Floodgate: inference for model-free variable importance

Dempster's Colloquium 2021

Lu Zhang

Department of Statistics
Harvard University

April 16, 2021

Zhang, L. and Janson, L., 2020. Floodgate: inference for model-free variable importance. arXiv preprint arXiv:2007.01283.

Collaborator

Lucas Janson

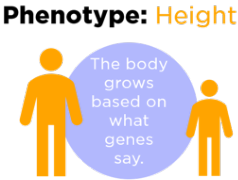
Overview

Problem setup

Setup: data (Y, X, Z) from some joint distribution.

- Y a outcome variable of interest (AKA response or dependent variable),
- X a explanatory variable of interest (AKA treatment, covariate, feature)
- $Z := (Z_1, \cdots, Z_p)$ a set of p further variables (AKA confounders, nuisance variables)

Examples:

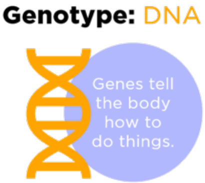


Problem setup

Setup: data (Y, X, Z) from some joint distribution.

- Y a outcome variable of interest (AKA response or dependent variable),
- X a explanatory variable of interest (AKA treatment, covariate, feature)
- $Z := (Z_1, \cdots, Z_p)$ a set of p further variables (AKA confounders, nuisance variables)

Examples:

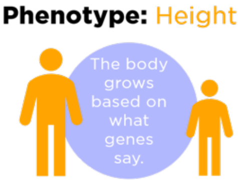


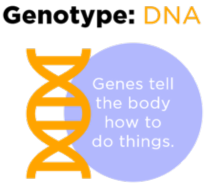
Problem setup

Setup: data (Y, X, Z) from some joint distribution.

- Y a outcome variable of interest (AKA response or dependent variable),
- X a explanatory variable of interest (AKA treatment, covariate, feature)
- $Z := (Z_1, \cdots, Z_p)$ a set of p further variables (AKA confounders, nuisance variables)

Examples:





Problem setup

Setup: data (Y, X, Z) from some joint distribution.

- Y a outcome variable of interest (AKA response or dependent variable),
- X a explanatory variable of interest (AKA treatment, covariate, feature)
- $Z := (Z_1, \cdots, Z_p)$ a set of p further variables (AKA confounders, nuisance variables)

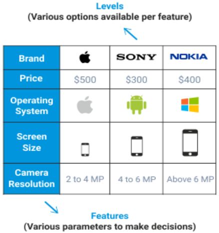
Examples:

Problem setup

Setup: data (Y, X, Z) from some joint distribution.

- Y a outcome variable of interest (AKA response or dependent variable),
- X a explanatory variable of interest (AKA treatment, covariate, feature)
- $Z := (Z_1, \cdots, Z_p)$ a set of p further variables (AKA confounders, nuisance variables)

Examples:

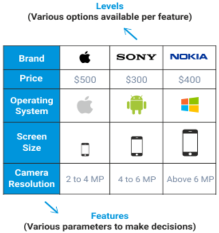


Problem setup

Setup: data (Y, X, Z) from some joint distribution.

- Y a outcome variable of interest (AKA response or dependent variable),
- X a explanatory variable of interest (AKA treatment, covariate, feature)
- $Z := (Z_1, \cdots, Z_p)$ a set of p further variables (AKA confounders, nuisance variables)

Examples:



Motivation

Question 1

Is the variable X important or not?

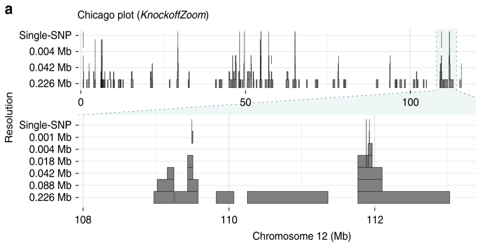


Figure: Select important groups of SNPs

Motivation

Question 1

Is the variable X important or not?

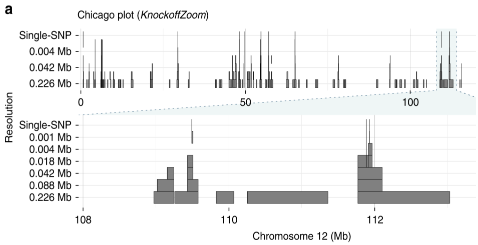


Figure: Select important groups of SNPs

Go Beyond

Motivation

Question 1

Is the variable X important or not?

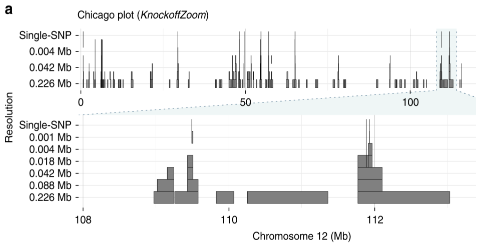


Figure: Select important groups of SNPs

Go Beyond

Question 1*

How important is the variable X?

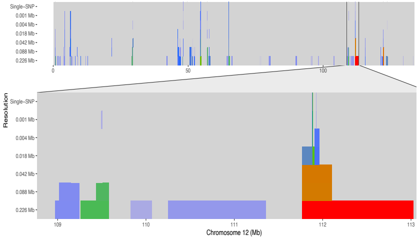


Figure: Infer the importance of a group of SNPs

Motivation

Q1: A desirable measure of variable importance (MOVI) (of the covariate X) should be

Motivation

Q1: A desirable measure of variable importance (MOVI) (of the covariate X) should be

Null-compatible : zero when $Y \perp\!\!\!\perp X \mid Z$.

Motivation

Q1: A desirable measure of variable importance (MOVI) (of the covariate X) should be

Null-compatible : zero when $Y \perp\!\!\!\perp X \mid Z$.

Sensitive : able to detect all interesting types of dependence.

Motivation

Q1: A desirable measure of variable importance (MOVI) (of the covariate X) should be

Null-compatible : zero when $Y \perp\!\!\!\perp X \mid Z$.

Sensitive : able to detect all interesting types of dependence.

Interpretable : simple functional of the data-generating distribution.

Motivation

Q1: A desirable measure of variable importance (MOVI) (of the covariate X) should be

Null-compatible : zero when $Y \perp\!\!\!\perp X \mid Z$.

Sensitive : able to detect all interesting types of dependence.

Interpretable : simple functional of the data-generating distribution.

Q2: A desirable inferential procedure for the MOVI should be

Motivation

Q1: A desirable measure of variable importance (MOVI) (of the covariate X) should be

Null-compatible : zero when $Y \perp\!\!\!\perp X \mid Z$.

Sensitive : able to detect all interesting types of dependence.

Interpretable : simple functional of the data-generating distribution.

Q2: A desirable inferential procedure for the MOVI should be

Motivation

Q1: A desirable measure of variable importance (MOVI) (of the covariate X) should be

- Null-compatible : zero when $Y \perp\!\!\!\perp X \mid Z$.
- Sensitive : able to detect all interesting types of dependence.
- Interpretable : simple functional of the data-generating distribution.

Q2: A desirable inferential procedure for the MOVI should be

Valid

Motivation

Q1: A desirable measure of variable importance (MOVI) (of the covariate X) should be

Null-compatible : zero when $Y \perp\!\!\!\perp X \mid Z$.

Sensitive : able to detect all interesting types of dependence.

Interpretable : simple functional of the data-generating distribution.

Q2: A desirable inferential procedure for the MOVI should be

Valid General

Motivation

Q1: A desirable measure of variable importance (MOVI) (of the covariate X) should be

Null-compatible : zero when $Y \perp\!\!\!\perp X \mid Z$.

Sensitive : able to detect all interesting types of dependence.

Interpretable : simple functional of the data-generating distribution.

Q2: A desirable inferential procedure for the MOVI should be

Valid General Accurate

Motivation

Q1: A desirable measure of variable importance (MOVI) (of the covariate X) should be

Null-compatible : zero when $Y \perp\!\!\!\perp X \mid Z$.

Sensitive : able to detect all interesting types of dependence.

Interpretable : simple functional of the data-generating distribution.

Q2: A desirable inferential procedure for the MOVI should be

Valid General Accurate Robust

Related work

- *Parametric approaches:* Bühlmann et al. (2013), Zhang and Zhang (2014), Javanmard and Montanari (2014), Bühlmann et al. (2015), Dezeure et al. (2017), Zhang and Cheng (2017), Van de Geer et al. (2014), Nickl et al. (2013).
- *Projection approaches:* Buja et al. (2015, 2019a,b), Rinaldo et al. (2019), Lee et al. (2016), Taylor et al. (2014), Berk et al. (2013), Buja and Brown (2014).
- *Semi-parametric approaches:* $\mathbb{E}\left[\mathrm{Cov}\left(Y, X \mid Z\right)\right]$; Robins et al. (2008, 2009); Li et al. (2011); Robins et al. (2017); Newey and Robins (2018), Shah and Peters (2018).
- *Random estimands:* Lei et al. (2018), Watson and Wright (2019), Rinaldo et al. (2019).
- *Same MOVI as us:* Saltelli et al. (2008), Williamson et al. (2017, 2020).

Related work

- *Parametric approaches:* Bühlmann et al. (2013), Zhang and Zhang (2014), Javanmard and Montanari (2014), Bühlmann et al. (2015), Dezeure et al. (2017), Zhang and Cheng (2017), Van de Geer et al. (2014), Nickl et al. (2013).
- *Projection approaches:* Buja et al. (2015, 2019a,b), Rinaldo et al. (2019), Lee et al. (2016), Taylor et al. (2014), Berk et al. (2013), Buja and Brown (2014).
- *Semi-parametric approaches:* $\mathbb{E}\left[\mathrm{Cov}\left(Y, X \mid Z\right)\right]$; Robins et al. (2008, 2009); Li et al. (2011); Robins et al. (2017); Newey and Robins (2018), Shah and Peters (2018).
- *Random estimands:* Lei et al. (2018), Watson and Wright (2019), Rinaldo et al. (2019).
- *Same MOVI as us:* Saltelli et al. (2008), Williamson et al. (2017, 2020).

Related work

- *Parametric approaches:* Bühlmann et al. (2013), Zhang and Zhang (2014), Javanmard and Montanari (2014), Bühlmann et al. (2015), Dezeure et al. (2017), Zhang and Cheng (2017), Van de Geer et al. (2014), Nickl et al. (2013).
- *Projection approaches:* Buja et al. (2015, 2019a,b), Rinaldo et al. (2019), Lee et al. (2016), Taylor et al. (2014), Berk et al. (2013), Buja and Brown (2014).
- *Semi-parametric approaches:* $\mathbb{E}\left[\mathrm{Cov}\left(Y, X \mid Z\right)\right]$; Robins et al. (2008, 2009); Li et al. (2011); Robins et al. (2017); Newey and Robins (2018), Shah and Peters (2018).
- *Random estimands:* Lei et al. (2018), Watson and Wright (2019), Rinaldo et al. (2019).
- *Same MOVI as us:* Saltelli et al. (2008), Williamson et al. (2017, 2020).

