Introduction to Envelope Models

Linquan Ma

Department of Statistics
University of Wisconsin-Madison

September 17, 2018

Outline

Regression with multiple responses

- Linear regression model can be written as:

$$\mathbf{y}_{1\times r} = \mathbf{x}_{1\times p}\boldsymbol{\beta}_{p\times r} + \boldsymbol{\epsilon}$$

where the error vector $\boldsymbol{\epsilon} \in \mathbb{R}^r$ is normally distributed with mean $\mathbf{0}$ and unknown parameter $\boldsymbol{\Sigma}$. When $\boldsymbol{\Sigma} > 0$, the model has a total of $pr + r(r+1)/2$ unknown parameters.

Regression with multiple responses

- Linear regression model can be written as:

$$\mathbf{y}_{1\times r} = \mathbf{x}_{1\times p}\boldsymbol{\beta}_{p\times r} + \boldsymbol{\epsilon}$$

where the error vector $\boldsymbol{\epsilon} \in \mathbb{R}^r$ is normally distributed with mean $\mathbf{0}$ and unknown parameter $\boldsymbol{\Sigma}$. When $\boldsymbol{\Sigma} > 0$, the model has a total of $pr + r(r+1)/2$ unknown parameters.

- Suppose we observe dataset $\Big((\mathbf{x}_1, \mathbf{y}_1), ..., (\mathbf{x}_n, \mathbf{y}_n)\Big)$, we can express them in a matrix form:

$$\mathbf{Y}_{n\times r} = \mathbf{X}_{n\times p}\boldsymbol{\beta}_{p\times r} + \boldsymbol{\epsilon}$$

Maximum likelihood estimator

- The log-likelihood of linear model:

$$l(\beta, \Sigma|\mathbf{Y}) = -\frac{n}{2}\det(\Sigma) - \frac{1}{2}(\mathbf{Y} - \mathbf{X}\beta)\Sigma^{-1}(\mathbf{Y} - \mathbf{X}\beta)^T + C$$

Maximum likelihood estimator

- The log-likelihood of linear model:

$$l(\beta, \Sigma|\mathbf{Y}) = -\frac{n}{2}\det(\Sigma) - \frac{1}{2}(\mathbf{Y} - \mathbf{X}\beta)\Sigma^{-1}(\mathbf{Y} - \mathbf{X}\beta)^T + C$$

- Setting partial derivative of β to $\mathbf{0}$, we have

$$\frac{\partial l}{\partial \beta} = \mathbf{X}^T\mathbf{Y}\Sigma^{-1} - \mathbf{X}^T\mathbf{X}\beta\Sigma^{-1} = 0$$

Maximum likelihood estimator

- The log-likelihood of linear model:

$$l(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\mathbf{Y}) = -\frac{n}{2}\det(\boldsymbol{\Sigma}) - \frac{1}{2}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T + C$$

- Setting partial derivative of $\boldsymbol{\beta}$ to $\mathbf{0}$, we have

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \mathbf{X}^T\mathbf{Y}\boldsymbol{\Sigma}^{-1} - \mathbf{X}^T\mathbf{X}\boldsymbol{\beta}\boldsymbol{\Sigma}^{-1} = 0$$

- Since $\boldsymbol{\Sigma}$ is positive definite, we can cancel it on both sides. Hence the MLE of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}}_{MLE} = (\mathbf{X}^T\mathbf{X})^{\dagger}\mathbf{X}^T\mathbf{Y}$$

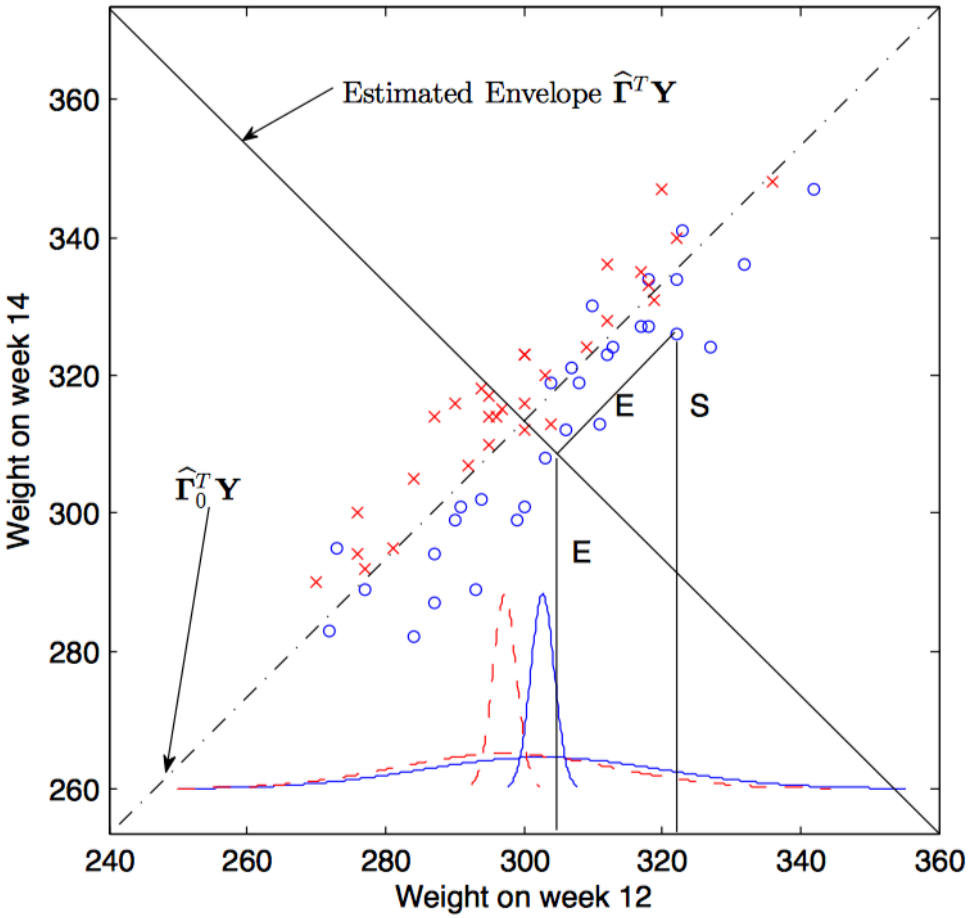
  where $\dagger$ indicates the Moore-Penrose inverse.

Remark

The estimation of $\boldsymbol{\beta}$ does not depend on $\boldsymbol{\Sigma}$ at all.

Outline

Motivation


Estimated Envelope $\widehat{\Gamma}^T\mathbf{Y}$
$\widehat{\Gamma}_0^T\mathbf{Y}$
E
S
E
Weight on week 14
Weight on week 12
360
340
320
300
280
260
240
260
280
300
320
340
360

Motivation

- From the graph, we see that some characteristics of the response vector could be unaffected by changes in the predictors.

