n^{-1}\sum_{i=1}^{n}\{Y_i - \overline{Y}_n\}^2}\right] - \left[1 - \frac{MSE(\hat{\mu}_s)}{n^{-1}\sum_{i=1}^{n}\{Y_i - \overline{Y}_n\}^2}\right]$$

Mean squared error (MSE) of linear regression function f:

$$MSE(f) = \frac{1}{n}\sum_{i=1}^{n}\{Y_i - f(X_i)\}^2$$

Flexible estimator variable importance?

Linear regression: excellent performance . . .

Flexible estimator variable importance?

Linear regression: excellent performance . . .

. . . if the true relationship is captured well by a linear model.

Flexible estimator variable importance?

Linear regression: excellent performance . . .

. . . if the true relationship is captured well by a linear model.

What if I have a lot of data?

Flexible estimator variable importance?

Linear regression: excellent performance . . .

. . . if the true relationship is captured well by a linear model.

What if I have a lot of data? fit a more flexible estimator!

Flexible estimator variable importance?

Linear regression: excellent performance ...

... if the true relationship is captured well by a linear model.

What if I have a lot of data? fit a more flexible estimator!

- better predictions?

Flexible estimator variable importance?

Linear regression: excellent performance . . .

. . . if the true relationship is captured well by a linear model.

What if I have a lot of data? fit a more flexible estimator!

- better predictions?
- how do I define importance?

Population variable importance

What if we could predict perfectly?

Population variable importance

What if we could predict perfectly?

Oracle prediction functions:

- $\mu_0(x) := E_{P_0}(Y \mid X = x)$
- $\mu_{0,s}(x) := E_{P_0}(Y \mid X_{-s} = x_{-s})$

Population variable importance

What if we could predict perfectly?

Oracle prediction functions:

- $\mu_0(x) := E_{P_0}(Y \mid X = x)$
- $\mu_{0,s}(x) := E_{P_0}(Y \mid X_{-s} = x_{-s})$

Define population importance in terms of $\mu_0, \mu_{0,s}$!

Population variable importance

$$
\begin{aligned}
DR^2_s(P_0) &:= R^2(\mu_0, P_0) - R^2(\mu_{0,s}, P_0) \\
R^2(\mu, P_0) &:= 1 - \frac{MSE(\mu, P_0)}{var_{P_0}(Y)} \\
MSE(\mu, P_0) &:= E_{P_0}\{Y - \mu(X)\}^2
\end{aligned}
$$

Population variable importance

$$
\begin{aligned}
DR_s^2(P_0) &:= R^2(\mu_0, P_0) - R^2(\mu_{0,s}, P_0) \\
R^2(\mu, P_0) &:= 1 - \frac{MSE(\mu, P_0)}{var_{P_0}(Y)} \\
MSE(\mu, P_0) &:= E_{P_0}\{Y - \mu(X)\}^2
\end{aligned}
$$

Large $R^2 \rightarrow$ high prediction accuracy

Measures predictiveness

Population variable importance

$$
\begin{aligned}
DR_s^2(P_0) &:= R^2(\mu_0, P_0) - R^2(\mu_{0,s}, P_0) \\
R^2(\mu, P_0) &:= 1 - \frac{MSE(\mu, P_0)}{var_{P_0}(Y)} \\
MSE(\mu, P_0) &:= E_{P_0}\{Y - \mu(X)\}^2
\end{aligned}
$$

Large $R^2 \rightarrow$ high prediction accuracy

Measures predictiveness

Variable importance: comparing population predictiveness!

Population variable importance

In general: consider function classes $\mathcal{F}_s \subseteq \mathcal{F} \subseteq L_2(P_0)$

Statistical framework for variable importance:

Population variable importance

In general: consider function classes $\mathcal{F}_s \subseteq \mathcal{F} \subseteq L_2(P_0)$

Statistical framework for variable importance:

- $V(f, P)$: predictiveness of $f \in \mathcal{F}$ under distribution P

Population variable importance

In general: consider function classes $\mathcal{F}_s \subseteq \mathcal{F} \subseteq L_2(P_0)$

Statistical framework for variable importance:

- $V(f, P)$: predictiveness of $f \in \mathcal{F}$ under distribution P
- $f_P^* := \arg\max_{f \in \mathcal{F}} V(f, P)$: predictiveness maximizer over $\mathcal{F}$

Population variable importance

In general: consider function classes $\mathcal{F}_s \subseteq \mathcal{F} \subseteq L_2(P_0)$

Statistical framework for variable importance:

- $V(f, P)$: predictiveness of $f \in \mathcal{F}$ under distribution P
- $f_P^* := \arg\max_{f \in \mathcal{F}} V(f, P)$: predictiveness maximizer over $\mathcal{F}$

Variable importance: $\Psi_s(P_0) := V(f_{P_0}^*, P_0) - V(f_{P_0,s}^*, P_0)$

Plug-in estimators of variable importance

$\psi_{0,s} \equiv \Psi_s(P_0) = V(f^*_{P_0}, P_0) - V(f^*_{P_0,s}, P_0)$

Natural estimator:

Plug-in estimators of variable importance

$\psi_{0,s} \equiv \Psi_s(P_0) = V(f^*_{P_0}, P_0) - V(f^*_{P_0,s}, P_0)$

Natural estimator:

1. obtain estimator $\hat{f}_n$ of $f^*_{P_0}$

Plug-in estimators of variable importance

$\psi_{0,s} \equiv \Psi_s(P_0) = V(f^*_{P_0}, P_0) - V(f^*_{P_0,s}, P_0)$

Natural estimator:

1. obtain estimator $\hat{f}_n$ of $f^*_{P_0}$
2. obtain estimator $\hat{f}_{n,s}$ of $f^*_{P_0,s}$

Plug-in estimators of variable importance

$\psi_{0,s} \equiv \Psi_s(P_0) = V(f^*_{P_0}, P_0) - V(f^*_{P_0,s}, P_0)$

Natural estimator:

1. obtain estimator $\hat{f}_n$ of $f^*_{P_0}$
2. obtain estimator $\hat{f}_{n,s}$ of $f^*_{P_0,s}$
3. Plug in: $\hat{\psi}_{n,s} := V(\hat{f}_n, \mathbb{P}_n) - V(\hat{f}_{n,s}, \mathbb{P}_n)$

Plug-in estimators of variable importance

$\psi_{0,s} \equiv \Psi_s(P_0) = V(f^*_{P_0}, P_0) - V(f^*_{P_0,s}, P_0)$

Natural estimator:

1. obtain estimator $\hat{f}_n$ of $f^*_{P_0}$
2. obtain estimator $\hat{f}_{n,s}$ of $f^*_{P_0,s}$
3. Plug in: $\hat{\psi}_{n,s} := V(\hat{f}_n, \mathbb{P}_n) - V(\hat{f}_{n,s}, \mathbb{P}_n)$

Questions:

- when is $\hat{\psi}_{n,s}$ regular and asymptotically linear?
- can we test $H_0 : \psi_{0,s} = 0$?

U parameters

Specify

$$V(f, P) = E_P G\{O_1, \ldots, O_m, f(O_1), \ldots, f(O_m)\} \equiv E_P G\{O, f(O)\}$$

Examples:

U parameters

Specify

$$V(f,P) = E_P G\{O_1,\ldots,O_m, f(O_1),\ldots,f(O_m)\} \equiv E_P G\{O, f(O)\}$$

Examples:

- $R^2(f,P) = 1 - \frac{MSE(f,P)}{var_P(Y)} = 1 - \frac{E_P\{Y-f(X)\}^2}{var_P(Y)}$

U parameters

Specify

$$V(f, P) = E_P G\{O_1, \ldots, O_m, f(O_1), \ldots, f(O_m)\} \equiv E_P G\{O, f(O)\}$$

Examples:

- $R^2(f, P) = 1 - \frac{MSE(f,P)}{var_P(Y)} = 1 - \frac{E_P\{Y-f(X)\}^2}{var_P(Y)}$
- Accuracy: $A(f, P) := 1 - E_P I\{Y \neq f(X)\}$

U parameters

Specify

$$V(f,P) = E_P G\{O_1,\ldots,O_m, f(O_1),\ldots,f(O_m)\} \equiv E_P G\{O, f(O)\}$$

Examples:

- $R^2(f,P) = 1 - \frac{MSE(f,P)}{var_P(Y)} = 1 - \frac{E_P\{Y - f(X)\}^2}{var_P(Y)}$
- Accuracy: $A(f,P) := 1 - E_P I\{Y \neq f(X)\}$
- AUC: $\frac{E_P E_P[I\{f(X_1) < f(X_2)\} I\{Y_1 = 0, Y_2 = 1\}]}{P(Y=0)P(Y=1)}$

U parameters

Specify

$$V(f,P) = E_P G\{O_1,\ldots,O_m, f(O_1),\ldots,f(O_m)\} \equiv E_P G\{O, f(O)\}$$

Examples:

- $R^2(f,P) = 1 - \frac{MSE(f,P)}{var_P(Y)} = 1 - \frac{E_P\{Y - f(X)\}^2}{var_P(Y)}$
- Accuracy: $A(f,P) := 1 - E_P I\{Y \neq f(X)\}$
- AUC: $\frac{E_P E_P[I\{f(X_1) < f(X_2)\} I\{Y_1 = 0, Y_2 = 1\}]}{P(Y=0)P(Y=1)}$

U parameter: $V(f_P^*, P)$

U parameters

Specify

$$V(f,P) = E_P G\{O_1,\ldots,O_m, f(O_1),\ldots,f(O_m)\} \equiv E_P G\{O, f(O)\}$$

Examples:

- $R^2(f,P) = 1 - \frac{MSE(f,P)}{var_P(Y)} = 1 - \frac{E_P\{Y-f(X)\}^2}{var_P(Y)}$
- Accuracy: $A(f,P) := 1 - E_P I\{Y \neq f(X)\}$
- AUC: $\frac{E_P E_P[I\{f(X_1)<f(X_2)\} I\{Y_1=0, Y_2=1\}]}{P(Y=0)P(Y=1)}$

U parameter: $V(f_P^*, P)$

U estimator: $V(\hat{f}_n, \mathbb{P}_n)$, plug-in estimator of a U parameter

Asymptotic distribution of U estimators

Goal: estimate $\Psi_s(P_0) = V(f^*_{P_0}, P_0) - V(f^*_{P_0,s}, P_0)$

Asymptotic distribution of U estimators

Goal: estimate $\Psi_s(P_0) = V(f^*_{P_0}, P_0) - V(f^*_{P_0,s}, P_0)$

Focus on $V(f^*_{P_0}, P_0)$

Asymptotic distribution of U estimators

Goal: estimate $\Psi_s(P_0) = V(f^*_{P_0}, P_0) - V(f^*_{P_0,s}, P_0)$

Focus on $V(f^*_{P_0}, P_0)$

Set $Pf = \int f(o)dP(o)$ for a P-measurable function f

Asymptotic distribution of U estimators

Goal: estimate $\Psi_s(P_0) = V(f^*_{P_0}, P_0) - V(f^*_{P_0,s}, P_0)$

Focus on $V(f^*_{P_0}, P_0)$

Set $Pf = \int f(o)dP(o)$ for a P-measurable function f

Linearize V about P_n using gradient D^*:

Asymptotic distribution of U estimators

Goal: estimate $\Psi_s(P_0) = V(f^*_{P_0}, P_0) - V(f^*_{P_0,s}, P_0)$

Focus on $V(f^*_{P_0}, P_0)$

Set $Pf = \int f(o)dP(o)$ for a P-measurable function f

Linearize V about P_n using gradient D^*:

$$\begin{aligned} V(\hat{f}_n, P_n) - V(f^*_{P_0}, P_0) &= (P_n - P_0)D^*(P_n) + R(P_n, P_0) \\ &= (\mathbb{P}_n - P_0)D^*(P_0) - \mathbb{P}_n D^*(P_n) \\ &\quad + (\mathbb{P}_n - P_0)\{D^*(P_n) - D^*(P_0)\} \\ &\quad + R(P_n, P_0) \end{aligned}$$