Related work

- *Parametric approaches:* Bühlmann et al. (2013), Zhang and Zhang (2014), Javanmard and Montanari (2014), Bühlmann et al. (2015), Dezeure et al. (2017), Zhang and Cheng (2017), Van de Geer et al. (2014), Nickl et al. (2013).
- *Projection approaches:* Buja et al. (2015, 2019a,b), Rinaldo et al. (2019), Lee et al. (2016), Taylor et al. (2014), Berk et al. (2013), Buja and Brown (2014).
- *Semi-parametric approaches:* $\mathbb{E}\left[\mathrm{Cov}\left(Y, X \mid Z\right)\right]$; Robins et al. (2008, 2009); Li et al. (2011); Robins et al. (2017); Newey and Robins (2018), Shah and Peters (2018).
- *Random estimands:* Lei et al. (2018), Watson and Wright (2019), Rinaldo et al. (2019).
- *Same MOVI as us:* Saltelli et al. (2008), Williamson et al. (2017, 2020).

Related work

- *Parametric approaches:* Bühlmann et al. (2013), Zhang and Zhang (2014), Javanmard and Montanari (2014), Bühlmann et al. (2015), Dezeure et al. (2017), Zhang and Cheng (2017), Van de Geer et al. (2014), Nickl et al. (2013).
- *Projection approaches:* Buja et al. (2015, 2019a,b), Rinaldo et al. (2019), Lee et al. (2016), Taylor et al. (2014), Berk et al. (2013), Buja and Brown (2014).
- *Semi-parametric approaches:* $\mathbb{E}\left[\mathrm{Cov}\left(Y, X \mid Z\right)\right]$; Robins et al. (2008, 2009); Li et al. (2011); Robins et al. (2017); Newey and Robins (2018), Shah and Peters (2018).
- *Random estimands:* Lei et al. (2018), Watson and Wright (2019), Rinaldo et al. (2019).
- *Same MOVI as us:* Saltelli et al. (2008), Williamson et al. (2017, 2020).

Preview of our results

A1 : present a MOVI, the mMSE gap.

- ✓ Zero under the conditional independence $Y \perp\!\!\!\perp X \mid Z$.
- ✓ Strictly positive unless $\mathbb{E}[Y \mid X, Z]$ has no dependence on X at all.
- ✓ Direct predictive, causal and explanatory interpretations.

A2 : propose a method for inference for it: **floodgate**.

- ✓ Asymptotically-valid inference.
- ✓ Does not make any parametric/smoothness/sparsity assumptions about $Y \mid X, Z$ and built around any regression estimator.
- ✓ Width of confidence bounds proportional to the predictive performance.
- ✓ Assume $P_{X|Z}$ known, quantified robustness to misspecification and extension allowing known up to a parametric model.

Preview of our results

A1 : present a MOVI, the mMSE gap.

Null compatible ☑ Zero under the conditional independence $Y \perp\!\!\!\perp X \mid Z$.

Sensitive ☑ Strictly positive unless $\mathbb{E}[Y \mid X, Z]$ has no dependence on X at all.

Interpretable ☑ Direct predictive, causal and explanatory interpretations.

A2 : propose a method for inference for it: **floodgate**.

Valid ☑ Asymptotically-valid inference.

General ☑ Does not make any parametric/smoothness/sparsity assumptions about $Y \mid X, Z$ and built around any regression estimator.

Accurate ☑ Width of confidence bounds proportional to the predictive performance.

Robust ☑ Assume $P_{X|Z}$ known; quantified robustness to misspecification and extension allowing known up to a parametric model.

Preview of our results

A1 : present a MOVI, the mMSE gap.

Null compatible ☑ Zero under the conditional independence $Y \perp\!\!\!\perp X \mid Z$.

Sensitive ☑ Strictly positive unless $\mathbb{E}[Y \mid X, Z]$ has no dependence on X at all.

Interpretable ☑ Direct predictive, causal and explanatory interpretations.

A2 : propose a method for inference for it: **floodgate**.

Valid ☑ Asymptotically-valid inference.

General ☑ Does not make any parametric/smoothness/sparsity assumptions about $Y \mid X, Z$ and built around any regression estimator.

Accurate ☑ Width of confidence bounds proportional to the predictive performance.

Robust ☑ Assume $P_{X|Z}$ known; quantified robustness to misspecification and extension allowing known up to a parametric model.

Preview of our results

A1 : present a MOVI, the mMSE gap.

Null compatible ☑ Zero under the conditional independence $Y \perp\!\!\!\perp X \mid Z$.

Sensitive ☑ Strictly positive unless $\mathbb{E}\left[Y \mid X, Z\right]$ has no dependence on X at all.

Interpretable ☑ Direct predictive, causal and explanatory interpretations.

A2 : propose a method for inference for it: **floodgate**.

Valid ☑ Asymptotically-valid inference.

General ☑ Does not make any parametric/smoothness/sparsity assumptions about $Y \mid X, Z$ and built around any regression estimator.

Accurate ☑ Width of confidence bounds proportional to the predictive performance.

Robust ☑ Assume $P_{X|Z}$ known; quantified robustness to misspecification and extension allowing known up to a parametric model.

Preview of our results

A1 : present a MOVI, the mMSE gap.

Null compatible ☑ Zero under the conditional independence $Y \perp\!\!\!\perp X \mid Z$.

Sensitive ☑ Strictly positive unless $\mathbb{E}[Y \mid X, Z]$ has no dependence on X at all.

Interpretable ☑ Direct predictive, causal and explanatory interpretations.

A2 : propose a method for inference for it: **floodgate**.

Valid ☑ Asymptotically-valid inference.

General ☑ Does not make any parametric/smoothness/sparsity assumptions about $Y \mid X, Z$ and built around any regression estimator.

Accurate ☑ Width of confidence bounds proportional to the predictive performance.

Robust ☑ Assume $P_{X|Z}$ known; quantified robustness to misspecification and extension allowing known up to a parametric model.

Preview of our results

A1 : present a MOVI, the mMSE gap.

Null compatible ☑ Zero under the conditional independence $Y \perp\!\!\!\perp X \mid Z$.

Sensitive ☑ Strictly positive unless $\mathbb{E}[Y \mid X, Z]$ has no dependence on X at all.

Interpretable ☑ Direct predictive, causal and explanatory interpretations.

A2 : propose a method for inference for it: **floodgate**.

Valid ☑ Asymptotically-valid inference.

General ☑ Does not make any parametric/smoothness/sparsity assumptions about $Y \mid X, Z$ and built around any regression estimator.

Accurate ☑ Width of confidence bounds proportional to the predictive performance.

Robust ☑ Assume $P_{X|Z}$ known; quantified robustness to misspecification and extension allowing known up to a parametric model.

Preview of our results

A1 : present a MOVI, the mMSE gap.

Null compatible ☑ Zero under the conditional independence $Y \perp\!\!\!\perp X \mid Z$.

Sensitive ☑ Strictly positive unless $\mathbb{E}[Y \mid X, Z]$ has no dependence on X at all.

Interpretable ☑ Direct predictive, causal and explanatory interpretations.

A2 : propose a method for inference for it: **floodgate**.

Valid ☑ Asymptotically-valid inference.

General ☑ Does not make any parametric/smoothness/sparsity assumptions about $Y \mid X, Z$ and built around any regression estimator.

Accurate ☑ Width of confidence bounds proportional to the predictive performance.

Robust ☑ Assume $P_{X|Z}$ known; quantified robustness to misspecification and extension allowing known up to a parametric model.

Preview of our results

A1 : present a MOVI, the mMSE gap.

Null compatible ☑ Zero under the conditional independence $Y \perp\!\!\!\perp X \mid Z$.

Sensitive ☑ Strictly positive unless $\mathbb{E}[Y \mid X, Z]$ has no dependence on X at all.

Interpretable ☑ Direct predictive, causal and explanatory interpretations.

A2 : propose a method for inference for it: **floodgate**.

Valid ☑ Asymptotically-valid inference.

General ☑ Does not make any parametric/smoothness/sparsity assumptions about $Y \mid X, Z$ and built around any regression estimator.

Accurate ☑ Width of confidence bounds proportional to the predictive performance.

Robust ☑ Assume $P_{X|Z}$ known; quantified robustness to misspecification and extension allowing known up to a parametric model.

Preview of our results

A1 : present a MOVI, the mMSE gap.

Null compatible ☑ Zero under the conditional independence $Y \perp\!\!\!\perp X \mid Z$.

Sensitive ☑ Strictly positive unless $\mathbb{E}[Y \mid X, Z]$ has no dependence on X at all.

Interpretable ☑ Direct predictive, causal and explanatory interpretations.

A2 : propose a method for inference for it: **floodgate**.

Valid ☑ Asymptotically-valid inference.

General ☑ Does not make any parametric/smoothness/sparsity assumptions about $Y \mid X, Z$ and built around any regression estimator.

Accurate ☑ Width of confidence bounds proportional to the predictive performance.

Robust ☑ Assume $P_{X|Z}$ known; quantified robustness to misspecification and extension allowing known up to a parametric model.

Preview of our results

A1 : present a MOVI, the mMSE gap.

Null compatible ☑ Zero under the conditional independence $Y \perp\!\!\!\perp X \mid Z$.

Sensitive ☑ Strictly positive unless $\mathbb{E}[Y \mid X, Z]$ has no dependence on X at all.

Interpretable ☑ Direct predictive, causal and explanatory interpretations.

A2 : propose a method for inference for it: **floodgate**.

Valid ☑ Asymptotically-valid inference.

General ☑ Does not make any parametric/smoothness/sparsity assumptions about $Y \mid X, Z$ and built around any regression estimator.

Accurate ☑ Width of confidence bounds proportional to the predictive performance.

Robust ☑ Assume $P_{X|Z}$ known; quantified robustness to misspecification and extension allowing known up to a parametric model.

Our target MOVI: the mMSE gap

Definition (mMSE Gap)

The minimum mean squared error (mMSE) gap for variable X is defined as

$$\mathcal{I}^2 = \mathbb{E}\left[(Y - \mathbb{E}[Y \mid Z])^2\right] - \mathbb{E}\left[(Y - \mathbb{E}[Y \mid X, Z])^2\right].$$

Our target MOVI: the mMSE gap

Definition (mMSE Gap)

The minimum mean squared error (mMSE) gap for variable X is defined as

$$\mathcal{I}^2 = \mathbb{E}\left[(Y - \mathbb{E}[Y \mid Z])^2\right] - \mathbb{E}\left[(Y - \mathbb{E}[Y \mid X, Z])^2\right].$$

$$\mathcal{I}^2 = 0 \iff \mathbb{E}[Y \mid X, Z] \overset{a.s.}{=} \mathbb{E}[Y \mid Z]$$

Our target MOVI: the mMSE gap

Definition (mMSE Gap)

The minimum mean squared error (mMSE) gap for variable X is defined as

$$\mathcal{I}^2 = \mathbb{E}\left[(Y - \mathbb{E}[Y \mid Z])^2\right] - \mathbb{E}\left[(Y - \mathbb{E}[Y \mid X, Z])^2\right].$$

$$\mathcal{I}^2 = 0 \iff \mathbb{E}[Y \mid X, Z] \overset{a.s.}{=} \mathbb{E}[Y \mid Z]$$

- **Predictive**: immediate from above.
- **Variance decomposition**: $\mathcal{I}^2 = \mathrm{Var}(\mathbb{E}[Y \mid X, Z]) - \mathrm{Var}(\mathbb{E}[Y \mid Z])$.
- **Causal**: $\mathcal{I}^2 = \frac{1}{2}\mathbb{E}_Z\left[\mathbb{E}_{x_1, x_2 \overset{i.i.d.}{\sim} P_{X|Z}}\left[(\mathbb{E}[Y \mid X = x_1, Z] - \mathbb{E}[Y \mid X = x_2, Z])^2\right]\right]$.
- Compact form: $\mathcal{I}^2 = \mathbb{E}[\mathrm{Var}(\mathbb{E}[Y \mid X, Z] \mid Z)]$.

