Floodgate: Inference for Model-free Variable Importance

Lu Zhang
Joint work with Lucas Janson

Department of Statistics
Harvard University

August 26, 2020

Summary

Setup : data (Y, X, Z).

- Y : response variable ; X : the variable of interest ; $Z := (Z_1, \cdots, Z_p)$ confounders.

Q : how important each covariate (X) is in this relationship ?
A : introduce floodgate, a new inferential approach for variable importance.

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

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

- Provide valid and robust lower confidence bounds for the mMSE gap.

$$f(\mu) := \frac{\mathbb{E}[\mathrm{Cov}(\mu^\star(X,Z), \mu(X,Z) \mid Z)]}{\sqrt{\mathbb{E}[\mathrm{Var}(\mu(X,Z) \mid Z)]}}, \quad f(\mu) \leq \mathcal{I} \text{ for any } \mu.$$

- Allow flexible regression algorithms, good prediction $\Rightarrow$ good accuracy.

Genomic application to UKBB data : colored “Chicago" plot.

Zhang, Lu, and Lucas Janson. "Floodgate : inference for model-free variable importance." arXiv preprint arXiv:2007.01283 (2020).

Extensions : relax the assumption ; a different MOVI for binary responses ; group variable importance ; different covariate distribution ; adjusting for multiplicity and selection effects.

Motivation

Motivation

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

Motivation

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

- Y a response variable of interest
- 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)

Motivation

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

- Y a response variable of interest
- 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)

Question : is the variable X important or not ?

Motivation

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

- Y a response variable of interest
- 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)

Question : is the variable X important or not ?

- Assuming parametric models : testing whether the coefficients are zero.

Motivation

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

- Y a response variable of interest
- 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)

Question : is the variable X important or not ?

- Assuming parametric models : testing whether the coefficients are zero.
- Conditional independence testing (without parametric assumption) :

$$Y \perp\!\!\!\perp X \mid Z$$

Motivation

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

- Y a response variable of interest
- 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)

Question : is the variable X important or not ?

- Assuming parametric models : testing whether the coefficients are zero.
- Conditional independence testing (without parametric assumption) :

$$Y \perp\!\!\!\perp X \mid Z$$

  - Kernel-based conditional independence tests.
  - Semi-parametric approaches.
  - Model-X approaches : model-X knockoffs, conditional knockoffs, conditional randomization tests, conditional permutation tests, hold-out randomization tests and so on.
  - Symmetry idea approaches : Gaussian mirrors and data splitting.

Motivation

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

- Y a response variable of interest
- 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)

Question : is the variable X important or not ?

- Assuming parametric models : testing whether the coefficients are zero.
- Conditional independence testing (without parametric assumption) :

$$Y \perp\!\!\!\perp X \mid Z$$

  - Kernel-based conditional independence tests.
  - Semi-parametric approaches.
  - Model-X approaches : model-X knockoffs, conditional knockoffs, conditional randomization tests, conditional permutation tests, hold-out randomization tests and so on.
  - Symmetry idea approaches : Gaussian mirrors and data splitting.

Go beyond : how important each covariate is in this relationship ?

Motivation

Applications : variable ranking, SNP-heritability ...

Motivation

Applications : variable ranking, SNP-heritability ...

Confidence intervals on the parameters ;

Motivation

Applications : variable ranking, SNP-heritability ...

Confidence intervals on the parameters ;What if no parametric assumption ?

Motivation

Applications : variable ranking, SNP-heritability ...

Confidence intervals on the parameters ;What if no parametric assumption ?

1. How to define a good measure of variable importance (MOVI) ?

Motivation

Applications : variable ranking, SNP-heritability ...

Confidence intervals on the parameters ;What if no parametric assumption ?

1. How to define a good measure of variable importance (MOVI) ?
2. How to provide inference for it ?

Motivation

Applications : variable ranking, SNP-heritability ...

