Ann Inst Stat Math (2009) 61:275–289
DOI 10.1007/s10463-007-0144-2

$\mathcal{M}$-decomposability and symmetric unimodal densities in one dimension

Nicholas Chia · Junji Nakano

Received: 7 November 2005 / Revised: 27 April 2007 / Published online: 11 August 2007

Abstract In this paper, we introduce the notion of $\mathcal{M}$*-decomposability* of probability density functions in one dimension. Using $\mathcal{M}$-decomposability, we derive an inequality that applies to all symmetric unimodal densities. Our inequality involves only the standard deviation of the densities concerned. The concept of $\mathcal{M}$-decomposability can be used as a non-parametric criterion for mode-finding and cluster analysis.

Keywords $\mathcal{M}$-decomposability · Symmetric unimodal densities · Inequality · Non-parametric criterion for clustering

1 Introduction

One important class of statistical distributions is the class of symmetric unimodal distributions, among which the Gaussian is perhaps the most commonly used. Unimodality and symmetric unimodality have been previously investigated by Anderson (1955) and Ibragimov (1956), among many others. Without the strongly assumptive functional constraints, the class of symmetric unimodal distributions is more general and flexible than the Gaussian and many others with specific functional forms.

A complimentary class of the symmetric unimodal distributions is the class of multimodal distributions. In this paper, we attempt to quantify the fundamental differences between the densities of unimodal and multimodal distributions. Intuitively,

N. Chia
Department of Statistical Science, School of Multidisciplinary Sciences,
The Graduate University for Advanced Studies, Tokyo, Japan

N. Chia (✉) · J. Nakano
The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu,
Minato-ku, Tokyo 106-8569, Japan
e-mail: sha@ism.ac.jp

it is possible to express a multimodal density as a mixture of functionally simpler, unimodal ones, such that the sum of the standard deviations of each unimodal density component is less than that of the original density. On the other hand, it may be difficult to achieve the same for unimodal densities. The main result of this paper is developed from this relatively simple observation.

In Sect. 2, we introduce the notion of $\mathcal{M}$-*decomposability* of probability density functions in one dimension. Here, the prefix '$\mathcal{M}$' can mean both '*multimodal*' and '*mixture*'. Examples are provided to illustrate the concept of $\mathcal{M}$-decomposability.

In Sect. 3, we demonstrate that unimodal densities can be approximated using a specially constructed mixture of uniform densities. In Sect. 4, we derive an inequality on symmetric unimodal densities. This is the main result of the paper. Using the main result obtained, we provide a concept demonstration of mode-finding of a multimodal density in Sect. 5. Concluding remarks are provided in Sect. 6.

2 $\mathcal{M}$-decomposability

In this paper, all probability density functions are in one dimension. We denote the mean and the standard deviation of a density f by μ_f and σ_f respectively. The density of the uniform distribution on the support $[a, b]$ is denoted by $\mathcal{U}(\cdot\,|a,\, b)$ for $(a < b)$. As for unimodality, we say that f is *unimodal* with mode m if there exists a real number m such that f is non-decreasing on $(-\infty, m)$ and non-increasing on (m, ∞). If f does not satisfy the above, we say that f is *multimodal*. Our definition of unimodality is commonly used in textbooks and is comparable with the definition given in Dharmadhikari and Joag-Dev (1987) and Kotz et al. (2005). If we also have $f(m - x) = f(m + x)$ on top of unimodality, we say that f is *symmetric unimodal* with mode m.

A density f can always be written as a two-component mixture, i.e. in the form

$$f(x) = \alpha\, g(x) + (1 - \alpha)\, h(x), \tag{1}$$

where $0 < \alpha < 1$. Conventionally, g and h are known as the component densities of f. In general, the number of component densities are not limited to two. In this paper, however, the focus is on the decomposition of a density into two components. Henceforth, a pair of $\{g, h\}$ satisfying Eq. (1) shall be called a *decomposition pair* of f. It is clear that there exist infinitely many possible decomposition pairs for a given f.

Definition 1 ($\mathcal{M}$-decomposability) For a given probability density function f, if there exists a decomposition pair $\{g, h\}$ such that

$$\sigma_f > \sigma_g + \sigma_h,$$

then f is defined to be $\mathcal{M}$-decomposable. Otherwise, f is $\mathcal{M}$-undecomposable. If, for all decomposition pairs $\{g, h\}$,

$$\sigma_f < \sigma_g + \sigma_h,$$

then f is strictly $\mathcal{M}$-undecomposable.

