Envelope method 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

November 15, 2018

Outlines

Regression with multiple responses

- Linear regression model can be written as:

$$\mathbf{Y}_{1\times r} = \mathbf{X}_{1\times p}\boldsymbol{\beta}_{p\times r} + \boldsymbol{\varepsilon}_{1\times r},$$

  where the error vector $\boldsymbol{\varepsilon}_{1\times r} \in \mathbb{R}^r$ is normally distributed with mean $\mathbf{0}$ and unknown parameter $\boldsymbol{\Sigma}$. When $\boldsymbol{\Sigma} > 0$, the model has a total number of $pr + r(r+1)/2$ unknown parameters.
- Suppose we observe the data set $\big((\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_n, \mathbf{y}_n)\big)$, we can write them as a matrix form:

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

Maximum likelihood estimator

- The log-likelihood of the 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 the 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 the partial derivative of β to $\mathbf{0}$, we have

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

Maximum likelihood estimator

- The log-likelihood of the 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 the partial derivative of β to $\mathbf{0}$, we have

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

- Since Σ is positive definite, we can cancel it on both sides. Hence the MLE of β is

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

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

Maximum likelihood estimator

- The log-likelihood of the 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 the partial derivative of β to $\mathbf{0}$, we have

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

- Since Σ is positive definite, we can cancel it on both sides. Hence the MLE of β is

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

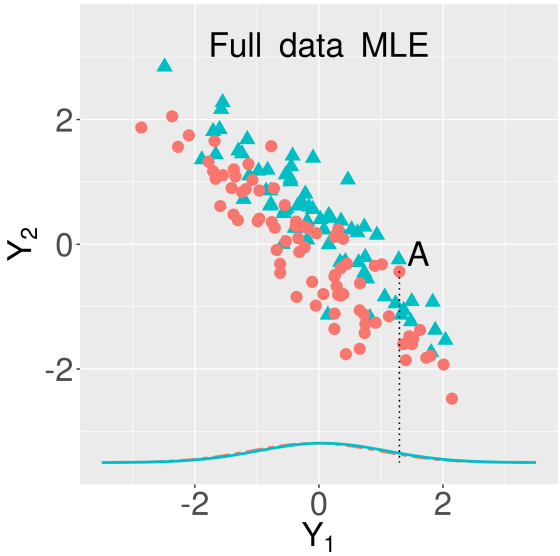
where $\dagger$ indicates the Moore-Penrose inverse.

Remark

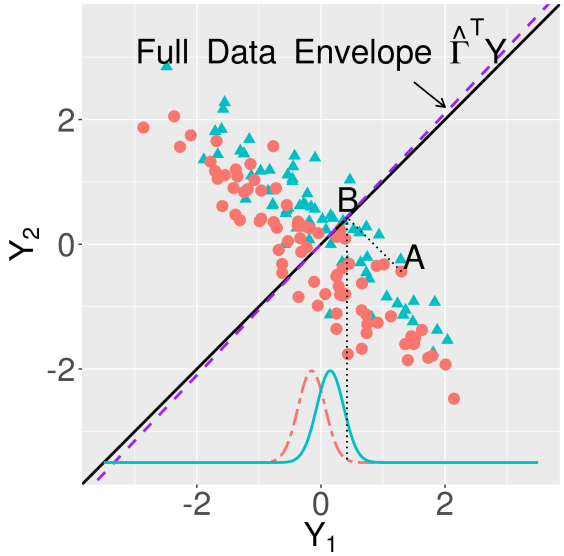
The estimator of β does not depend on Σ at all.

Motivation

Figure: Intuitive illustration of the envelope method. The density curves of the two groups using envelope method are shown at the bottom of each subfigure.



