SUPPLEMENTARY MATERIAL

# Supplementary material for "EM-test for homogeneity in a two-sample problem with a mixture structure"

Guanfu Liu[a], Yuejiao Fu[b], Jianjun Zhang[c], Xiaolong Pu[d] and Boying Wang[d]

[a]School of Statistics and Information, Shanghai University of International Business and Economics, Shanghai 201620, China; [b]Department of Mathematics and Statistics, York University, Toronto M3J 1P3, Canada; [c]Department of Mathematics, University of North Texas, Denton, TX 76201, USA; [d]School of Statistics, East China Normal University, Shanghai 200241, China

**ARTICLE HISTORY**
Compiled June 20, 2019

**ABSTRACT**
This is a supplementary material to the corresponding paper submitted to the *Journal of Applied Statistics*. It contains the regularity conditions, the proofs of Theorems 2.1–2.2 and Theorem 2.5, and R codes for sample size calculations.

## 1. Regularity conditions

For smooth reading, we list the regularity conditions of the main paper in the following. In the proofs, the expectation are taken under the distribution $f(x; \mu_0, \sigma_0)$.

- B0. (IID condition) In the model (1), the first sample $x_{11}, \ldots, x_{n_1}$ are independent and identically distributed (i.i.d.), and the second sample $x_{21}, \ldots, x_{n_2}$ are i.i.d., and the two samples are independent.
- B1. (Wald's integrability conditions) (i) $E\{|\log f(x; 0, 1)|\} < \infty$; (ii) $\lim_{x\to\infty} f(x; 0, 1) = 0$.
- B2. (Smoothness) The support of $f(x; \mu, \sigma)$ is $(-\infty, \infty)$, and it is three times continuously differentiable with respect to $\mu$ and $\sigma$.
- B3. (Identifiability) For any two mixing distribution functions $\Psi_1$ and $\Psi_2$ with two supporting points such that $\int f(x; \mu, \sigma)\, d\Psi_1(\mu, \sigma) = \int f(x; \mu, \sigma)\, d\Psi_2(\mu, \sigma)$ for all $x$, we must have $\Psi_1 = \Psi_2$.
- B4. (Uniform boundedness) There exists a function $g$ with finite expectation such that

$$\left|\frac{\partial^{(h+l)} f(x; \mu_0, \sigma_0)/\partial\mu^h \partial\sigma^l}{f(x; \mu_0, \sigma_0)}\right|^3 \le g(x), \text{ for } h + l \le 2,$$

where $h$ and $l$ are two nonnegative integers. Moreover, there exists a positive $\epsilon$

CONTANT Author: Jianjun Zhang. Email: Jianjun.Zhang@unt.edu

such that

$$\sup_{|\mu-\mu_0|^2+|\sigma-\sigma_0|^2\le\epsilon}\left|\frac{\partial^{(h+l)}f(x;\mu,\sigma)/\partial\mu^h\partial\sigma^l}{f(x;\mu_0,\sigma_0)}\right|^3 \le g(x), \text{ for } h+l=3.$$

B5. (Positive definiteness) The covariance matrix of $(U,V)$ is positive definite, where

$$U=\frac{\partial f(x_{11};\mu_0,\sigma_0)/\partial\mu}{f(x_{11};\mu_0,\sigma_0)} \text{ and } V=\frac{\partial f(x_{11};\mu_0,\sigma_0)/\partial\sigma}{f(x_{11};\mu_0,\sigma_0)}.$$

B6. (Tail condition) There exists positive constants $v_0$, $v_1$ and $\beta_0$ with $\beta_0>1$ such that $f(x;0,1)\le\min\{v_0,v_1|x|^{-\beta_0}\}$.

B7. (Upper bound function) There exist a nonnegative function $s(x;\mu,\sigma)$ which satisfies Condition B1 and is continuous in $(\mu,\sigma)$, a positive number $a$ with $1/\beta_0<a<1$, a positive number $b$, and a positive number $\epsilon^*$ with $\epsilon^*<1$ such that for $\sigma\in(0,\epsilon^*\sigma_0)$, $s(x;\mu,\sigma)$ is uniformly bounded, $\int s(x;\mu,\sigma)dx<1$, and

$$f(x;\mu,\sigma)\le\begin{cases}\sigma^{-1}s(x;\mu,\sigma), & \text{if } |x-\mu|\le\sigma^{1-a}\\ \sigma^{b}s(x;\mu,\sigma), & \text{if } |x-\mu|>\sigma^{1-a}\end{cases}.$$

C1. $p(\lambda)$ is a continuous function such that it is maximized at $\lambda=1$ and goes to negative infinity as $\lambda\to0$.

C2. $\sup_{\sigma>0}\max\{p_n(\sigma),0\}=o(n)$ and $p_n(\sigma)=o(n)$ for any $\sigma$.

