Supplementary Material for "Fast Bootstrapping Nonparametric Maximum Likelihood for Latent Mixture Models"

Shijie Wang, Minsuk Shin, and Ray Bai

A. Sensitivity Analysis for Generator Architecture

We conduct a sensitivity analysis to study whether different choices for the number of hidden layers L or the number of hidden neurons h in the generator G affects GB-NPMLE's performance. We first fix $L = 2$ and vary $h \in \{50, 250, 500, 750\}$. Next, we fix $h = 500$ and vary $L \in \{1, 2, 3, 4\}$.

We repeat all of the simulations described in Section III-A of the main article 20 times and record the average performance metrics (i.e. $W_1(\pi, \widehat{\pi})$ and ISE). Our results are summarized in Table I. We do not observe any significant differences in the performance of GB-NPMLE for these different combinations of (L, h). In practice, we recommend a default choice of $L = 2$ and $h = 500$ for satisfactory performance.

TABLE I
SENSITIVITY ANALYSIS RESULTS

$L = 2$	$h = 50$		$h = 250$		$h = 500$		$h = 750$	
Model	$W_1(\pi, \widehat{\pi})$	ISE	$W_1(\pi, \widehat{\pi})$	ISE	$W_1(\pi, \widehat{\pi})$	ISE	$W_1(\pi, \widehat{\pi})$	ISE
GMM	0.334	0.008	0.326	0.008	0.348	0.008	0.329	0.008
GaMM	0.032	0.223	0.032	0.282	0.032	0.263	0.032	0.245
PMM	0.413	0.033	0.380	0.028	0.400	0.045	0.386	0.030

$h = 500$	$L = 1$		$L = 2$		$L = 3$		$L = 4$	
Model	$W_1(\pi, \widehat{\pi})$	ISE	$W_1(\pi, \widehat{\pi})$	ISE	$W_1(\pi, \widehat{\pi})$	ISE	$W_1(\pi, \widehat{\pi})$	ISE
GMM	0.330	0.008	0.348	0.008	0.331	0.008	0.332	0.008
GaMM	0.032	0.292	0.032	0.263	0.032	0.304	0.032	0.270
PMM	0.400	0.029	0.400	0.045	0.398	0.035	0.393	0.036

B. More Simulation Studies

We investigate the performance of GB-NPMLE under two additional settings for the prior π: (iv) a trimodal density, and (v) a bounded and skewed left density. Namely, we consider:

(iv) Gaussian trimodal mixture (**GMM-tri**): $y \mid \theta \sim \mathcal{N}(\theta, 1)$ and $\theta = 0.2\mathcal{N}(-4, 0.5) + 0.6\mathcal{N}(0, 1) + 0.2\mathcal{N}(4, 0.5)$;

(v) Binomial-beta mixture (**BBM**): $y \mid \theta \sim \text{Binomial}(10, \theta)$ and $\theta \sim \text{Beta}(3, 2)$.

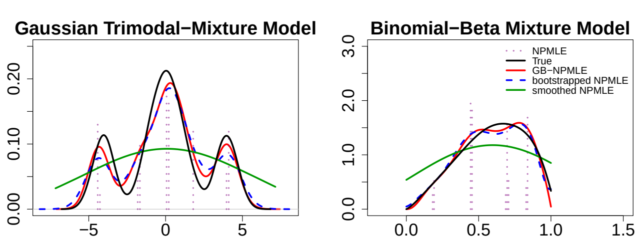

Fig. 1. Results from one replication of Simulations (iv)-(v). In addition to GB-NPMLE (solid red), bootstrapped NPMLE (dashed blue), and smoothed NPMLE (solid green), we also plot the true density (solid black) and the classical discrete NPMLE (dotted purple).

The authors S. Wang and R. Bai are with the Department of Statistics, University of South Carolina, SC 29208 USA (e-mail: shijiew@email.sc.edu; rbai@mailbox.sc.edu). M. Shin is with Gauss Labs, CA 94301 USA (e-mail: minsuk000@gmail.com).

As shown in Fig. 1, GB-NPMLE approximates the bootstrapped NPMLE quite well for both GMM-tri and BBM. GB-NPMLE and bootstrapped NPMLE also provide density estimates that are close to the true π, whereas the smoothed NPMLE with bandwidth chosen from LOOCV is once again oversmoothed. A performance comparison of GB-NPMLE, bootstrapped NPMLE, and smoothed NPMLE based on the average of 20 replications is summarized in Table II.

TABLE II
COMPARISONS OF PERFORMANCE OF DIFFERENT NPMLE METHODS

	GB-NPMLE		Bootstrapped NPMLE		Smoothed NPMLE	
Model	$W_1(\pi, \widehat{\pi})$	ISE	$W_1(\pi, \widehat{\pi})$	ISE	$W_1(\pi, \widehat{\pi})$	ISE
GMM-tri	0.213	0.058	0.273	0.071	0.908	0.031
BBM	0.033	0.059	0.035	0.055	0.109	0.320

C. Convergence Analysis

To empirically check the convergence of the proposed two-stage algorithm, we plot the GB-NPMLE loss vs. epoch number for training the generator G in Stage I. For Stage II, we plot the log-likelihood vs. MCEM iteration number for learning the multinomial weights $\boldsymbol{\tau}$. Fig. 2 plots these solution paths for all 40 replications of the Gaussian and Gamma mixture models (i.e. Experiments (i) and (ii) in Section III-A of the main article), denoted as GMM and GaMM respectively.

Fig. 2 shows that Stage I converges quickly, typically within 50 epochs. Meanwhile, Stage II converges even faster – typically within one or two iterations, resulting in a very flat solution path. The convergence plots for our other numerical experiments were very similar to those in Fig. 2. Although empirical evidence suggests that the GB-NPMLE two-stage algorithm converges quickly, a rigorous theoretical analysis of convergence rate should be done for future work.

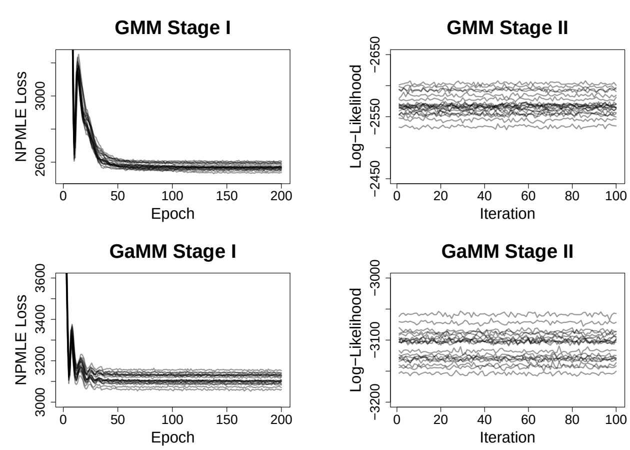

Fig. 2. Convergence plots for Stage I and Stage II of the two-stage algorithm under the GMM and GaMM models (Section III-A of the main article).