Selecting Moment Conditions in the GMM for Mixture Models

Zhiyue Huang
MRC Biostatistics Unit
Cambridge, UK

Introduction

- Consider the class of nonparametric mixture models

$$f_{\text{Mix}}(x;Q) = \int_a^b f(x;\theta)\mathrm{d}Q, \quad x \in \mathcal{S}, \tag{1}$$

where for each $\theta \in [a,b]$, the component distribution $f(x;\theta)$ is in the exponential family with the sample space $\mathcal{S}$, and the mixing distribution $Q(\theta)$ is a probability measure over a known, compact set $[a,b]$.

- Huang (2015) proposed the generalized method of moments (GMM) for mixture models. By choosing generalized moment conditions, the GMM estimator could be robust to the outliers at the cost of losing efficiency.
- Our main objective is to investigate the robustness and efficiency in the GMM, when two different classes of generalized moment conditions are used.

The GMM for Mixture Models

- Assume that, for each $\theta \in [a,b]$, there exists a strictly positive function of $x \in \mathcal{S}$, say $g(x)$, such that $f(x;\theta)g^{-1}(x) \in L^2(\mathcal{S}, \nu_0)$, where $\mathrm{d}\nu_0 = \mathrm{d}x$. Let $\{\psi_j(x)\}_{j=1}^{\infty}$ be the eigenfunctions associated with the eigenvalues $\lambda_1 > \lambda_2 > \cdots > 0$, of the positive definite integral operator

$$(Ah)(x) = \int_{\mathcal{S}} h(x')K(x,x')\mathrm{d}x', \tag{2}$$

where

$$K(x,x') = \int_a^b \frac{f(x;\theta)}{g(x)}\frac{f(x';\theta)}{g(x')}\mathrm{d}\theta. \tag{3}$$

- Under regularity conditions, we have the expansion

$$f(x;\theta) = \sum_{j=1}^{\infty} u_j(\theta)\psi_j(x)g(x), \tag{4}$$

where for each j,

$$u_j(\theta) = \int_{\mathcal{S}} \psi_j(x)f(x;\theta)g^{-1}(x)\mathrm{d}x. \tag{5}$$

- Assuming that the order of the integral and the infinite sum is exchangeable, we have

$$f_{\text{Mix}}(x;Q) = \sum_{j=1}^{\infty} m_j\psi_j(x)g(x), \tag{6}$$

where for each j,

$$m_j(Q) = \int_a^b u_j(\theta)\mathrm{d}Q = \mathrm{E}_X\left[\psi_j(x)g^{-1}(x)\right]. \tag{7}$$

- It follows a truncation approximation of $f_{\text{Mix}}(x;Q)$

$$f_{\text{T}}(x;\boldsymbol{m}_{(J)}) = \sum_{j=1}^{J} m_j\psi_j(x)g(x), \tag{8}$$

where $\boldsymbol{m}_{(J)} \in \mathrm{R}^J$ is the vector of the generalized moments induced by $\{u_j(\theta)\}_{j=1}^{J}$.

Definition 1 (GMM for mixture models)**.**
Given a random sample $X_1, \ldots, X_N$ from $f_{\text{Mix}}(x;Q^*)$, the GMM estimator of $\boldsymbol{m}_{(J)} \in \mathrm{R}^J$ is

$$\hat{\boldsymbol{m}}_{(J)} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{\phi}_{(J)}(X_n), \tag{9}$$

where J is a fixed number and

$$\boldsymbol{\phi}_{(J)}(x) = (\psi_1(x)g^{-1}(x), \ldots, \psi_J(x)g^{-1}(x))^{\mathrm{T}} \in \mathrm{R}^J. \tag{10}$$

The GMM estimator for $f_{\text{Mix}}(x;Q)$ is

$$f_{\text{T}}(x;\hat{\boldsymbol{m}}_{(J)}) = g(x)\frac{1}{N}\sum_{n=1}^{N} \boldsymbol{\phi}_{(J)}^{\mathrm{T}}(X_n)\boldsymbol{\psi}_{(J)}(x). \tag{11}$$

- If J diverges with the sample size N, the GMM estimator $f_{\text{T}}(x;\boldsymbol{m}_{(J)})$ point-wisely converges to $f_{\text{Mix}}(x;Q^*)$ in probability.
- When the sample size N is finite, the number J can be chosen by balancing the bias and variance of $f_{\text{T}}(x;\hat{\boldsymbol{m}}_{(J)})$.