C3. $p_n'(\sigma)=o_p(n^{1/2})$, for all $\sigma>0$, where $p_n'(\sigma)$ is the derivative function with respect to $\sigma$.

C4. When $0<\sigma\le 8/(n_2M_0)$ and $n_2$ is large, $p_n(\sigma)\le 4(\log n_2)^2\log(\sigma)$. Here $M_0=\max\{\sup_x f(x;\mu_0,\sigma_0),8\}$.

C5. $p_n(b_1\sigma;b_1X_1+b_0,...,b_1X_n+b_0)=p_n(\sigma;X_1,...,X_n)$.

## 2. Some useful lemmas

Since the EM-test is location-scale invariant, we assume $(\mu_0,\sigma_0)=(0,1)$ for the convenience of presentation.

**Lemma 2.1.** *(Consistency with non-zero mixing proportion) Assume the same conditions as in Theorem 2.1. Let $(\bar\lambda,\bar\mu,\bar\sigma_1,\bar\sigma_2)$ be any estimator of $(\lambda,\mu,\sigma_1,\sigma_2)$. If*

$$pl_n(\bar\lambda,\bar\mu,\bar\sigma_1,\bar\sigma_2)-pl_n(1,0,1,1)>c>-\infty$$

*hold and $\bar\lambda\in[\delta,1]$ for some $\delta\in(0,1)$, then under the null model $f(x;0,1)$, we have $\bar\mu=o_p(1)$ and $\bar\sigma_1-1=o_p(1),\bar\sigma_2-1=o_p(1)$.*

The proof of Lemma 2.1 is similar to the proof of Lemma 1 in [4], hence we omit it. In the next lemma, the conclusion of Lemma 2.1 is strengthened.

**Lemma 2.2.** *(Convergence rate with non-zero mixing proportion) Assume the same conditions as in Lemma 1.1. If $\bar\lambda-\lambda_0=o_p(1)$ for $\lambda_0\in(0,1]$, then $\bar\mu=O_p(n^{-1/2})$, $\bar\sigma_h-1=O_p(n^{-1/2})$, $h=1,2$.*

***Proof.*** Let

$$l_{n1}(\mu, \sigma_1) = \sum_{i=1}^{n_1} \log f(x_{1i}; \mu, \sigma_1),$$
$$l_{n2}(\lambda, \mu, \sigma_1, \sigma_2) = \sum_{i=1}^{n_2} \log \{(1-\lambda) f(x_{2i}; \mu, \sigma_1) + \lambda f(x_{2i}; \mu, \sigma_2)\}.$$

Further let

$R_{n1}(\mu, \sigma_1) = l_{n1}(\mu, \sigma_1) - l_{n1}(0, 1)$ and $R_{n2}(\lambda, \mu, \sigma_1, \sigma_2) = l_{n2}(\lambda, \mu, \sigma_1, \sigma_2) - l_{n2}(1, 0, 1, 1)$.

Then

$$\begin{aligned} pl_n(\bar\lambda, \bar\mu, \bar\sigma_1, \bar\sigma_2) - pl_n(1, 0, 1, 1) = {} & R_{n1}(\mu, \sigma_1) + R_{n2}(\lambda, \mu, \sigma_1, \sigma_2) \\ & + p(\bar\lambda) - p(1) + p_n(\bar\sigma_2) - p_n(1). \end{aligned}$$

Next we derive an upper bound for $pl_n(\bar\lambda, \bar\mu, \bar\sigma_1, \bar\sigma_2) - pl_n(1, 0, 1, 1)$. Together with the lower bound $c$, we get the order assessment of $\bar\mu$ and $\bar\sigma_h (h = 1, 2)$.

We first find an approximation for $R_{n1}(\bar\mu, \bar\sigma_1)$. From Lemma 2.1, we have the consistency results $\bar\mu = o_p(1)$, $\bar\sigma_h - 1 = o_p(1)$, $h = 1, 2$. Applying the second Taylor expansion to $R_{n1}(\bar\mu, \bar\sigma_1)$ around $(0, 1)$, and the weak law of large numbers with Conditions B2 and B4, we get that

$$\begin{aligned} R_{n1}(\bar\mu, \bar\sigma_1) &= l_{n1}(\bar\mu, \bar\sigma_1) - l_{n1}(0, 1) \\ &= \frac{\partial l_{n1}(0,1)}{\partial \mu}\bar\mu + \frac{\partial l_{n1}(0,1)}{\partial \sigma_1}(\bar\sigma_1 - 1) \\ &\quad + \frac{1}{2}\frac{\partial^2 l_{n1}(0,1)}{\partial^2 \mu}\bar\mu^2 + \frac{\partial^2 l_{n1}(0,1)}{\partial \mu \partial \sigma_1}\bar\mu(\bar\sigma_1 - 1) + \frac{1}{2}\frac{\partial^2 l_{n1}(0,1)}{\partial^2 \sigma_1}(\bar\sigma_1 - 1)^2 \\ &\quad + o_p(n)\{\bar\mu^2 + (\bar\sigma_1 - 1)^2\}. \end{aligned} \tag{1}$$