Asymptotic distribution of U estimators

Goal: estimate $\Psi_s(P_0) = V(f^*_{P_0}, P_0) - V(f^*_{P_0,s}, P_0)$

Focus on $V(f^*_{P_0}, P_0)$

Set $Pf = \int f(o)dP(o)$ for a P-measurable function f

Linearize V about P_n using gradient D^*:

$$
\begin{aligned}
V(\hat{f}_n, P_n) - V(f^*_{P_0}, P_0) &= (P_n - P_0)D^*(P_n) + R(P_n, P_0) \\
&= (\mathbb{P}_n - P_0)D^*(P_0) - \mathbb{P}_n D^*(P_n) \\
&\quad + (\mathbb{P}_n - P_0)\{D^*(P_n) - D^*(P_0)\} \\
&\quad + R(P_n, P_0)
\end{aligned}
$$

- linear term; (1st order)

Asymptotic distribution of U estimators

Goal: estimate $\Psi_s(P_0) = V(f^*_{P_0}, P_0) - V(f^*_{P_0,s}, P_0)$

Focus on $V(f^*_{P_0}, P_0)$

Set $Pf = \int f(o)dP(o)$ for a P-measurable function f

Linearize V about P_n using gradient D^*:

$$\begin{aligned} V(\hat{f}_n, P_n) - V(f^*_{P_0}, P_0) &= (P_n - P_0)D^*(P_n) + R(P_n, P_0) \\ &= (\mathbb{P}_n - P_0)D^*(P_0) - \mathbb{P}_n D^*(P_n) \\ &\quad + (\mathbb{P}_n - P_0)\{D^*(P_n) - D^*(P_0)\} \\ &\quad + R(P_n, P_0) \end{aligned}$$

- linear term; (1st order)
- empirical process term; (2nd order)

Asymptotic distribution of U estimators

Goal: estimate $\Psi_s(P_0) = V(f^*_{P_0}, P_0) - V(f^*_{P_0,s}, P_0)$

Focus on $V(f^*_{P_0}, P_0)$

Set $Pf = \int f(o)dP(o)$ for a P-measurable function f

Linearize V about P_n using gradient D^*:

$$\begin{aligned} V(\hat{f}_n, P_n) - V(f^*_{P_0}, P_0) &= (P_n - P_0)D^*(P_n) + R(P_n, P_0) \\ &= (\mathbb{P}_n - P_0)D^*(P_0) - \mathbb{P}_n D^*(P_n) \\ &\quad + (\mathbb{P}_n - P_0)\{D^*(P_n) - D^*(P_0)\} \\ &\quad + R(P_n, P_0) \end{aligned}$$

- linear term; (1st order)
- empirical process term; (2nd order)
- remainder term; (2nd order)

Asymptotic distribution of U estimators

Goal: estimate $\Psi_s(P_0) = V(f^*_{P_0}, P_0) - V(f^*_{P_0,s}, P_0)$

Focus on $V(f^*_{P_0}, P_0)$

Set $Pf = \int f(o)dP(o)$ for a P-measurable function f

Linearize V about P_n using gradient D^*:

$$\begin{aligned} V(\hat{f}_n, P_n) - V(f^*_{P_0}, P_0) &= (P_n - P_0)D^*(P_n) + R(P_n, P_0) \\ &= {\color{cyan}(\mathbb{P}_n - P_0)D^*(P_0)} - {\color{red}\mathbb{P}_n D^*(P_n)} \\ &\quad + {\color{blue}(\mathbb{P}_n - P_0)\{D^*(P_n) - D^*(P_0)\}} \\ &\quad + R(P_n, P_0) \end{aligned}$$

- linear term; (1st order)
- empirical process term; (2nd order)
- remainder term; (2nd order)
- problem term! (1/2-ish order, usually)

Asymptotic distribution of U estimators

If V is a U parameter, then

Asymptotic distribution of U estimators

If V is a U parameter, then

- $\mathbb{P}_n D^*(P_n) = 0$

Asymptotic distribution of U estimators

If V is a U parameter, then

- $\mathbb{P}_n D^*(P_n) = 0$
- D^* is the efficient influence function (EIF)

Asymptotic distribution of U estimators

If V is a U parameter, then

- $\mathbb{P}_n D^*(P_n) = 0$
- D^* is the efficient influence function (EIF)

