Latent Gaussian Mixture Models for Nationwide Kidney Transplant Center Evaluation

Presented by: Yehua Li

Joint work with Lanfeng Pan, Kevin He, Yanming Li and Yi Li

Iowa State University and University of Michigan

October 13, 2016

Outline

1. Introduction
2. Methods
3. Simulation
4. Application

Outline

1 Introduction

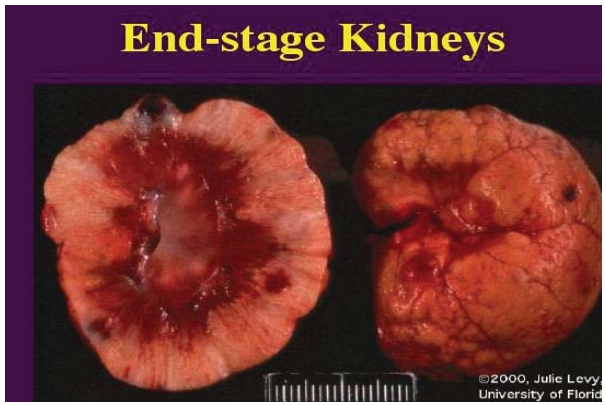
2 Methods

3 Simulation

4 Application

Introduction

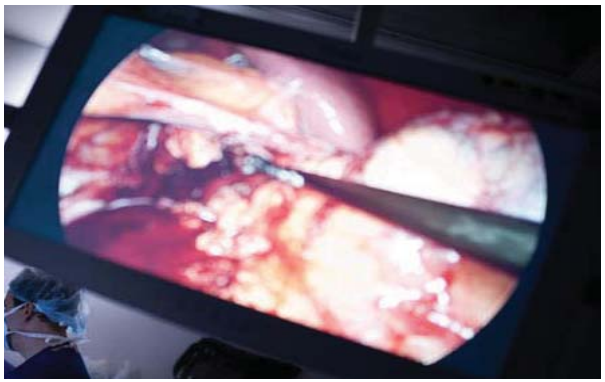
End Stage Renal Disease (ESRD)



- K idneys are functioning below 10 percent of their normal function.
- Progressive and Costly
  - Total patients: 620,000.
  - New cases: 120,000 a year.
  - Total cost per year: $50.0 billion.

Introduction

Kidney Transplantation



- The most desired and cost-effective modality of renal replacement therapy.
- Conducted at 296 transplant centers nationwide.

Evaluation of Quality of Care

- Practice patterns vary across transplant centers.
- Variation reflects the differential practice preferences of physicians, surgeons, therapists, dietitians, etc. from a specific center.
- Important to evaluate transplant centers for quality improvement.
- Post-transplant survival is an important indicator for quality of care delivered by transplant centers.

Data Information

- Data are obtained from the Organ Procurement and Transplantation Network (OPTN)
- All adult transplant patients with the transplant date between January 1987 and December 2008.
- 269,386 patients from 296 transplant centers.
- **X**: demographics of patients (age, gender, BMI), donor information (cold ischemic time).
- Y: 5 year survival status (1 = death, -1=survive) .
- The overall failure rate within 5 years of transplantation is 27.59%.

Model

Generalized linear mixed model

Y_{ik}: observed outcome for the kth patient from the ith center; $\boldsymbol{X}_{ik}$ is a subject-specific covariate.

$$f(Y_{ik}|\boldsymbol{X}_{ik}, \gamma_i; \beta, \varphi) = \exp\left\{\frac{Y_{ik}\xi_{ik} + b(\xi_{ik})}{a(\varphi)} + d(Y_{ik}, \varphi)\right\}.$$

- $\xi_{ik} = \boldsymbol{X}_{ik}^T\boldsymbol{\beta} + \gamma_i$.
- γ_i: effect for center i.
- We use logistic link in our application.

Fixed Effects Models

- Fixed effects provide more precise estimation of true effects for centers with extreme values.
- Variation is large for centers with fewer patients.

Generalized Linear Mixed Effects Models (GLMM)

- Some policy researchers advocate modeling the center effects as Gaussian random effects.
- Most GLMM literature assumes Gaussian random effects.
  - Ignores the heterogeneity among the medical providers.
  - Estimates are shrunk toward the overall mean, and are biased.

Our Proposal: Finite Gaussian Mixture Models

- Allow subgroup (cluster) structures among the centers.
- Model center effects using a finite Gaussian mixture model.
- A compromise between the random effects and fixed effects models.
  - Reduce to the random effects model when there is only one component in the mixture distribution.
  - Become the fixed effects model if each transplant center is a cluster.
- Gaussian mixture models are weakly identifiable with unbounded Fisher information.
- Limited work on GLMM with finite Gaussian mixture random effects.
- Extend recent work by Chen et al. (2012) and Kasahara and Shimotsu (2015) to make inference on the latent Gaussian mixture model.

Outline

1 Introduction

2 Methods

3 Simulation

4 Application

Gaussian mixture random effects with C components

Assume transplant centers belong to C subpopulations

- $\gamma_i \sim \sum_{c=1}^{C} \pi_c N(\mu_c, \sigma_c^2)$.
- $\pi_c \in (0, 1]$ is the weight for subpopulation c, $\sum_{c=1}^{C} \pi_c = 1$.
- $\boldsymbol{\theta}_\gamma = (\mu_1, \ldots, \mu_C, \sigma_1, \ldots, \sigma_C, \pi_1, \ldots, \pi_C)^T$ is the collection of parameters in $g(\gamma)$.

Latent Random Vector

- Define $\boldsymbol{L}_i = (L_{i1}, \ldots, L_{iC})^T$ as a latent random vector of subpopulation memberships.
- $L_{ic} = 1$ if γ_i belongs to component c and $L_{ic} = 0$ otherwise.
- $\boldsymbol{L}_i \sim \text{Multinomial}(\pi_1, \ldots, \pi_C)$.

Log Complete Likelihood

Log complete likelihood

$$\ell_{comp}(\boldsymbol{\theta}; \mathbf{Y}, \mathbf{X}, \boldsymbol{\gamma}, \mathbf{L}) = \sum_{i=1}^{n} \ell_{i,comp}(\boldsymbol{\theta}; Y_i, X_i, \gamma_i, L_i),$$

- $\ell_{i,comp}(\boldsymbol{\theta}; Y_i, X_i, \gamma_i, L_i) = \log f(\boldsymbol{Y}_i | \boldsymbol{X}_i, \gamma_i; \boldsymbol{\theta}_y) + \sum_{c=1}^{C} L_{ic} \log\{\pi_c f_c(\gamma_i | \mu_c, \sigma_c)\}$.
- $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)^T$.

