From Latent Variables to Mixture Models to Inference in Subgroup Analysis

Xuming He

Department of Statistics
University of Michigan
October 14, 2016

Introduction

A personal “encounter” with Professor Jayaram Sethuraman:

- Bahadur efficiency
- Markov chain simulation methods
- Bayesian inference
- Rank-based methods

Introduction

- Latent variables are valuable in model-based clustering and inference.
- Mixture models offer a flexible parametric approach to data clustering, subgroup identification, and prior construction.
- The Expectation-Maximization (EM) algorithms are naturally suited estimation associated with latent variables and mixture models.

Introduction

- Subgroup analysis:
  looking for subpopulations with distinct response characteristics.
- Subgroup analysis is commonly performed: (Sun et al., 2012)
  - 469 trials in 2007 in core medical journals,
  - 207 trials (44%) contain subgroup analysis,
  - 83 (41%) of them claim subgroup effects,
  - among the 83 trials, 46 (56%) of them made one claim, 37 (44%) made two or more claims.

Introduction

- Subgroup analysis:
  looking for subpopulations with distinct response characteristics.
- Subgroup analysis is commonly performed: (Sun et al., 2012)
  - 469 trials in 2007 in core medical journals,
  - 207 trials (44%) contain subgroup analysis,
  - 83 (41%) of them claim subgroup effects,
  - among the 83 trials, 46 (56%) of them made one claim, 37 (44%) made two or more claims.

Introduction

- “Subgroup analysis in clinical trials: fun to look at – but do not believe them!” (Sleight, 2000).
- Special issue on subgroup analysis in clinical trials (J Biopharm Stat. 2014).
- **21,000 entries since 2015** by Google Scholar

Issues with subgroup analysis:

- Unplanned (exploratory):
  - False discovery: exhaustive search, selective reporting.
- Planned:
  - Multiple comparisons (type-1 errors uncontrolled).
  - Lack of power.

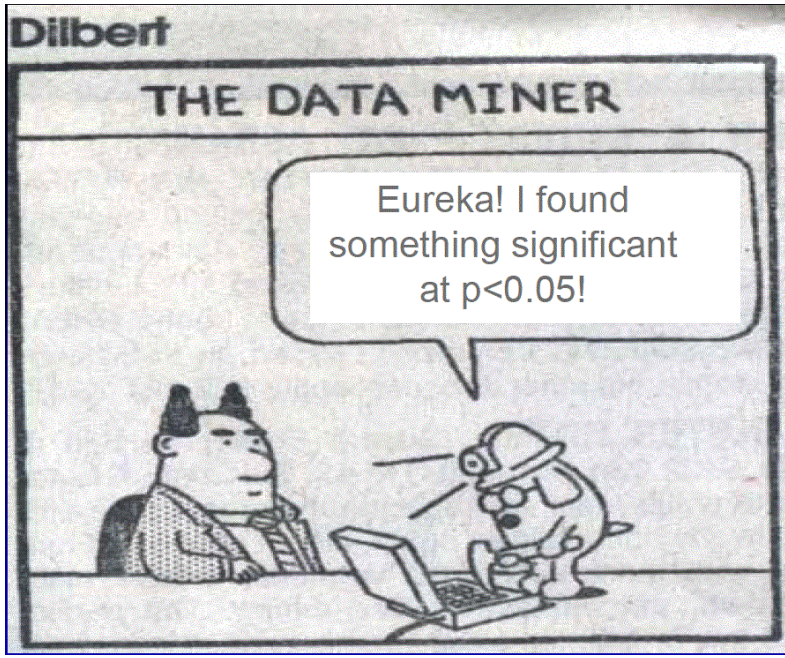


Figure: *from* `http://obssr.od.nih.gov/pdf/8-7-02FrasureSmith.pdf`

Outline

Confirmatory Subgroup Analysis

Key statistical considerations with pre-specified subgroups (Planned),

- Control overall Type I error rate.
- Improve power.

Confirmatory Subgroup Analysis

Strategies for multiplicity adjustment:

- Consider the logical relationship among the null hypothesis: e.g., the gatekeeping procedures in Dmitrienko et al. (2011).
- Consider the correlation among test statistics: Song and Chi (2007), Alosh and Huque (2009).
- Consider a change-plane model with semiparametric specification: Fan, Song, and Lu (2016).

Recent Work on Subgroup Identifications

Goal: to find a subgroup (subgroups) with an enhanced treatment effect

- Treat subgroup identification as variable selection by interactions.
  - For example, Bayesian methods of Jones et al. (2011).
- Interaction tree procedure of Su et al. (2009).