Motivation

- From the graph, we see that some characteristics of the response vector could be unaffected by changes in the predictors.
- For the ease of notation later, we slightly change the model as

$$\mathbf{y}_{r\times 1} = \boldsymbol{\beta}_{r\times p}\mathbf{x}_{p\times 1} + \boldsymbol{\epsilon}$$

Motivation

- From the graph, we see that some characteristics of the response vector could be unaffected by changes in the predictors.
- For the ease of notation later, we slightly change the model as

$$\mathbf{y}_{r\times 1} = \boldsymbol{\beta}_{r\times p}\mathbf{x}_{p\times 1} + \boldsymbol{\epsilon}$$

- Consider a subspace $\mathcal{E} \subseteq \mathbb{R}^r$ so that ($\mathbf{P}_{\mathcal{E}}$ = projection onto $\mathcal{E}$, $\mathbf{Q}_{\mathcal{E}} = \mathbf{I} - \mathbf{P}_{\mathcal{E}}$)

$$\mathbf{Q}_{\mathcal{E}}\mathbf{Y}|(\mathbf{X} = \mathbf{x}_1) \sim \mathbf{Q}_{\mathcal{E}}\mathbf{Y}|(\mathbf{X} = \mathbf{x}_2), \forall(\mathbf{x}_1, \mathbf{x}_2) \Longleftrightarrow \operatorname{span}(\boldsymbol{\beta}) \subset \mathcal{E}$$

$$\mathbf{P}_{\mathcal{E}}\mathbf{Y} \perp\!\!\!\perp \mathbf{Q}_{\mathcal{E}}\mathbf{Y}|\mathbf{X} \Longleftrightarrow \boldsymbol{\Sigma} = \mathbf{P}_{\mathcal{E}}\boldsymbol{\Sigma}\mathbf{P}_{\mathcal{E}} + \mathbf{Q}_{\mathcal{E}}\boldsymbol{\Sigma}\mathbf{Q}_{\mathcal{E}}$$

Motivation

- From the graph, we see that some characteristics of the response vector could be unaffected by changes in the predictors.
- For the ease of notation later, we slightly change the model as

$$\mathbf{y}_{r\times 1} = \boldsymbol{\beta}_{r\times p}\mathbf{x}_{p\times 1} + \boldsymbol{\epsilon}$$

- Consider a subspace $\mathcal{E} \subseteq \mathbb{R}^r$ so that ($\mathbf{P}_{\mathcal{E}}$ = projection onto $\mathcal{E}$, $\mathbf{Q}_{\mathcal{E}} = \mathbf{I} - \mathbf{P}_{\mathcal{E}}$)

$$\mathbf{Q}_{\mathcal{E}}\mathbf{Y}|(\mathbf{X} = \mathbf{x}_1) \sim \mathbf{Q}_{\mathcal{E}}\mathbf{Y}|(\mathbf{X} = \mathbf{x}_2), \forall(\mathbf{x}_1, \mathbf{x}_2) \Longleftrightarrow \operatorname{span}(\boldsymbol{\beta}) \subset \mathcal{E}$$

$$\mathbf{P}_{\mathcal{E}}\mathbf{Y} \perp\!\!\!\perp \mathbf{Q}_{\mathcal{E}}\mathbf{Y}|\mathbf{X} \Longleftrightarrow \boldsymbol{\Sigma} = \mathbf{P}_{\mathcal{E}}\boldsymbol{\Sigma}\mathbf{P}_{\mathcal{E}} + \mathbf{Q}_{\mathcal{E}}\boldsymbol{\Sigma}\mathbf{Q}_{\mathcal{E}}$$

- This implies the impact of $\mathbf{X}$ on $\mathbf{Y}$ is concentrated only in $\mathbf{P}_{\mathcal{E}}\mathbf{Y}$. We refer to $\mathbf{P}_{\mathcal{E}}\mathbf{Y}$ and $\mathbf{Q}_{\mathcal{E}}\mathbf{Y}$ informally as material and immaterial part of $\mathbf{Y}$.

Outline

Invariant and reducing subspaces

- A subspace $\mathcal{R}$ of $\mathbb{R}^r$ is an invariant subspace of $\mathbf{M} \in \mathbb{R}^{r \times r}$ if $\mathbf{M}\mathcal{R} \subseteq \mathcal{R}$.
- $\mathcal{R}$ is a reducing subspace of $\mathbf{M}$, if, in addition, $\mathbf{M}\mathcal{R}^{\perp} \subseteq \mathcal{R}^{\perp}$.

Invariant and reducing subspaces

- A subspace $\mathcal{R}$ of $\mathbb{R}^r$ is an invariant subspace of $\mathbf{M} \in \mathbb{R}^{r \times r}$ if $\mathbf{M}\mathcal{R} \subseteq \mathcal{R}$.
- $\mathcal{R}$ is a reducing subspace of $\mathbf{M}$, if, in addition, $\mathbf{M}\mathcal{R}^{\perp} \subseteq \mathcal{R}^{\perp}$.

Proposition 1.

$\mathcal{R}$ reduces $\mathbf{M} \in \mathbb{R}^{r \times r}$ if and only if $\mathbf{M}$ can be written in the form:

$$\mathbf{M} = \mathbf{P}_{\mathcal{R}}\mathbf{M}\mathbf{P}_{\mathcal{R}} + \mathbf{Q}_{\mathcal{R}}\mathbf{M}\mathbf{Q}_{\mathcal{R}}$$

Invariant and reducing subspaces

- A subspace $\mathcal{R}$ of $\mathbb{R}^r$ is an invariant subspace of $\mathbf{M} \in \mathbb{R}^{r\times r}$ if $\mathbf{M}\mathcal{R} \subseteq \mathcal{R}$.
- $\mathcal{R}$ is a reducing subspace of $\mathbf{M}$, if, in addition, $\mathbf{M}\mathcal{R}^\perp \subseteq \mathcal{R}^\perp$.

Proposition 1.

$\mathcal{R}$ reduces $\mathbf{M} \in \mathbb{R}^{r\times r}$ if and only if $\mathbf{M}$ can be written in the form:

$$\mathbf{M} = \mathbf{P}_\mathcal{R}\mathbf{M}\mathbf{P}_\mathcal{R} + \mathbf{Q}_\mathcal{R}\mathbf{M}\mathbf{Q}_\mathcal{R}$$

Proposition 2.

Assume that $\mathbf{M}_{r\times r}$ is symmetric and has $q \leq r$ distinct eigenvalues, and let $\mathbf{P}_i, i = 1, ..., q$, indicate the projection onto the corresponding eigenspaces. Then $\mathcal{R}$ reduces $\mathbf{M}$ if and only if $\mathcal{R} = \oplus_{i=1}^q \mathbf{P}_i\mathcal{R}$

M-envelopes

Definition 1. (Cook et al. (2010))

Let $\mathbf{M}$ be a symmetric matrix, and let $\mathcal{S} \subseteq \mathrm{span}(\mathbf{M})$. The $\mathbf{M}$-envelope of $\mathcal{S}$, to be written as $\mathcal{E}_{\mathbf{M}}(\mathcal{S})$, is the intersection of all reducing subspaces of $\mathbf{M}$ that contains $\mathcal{S}$.