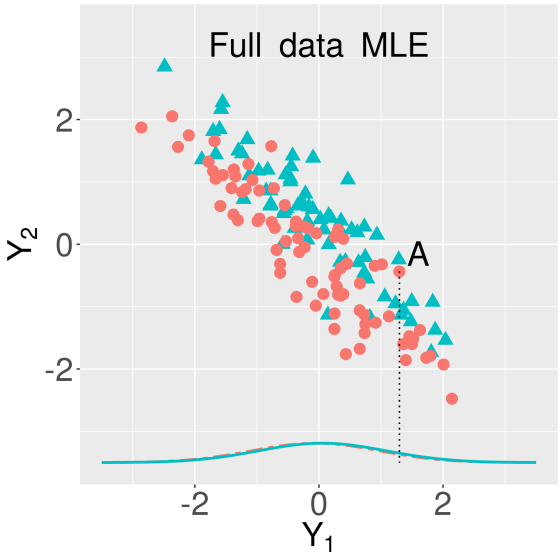
(a) MLE



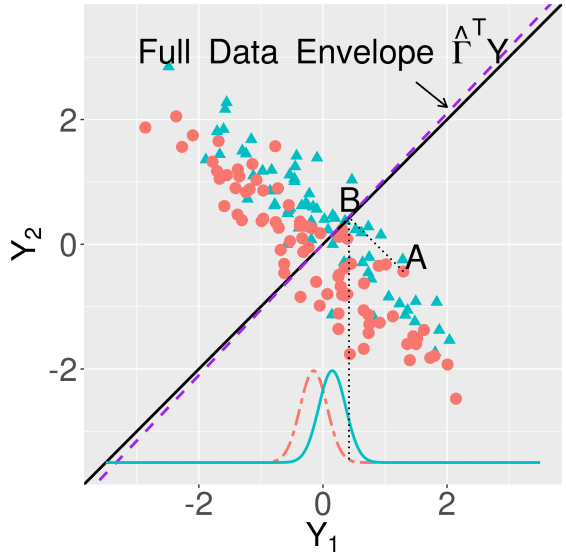
(b) Envelope estimation

Motivation

Figure: Intuitive illustration of the envelope method. The density curves of the two groups using envelope method are shown at the bottom of each subfigure.



(a) MLE



(b) Envelope estimation

- Suppose we want to test whether $\beta_1 = 0$. Using the standard method, we get $p-$value$= 0.37$; Using envelope estimation, the $p-$value is less than 0.001.

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

$$\mathbf{Y}_{r\times 1} = \beta_{r\times p}\mathbf{X}_{p\times 1} + \varepsilon_{r\times 1}.$$

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

$$\mathbf{Y}_{r\times 1} = \boldsymbol{\beta}_{r\times p}\mathbf{X}_{p\times 1} + \boldsymbol{\varepsilon}_{r\times 1}.$$

- Consider a subspace $\mathcal{E} \subseteq \mathbb{R}^r$ such 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

$$\mathbf{Y}_{r\times 1} = \boldsymbol{\beta}_{r\times p}\mathbf{X}_{p\times 1} + \boldsymbol{\varepsilon}_{r\times 1}.$$

- Consider a subspace $\mathcal{E} \subseteq \mathbb{R}^r$ such 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 $\mathbf{Q}_{\mathcal{E}}\mathbf{Y}$ as material and immaterial part of $\mathbf{Y}$.

Motivation

- Suppose $\mathbf{\Gamma}_{r\times u}$ is a semi-orthogonal matrix ($\mathbf{\Gamma}^T\mathbf{\Gamma} = \mathbf{I}_{u\times u}$) such that $\mathrm{span}(\mathbf{\Gamma}) = \mathcal{E}$, then the previous two conditions are equivalent to

$$\mathrm{Span}(\boldsymbol{\beta}) \subseteq \mathrm{Span}(\mathbf{\Gamma})$$

$$\mathbf{\Sigma} = \mathbf{P}_{\mathbf{\Gamma}}\mathbf{\Sigma}\mathbf{P}_{\mathbf{\Gamma}} + \mathbf{Q}_{\mathbf{\Gamma}}\mathbf{\Sigma}\mathbf{Q}_{\mathbf{\Gamma}}$$

Motivation

- Suppose $\mathbf{\Gamma}_{r\times u}$ is a semi-orthogonal matrix ($\mathbf{\Gamma}^T\mathbf{\Gamma} = \mathbf{I}_{u\times u}$) such that $\mathrm{span}(\mathbf{\Gamma}) = \mathcal{E}$, then the previous two conditions are equivalent to

$$\mathrm{Span}(\boldsymbol{\beta}) \subseteq \mathrm{Span}(\mathbf{\Gamma})$$

$$\mathbf{\Sigma} = \mathbf{P}_{\mathbf{\Gamma}}\mathbf{\Sigma}\mathbf{P}_{\mathbf{\Gamma}} + \mathbf{Q}_{\mathbf{\Gamma}}\mathbf{\Sigma}\mathbf{Q}_{\mathbf{\Gamma}}$$