Under optimization* and convergence rate† conditions,

Asymptotic distribution of U estimators

If V is a U parameter, then

- $\mathbb{P}_n D^*(P_n) = 0$
- D^* is the efficient influence function (EIF)

Under optimization* and convergence rate† conditions,

- $\hat{\psi}_{n,s}$ asymptotically linear with influence function D^*

Asymptotic distribution of U estimators

If V is a U parameter, then

- $\mathbb{P}_n D^*(P_n) = 0$
- D^* is the efficient influence function (EIF)

Under optimization* and convergence rate† conditions,

- $\hat{\psi}_{n,s}$ asymptotically linear with influence function D^*
- $\hat{\psi}_{n,s}$ regular

Asymptotic distribution of U estimators

If V is a U parameter, then

- $\mathbb{P}_n D^*(P_n) = 0$
- D^* is the efficient influence function (EIF)

Under optimization* and convergence rate† conditions,

- $\hat{\psi}_{n,s}$ asymptotically linear with influence function D^*
- $\hat{\psi}_{n,s}$ regular
- If $\sigma^2_{0,s} := E_{P_0}\{D^*(P_0)(O)^2\} > 0$, then

$$\sqrt{n}(\hat{\psi}_{n,s} - \psi_{0,s}) \rightarrow_d Z \sim N(0, \sigma^2_{0,s})$$

Asymptotic distribution of U estimators

If V is a U parameter, then

- $\mathbb{P}_n D^*(P_n) = 0$
- D^* is the efficient influence function (EIF)

Under optimization* and convergence rate† conditions,

- $\hat{\psi}_{n,s}$ asymptotically linear with influence function D^*
- $\hat{\psi}_{n,s}$ regular
- If $\sigma^2_{0,s} := E_{P_0}\{D^*(P_0)(O)^2\} > 0$, then

$$\sqrt{n}(\hat{\psi}_{n,s} - \psi_{0,s}) \to_d Z \sim N(0, \sigma^2_{0,s})$$

*: derivative of V at $f^*_{P_0}$ is 0

†: $E_{P_0}\{\hat{f}_n(O) - f^*_{P_0}(O)\}^2 = o_P(n^{-1/2})$

Estimation and hypothesis testing

For R^2, AUC, accuracy, . . .

Estimation and hypothesis testing

For R^2, AUC, accuracy, $\ldots\, \hat{\psi}_{n,s}$ is efficient

Estimation and hypothesis testing

For R^2, AUC, accuracy, $\ldots \hat{\psi}_{n,s}$ is efficient

Hypothesis testing: under $H_0 : \psi_{0,s} = 0$, $D^*(P_0)(o) = 0$

Estimation and hypothesis testing

For R^2, AUC, accuracy, $\ldots \hat{\psi}_{n,s}$ is efficient

Hypothesis testing: under $H_0 : \psi_{0,s} = 0$, $D^*(P_0)(o) = 0$

Instead:

1. Split data in two parts

Estimation and hypothesis testing

For R^2, AUC, accuracy, $\ldots$ $\hat{\psi}_{n,s}$ is efficient

Hypothesis testing: under $H_0 : \psi_{0,s} = 0$, $D^*(P_0)(o) = 0$

Instead:

1. Split data in two parts
2. Obtain CI for $V(f^*_{P_0}, P_0)$

Estimation and hypothesis testing

For R^2, AUC, accuracy, $\ldots \hat{\psi}_{n,s}$ is efficient

Hypothesis testing: under $H_0 : \psi_{0,s} = 0$, $D^*(P_0)(o) = 0$

Instead:

1. Split data in two parts
2. Obtain CI for $V(f^*_{P_0}, P_0)$
3. Obtain CI for $V(f^*_{P_0,s}, P_0)$

Estimation and hypothesis testing

For R^2, AUC, accuracy, $\ldots \hat{\psi}_{n,s}$ is efficient

Hypothesis testing: under $H_0 : \psi_{0,s} = 0$, $D^*(P_0)(o) = 0$

Instead:

1. Split data in two parts
2. Obtain CI for $V(f^*_{P_0}, P_0)$
3. Obtain CI for $V(f^*_{P_0,s}, P_0)$
4. If CIs do not overlap, reject H_0