Confidence intervals on the parameters ;What if no parametric assumption ?

1. How to define a good measure of variable importance (MOVI) ?
2. How to provide inference for it ?

A desirable MOVI (of the covariate X) should have

Validity : zero when $Y \perp\!\!\!\perp X \mid Z$.

Motivation

Applications : variable ranking, SNP-heritability ...

Confidence intervals on the parameters ;What if no parametric assumption ?

1. How to define a good measure of variable importance (MOVI) ?
2. How to provide inference for it ?

A desirable MOVI (of the covariate X) should have

Validity : zero when $Y \perp\!\!\!\perp X \mid Z$.

Sensitivity : able to detect nonlinear effects and interactions.

Motivation

Applications : variable ranking, SNP-heritability ...

Confidence intervals on the parameters ;What if no parametric assumption ?

1. How to define a good measure of variable importance (MOVI) ?
2. How to provide inference for it ?

A desirable MOVI (of the covariate X) should have

Validity : zero when $Y \perp\!\!\!\perp X \mid Z$.

Sensitivity : able to detect nonlinear effects and interactions.

Interpretability : interpretable and ready for scientists and practitioners' use.

Motivation

Applications : variable ranking, SNP-heritability ...

Confidence intervals on the parameters ;What if no parametric assumption ?

1. How to define a good measure of variable importance (MOVI) ?
2. How to provide inference for it ?

A desirable MOVI (of the covariate X) should have

Validity : zero when $Y \perp\!\!\!\perp X \mid Z$.

Sensitivity : able to detect nonlinear effects and interactions.

Interpretability : interpretable and ready for scientists and practitioners' use.

A desirable inferential procedure for the MOVI should be

Motivation

Applications : variable ranking, SNP-heritability ...

Confidence intervals on the parameters ;What if no parametric assumption ?

1. How to define a good measure of variable importance (MOVI) ?
2. How to provide inference for it ?

A desirable MOVI (of the covariate X) should have

Validity : zero when $Y \perp\!\!\!\perp X \mid Z$.

Sensitivity : able to detect nonlinear effects and interactions.

Interpretability : interpretable and ready for scientists and practitioners' use.

A desirable inferential procedure for the MOVI should be

- General

Motivation

Applications : variable ranking, SNP-heritability ...

Confidence intervals on the parameters ;What if no parametric assumption ?

1. How to define a good measure of variable importance (MOVI) ?
2. How to provide inference for it ?

A desirable MOVI (of the covariate X) should have

Validity : zero when $Y \perp\!\!\!\perp X \mid Z$.

Sensitivity : able to detect nonlinear effects and interactions.

Interpretability : interpretable and ready for scientists and practitioners' use.

A desirable inferential procedure for the MOVI should be

- General
- Accurate

Motivation

Applications : variable ranking, SNP-heritability ...

Confidence intervals on the parameters ;What if no parametric assumption ?

1. How to define a good measure of variable importance (MOVI) ?
2. How to provide inference for it ?

A desirable MOVI (of the covariate X) should have

Validity : zero when $Y \perp\!\!\!\perp X \mid Z$.

Sensitivity : able to detect nonlinear effects and interactions.

Interpretability : interpretable and ready for scientists and practitioners' use.

A desirable inferential procedure for the MOVI should be

- General
- Accurate
- Robust

Literature review

- 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), Sur and Candès (2019), Zhao et al. (2020) ...
- 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).
- Random parameters : Lei et al. (2018), Fisher et al. (2018), Watson and Wright (2019), Rinaldo et al. (2019).
- Semi-parametric approaches : Robins et al. (2008, 2009); Li et al. (2011); Robins et al. (2017); Newey and Robins (2018), Shah and Peters (2018).
- A very recent MOVI : Azadkia and Chatterjee (2019).
- Same MOVI as us : Williamson et al. (2017).

Our target MOVI : the mMSE gap

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.