Penalized complete data likelihood

Penalized complete data likelihood

$$\ell_{pen}(\boldsymbol{\theta}; \mathbf{Y}, \mathbf{X}, \gamma, \mathbf{L}) = \ell_{comp}(\boldsymbol{\theta}; \mathbf{Y}, \mathbf{X}, \gamma, \mathbf{L}) + \sum_c p_n(\sigma_c^2; \hat{\sigma}_{c,pilot}^2).$$

- Penalty proposed by Chen and Li (2009)

$$p_n(\sigma^2; \hat{\sigma}_{pilot}^2) = -a_n(\hat{\sigma}_{pilot}^2/\sigma^2 + \log(\sigma^2/\hat{\sigma}_{pilot}^2) - 1),$$

- $\hat{\sigma}_{pilot}^2$ is an initial estimate for the variance of γ and $a_n = o_p(n^{1/4})$.
- The penalty prevents any σ_c^2 from converging to 0.

EM Algorithm

- E-step: calculate a conditional expectation of the complete log-likelihood at the t-th iteration:

$$E\left[\ell_{i,comp}(\boldsymbol{\theta}; Y_i, \boldsymbol{X}_i, \gamma_i, L_i) | Y_i, \boldsymbol{X}_i, \boldsymbol{\theta}^{(t-1)}\right]$$

- M-step: maximize with respect to $\boldsymbol{\theta}$.

Parameter Space and Its Closure

- The parameter space for a model with exactly C components is

$$\Theta_C = \{\boldsymbol{\theta} \mid \boldsymbol{\beta} \in \mathbb{R}^p, \sum_{c=1}^{C} \pi_c = 1, 0 < \pi_c < 1, \mu_1 < \mu_2 < \cdots < \mu_C, \sigma_c > 0, c = 1, 2, \ldots, C\}.$$

- The closure of Θ_C is

$$\bar{\Theta}_C = \{\mid \boldsymbol{\beta} \in \mathbb{R}^p, \sum_{c=1}^{C} \pi_c = 1, 0 \leq \pi_c \leq 1, \mu_1 \leq \mu_2 \leq \cdots \leq \mu_C, \sigma_c \geq 0, \ c = 1, 2, \ldots, C\},$$

  which also includes the over-fitted models, i.e models with less than C components.

Consistency of Estimators

- Parameterization in $\bar{\Theta}_C$ is not unique.
- Let $\boldsymbol{\theta}_0 \in \bar{\Theta}_C$ be the true parameter, define the equivalent class of $\boldsymbol{\theta}_0$

$$\mathbb{F} = \left\{ \boldsymbol{\theta} \in \bar{\Theta}_C;\ \int_{-\infty}^{(\mathbf{x}',\mathbf{y}')} f(\mathbf{x},\mathbf{y}|\boldsymbol{\theta}) d\mu(\mathbf{x},\mathbf{y}) = \int_{-\infty}^{(\mathbf{x}',\mathbf{y}')} f(\mathbf{x},\mathbf{y},|\boldsymbol{\theta}_0) d\mu(\mathbf{x},\mathbf{y}) \text{ for any } (\mathbf{x}',\mathbf{y}') \right\}.$$

Proposition

Under suitable assumptions, $\widehat{\boldsymbol{\theta}}_C$ is consistent in the sense $\inf_{\boldsymbol{\theta}^* \in \mathbb{F}} \|\widehat{\boldsymbol{\theta}}_C - \boldsymbol{\theta}^*\| \to 0$ in probability

Detect Number of Components: Locally Restricted Alternative Space

- Test $H_0 : C_0 = C$ sequentially for $C = 1, 2, \ldots$ and stop when fail to reject a hypothesis.
- A Bonferroni procedure can be used to control the family-wise error rate.
- With non-identifiability, the regular asymptotic theory for likelihood ratio tests does not apply.
- Most importance one is the homogeneity hypothesis $H_0 : C_0 = 1$ vs $H_1 : C_0 = 2$.
- We use a locally restricted likelihood ratio test similar to those Chen et al. (2012) and Kasahara and Shimotsu (2015).

Homogeneity test

- Test $H_0 : C_0 = 1$ vs $H_1 : C_0 = 2$.
- The null model is not unique under the alternative hypothesis.
- For any $\tau \in (0, 0.5]$, define a restricted parameter space

$$\bar{\Theta}_2(\tau) = \{\boldsymbol{\theta} \in \bar{\Theta}_2 \mid \pi_1 = \tau, \pi_2 = 1 - \tau\},$$

- In $\bar{\Theta}_2(\tau)$, the null model is uniquely parameterized by $\boldsymbol{\theta}_{\gamma,0}(\tau) = (\mu_\gamma, \mu_\gamma, \sigma_\gamma, \sigma_\gamma, \tau, 1-\tau)^{\mathrm{T}}$.
- Restricted alternative estimator $\widehat{\boldsymbol{\theta}}_{full}(\tau) = \arg\max_{\boldsymbol{\theta} \in \bar{\Theta}_2(\tau)} \ell_{pen}(\boldsymbol{\theta})$.

Proposition

Under $H_0 : C_0 = 1$ for any fixed $\tau \in (0, 0.5]$, $\widehat{\boldsymbol{\beta}}_{full}(\tau) - \boldsymbol{\beta}_0 = O_p(n^{-1/2})$, and $\widehat{\boldsymbol{\theta}}_{\gamma,full}(\tau) - \boldsymbol{\theta}_{\gamma,0}(\tau) = O_p(n^{-1/4})$.

Locally Restricted Likelihood Ratio Test

- Let the reduced model estimator be $\widehat{\boldsymbol{\theta}}_{red} = \arg\max_{\boldsymbol{\theta} \in \bar{\Theta}_1} \ell_{pen}(\boldsymbol{\theta})$.
- Let $\mathcal{T} = (0, 0.5]$, define the test statistic $\widetilde{T}_1 = \max_{\tau \in \boldsymbol{\mathcal{T}}} T_1(\tau)$, where $T_1(\tau) = 2[l_n\{\widehat{\boldsymbol{\theta}}_{full}(\tau)\} - l_n(\widehat{\boldsymbol{\theta}}_{red})]$ and

$$l_n(\boldsymbol{\theta}; \mathbf{Y}, \mathbf{X}) = \sum_{i=1}^{n} \log \int \left\{ \prod_{k=1}^{n_i} f(Y_{ik} | \mathbf{X}_{ik}, \gamma_i; \boldsymbol{\theta}_y) g(\gamma_i | \boldsymbol{\theta}_\gamma) \right\} d\gamma_i.$$

Proposition

Under $H_0 : C_0 = 1$ and suitable assumptions, $\widetilde{T}_1 \xrightarrow{d} \chi^2(2)$ as $n \to \infty$.

