Complete proof of the identifiability as a supplementary to

Mixture of Shifted Binomial Distributions for Rating Data

Shaoting Li and Jiahua Chen

Abstract

This supplementary material contains full proof of the identifiability of the SBM, while the paper proves the identifiability in representative cases.

To ensure the readability of the supplementary file, we include some materials in the paper so that the readers need not go back to find the meaning of some notation and facts. The probability mass function of the proposed SBM model is given by

$$\begin{aligned}\mathrm{SBM}(r; m, G) &= \pi \binom{m-1}{m-r} \theta_1^{m-r}(1-\theta_1)^{r-1} + (1-\pi)\binom{m-s}{m-r}\theta_2^{m-r}(1-\theta_2)^{r-s} \\ &= \pi\, B_1(r; m, \theta_1) + (1-\pi)\, B_s(r; m, \theta_2) \qquad (1)\end{aligned}$$

for $r = 1, 2, \cdots, m$.

Identifiability Theorem: Suppose $G, G' \in \mathbb{G}$ and $m \geq 5$. Then, $\mathrm{SBM}(m, G) = \mathrm{SBM}(m, G')$ if and only if $G = G'$.

We write two mixing distributions as

$$G = \pi\{(1, \theta_1)\} + (1-\pi)\{(s, \theta_2)\}; \quad G' = \alpha\{(1, \eta_1)\} + (1-\alpha)\{(t, \eta_2)\}.$$

For a real x and positive integer k, we define its kth factorial $x^{(k)} = x(x-1)\cdots(x-k+1)$. Further, we let $x^{(0)} = 1$ when $x \neq 0$ and $0^{(0)} = 0$. When X has a binomial distribution with the parameters m and θ, it is easy to verify that $E\{X^{(k)}\} = m^{(k)}\theta^k$. When R has a shifted binomial distribution, $m - R$ has a binomial distribution. Hence, when R has $\mathrm{SBM}(m, G)$ or $\mathrm{SBM}(m, G')$ distributions, the factorial moments of $(m-R)$ are found to be

$$\begin{aligned}E_G\{(m-R)^{(k)}\} &= \pi(m-1)^{(k)}\theta_1^k + (1-\pi)(m-s)^{(k)}\theta_2^k, \\ E_{G'}\{(m-R)^{(k)}\} &= \alpha(m-1)^{(k)}\eta_1^k + (1-\alpha)(m-t)^{(k)}\eta_2^k.\end{aligned}$$

When $\mathrm{SBM}(m, G) = \mathrm{SBM}(m, G')$, we have

$$\pi(m-1)^{(k)}\theta_1^k + (1-\pi)(m-s)^{(k)}\theta_2^k = \alpha(m-1)^{(k)}\eta_1^k + (1-\alpha)(m-t)^{(k)}\eta_2^k$$

for $k = 0, 1, \ldots, m-1$. Introduce column vectors

$$\mathbf{a}_1 = (a_{10}, a_{11}, \ldots, a_{1,m-1})^\top,\ \mathbf{a}_2 = (a_{20}, a_{21}, \ldots, a_{2,m-1})^\top,$$
$$\mathbf{b}_1 = (b_{10}, b_{11}, \ldots, b_{1,m-1})^\top,\ \mathbf{b}_2 = (b_{20}, b_{21}, \ldots, b_{2,m-1})^\top$$

with its entries being

$$a_{1k} = \pi(m-1)^{(k)}, \quad a_{2k} = (1-\pi)(m-s)^{(k)}; \tag{2}$$
$$b_{1k} = \alpha(m-1)^{(k)}, \quad b_{2k} = (1-\alpha)(m-t)^{(k)}. \tag{3}$$

We summarize the moment equations by a matrix equation:

$$\begin{bmatrix} \mathbf{a}_1 & -\mathbf{b}_1 & \mathbf{a}_2 & -\mathbf{b}_2 \end{bmatrix} \times \begin{bmatrix} 1 & \theta_1 & \theta_1^2 & \ldots & \theta_1^{m-1} \\ 1 & \eta_1 & \eta_1^2 & \ldots & \eta_1^{m-1} \\ 1 & \theta_2 & \theta_2^2 & \ldots & \theta_2^{m-1} \\ 1 & \eta_2 & \eta_2^2 & \ldots & \eta_2^{m-1} \end{bmatrix} = 0.$$

The second matrix in the above equation is a Vandermonde matrix. Such a matrix has full row rank 4 when $m \geq 5$ and θ_1, η_1, θ_2, η_2 have distinct values. When $m \geq 5$ and the above matrix equation holds, we show $G = G'$. The proof is straightforward as claimed in the main paper. The task here is go over each of many trivial and tedious cases. We form cases by how many distinct values θ_1, η_1, θ_2, η_2 assume and organize this supplement accordingly.