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** : $\mathcal{I}^2 = \mathrm{Var}(\mathbb{E}[Y \mid X, Z]) - \mathrm{Var}(\mathbb{E}[Y \mid Z])$.

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** : $\mathcal{I}^2 = \frac{1}{2}\mathbb{E}_{x_1, x_2 \overset{i.i.d.}{\sim} P_{X \mid Z}}\left[(\mathbb{E}[Y \mid X = x_1, Z] - \mathbb{E}[Y \mid X = x_2, Z])^2\right]$.

Our target MOVI : the mMSE gap

Definition (mMSE Gap)

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

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

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

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

Our target MOVI : the mMSE gap

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.

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** : $\mathcal{I}^2 = \mathrm{Var}(\mathbb{E}[Y \mid X, Z]) - \mathrm{Var}(\mathbb{E}[Y \mid Z])$.

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** : $\mathcal{I}^2 = \frac{1}{2}\mathbb{E}_{x_1, x_2 \overset{i.i.d.}{\sim} P_{X \mid Z}}\left[(\mathbb{E}[Y \mid X = x_1, Z] - \mathbb{E}[Y \mid X = x_2, Z])^2\right]$.

Our target MOVI : the mMSE gap

Definition (mMSE Gap)

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

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

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

- **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** : $\mathcal{I}^2 = \frac{1}{2}\mathbb{E}_{x_1, x_2 \overset{i.i.d.}{\sim} P_{X \mid Z}}\left[(\mathbb{E}[Y \mid X = x_1, Z] - \mathbb{E}[Y \mid X = x_2, Z])^2\right]$.
- 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 $\mu^\star(x, z) := \mathbb{E}[Y \mid X = x, Z = z]$

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

Challenges :

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

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

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

Challenges :

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

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

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

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

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

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

Challenges :

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

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

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

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

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

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

Challenges :

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

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

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

- construct a functional f such that

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

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

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

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

Challenges :

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

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

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

- construct a functional f such that

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

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

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

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

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

Challenges :

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

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

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

- construct a functional f such that

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

- know how to obtain LCB $L(\mu)$ of $f(\mu)$ for any μ.
- (Ideally) the functional f also satisfies $f(\mu^\star) = \mathcal{I}$.

Floodgate LCB

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

Floodgate LCB

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

Lemma (Zhang and Janson (2020))

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

Floodgate LCB

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

Lemma (Zhang and Janson (2020))

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

Ingredients of our model-X inferential procedure :

Floodgate LCB

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

Lemma (Zhang and Janson (2020))

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

Ingredients of our model-X inferential procedure :

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

Floodgate LCB

Our choice of functional : $f(\mu) := \frac{\mathbb{E}\big[\mathrm{Cov}(\mu^\star(X,Z),\mu(X,Z)\mid Z)\big]}{\sqrt{\mathbb{E}[\mathrm{Var}(\mu(X,Z)\mid Z)]}}$

Lemma (Zhang and Janson (2020))

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

Ingredients of our model-X inferential procedure :

1. $(Y_i, X_i, Z_i)_{i=1}^n$.
2. μ (can be fitted from a separate dataset e.g. sample splitting).

Floodgate LCB

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

Lemma (Zhang and Janson (2020))

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

Ingredients of our model-X inferential procedure :

1. $(Y_i, X_i, Z_i)_{i=1}^n$.
2. μ (can be fitted from a separate dataset e.g. sample splitting).
3. Assume $P_{X|Z}$ known (also have robustness analysis and assumption relaxation).

Floodgate LCB

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

Lemma (Zhang and Janson (2020))

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

Ingredients of our model-X inferential procedure :

1. $(Y_i, X_i, Z_i)_{i=1}^n$.
2. μ (can be fitted from a separate dataset e.g. sample splitting).
3. Assume $P_{X|Z}$ known (also have robustness analysis and assumption relaxation).