Our target MOVI: the mMSE gap

Definition (mMSE Gap)

The minimum mean squared error (mMSE) gap for variable X is defined as

$$\mathcal{I}^2 = \mathbb{E}\left[(Y - \mathbb{E}[Y \mid Z])^2\right] - \mathbb{E}\left[(Y - \mathbb{E}[Y \mid X, Z])^2\right].$$

$$\mathcal{I}^2 = 0 \iff \mathbb{E}[Y \mid X, Z] \overset{a.s.}{=} \mathbb{E}[Y \mid Z]$$

- **Predictive**: immediate from above.
- **Variance decomposition**: $\mathcal{I}^2 = \mathrm{Var}(\mathbb{E}[Y \mid X, Z]) - \mathrm{Var}(\mathbb{E}[Y \mid Z])$.
- **Causal**: $\mathcal{I}^2 = \frac{1}{2}\mathbb{E}_Z\left[\mathbb{E}_{x_1, x_2 \overset{i.i.d.}{\sim} P_{X|Z}}\left[(\mathbb{E}[Y \mid X = x_1, Z] - \mathbb{E}[Y \mid X = x_2, Z])^2\right]\right]$.
- Compact form: $\mathcal{I}^2 = \mathbb{E}[\mathrm{Var}(\mathbb{E}[Y \mid X, Z] \mid Z)]$.

Our target MOVI: the mMSE gap

Definition (mMSE Gap)

The minimum mean squared error (mMSE) gap for variable X is defined as

$$\mathcal{I}^2 = \mathbb{E}\left[(Y - \mathbb{E}[Y \mid Z])^2\right] - \mathbb{E}\left[(Y - \mathbb{E}[Y \mid X, Z])^2\right].$$

$$\mathcal{I}^2 = 0 \iff \mathbb{E}[Y \mid X, Z] \overset{a.s.}{=} \mathbb{E}[Y \mid Z]$$

- **Predictive**: immediate from above.
- **Variance decomposition**: $\mathcal{I}^2 = \mathrm{Var}(\mathbb{E}[Y \mid X, Z]) - \mathrm{Var}(\mathbb{E}[Y \mid Z])$.
- **Causal**: $\mathcal{I}^2 = \frac{1}{2}\mathbb{E}_Z\left[\mathbb{E}_{x_1, x_2 \overset{i.i.d.}{\sim} P_{X|Z}}\left[(\mathbb{E}[Y \mid X = x_1, Z] - \mathbb{E}[Y \mid X = x_2, Z])^2\right]\right]$.
- Compact form: $\mathcal{I}^2 = \mathbb{E}[\mathrm{Var}(\mathbb{E}[Y \mid X, Z] \mid Z)]$.

Our target MOVI: the mMSE gap

Definition (mMSE Gap)

The minimum mean squared error (mMSE) gap for variable X is defined as

$$\mathcal{I}^2 = \mathbb{E}\left[(Y - \mathbb{E}[Y \mid Z])^2\right] - \mathbb{E}\left[(Y - \mathbb{E}[Y \mid X, Z])^2\right].$$

$$\mathcal{I}^2 = 0 \iff \mathbb{E}[Y \mid X, Z] \overset{a.s.}{=} \mathbb{E}[Y \mid Z]$$

- **Predictive**: immediate from above.
- **Variance decomposition**: $\mathcal{I}^2 = \mathrm{Var}(\mathbb{E}[Y \mid X, Z]) - \mathrm{Var}(\mathbb{E}[Y \mid Z])$.
- **Causal**: $\mathcal{I}^2 = \frac{1}{2}\mathbb{E}_Z\left[\mathbb{E}_{x_1, x_2 \overset{i.i.d.}{\sim} P_{X|Z}}\left[(\mathbb{E}[Y \mid X = x_1, Z] - \mathbb{E}[Y \mid X = x_2, Z])^2\right]\right]$.
- Compact form: $\mathcal{I}^2 = \mathbb{E}[\mathrm{Var}(\mathbb{E}[Y \mid X, Z] \mid Z)]$.

How to do inference on $\mathcal{I}$?

True regression function $\mu^\star(x, z) := \mathbb{E}[Y \mid X = x, Z = z]$

How to do inference on $\mathcal{I}$?

True regression function $\mu^\star(x, z) := \mathbb{E}\left[Y \mid X = x, Z = z\right]$

$$\Rightarrow \quad \mathcal{I}^2 = \mathbb{E}\left[\mathrm{Var}(\mu^\star(X, Z) \mid Z)\right] = \mathbb{E}\left[(\mu^\star(X, Z) - \mathbb{E}\left[\mu^\star(X, Z) \mid Z\right])^2\right]$$

Challenges:

- $\mu^\star$ unknown.
- Nonlinearity in the above functional.

How to do inference on $\mathcal{I}$?

True regression function $\mu^\star(x,z) := \mathbb{E}\left[Y \mid X = x, Z = z\right]$

$$\Rightarrow \quad \mathcal{I}^2 = \mathbb{E}\left[\mathrm{Var}(\mu^\star(X,Z) \mid Z)\right] = \mathbb{E}\left[\left(\mu^\star(X,Z) - \mathbb{E}\left[\mu^\star(X,Z) \mid Z\right]\right)^2\right]$$

Challenges:

- $\mu^\star$ unknown.
- Nonlinearity in the above functional.

Possible solution:

How to do inference on $\mathcal{I}$?

True regression function $\mu^\star(x,z) := \mathbb{E}\left[Y \mid X = x, Z = z\right]$

$$\Rightarrow \quad \mathcal{I}^2 = \mathbb{E}\left[\mathrm{Var}(\mu^\star(X,Z) \mid Z)\right] = \mathbb{E}\left[\left(\mu^\star(X,Z) - \mathbb{E}\left[\mu^\star(X,Z) \mid Z\right]\right)^2\right]$$

Challenges:

- $\mu^\star$ unknown.
- Nonlinearity in the above functional.

Possible solution: assume we have a good estimator μ of $\mu^\star$?

How to do inference on $\mathcal{I}$?

True regression function $\mu^\star(x,z) := \mathbb{E}\left[Y \mid X = x, Z = z\right]$

$$\Rightarrow \quad \mathcal{I}^2 = \mathbb{E}\left[\mathrm{Var}(\mu^\star(X,Z) \mid Z)\right] = \mathbb{E}\left[\left(\mu^\star(X,Z) - \mathbb{E}\left[\mu^\star(X,Z) \mid Z\right]\right)^2\right]$$

Challenges:

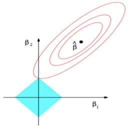
- $\mu^\star$ unknown.
- Nonlinearity in the above functional.

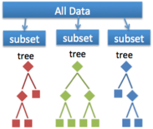
Possible solution: ~~assume we have a good estimator μ of $\mu^\star$?~~

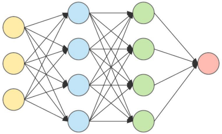
How to do inference on $\mathcal{I}$?

Possible solution: ~~assume we have a good estimator μ of $\mu^\star$?~~

- Only known for limited class of estimators and data-generating distributions.
- Precludes most modern machine learning algorithms and methods that integrate hard-to-quantify domain knowledge.







Floodgate

Our approach: construct a lower confidence bound (LCB) for $\mathcal{I}$ via floodgate, i.e.

Floodgate

Our approach: construct a lower confidence bound (LCB) for $\mathcal{I}$ via floodgate, i.e.

P1 construct a functional $\boldsymbol{f}$ such that

$$\boldsymbol{f}(\mu) \leq \mathcal{I} \quad \text{for any } \mu.$$

Floodgate

Our approach: construct a lower confidence bound (LCB) for $\mathcal{I}$ via floodgate, i.e.

P1 construct a functional $\boldsymbol{f}$ such that

$$\boldsymbol{f}(\mu) \leq \mathcal{I} \quad \text{for any } \mu.$$

P2 know how to obtain LCB $L(\mu)$ of $\boldsymbol{f}(\mu)$ for any μ.

Floodgate

Our approach: construct a lower confidence bound (LCB) for $\mathcal{I}$ via floodgate, i.e.

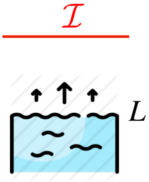
P1 construct a functional f such that

$$f(\mu) \le \mathcal{I} \quad \text{for any } \mu.$$

P2 know how to obtain LCB $L(\mu)$ of $f(\mu)$ for any μ.

P3 (Ideally) the functional f also satisfies $f(\mu^\star) = \mathcal{I}$.

Floodgate



Our approach: construct a lower confidence bound (LCB) for $\mathcal{I}$ via floodgate, i.e.

P1 construct a functional f such that

$$f(\mu) \leq \mathcal{I} \quad \text{for any } \mu.$$

P2 know how to obtain LCB $L(\mu)$ of $f(\mu)$ for any μ.

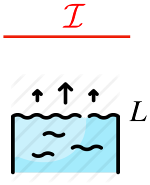
P3 (Ideally) the functional f also satisfies $f(\mu^\star) = \mathcal{I}$.

Floodgate

Our approach: construct a lower confidence bound (LCB) for $\mathcal{I}$ via floodgate, i.e.

P1 construct a functional f such that

$$f(\mu) \leq \mathcal{I} \quad \text{for any } \mu.$$

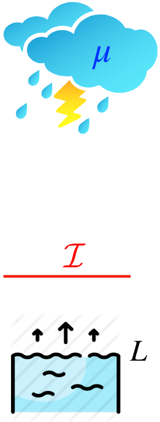


P2 know how to obtain LCB $L(\mu)$ of $f(\mu)$ for any μ.

P3 (Ideally) the functional f also satisfies $f(\mu^\star) = \mathcal{I}$.

Floodgate

Floodgate



Our approach: construct a lower confidence bound (LCB) for $\mathcal{I}$ via floodgate, i.e.

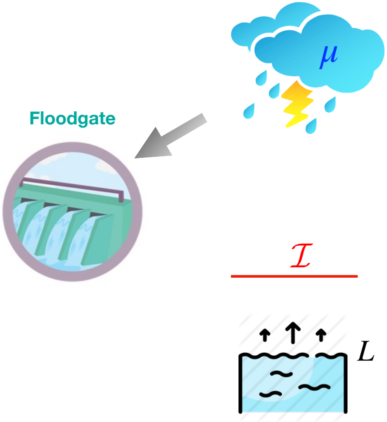
P1 construct a functional f such that

$$f(\mu) \leq \mathcal{I} \quad \text{for any } \mu.$$

P2 know how to obtain LCB $L(\mu)$ of $f(\mu)$ for any μ.

P3 (Ideally) the functional f also satisfies $f(\mu^{\star}) = \mathcal{I}$.

Floodgate



Our approach: construct a lower confidence bound (LCB) for $\mathcal{I}$ via floodgate, i.e.