Example

Assume $\mathbf{Y} = (Y_1, Y_2)^T$, where $Y_1 = \boldsymbol{\beta}\mathbf{X} + \varepsilon_1$, and $Y_2 = -\boldsymbol{\beta}\mathbf{X} + \varepsilon_2$. Suppose ε_1 and ε_2 follow two normal distributions and they are independent of each other.

Motivation

- Suppose $\mathbf{\Gamma}_{r\times u}$ is a semi-orthogonal matrix ($\mathbf{\Gamma}^T\mathbf{\Gamma} = \mathbf{I}_{u\times u}$) such that $\mathrm{span}(\mathbf{\Gamma}) = \mathcal{E}$, then the previous two conditions are equivalent to

$$\mathrm{Span}(\boldsymbol{\beta}) \subseteq \mathrm{Span}(\mathbf{\Gamma})$$

$$\mathbf{\Sigma} = \mathbf{P}_{\mathbf{\Gamma}}\mathbf{\Sigma}\mathbf{P}_{\mathbf{\Gamma}} + \mathbf{Q}_{\mathbf{\Gamma}}\mathbf{\Sigma}\mathbf{Q}_{\mathbf{\Gamma}}$$

Example

Assume $\mathbf{Y} = (Y_1, Y_2)^T$, where $Y_1 = \boldsymbol{\beta}\mathbf{X} + \varepsilon_1$, and $Y_2 = -\boldsymbol{\beta}\mathbf{X} + \varepsilon_2$. Suppose ε_1 and ε_2 follow two normal distributions and they are independent of each other.
Then, the predictors $\mathbf{X}$ do not affect the summation of response $Y_1 + Y_2 = \varepsilon_1 + \varepsilon_2$. Also, $Y_1 + Y_2$ is independent of $Y_1 - Y_2$, so that $Y_1 + Y_2$ can be completely discarded. That is, the regression of $\mathbf{Y}$ on $\mathbf{X}$ can be replaced with the regression of $Y_1 - Y_2$ on $\mathbf{X}$.

Motivation

- Suppose $\mathbf{\Gamma}_{r\times u}$ is a semi-orthogonal matrix ($\mathbf{\Gamma}^T\mathbf{\Gamma} = \mathbf{I}_{u\times u}$) such that $\mathrm{span}(\mathbf{\Gamma}) = \mathcal{E}$, then the previous two conditions are equivalent to

$$\mathrm{Span}(\boldsymbol{\beta}) \subseteq \mathrm{Span}(\mathbf{\Gamma})$$

$$\mathbf{\Sigma} = \mathbf{P}_{\mathbf{\Gamma}}\mathbf{\Sigma}\mathbf{P}_{\mathbf{\Gamma}} + \mathbf{Q}_{\mathbf{\Gamma}}\mathbf{\Sigma}\mathbf{Q}_{\mathbf{\Gamma}}$$

Example

Assume $\mathbf{Y} = (Y_1, Y_2)^T$, where $Y_1 = \boldsymbol{\beta}\mathbf{X} + \varepsilon_1$, and $Y_2 = -\boldsymbol{\beta}\mathbf{X} + \varepsilon_2$. Suppose ε_1 and ε_2 follow two normal distributions and they are independent of each other.
Then, the predictors $\mathbf{X}$ do not affect the summation of response $Y_1 + Y_2 = \varepsilon_1 + \varepsilon_2$. Also, $Y_1 + Y_2$ is independent of $Y_1 - Y_2$, so that $Y_1 + Y_2$ can be completely discarded. That is, the regression of $\mathbf{Y}$ on $\mathbf{X}$ can be replaced with the regression of $Y_1 - Y_2$ on $\mathbf{X}$.
In this example, $\mathbf{\Gamma} = (1, -1)^T/\sqrt{2}$ and $\mathbf{\Gamma}_0 = (1, 1)^T/\sqrt{2}$