Test for $H_0 : C_0 = C$ vs $H_1 : C_0 = C + 1$ for $C \geq 2$

- Reduced model estimator is $\widehat{\boldsymbol{\theta}}_{red} = \arg\max_{\boldsymbol{\theta} \in \bar{\Theta}_C} l_{pen}(\boldsymbol{\theta})$
- Take turns to test if the cth component of the mixture can be further split into two.
- Define non-overlapping intervals $D_1, \ldots, D_C$ such that $\mu_{c,0} \in D_c$. For a fixed $\tau \in (0, 0.5]$ and $c \in \{1, \ldots, C\}$, define neighborhoods in $\bar{\Theta}_{C+1}$

$$\begin{aligned} \mathcal{N}_{C+1}(c, \tau) \quad = \quad \{\boldsymbol{\theta} \in \bar{\Theta}_{C+1} \mid \tfrac{\pi_c}{\pi_c + \pi_{c+1}} = \tau; \quad & \mu_{c'} \in D_{c'} \text{ for } c' < c; \\ & \mu_c, \mu_{c+1} \in D_c; \quad \mu_{c'} \in D_{c'-1} \text{ for } c' > c+1\}. \end{aligned}$$

- $\widehat{\boldsymbol{\theta}}_{full}(c, \tau) = \arg\max_{\boldsymbol{\theta} \in \mathcal{N}_{C+1}(c,\tau)} l_{pen}(\boldsymbol{\theta})$.

Proposition

Under $H_0 : C_0 = C$, for any fixed $\tau \in (0, 0.5]$,

$$\begin{aligned} \widehat{\mu}_{c,full}(c, \tau) - \mu_{c,0} &= O_p(n^{-1/4}), \quad \widehat{\mu}_{c+1,full}(c, \tau) - \mu_{c,0} = O_p(n^{-1/4}), \\ \widehat{\sigma}_{c,full}(c, \tau) - \sigma_{c,0} &= O_p(n^{-1/4}), \quad \widehat{\sigma}_{c+1,full}(c, \tau) - \sigma_{c,0} = O_p(n^{-1/4}), \end{aligned}$$

and all other estimators converge to the true values in a root-n rate.

Test for $H_0 : C_0 = C$ vs $H_1 : C_0 = C + 1$ for $C \geq 2$ (cont.)

- For a fixed $\tau \in (0, 0.5]$, define $T_C(\tau) = \max_{c \in \{1,2,\ldots,C\}} T_C(c, \tau)$, where $T_C(c, \tau) = 2[l_n\{\widehat{\boldsymbol{\theta}}_{full}(c, \tau)\} - l_n(\widehat{\boldsymbol{\theta}}_{red})]$.
- Let $\mathcal{T}$ be any subset of $(0, 0.5]$, define $\widetilde{T}_C = \max_{\tau \in \mathcal{T}} T_C(\tau)$.

Proposition

The asymptotic distribution for $\widetilde{T}_C$ is that of the maximum of C correlated χ^2_2 random variables.

- We use a simulation procedure to approximate the asymptotic distribution of $\widetilde{T}_C$, where the correlations between the χ^2 variables are computed using the theoretical expression.

Transplant center evaluation

- Following Efron (2004), identify the "empirical null" distribution of γ as a subset of components in the mixture density, $g_0(\gamma|\boldsymbol{\theta}_\gamma) = \sum_{c\in\mathcal{C}_0} \pi_c f_c(\gamma|\mu_c, \sigma_c) / \sum_{c\in\mathcal{C}_0} \pi_c$ where $\mathcal{C}_0 \subset \{1, 2, \ldots, C\}$.
- For each transplant center i, test if it belongs to one of the components in $\mathcal{C}_0$, or $H_{i0} : \sum_{c\in\mathcal{C}_0} L_{ic} = 1$, $i = 1, \ldots, n$.
- Suppose $\mathcal{C}_0$ consists of centers of average performance, then center i is considered "interesting" (either outperforming or underperforming) if H_{i0} is rejected.

Transplant center evaluation (cont.)

- We address this multiple hypothesis testing problem using local FDR

$$\begin{aligned} lFDR_i &= P(\textstyle\sum_{c\in\mathcal{C}_0} L_{ic} = 1 | \boldsymbol{X}_i, \boldsymbol{Y}_i) \\ &= \frac{\sum_{c\in\mathcal{C}_0} \pi_c \int f(\boldsymbol{Y}_i|\boldsymbol{X}_i, \gamma; \boldsymbol{\beta}) f_c(\gamma|\mu_c, \sigma_c) d\gamma}{\int f(\boldsymbol{Y}_i|\boldsymbol{X}_i, \gamma; \boldsymbol{\beta}) g(\gamma|\boldsymbol{\theta}_\gamma) d\gamma}. \end{aligned}$$

- **Benjamini-Hochberg** procedure: Let $lFDR_{(1)} \leq lFDR_{(2)} \leq \cdots \leq lFDR_{(n)}$ be the ranked lFDR values. For any $\alpha > 0$, let $k = \max_i\{\frac{1}{i}\sum_{j=1}^{i} lFDR_{(j)} \leq \alpha\}$ and our FDR control procedure is to reject all H_{i0} with the rank of $lFDR_i$ less or equal to k.

Transplant center evaluation (cont.)

- We address this multiple hypothesis testing problem using local FDR

$$\begin{aligned} lFDR_i &= P(\textstyle\sum_{c\in\mathcal{C}_0} L_{ic} = 1 | \boldsymbol{X}_i, \boldsymbol{Y}_i) \\ &= \frac{\sum_{c\in\mathcal{C}_0} \pi_c \int f(\boldsymbol{Y}_i|\boldsymbol{X}_i, \gamma; \boldsymbol{\beta}) f_c(\gamma|\mu_c, \sigma_c) d\gamma}{\int f(\boldsymbol{Y}_i|\boldsymbol{X}_i, \gamma; \boldsymbol{\beta}) g(\gamma|\boldsymbol{\theta}_\gamma) d\gamma}. \end{aligned}$$

- **Benjamini-Hochberg** procedure: Let $lFDR_{(1)} \leq lFDR_{(2)} \leq \cdots \leq lFDR_{(n)}$ be the ranked lFDR values. For any $\alpha > 0$, let $k = \max_i\{\frac{1}{i}\sum_{j=1}^{i} lFDR_{(j)} \leq \alpha\}$ and our FDR control procedure is to reject all H_{i0} with the rank of $lFDR_i$ less or equal to k.