Let

$$Y_{hi} = \frac{\partial f(x_{hi}; 0, 1)/\partial \mu}{f(x_{hi}; 0, 1)} \quad \text{and} \quad Z_{hi} = \frac{\partial f(x_{hi}; 0, 1)/\partial \sigma}{f(x_{hi}; 0, 1)}, \ h = 1, 2, \ i = 1, 2, \ldots, n_h.$$

Hence,

$$\frac{\partial l_{n1}(0,1)}{\partial \mu}\bar\mu + \frac{\partial l_{n1}(0,1)}{\partial \sigma_1}(\bar\sigma_1 - 1) = \sum_{i=1}^{n_1}\{\bar\mu Y_{1i} + (\bar\sigma_1 - 1) Z_{1i}\}. \tag{2}$$

Further by the weak law of large numbers, we get

$$\begin{aligned}\frac{\partial^2 l_{n1}(0,1)}{\partial^2\mu} &= \sum_{i=1}^{n_1}\left[\frac{\partial^2 f(x_{hi};0,1)/\partial^2\mu}{f(x_{hi};0,1)} - \left\{\frac{\partial f(x_{hi};0,1)/\partial\mu}{f(x_{hi};0,1)}\right\}^2\right] \\ &= -\sum_{i=1}^{n_1} Y_{1i}^2 + o_p(n). \end{aligned} \tag{3}$$

Similarly to (3), we have that

$$\frac{\partial^2 l_{n1}(0,1)}{\partial\mu\partial\sigma_1} = -\sum_{i=1}^{n_1} Y_{1i}Z_{1i} + o_p(n) \text{ and } \frac{\partial^2 l_{n1}(0,1)}{\partial^2\sigma_1} = -\sum_{i=1}^{n_1} Z_{1i}^2 + o_p(n). \tag{4}$$

By Condition B5 and the weak law of large numbers, we get that

$$\left[\sum_{i=1}^{n_1}\{\bar{\mu}Y_{1i} + (\bar{\sigma}_1 - 1)Z_{1i}\}^2\right] o_p(1) = o_p(n)\{\bar{\mu}^2 + (\bar{\sigma}_1 - 1)^2\}. \tag{5}$$

Combining (1)–(5), we get

$$R_{n1}(\bar{\mu},\bar{\sigma}_1) = \sum_{i=1}^{n_1}\{\bar{\mu}Y_{1i} + (\bar{\sigma}_1 - 1)Z_{1i}\} - \frac{1}{2}\left[\sum_{i=1}^{n_1}\{\bar{\mu}Y_{1i} + (\bar{\sigma}_1 - 1)Z_{1i}\}^2\right]\{1 + o_p(1)\}. \tag{6}$$

By Condition C3, we have

$$|p_n(\bar{\sigma}_2) - p_n(1)| = |o_p(n^{1/2})(\bar{\sigma}_2 - 1)| \le |o_p(1)| + |o_p(n)|(\bar{\sigma}_2 - 1)^2. \tag{7}$$

We now find an upper bound for $R_{n2}(\lambda,\mu,\sigma_1,\sigma_2)$. Write $R_{n2}(\lambda,\mu,\sigma_1,\sigma_2) = \sum_{i=1}^{n_2}\log(1+\delta_i)$, where

$$\delta_i = \frac{(1-\bar{\lambda})\{f(x_{2i};\bar{\mu},\bar{\sigma}_1) - f(x_{2i};0,1)\} + \bar{\lambda}\{f(x_{2i};\bar{\mu},\bar{\sigma}_2) - f(x_{2i};0,1)\}}{f(x_{2i};0,1)}.$$

By the inequality $\log(1+x) \le x - x^2/2 + x^3/3$, we have

$$R_{n2}(\bar{\lambda},\bar{\mu},\bar{\sigma}_1,\bar{\sigma}_2) \le \sum_{i=1}^{n_2}\delta_i - \frac{1}{2}\sum_{i=1}^{n_2}\delta_i^2 + \frac{1}{3}\sum_{i=1}^{n_2}\delta_i^3.$$