Conclusions

- The GMM estimators $f_{\text{T}}(x;\hat{\boldsymbol{m}}_{(J)})$ can have different properties, when different $g(x)$ are used.
- When $g_0 = 1$, the resulting GMM estimator is robust to the outliers in data.
- When $g_0 = f_0^{1/2}(x)$, the GMM estimator for $f_{\text{Mix}}(x;Q)$ is more efficient.
- In either of the considered cases, the bias of the GMM estimator is small and decays with the increase of J.

Forthcoming Research

In this poster, we only considered two possible choices of $g(x)$. More numerical studies are needed for other choices of $g(x)$. Theoretical results are also needed to provide a guideline in selecting $g(x)$.

References

Z. Huang. (2015). *The Generalized Method of Moments for Mixture and Mixed Models* (Unpublished doctoral thesis). University of Waterloo, Canada.

Example of the Mixture of Poisson Distributions

- Consider the mixture of Poisson distributions, $\text{Pois}(x;\theta)$, $\theta \in [a,b]$, where $a = 0.1$ and $b = 10$.
- Compare the performances of the two classes of generalized moment conditions:
  1. $g(x) = 1$;
  2. $g(x) = f_0^{1/2}(x)$, where $f_0(x) = (b-a)^{-1}\int_a^b f(x;\theta)\mathrm{d}\theta$.

Robustness to the outliers

Figure 1 shows the functions $\psi_j(x)g^{-1}(x)$, $j = 1,2,3,4$, from the two classes of generalized moment conditions.

- When $g(x) = 1$, the functions $\psi_j(x)g^{-1}(x)$ are bounded. Thus, the robustness to the outliers is expected.
- When $g(x) = f_0^{1/2}(x)$, the functions $\psi_j(x)g^{-1}(x)$ diverges as x goes to infinity. Thus, the GMM estimator $\hat{\boldsymbol{m}}_{(J)}$ is not robust to the outliers.