Proposition

Under the latent Gaussian mixture model, the above procedure controls FDR at level α.

Outline

1 Introduction

2 Methods

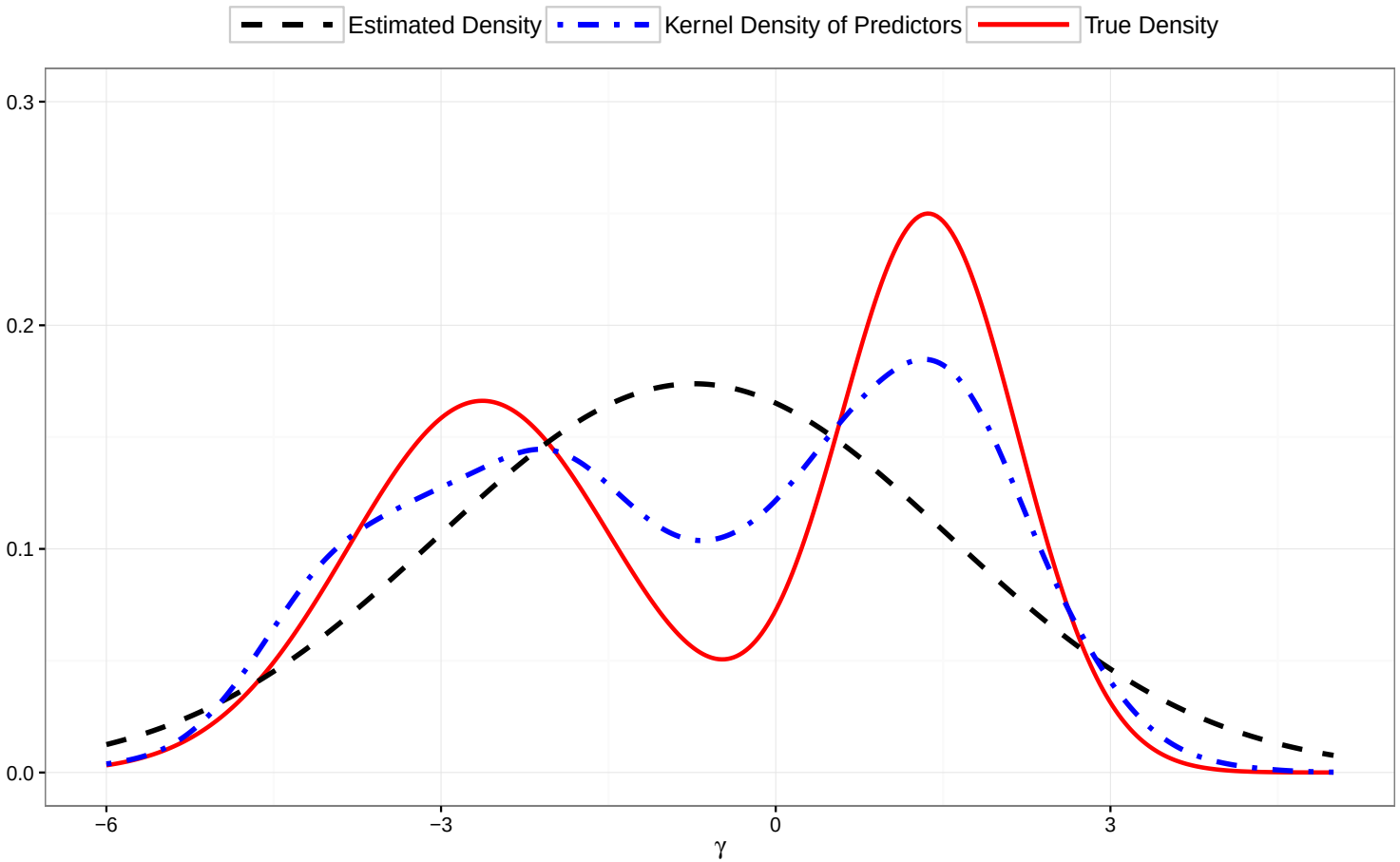
3 Simulation

4 Application

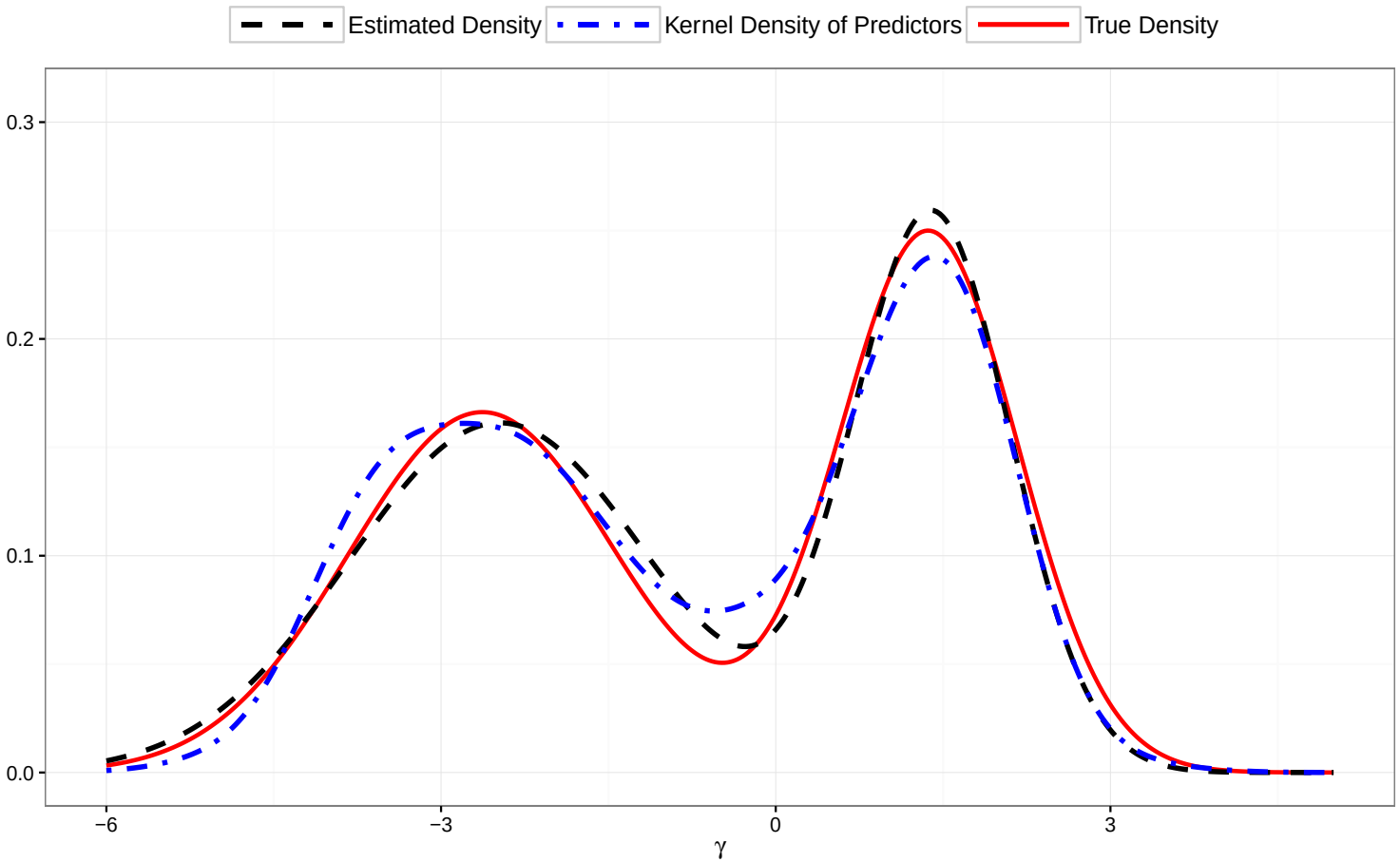
Simulation Setup

- We use the same number of centers as in real data.
- Roughly 1/10 of the number subject per center.
- Y_{ik} are generated from logistic regression model and γ_i are generated from various finite Gaussian mixture models.