Recent Work on Subgroup Identifications

- Lipkovich et al. (2011) and Lipkovich and Dmitrienko (2014): subgroup identification based on differential effect search (SIDES)- a recursive partioning method for establishing response to treatment in patient subpopulations.
- Foster et al. (2011): the “Virtual Twins” method.
- Cai et al. (2011) and Zhao et al. (2013): a parametric scoring system based on multiple covariates; subgroup based on estimated scores.

A Model-based Subgroup Inference

Key statistical considerations of our work:

- to develop a confirmatory test for the existence of subgroups,
- to characterize potential subgroups if possible.

Reference: Shen and He (2015)

ACTG320 Data

The AIDS Clinical Trials Group 320 (ACTG320) Study:

- For HIV-1 infected patients: the standard two-drug combination (zidovudine and lamivudine) VS. a new three-drug combination (adding a protease inhibitor indinavir).
- Outcome variable: CD4 counts (cells/mm^3).

ACTG320 Data

ACTG320 Project:

- The new treatment was shown to be effective ($n = 1146$) for increasing or inhibiting the decrease of the *CD*4 cell counts.
- Outcome: the change of the *CD*4 counts at the 24th week.
- Mean difference 81 cells/mm^3, highly significant.

ACTG320 Data

- Question: is there a (predictable) subgroup of patients who benefits much more from the treatment?
- From earlier studies, baseline RNA, CD4, Age are shown to be related to HIV disease.
- RNA: plasma (HIV-1) RNA concentrations – $\log_{10}$ copies/ml.

A Model-based Subgroup Inference

Our Goals:

- Primary goal: to test the existence of clinically meaningful subgroups through joint modeling of the response and the subgroup membership.

- Secondary goal: to characterize the subgroup membership if possible.