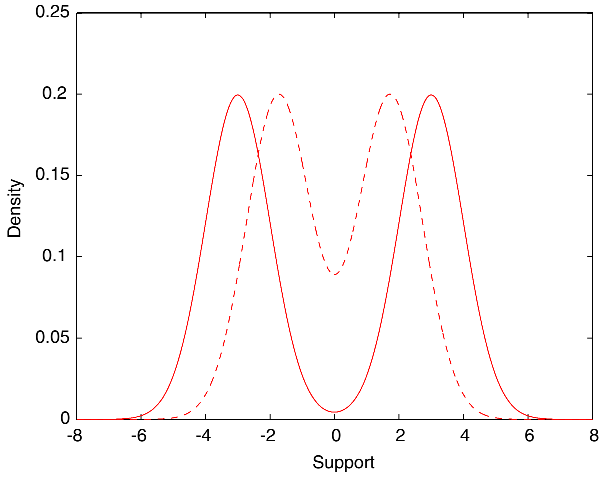

Fig. 1 Densities in Example 1: p with $m = 3$ and $\sqrt{3}$; denoted by *solid* and *broken lines* respectively

Example 1 (Mixture density of 2 Gaussians) Let p be a mixture of two Gaussians such that

$$p(x) = 0.5\mathcal{N}(x|-m,\ 1) + 0.5\mathcal{N}(x|m,\ 1).$$

Here, $\mathcal{N}(\cdot|\mu, \sigma)$ denotes the density of the Gaussian with mean μ and standard deviation σ, and $m \geq 0$. The original density p has a standard deviation σ_p which is $\sqrt{1+m^2}$. One possible decomposition pair $\{q, r\}$ is easily obtained by setting $q(x) = \mathcal{N}(x|-m,\ 1)$ and $r(x) = \mathcal{N}(x|m, 1)$, yielding $\sigma_q + \sigma_r = 2$. If $m > \sqrt{3}$, then $\sigma_p > \sigma_q + \sigma_r$ and accordingly p is $\mathcal{M}$-decomposable. Figure 1 shows the densities of p with $m = 3$ and $m = \sqrt{3}$. The density of p with $m = 3$ is an example of an $\mathcal{M}$-decomposable density.

From the above argument, a density is likely to be $\mathcal{M}$-decomposable if it is a mixture of two distantly located densities. In Example 1, p is $\mathcal{M}$-decomposable for all $m > \sqrt{3}$ by considering the given decomposition pair $\{q, r\}$. It is actually possible to find another decomposition pair $\{q^*, r^*\}$ of p such that $\sigma_{q^*} + \sigma_{r^*} < 2$. For example, set q^* to be p truncated above 0 (hence r^* is p truncated below 0). Then, regardless of m, we must have $\sigma_{q^*} = \sigma_{r^*} < 1$. We are therefore able to conclude that when $m = \sqrt{3}$, p is $\mathcal{M}$-decomposable as well. For $0 < m < \sqrt{3}$, it is difficult to determine the $\mathcal{M}$-decomposability of p.

Next, we present a class of $\mathcal{M}$-undecomposable density.

Theorem 1 *All uniform densities are $\mathcal{M}$-undecomposable.*

To prove Theorem 1, we need to establish the following lemma first.

Lemma 1 (Density with minimum variance) *Let f be a probability density function such that $f(x) \leq M_f < \infty$ for all x. Then*

$$\sigma_f \geq \frac{1}{M_f\sqrt{12}}.$$

Identity holds if and only if f is $\mathcal{U}(\cdot\,|t, t + 1/M_f)$ for real t's.

Proof Set $\mu_f = 0$ without loss of generality. Let the density of u be

$$u(x) = \mathcal{U}\left(x| -\frac{1}{2M_f}, \frac{1}{2M_f}\right).$$

Therefore, $\mu_u = 0$ and $\sigma_u = 1/(M_f\sqrt{12})$. It is also clear that

$$f(x) \begin{cases} \leq u(x) & \text{when } |x| \leq \frac{1}{2\,M_f}\,; \\ \geq u(x) & \text{when } |x| > \frac{1}{2\,M_f}. \end{cases}$$