Assume One Component

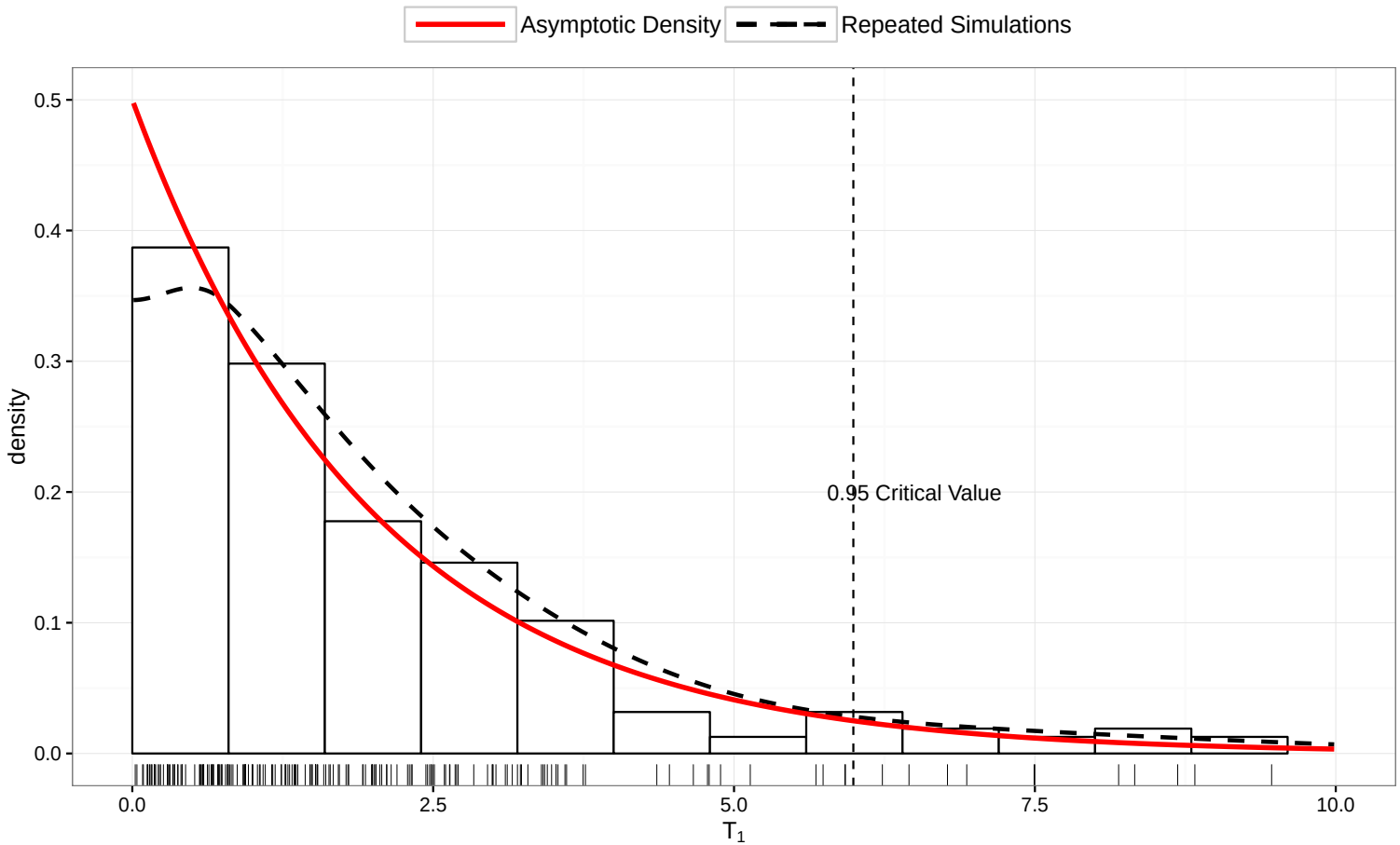


Assume Two Components



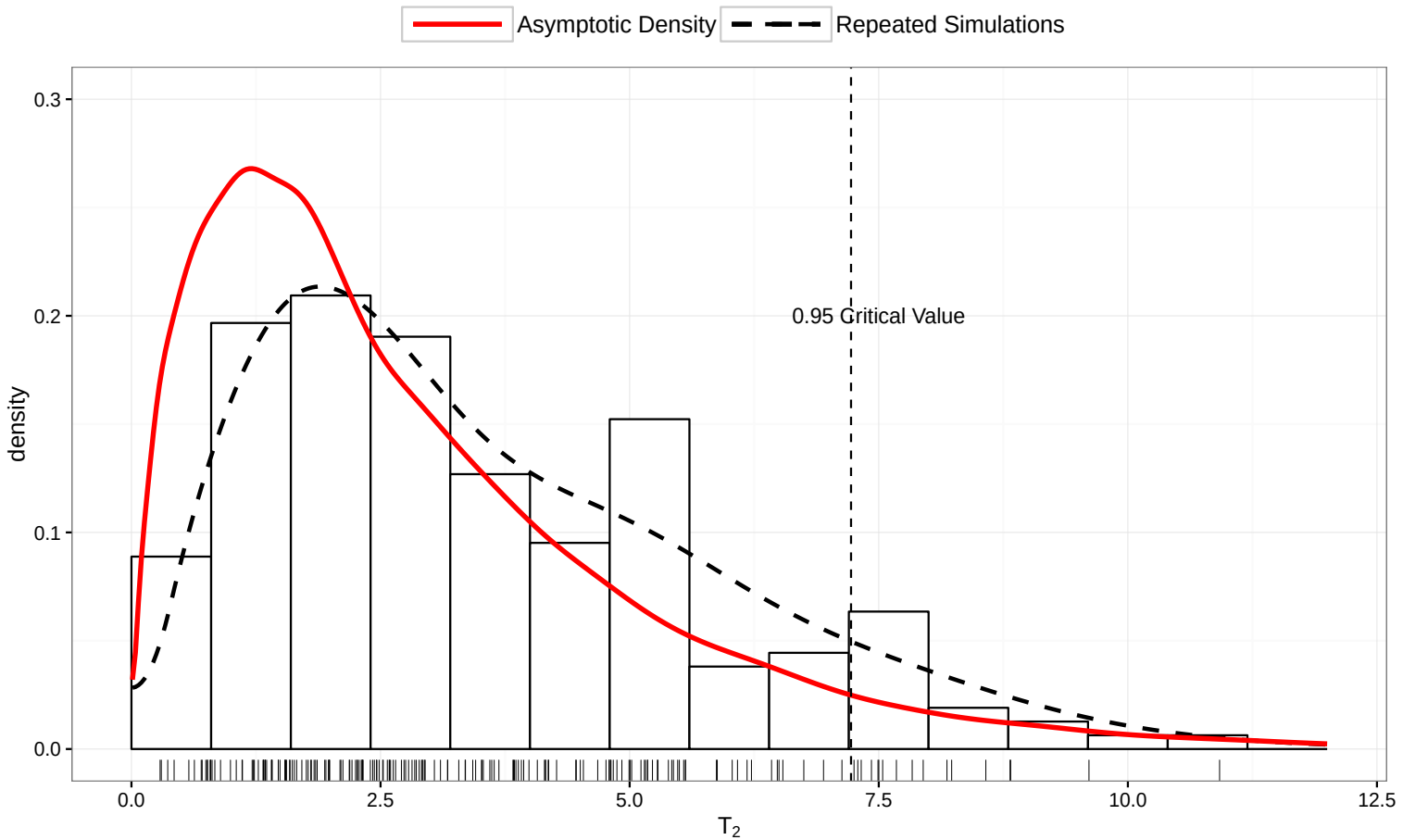
Asymptotic Distribution

Theoretical vs. empirical distribution of $\widetilde{T}_1$ when $H_0 : C = 1$ is true.



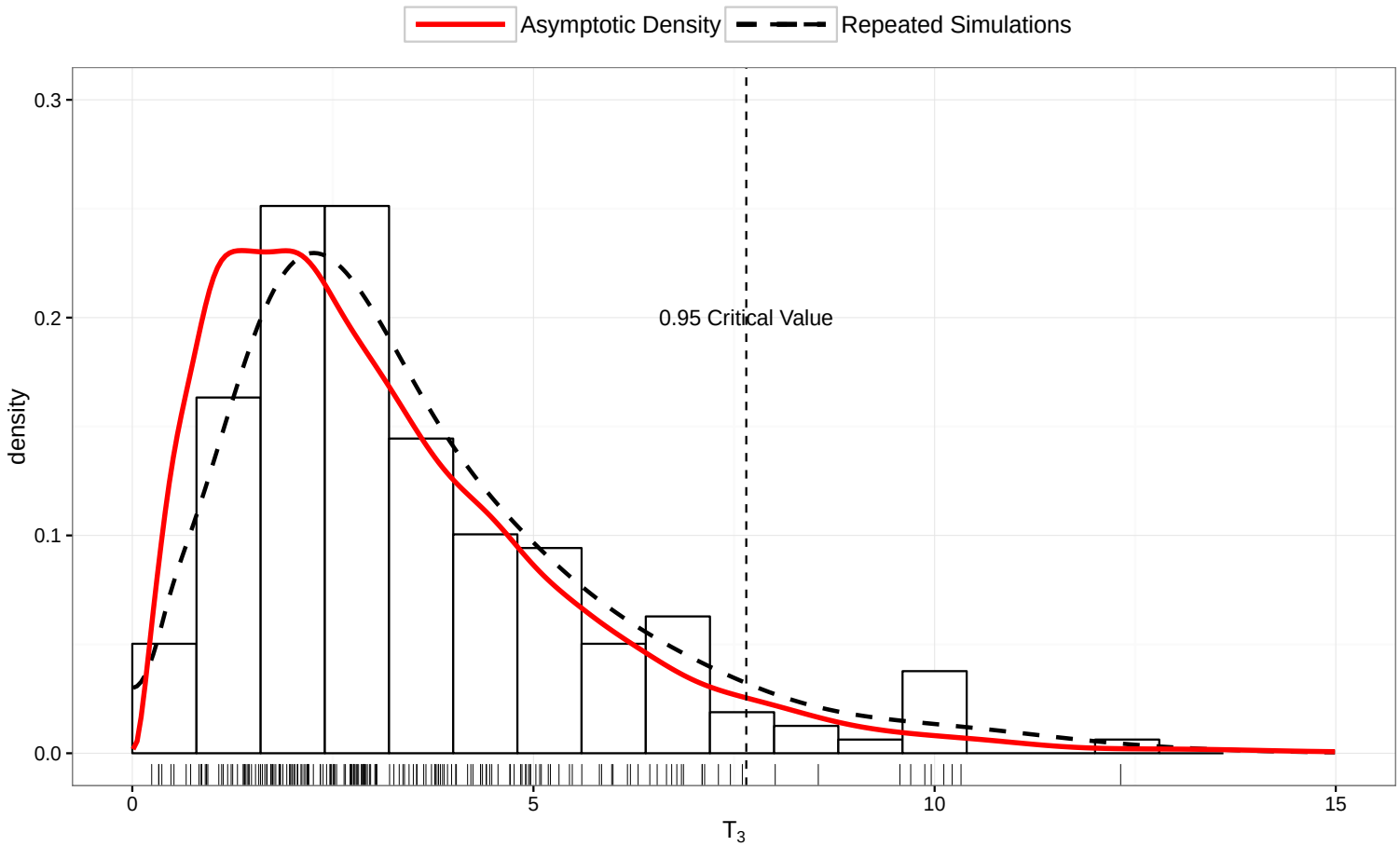
Asymptotic Distribution

Theoretical vs. empirical distribution of $\widetilde{T}_2$ when $H_0 : C = 2$ is true.