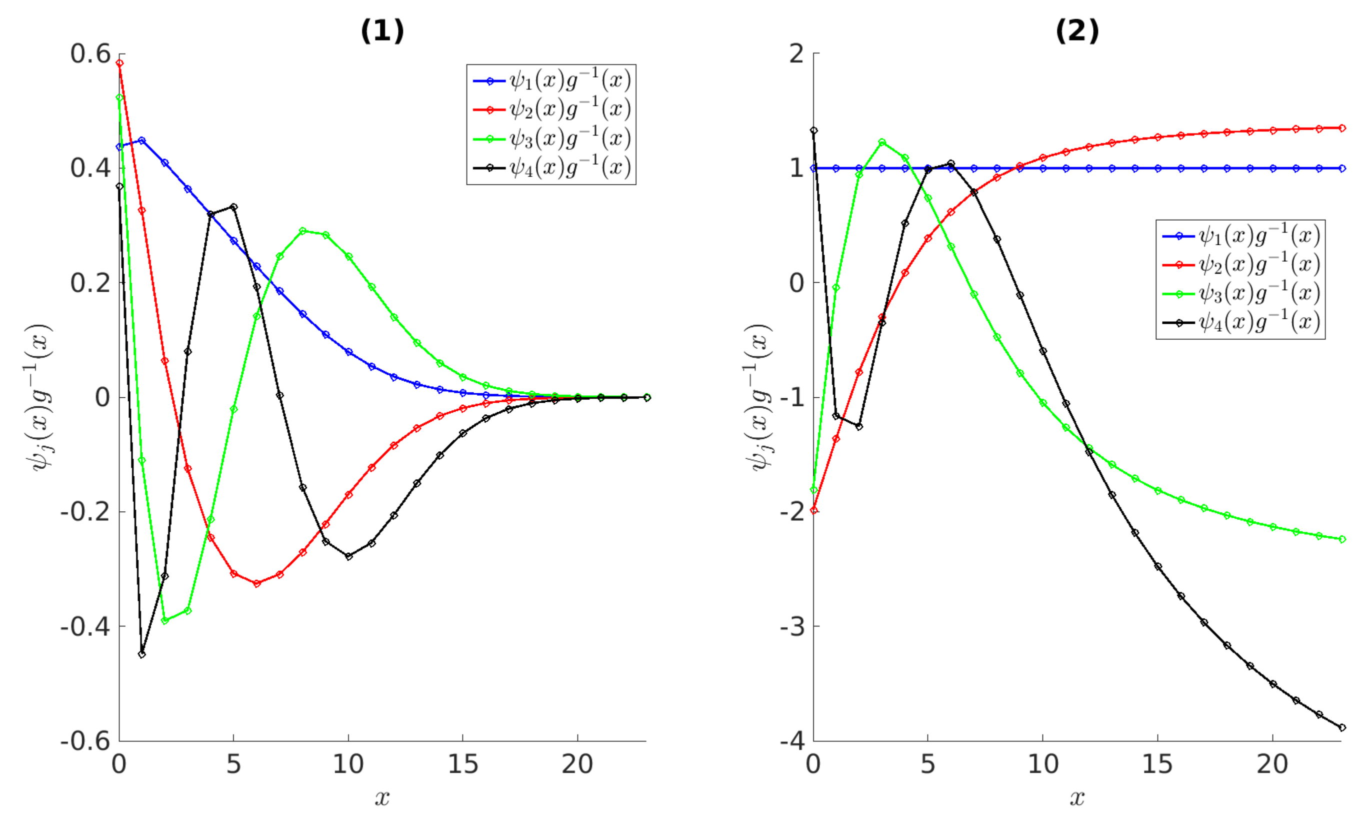

Figure 1: The functions $\psi_j(x)g^{-1}(x)$, $j = 1,2,3,4$, when (1) $g(x) = 1$ and (2) $g(x) = f_0^{1/2}(x)$.

Bias from the truncation approximation

Figure 2 presents the approximation errors $f(x;\theta) - f_{\text{T}}(x;\boldsymbol{u}_{(J)}(\theta))$ over $\mathcal{S} \times [a,b]$, when $J = 5$.

- In both cases, $f_{\text{T}}(x;\boldsymbol{u}_{(J)}(\theta))$ appropriately approximate $f(x;\theta)$ in pointwise.
- In the case $g(x) = f_0^{1/2}(x)$, the approximation $f_{\text{T}}(x;\boldsymbol{u}_{(J)}(\theta))$ has smaller bias, when $x > 10$.