M-envelopes

Definition 1. (Cook et al. (2010))

Let $\mathbf{M}$ be a symmetric matrix, and let $\mathcal{S} \subseteq \text{span}(\mathbf{M})$. The $\mathbf{M}$-envelope of $\mathcal{S}$, to be written as $\mathcal{E}_{\mathbf{M}}(\mathcal{S})$, is the intersection of all reducing subspaces of $\mathbf{M}$ that contains $\mathcal{S}$.

Proposition 3.

$$\mathcal{E}_{\mathbf{M}}(\mathcal{S}) = \oplus_{i=1}^{q} \mathbf{P}_i \mathcal{S}$$

where $\mathbf{P}_i$ is the projection matrix to the i-th eigenspace of $\mathbf{M}$.

Outline

- Now we want to refine the regression model by using an envelope to connect $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$.
- Let the columns of the semi-orthogonal matrices $\boldsymbol{\Gamma} \in \mathbb{R}^{u \times r}$ be the base of $\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B})$, and $\boldsymbol{\Gamma}_0 \in \mathbb{R}^{(r-u) \times r}$ be its orthogonal complements. Then, there exist an $\boldsymbol{\eta} \in \mathbb{R}^{u \times p}$ such that $\boldsymbol{\beta} = \boldsymbol{\Gamma}\boldsymbol{\eta}$ where $\boldsymbol{\eta}$ contains the coordinates of $\boldsymbol{\beta}$ relative to $\boldsymbol{\Gamma}$.

Envelope model:

$$\mathbf{Y} = \boldsymbol{\Gamma}\boldsymbol{\eta}\mathbf{X} + \boldsymbol{\epsilon}, \quad \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T$$

- Estimation of the parameters can be carried out by maximum likelihood with u determined by AIC, BIC or other methods.

Maximum likelihood estimation

- The estimated envelope $\hat{\mathcal{E}}_{\mathbf{\Sigma}}(\mathcal{B})$ can be represented as

$$\hat{\mathcal{E}}_{\mathbf{\Sigma}}(\mathcal{B}) = \arg\min_{\boldsymbol{\delta}} (\log |\mathbf{P}_{\boldsymbol{\delta}} \mathbf{S}_{\mathbf{Y}|\mathbf{X}} \mathbf{P}_{\boldsymbol{\delta}}|_0 + \log |\mathbf{Q}_{\boldsymbol{\delta}} \mathbf{S}_{\mathbf{Y}} \mathbf{Q}_{\boldsymbol{\delta}}|_0)$$

where $|\cdot|_0$ means the product of the non-zero eigenvalues and $\boldsymbol{\delta}$ is a u-dim subspace of $\mathbb{R}^r$. $\mathbf{S}_{\mathbf{Y}|\mathbf{X}}$ and $\mathbf{S}_{\mathbf{Y}}$ are the sample version of $\mathbf{\Sigma}$ and $\mathrm{var}(\mathbf{Y})$.

Maximum likelihood estimation

- The estimated envelope $\hat{\mathcal{E}}_{\mathbf{\Sigma}}(\mathcal{B})$ can be represented as

$$\hat{\mathcal{E}}_{\mathbf{\Sigma}}(\mathcal{B}) = \arg\min_{\boldsymbol{\delta}} (\log |\mathbf{P}_{\boldsymbol{\delta}} \mathbf{S}_{\mathbf{Y}|\mathbf{X}} \mathbf{P}_{\boldsymbol{\delta}}|_0 + \log |\mathbf{Q}_{\boldsymbol{\delta}} \mathbf{S}_{\mathbf{Y}} \mathbf{Q}_{\boldsymbol{\delta}}|_0)$$

  where $|\cdot|_0$ means the product of the non-zero eigenvalues and $\boldsymbol{\delta}$ is a u-dim subspace of $\mathbb{R}^r$. $\mathbf{S}_{\mathbf{Y}|\mathbf{X}}$ and $\mathbf{S}_{\mathbf{Y}}$ are the sample version of $\mathbf{\Sigma}$ and $\mathrm{var}(\mathbf{Y})$.
- $\hat{\mathcal{E}}_{\mathbf{\Sigma}}(\mathcal{B})$ can be estimated through a 1-D algorithm proposed by Cook and Zhang (2016).

Maximum likelihood estimation

- The estimated envelope $\hat{\mathcal{E}}_{\mathbf{\Sigma}}(\mathcal{B})$ can be represented as

$$\hat{\mathcal{E}}_{\mathbf{\Sigma}}(\mathcal{B}) = \arg\min_{\boldsymbol{\delta}}(\log|\mathbf{P}_{\boldsymbol{\delta}}\mathbf{S}_{\mathbf{Y}|\mathbf{X}}\mathbf{P}_{\boldsymbol{\delta}}|_0 + \log|\mathbf{Q}_{\boldsymbol{\delta}}\mathbf{S}_{\mathbf{Y}}\mathbf{Q}_{\boldsymbol{\delta}}|_0)$$

  where $|\cdot|_0$ means the product of the non-zero eigenvalues and $\boldsymbol{\delta}$ is a u-dim subspace of $\mathbb{R}^r$. $\mathbf{S}_{\mathbf{Y}|\mathbf{X}}$ and $\mathbf{S}_{\mathbf{Y}}$ are the sample version of $\mathbf{\Sigma}$ and $\mathrm{var}(\mathbf{Y})$.
- $\hat{\mathcal{E}}_{\mathbf{\Sigma}}(\mathcal{B})$ can be estimated through a 1-D algorithm proposed by Cook and Zhang (2016).
- Let $\hat{\mathbf{\Gamma}}$ be a basis for $\hat{\mathcal{E}}_{\mathbf{\Sigma}}(\mathcal{B})$, estimators of other parameters are:
  - $\hat{\boldsymbol{\beta}}_{env} = \mathbf{P}_{\hat{\mathbf{\Gamma}}}\hat{\boldsymbol{\beta}}_{ols}$, which is $\sqrt{n}$ consistent and asymptotically normal.
  - $\hat{\mathbf{\Sigma}} = \mathbf{P}_{\hat{\mathbf{\Gamma}}}\mathbf{S}_{\mathbf{Y}|\mathbf{X}}\mathbf{P}_{\hat{\mathbf{\Gamma}}} + \mathbf{Q}_{\hat{\mathbf{\Gamma}}}\mathbf{S}_{\mathbf{Y}|\mathbf{X}}\mathbf{Q}_{\hat{\mathbf{\Gamma}}}$

Efficiency gain

Proposition 4. (Cook et al. (2010))

$$\mathrm{avar}(\sqrt{n}\mathrm{vec}[\hat{\boldsymbol{\beta}}_{env}]) \leq \mathrm{avar}(\sqrt{n}\mathrm{vec}[\hat{\boldsymbol{\beta}}_{std}])$$

where $\mathrm{avar}(\cdot)$ stands for the asymptotic covariance and $\mathrm{vec}(\cdot)$ stands for the vectorization of a matrix.