Notations

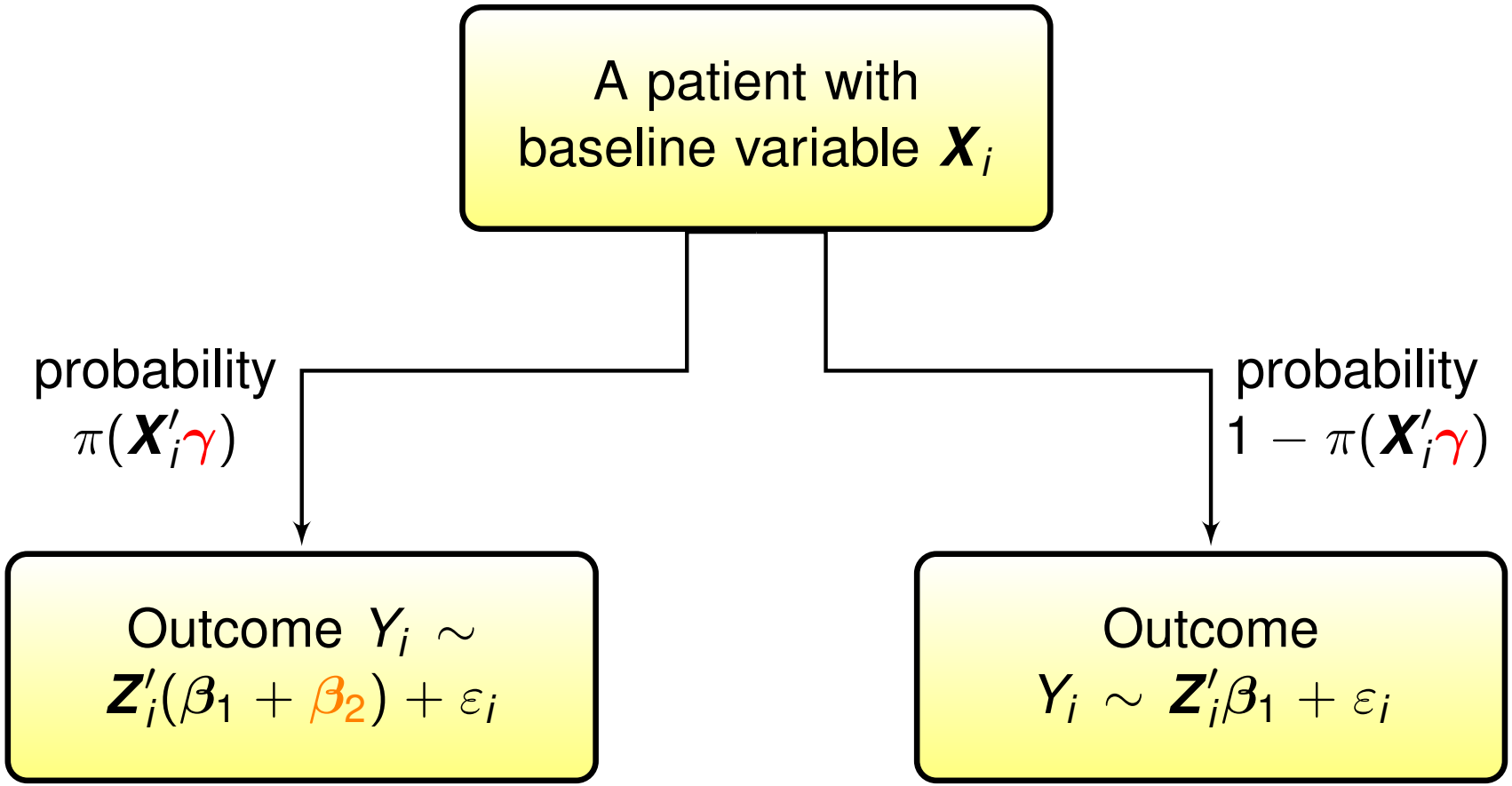
For $i = 1, ..., n$,

- $Y_i \in \mathbb{R}$, outcome variable;
- $\boldsymbol{X}_i \in \mathbb{R}^{q_2}$, baseline variables;
- $\boldsymbol{Z}_i \in \mathbb{R}^{q_1}$, covariates including the treatment indicator t_i;
- Let $\pi(\cdot)$ be the logistic function:

$$\pi(a) = \frac{1}{1 + \exp(-a)};$$

- The first elements of $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are 1 (for intercepts);
- $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ may have overlapping components.

Proposed Model



Structured Logistic-Normal Mixture Model

$$
\begin{aligned}
P(\delta_i = 1|\boldsymbol{X}_i) &= \pi(\boldsymbol{X}_i^T\gamma) \equiv \frac{\exp(\boldsymbol{X}_i^T\gamma)}{1+\exp(\boldsymbol{X}_i^T\gamma)}, \\
Y_i|(\boldsymbol{Z}_i, \boldsymbol{X}_i, \delta_i) &= \boldsymbol{Z}_i^T(\beta_1 + \beta_2\delta_i) + \varepsilon_i,
\end{aligned} \tag{1}
$$

where

- $\delta_i \in \{0, 1\}$; $\varepsilon_i \sim N(0, \sigma^2)$ white noise.
- Parameters $\gamma \in \mathbb{R}^{q_2}$, $\beta_1 \in \mathbb{R}^{q_1}$, $\beta_2 \in \mathbb{R}^{q_1}$, and $\sigma^2 \in \mathbb{R}^+$.
- Let $\theta = (\beta_1, \sigma, \beta_2)$.

Structured Logistic-Normal Mixture Model

$$
\begin{aligned}
P(\delta_i = 1|\boldsymbol{X}_i) &= \pi(\boldsymbol{X}_i^T\gamma) \equiv \frac{\exp(\boldsymbol{X}_i^T\gamma)}{1+\exp(\boldsymbol{X}_i^T\gamma)}, \\
Y_i|(\boldsymbol{Z}_i, \boldsymbol{X}_i, \delta_i) &= \boldsymbol{Z}_i^T(\beta_1 + \beta_2\delta_i) + \varepsilon_i,
\end{aligned} \tag{2}
$$

In the motivating example,

- $\boldsymbol{X}_i$ contains baseline RNA, CD4 counts, and Age;
- $\boldsymbol{Z}_i$ contains those in $\boldsymbol{X}_i$ and the treatment indicator;
- $\boldsymbol{Y}_i$ is the CD4 count changes at the 24th week.

Structured Logistic-Normal Mixture Model

Model Specifics:

- If subgroups do exist, the *EM* algorithm can be used for estimation.
- When $\boldsymbol{X} = \boldsymbol{Z}$, the model falls into the general model of Mixture of Experts (Jordan and Jacobs (1994), Yuksel et al. (2012)), popular in machine learning mainly for prediction.

Testing for the Existence of Subgroups

- Hypothesis on the existence of subgroups:

$$H_0 : \beta_2 = 0$$

About the null hypothesis:

- We are testing if the whole vector $\beta_2 = 0$, not just the treatment effect difference. The most challenging step!
- Once H_0 is rejected, quantify the treatment effect difference β_t is straightforward.

Testing for the Existence of Subgroups

- Hypothesis on the existence of subgroups:

$$H_0 : \beta_2 = 0$$

A non-standard problem:

- When $\beta_2 = 0$, γ is not identifiable.
- The null model is not an interior point in the set of alternative models.
- The likelihood ratio test(LRT) statistic may not have a χ^2 limiting distribution .