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

Floodgate LCB

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

Lemma (Zhang and Janson (2020))

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

Ingredients of our model-X inferential procedure :

1. $(Y_i, X_i, Z_i)_{i=1}^n$.
2. μ (can be fitted from a separate dataset e.g. sample splitting).
3. Assume $P_{X|Z}$ known (also have robustness analysis and assumption relaxation).

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

- By Delta method, we can construct CLT-based LCB for $f(\mu)$: $L_n^\alpha(\mu)$ (with confidence level α).

Floodgate LCB

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

Lemma (Zhang and Janson (2020))

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

Ingredients of our model-X inferential procedure :

1. $(Y_i, X_i, Z_i)_{i=1}^n$.
2. μ (can be fitted from a separate dataset e.g. sample splitting).
3. Assume $P_{X|Z}$ known (also have robustness analysis and assumption relaxation).

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

- By Delta method, we can construct CLT-based LCB for $f(\mu)$: $L_n^\alpha(\mu)$ (with confidence level α).
- under certain fitted models, compute $\mathbb{E}\left[\mu(X,Z) \mid Z\right]$, $\mathrm{Var}\left(\mu(X,Z) \mid Z\right)$ analytically.

Floodgate LCB

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

Lemma (Zhang and Janson (2020))

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

Ingredients of our model-X inferential procedure :

1. $(Y_i, X_i, Z_i)_{i=1}^n$.
2. μ (can be fitted from a separate dataset e.g. sample splitting).
3. Assume $P_{X|Z}$ known (also have robustness analysis and assumption relaxation).

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

- By Delta method, we can construct CLT-based LCB for $f(\mu)$: $L_n^\alpha(\mu)$ (with confidence level α).
- under certain fitted models, compute $\mathbb{E}\left[\mu(X,Z) \mid Z\right]$, $\mathrm{Var}\left(\mu(X,Z) \mid Z\right)$ analytically.
- Generally, draw $\tilde{X}^{(k)}, k = 1, \cdots, K$ from $P_{X|Z}$, conditionally independently of X, Y then plug-in the Monte Carlo estimators.

Asymptotic validity

Theorem (Zhang and Janson (2020))

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

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

Asymptotic validity

Theorem (Zhang and Janson (2020))

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))

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

Asymptotic validity

Theorem (Zhang and Janson (2020))

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).
- Constant in $O(n^{-1/2})$ has complicated dependence on μ and $P_{(Y,X,Z)}$.
- Invariance of the floodgate procedure : e.g. $\mu(x, z) = ax + g(z)$, constant only depends on sign(a) and bivariate distribution of

$$\left(Y, \frac{X - \mathbb{E}[X \mid Z]}{\sqrt{\mathrm{Var}\left(X - \mathbb{E}[X \mid Z]\right)}}\right).$$

Asymptotic validity

Theorem (Zhang and Janson (2020))

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).
- Constant in $O(n^{-1/2})$ has complicated dependence on μ and $P_{(Y,X,Z)}$.
- Invariance of the floodgate procedure : e.g. $\mu(x, z) = ax + g(z)$, constant only depends on sign(a) and bivariate distribution of

$$\left(Y, \frac{X - \mathbb{E}\left[X \mid Z\right]}{\sqrt{\mathrm{Var}\left(X - \mathbb{E}\left[X \mid Z\right]\right)}}\right).$$

- Suggests floodgate may be robust to μ and high-dimensionality.