Let $\bar{m} = (1-\bar{\lambda})(\bar{\sigma}_1 - 1) + \bar{\lambda}(\bar{\sigma}_2 - 1)$. By the consistency $\bar{\sigma}_h - 1 = o_p(1)$, $h = 1,2$, we get $\bar{m} = o_p(1)$. Applying the first order Taylor expansion to $f(x_{2i};\bar{\mu},\bar{\sigma}_h)$ around $(0,1)$, we have $\delta_i = \bar{\mu}Y_{2i} + \bar{m}Z_{2i} + \varepsilon_{ni}$. Let $\varepsilon_n = \sum_{i=1}^{n_2}\varepsilon_{ni}$. The remainder term $\varepsilon_n$ satisfies

$$\varepsilon_n = O_p(n_2^{1/2})\{\bar{\mu}^2 + \sum_{h=1}^{2}(\bar{\sigma}_h - 1)^2\} = O_p(n^{1/2})\{\bar{\mu}^2 + \sum_{h=1}^{2}(\bar{\sigma}_h - 1)^2\}.$$

From Condition B4 and the weak law of large numbers, after some straightforward algebra, we get

$$\begin{aligned} R_{n2}(\bar{\lambda}, \bar{\mu}, \bar{\sigma}_1, \bar{\sigma}_2) \leq & \sum_{i=1}^{n_2} \{\bar{\mu} Y_{2i} + \bar{m} Z_{2i}\} - \frac{1}{2} \sum_{i=1}^{n_2} \{\bar{\mu} Y_{2i} + \bar{m} Z_{2i}\}^2 \{1 + o_p(1)\} \\ & + \frac{1}{3} \sum_{i=1}^{n_2} \{\bar{\mu} Y_{2i} + \bar{m} Z_{2i}\}^3 + O_p(\varepsilon_n) \end{aligned}$$

For the cubic term in the upper bound of $R_{n2}(\bar{\lambda}, \bar{\mu}, \bar{\sigma}_1, \bar{\sigma}_2)$, we have that

$$\begin{aligned} \sum_{i=1}^{n_2} \{\bar{\mu} Y_{2i} + \bar{m} Z_{2i}\}^3 &= \sum_{i=1}^{n_2} \{\bar{\mu}^3 Y_{2i}^3 + 3\bar{\mu}^2 \bar{m} Y_{2i}^2 Z_{2i} + 3\bar{\mu}\bar{m}^2 Y_{2i} Z_{2i}^2 + \bar{m}^3 Z_{2i}^3\} \\ &= o_p(1) \left[ \sum_{i=1}^{n_2} \{\bar{\mu}^2 Y_{2i}^3 + 3\bar{\mu}^2 Y_{2i}^2 Z_{2i} + 3\bar{m}^2 Y_{2i} Z_{2i}^2 + \bar{m}^2 Z_{2i}^3\} \right] \\ &= o_p(n)(\bar{\mu}^2 + \bar{m}^2). \end{aligned}$$

Hence

$$R_{n2}(\bar{\lambda}, \bar{\mu}, \bar{\sigma}_1, \bar{\sigma}_2) \leq \sum_{i=1}^{n_2} \{\bar{\mu} Y_{2i} + \bar{m} Z_{2i}\} - \frac{1}{2} \sum_{i=1}^{n_2} \{\bar{\mu} Y_{2i} + \bar{m} Z_{2i}\}^2 \{1 + o_p(1)\} + O_p(\varepsilon_n). \quad (8)$$

Combining (6)–(8) and Condition C1, we get

$$\begin{aligned} pl_n(\bar{\lambda}, \bar{\mu}, \bar{\sigma}_1, \bar{\sigma}_2) - pl_n(1, 0, 1, 1) \leq & \sum_{i=1}^{n_1} \{\bar{\mu} Y_{1i} + (\bar{\sigma}_1 - 1) Z_{1i}\} + \sum_{i=1}^{n_2} (\bar{\mu} Y_{2i} + \bar{m} Z_{2i}) \\ & - \frac{1}{2} \left[ \sum_{i=1}^{n_1} \{\bar{\mu} Y_{1i} + (\bar{\sigma}_1 - 1) Z_{1i}\}^2 + \sum_{i=1}^{n_2} (\bar{\mu} Y_{2i} + \bar{m} Z_{2i})^2 \right] \{1 + o_p(1)\} \\ & + O_p(\varepsilon_n) + o_p(1). \end{aligned}$$

Condition $\bar{\lambda} - \lambda_0 = o_p(1)$ with $\lambda_0 \in (0, 1]$ implies that

$$O_p(\varepsilon_n) = O_p(n^{1/2})\{\bar{\mu}^2 + (\bar{\sigma}_1 - 1)^2 + \bar{m}^2\} = o_p(n)\{\bar{\mu}^2 + (\bar{\sigma}_1 - 1)^2 + \bar{m}^2\}.$$