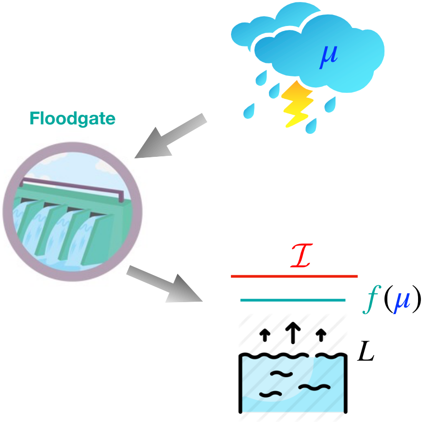
P1 construct a functional f such that

$$f(\mu) \leq \mathcal{I} \quad \text{for any } \mu.$$

P2 know how to obtain LCB $L(\mu)$ of $f(\mu)$ for any μ.

P3 (Ideally) the functional f also satisfies $f(\mu^\star) = \mathcal{I}$.

Floodgate



Our approach: construct a lower confidence bound (LCB) for $\mathcal{I}$ via floodgate, i.e.

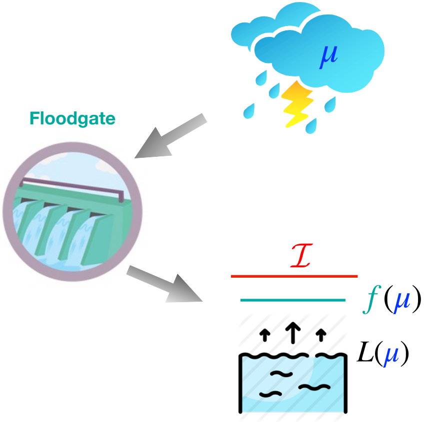
P1 construct a functional f such that

$$f(\mu) \leq \mathcal{I} \quad \text{for any } \mu.$$

P2 know how to obtain LCB $L(\mu)$ of $f(\mu)$ for any μ.

P3 (Ideally) the functional f also satisfies $f(\mu^\star) = \mathcal{I}$.

Floodgate



Our approach: construct a lower confidence bound (LCB) for $\mathcal{I}$ via floodgate, i.e.

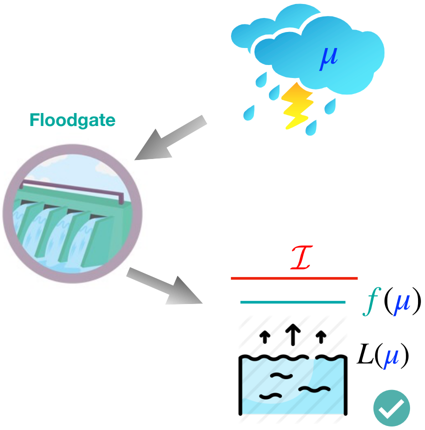
P1 construct a functional f such that

$$f(\mu) \leq \mathcal{I} \quad \text{for any } \mu.$$

P2 know how to obtain LCB $L(\mu)$ of $f(\mu)$ for any μ.

P3 (Ideally) the functional f also satisfies $f(\mu^{\star}) = \mathcal{I}$.

Floodgate



Our approach: construct a lower confidence bound (LCB) for $\mathcal{I}$ via floodgate, i.e.

P1 construct a functional $\boldsymbol{f}$ such that

$$\boldsymbol{f}(\mu) \leq \mathcal{I} \quad \text{for any } \mu.$$

P2 know how to obtain LCB $L(\mu)$ of $\boldsymbol{f}(\mu)$ for any μ.

P3 (Ideally) the functional $\boldsymbol{f}$ also satisfies $\boldsymbol{f}(\mu^\star) = \mathcal{I}$.

Our choice of Floodgate functional

- Our choice:

$$f(\mu) := \frac{\mathbb{E}\left[\mathrm{Cov}(\mu^\star(X,Z), \mu(X,Z) \mid Z)\right]}{\sqrt{\mathbb{E}\left[\mathrm{Var}(\mu(X,Z) \mid Z)\right]}}$$

P1

P2

P3

Our choice of Floodgate functional

- Our choice:

$$f(\mu) := \frac{\mathbb{E}\left[\text{Cov}(\mu^\star(X,Z), \mu(X,Z) \mid Z)\right]}{\sqrt{\mathbb{E}\left[\text{Var}(\mu(X,Z) \mid Z)\right]}}$$

P1 ☑

P2

P3 ☑

Lemma (Zhang and Janson (2020))

For any μ such that $f(\mu)$ exists, we have $f(\mu) \leq \mathcal{I}$ and $f(\mu^\star) = \mathcal{I}$.

Our choice of Floodgate functional

- Our choice:

$$f(\mu) := \frac{\mathbb{E}\left[\mathrm{Cov}(\mu^\star(X,Z), \mu(X,Z) \mid Z)\right]}{\sqrt{\mathbb{E}\left[\mathrm{Var}(\mu(X,Z) \mid Z)\right]}}$$

P1 ☑

P2 ☑

P3 ☑

Lemma (Zhang and Janson (2020))

For any μ such that $f(\mu)$ exists, we have $f(\mu) \le \mathcal{I}$ and $f(\mu^\star) = \mathcal{I}$.

- How to obtain LCB $L(\mu)$ of $f(\mu)$ for any μ?

$$f(\mu) = \frac{\mathbb{E}\left[Y\left(\mu(X,Z) - \mathbb{E}\left[\mu(X,Z) \mid Z\right]\right)\right]}{\sqrt{\mathbb{E}\left[\mathrm{Var}\left(\mu(X,Z) \mid Z\right)\right]}} = \frac{\text{a linear functional of } P_{(Y,X,Z)}}{\sqrt{\text{a linear functional of } P_Z}}$$

Inferential procedures

Input: $\mathcal{D} = \{(Y_i, X_i, Z_i)\}_{i=1}^n$; $\mathcal{D}'$; any regression algorithm $\mathcal{A}$; assume $P_{X|Z}$ known.

Inferential procedures

Input: $\mathcal{D} = \{(Y_i, X_i, Z_i)\}_{i=1}^n$; $\mathcal{D}'$; any regression algorithm $\mathcal{A}$; assume $P_{X|Z}$ known.

1. Obtain $\mu = \mathcal{A}(\mathcal{D}')$ from the separate dataset $\mathcal{D}'$.

Inferential procedures

Input: $\mathcal{D} = \{(Y_i, X_i, Z_i)\}_{i=1}^n$; $\mathcal{D}'$; any regression algorithm $\mathcal{A}$; assume $P_{X|Z}$ known.

1. Obtain $\mu = \mathcal{A}(\mathcal{D}')$ from the separate dataset $\mathcal{D}'$.
2. Compute $\mathbb{E}\left[\mu(X, Z) \mid Z\right], \mathrm{Var}\left(\mu(X, Z) \mid Z\right)$.

Inferential procedures

Input: $\mathcal{D} = \{(Y_i, X_i, Z_i)\}_{i=1}^n$; $\mathcal{D}'$; any regression algorithm $\mathcal{A}$; assume $P_{X|Z}$ known.

1. Obtain $\mu = \mathcal{A}(\mathcal{D}')$ from the separate dataset $\mathcal{D}'$.
2. Compute $\mathbb{E}\left[\mu(X,Z) \mid Z\right], \mathrm{Var}\left(\mu(X,Z) \mid Z\right)$.

Inferential procedures

Input: $\mathcal{D} = \{(Y_i, X_i, Z_i)\}_{i=1}^n$; $\mathcal{D}'$; any regression algorithm $\mathcal{A}$; assume $P_{X|Z}$ known.

1. Obtain $\mu = \mathcal{A}(\mathcal{D}')$ from the separate dataset $\mathcal{D}'$.
2. Compute $\mathbb{E}\left[\mu(X, Z) \mid Z\right], \mathrm{Var}\left(\mu(X, Z) \mid Z\right)$.
3. Construct CLT-based LCB for $f(\mu)$: $L_n^\alpha(\mu)$ (with confidence level α) by Delta method.

$$\begin{pmatrix} \frac{1}{n}\sum_{i=1}^n Y_i\big(\mu(X_i, Z_i) - \mathbb{E}\left[\mu(X_i, Z_i) \mid Z_i\right]\big) \\ \frac{1}{n}\sum_{i=1}^n \mathrm{Var}\left(\mu(X_i, Z_i) \mid Z_i\right) \end{pmatrix} \xrightarrow{\text{asympt.}} \mathcal{N}\begin{pmatrix} \mathbb{E}\left[\mathrm{Cov}(\mu^\star(X, Z), \mu(X, Z) \mid Z)\right] \\ \mathbb{E}\left[\mathrm{Var}(\mu(X, Z) \mid Z)\right] \end{pmatrix}$$

Asymptotic validity

Theorem (Zhang and Janson (2020); informal)

Under mild moment conditions on Y and $\mu(X, Z)$, we have

$$\mathbb{P}\left(L_n^{\alpha}(\mu) \leq \mathcal{I}\right) \geq 1 - \alpha - O(n^{-1/2}).$$

- Point-wise result: the convergence rate result builds on recent Berry–Esseen type bounds for Delta method (Pinelis et al., 2016).

Asymptotic validity

Theorem (Zhang and Janson (2020); informal)

Under mild moment conditions on Y and $\mu(X, Z)$, we have

$$\mathbb{P}\left(L_n^\alpha(\mu) \leq \mathcal{I}\right) \geq 1 - \alpha - O(n^{-1/2}).$$

- Constant in $O(n^{-1/2})$ has complicated dependence on μ and $P_{(Y,X,Z)}$.

Asymptotic validity

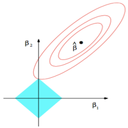
Theorem (Zhang and Janson (2020); informal)

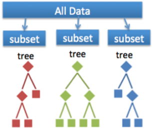
Under mild moment conditions on Y and $\mu(X, Z)$, we have

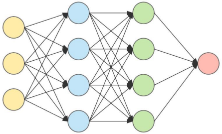
$$\mathbb{P}\left(L_n^{\alpha}(\mu) \leq \mathcal{I}\right) \geq 1 - \alpha - O(n^{-1/2}).$$

- Invariance of the floodgate procedure: e.g., $\mu(x, z) = ax + g(z)$, constant only depends on sign(a) and bivariate distribution of $\left(Y, \frac{X - \mathbb{E}[X \mid Z]}{\sqrt{\mathrm{Var}(X - \mathbb{E}[X \mid Z])}}\right)$.

Computation

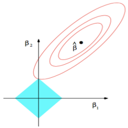


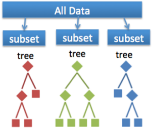


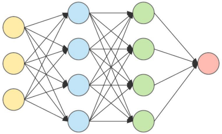


- Allow the user to choose machine learning algorithms or integrate domain knowledge.

Computation

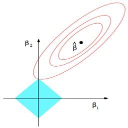


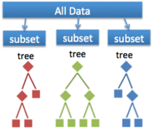


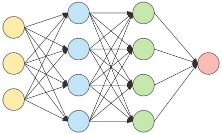


- Allow the user to choose machine learning algorithms or integrate domain knowledge.
- Only involve one time of model fitting.

Computation







- Allow the user to choose machine learning algorithms or integrate domain knowledge.
- Only involve one time of model fitting.
- Under certain fitted models, can compute $\mathbb{E}\left[\mu(X,Z) \mid Z\right], \mathrm{Var}\left(\mu(X,Z) \mid Z\right)$ analytically, e.g., partial linear model with Gaussian design.