Testing for the Existence of Subgroups

- Hypothesis on the existence of subgroups:

$$H_0 : \beta_2 = 0$$

A non-standard problem:

- When $\beta_2 = 0$, γ is not identifiable.
- The null model is not an interior point in the set of alternative models.
- The likelihood ratio test(LRT) statistic may not have a χ^2 limiting distribution .

Simple Model without Covariates

Consider the simplest case: to test $\beta_2 = 0$ in

$$\boldsymbol{Y} = pN(\beta_1 + \beta_2, \boldsymbol{\Sigma}) + (1-p)N(\beta_1, \boldsymbol{\Sigma}) \tag{3}$$

where $\boldsymbol{Y}, \beta_1, \beta_2 \in \mathbb{R}^q$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{q \times q}$,.

Existing Results for the Simple Model

Goeffinet et al. (1992) develops the limiting distribution for the LRT for given $p \in (0, 0.5]$:

- If $q = 1$,
  - if $p \neq 0.5$ and Σ unknown, LRT $\Rightarrow \chi^2_1$.
  - o.w., $0.5\chi^2_0 + 0.5\chi^2_1$.
- If $q = 2$,
  - if Σ known, LRT $\Rightarrow 0.5\chi^2_0 + 0.5M^2$, where $M = M_1 + \sqrt{M_2^2 + M_3^2}$, M_1, M_2 and $M_3 \sim N(0, 1)$.
  - o.w., no analytical results.

Existing Results for the Simple Model

Goeffinet et al. (1992) develops the limiting distribution for the LRT for given $p \in (0, 0.5]$:

- If $q = 1$,
  - if $p \neq 0.5$ and Σ unknown, LRT $\Rightarrow \chi^2_1$.
  - o.w., $0.5\chi^2_0 + 0.5\chi^2_1$.
- If $q = 2$,
  - if Σ known, LRT $\Rightarrow 0.5\chi^2_0 + 0.5M^2$, where $M = M_1 + \sqrt{M_2^2 + M_3^2}$, M_1, M_2 and $M_3 \sim N(0, 1)$.
  - o.w., no analytical results.

- Without covariates, for a given p, the LRT of $\beta_2 = 0$ has a complicated limiting distribution.
- For those with known limiting distributions, the convergence is very slow, especially when p is close to 0.5.

- What will happen when we have covariates in both the subgroup means and the mixing proportions?
- The covariates bring us both convenience and challenges.

Fisher Information Matrix

For our model, given γ, under the null that $\theta_0 = (\beta_0, \sigma_0, 0_{q_1 \times 1})$,

- The Fisher information matrix is

$$I_\gamma(\theta_0) = \frac{1}{\sigma_0^2} \begin{pmatrix} A & 0 & B(\gamma) \\ 0 & 2 & 0 \\ B(\gamma) & 0 & C(\gamma) \end{pmatrix}, \tag{4}$$

where

- $A = \mathbb{E}(\boldsymbol{Z}\boldsymbol{Z}^T)$
- $B(\gamma) = \mathbb{E}(\pi(\boldsymbol{X}^T\gamma)\boldsymbol{Z}\boldsymbol{Z}^T)$
- $C(\gamma) = \mathbb{E}(\pi^2(\boldsymbol{X}^T\gamma)\boldsymbol{Z}\boldsymbol{Z}^T)$

LRT Given γ

Given $\gamma = (\gamma_{-\boldsymbol{X}}, \gamma_{\boldsymbol{X}})$ with $\gamma_{\boldsymbol{X}} \neq 0$, if $\boldsymbol{X}$ is linearly independent, then

- The regularity conditions are satisfied;
- The likelihood ratio test (LRT) of $H_0 : \beta_2 = 0$ given γ, $T(\gamma) \Rightarrow \chi^2_{q_1}$.
- For each γ,

$$T(\gamma) = \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(Y_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \gamma) \right\|^2 + o_p(1), \qquad (5)$$

where $\psi(\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{X}; \gamma) = \frac{1}{\sigma_0^2} I_{\gamma 22 \cdot 1}^{-1/2} \{\pi(\boldsymbol{X}^T \gamma) I_k - B(\gamma) A^{-1}\} (Y - \boldsymbol{Z}^T \beta_0) \boldsymbol{Z}$.

LRT Given γ

- If several γ's are chosen, the maximum statistic

$$\max_{1 \le j \le J} T(\gamma_j)$$

converges to a limiting distribution, for any $J \geq 1$, where $\gamma_j = (\gamma_{j,-x}, \gamma_{j,x})$ with $\gamma_{j,x} \neq 0$.