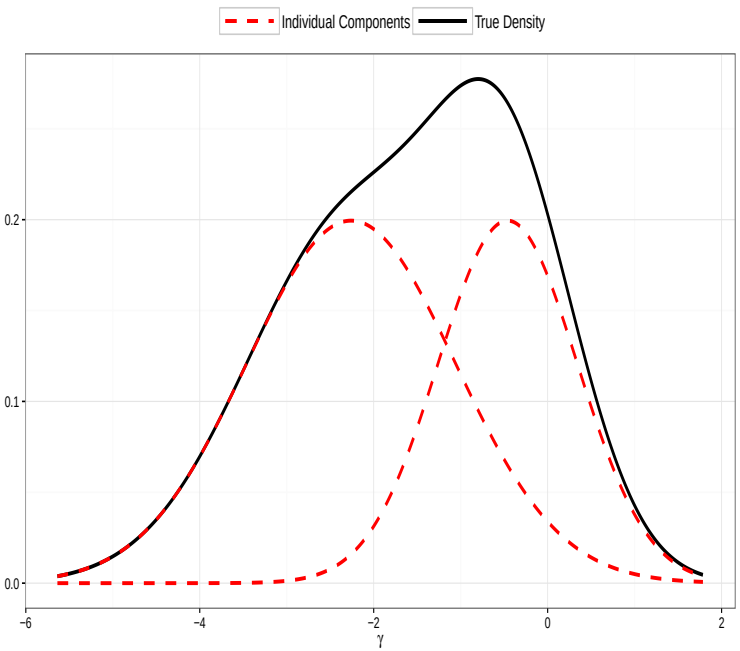
Asymptotic Distribution

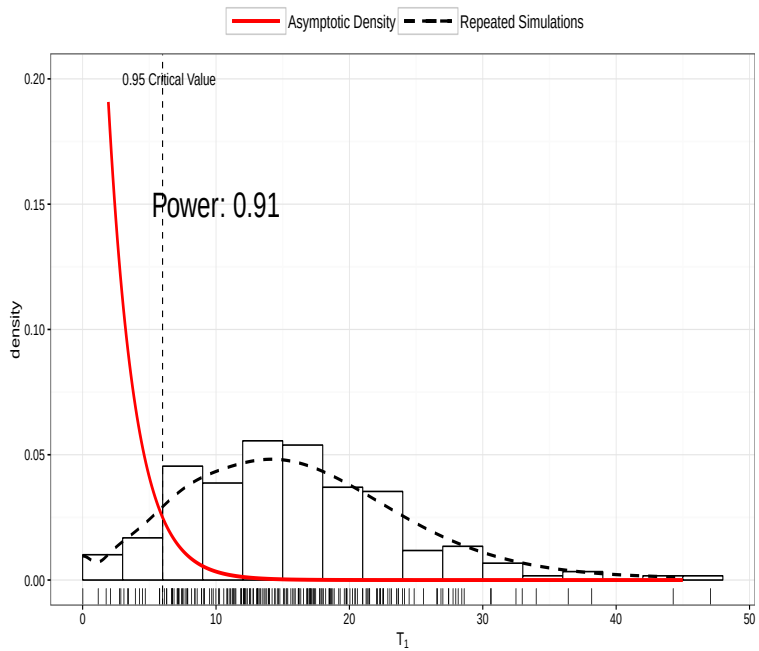
Theoretical vs. empirical distribution of $\widetilde{T}_3$ when $H_0 : C = 3$ is true.



Power

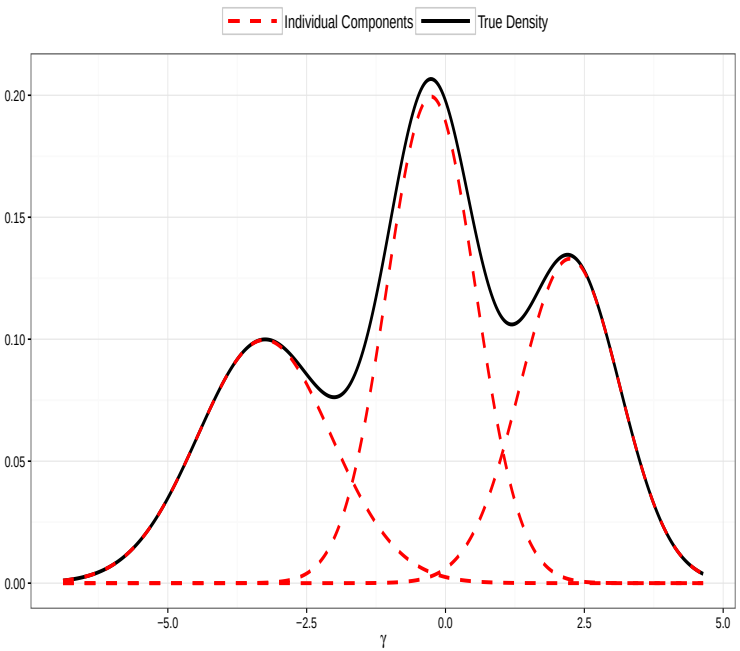
Testing $H_0 : C = 1$ when the true distribution is a 2-component mixture.