General computation

How to compute $\mathbb{E}\left[\mu(X,Z) \mid Z\right], \mathrm{Var}\left(\mu(X,Z) \mid Z\right)$ in a general way (e.g., μ is fitted based on random forest or neural networks)?

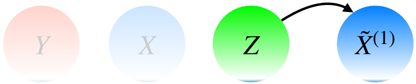
General computation

How to compute $\mathbb{E}\left[\mu(X, Z) \mid Z\right], \mathrm{Var}\left(\mu(X, Z) \mid Z\right)$ in a general way (e.g., μ is fitted based on random forest or neural networks)?

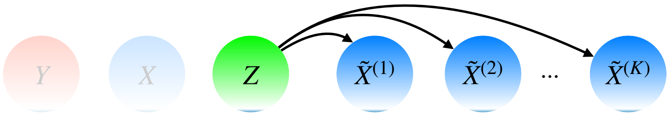
- Sample $\tilde{X}$ from $P_{X|Z}$, conditionally independently of X, Y.

General computation

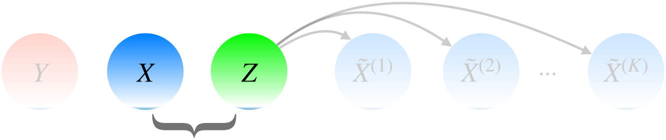
General computation

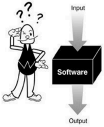


General computation



General computation

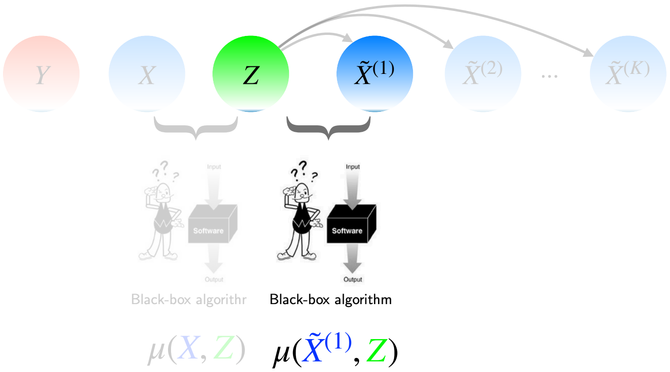




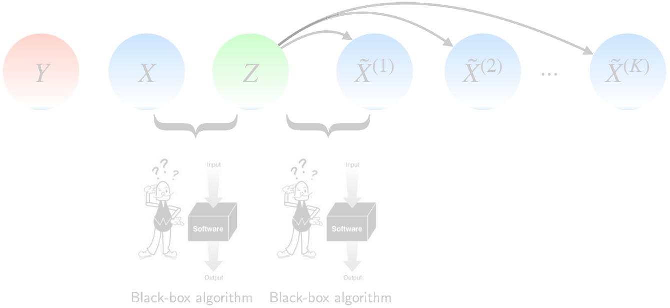
Black-box algorithm

$$\mu(X, Z)$$

General computation

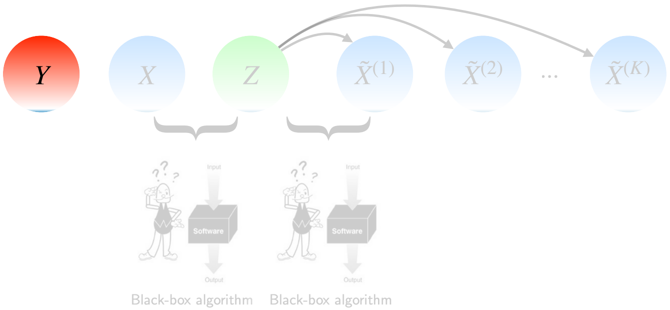


General computation



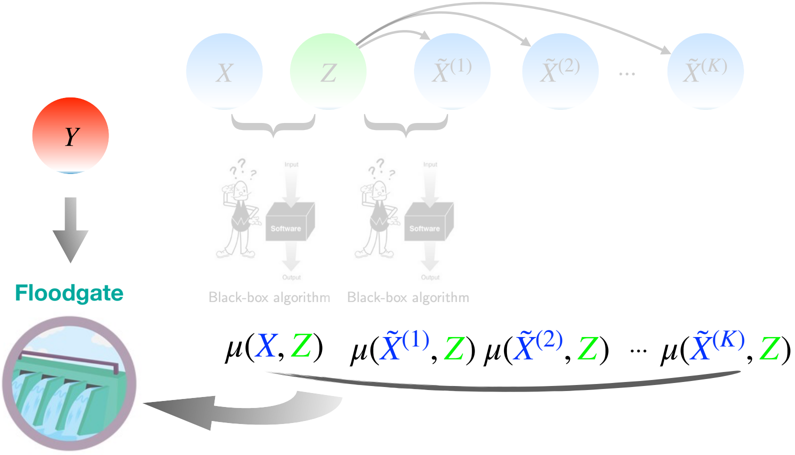
$\mu(X, Z)$ $\mu(\tilde{X}^{(1)}, Z)$

General computation



$\mu(X, Z)$ $\mu(\tilde{X}^{(1)}, Z)$ $\mu(\tilde{X}^{(2)}, Z)$ $\cdots$ $\mu(\tilde{X}^{(K)}, Z)$

General computation



General computation

How to compute $\mathbb{E}\left[\mu(X,Z) \mid Z\right], \mathrm{Var}\left(\mu(X,Z) \mid Z\right)$ in a general way (e.g., μ is fitted based on random forest or neural networks)?

- Sample $\tilde{X}$ from $P_{X|Z}$, conditionally independently of X, Y.
- We know

$$\mathbb{E}\left[Y(\mu(X,Z) - \mu(\tilde{X},Z))\right] = \mathbb{E}\left[Y\left(\mu(X,Z) - \mathbb{E}\left[\mu(X,Z) \mid Z\right]\right)\right]$$

$$\frac{1}{2}\,\mathbb{E}\left[\left(\mu(X,Z) - \mu(\tilde{X},Z)\right)^2\right] = \mathbb{E}\left[\mathrm{Var}(\mu(X,Z) \mid Z)\right]$$

Upper confidence bound

Suppose there were no Z.

- $\mathcal{I}^2 = \mathrm{Var}\left(\mathbb{E}\left[Y \mid X\right]\right)$.
- $\mathrm{Var}\left(Y\right)$ is a trivial UCB, as $\mathcal{I}^2 \leq \mathrm{Var}\left(Y\right)$.

Theorem (Zhang and Janson (2020); informal)

Under our assumptions, any asymptotically-valid UCB$_\alpha$ will asymptotically be $\geq \mathrm{Var}\left(Y\right)$ with probability at least $1 - \alpha$.

Valid, nontrivial UCB **impossible** without structure on $Y \mid X$.

Intuition behind the UCB result

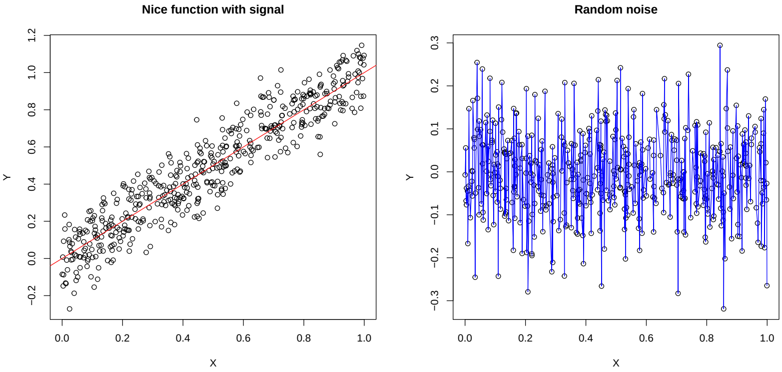


Figure: Left: $Y = X + \mathcal{N}(0, 0.1)$; Right: $Y \sim \mathcal{N}(0, 0.1)$.

Statistical accuracy

Statistical accuracy

Floodgate procedure is invariant with respect to a "equivalent" function class of μ,

$$S_\mu = \{c\mu(x, z) + g(z) : c > 0, g : \mathbb{R}^p \to \mathbb{R}\}.$$

Statistical accuracy

Floodgate procedure is invariant with respect to a "equivalent" function class of μ,

$$S_\mu = \{c\mu(x,z) + g(z) : c > 0, g : \mathbb{R}^p \to \mathbb{R}\}.$$

- If the true model is partial linear: $\mu^\star(x,z) = a^\star x + g^\star(z)$, only need to know sign$(a^\star)$.

Statistical accuracy

Floodgate procedure is invariant with respect to a "equivalent" function class of μ,

$$S_\mu = \{c\mu(x,z) + g(z) : c > 0, g : \mathbb{R}^p \to \mathbb{R}\}.$$

- If the true model is partial linear: $\mu^\star(x,z) = a^\star x + g^\star(z)$, only need to know sign$(a^\star)$.

Theorem (Zhang and Janson (2020); informal)

Under mild moment conditions on Y and noises, for μ_n with well-behaved moments,

$$\mathcal{I} - L_n^\alpha(\mu_n) = O_p\left(\inf_{\mu \in S_{\mu_n}} \mathbb{E}\left[(\mu(X,Z) - \mu^\star(X,Z))^2\right] + n^{-1/2}\right).$$

Statistical accuracy

$$S_\mu = \{c\mu(x,z) + g(z) : c > 0, g : \mathbb{R}^p \to \mathbb{R}\}.$$

Theorem (Zhang and Janson (2020); informal)

$$\mathcal{I} - L_n^\alpha(\mu_n) = O_p\left(\inf_{\mu \in S_{\mu_n}} \mathbb{E}\left[(\mu(X,Z) - \mu^\star(X,Z))^2\right] \qquad \right).$$

Statistical accuracy

$$S_\mu = \{c\mu(x, z) + g(z) : c > 0, g : \mathbb{R}^p \to \mathbb{R}\}.$$

Theorem (Zhang and Janson (2020); informal)

$$\mathcal{I} - L_n^\alpha(\mu_n) = O_p\left(\qquad\qquad\qquad\qquad + n^{-1/2} \right).$$

Statistical accuracy

$$S_\mu = \{c\mu(x,z) + g(z) : c > 0, g : \mathbb{R}^p \to \mathbb{R}\}.$$

Theorem (Zhang and Janson (2020); informal)

$$\mathcal{I} - L_n^\alpha(\mu_n) = O_p\left(\inf_{\mu \in S_{\mu_n}} \mathbb{E}\left[(\mu(X,Z) - \mu^\star(X,Z))^2\right] + n^{-1/2}\right).$$

Floodgate is **adaptive** to the accuracy of μ_n

(through the MSE of the best element of its equivalence class S_{μ_n})

Robustness

Suppose $P_{X|Z}$ unknown, we instead use its estimate $Q^{(n)}_{X|Z}$ to run floodgate.

Theorem (Zhang and Janson (2020); informal)

Under moment conditions on Y and noises, for μ_n with well-behaved moments under both the true distribution P and the specified one $Q^{(n)}$, then for floodgate with $Q^{(n)}_{X|Z}$ we have

$$\mathbb{P}\left(L_n^{\alpha}(\mu_n) \leq \mathcal{I} + \Delta_n\right) \geq 1 - \alpha - O(n^{-1/2}),$$

where

Robustness

Suppose $P_{X|Z}$ unknown, we instead use its estimate $Q^{(n)}_{X|Z}$ to run floodgate.