- If the chosen γ_j's are away from the true one, we may have lack of power.

LRT Given γ

- If several γ's are chosen, the maximum statistic

$$\max_{1 \le j \le J} T(\gamma_j)$$

converges to a limiting distribution, for any $J \geq 1$, where $\gamma_j = (\gamma_{j,-x}, \gamma_{j,x})$ with $\gamma_{j,x} \neq 0$.

- If the chosen γ_j's are away from the true one, we may have lack of power.

A Better Solution

- Start from a small set $\{\gamma_j : j = 1, \cdots, J\}$;
- For each γ_j as a starting value, perform the EM iterations to find θ and γ for the joint likelihood but stop after just K iterations;
- Compute the LRT-type statistic based on the final values from K EM-iterations: $EM_j^{(K)}$;
- Take the maximum statistic over the initial values of γ_j's:

$$\max_{1 \le j \le J} EM_j^{(K)}.$$

A Model-based Subgroup Inference

- Why iterate? to target the true γ under the alternative.
- Why stop? Iteration towards local maximum leads to more complicated distributions. (Zhu and Zhang, 2004)

- Computationally easier and more stable without having to worry about the convergence of the EM algorithm;
- Based on the EM tests proposed by Chen and Li (2009) for simple Gaussian mixtures, but the effect of X leads to simplifications as well as challenges.

A Model-based Subgroup Inference

- Why iterate? to target the true γ under the alternative.
- Why stop? Iteration towards local maximum leads to more complicated distributions. (Zhu and Zhang, 2004)

- Computationally easier and more stable without having to worry about the convergence of the EM algorithm;
- Based on the EM tests proposed by Chen and Li (2009) for simple Gaussian mixtures, but the effect of X leads to simplifications as well as challenges.

Convergence Theorem

Theorem

Under the null hypothesis and if $\mathbb{E}[\boldsymbol{Z}\boldsymbol{Z}^T] > 0$ *and* $\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T] > 0$, *and for any* K, *the EM test statistic* $EM^{(K)}$ *converges to a fixed distribution as* $n \to \infty$.

More specifically, uniformly in $\gamma \in \Gamma$,

$$EM_j^{(K)} = ||\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(Y_i, \boldsymbol{Z}_i, \boldsymbol{X}_i; \gamma_j)||^2 + o_p(1), \qquad (6)$$

where
$\psi(\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{X}; \gamma) = \frac{1}{\sigma_0^2} I_{\gamma 22 \cdot 1}^{-1/2} \{\pi(\boldsymbol{X}^T\gamma) I_k - B(\gamma)A^{-1}\}(Y - \boldsymbol{Z}^T\beta_0)\boldsymbol{Z}.$

Local Power

Consider the hypothesis testing problem of

$$\begin{aligned} H_0 : &\quad \beta_2 = \mathbf{0}, \ v.s. \\ H_a : &\quad \beta_2 = n^{-1/2}\boldsymbol{h}. \end{aligned} \tag{7}$$

Theorem

Under H_a, for $J = 1$, the test statistic $T_K(\gamma) := EM^{(K)}$, with any value $\gamma \in \Gamma$ and for any positive integer K, converges to a noncentral chi-square distribution with the degree of freedom q_1 and the noncentrality parameter

$$\lambda(\gamma) = \sigma_0^{-2} || \boldsymbol{I}_{\gamma 22 \cdot 1}^{-1/2} (\mathbb{E}(\pi(\boldsymbol{X}^T\gamma)\pi(\boldsymbol{X}^T\gamma_0)\boldsymbol{Z}\boldsymbol{Z}^T) - B(\gamma)A^{-1}B(\gamma_0))\boldsymbol{h}||^2. \tag{8}$$

Local Power

In particular, when $\gamma = \gamma_0$, we have

$$\lambda(\gamma_0) = \sigma_0^{-2} \boldsymbol{h}^T (C(\gamma_0) - B(\gamma_0) A^{-1} B(\gamma_0)) \boldsymbol{h}. \tag{9}$$

We can verify that $\lambda(\gamma)$ achieves its maximum value at $\gamma = \gamma_0$.

The choice of J and K

- With two γ_j's close to each other, they tend to give similar results after iterations.
- From empirical experience, higher values of J or K **do not** bring sufficient gain in power.
- For example, for a small q_2, choose J such γ_j's **cover all quadrants**.
- If q_2 is large, we may choose a small number of $\gamma_j \in \Gamma$ **randomly**.
- A small value of $K = 3$ generally works well.