Formal definition

- The semi-orthogonal matrix $\mathbf{\Gamma}$ is not unique. For example, identity matrix $\mathbf{I}_r$ satisfies the two conditions trivially, i.e.,

$$\mathrm{Span}(\boldsymbol{\beta}) \subseteq \mathrm{Span}(\mathbf{I}_r)$$

$$\mathbf{\Sigma} = \mathbf{P}_{\mathbf{I}_r}\mathbf{\Sigma}\mathbf{P}_{\mathbf{I}_r} + \mathbf{Q}_{\mathbf{I}_r}\mathbf{\Sigma}\mathbf{Q}_{\mathbf{I}_r}$$

Formal definition

- The semi-orthogonal matrix $\mathbf{\Gamma}$ is not unique. For example, identity matrix $\mathbf{I}_r$ satisfies the two conditions trivially, i.e.,

$$\mathrm{Span}(\boldsymbol{\beta}) \subseteq \mathrm{Span}(\mathbf{I}_r)$$

$$\mathbf{\Sigma} = \mathbf{P}_{\mathbf{I}_r}\mathbf{\Sigma}\mathbf{P}_{\mathbf{I}_r} + \mathbf{Q}_{\mathbf{I}_r}\mathbf{\Sigma}\mathbf{Q}_{\mathbf{I}_r}$$

Definition 1. (Cook et al., 2010)

The intersection of all subspaces $\mathcal{E}$ with properties $\mathrm{Span}(\boldsymbol{\beta}) \subseteq \mathrm{Span}(\mathbf{\Gamma})$ and $\mathbf{\Sigma} = \mathbf{P}_{\mathcal{E}}\mathbf{\Sigma}\mathbf{P}_{\mathcal{E}} + \mathbf{Q}_{\mathcal{E}}\mathbf{\Sigma}\mathbf{Q}_{\mathcal{E}}$ is defined as the $\mathbf{\Sigma}$-envelope of $\boldsymbol{\beta}$, denoted by $\mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B})$. $u = \dim(\mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B}))$ is called the envelope dimension.

Formal definition

- The semi-orthogonal matrix $\mathbf{\Gamma}$ is not unique. For example, identity matrix $\mathbf{I}_r$ satisfies the two conditions trivially, i.e.,

$$\mathrm{Span}(\boldsymbol{\beta}) \subseteq \mathrm{Span}(\mathbf{I}_r)$$

$$\mathbf{\Sigma} = \mathbf{P}_{\mathbf{I}_r}\mathbf{\Sigma}\mathbf{P}_{\mathbf{I}_r} + \mathbf{Q}_{\mathbf{I}_r}\mathbf{\Sigma}\mathbf{Q}_{\mathbf{I}_r}$$

Definition 1. (Cook et al., 2010)

The intersection of all subspaces $\mathcal{E}$ with properties $\mathrm{Span}(\boldsymbol{\beta}) \subseteq \mathrm{Span}(\mathbf{\Gamma})$ and $\mathbf{\Sigma} = \mathbf{P}_{\mathcal{E}}\mathbf{\Sigma}\mathbf{P}_{\mathcal{E}} + \mathbf{Q}_{\mathcal{E}}\mathbf{\Sigma}\mathbf{Q}_{\mathcal{E}}$ is defined as the $\mathbf{\Sigma}$-envelope of $\boldsymbol{\beta}$, denoted by $\mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B})$. $u = \dim(\mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B}))$ is called the envelope dimension.

Proposition 1. (Cook et al., 2010)

Assume that $\mathbf{\Sigma}_{r\times r}$ is symmetric and has $q \leq r$ distinct eigenvalues. Let $\mathbf{P}_i$, $i = 1, \ldots, q$, indicate the projection onto the corresponding eigenspaces. Then,

$$\mathcal{E}_{\mathbf{\Sigma}}(\mathcal{B}) = \oplus_{q=1}^{q}\mathbf{P}_i\mathcal{B}$$

The envelope model

- Now we want to refine the multivariate regression model by using the $\mathbf{\Sigma}$-envelope of $\boldsymbol{\beta}$ to connect $\boldsymbol{\beta}$ and $\mathbf{\Sigma}$.