Theorem (Zhang and Janson (2020); informal)

Under moment conditions on Y and noises, for μ_n with well-behaved moments under both the true distribution P and the specified one $Q^{(n)}$, then for floodgate with $Q^{(n)}_{X|Z}$ we have

$$\mathbb{P}\left(L_n^\alpha(\mu_n) \leq \mathcal{I} + \Delta_n\right) \geq 1 - \alpha - O(n^{-1/2}),$$

where

$$\Delta_n \leq c_1 \sqrt{\mathbb{E}\left[\chi^2\left(P_{X|Z} \,||\, Q^{(n)}_{X|Z}\right)\right]}$$

Robustness

Suppose $P_{X|Z}$ unknown, we instead use its estimate $Q^{(n)}_{X|Z}$ to run floodgate.

Theorem (Zhang and Janson (2020); informal)

Under moment conditions on Y and noises, for μ_n with well-behaved moments under both the true distribution P and the specified one $Q^{(n)}$, then for floodgate with $Q^{(n)}_{X|Z}$ we have

$$\mathbb{P}\left(L_n^{\alpha}(\mu_n) \leq \mathcal{I} + \Delta_n\right) \geq 1 - \alpha - O(n^{-1/2}),$$

where

$$\Delta_n \leq c_1 \sqrt{\mathbb{E}\left[\chi^2\left(P_{X|Z} \,||\, Q^{(n)}_{X|Z}\right)\right]} - c_2\, \mathbb{E}\left[(\bar{\mu}_n(X,Z) - \mu^{\star}(X,Z))^2\right]$$

where $\bar{\mu}_n$ is a particular representative of S_{μ_n} and $\chi^2(\cdot\,||\,\cdot)$ denotes the χ^2 divergence.

Robustness

Theorem (Zhang and Janson (2020); informal)

$$\mathbb{P}\left(L_n^{\alpha}(\mu_n) \leq \mathcal{I} + \Delta_n\right) \geq 1 - \alpha - O(n^{-1/2}),$$

where

$$\Delta_n \leq c_1 \sqrt{\mathbb{E}\left[\chi^2\left(P_{X|Z} \,||\, Q_{X|Z}^{(n)}\right)\right]} \; - \; c_2 \, \mathbb{E}\left[(\bar{\mu}_n(X,Z) - \mu^{\star}(X,Z))^2\right]$$

where $\bar{\mu}_n$ is a particular representative of S_{μ_n} and $\chi^2(\cdot \,||\, \cdot)$ denotes the χ^2 divergence.

Robustness

Theorem (Zhang and Janson (2020); informal)

$$\mathbb{P}\left(L_n^\alpha(\mu_n) \leq \mathcal{I} + \Delta_n\right) \geq 1 - \alpha - O(n^{-1/2}),$$

where

$$\Delta_n \leq c_1 \sqrt{\mathbb{E}\left[\chi^2\left(P_{X|Z} \,||\, Q_{X|Z}^{(n)}\right)\right]} \;-\; c_2\, \mathbb{E}\left[(\bar{\mu}_n(X,Z) - \mu^\star(X,Z))^2\right]$$

where $\bar{\mu}_n$ is a particular representative of S_{μ_n} and $\chi^2(\cdot\,||\,\cdot)$ denotes the χ^2 divergence.

Note: by definition of $\bar{\mu}_n$, we have: $\mathcal{I} = 0 \quad \Rightarrow \mathbb{E}\left[(\bar{\mu}_n(X,Z) - \mu^\star(X,Z))^2\right] = 0$.

Robustness

Theorem (Zhang and Janson (2020); informal)

$$\mathbb{P}\left(L_n^{\alpha}(\mu_n) \leq \mathcal{I} + \Delta_n\right) \geq 1 - \alpha - O(n^{-1/2}),$$

where

$$\Delta_n \leq c_1 \sqrt{\mathbb{E}\left[\chi^2\left(P_{X|Z} \,||\, Q_{X|Z}^{(n)}\right)\right]} - c_2\, \mathbb{E}\left[(\bar{\mu}_n(X,Z) - \mu^{\star}(X,Z))^2\right]$$

where $\bar{\mu}_n$ *is a particular representative of* S_{μ_n} *and* $\chi^2(\cdot\,||\,\cdot)$ *denotes the* χ^2 *divergence.*

Note: by definition of $\bar{\mu}_n$, we have: $\mathcal{I} = 0 \quad \Rightarrow \mathbb{E}\left[(\bar{\mu}_n(X,Z) - \mu^{\star}(X,Z))^2\right] = 0$.

Floodgate is robust if $P_{X|Z}$ well-estimated.

Robustness

Theorem (Zhang and Janson (2020); informal)

$$\mathbb{P}\left(L_n^{\alpha}(\mu_n) \leq \mathcal{I} + \Delta_n\right) \geq 1 - \alpha - O(n^{-1/2}),$$

where

$$\Delta_n \leq c_1 \sqrt{\mathbb{E}\left[\chi^2\left(P_{X|Z} \,||\, Q_{X|Z}^{(n)}\right)\right]} - c_2\, \mathbb{E}\left[(\bar{\mu}_n(X,Z) - \mu^{\star}(X,Z))^2\right]$$

where $\bar{\mu}_n$ is a particular representative of S_{μ_n} and $\chi^2(\cdot \,||\, \cdot)$ denotes the χ^2 divergence.

Note: by definition of $\bar{\mu}_n$, we have: $\mathcal{I} = 0 \quad \Rightarrow \quad \mathbb{E}\left[(\bar{\mu}_n(X,Z) - \mu^{\star}(X,Z))^2\right] = 0$.

Floodgate is robust if $P_{X|Z}$ well-estimated.

If $\mathcal{I} > 0$, floodgate is robust if $P_{X|Z}$ better-estimated than $\mathbb{E}[Y \mid X, Z]$.

Simulation setup

- $n = 1100$, $p = 1000$, and a sparsity of 30 unless stated otherwise.
- Linear: $X \sim \mathcal{N}(0, \Sigma)$, AR(1); $Y = X\beta + \mathcal{N}(0, 1)$, $||\beta||_0 = 30$, $|\beta_j| \in \{0, \frac{\text{amplitude}}{\sqrt{n}}\}$.
- Nonlinear: each component chosen from below; up to 3rd order interactions.

$$\sin(\pi x),\ \cos(\pi x),\ \sin(\pi x/2),\ \cos(\pi x)I(x > 0),\ x\sin(\pi x),\ x,\ |x|,\ x^2,\ x^3,\ e^x - 1.$$

- Number of replicates: 64.
- Default sample splitting proportion: 0.50.
- Four fitting algorithms: LASSO, Ridge, SAM, Random Forest.
- Number of null samples: $K = 500$.

Splitting proportion

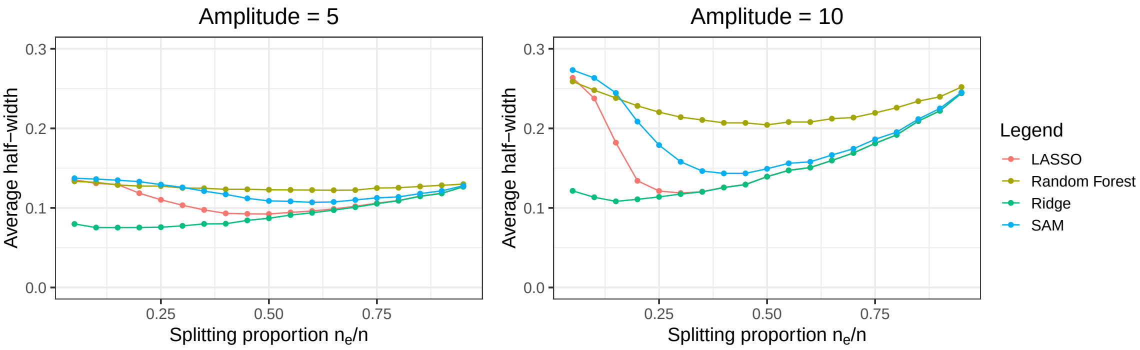


Figure: Linear setting.

Splitting proportion

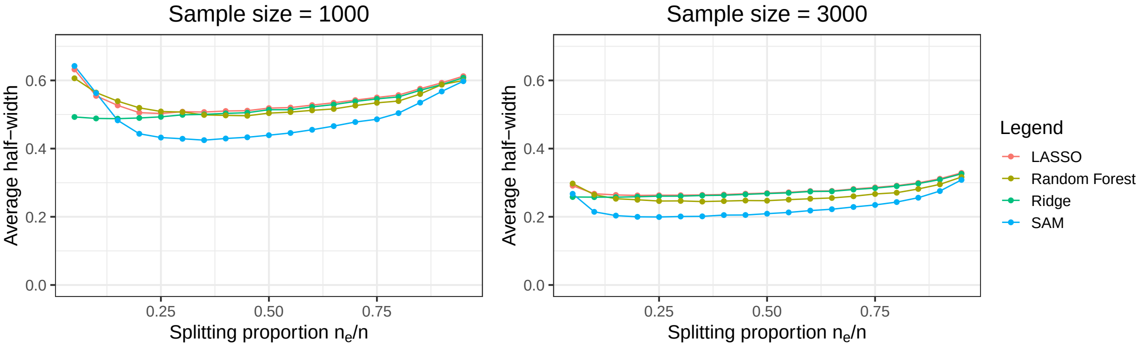


Figure: Nonlinear setting.

Covariate dimension

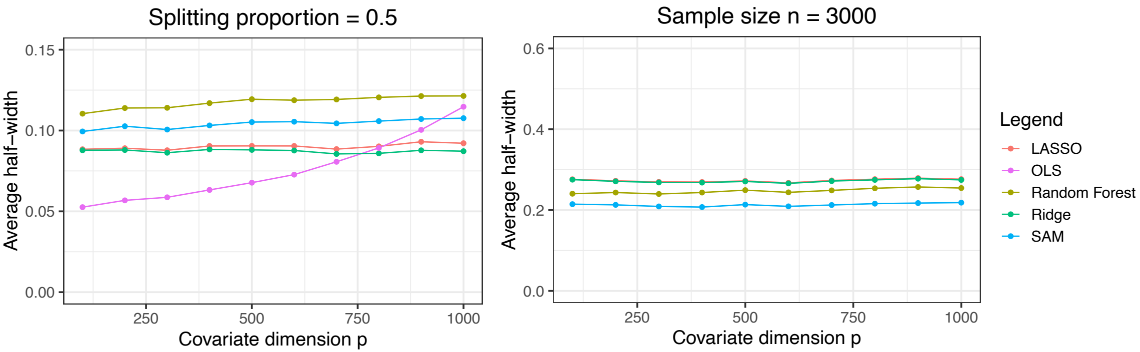


Figure: Left: Linear setting; Right: Nonlinear setting.