Since $\mu_f = \mu_u = 0$, we obtain

$$\begin{aligned} \sigma_f^2 - \sigma_u^2 &= \int x^2 \{f(x) - u(x)\}\,\mathrm{d}x \\ &= \int_{|x| \leq \frac{1}{2\,M_f}} x^2\,\{\underbrace{f(x) - u(x)}_{\leq 0}\}\,\mathrm{d}x + \int_{|x| > \frac{1}{2\,M_f}} x^2\,\{\underbrace{f(x) - u(x)}_{\geq 0}\}\,\mathrm{d}x \quad (*) \\ &\geq \frac{1}{4\,M_f^2}\int \{f(x) - u(x)\}\,\mathrm{d}x = 0. \end{aligned}$$

Therefore, $\sigma_f^2 \geq \sigma_u^2$ and hence $\sigma_f \geq \sigma_u$. Identity holds if and only if both terms of $(*)$ equal to 0, that is $f(x) = u(x)$. □

Using Lemma 1, we are ready to prove Theorem 1.

Proof of Theorem 1 Let u be a uniform density. We need to prove that for any decomposition pair $\{v, w\}$ of u,

$$\sigma_u \leq \sigma_v + \sigma_w.$$

Without loss of generality, set $\max(u) = M$ or, equivalently, $\sigma_u = 1/(M\sqrt{12})$. Since $u(x) = \alpha\,v(x) + (1-\alpha)\,w(x)$, we have

$$v(x) \leq \frac{u(x)}{\alpha} \leq \frac{M}{\alpha}; \quad w(x) \leq \frac{u(x)}{1-\alpha} \leq \frac{M}{1-\alpha}. \tag{2}$$

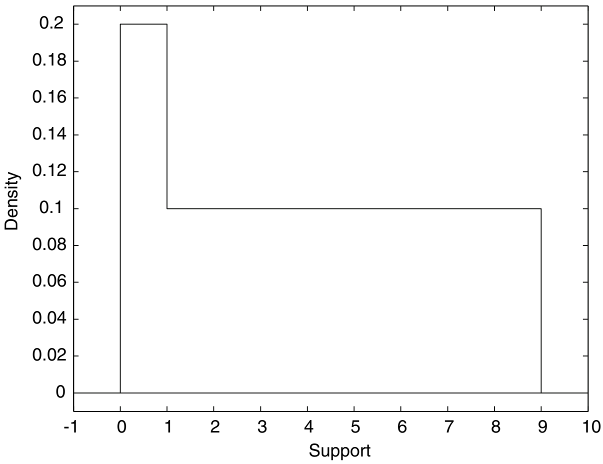

Fig. 2 Density p which is shown in Example 2

Using Lemma 1, the standard deviations of v and w must satisfy

$$\sigma_v \geq \frac{\alpha}{M\sqrt{12}} = \alpha\,\sigma_u; \quad \sigma_w \geq \frac{1-\alpha}{M\sqrt{12}} = (1-\alpha)\,\sigma_u; \tag{3}$$

yielding

$$\sigma_v + \sigma_w \geq \sigma_u. \tag{4}$$

Remark 1 For identity in Eq. (4) to hold, equality has to hold for both cases in Eq. (3). From Lemma 1, this occurs if and only if v is uniform with $\max(v) = M/\alpha$ and w is uniform with $\max(w) = M/(1-\alpha)$. The original density u can be written as $u(x) = \mathcal{U}(x|a,b)$ where $b = a + 1/M$. Identity holds in Eq. (4) if and only if v and w are such that $v(x) = \mathcal{U}(x|a,c)$ and $w(x) = \mathcal{U}(x|c,b)$ where $a < c < b$.

The uniform distribution forms a natural divider between unimodal and multimodal distributions. When the density is cup-shaped with depression occurring near the centre, we have a multimodal distribution. On the other hand, if the density is bell-shaped, with the mode located around the middle, an unimodal distribution is formed. Intuitively, unimodal densities are more likely to be $\mathcal{M}$-undecomposable. In the next example, we investigate the $\mathcal{M}$-decomposability of a skewed unimodal density.

Example 2 (L-shaped density) Let the probability density function p be

$$p(x) = 0.1\,\mathcal{U}(x|0,1) + 0.9\,\mathcal{U}(x|0,9),$$

as depicted in Fig. 2. The standard deviation of p is $\sigma_p = \sqrt{2257/300} > 2.742$. One can also write p as $p(x) = 0.2\,q(x) + 0.8\,r(x)$, where $q(x) = \mathcal{U}(x|0,1)$ and $r(x) = \mathcal{U}(x|1,9)$. Now, we can easily compute $\sigma_q = \sqrt{1/12} < 0.289$ and $\sigma_r = \sqrt{16/3} < 2.310$. Hence, $\sigma_q + \sigma_r < 2.599 < \sigma_p$ and thus p is $\mathcal{M}$-decomposable.