Hence, by the weak law of large numbers with Condition B5,

$$
\begin{aligned}
c &\le pl_n(\bar\lambda,\bar\mu,\bar\sigma_1,\bar\sigma_2) - pl_n(1,0,1,1)\\
&\le \sum_{i=1}^{n_1}\{\bar\mu Y_{1i} + (\bar\sigma_1 - 1)Z_{1i}\} + \sum_{i=1}^{n_2}(\bar\mu Y_{2i} + \bar m Z_{2i})\\
&\quad - \frac{1}{2}\left[\sum_{i=1}^{n_1}\{\bar\mu Y_{1i} + (\bar\sigma_1 - 1)Z_{1i}\}^2 + \sum_{i=1}^{n_2}(\bar\mu Y_{2i} + \bar m Z_{2i})^2\right][1+o_p(1)] + o_p(1)\\
&= \bar\mu\left\{\sum_{i=1}^{n_1} Y_{1i} + \sum_{i=1}^{n_2} Y_{2i}\right\} + (\bar\sigma_1 - 1)\sum_{i=1}^{n_1} Z_{1i} + \bar m \sum_{i=1}^{n_2} Z_{2i}\\
&\quad - \frac{1}{2}\left[\bar\mu^2\left\{\sum_{i=1}^{n_1} Y_{1i}^2 + \sum_{i=1}^{n_2} Y_{2i}^2\right\} + (\bar\sigma_1 - 1)^2\sum_{i=1}^{n_1} Z_{1i}^2 + \bar m^2 \sum_{i=1}^{n_2} Z_{2i}\right.\\
&\quad \left. + 2\bar\mu(\bar\sigma_1 - 1)\sum_{i=1}^{n_1} Y_{1i}Z_{1i} + 2\bar\mu\bar m\sum_{i=1}^{n_2} Y_{2i}Z_{2i}\right]\{1+o_p(1)\} + o_p(1)\\
&\le \frac{1}{2}\boldsymbol{U_n'}\mathbf{W^{-1}}\boldsymbol{U_n} + o_p(1).
\end{aligned}
\tag{9}
$$

where

$$
\boldsymbol{U_n} = n^{-1/2}\begin{pmatrix} \sum_{i=1}^{n_1} Y_{1i} + \sum_{i=1}^{n_2} Y_{2i} \\ \sum_{i=1}^{n_1} Z_{1i} \\ \sum_{i=1}^{n_2} Z_{2i} \end{pmatrix} \quad \text{and} \quad \mathbf{W} = \begin{pmatrix} \sigma_Y^2 & \rho_1\sigma_{Y,Z} & \rho_2\sigma_{Y,Z} \\ \rho_1\sigma_{Y,Z} & \rho_1\sigma_Z^2 & 0 \\ \rho_2\sigma_{Y,Z} & 0 & \rho_2\sigma_Z^2 \end{pmatrix},
$$

with $\sigma_Y^2 = \mathrm{Var}(Y_{11})$, $\sigma_Z^2 = \mathrm{Var}(Z_{11})$, $\sigma_{Y,Z} = \mathrm{Cov}(Y_{11}, Z_{11})$. Therefore

$$
\bar\mu = O_p(n^{-1/2}),\ \bar\sigma_1 - 1 = O_p(n^{-1/2}),\ \bar m = O_p(n^{-1/2}). \tag{10}
$$

Any values of $(\bar\mu, \bar\sigma - 1, \bar m)$ out of this range in (10) will violate the inequality. With the condition that $\bar\lambda - \lambda_0 = o_p(1)$, for some $\lambda_0 \in (0, 1]$, we have

$$
\bar\mu = O_p(n^{-1/2}),\ \bar\sigma_1 - 1 = O_p(n^{-1/2}),\ \bar\sigma_2 - 1 = O_p(n^{-1/2}).
$$

□

Let $(\bar\lambda,\bar\mu,\bar\sigma_1,\bar\sigma_2)$ be estimator of $(\lambda,\mu,\sigma_1,\sigma_2)$ as before, and let

$$
\bar w_i = \frac{\bar\lambda f(x_{2i};\bar\mu,\bar\sigma_2)}{(1-\bar\lambda)f(x_{2i};\bar\mu,\bar\sigma_1) + \bar\lambda f(x_{2i};\bar\mu,\bar\sigma_2)}.
$$

Define

$$
H_n(\lambda) = (n_2 - \sum_{i=1}^{n_2}\bar w_i)\log(1-\lambda) + \sum_{i=1}^{n_2}\bar w_i\log\lambda + p(\lambda).
$$

The EM-test updates $\lambda$ by search $\bar{\lambda}^* = \arg\max_\lambda H_n(\lambda)$. Then, we have the following lemma.

**Lemma 2.3.** *Under the same conditions as in Lemma 2.2, and if* $\bar{\lambda} - \lambda_0 = o_p(1)$ *for some* $\lambda_0 \in (0, 1]$*, then* $\bar{\lambda}^* - \bar{\lambda}_0 = o_p(1)$.