Robustness

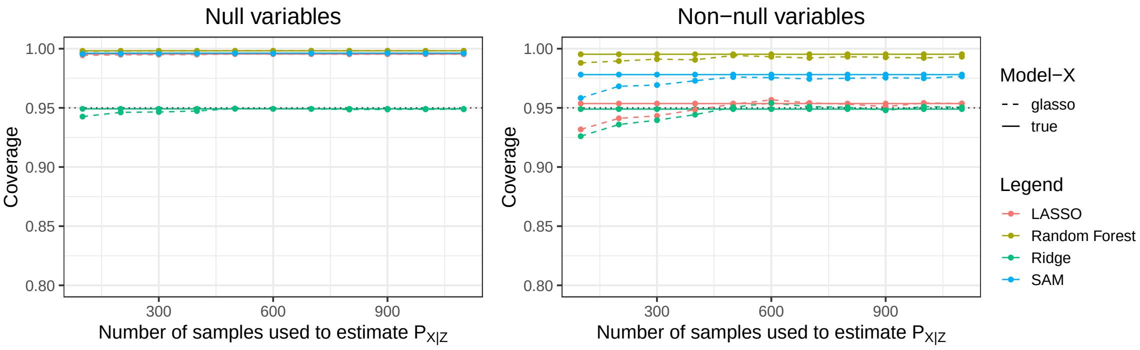


Figure: Linear setting.

Robustness

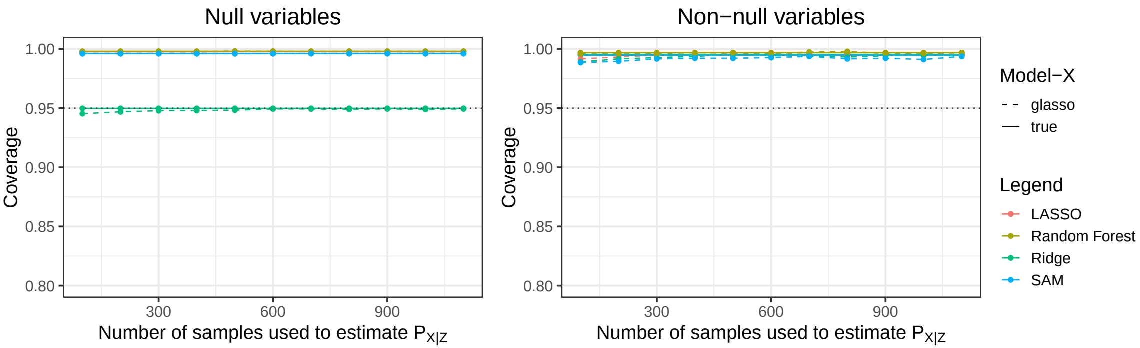


Figure: Nonlinear setting.

Genomic study of platelet count

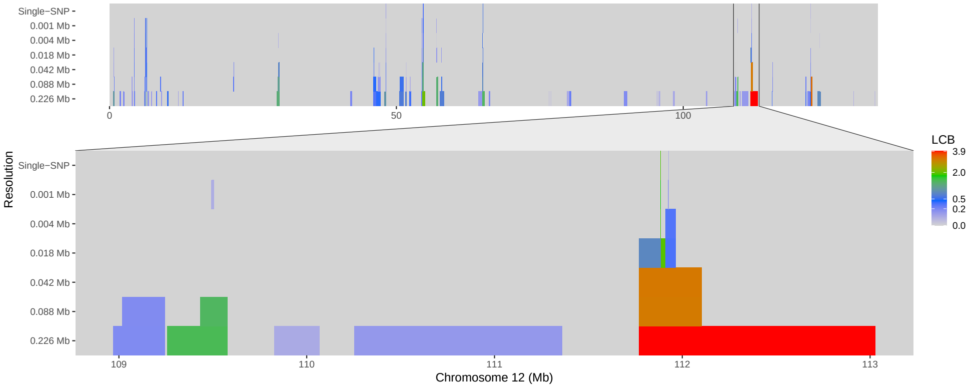


Figure: Colored Chicago plot (Sesia et al., 2020) with the color of each point representing the floodgate LCB for the importance of a group of SNPs on Chromosome 12 in the UK Biobank data at different resolutions (y-axis). Bottom plot shows a zoomed-in region of strong importance.

Extensions

Questions	Answers

Extensions

Questions	**Answers**
What if only know a model for $P_{X\|Z}$?	☑ Co-sufficient floodgate.

Extensions

Questions	Answers
What if only know a model for $P_{X\|Z}$?	☑ Co-sufficient floodgate.
Beyond the mMSE gap?	☑ Floodgate for MACM gap.

Extensions

Questions	**Answers**
What if only know a model for $P_{X\|Z}$?	☑ Co-sufficient floodgate.
Beyond the mMSE gap?	☑ Floodgate for MACM gap.
Inferring the MOVI w.r.t $\{X_j\}_{j \in \mathcal{G}}$?	☑ Easily extends.

Extensions

Questions	**Answers**
What if only know a model for $P_{X\|Z}$?	☑ Co-sufficient floodgate.
Beyond the mMSE gap?	☑ Floodgate for MACM gap.
Inferring the MOVI w.r.t $\{X_j\}_{j\in\mathcal{G}}$?	☑ Easily extends.
Transporting inference from $P_{(X,Z)}$ to $Q_{(X,Z)}$?	☑ Easily extends.

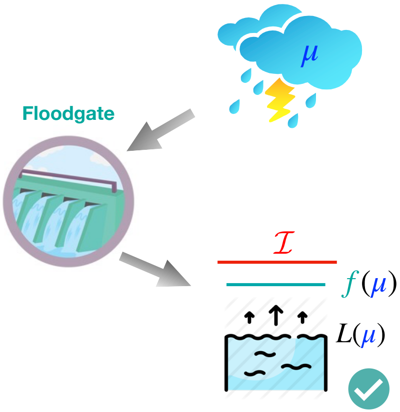
Extensions

Questions	**Answers**
What if only know a model for $P_{X\|Z}$?	☑ Co-sufficient floodgate.
Beyond the mMSE gap?	☑ Floodgate for MACM gap.
Inferring the MOVI w.r.t $\{X_j\}_{j\in\mathcal{G}}$?	☑ Easily extends.
Transporting inference from $P_{(X,Z)}$ to $Q_{(X,Z)}$?	☑ Easily extends.
Adjusting for multiplicity and selection effects?	☑ Has answers.

Extensions

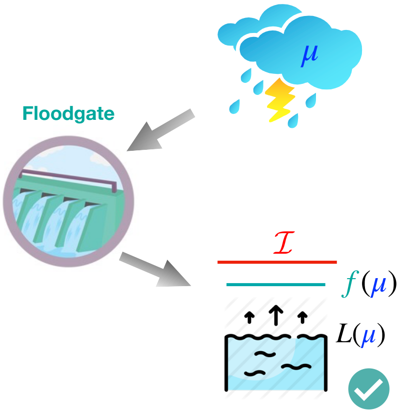
Questions	**Answers**
What if only know a model for $P_{X\|Z}$?	☑ Co-sufficient floodgate.
Beyond the mMSE gap?	☑ Floodgate for MACM gap.
Inferring the MOVI w.r.t $\{X_j\}_{j\in\mathcal{G}}$?	☑ Easily extends.
Transporting inference from $P_{(X,Z)}$ to $Q_{(X,Z)}$?	☑ Easily extends.
Adjusting for multiplicity and selection effects?	☑ Has answers.
Inferring the normalized mMSE gap $\frac{\mathcal{I}}{\sqrt{\mathrm{Var}(Y)}}$?	☑ Easily extends.

Summary



Floodgate: a new inferential approach for variable importance.

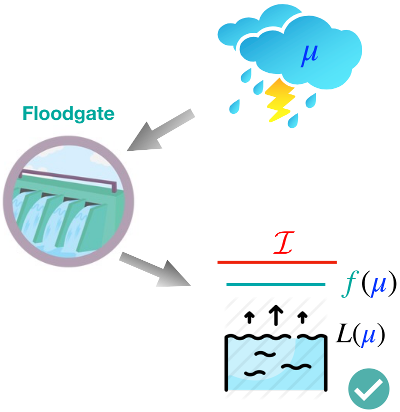
Summary



Floodgate: a new inferential approach for variable importance.

- Focus on an interpretable, sensitive and model-free MOVI: the mMSE gap.

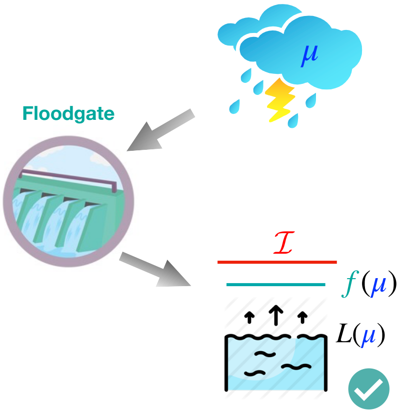
Summary



Floodgate: a new inferential approach for variable importance.

- Provide valid and robust LCBs for the mMSE gap.

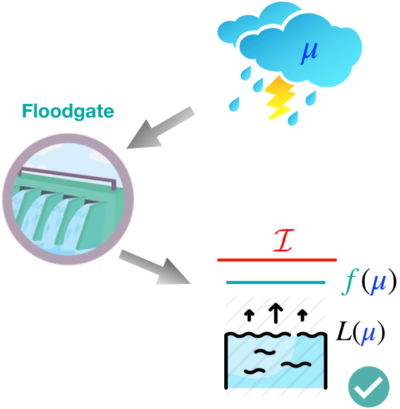
Summary



Floodgate: a new inferential approach for variable importance.

- Allow flexible regression algorithms, and is adaptive to the MSE.

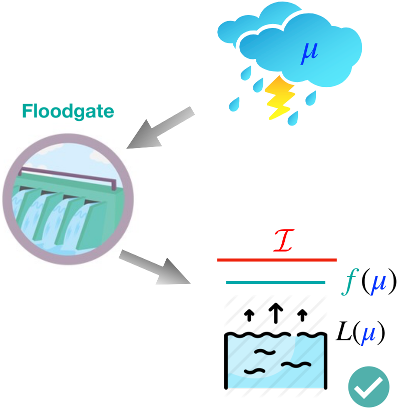
Discussion: beyond this paper



Floodgate: a new inferential approach for

____?

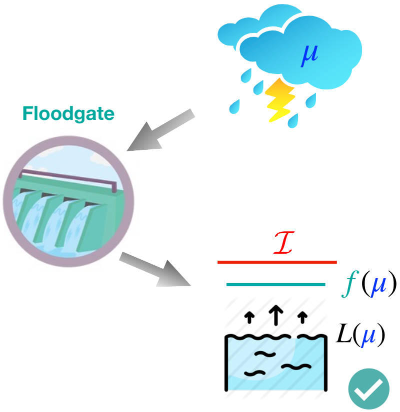
Discussion: beyond this paper



Floodgate: a new inferential approach for ________________________ ?

- How to characterize a class of feasible model-free targets?

Discussion: beyond this paper

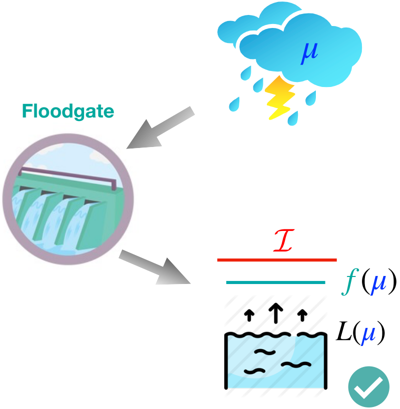


Floodgate: a new inferential approach for

________________________________?

- How to characterize a class of feasible model-free targets?
- How to construct floodgate functional f?

Discussion: beyond this paper



Floodgate: a new inferential approach for ________________?