The envelope model

- Now we want to refine the multivariate regression model by using the $\boldsymbol{\Sigma}$-envelope of $\boldsymbol{\beta}$ 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 $\boldsymbol{\Gamma}_0 \in \mathbb{R}^{r \times (u-r)}$ be its orthogonal complements.

The envelope model

- Now we want to refine the multivariate regression model by using the $\boldsymbol{\Sigma}$-envelope of $\boldsymbol{\beta}$ 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 $\boldsymbol{\Gamma}_0 \in \mathbb{R}^{r \times (u-r)}$ be its orthogonal complements.
- There exist $\boldsymbol{\eta} \in \mathbb{R}^{u \times p}$ such that $\boldsymbol{\beta} = \boldsymbol{\Gamma}\boldsymbol{\eta}$, where $\boldsymbol{\eta}$ contains the coordinates of $\boldsymbol{\beta}$ relative to $\boldsymbol{\Gamma}$.

The envelope model

- Now we want to refine the multivariate regression model by using the $\boldsymbol{\Sigma}$-envelope of $\boldsymbol{\beta}$ 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 $\boldsymbol{\Gamma}_0 \in \mathbb{R}^{r \times (u-r)}$ be its orthogonal complements.
- There exist $\boldsymbol{\eta} \in \mathbb{R}^{u \times p}$ such that $\boldsymbol{\beta} = \boldsymbol{\Gamma}\boldsymbol{\eta}$, where $\boldsymbol{\eta}$ contains the coordinates of $\boldsymbol{\beta}$ relative to $\boldsymbol{\Gamma}$.

Envelope model

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

Estimation of the parameters can be carried out by maximum likelihood, and u can be selected 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 where $|\cdot|_0$ means the product of the non-zero eigenvalues and δ 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 $\hat{\mathbf{\Gamma}}$ denote the basis of $\hat{\mathcal{E}}_{\mathbf{\Sigma}}(\mathcal{B})$, then estimators of regression parameters are
  - $\hat{\beta}_{env} = \mathbf{P}_{\hat{\mathbf{\Gamma}}}\hat{\beta}_{std}$, which is $\sqrt{n}$-consistent and asymptotically normal.
  - $\mathbf{\Sigma} = \mathbf{P}_{\hat{\mathbf{\Gamma}}}\mathbf{S}_{\mathbf{Y}|\mathbf{X}}\mathbf{P}_{\hat{\mathbf{\Gamma}}} + \mathbf{Q}_{\hat{\mathbf{\Gamma}}}\mathbf{S}_{\mathbf{Y}}\mathbf{Q}_{\hat{\mathbf{\Gamma}}}$

Efficiency gain

Proposition 2. (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 avar$(\cdot)$ stands for the asymptotic covariance and vec$(\cdot)$ stands for the vectorization of a matrix.

Remarks

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

An R 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)
```

An R 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)
```

An R 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

1. Introduction to the envelope model
   - Regression with multiple responses
   - Motivation
   - Formal definition
   - The envelope model
   - An R Example

2. Envelope method with ignorable missing data
   - Motivation and preliminary
   - EM envelope algorithm
   - Simulations
   - Real data analysis