For the proof of Lemma 2.3 is similar to that of Lemma A3 of [3] and hence is omitted.

## 3. The proofs of Theorems 2.1–2.2 and Theorem 2.5

*The proof of Theorem 2.1*

***Proof.*** For any $k \le K$, due to the monotonicity property of the EM algorithm that the penalized likelihood increases after each iteration ([1, 8, 5]), we have

$$pl_n(\lambda_j^{(k)}, \mu_j^{(k)}, \sigma_{1j}^{(k)}, \sigma_{2j}^{(k)}) \ge pl_n(\lambda_j^{(1)}, \mu_j^{(1)}, \sigma_{1j}^{(1)}, \sigma_{2j}^{(1)}) \ge pl_n(\lambda_j, 0, 1, 1).$$

Further, $pl_n(\lambda_j^{(k)}, \mu_j^{(k)}, \sigma_{1j}^{(k)}, \sigma_{2j}^{(k)}) - pl_n(1, 0, 1, 1) \ge p(\lambda_j) - p(1) > -\infty$. Then by Lemmas 2.1–2.3, Theorem 2.1 holds. □

*The proof of Theorem 2.2*

***Proof.*** Under Conditions B2, B4 and B5, applying some of the classic results about regular models ([6]), we have

$$\sup_{\mu,\sigma} pl_n(1, \mu, \sigma, \sigma) - pl_n(1, 0, 1, 1) = \frac{1}{2}\boldsymbol{U_n^{*\prime}}\mathbf{W}^{*-\mathbf{1}}\boldsymbol{U_n^*} + o_p(1), \tag{11}$$

where

$$\boldsymbol{U_n^*} = n^{-1/2}\begin{pmatrix} \sum_{i=1}^{n_1} Y_{1i} + \sum_{i=1}^{n_2} Y_{2i} \\ \sum_{i=1}^{n_1} Z_{1i} + \sum_{i=1}^{n_2} Z_{2i} \end{pmatrix}, \qquad \mathbf{W}^* = \begin{pmatrix} \sigma_Y^2 & \sigma_{Y,Z} \\ \sigma_{Y,Z} & \sigma_Z^2 \end{pmatrix}.$$

From Theorem 2.1 and (9), we have

$$pl_n(\lambda_j^{(K)}, \mu_j^{(K)}, \sigma_{1j}^{(K)}, \sigma_{2j}^{(K)}) - pl_n(1, 0, 1, 1) \le \frac{1}{2}\boldsymbol{U_n'}\mathbf{W^{-1}}\boldsymbol{U_n} + o_p(1) \tag{12}$$

Hence, combining (11) and (12), we get

$$\begin{aligned} M_n^{(K)}(\lambda_j) &= 2\{pl_n(\lambda_j^{(K)}, \mu_j^{(K)}, \sigma_{1j}^{(K)}, \sigma_{2j}^{(K)}) - pl_n(1, 0, 1, 1)\} \\ &\quad - 2\{\sup_{\mu,\sigma} pl_n(1, \mu, \sigma, \sigma) - pl_n(1, 0, 1, 1)\} \\ &\le \boldsymbol{U_n'}\mathbf{W^{-1}}\boldsymbol{U_n} - \boldsymbol{U_n^{*\prime}}\mathbf{W}^{*-\mathbf{1}}\boldsymbol{U_n^*} + o_p(1), \end{aligned}$$

where the presentations of $\mathbf{W^{-1}}$ and $\mathbf{W^{*-1}}$ are provided in the end of the proof. The above equality can further be simplified as

$$M_n^{(K)}(\lambda_j) \leq \sigma_Z^{-2} T_n^2 + o_p(1),$$

where $T_n = n^{-1/2}\left(\sqrt{\frac{\rho_2}{\rho_1}}\sum_{i=1}^{n_1} Z_{1i} - \sqrt{\frac{\rho_1}{\rho_2}}\sum_{i=1}^{n_2} Z_{2i}\right)$. Since the upper bound does not depend on $\lambda_j$, we further have

$$EM_n^{(K)} \leq \sigma_Z^{-2} T_n^2 + o_p(1). \tag{13}$$

Next, we show the upper bound for $EM_n^{(K)}$. Let $(\tilde{\mu}, \tilde{\sigma}_1 - 1, \tilde{\sigma}_2 - 1) = n^{-1/2}\mathbf{W^{-1}}\boldsymbol{U_n}$. Since the EM-iteration always increases the penalized likelihood and $\lambda_1 = 1$, we have that

$$EM_n^{(K)} \geq M_n^{(K)}(\lambda_1) \geq M_n^{(1)}(\lambda_1) \geq 2\{pl_n(\lambda_1, \tilde{\mu}, \tilde{\sigma}_1, \tilde{\sigma}_2) - \sup_{\mu,\sigma} pl_n(1, \mu, \sigma, \sigma)\}. \tag{14}$$