The Bootstrap Method

- Calculate the test statistic from the data $\{(Y_i, \boldsymbol{Z}_i, \boldsymbol{X}_i) : i = 1, \ldots, n\}$;
- From the J parameter values obtained at the Kth iteration from different initial γ's, choose the one that maximizes $EM_j^{(K)}$: $\Rightarrow (\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}^2)$;
- For $b = 1, \ldots B$,
  - Sample $Y_i^b \sim N(\boldsymbol{Z}_i^T \hat{\beta}_1, \hat{\sigma}^2)$ independently;
  - Calculate the test statistic from each bootstrap sample $\{(Y_i^b, \boldsymbol{Z}_i, \boldsymbol{X}_i) : i = 1, \ldots, n\}$.

EM test simulations: Type-1 errors

$$
\begin{aligned}
P(\delta_i = 1|\boldsymbol{X}_i) &= \pi(\boldsymbol{X}_i^T\gamma) \equiv \frac{\exp(\boldsymbol{X}_i^T\gamma)}{1+\exp(\boldsymbol{X}_i^T\gamma)}, \\
Y_i|(\boldsymbol{Z}_i, \boldsymbol{X}_i, \delta_i) &= \boldsymbol{Z}_i^T(\beta_1 + \beta_2\delta_i) + \varepsilon_i,
\end{aligned} \tag{1}
$$

- $q_1 = 3, q_2 = 2, \boldsymbol{Z} = (1, t, x), \boldsymbol{X} = (1, x)$.
- $x \sim N(-1, 1)$ and $t \sim$ Bernoulli(0.5).
- $\beta_1 = (1, 0, 2)$, $\beta_2 = (0, 0, 0)$.
- The error $\varepsilon \sim N(0, 0.5^2)$.
- The *EM* test uses starting values $\{(1, -2), (1, 2)\}$.

EM test simulations

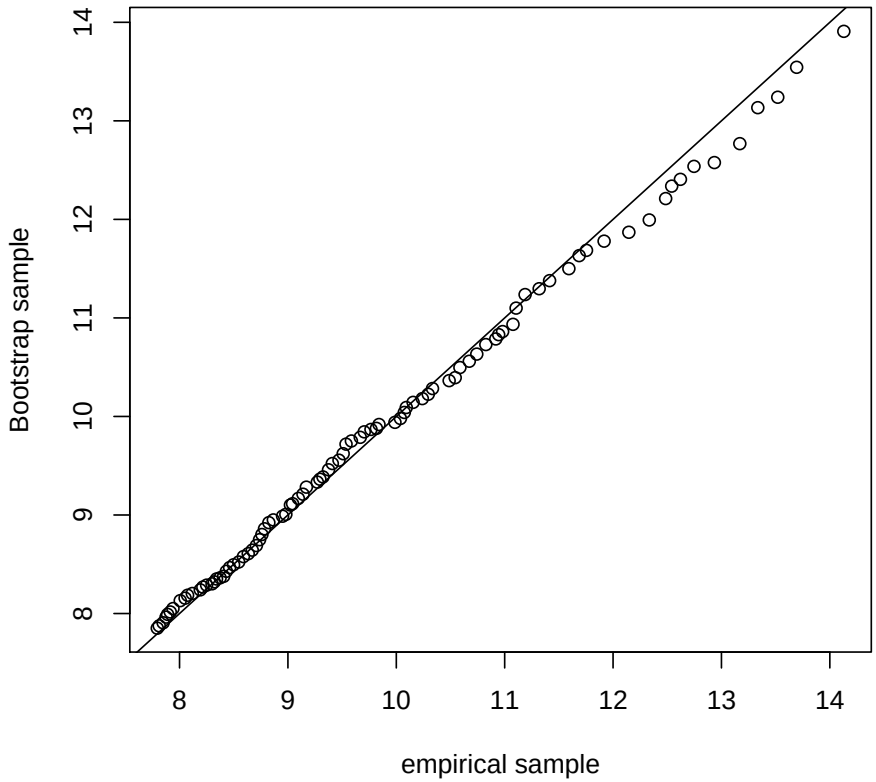


Figure: *Q-Q plot of the upper 10% quantiles of the Bootstrap sample and the empirical sample of the test statistic under the null.*

EM test simulations

Table: *Type 1 errors of the EM tests with bootstrap approximations.*

n	Nominal level α	$EM^{(0)}$	$EM^{(3)}$	$EM^{(9)}$
n=60	0.01	0.011	0.014	0.012
	0.05	0.046	0.055	0.056
	0.10	0.094	0.097	0.099
n=100	0.01	0.010	0.010	0.011
	0.05	0.050	0.051	0.052
	0.10	0.100	0.100	0.095

EM test simulations

- With the sample size as small as 60, the Type 1 errors are quite close to their nominal values, regardless of our choice of $K \in \{0, 3, 9\}$.
- LRT(Oracle): LRT with the fixed true γ.
- It is used as a benchmark (with the true γ) to gauge the performance of the proposed *EM* test.