Thus, we have a skewed unimodal density p which is $\mathcal{M}$-decomposable. As such, we conclude that not all unimodal densities are $\mathcal{M}$-undecomposable.

3 Representation of unimodal densities

From Theorem 1, all uniform densities are $\mathcal{M}$-undecomposable. We shall proceed to show that the class of $\mathcal{M}$-undecomposable densities can be extended to include symmetric unimodal densities. For this purpose, we need to represent symmetric unimodal densities via a mixture of uniform densities in a special way presented in this section.

Theorem 2 (Representation of unimodal densities via uniforms) *Let f be an unimodal density whose k^{th} moment is finite and is equal to M, where k is even. Then, for all $\epsilon > 0$, it is possible to construct $g_n = \sum_{i=1}^{n} \omega_i u_i$, a mixture of uniforms such that*

$$\left| \int_{-\infty}^{\infty} x^k \, g_n(x) \, \mathrm{d}x - M \right| < \epsilon.$$

Here, each u_i is the density of the uniform on the interval $I_{i,n}$ satisfying $I_{1,n} \supseteq I_{2,n} \supseteq \cdots \supseteq I_{n,n}$, and the weight ω_i, corresponding to u_i, is proportional to the length of the interval $I_{i,n}$.

Proof of Theorem 2 We can define the following functions on non-negative values of y, for a given f:

$$p(y) = \int_{-\infty}^{\infty} \min\{f(x), y\} \, \mathrm{d}x \, ; \quad q(y) = \int_{-\infty}^{\infty} x^k \min\{f(x), y\} \, \mathrm{d}x.$$

Then, both p and q are increasing with $p(0) = q(0) = 0$. If f is unbounded, then p and q are strictly increasing for all y with $\lim_{y\to\infty} p(y) = 1$ and $\lim_{y\to\infty} q(y) = M$. If f is bounded such that $\max(f) = F$, then p and q are strictly increasing for $0 \le y \le F$ and $p(F) = 1$ and $q(F) = M$.

We can rewrite f as a sum of two positive functions in the form

$$f(x) = f^{(1)}(x) + f^{(2)}(x), \tag{5}$$

where $f^{(1)}(x) = \min\{f(x), Y\}$ and Y is positive. For a given $\epsilon_1 > 0$, it is possible to choose Y such that

$$1 - \epsilon_1 < \int_{-\infty}^{\infty} f^{(1)}(x) \, \mathrm{d}x = p(Y) < 1,$$

$$M - \epsilon_1 < \int_{-\infty}^{\infty} x^k \, f^{(1)}(x) \, \mathrm{d}x = q(Y) < M. \tag{6}$$

The above "slicing" ensures that the function $f^{(1)}$ is bounded from above by Y. Let $h = Y/n$. Define two sets of real numbers $\{a_{n,1}, \ldots, a_{n,n}\}$ and $\{b_{n,1}, \ldots, b_{n,n}\}$ by

$$a_{n,j} = \inf\{x | f(x) \ge jh\} \quad \text{and} \quad b_{n,j} = \sup\{x | f(x) \ge jh\}.$$

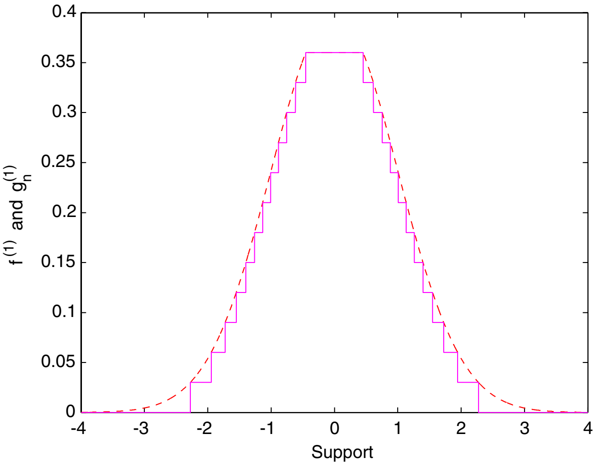

Fig. 3 $f^{(1)}$ dominating $g_n^{(1)}$. Functions $f^{(1)}$ and $g_n^{(1)}$ denoted by *broken* and *solid lines* respectively