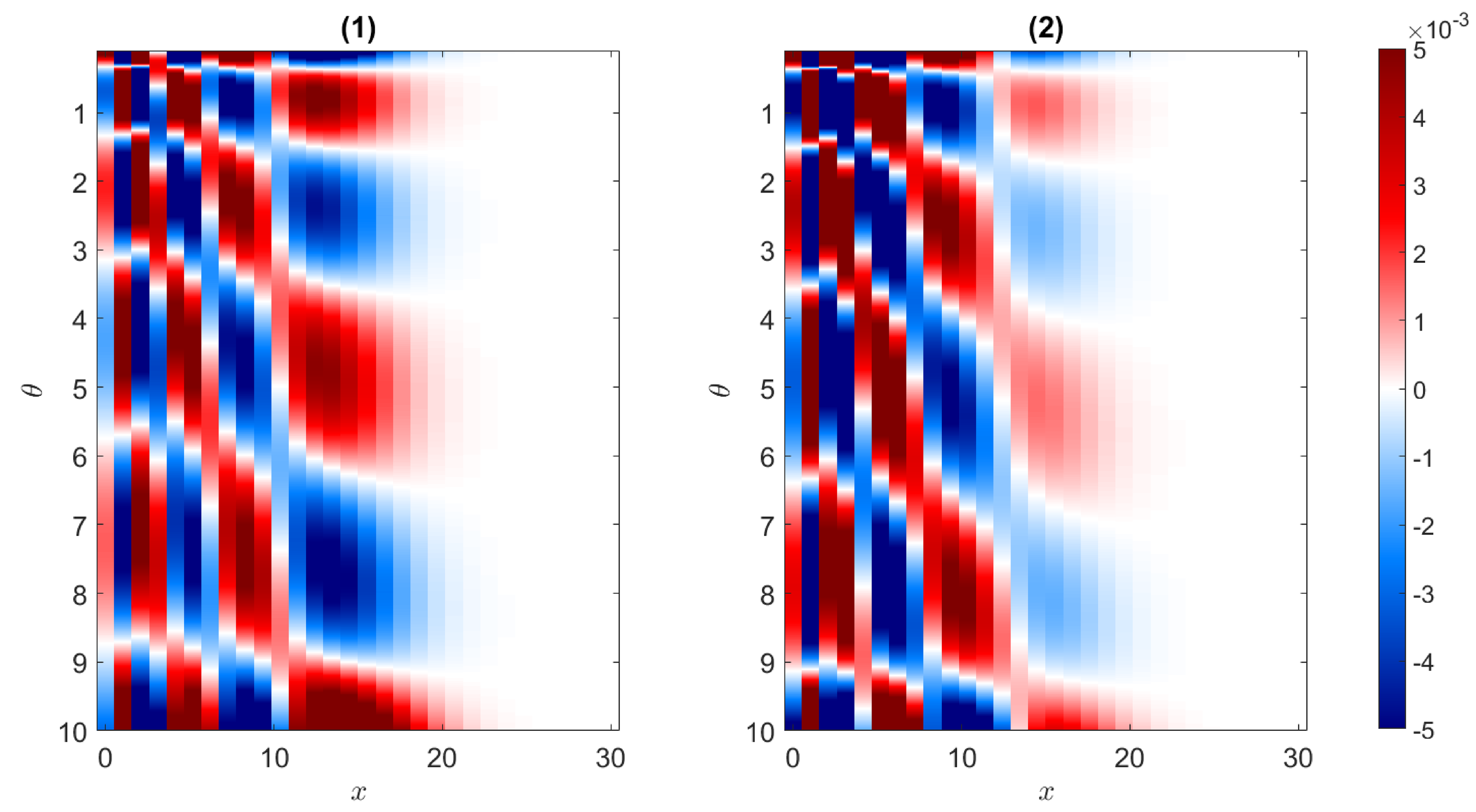

Figure 2: The approximation errors of $f_{\text{T}}(x;\boldsymbol{u}_{(J)}(\theta))$, $J = 5$, over $\mathcal{S} \times [a,b]$, when (1) $g(x) = 1$ and (2) $g(x) = f_0^{1/2}(x)$.

Variance of $f_{\text{T}}(x;\hat{\boldsymbol{m}}_{(J)})$

Figure 3 shows the standard deviation of $f_{\text{T}}(x;\hat{\boldsymbol{m}}_{(J)})$ when the true mixing distribution is

- $Q_1 = \text{Unif}([2,8])$;
- $Q_2 = 0.5\delta(\theta = 3) + 0.5\delta(\theta = 5)$;
- $Q_3 = 0.5\delta(\theta = 8.95) + 0.5\delta(\theta = 9.05)$;
- $Q_4 = 0.99\delta(\theta = 3) + 0.01\delta(\theta = 5)$.