Motivation and preliminary

- In the 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

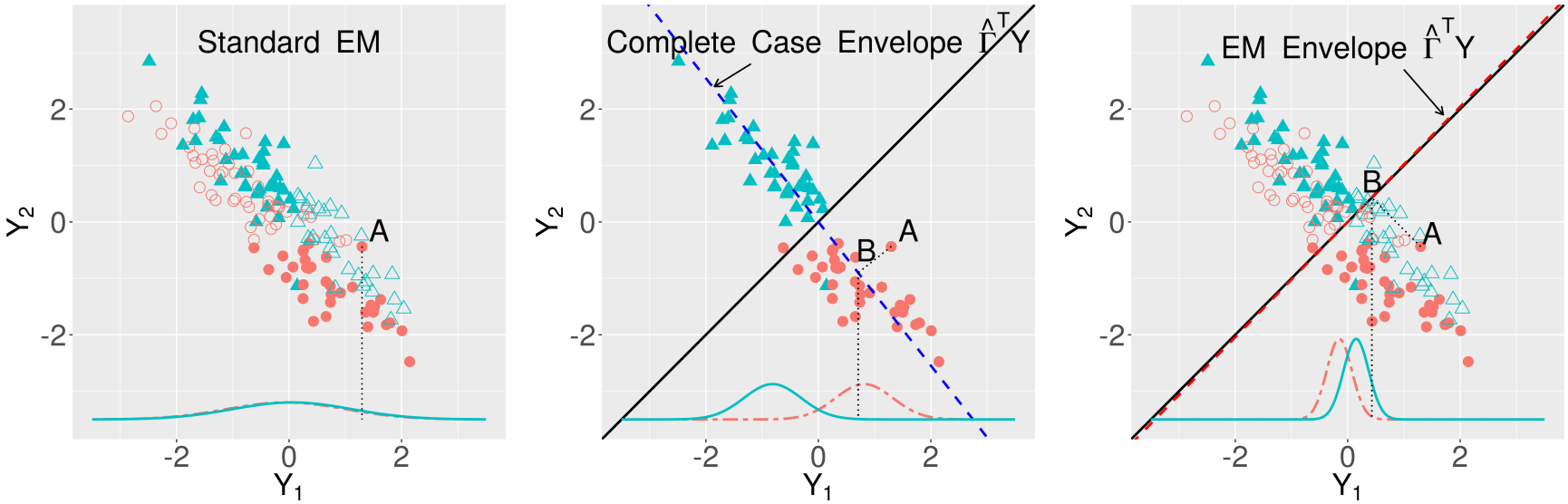
- In the big data setting where large amount of responses and predictors are collected, it is common that the responses or the predictors or both suffer from missingness.
- Types of missing data:
  - MCAR: Missingness is independent both of observed and unobserved data.
  - MAR: Missingness is independent 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 the 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

Figure: Intuitive illustration of the envelope method in the presence of missing data. Hollow circle dots or triangles indicate one of the component of $\mathbf{Y}$ is missing: the hollow triangle has Y_1 missing, and the hollow circle dot has Y_2 missing. The density curves of the two groups using different methods are shown at the bottom of each subfigure.



(a) Standard EM (b) CC envelope (c) EM envelope

EM algorithm

$$
\begin{aligned}
& l(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{y}) = \log(f_{y|x}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})) + \log(f_x(\boldsymbol{x}|\boldsymbol{\theta})) \\
= & \sum_{i=1}^{n} \Big(-\frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\boldsymbol{y_i} - \boldsymbol{x_i}\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{y_i} - \boldsymbol{x_i}\boldsymbol{\beta}) + \log(f_x(\boldsymbol{x}_i|\boldsymbol{\rho})) \Big) + C
\end{aligned}
$$

EM algorithm

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

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

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

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

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

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.
- $\Sigma_{1,t+1} = \frac{1}{n}\boldsymbol{P}_{\boldsymbol{\Gamma}_t}(\boldsymbol{A}_{1,t} - \boldsymbol{A}_{2,t}\boldsymbol{A}_{3,t}^{-1}\boldsymbol{A}_{2,t}')\boldsymbol{P}_{\boldsymbol{\Gamma}_t}$;
  $\rho_{t+1} = \arg\max_{\rho\in\Theta} \mathbb{E}(\log(f_x(\boldsymbol{x}_i|\rho))|\boldsymbol{D}_{obs}, \theta_t)$;
  $\beta_{t+1} = \boldsymbol{A}_{3,t}^{-1}\boldsymbol{A}_{2,t}'\boldsymbol{P}_{\boldsymbol{\Sigma}_{1,t+1}}$;
  $\Sigma_{t+1} = \Sigma_{1,t+1} + \frac{1}{n}\boldsymbol{Q}_{\boldsymbol{\Gamma}_t}\boldsymbol{A}_{1,t}\boldsymbol{Q}_{\boldsymbol{\Gamma}_t}$

Algorithm 0: 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: $\boldsymbol{\Sigma}_t = \boldsymbol{I}_q$, $\beta_t = \mathbf{0}$, $\boldsymbol{\rho}_t = \boldsymbol{\rho}_0$, $\boldsymbol{\theta}_t = (\boldsymbol{\Sigma}_t, \beta_t, \boldsymbol{\rho}_t)$, $\Delta = 1$;