Efficiency gain

Proposition 4. (Cook et al. (2010))

$$\mathrm{avar}(\sqrt{n}\mathrm{vec}[\hat{\beta}_{env}]) \leq \mathrm{avar}(\sqrt{n}\mathrm{vec}[\hat{\beta}_{std}])$$

where $\mathrm{avar}(\cdot)$ stands for the asymptotic covariance and $\mathrm{vec}(\cdot)$ stands for the vectorization of a matrix.

Remarks

- Envelope methods are never worse than standard methods.
- Envelope methods will provide the most gain in efficiency when $\mathcal{E}_{\Sigma}(\mathcal{B})$ can be constructed from eigenspaces of Σ with relatively small eigenvalues.

Outline

Example

Illustration of an example using R:

```
require(envlp)
set.seed(0411)
num = 200
env_dim <- 5
p = 5
q = 20
sq_err_env <- NULL
sq_err_std <- NULL
for (i in 1:10) {
  GAMMA <- matrix(runif(env_dim * q), nrow = q)
  beta0 <- matrix(runif(p * q, -10, 10), nrow = p)
  beta <- beta0 %*% P(GAMMA)
  Omega <- 0.1 * diag(nrow(GAMMA))
  Omega0 <- 1000 * diag(nrow(GAMMA))
  Sigma_y <- P(GAMMA) %*% Omega %*% P(GAMMA) + Q(GAMMA) %*% Omega0 %*% Q(GAMMA)
  A <- matrix(runif(p ^ 2, -10, 10), nrow = p)
  mu_x <- runif(p, -10, 10)
  Sigma_x <- A %*% t(A)
  X <- mvrnorm(num, mu_x, Sigma_x)
  Y <- X %*% beta + mvrnorm(num, rep(0, q), Sigma_y)
```

Example

```
  u = u.env(X, Y)$u.bic
  env_beta <- t(env(X, Y, u)$beta)
  std_beta <- solve(crossprod(X)) %*% crossprod(X, Y)
  sum((std_beta - beta)^2)
  sq_err_env <- c(sum((env_beta - beta)^2), sq_err_env)
  sq_err_std <- c(sum((std_beta - beta)^2), sq_err_std)
}
mean(sq_err_env)
mean(sq_err_std)
```

Example

```
  u = u.env(X, Y)$u.bic
  env_beta <- t(env(X, Y, u)$beta)
  std_beta <- solve(crossprod(X)) %*% crossprod(X, Y)
  sum((std_beta - beta)^2)
  sq_err_env <- c(sum((env_beta - beta)^2), sq_err_env)
  sq_err_std <- c(sum((std_beta - beta)^2), sq_err_std)
}
mean(sq_err_env)
mean(sq_err_std)
```

- Output: 0.007500058 14.07488

Outline

Envelope methods with ignorable missing data

Linquan Ma [1] Lan Liu[2] Wei Yang[3]

[1]Department of Statistics, University of Wisconsin-Madison

[2]School of Statistics, University of Minnesota

[3]Perelman School of Medicine, University of Pennsylvania

September 17, 2018

Motivation and preliminary

- In big data setting where large amount of responses and predictors are collected, it is common that the responses or the predictors or both suffer from missingness.

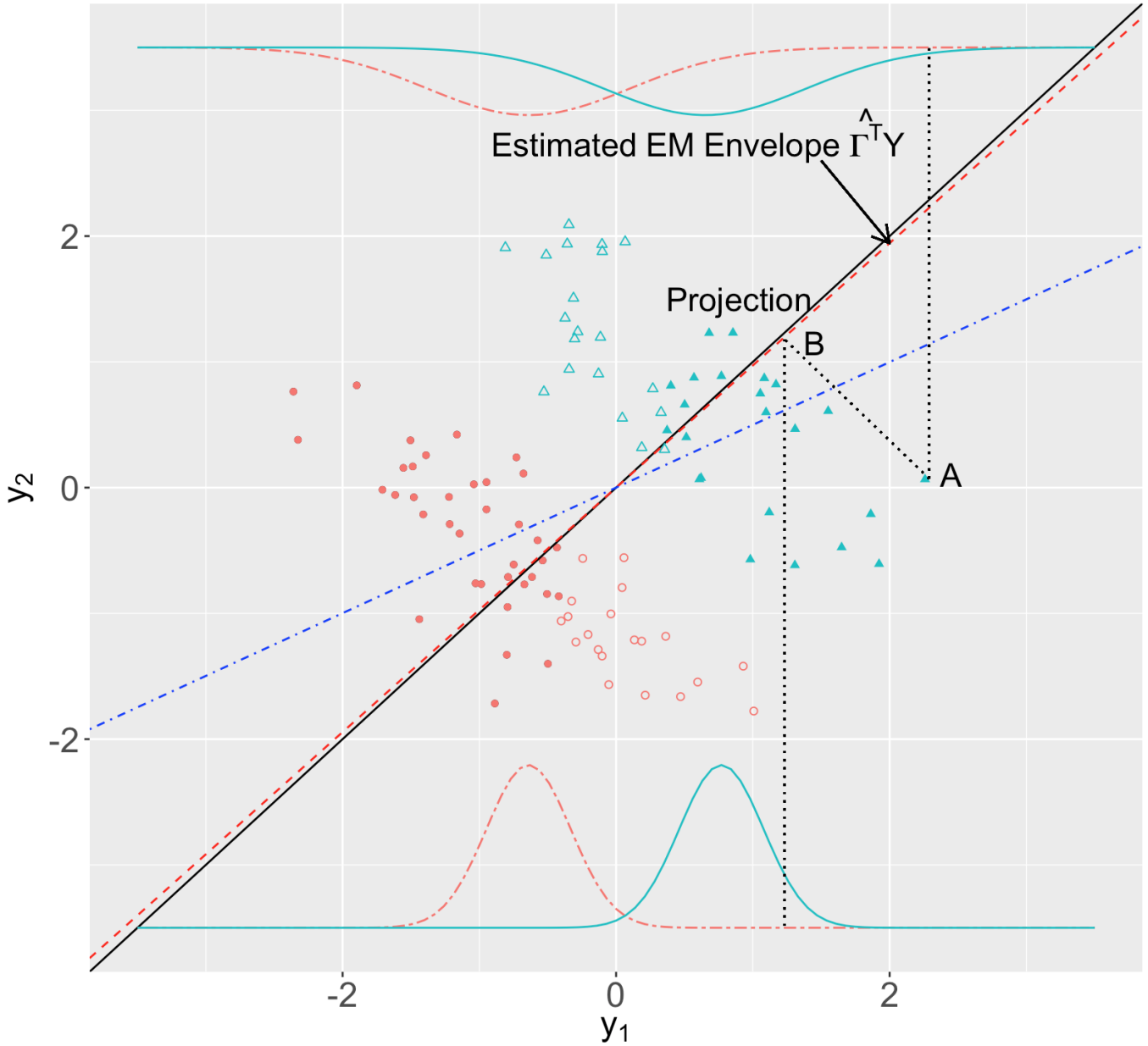
Motivation and preliminary

- In big data setting where large amount of responses and predictors are collected, it is common that the responses or the predictors or both suffer from missingness.
- Types of missing data:
  - MCAR: Missingness is independent both of observed and unobserved data.
  - MAR: Missingness is independent of unobserved data.
  - MNAR: is data that is not MAR, is also known as nonignorable missing.
- The former two mechanisms are addressed as ignorable missing since the missing information could be partially recovered using the observed data.