Let $I_{n,j}$ denote the interval $[a_{n,j}, b_{n,j}]$ and let $u_{n,j}$ be the density of the uniform on the interval $I_{n,j}$. By construction, a's are monotone non-decreasing and b's are monotone non-increasing, ensuring that $I_{n,1} \supseteq I_{n,2} \supseteq \cdots \supseteq I_{n,n}$. Setting

$$\omega_{n,j} = \frac{b_{n,j} - a_{n,j}}{\sum_{i=1}^{n}(b_{n,i} - a_{n,i})},$$

we create a density g_n such that $g_n(x) = \sum_{j=1}^{n} \omega_{n,j}\, u_{n,j}(x)$. Next, rewrite g_n as a sum of two positive functions in the form of

$$g_n(x) = g_n^{(1)}(x) + g_n^{(2)}(x), \tag{7}$$

where $g_n^{(1)}(x) = \sum_{j=1}^{n}(b_{n,j} - a_{n,j})\, h\, u_{n,j}(x)$. Here, all three functions g_n, $g_n^{(1)}$ and $g_n^{(2)}$ are proportional to one another. Each uniform component $(b_{n,j} - a_{n,j})\, h\, u_{n,j}$ of $g_n^{(1)}$ has thickness h. As depicted in Fig. 3, we have constructed $g_n^{(1)}$ such that it is dominated everywhere by $f^{(1)}$. Unimodality of f ensures that

$$0 \le f^{(1)}(x) - g_n^{(1)}(x) \le \min(f(x), h) \le h.$$

It is then possible to choose n (and hence h) such that

$$\int_{-\infty}^{\infty} \left| g_n^{(1)}(x) - f^{(1)}(x) \right| \,\mathrm{d}x = \int_{-\infty}^{\infty} \left\{ f^{(1)}(x) - g_n^{(1)} \right\} \mathrm{d}x = p(h) < \epsilon_1 \tag{8}$$

$$\int_{-\infty}^{\infty} \left| x^k\, g_n^{(1)}(x) - x^k\, f^{(1)}(x) \right| \,\mathrm{d}x = \int_{-\infty}^{\infty} x^k \left\{ f^{(1)}(x) - g_n^{(1)}(x) \right\} \mathrm{d}x = q(h) < \epsilon_1. \tag{9}$$

Applying the triangle inequality on integrals twice, we have

$$\left|\int_{-\infty}^{\infty} x^k\, g_n(x)\,\mathrm{d}x - M\right| \leq \int_{-\infty}^{\infty} \left|x^k\, g_n(x) - x^k\, f(x)\right|\,\mathrm{d}x$$
$$\leq \int_{-\infty}^{\infty} \left|x^k\, g_n^{(1)}(x) - x^k\, f^{(1)}(x)\right|\,\mathrm{d}x + \int_{-\infty}^{\infty} \left|x^k\, f^{(2)}(x)\right|\,\mathrm{d}x + \int_{-\infty}^{\infty} \left|x^k\, g_n^{(2)}(x)\right|\,\mathrm{d}x. \tag{10}$$

The first term on the last inequality is less than ϵ_1, from Eq. (9). The second term is also less than ϵ_1, ensured by Eqs. (5) and (6). To quantify the third term, note that from Eqs. (7) and (8),

$$\int_{-\infty}^{\infty} g_n^{(2)}(x)\,\mathrm{d}x = 1 - \int_{-\infty}^{\infty} g_n^{(1)}(x)\,\mathrm{d}x < 1 - \int_{-\infty}^{\infty} f^{(1)}(x)\,\mathrm{d}x + \epsilon_1 < 2\,\epsilon_1$$

and therefore,

$$\int_{-\infty}^{\infty} g_n^{(1)}(x)\,\mathrm{d}x > 1 - 2\,\epsilon_1.$$

Furthermore, since $g_n^{(1)}$ and $g_n^{(2)}$ are proportional,

$$g_n^{(2)}(x) < \frac{2\,\epsilon_1}{1-2\,\epsilon_1} \times g_n^{(1)}(x) < \frac{2\,\epsilon_1}{1-2\,\epsilon_1} \times f(x)$$

and hence

$$\int_{-\infty}^{\infty} x^k\, g_n^{(2)}(x)\,\mathrm{d}x < \frac{2\,\epsilon_1}{1-2\,\epsilon_1} \times \int_{-\infty}^{\infty} x^k\, f(x)\,\mathrm{d}x = \frac{2\,\epsilon_1}{1-2\,\epsilon_1} \times M.$$