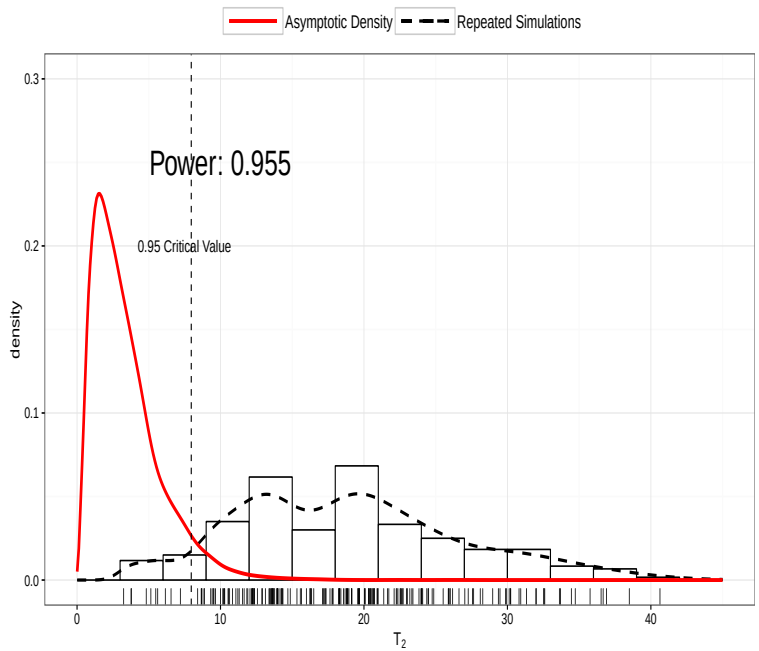




Power

Testing $H_0 : C = 2$ when the true distribution is a 3-component mixture.





Model selection: BIC vs. Sequential test

- Sequential test procedure with $\alpha = 0.05$ vs. a Naive BIC.
- For the 2-component mixture model, BIC picks the correct model 39% of the time (usually underfit).
- For the 3-component model, BIC chooses the correct number of components 50.5% of the time.
- The sequential test chooses the correct number of components 88.5% and 86% of the time for the two models.

Outline

1 Introduction

2 Methods

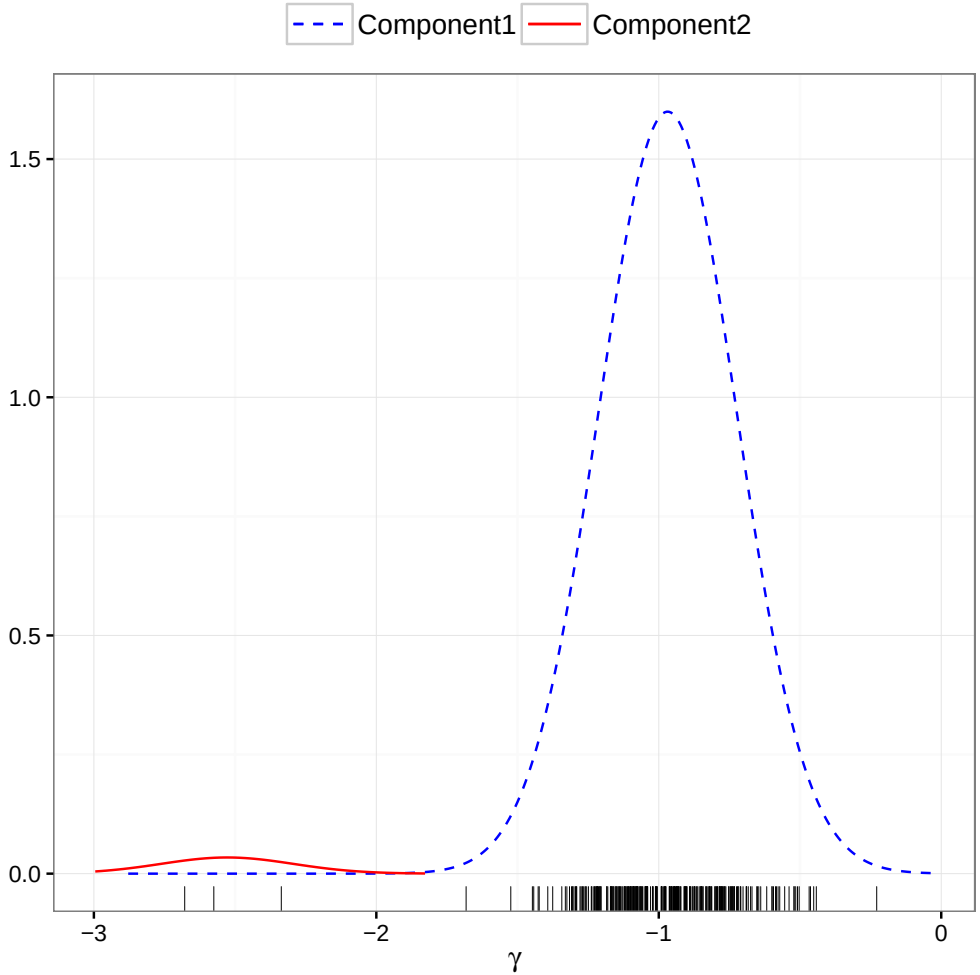
3 Simulation

4 Application

A Cohort of Patients from United Network for Organ Sharing

- Transplant patients with the transplant date between January 1987 and December 2008.
- $N = 269,386$ patients from 296 transplant centers.
- **X**: demographics of patients, donor information.
  - Age, gender (1=male, 0=female), BMI and cold ischemic time.
- **Y**: 5 year survival status (1=death and -1=survival).
  - The overall failure rate within 5 years of transplantation is 27.59%.

Data Analysis



Data Analysis (Cont.)

- The density for the predicted center effects is left skewed with a minor mixture component on the left tail.
- Three transplant centers in the small component are flagged as performing significantly better than the majority.

Characteristics of flagged centers

Center id	lFDR	$\hat{\gamma}$	Sample Size	Survival Rate
#287	0.0013	-2.6784	114	0.973
#10	0.0061	-2.5753	125	0.944
#28	0.0736	-2.3364	120	0.841

Data Analysis (Cont.)

Fixed effects estimation

	Estimate	Std. Error	z-value	p-value
cold ischemic time	0.019503	0.0003048	63.9869	<1e-99
age	0.007112	0.0002117	33.5890	<1e-99
sex	0.030928	0.0094616	3.2688	0.0011
BMI (22,25]	0.077860	0.0154998	5.0232	<1e-6
BMI (25,30]	0.120536	0.0129628	9.2986	<1e-19
BMI 30+	0.225015	0.0148196	15.1836	<1e-51

- The longer the cold ischemic time, the older the patient at transplantation, the higher the post-transplantation death rate.
- Males and patients with high BMI have higher death rates.
- We also included indicator for time periods when the surgery was performed. 5-year death rate is decaying over time due to new technology used in surgery and recovery.

Summary

- Model transplant center effects by using finite Gaussian mixture random effects.
- Study likelihood ratio tests to determine the number of components in the Gaussian mixture distribution.
- The fitted mixture model provides a convenient means of classifying transplant centers.

Thank You!