Motivation and preliminary

- In big data setting where large amount of responses and predictors are collected, it is common that the responses or the predictors or both suffer from missingness.
- Types of missing data:
  - MCAR: Missingness is independent both of observed and unobserved data.
  - MAR: Missingness is independent of unobserved data.
  - MNAR: is data that is not MAR, is also known as nonignorable missing.
- The former two mechanisms are addressed as ignorable missing since the missing information could be partially recovered using the observed data.
- Complete case analysis will introduce bias even if data is ignorable missing.

Motivation and preliminary



EM algorithm

$$l(\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{Y}) = \log(f_{y|x}(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\theta})) + \log(f_x(\boldsymbol{X}|\boldsymbol{\theta}))$$

$$= \sum_{i=1}^{n}\Big(-\frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\boldsymbol{y_i}-\boldsymbol{x_i}\boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y_i}-\boldsymbol{x_i}\boldsymbol{\beta})' + \log(f_x(\boldsymbol{x}_i|\boldsymbol{\rho}))\Big) + C$$

EM algorithm

$$l(\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{Y}) = \log(f_{y|x}(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\theta})) + \log(f_x(\boldsymbol{X}|\boldsymbol{\theta}))$$

$$= \sum_{i=1}^{n}\Big(-\frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\boldsymbol{y_i}-\boldsymbol{x_i}\boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y_i}-\boldsymbol{x_i}\boldsymbol{\beta})' + \log(f_x(\boldsymbol{x}_i|\boldsymbol{\rho}))\Big) + C$$

E-Step: Let $\boldsymbol{\theta} = (\boldsymbol{\eta}, \boldsymbol{\Gamma}, \boldsymbol{\Omega}_0, \boldsymbol{\Omega}_1, \boldsymbol{\rho})$ and let $\boldsymbol{\theta}_t$ denote the current estimate of the parameter $\boldsymbol{\theta}$. The E-step evaluate the expectation of full data likelihood given the current parameter estimates as

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}_t) = E[l_{full}(\boldsymbol{\theta}|\mathbf{X},\mathbf{Y})|\boldsymbol{D}_{obs},\boldsymbol{\theta}_t] = \int l_{full}(\boldsymbol{\theta}|L) f(\boldsymbol{D}_{mis}|\boldsymbol{D}_{obs},\boldsymbol{\theta}_t) d\boldsymbol{D}_{mis}.$$

EM algorithm

$$l(\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{Y}) = \log(f_{y|x}(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\theta})) + \log(f_x(\boldsymbol{X}|\boldsymbol{\theta}))$$

$$= \sum_{i=1}^{n}\Big(-\frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\boldsymbol{y_i}-\boldsymbol{x_i}\boldsymbol{\beta})\boldsymbol{\Sigma}^{-1}(\boldsymbol{y_i}-\boldsymbol{x_i}\boldsymbol{\beta})' + \log(f_x(\boldsymbol{x}_i|\boldsymbol{\rho}))\Big) + C$$

E-Step: Let $\boldsymbol{\theta} = (\boldsymbol{\eta}, \boldsymbol{\Gamma}, \boldsymbol{\Omega}_0, \boldsymbol{\Omega}_1, \boldsymbol{\rho})$ and let $\boldsymbol{\theta}_t$ denote the current estimate of the parameter $\boldsymbol{\theta}$. The E-step evaluate the expectation of full data likelihood given the current parameter estimates as

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}_t) = E[l_{full}(\boldsymbol{\theta}|\mathbf{X},\mathbf{Y})|\boldsymbol{D}_{obs},\boldsymbol{\theta}_t] = \int l_{full}(\boldsymbol{\theta}|L) f(\boldsymbol{D}_{mis}|\boldsymbol{D}_{obs},\boldsymbol{\theta}_t) d\boldsymbol{D}_{mis}.$$

M-Step: The M-step computes $\boldsymbol{\theta}^{(t+1)}$ by maximizing the expected log-likelihood obtained in the E-step:

$$Q(\boldsymbol{\theta}^{(t+1)}|\boldsymbol{\theta}_t) \geq Q(\boldsymbol{\theta}|\boldsymbol{\theta}_t), \qquad \text{for all } \boldsymbol{\theta}.$$

We iterate the E- and M-steps until convergence.

Remarks

- Since solving for envelope involves reparametrization of the covariance matrix, i.e. $\boldsymbol{\Sigma} = \mathbf{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Sigma}\mathbf{P}_{\boldsymbol{\Gamma}} + \mathbf{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Sigma}\mathbf{Q}_{\boldsymbol{\Gamma}}$, hence, it is non-trivial to combine EM algorithm and envelope moedels.

Remarks

- Since solving for envelope involves reparametrization of the covariance matrix, i.e. $\boldsymbol{\Sigma} = \mathbf{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Sigma}\mathbf{P}_{\boldsymbol{\Gamma}} + \mathbf{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Sigma}\mathbf{Q}_{\boldsymbol{\Gamma}}$, hence, it is non-trivial to combine EM algorithm and envelope moedels.
- In E-step, we need to calculate the conditional expectations:
  $\boldsymbol{A}_{i1,t} = \mathbb{E}(\boldsymbol{y}_i'\boldsymbol{y}_i|\theta_t, \boldsymbol{D}_{obs})$, $\boldsymbol{A}_{i2,t} = \mathbb{E}(\boldsymbol{y}_i'\boldsymbol{x}_i|\theta_t, \boldsymbol{D}_{obs})$,
  $\boldsymbol{A}_{i3,t} = \mathbb{E}(\boldsymbol{x}_i'\boldsymbol{x}_i|\theta_t, \boldsymbol{D}_{obs})$, $\boldsymbol{A}_{i4,t} = \mathbb{E}(\boldsymbol{x}_i'|\theta, \boldsymbol{D}_{obs})$,
  Denote $\boldsymbol{A}_{j,t} = \sum_{i=1}^{n} \boldsymbol{A}_{ij,t}$, $j = 1, 2, 3, 4$

Remarks

- Since solving for envelope involves reparametrization of the covariance matrix, i.e. $\Sigma = \mathbf{P}_{\mathbf{\Gamma}}\Sigma\mathbf{P}_{\mathbf{\Gamma}} + \mathbf{Q}_{\mathbf{\Gamma}}\Sigma\mathbf{Q}_{\mathbf{\Gamma}}$, hence, it is non-trivial to combine EM algorithm and envelope moedels.
- In E-step, we need to calculate the conditional expectations:
  $\boldsymbol{A}_{i1,t} = \mathbb{E}(\boldsymbol{y}_i'\boldsymbol{y}_i|\theta_t, \boldsymbol{D}_{obs})$, $\boldsymbol{A}_{i2,t} = \mathbb{E}(\boldsymbol{y}_i'\boldsymbol{x}_i|\theta_t, \boldsymbol{D}_{obs})$,
  $\boldsymbol{A}_{i3,t} = \mathbb{E}(\boldsymbol{x}_i'\boldsymbol{x}_i|\theta_t, \boldsymbol{D}_{obs})$, $\boldsymbol{A}_{i4,t} = \mathbb{E}(\boldsymbol{x}_i'|\theta, \boldsymbol{D}_{obs})$,
  Denote $\boldsymbol{A}_{j,t} = \sum_{i=1}^{n} \boldsymbol{A}_{ij,t}$, $j = 1, 2, 3, 4$