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 $\boldsymbol{\theta}_t$;
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: $\boldsymbol{\rho}_{t+1} = \arg\max_{\boldsymbol{\rho}\in\Theta} \mathbb{E}(\log(f_x(\boldsymbol{x}_i|\boldsymbol{\rho}))|\boldsymbol{D}_{obs}, \boldsymbol{\theta}_t)$, $\beta_{t+1} = \boldsymbol{A}_{3,t}^{-1}\boldsymbol{A}_{2,t}'\boldsymbol{P}_{\boldsymbol{\Sigma}_{1,t+1}}$, $\boldsymbol{\Sigma}_{t+1} = \boldsymbol{\Sigma}_{1,t+1} + \frac{1}{n}\boldsymbol{Q}_{\boldsymbol{\Gamma}_t}\boldsymbol{A}_{1,t}\boldsymbol{Q}_{\boldsymbol{\Gamma}_t}$;
4. Set $\Delta = \|\beta_{t+1} - \beta_t\|_1$, $\boldsymbol{\theta}_t = (\boldsymbol{\Sigma}_{t+1}, \beta_{t+1}, \boldsymbol{\rho}_{t+1})$;

end

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

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.

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.
- 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.
- 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%.

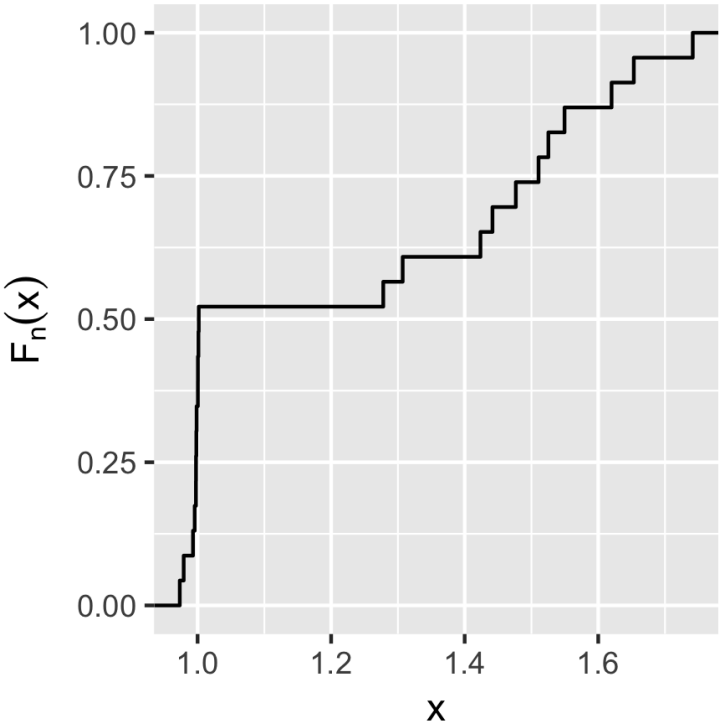
Real data analysis

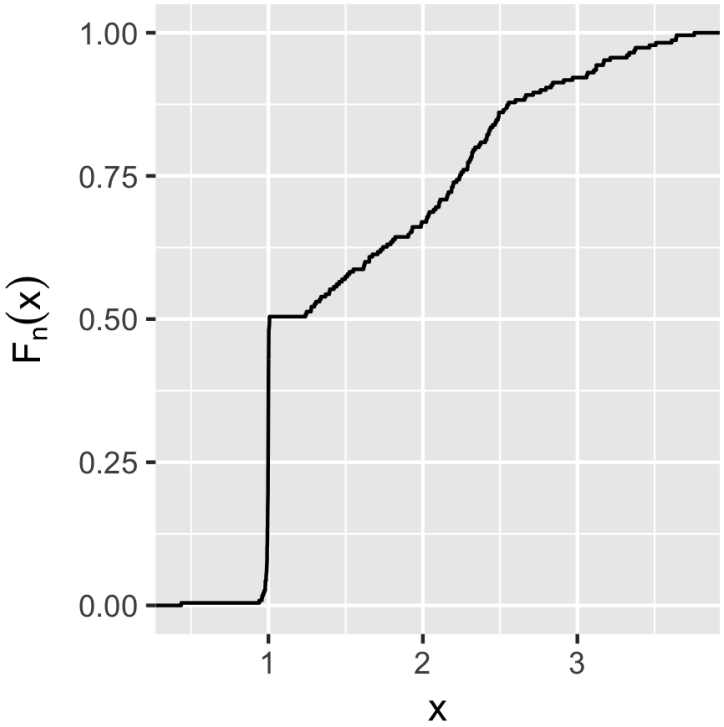
In the regression setting

- 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 the 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 these two variables as predictors rather than outcomes.