Choosing $\epsilon_1 < 1/4$, the right side of Eq. (10) becomes less than $2\epsilon_1(1+2M)$. Therefore, starting with any $\epsilon > 0$ and setting

$$\epsilon_1 = \min\left\{\frac{1}{4}, \frac{\epsilon}{2(1+2M)}\right\},$$

we attain

$$\left|\int_{-\infty}^{\infty} x^k\, g_n(x)\,\mathrm{d}x - M\right| < \epsilon$$

with the constructed g_n. □

4 Symmetric unimodal densities

Theorem 3 *(Inequality on symmetric unimodal densities) Let f be a symmetric unimodal density with finite variance. Then, for any decomposition pair $\{g, h\}$,*

$$\sigma_f \leq \sigma_g + \sigma_h.$$

Proof From Theorem 2, it is possible to approximate f as a mixture of uniform components as shown below, such that the variances converge:

$$f(x) = \frac{k_1}{k_1 + \cdots + k_n}\mathcal{U}(x|-k_1, k_1) + \cdots + \frac{k_n}{k_1 + \cdots + k_n}\mathcal{U}(x|-k_n, k_n). \quad (11)$$

Without loss of generality, we have set the mean m to 0. As f and all uniforms appearing in Eq. (11) above have means fixed at 0, the variance of f is computed to be

$$\sigma_f^2 = \int x^2 f(x)\,\mathrm{d}x = \frac{k_1^3 + \cdots + k_n^3}{3\,(k_1 + \cdots + k_n)}. \quad (12)$$

As a result of f being decomposed into a mixture of two densities g and h, each uniform component is consequently broken up into a mixture of two densities as well. The ith uniform component becomes

$$\mathcal{U}(x|-k_i, k_i) = \alpha_i\, v_i(x) + (1 - \alpha_i)\, w_i(x), \quad (13)$$

where α_is are real numbers such that $0 \leq \alpha_i \leq 1$. Here, we allow *some but not all of* α_i's to assume the trivial values of 0 or 1 to ensure the generality of the separation of f. Using Eqs. (11) and (13), we can rewrite f in terms of u_is and v_is as follows:

$$f(x) = \left\{\frac{k_1\,\alpha_1}{k_1 + \cdots + k_n}\, v_1(x) + \cdots + \frac{k_n\,\alpha_n}{k_1 + \cdots + k_n}\, v_n(x)\right\}$$
$$+ \left\{\frac{k_1\,(1-\alpha_1)}{k_1 + \cdots + k_n}\, w_1(x) + \cdots + \frac{k_n\,(1-\alpha_n)}{k_1 + \cdots + k_n}\, w_n(x)\right\}.$$

Assigning $\alpha\, g(x)$ and $(1-\alpha)\, h(x)$ the first and second terms respectively, we have

$$\alpha\, g(x) = \frac{k_1\,\alpha_1}{k_1 + \cdots + k_n}\, v_1(x) + \cdots + \frac{k_n\,\alpha_n}{k_1 + \cdots + k_n}\, v_n(x),$$
$$(1-\alpha)\, h(x) = \frac{k_1\,(1-\alpha_1)}{k_1 + \cdots + k_n}\, w_1(x) + \cdots + \frac{k_n\,(1-\alpha_n)}{k_1 + \cdots + k_n}\, w_n(x),$$

or more compactly,

$$g(x) \;\propto\; k_1\alpha_1\, v_1(x) + \cdots + k_n\alpha_n\, v_n(x) = l_1\, v_1(x) + \cdots + l_n\, v_n(x), \quad (14)$$

$$h(x) \propto k_1(1-\alpha_1)\, w_1(x) + \cdots + k_n(1-\alpha_n)\, w_n(x) = m_1\, w_1(x) + \cdots + m_n\, w_n(x)$$

where $l_i \equiv k_i \alpha_i$ and $m_i \equiv k_i(1-\alpha_i)$. Note that

$$l_i + m_i = k_i$$

for all $1 \le i \le n$. By the choice of α's, we circumvent the trivial situation where $g(x) = 0$ or $h(x) = 0$ as at least one l must be neither 0 nor 1. The same applies to m's.