1 The first Case

All $\theta_1, \theta_2, \eta_1, \eta_2$ have distinct values.

In this case, the Vandermonde matrix has full row rank 4. The linear combination of four rows is not zero unless all coefficients are zero. Therefore, $\mathbf{a}_1 = \mathbf{b}_1 = \mathbf{a}_2 = \mathbf{b}_2 = 0$.

This leads to equations on the mixing proportions and shift parameters:

$$\pi(m-1)^{(k)} = (1-\pi)(m-s)^{(k)} = \alpha(m-1)^{(k)} = (1-\alpha)(m-t)^{(k)} = 0$$

for $k = 0, 1, \ldots, m-1$ in view of (2) and (3). Note that all the terms are non-negative and $0^{(0)} = 0$. There exists only one solution: $\pi = \alpha = 0$ and $s = t = m$. The corresponding mixing distributions are as follows

$$G = 0\{(1, \theta_1)\} + 1\{(m, 0)\}; \quad G' = 0\{(1, \eta_1)\} + 1\{(m, 0)\}.$$

It is seen $G = G'$ regardless $\theta_1 \neq \eta_1$.

2 The Second case

Parameters $\theta_1, \theta_2, \eta_1, \eta_2$ assume 3 distinct values.

In this case, there can be only three sub-cases due to symmetry: (a) $\theta_1 = \theta_2$, (b) $\theta_1 = \eta_1$, and (c) $\theta_2 = \eta_2$.

In sub-case (a), with $\theta_1 = \theta_2$, the matrix equation is simplified to the following

$$\begin{bmatrix} \mathbf{a}_1 + \mathbf{a}_2 & -\mathbf{b}_1 & -\mathbf{b}_2 \end{bmatrix} \times \begin{bmatrix} 1 & \theta_1 & \theta_1^2 & \ldots & \theta_1^{m-1} \\ 1 & \eta_1 & \eta_1^2 & \ldots & \eta_1^{m-1} \\ 1 & \eta_2 & \eta_2^2 & \ldots & \eta_2^{m-1} \end{bmatrix} = 0.$$

The linear independence of the three rows imply $\mathbf{a}_1 + \mathbf{a}_2 = \mathbf{b}_1 = \mathbf{b}_2 = 0$. They imply

$$\pi(m-1)^{(k)} + (1-\pi)(m-s)^{(k)} = \alpha(m-1)^{(k)} = (1-\alpha)(m-t)^{(k)} = 0$$

for $k = 0, 1, \ldots, m-1$ in view of (2) and (3). Note again that all the terms are non-negative. Hence, it is the same as

$$\pi(m-1)^{(k)} = (1-\pi)(m-s)^{(k)} = \alpha(m-1)^{(k)} = (1-\alpha)(m-t)^{(k)} = 0,$$

which implies $G = G'$, as shown in the first case.

In sub-case (b) with $\theta_1 = \eta_1$, the matrix equation can be simplified in a similar way.

This time, we get $\mathbf{a}_1 - \mathbf{b}_1 = \mathbf{a}_2 = \mathbf{b}_2 = 0$. They imply

$$\pi(m-1)^{(k)} - \alpha(m-1)^{(k)} = (1-\pi)(m-s)^{(k)} = (1-\alpha)(m-t)^{(k)} = 0$$

for $k = 0, 1, \ldots, m-1$ in view of (2) and (3). The first term implies $\pi = \alpha$. If $\pi = \alpha = 1$, it leads to $G = G'$ because $\theta_1 = \eta_1$. If $\pi = \alpha < 1$, the second and third term are zero only if $s = t = m$. This leads to the conventional value $\theta_2 = \eta_2 = 0$ as required by $\mathbb{G}$. The corresponding mixing distributions are as follows:

$$G = G' = \pi\{(1, \theta_1)\} + (1-\pi)\{(m, 0)\}.$$

That is, we also have $G = G'$.