Parameter updates

- Using the 1-D algorithm proposed by Cook and Zhang (2016) to estimate $\mathbf{\Gamma}_t$ based on θ_t.
- $\boldsymbol{\Sigma}_{1,t+1} = \frac{1}{n}\boldsymbol{P}_{\boldsymbol{\Gamma}_t}(\boldsymbol{A}_{1,t} - \boldsymbol{A}_{2,t}\boldsymbol{A}_{3,t}^{-1}\boldsymbol{A}_{2,t}')\boldsymbol{P}_{\boldsymbol{\Gamma}_t}$;
  $\boldsymbol{\rho}_{t+1} = \arg\max_{\boldsymbol{\rho}\in\Theta} \mathbb{E}(\log(f_x(\boldsymbol{x}_i|\boldsymbol{\rho}))|\boldsymbol{D}_{obs}, \theta_t)$;
  $\boldsymbol{\beta}_{t+1} = \boldsymbol{A}_{3,t}^{-1}\boldsymbol{A}_{2,t}'\boldsymbol{P}_{\boldsymbol{\Sigma}_{1,t+1}}$;
  $\boldsymbol{\Sigma}_{t+1} = \boldsymbol{\Sigma}_{1,t+1} + \frac{1}{n}\boldsymbol{Q}_{\boldsymbol{\Gamma}_t}\boldsymbol{A}_{1,t}\boldsymbol{Q}_{\boldsymbol{\Gamma}_t}$

Algorithm 1: The EM envelope algorithm

Data: n observations with MAR p predictors and q responses .

Result: Finding the estimator $\hat{\beta}_{em\_env}$.

for $k = 1, 2, \ldots, q$ **do**

initialization: $\boldsymbol{\Sigma}_t = \boldsymbol{I}_q$, $\beta_t = \mathbf{0}$, $\rho_t = \rho_0$, $\theta_t = (\boldsymbol{\Sigma}_t, \beta_t, \rho_t)$, $\Delta = 1$;

while $\Delta > \delta$ **do**

1. Calculate $\boldsymbol{A}_{1,t} = \sum_{i=1}^{n} \boldsymbol{A}_{i1,t}$, $\boldsymbol{A}_{2,t} = \sum_{i=1}^{n} \boldsymbol{A}_{i2,t}$, $\boldsymbol{A}_{3,t} = \sum_{i=1}^{n} \boldsymbol{A}_{i3,t}$ based on θ_t;
2. Using 1-D algorithm to calculate $\boldsymbol{\Gamma}_t$, then $\boldsymbol{\Sigma}_{1,t+1} = \frac{1}{n}\boldsymbol{P}_{\boldsymbol{\Gamma}_t}(\boldsymbol{A}_{1,t} - \boldsymbol{A}_{2,t}\boldsymbol{A}_{3,t}^{-1}\boldsymbol{A}_{2,t}')\boldsymbol{P}_{\boldsymbol{\Gamma}_t}$;
3. Update: $\rho_{t+1} = \arg\max_{\rho \in \Theta} \mathbb{E}(\log(f_x(\boldsymbol{x}_i|\rho))|\boldsymbol{D}_{obs}, \theta_t)$, $\beta_{t+1} = \boldsymbol{A}_{3,t}^{-1}\boldsymbol{A}_{2,t}'\boldsymbol{P}_{\boldsymbol{\Sigma}_{1,t+1}}$, $\boldsymbol{\Sigma}_{t+1} = \boldsymbol{\Sigma}_{1,t+1} + \frac{1}{n}\boldsymbol{Q}_{\boldsymbol{\Gamma}_t}\boldsymbol{A}_{1,t}\boldsymbol{Q}_{\boldsymbol{\Gamma}_t}$;
4. Set $\Delta = \|\beta_{t+1} - \beta_t\|_1$, $\theta_t = (\boldsymbol{\Sigma}_{t+1}, \beta_{t+1}, \rho_{t+1})$;

end

$\mathrm{BIC}_k = -2Q(\hat{\theta}|\hat{\theta}) + pu \log n$, $\hat{\beta}_k = \beta_{t+1}$

end

Find u such that BIC_k is minimum. The corresponding $\hat{\beta}_u$ is the EM envelope estimator.

EM envelope algorithm

Proposition 1.

Denote $\hat{\beta}_{env}$ as the estimator by EM envelope algorithm, and $\hat{\beta}_{std}$ as the estimator by standard EM algorithm. Then

$$\sqrt{n}(\mathrm{vec}(\hat{\beta}_{env}) - \mathrm{vec}(\beta)) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_{env}),$$
$$\sqrt{n}(\mathrm{vec}(\hat{\beta}_{std}) - \mathrm{vec}(\beta)) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_{std}),$$
$$\text{where } \mathbf{V}_{env} \leq \mathbf{V}_{std}.$$

Simulations

We run simulations to compare four different methods: standard complete case analysis, complete case envelope, standard EM, and EM envelope.

- Suppose $\mathbf{X}_n \sim N(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$, $\mathbf{Y}_n \sim N(\mathbf{X}_n\boldsymbol{\beta}, \boldsymbol{\Sigma}_\epsilon)$
- Generate $n = 500$ samples, each has $q = 20$ responses, and $p = 5$ covariates with envelope dimension $u = 3$.
- Generate $\boldsymbol{\Gamma}$, $\boldsymbol{\beta}_0$, $\boldsymbol{\mu}_x$, $\boldsymbol{\Sigma}_x$ at random, and set $\boldsymbol{\beta} = \boldsymbol{\beta}_0\mathbf{P}_{\boldsymbol{\Gamma}}$.
- $\boldsymbol{\Sigma}_\epsilon = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}' + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0'$. Fix $\boldsymbol{\Omega} = 0.1\mathbf{I}_q$. We run two simulations when setting $\boldsymbol{\Omega}_0 = 1000\mathbf{I}_q$ and $\boldsymbol{\Omega}_0 = 10\mathbf{I}_q$.