Note that it is easy to verify that $\tilde{\mu} = O_p(n^{-1/2}), \tilde{\sigma}_1 - 1 = O_p(n^{-1/2}), \tilde{\sigma}_2 - 1 = O_p(n^{-1/2})$. With this order assessment and applying the second order Taylor expansion, we have that

$$2\{pl_n(\lambda_1, \tilde{\mu}, \tilde{\sigma}_1, \tilde{\sigma}_2) - \sup_{\mu,\sigma} pl_n(1, \mu, \sigma, \sigma)\} = \sigma_Z^{-2} T_n^2 + o_p(1). \tag{15}$$

From (13)–(15), we get $EM_n^{(K)} = \sigma_Z^{-2} T_n^2 + o_p(1)$. By central limit theorem, $\sigma_Z^{-1} T_n$ converges to $N(0,1)$ in distribution. Therefore, Consequently, the null limiting distribution of $EM_n^{(K)}$ is $\chi_1^2$.

The presentations of $\mathbf{W^{-1}}$ and $\mathbf{W^{*-1}}$ are as follows.

$$\mathbf{W^{-1}} = \frac{1}{\sigma_Y^2\sigma_Z^2 - \sigma_{Y,Z}^2}\begin{pmatrix} \sigma_Z^2 & -\sigma_{Y,Z} & -\sigma_{Y,Z} \\ -\sigma_{Y,Z} & \frac{\sigma_Y^2}{\rho_1} - \frac{\rho_2\sigma_{Y,Z}^2}{\rho_1\sigma_Z^2} & \frac{\sigma_{Y,Z}^2}{\sigma_Z^2} \\ -\sigma_{Y,Z} & \frac{\sigma_{Y,Z}^2}{\sigma_Z^2} & \frac{\sigma_Y^2}{\rho_2} - \frac{\rho_1\sigma_{Y,Z}^2}{\rho_2\sigma_Z^2} \end{pmatrix},$$
$$\mathbf{W^{*-1}} = \frac{1}{\sigma_Y^2\sigma_Z^2 - \sigma_{Y,Z}^2}\begin{pmatrix} \sigma_Z^2 & -\sigma_{Y,Z} \\ -\sigma_{Y,Z} & \sigma_Y^2 \end{pmatrix}.$$

□

*The proof of Theorem 2.5*

***Proof.*** *The proof of (i)*

Without loss of generality, we assume that the null model is $f(x; 0, 1)$. Then $E(V^2) = \sigma_0^{-2}\sigma_Z^2$. Further the local alternative $H_{a1}^n$ in (6) of the main paper becomes

$$H_{a1}^n : \lambda = \lambda_0,\ (\mu_1, \sigma_1) = (0, 1),\ (\mu_2, \sigma_2) = (0, 1 + n_2^{-1/2}\Delta_1/\sigma_0).$$

Let

$$\Lambda_n = \sum_{i=1}^{n_2} \log \frac{(1-\lambda_0) f(x_{2i}; 0, 1) + \lambda_0 f(x_{2i}; 0, 1 + n_2^{-1/2}\Delta_1/\sigma_0)}{f(x_{2i}; 0, 1)}.$$

Under Conditions B2 and B4, applying second order approximation, we can verify that under the null model,

$$\Lambda_n = \lambda_0 \sigma_0^{-1} n_2^{-1/2} \sum_{i=1}^{n_2} \Delta_1 Z_{2i} - 0.5\lambda_0^2 \sigma_0^{-2} \sigma_Z^2 \Delta_1^2 + o_p(1).$$

Hence under the null model, $\Lambda_n \xrightarrow{d} N(-0.5c^2, c^2)$, where $c^2 = \lambda_0^2 \sigma_0^{-2} \Delta_1^2 \sigma_Z^2$. Therefore, the local alternative $H_{a1}^n$ is contiguous to the null distribution ([2] and Example 6.5 of [7]).