Experiment: binary outcome, bivariate feature vector

$Y \sim Bern(0.6)$; $X_1 \mid Y \sim N(\mu_1, \Sigma)$

Under H_0, $X_2 \mid Y \sim N(0, \Sigma)$

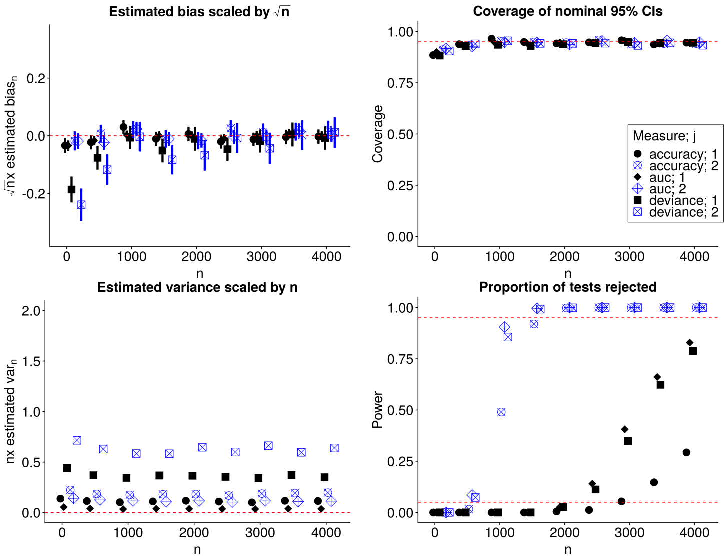
Under H_1, $X_2 \mid Y \sim N(\mu_2, \Sigma)$

Investigate:

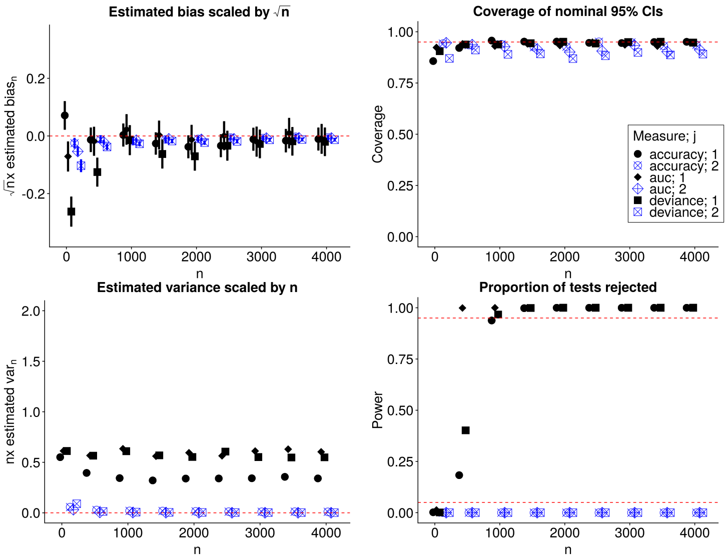
- scaled bias
- coverage of nominal 95% CIs
- type I error (or power)

Estimate using cross-validation and regression stacking

Experiment: results under the alternative



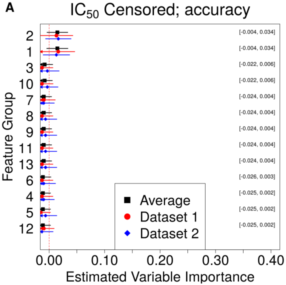
Experiment: results under the null

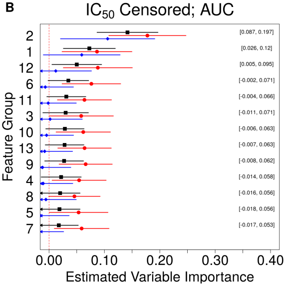


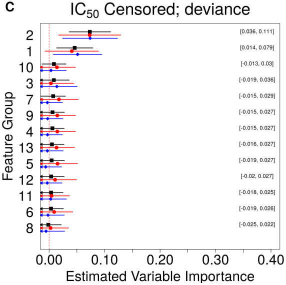
Studying an antibody against HIV-1 infection

- 611 HIV-1 pseudoviruses, split into two datasets
- Outcome: neutralization sensitivity to antibody
- 493 individual features, 12 groups of interest
- Estimate using cross-validation and regression stacking