Next, using Eq. (13) and Theorem 1, we obtain

$$\sigma_{v_i} \ge \frac{k_i \alpha_i}{\sqrt{3}} = \frac{l_i}{\sqrt{3}} \quad \text{and} \quad \sigma_{w_i} \ge \frac{k_i(1-\alpha_i)}{\sqrt{3}} = \frac{m_i}{\sqrt{3}}.$$

From Eq. (14), μ_g, the mean of g can be expressed in terms of means of v_i's as

$$\mu_g = \frac{l_1\, \mu_{v_1} + \cdots + l_n\, \mu_{v_n}}{l_1 + \cdots + l_n}.$$

Consequently, the variance of g becomes

$$\begin{aligned}
\sigma_g^2 &= \int x^2\, g(x)\, \mathrm{d}x - \mu_g^2 \\
&= \frac{l_1 \sigma_{v_1}^2 + \cdots + l_n \sigma_{v_n}^2}{l_1 + \cdots + l_n} + \left\{ \frac{l_1 \mu_{v_1}^2 + \cdots + l_n \mu_{v_n}^2}{l_1 + \cdots + l_n} - \left(\frac{l_1\, \mu_{v_1} + \cdots + l_n\, \mu_{v_n}}{l_1 + \cdots + l_n} \right)^2 \right\} \\
&\ge \frac{l_1 \sigma_{v_1}^2 + \cdots + l_n \sigma_{v_n}^2}{l_1 + \cdots + l_n} \ge \frac{l_1^3 + \cdots + l_n^3}{3\,(l_1 + \cdots + l_n)}.
\end{aligned} \tag{15}$$

The first inequality in Eq. (15) is the result of Jensen's inequality, ensuring that

$$\frac{l_1 \mu_{v_1}^2 + \cdots + l_n \mu_{v_n}^2}{l_1 + \cdots + l_n} \ge (\frac{l_1\, \mu_{v_1} + \cdots + l_n\, \mu_{v_n}}{l_1 + \cdots + l_n})^2.$$

Similarly, the variance of h can be bounded from below as

$$\sigma_h^2 \ge \frac{m_1^3 + \cdots + m_n^3}{3\,(m_1 + \cdots + m_n)}, \tag{16}$$

yielding,

$$\sigma_g + \sigma_h \ge \frac{1}{\sqrt{3}} \cdot \left\{ \left(\frac{l_1^3 + \cdots + l_n^3}{l_1 + \cdots + l_n} \right)^{\frac{1}{2}} + \left(\frac{m_1^3 + \cdots + m_n^3}{m_1 + \cdots + m_n} \right)^{\frac{1}{2}} \right\}.$$

From Eq. (12), we have

$$\sigma_f = \frac{1}{\sqrt{3}} \cdot \left(\frac{k_1^3 + \cdots + k_n^3}{k_1 + \cdots + k_n} \right)^{\frac{1}{2}}.$$

Therefore, Lemma 2 which follows immediately below is a sufficient condition for the inequality $\sigma_f \le \sigma_g + \sigma_h$ to hold. We are now only left with proof of Lemma 2 to complete the proof of Theorem 3.

Lemma 2 *Let a_i, b_i, c_i be sequences of non-negative real numbers such that for all i, $a_i = b_i + c_i$ and $a_i > 0$. Then the following inequality holds for any positive integer n:*

$$\left(\frac{a_1^3 + \cdots + a_n^3}{a_1 + \cdots + a_n} \right)^{\frac{1}{2}} \le \left(\frac{b_1^3 + \cdots + b_n^3}{b_1 + \cdots + b_n} \right)^{\frac{1}{2}} + \left(\frac{c_1^3 + \cdots + c_n^3}{c_1 + \cdots + c_n} \right)^{\frac{1}{2}}.$$

Equality holds if and only if the sequences a_i, b_i and c_i are linearly dependent.

Proof We prove the inequality in the spirit of Hardy et al. (1988) and Pòlya and Szegö (1972).

Set $\mathbf{x} \equiv [x_1, \ldots, x_n]^T$, $\mathbf{y} \equiv [y_1, \ldots, y_n]^T$ and $\mathbf{z} \equiv [z_1, \ldots, z_n]^T$. and similarly for $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$. Let $\mathbf{x} = t\,\mathbf{y} + (1-t)\,\mathbf{z}$, *i.e.* $x_i = t\,y_i + (1-t)\,z_i$ for all i. Furthermore, define the function ψ as follows:

$$\psi(\mathbf{x}) = \left(\frac{x_1^3 + \cdots + x_n^3}{x_1 + \cdots + x_n} \right)^{\frac{1}{2}} \tag{17}$$

and set $\phi(t) = \psi(t\,\mathbf{y} + (1-t)\,\mathbf{z}) \equiv \psi(\mathbf{x})$ where $0 \le t \le 1$. It suffices to prove that $\phi''(t) \ge 0$ for $0 \le t \le 1$. This is an immediate consequence of Jensen's inequality as $\phi''(t) \ge 0$ implies $\phi(t) \le t\,\phi(0) + (1-t)\,\phi(1)$. Setting $t = 1/2$, we have

$$\psi\left(\frac{\mathbf{y}}{2} + \frac{\mathbf{z}}{2}\right) \le \frac{1}{2}\,\psi(\mathbf{y}) + \frac{1}{2}\,\psi(\mathbf{z}).$$