When $g(x) = f_0^{1/2}(x)$, the GMM estimator $f_{\text{T}}(x;\hat{\boldsymbol{m}}_{(J)})$ has smaller variances over $\mathcal{S}$.

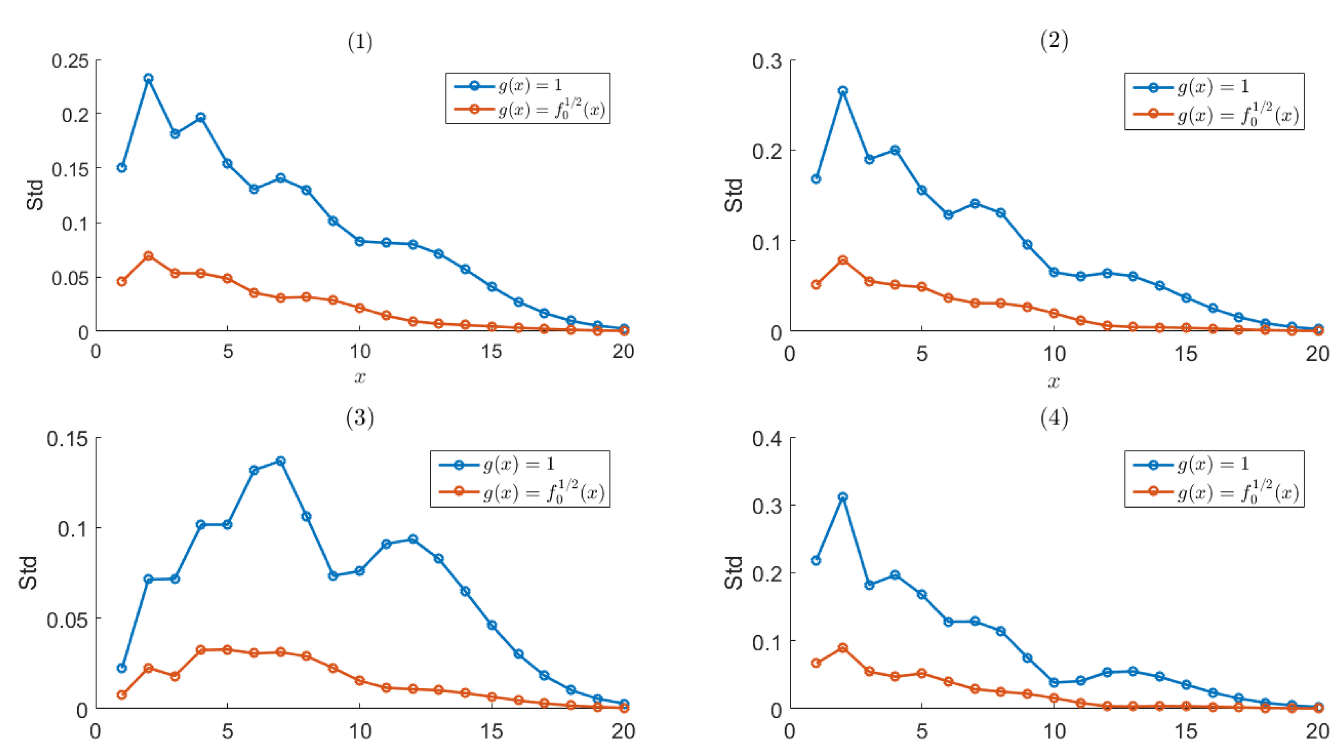

Figure 3: The variance of $f_{\text{T}}(x;\hat{\boldsymbol{m}}_{(J)})$, $J = 5$, when the random sample is from $f_{\text{Mix}}(x;Q_i)$, $i = 1,2,3,4$.