Results

Table: Summary of MSE when $\boldsymbol{\Omega}_0 = 1000\boldsymbol{I}_q$

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
$\hat{\boldsymbol{\beta}}_{em\_env}$	1.64e-05	3.58e-05	4.44e-05	1.03e-03	5.70e-05	8.66e-02
$\hat{\boldsymbol{\beta}}_{comp\_env}$	3.28e-04	8.16e-04	1.27e-03	1.47e-02	2.52e-03	4.33
$\hat{\boldsymbol{\beta}}_{em\_std}$	2.37e-02	4.41e-02	5.34e-02	5.47e-02	6.38e-02	0.12
$\hat{\boldsymbol{\beta}}_{comp\_std}$	0.82	3.43	4.95	79.3	9.79	3.64e+04

Table: Summary of MSE when $\boldsymbol{\Omega}_0 = 10\boldsymbol{I}_q$

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
$\hat{\boldsymbol{\beta}}_{em\_env}$	4.54e-05	9.08e-05	1.06e-04	1.36e-04	1.25e-04	1.05e-03
$\hat{\boldsymbol{\beta}}_{comp\_env}$	1.69e-03	4.50e-03	6.16e-03	0.590	1.24e-02	3.95e+02
$\hat{\boldsymbol{\beta}}_{em\_std}$	2.17e-04	4.52e-04	5.42e-04	5.62e-04	6.49e-04	1.34e-03
$\hat{\boldsymbol{\beta}}_{comp\_std}$	1.14e-02	3.42e-02	5.06e-02	6.02	9.55e-02	4.60e+03

Real data analysis

We applied our proposed method to a Chronic Renal Insufficiency Cohort (CRIC) study.

Real data analysis

We applied our proposed method to a Chronic Renal Insufficiency Cohort (CRIC) study.

More about the data set

- The CRIC study recruited 3939 participants from April 8, 2003 through September 3, 2008 and continued through March 31, 2013 (Feldman et al., 2003).

Real data analysis

We applied our proposed method to a Chronic Renal Insufficiency Cohort (CRIC) study.

More about the data set

- The CRIC study recruited 3939 participants from April 8, 2003 through September 3, 2008 and continued through March 31, 2013 (Feldman et al., 2003).
- The study cohort is racially and ethnically diverse group aged from 21 to 74 years with mild to moderate chronic kidney disease (CKD).

Real data analysis

We applied our proposed method to a Chronic Renal Insufficiency Cohort (CRIC) study.

More about the data set

- The CRIC study recruited 3939 participants from April 8, 2003 through September 3, 2008 and continued through March 31, 2013 (Feldman et al., 2003).
- The study cohort is racially and ethnically diverse group aged from 21 to 74 years with mild to moderate chronic kidney disease (CKD).
- It is of interest to investigate the difference in the distributions of baseline biomarkers among patients who develop end-stage renal diseases (ESRD) or not.

Real data analysis

We applied our proposed method to a Chronic Renal Insufficiency Cohort (CRIC) study.

More about the data set

- The CRIC study recruited 3939 participants from April 8, 2003 through September 3, 2008 and continued through March 31, 2013 (Feldman et al., 2003).
- The study cohort is racially and ethnically diverse group aged from 21 to 74 years with mild to moderate chronic kidney disease (CKD).
- It is of interest to investigate the difference in the distributions of baseline biomarkers among patients who develop end-stage renal diseases (ESRD) or not.

Real data analysis

In the regression setting

- Predictors: ESRD status, gender, age, race, systolic and diastolic blood pressures, and hemoglobin from CBC lab data. ($p = 10$)

Real data analysis

In the regression setting

- Predictors: ESRD status, gender, age, race, systolic and diastolic blood pressures, and hemoglobin from CBC lab data. ($p = 10$)
- Responses: Biomarkers, which are urine albumin, urine creatinine, high sensitivity C-reactive protein (HS_CRP), brain natriuretic peptide (BNP), chemokine ligand 12 (CXCL12), fetuin A, fractalkine, myeloperoxidase (MPO), neutrophil gelatinase associated lipocalin (NGAL), fibrinogen, troponini, urine calcium, urine sodium, urine potassium, urine phosphate, high sensitive troponin T (TNTHS), aldosterone, C-peptide, insulin value, total parathyroid hormone (Total PTH), CO2, 24-hour urine protein, estimated glomerular filtration rate. ($q = 23$)
- All the biomarkers have some missingness ranging from 0% to 6%.

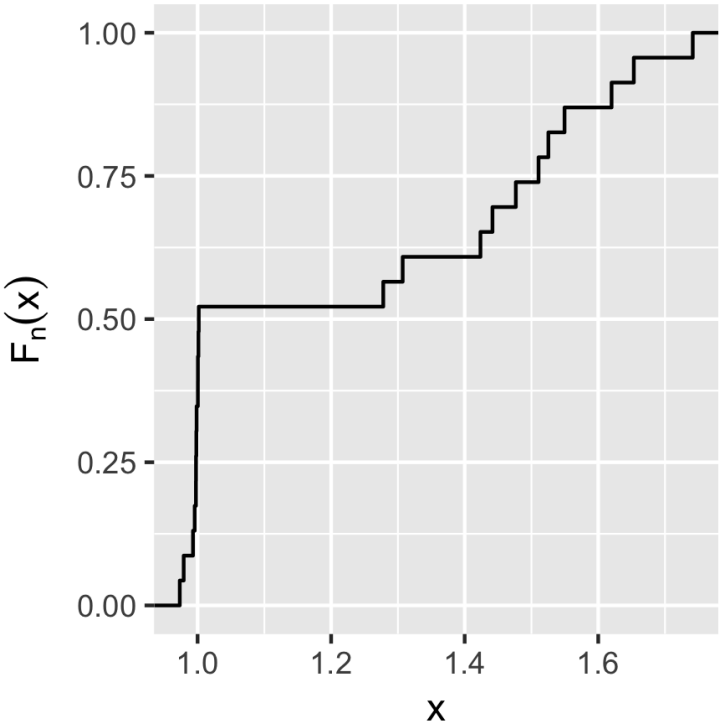
Real data analysis

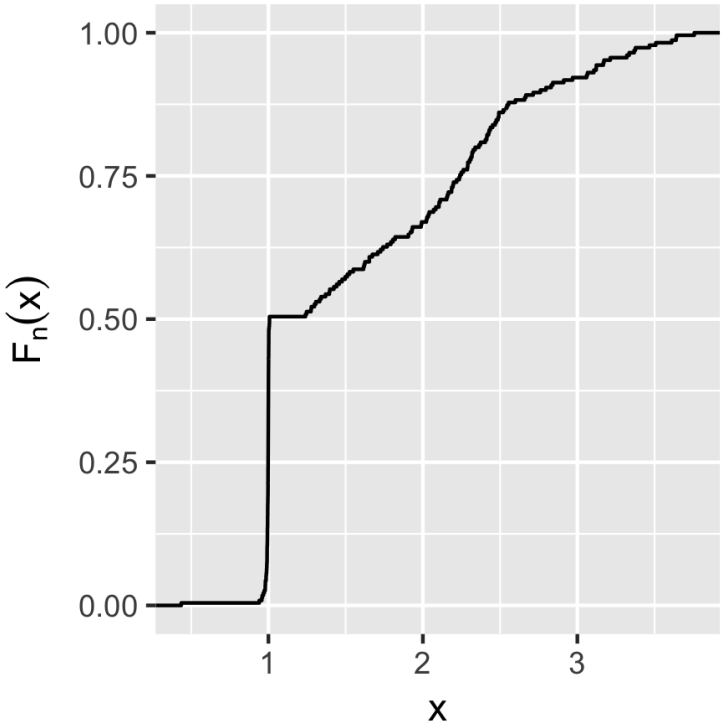
In the regression setting

- Predictors: ESRD status, gender, age, race, systolic and diastolic blood pressures, and hemoglobin from CBC lab data. ($p = 10$)
- Responses: Biomarkers, which are urine albumin, urine creatinine, high sensitivity C-reactive protein (HS_CRP), brain natriuretic peptide (BNP), chemokine ligand 12 (CXCL12), fetuin A, fractalkine, myeloperoxidase (MPO), neutrophil gelatinase associated lipocalin (NGAL), fibrinogen, troponini, urine calcium, urine sodium, urine potassium, urine phosphate, high sensitive troponin T (TNTHS), aldosterone, C-peptide, insulin value, total parathyroid hormone (Total PTH), CO2, 24-hour urine protein, estimated glomerular filtration rate. ($q = 23$)
- All the biomarkers have some missingness ranging from 0% to 6%.
- We compared EM envelope and the standard EM to examine their performance.