By Le Cam's contiguity theory, the limiting distribution of $\widetilde{EM}_n^{(K)}$ under $H_{a1}^n$ is determined by the joint limiting distribution of $\sigma_Z^{-1} T_n$ and $\Lambda_n$ under the null model. By central limit theorem and Slutsky's theorem, the joint limiting distribution of $\sigma_Z^{-1} T_n$ and $\Lambda_n$ under the null model is multivariate normal

$$\mathcal{N}_2 \left( \begin{pmatrix} 0 \\ -0.5c^2 \end{pmatrix}, \begin{pmatrix} 1 & -\sqrt{\rho_1}\lambda_0 \sigma_0^{-1} \sigma_Z \Delta_1 \\ -\sqrt{\rho_1}\lambda_0 \sigma_0^{-1} \sigma_Z \Delta_1 & c^2 \end{pmatrix} \right).$$

By Le Cam's third lemma ([7]), we have under $H_{a1}^n$,

$$\sigma_Z^{-1} T_n \xrightarrow{d} N(-\sqrt{\rho_1}\lambda_0 \sigma_0^{-1} \sigma_Z \Delta_1, 1).$$

Since $EM_n^{(K)} = \sigma_Z^{-1} T_n + o_p(1)$ holds under the null, by Le Cam's first lemma ([7]), $EM_n^{(K)} = \sigma_Z^{-1} T_n + o_p(1)$ still holds under $H_{a1}^n$. Therefore, the limiting distribution of $EM_n^{(K)}$ under the local alternative $H_{a1}^n$ is $\chi_1^2(c_1^2)$, where $c_1^2 = \lambda_0^2 \rho_1 \sigma_0^{-2} \Delta_1^2 \sigma_Z^2 = \lambda_0^2 \rho_1 \Delta_1^2 \mathrm{E}(V^2)$.

*The proof of (ii)*

The proof for part (ii) is similar to that of (i), hence we omit it. □

## 4. R codes for sample size calculation

Given the null model $H_0$, the local alternative model $H_{a1}^n$ and given $\rho_1$, the following R functions *size.norm()* and *size.logis()* calculate the required sample sizes $(n_1, n_2)$ to reject the null hypothesis with the target power $1-\beta$ at the significance level $\alpha$ for the normal kernel and logistic kernel, respectively.

For example, suppose $\lambda_0 = 0.5$, $(\mu_1, \sigma_1) = (0, 1)$, $(\mu_2, \sigma_2) = (0, 1.5)$, and $\rho_1 = 1/3$. If the target power is 80% at the 5% significance level, the required sample sizes are found to be $(n_1, n_2) = (94, 189)$ under the normal kernel and $(n_1, n_2) = (129, 258)$ under the logistic kernel by using R functions *size.norm()* and *size.logis()*.

```
size.norm <- function(lambda0,rho1,sigma1,sigma2,alpha,target_power){
  n2 <- 2
  power0 <- target_power
```

```
  diff_power <- 1
  while(diff_power>0.001){
    Delta1 <- sqrt(n2)*(sigma2-sigma1)
    c0_squ <- lambda0^2*rho1/(sigma1^2)*(2*Delta1^2)
    power1 <- pchisq(qchisq(1-alpha,1),1,ncp = c0_squ,lower.tail = F)
    diff_power <- power0-power1
    n2 <- n2+1
  }
  n1 <- round(rho1/(1-rho1)*n2,0)
  data.frame(n1=n1,n2=n2,row.names = "sample size")
}


size.logis <- function(lambda0,rho1,sigma1,sigma2,alpha,target_power){
  n2 <- 2
  power0 <- target_power
  diff_power <- 1
  while(diff_power>0.001){
    Delta1 <- sqrt(n2)*(sigma2-sigma1)
    c0_squ <- lambda0^2*rho1/(sigma1^2)*(Delta1^2*(3+pi^2)/9)
    power1 <- pchisq(qchisq(1-alpha,1),1,ncp = c0_squ,lower.tail = F)
    diff_power <- power0-power1
    n2 <- n2+1
  }
  n1 <- round(rho1/(1-rho1)*n2,0)
  data.frame(n1=n1,n2=n2,row.names = "sample size")
}

> size.norm(0.5,1/3,1,1.5,0.05,0.8)
            n1  n2
sample size 94 189

> size.logis(0.5,1/3,1,1.5,0.05,0.8)
             n1  n2
sample size 132 264
```

## References


[1] A.P. Dempster, N.M. Laird and D.B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm*, J. Roy. Statist. Soc. Ser. B 39 (1977), pp. 1–38.

[2] L.M. Le Cam and G.L. Yang, *Asymptotics in Statistics : Some Basic Concepts*, Springer-Verlag, New York, 1990.

[3] P. Li, J. Chen and P. Marriott, *Non-finite Fisher information and homogeneity: an EM approach*, Biometrika 96 (2009), pp.411–426.

[4] G. Liu, Y. Fu, P. Li and X. Pu, *Using differential variability to increase the power of the homogeneity test in a two-sample problem*, Statist. Sinica 28 (2018), pp. 27–41.

[5] G.J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*,Wiley, New York, 1997.

[6] R.J. Serfling, *Approximation Theorem of Mathematical Statistics*, Wiley, New York, 1980.

[7] A.W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press, New York, 2000.

[8] C.F.J. Wu, *On the Convergence Properties of the EM Algorithm*, Ann. Statist. 11 (1983), pp. 95–103.