Studying an antibody against HIV-1 infection







D Feature group definitions

1	VRC01 binding footprint
2	CD4 binding sites
3	Sites with sufficient exposed surface area
4	Sites identified as important for glycosylation
5	Sites with residues that covary with the VRC01 binding footprint
6	Sites associated with VRC01-specific potential N-linked glycosylation (PNGS) effects
7	gp41 sites important for VRC01 binding
8	Sites for indicating N-linked glycosylation
9	Majority virus subtypes
10	Region-specific counts of PNGS
11	Viral geometry
12	Cysteine counts
13	Steric bulk at critical locations

Conclusions

Variable importance: comparing population predictiveness

Conclusions

Variable importance: comparing population predictiveness

Plug-in estimators of R^2, AUC, accuracy, . . . efficient

Conclusions

Variable importance: comparing population predictiveness

Plug-in estimators of R^2, AUC, accuracy, ... efficient

Asymptotically valid CIs and hypothesis testing ...

Conclusions

Variable importance: comparing population predictiveness

Plug-in estimators of R^2, AUC, accuracy, ... efficient

Asymptotically valid CIs and hypothesis testing ...

... based on flexible estimators

Conclusions

Variable importance: comparing population predictiveness

Plug-in estimators of R^2, AUC, accuracy, ... efficient

Asymptotically valid CIs and hypothesis testing ...

... based on flexible estimators

`github.com/bdwilliamson/vimp`

Appendix: Asymptotic distribution of U estimators

Problem is simple for fixed f:

Appendix: Asymptotic distribution of U estimators

Problem is simple for fixed f:

- Only need to estimate P, use empirical $\mathbb{P}_n$

Appendix: Asymptotic distribution of U estimators

Problem is simple for fixed f:

- Only need to estimate P, use empirical $\mathbb{P}_n$

$$V(f, \mathbb{P}_n) - V(f, P) = \frac{1}{n} \sum_{i=1}^{n} [G\{O_i, f(O_i)\} - E_P G\{O_i, f(O_i)\}]$$

Appendix: Asymptotic distribution of U estimators

Problem is simple for fixed f:

- Only need to estimate P, use empirical $\mathbb{P}_n$

$$V(f, \mathbb{P}_n) - V(f, P) = \frac{1}{n} \sum_{i=1}^{n} [G\{O_i, f(O_i)\} - E_P G\{O_i, f(O_i)\}]$$

$V(f, P)$ is **linear** with EIF

$$D(f, P)(o) := G\{o, f(o)\} - E_P G\{O, f(O)\}$$

Appendix: Asymptotic distribution of U estimators

For $\hat{f}_n$, using optimality implies that for a path P_ϵ with $P_{\epsilon=0} = P_0$,

$$\frac{\partial}{\partial \epsilon} V(f^*_{P_\epsilon}, P)\Big|_{\epsilon=0} = 0$$

Appendix: Asymptotic distribution of U estimators

For $\hat{f}_n$, using optimality implies that for a path P_ϵ with $P_{\epsilon=0} = P_0$,

$$\frac{\partial}{\partial \epsilon} V(f^*_{P_\epsilon}, P)\bigg|_{\epsilon=0} = 0$$

So we can treat $f^*_{P_0}$ as known; EIF of $V(f^*_{P_0}, P_0) = D^*(P)(o) = D(f^*_P, P)(o)$

Appendix: Asymptotic distribution of U estimators

For $\hat{f}_n$, using optimality implies that for a path P_ϵ with $P_{\epsilon=0} = P_0$,

$$\frac{\partial}{\partial \epsilon} V(f^*_{P_\epsilon}, P)\bigg|_{\epsilon=0} = 0$$

So we can treat $f^*_{P_0}$ as known; EIF of $V(f^*_{P_0}, P_0) = D^*(P)(o) = D(f^*_P, P)(o)$

$$\mathbb{P}_n D^*(P_n) = \frac{1}{n}\sum_{i=1}^n \left[G\{O_i, \hat{f}_n(O_i)\} - \frac{1}{n}\sum_{i=1}^n G\{O_i, \hat{f}_n(O_i)\} \right] = 0$$