In sub-case (c) with $\theta_2 = \eta_2$, the matrix equation leads to $\mathbf{a}_1 = \mathbf{b}_1 = \mathbf{a}_2 - \mathbf{b}_2 = 0$. They imply

$$\pi(m-1)^{(k)} = \alpha(m-1)^{(k)} = (1-\pi)(m-s)^{(k)} - (1-\alpha)(m-t)^{(k)} = 0$$

for $k = 0, 1, \ldots, m-1$ in view of (2) and (3). The first two entries imply $\pi = \alpha = 0$. Subsequently, the third entry has solution $s = t$. With $\theta_2 = \eta_2$, the resulting G and G' are given by

$$G = 0\{(1, \theta_1)\} + 1\{(s, \theta_2)\}; \quad G' = 0\{(1, \eta_1)\} + 1\{(s, \theta_2)\}.$$

It is seen that $G = G'$ regardless of $\theta_1 \neq \eta_1$.

3 The third case

Parameters $\theta_1, \theta_2, \eta_1, \eta_2$ assume two distinct values.

In spite of symmetry, there are still five different sub-cases: (a) $\theta_1 = \theta_2 = \eta_1 \neq \eta_2$, (b) $\theta_1 = \theta_2 = \eta_2 \neq \eta_1$, (c) $\theta_1 = \theta_2 \neq \eta_1 = \eta_2$, (d) $\theta_1 = \eta_1 \neq \theta_2 = \eta_2$, and (e) $\theta_1 = \eta_2 \neq \theta_2 = \eta_1$. They are formed by having one of $\theta_1, \theta_2, \eta_1, \eta_2$ singled out to have a different value or by having their values paired up.

For sub-case (a) $\theta_1 = \theta_2 = \eta_1 \neq \eta_2$, the simplified matrix equation implies $\mathbf{a}_1 - \mathbf{b}_1 +$

$\mathbf{a}_2 = \mathbf{b}_2 = 0$. They imply

$$\pi(m-1)^{(k)} - \alpha(m-1)^{(k)} + (1-\pi)(m-s)^{(k)} = (1-\alpha)(m-t)^{(k)} = 0$$

for $k = 0, 1, \ldots, m-1$ in view of (2) and (3). Suppose $s < m$. Letting $k = 0$ in the first entry, we obtain

$$\pi(m-1)^{(0)} - \alpha(m-1)^{(0)} + (1-\pi)(m-s)^{(0)} = 1-\alpha = 0,$$

so we have $\alpha = 1$. With $\alpha = 1$ and letting $k = 1$, we get

$$\pi(m-1)^{(1)} - (m-1)^{(1)} + (1-\pi)(m-s)^{(1)} = (1-\pi)(1-s) = 0.$$

The equality holds only when $s = 1$ or $\pi = 1$. Both result in G and G' given by

$$G = \pi\{(1, \theta_1)\} + (1-\pi)\{(1, \theta_1)\}; \quad G' = 1\{(1, \theta_1)\} + 0\{(t, \eta_2)\}.$$

It is seen that $G = G'$ in spite of their different appearances. Suppose $s = m$ in the first place. Then $\pi(m-1)^{(k)} - \alpha(m-1)^{(k)} + (1-\pi)(m-s)^{(k)} = 0$, which implies $\pi = \alpha$. If $\pi = \alpha = 1$, then

$$G = 1\{(1, \theta_1)\} + 0\{(s, \theta_1)\}; \quad G' = 1\{(1, \theta_1)\} + 0\{(t, \eta_2)\}$$

and $G = G'$. If $\pi = \alpha < 1$, then $(1-\alpha)(m-t)^{(k)} = 0$ implies $t = m$. Hence, $G = G'$.

In the sub-case (b), $\theta_1 = \theta_2 = \eta_2 \neq \eta_1$, the solution to the simplified matrix equation must satisfy $\mathbf{a}_1 + \mathbf{a}_2 - \mathbf{b}_2 = \mathbf{b}_1 = 0$. From $\mathbf{b}_1 = 0$, we find $\alpha = 0$. From $\mathbf{a}_1 - \mathbf{b}_1 - \mathbf{b}_2 = 0$ and examine the situation of $k = 0$, we get we also have

$$\pi - 1 = 0.$$

With $\alpha = \pi = 1$ and $\theta_1 = \theta_2$, we get $G = G' = 1\{1, \theta_1\} + 0\{1, \theta_2\}$.