Statistical accuracy

1. through the best element of its equivalent class S_μ in terms of MSE

Statistical accuracy

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

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

1. through the best element of its equivalent class S_μ in terms of MSE

Statistical accuracy

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

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

Theorem (Zhang and Janson (2020))

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

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

1. through the best element of its equivalent class S_μ in terms of MSE

Statistical accuracy

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

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

Theorem (Zhang and Janson (2020))

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

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

Inferential accuracy is directly related to the MSE of μ_n [1]

1. through the best element of its equivalent class S_μ in terms of MSE

Statistical accuracy

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

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

Theorem (Zhang and Janson (2020))

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

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

Inferential accuracy is directly related to the MSE of μ_n [1]

Good predictive performance $\Longrightarrow$ Good inferential accuracy

1. through the best element of its equivalent class S_μ in terms of MSE

Robustness

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

Theorem (Zhang and Janson (2020))

Under moment conditions on Y and noises, for μ_n with well-behaved moments under both the true distribution and the specified one, we have

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

where

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

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

2. When $\mathcal{I} > 0$

Robustness

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

Theorem (Zhang and Janson (2020))

Under moment conditions on Y and noises, for μ_n with well-behaved moments under both the true distribution and the specified one, we have

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

where

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

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

$P_{X|Z}$ is better estimated than $\mathbb{E}\left[Y \mid X, Z\right] \Longrightarrow$[2] Floodgate is robust

2. When $\mathcal{I} > 0$

Application to genomic study of platelet count

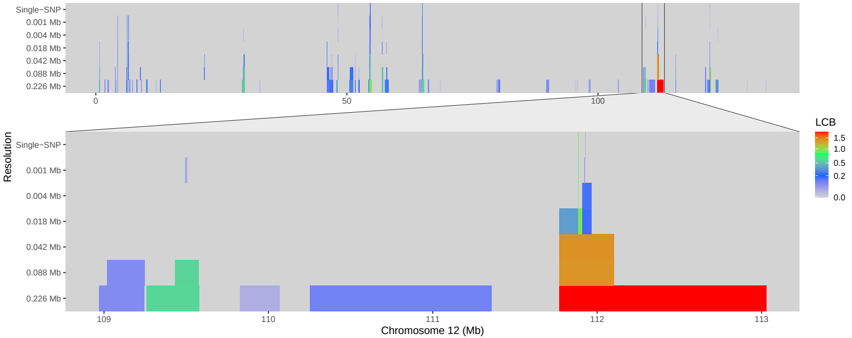

FIGURE – Colored Chicago plot* with the color of each point representing the floodgate LCB for the importance of a group of SNPs on Chromosome 12 in the UK Biobank data at different resolutions (y-axis). Bottom plot shows a zoomed-in region of strong importance.

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

Takeaways

Floodgate : a new inferential approach for variable importance.

- Focus on an interpretable, sensitive and nonparametric measure of variable importance : the mMSE gap.
- Provide valid and robust lower confidence bounds for the mMSE gap.
- Allow flexible regression algorithms, good predictive performance leads to good inferential accuracy.

See more extensions in our paper :

1. Co-sufficient floodgate relaxes the assumptions to only knowing a model for $P_{X|Z}$
2. Floodgate for a different measure of variable importance.
3. Inference on group variable importance.
4. Transporting floodgate inference to a different covariate distribution.
5. Adjusting for multiplicity and selection effects.

Zhang, Lu, and Lucas Janson. "Floodgate : inference for model-free variable importance." arXiv preprint arXiv:2007.01283 (2020).

Azadkia, M. and Chatterjee, S. (2019). A simple measure of conditional dependence. arXiv preprint arXiv :1910.12327.

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.

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.

Fisher, A., Rudin, C., and Dominici, F. (2018). Model class reliance : Variable importance measures for any machine learning model class, from the" rashomon" perspective. arXiv preprint arXiv :1801.01489, 68.

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.

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.

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.

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

Sur, P. and Candès, E. J. (2019). A modern maximum-likelihood theory for high-dimensional logistic regression. Proceedings of the National Academy of Sciences, 116(29) :14516–14525.

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.

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.

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.

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.

Zhao, Q., Sur, P., and Candes, E. J. (2020). The asymptotic distribution of the mle in high-dimensional logistic models : Arbitrary covariance. arXiv preprint arXiv :2001.09351.