- How to characterize a class of feasible model-free targets?
- How to construct floodgate functional f?
- How to obtain LCBs for $f(\cdot)$ under reasonable conditions?

Appendix

MACM gap

Definition (Mean absolute conditional mean gap)

The *mean absolute conditional mean (MACM) gap* for variable X is defined as

$$\mathcal{I}_{\ell_1} = \mathbb{E}\left[\left|\mathbb{E}\left[Y \mid Z\right] - \mathbb{E}\left[Y \mid X, Z\right]\right|\right]$$

whenever all the above expectations exist.

The subscript in $\mathcal{I}_{\ell_1}$ reflects its similarity to $\mathcal{I}^2 = \mathbb{E}\left[\left(\mathbb{E}\left[Y \mid Z\right] - \mathbb{E}\left[Y \mid X, Z\right]\right)^2\right]$ except with the square replaced by the absolute value (also known as the ℓ_1 norm).

Covariate dimension

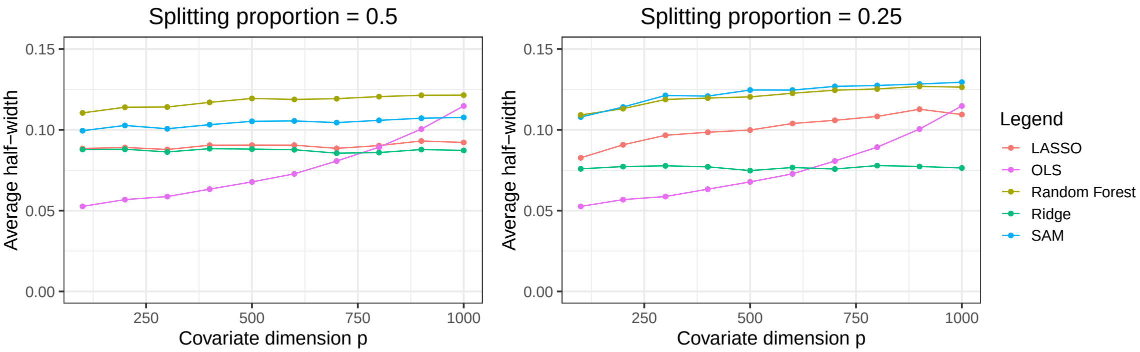


Figure: Linear setting.

Covariate dimension

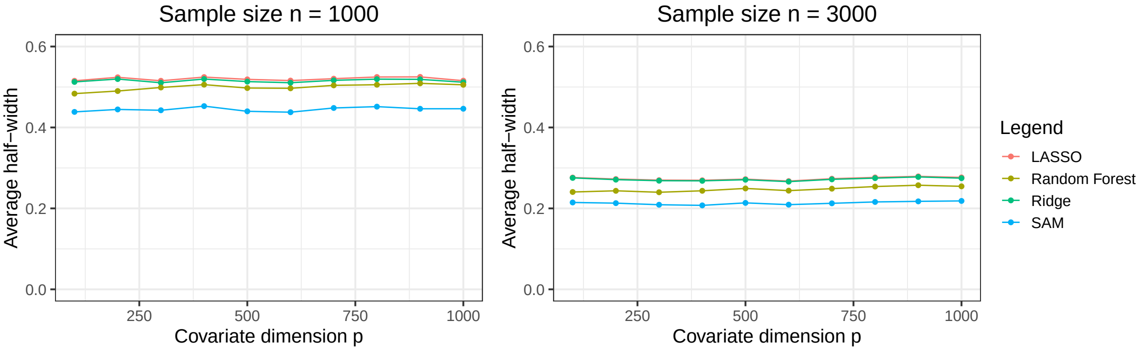


Figure: Nonlinear setting.

Sample size

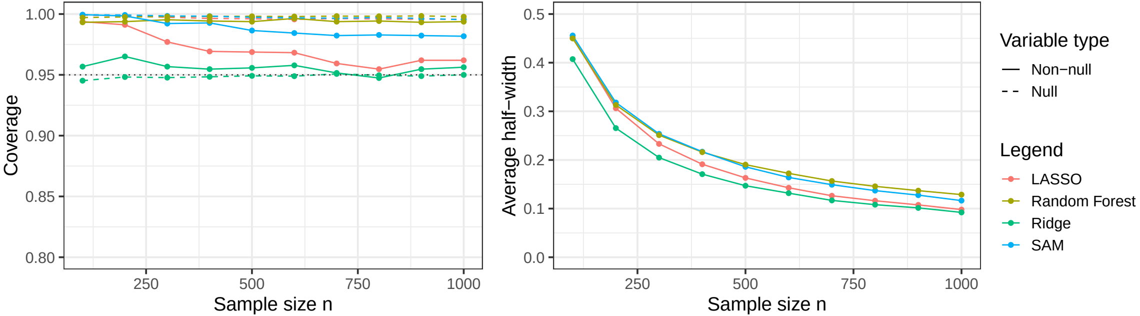


Figure: Linear setting.

Sample size

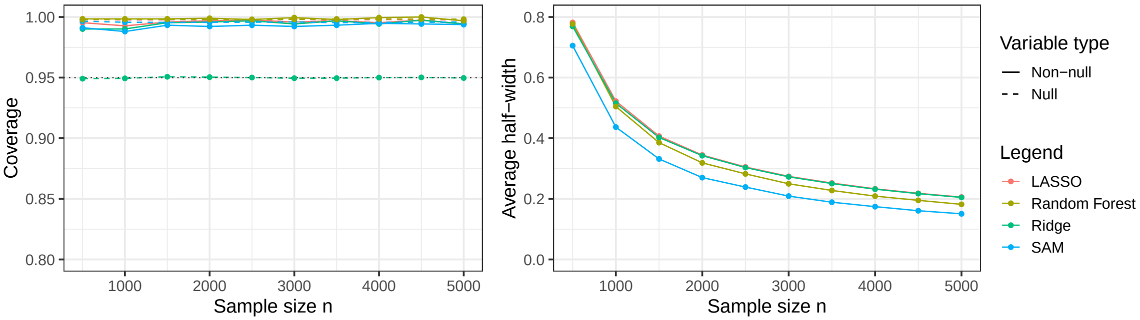


Figure: Nonlinear setting.

References I

Berk, R., Brown, L., Buja, A., Zhang, K., Zhao, L., et al. (2013). Valid post-selection inference. *The Annals of Statistics*, 41(2):802–837.

Bühlmann, P. et al. (2013). Statistical significance in high-dimensional linear models. *Bernoulli*, 19(4):1212–1242.

Bühlmann, P., van de Geer, S., et al. (2015). High-dimensional inference in misspecified linear models. *Electronic Journal of Statistics*, 9(1):1449–1473.

Buja, A., Berk, R. A., Brown, L. D., George, E. I., Pitkin, E., Traskin, M., Zhao, L., and Zhang, K. (2015). Models as approximations-a conspiracy of random regressors and model deviations against classical inference in regression. *Statistical Science*, page 1.

References II

Buja, A. and Brown, L. (2014). Discussion:" a significance test for the lasso". *The Annals of Statistics*, 42(2):509–517.

Buja, A., Brown, L., Berk, R., George, E., Pitkin, E., Traskin, M., Zhang, K., Zhao, L., et al. (2019a). Models as approximations i: Consequences illustrated with linear regression. *Statistical Science*, 34(4):523–544.

Buja, A., Brown, L., Kuchibhotla, A. K., Berk, R., George, E., Zhao, L., et al. (2019b). Models as approximations ii: A model-free theory of parametric regression. *Statistical Science*, 34(4):545–565.

Dezeure, R., Bühlmann, P., and Zhang, C.-H. (2017). High-dimensional simultaneous inference with the bootstrap. *Test*, 26(4):685–719.

References III

Javanmard, A. and Montanari, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *The Journal of Machine Learning Research*, 15(1):2869–2909.

Lee, J. D., Sun, D. L., Sun, Y., Taylor, J. E., et al. (2016). Exact post-selection inference, with application to the lasso. *The Annals of Statistics*, 44(3):907–927.

Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J., and Wasserman, L. (2018). Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113(523):1094–1111.

Li, L., Tchetgen, E. T., van der Vaart, A., and Robins, J. M. (2011). Higher order inference on a treatment effect under low regularity conditions. *Statistics & probability letters*, 81(7):821–828.

References IV

Newey, W. K. and Robins, J. R. (2018). Cross-fitting and fast remainder rates for semiparametric estimation. *arXiv preprint arXiv:1801.09138*.

Nickl, R., Van De Geer, S., et al. (2013). Confidence sets in sparse regression. *The Annals of Statistics*, 41(6):2852–2876.

Pinelis, I., Molzon, R., et al. (2016). Optimal-order bounds on the rate of convergence to normality in the multivariate delta method. *Electronic Journal of Statistics*, 10(1):1001–1063.

Rinaldo, A., Wasserman, L., G'Sell, M., et al. (2019). Bootstrapping and sample splitting for high-dimensional, assumption-lean inference. *The Annals of Statistics*, 47(6):3438–3469.

References V

Robins, J., Li, L., Tchetgen, E., van der Vaart, A., et al. (2008). Higher order influence functions and minimax estimation of nonlinear functionals. In *Probability and statistics: essays in honor of David A. Freedman*, pages 335–421. Institute of Mathematical Statistics.

Robins, J., Tchetgen, E. T., Li, L., and van der Vaart, A. (2009). Semiparametric minimax rates. *Electronic journal of statistics*, 3:1305.

Robins, J. M., Li, L., Mukherjee, R., Tchetgen, E. T., van der Vaart, A., et al. (2017). Minimax estimation of a functional on a structured high-dimensional model. *The Annals of Statistics*, 45(5):1951–1987.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global sensitivity analysis: the primer*. John Wiley & Sons.

References VI

Sesia, M., Katsevich, E., Bates, S., Candès, E., and Sabatti, C. (2020). Multi-resolution localization of causal variants across the genome. *Nature communications*, 11(1):1–10.

Shah, R. D. and Peters, J. (2018). The hardness of conditional independence testing and the generalised covariance measure. *arXiv preprint arXiv:1804.07203*.

Taylor, J., Lockhart, R., Tibshirani, R. J., and Tibshirani, R. (2014). Exact post-selection inference for forward stepwise and least angle regression. *arXiv preprint arXiv:1401.3889*, 7:10–1.

Van de Geer, S., Bühlmann, P., Ritov, Y., Dezeure, R., et al. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3):1166–1202.

References VII

Watson, D. S. and Wright, M. N. (2019). Testing conditional predictive independence in supervised learning algorithms. *arXiv preprint arXiv:1901.09917*.

Williamson, B. D., Gilbert, P. B., Simon, N., and Carone, M. (2017). Nonparametric variable importance assessment using machine learning techniques. *UW Biostatistics Working Paper Series. Working Paper 422*.

Williamson, B. D., Gilbert, P. B., Simon, N. R., and Carone, M. (2020). A unified approach for inference on algorithm-agnostic variable importance. *arXiv preprint arXiv:2004.03683*.

Zhang, C.-H. and Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):217–242.

References VIII

Zhang, L. and Janson, L. (2020). Floodgate: Inference for model-free variable importance. *arXiv preprint arXiv:2007.01283*.

Zhang, X. and Cheng, G. (2017). Simultaneous inference for high-dimensional linear models. *Journal of the American Statistical Association*, 112(518):757–768.