EM test simulations

- With the sample size as small as 60, the Type 1 errors are quite close to their nominal values, regardless of our choice of $K \in \{0, 3, 9\}$.
- LRT(Oracle): LRT with the fixed true γ.
- It is used as a benchmark (with the true γ) to gauge the performance of the proposed *EM* test.

EM test simulations: Power Comparison

$$
\begin{aligned}
P(\delta_i = 1|\boldsymbol{X}_i) &= \pi(\boldsymbol{X}_i^T\gamma) \equiv \frac{\exp(\boldsymbol{X}_i^T\gamma)}{1+\exp(\boldsymbol{X}_i^T\gamma)}, \\
Y_i|(\boldsymbol{Z}_i, \boldsymbol{X}_i, \delta_i) &= \boldsymbol{Z}_i^T(\beta_1 + \beta_2\delta_i) + \varepsilon_i,
\end{aligned} \tag{1}
$$

- $q_1 = 3, q_2 = 2, \boldsymbol{Z} = (1, t, x), \boldsymbol{X} = (1, x)$.
- $x \sim N(-1, 1)$ and $z \sim$ Bernoulli(0.5).
- $\beta_1 = (1, 0, 2)$, $\beta_2 = (1, \beta_t, \beta_x)$, and $\gamma = (1, \gamma x)$.
- The error $\varepsilon \sim N(0, 0.5^2)$.
- The *EM* test uses starting values $\{(1, -2), (1, 2)\}$.

EM test simulations: Power Comparison

Table: *Power (%) of the EM test at 5% level.* $\gamma = (1, \gamma_x)$, $\beta_2 = (1, \beta_t, \beta_x)$, *and* $n = 100$.

β_t	β_x	γ_x	LRT(Oracle)	$EM^{(3)}$	$EM^{(9)}$
0.5	1	1	97.4	97.0	97.4
0.5	0	1	68.4	52.0	56.8
0.5	1	0	85.4	59.2	66.8
1.0	1	1	97.2	99.4	99.8
1.0	0	1	94.0	91.4	94.6
1.0	1	0	90.2	82.4	84.6

EM test simulations: Power Comparison

Table: *Power (%) of the EM test at 5% level.* $\gamma = (1, \gamma_x)$, $\beta_2 = (1, \beta_t, \beta_x)$, *and* $n = 100$.

β_t	β_x	γ_x	LRT(Oracle)	$EM^{(3)}$	$EM^{(9)}$
0.5	1	1	97.4	97.0	97.4
0.5	0	1	68.4	52.0	56.8
0.5	1	0	85.4	59.2	66.8
1.0	1	1	97.2	99.4	99.8
1.0	0	1	94.0	91.4	94.6
1.0	1	0	90.2	82.4	84.6

Conclusions from simulations

- In finite samples, the EM test is powerful.
- A small number of starting values and a small K work well; so the test is computationally efficient.
- The bootstrap works well for finding critical values or p values.

ACTG320 Study

- $\boldsymbol{X}_i$ contains baseline RNA (rna.0), CD4 counts (cd4.0), and Age.
- $\boldsymbol{Z}_i$ contains those in $\boldsymbol{X}_i$ and the treatment indicator.
- $\boldsymbol{Y}_i$ is the CD4 count changes at the 24th week.
- Primary goal: to confirm if the new three-drug combination has an enhanced treatment effect in a subgroup.
- Secondary goal: to predict the subgroup membership if the subgroup exists.

ACTG320 Study

- From the EM test with $K = 3$, the p-value < 0.001.
- We reject the null hypothesis of no subgroups.
- The treatment effect difference is estimated as 112.98 with s.e. 15.79.

ACTG320 Study

Table: *Coefficient Estimates.*

Coefficients	t	log(*cd*4.0)	$\log_{10}$(*rna*.0)	Age
β_1	41.24	-1.97	6.10	0.69
$\beta_1 + \beta_2$	154.22	20.19	-18.56	-0.23

The score – probability for a subject to fall into the second group is

$$S_1(X) = \pi(-7.89 + 0.44 \log(cd4.0) + 1.10 \log_{10}(rna.0) - 0.02 Age).$$

ACTG320 Study: Two Scores

Compare our scores to the score from Zhao et al. (2013): estimate of $S_2(X) = \mathrm{E}(\mathrm{Y}|\mathrm{t} = 1, \mathrm{X}) - \mathrm{E}(\mathrm{Y}|\mathrm{t} = 0, \mathrm{X})$.