Real data analysis

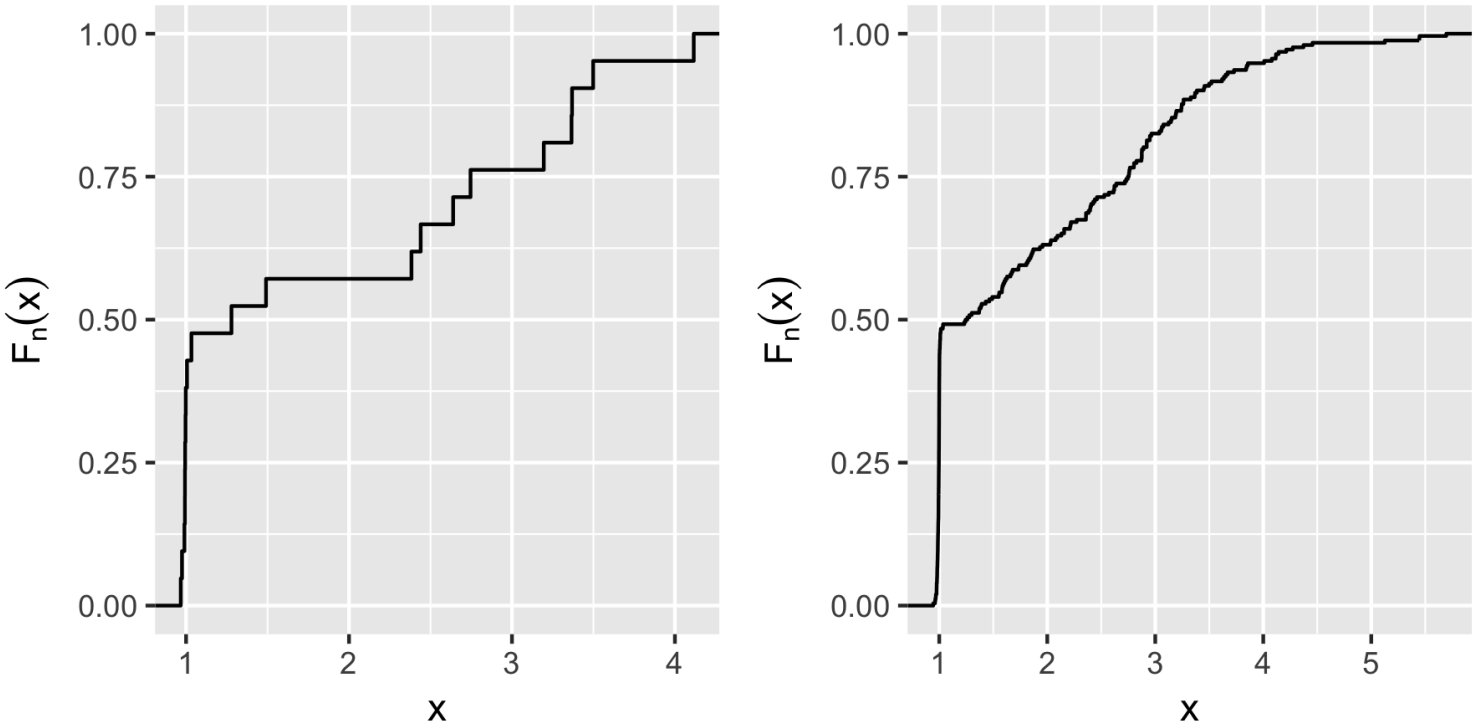
- It is found in the literature that although many novel biomarkers are found to be marginally significant associate with ESRD status, such association is muted after adjusting for glomerular filtration rate (GFR) and the amount of urine protein excreted in 24 hours.
- Thus, in the subsequent analysis, we use these 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., Li, B., and Chiaromonte, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression. *Statistica Sinica*, pages 927–960.

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