Real data analysis

- Using BIC, we chose envelope dimension u = 15.
- These two methods found the same set of biomarkers significant. However, using bootstrap method, the standard error of regression parameter are usually smaller when using EM envelope.





Real data analysis

- It is found in the literature that although many novel biomarkers are found to be marginally significant associate with ESRD status, such association is muted after adjusting for glomerular filtration rate (GFR) and the amount of urine protein excreted in 24 hours.

Real data analysis

- It is found in the literature that although many novel biomarkers are found to be marginally significant associate with ESRD status, such association is muted after adjusting for glomerular filtration rate (GFR) and the amount of urine protein excreted in 24 hours.
- Thus, in the subsequent analysis, we use the two variables as predictors rather than outcomes.

Real data analysis

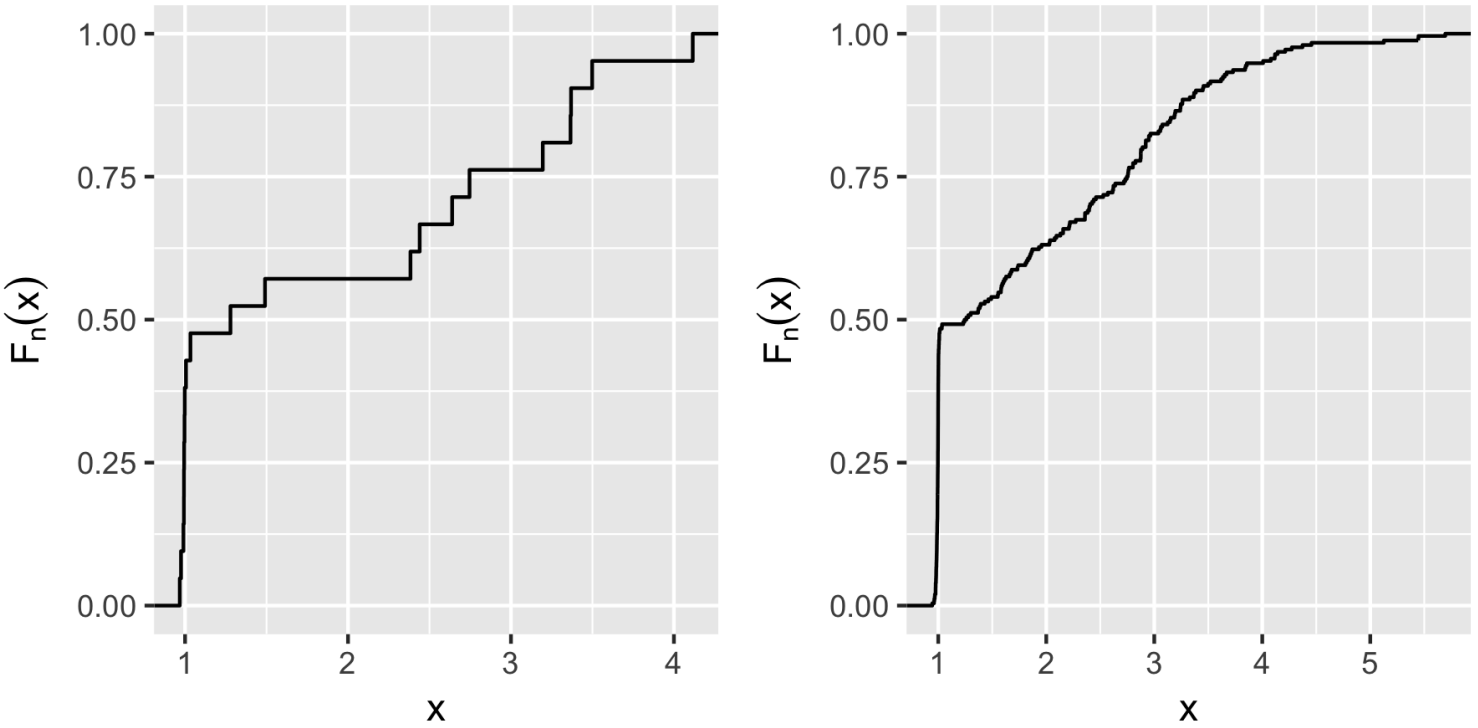
- It is found in the literature that although many novel biomarkers are found to be marginally significant associate with ESRD status, such association is muted after adjusting for glomerular filtration rate (GFR) and the amount of urine protein excreted in 24 hours.
- Thus, in the subsequent analysis, we use the two variables as predictors rather than outcomes.
- The estimated envelope dimension is $u = 17$.

Table: The point estimates, bootstrap standard errors, confidence intervals and $p-$values for biomarkers adjusted for the established biomarkers

	Our Method					Standard EM				
	$\hat{\beta}$	$\widehat{\mathrm{SE}}$	2.5%	97.5%	$p-$value	$\hat{\beta}$	$\widehat{\mathrm{SE}}$	2.5%	97.5%	$p-$value
HS_CRP	-0.04	0.02	-0.07	-2e-3	0.05	-0.12	0.07	-0.28	0.02	0.10
NGAL	-0.01	0.03	-0.07	0.04	0.69	0.18	0.07	0.06	0.31	6e-3
ALDOSTERONE	0.06	0.02	0.02	0.09	2e-3	0.04	0.04	-0.04	0.13	0.31
C_PEPTIDE	-0.10	0.04	-0.17	-0.03	9e-3	0.21	0.12	-0.02	0.44	0.08

Real data analysis

Figure: The Empirical Cumulative Distribution of ratio between standard error of EM OLS and EM envelope for the coefficient corresponds to ESRD (left) and for all coefficients (right), adjusting for established biomarkers.



Thank you!

Cook, R. D. and Zhang, X. (2016). Algorithms for envelope estimation. *Journal of Computational and Graphical Statistics*, 25(1):284–300.

Feldman, H., Appel, L., Chertow, G., Cifelli, D., Cizman, B., Daugirdas, J., Fink, J., Franklin-Becker, E., Go, A., Hamm, L., et al. (2003). The chronic renal insufficiency cohort (cric) study: design and methods. *Journal of the American Society of Nephrology*, 14:S148–S153.