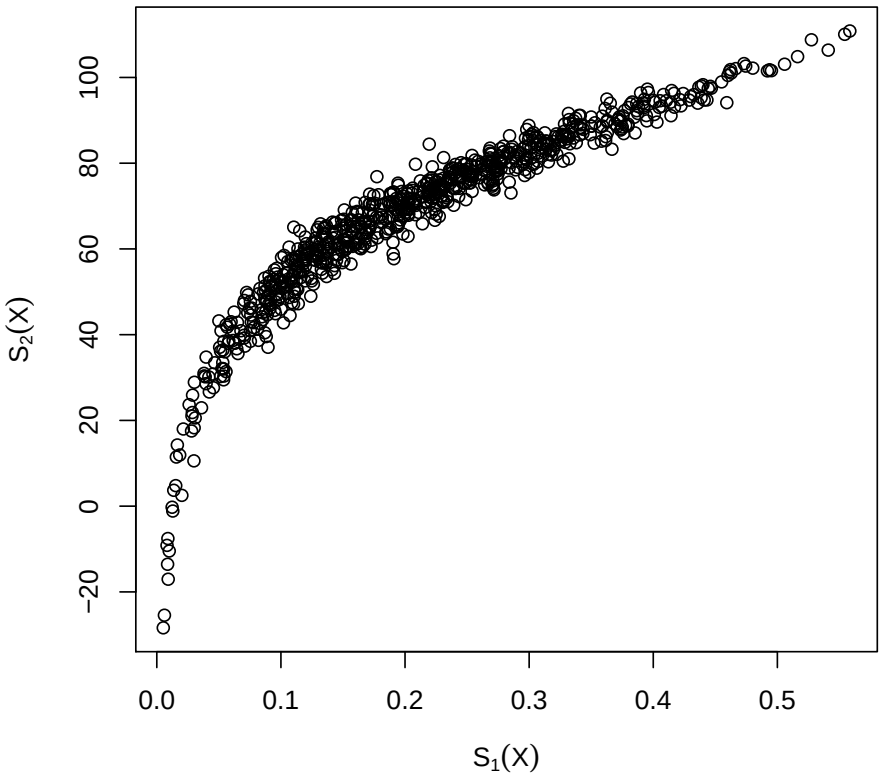


Figure: *Scatter plot of $S_2(X)$ versus $S_1(X)$*

ACTG320 Study

- They have a high rank correlation of 0.98.
- Two scoring systems would make consistent treatment assignments if we determine the treatment by the rank of the scores.
- Our model captures the subgroup characteristics well.

Summary

- We propose a structured logistic-normal mixture model with latent variables, for a **confirmatory** subgroup analysis, and **predictive modeling of subgroup membership** simultaneously.
- The proposed EM test helps **reduce false positive** subgroup identification.
- **More work to do:** what if X and Z are of high dimensions?

Summary

- We propose a structured logistic-normal mixture model with latent variables, for a **confirmatory** subgroup analysis, and **predictive modeling of subgroup membership** simultaneously.
- The proposed EM test helps **reduce false positive** subgroup identification.
- **More work to do:** what if X and Z are of high dimensions?

Model selection

- Penalized likelihood with re-parametrization $(\beta/\sigma, \gamma/\sigma, 1/\sigma)$, with penalizes on the first two parts.
- **Challenges:** lack of convexity.
- Results: The penalized estimate of the model converges, but thresholding is needed to select variables (PhD dissertation of Yingchuan Wang).

Unequal component variance

- The maximum likelihood estimator is not well-defined in this case; but penalization on

$$\frac{S_n^2}{\sigma^2} + \log\left(\frac{S_n^2}{\sigma^2}\right)$$

  for each σ^2 can help, where S_n^2 is the variance estimate from the equal variance model.

- **Challenges:** need to ensure no variance estimate approaching zero.

- Results: The EM test works as usual with improved power if σ_2/σ_1 is high (Shen, Wang and He, 2016).

Unknown number of subgroups

- Sequential testing or model selection?
- A new model selection criterion might be needed.
- Bayesian methods might be helpful.

A bigger issue:

Predictive enrichment study designs:

European Medicines Agency (2010): a concept paper for developing guidelines for the use of subgroup analysis in clinical trials.

FDA (2012): a draft guideline that discussed enrichment designs for clinical trials.

It is the way to go, even if it will be a long way.

Thank you!

We also acknowledge partial support from NSF, NIH and Takeda Pharmaceutical Company.