In the sub-case (c), $\theta_1 = \theta_2 \neq \eta_1 = \eta_2$, the solution to the simplified matrix equation must satisfy $\mathbf{a}_1 - \mathbf{a}_2 = \mathbf{b}_1 - \mathbf{b}_2 = 0$. By inspecting the expressions of $\mathbf{a}_1$ and $\mathbf{a}_2$, we know that it can only happen only if $s = 1$. With $\theta_1 = \theta_2$ given previously and $s = 1$ obtained now, we find $G = 1\{1, \theta_1\}$. The same derivation will lead to $G' = 1\{1, \eta_1\}$. Hence, we have must $G = G'$ when two mixtures are the same.

In the sub-case (d) $\theta_1 = \eta_1 \neq \theta_2 = \eta_2$, the solution to the simplified matrix equation must satisfy $\mathbf{a}_1 - \mathbf{b}_1 = \mathbf{a}_2 - \mathbf{b}_2 = 0$. They imply

$$\pi(m-1)^{(k)} - \alpha(m-1)^{(k)} = (1-\pi)(m-s)^{(k)} - (1-\alpha)(m-t)^{(k)} = 0$$

for $k = 0, 1, \ldots, m-1$ in view of (2) and (3). The first entry implies $\pi = \alpha$. After which, the second entry requires $s = t$. Hence $G = G'$.

In the sub-case (e) $\theta_1 = \eta_2 \neq \theta_2 = \eta_1$, the solution to the simplified matrix equation must satisfy $\mathbf{a}_1 - \mathbf{b}_2 = \mathbf{a}_2 - \mathbf{b}_1 = 0$. If $\pi = 0$, then $\mathbf{a}_1 - \mathbf{b}_2$ implies $\alpha = 1$. Because in addition $\mathbf{a}_2 = \mathbf{b}_1$, it implies $s = 1$. The corresponding mixing distributions are

$$G = 0\{(1, \theta_1)\} + 1\{(1, \theta_2)\}$$

and

$$G' = 1\{(1, \eta_1)\} + 0\{(1, \eta_2)\}$$

While the expressions of G and G' appear different, they are the same because $\eta_1 = \theta_2$. Due to symmetry, $\alpha = 0$ will also lead to $G = G'$.

When both $\pi \neq 0$ and $\alpha \neq 0$, we find that $\mathbf{a}_1 = \mathbf{b}_2 = 0$ and $\mathbf{a}_2 = \mathbf{b}_1$ imply both $s = 1$ and $t = 1$.

$$G = \pi\{(1, \theta_1)\} + (1-\pi)\{(1, \theta_2)\}$$

and

$$G' = \alpha\{(1, \eta_1)\} + (1-\alpha)\{(1, \eta_2)\}.$$

The make both SBM(m, G) and SBM(m, G') plain two component binomial mixtures. Hence, we must have $G = G'$ which holds when $1 - \pi = \alpha$.

We have now exhausted all five sub-cases and therefore completed the proof of the third case.

4 The fourth case

. The parameter values satisfy $\theta_1 = \theta_2 = \eta_1 = \eta_2$.

The solution to the simplified matrix equation must satisfy $\mathbf{a}_1 + \mathbf{a}_2 = \mathbf{b}_1 + \mathbf{b}_2$. It implies

$$\pi(m-1)^{(k)} + (1-\pi)(m-s)^{(k)} = \alpha(m-1)^{(k)} + (1-\alpha)(m-t)^{(k)}$$

for $k = 0, 1, \ldots, m-1$ in view of (2) and (3).

(i) If $s, t > 1$, then $(m-s)^{(k)} = (m-t)^{(k)} = 0$ but $(m-1)^{(k)} = (m-1)!$ when $k = m-1$. The above equation becomes $\pi(m-1)! = \alpha(m-1)!$ so $\pi = \alpha$. After which, the equation becomes $(m-s)^{(k)} = (m-t)^{(k)}$, implying $s = t$. Hence, $G = G'$.

(ii) If $s = 1$ instead, the left side of the above equation is equal to $(m-1)^{(k)}$, and the equation becomes $(m-1)^{(k)} = \alpha(m-1)^{(k)} + (1-\alpha)(m-t)^{(k)}$. The solutions to the equation are $t = 1$ or $\alpha = 1$. They all imply $G = G'$.

5 Conclusion

We have exhausted all cases, therefore have established the identifiability of SBM.