Denoting by $\mathbf{y} = \mathbf{b},\ \mathbf{z} = \mathbf{c}$, this becomes $2\,\psi(\mathbf{a}/2) \le \psi(\mathbf{b}) + \psi(\mathbf{c})$. Using Eq. (17),

$$\psi\left(\frac{\mathbf{a}}{2}\right) = \left(\frac{1}{2}\right)^{(3-1)\cdot\frac{1}{2}} \cdot \psi(\mathbf{a}) = \frac{1}{2} \cdot \psi(\mathbf{a}).$$

Therefore $\phi''(t) \ge 0$ implies $\psi(\mathbf{a}) \le \psi(\mathbf{b}) + \psi(\mathbf{c})$ as required. Equality holds if and only if $\phi''(t) = 0$.

We define ϕ as

$$\phi(t) = \psi(\mathbf{x}) = (\Sigma\, x_i^3)^{\frac{1}{2}}\,(\Sigma\, x_j)^{-\frac{1}{2}}.$$

Differentiating once with respect to t, we have

$$\phi'(t) = \left(\frac{3}{2}\right) \cdot \phi(t) \cdot [\Sigma\, x_i^3]^{-1} \cdot [\Sigma\, x_k^2\,(y_k - z_k)] - \left(\frac{1}{2}\right) \cdot \phi(t) \cdot [\Sigma x_j]^{-1} \cdot [\Sigma (y_k - z_k)].$$

Differentiating again with respect to t, we have

$$\begin{aligned}\phi''(t) = &\left(\frac{3}{2}\right) \cdot \phi'(t) \cdot [\Sigma\, x_i^3]^{-1} \cdot [\Sigma\, x_k^2\,(y_k - z_k)] \\ &+ \left(\frac{3}{2}\right) \cdot \phi(t) \cdot (-1) \cdot [\Sigma\, x_i^3]^{-2} \cdot (3) \cdot [\Sigma\, x_k^2\,(y_k - z_k)]^2 \\ &+ \left(\frac{3}{2}\right) \cdot \phi(t) \cdot [\Sigma\, x_i^3]^{-1} \cdot (2) \cdot [\Sigma\, x_k\,(y_k - z_k)^2] \\ &- \left(\frac{1}{2}\right) \cdot \phi'(t) \cdot [\Sigma x_j]^{-1} \cdot [\Sigma (y_k - z_k)] \\ &- \left(\frac{1}{2}\right) \cdot \phi(t) \cdot (-1) \cdot [\Sigma x_j]^{-2} \cdot [\Sigma (y_k - z_k)]^2.\end{aligned}$$

After some rearrangements, we have

$$\begin{aligned}\frac{\phi''(t)}{\phi(t)} = &\frac{3}{4} \cdot \underbrace{\{[\Sigma x_j]^{-1} \cdot [\Sigma (y_k - z_k)] - [\Sigma x_i^3]^{-1} \cdot [\Sigma x_k^2\,(y_k - z_k)]\}^2}_{A} \\ &+ (3) \cdot [\Sigma x_i^3]^{-2} \cdot \underbrace{\{[\Sigma x_i^3] \cdot [\Sigma x_j (y_j - z_j)^2] - [\Sigma x_k^2\,(y_k - z_k)]^2\}}_{B}. \qquad (18)\end{aligned}$$

Here, term A is a square and therefore $A \geq 0$. To prove that $B \geq 0$, set $p_i^2 = x_i^3$ and $q_j^2 = x_j (y_j - z_j)^2$, and therefore we obtain

$$B = [\Sigma p_i^2] \cdot [\Sigma q_j^2] - [\Sigma p_k\, q_k]^2 \geq 0, \qquad (19)$$

as an immediate consequence of Cauchy–Schwarz's inequality. As such, $\phi''(t) \geq 0$, due to the non-negativeness of x_i, y_i and z_i.

Next, for $B = 0$ to hold in Eq. (19), there must exist a real number s such that $p_i = s\, q_i$ for all i, implying that $x_i = s\,(y_i - z_i)$. When this happens, term A in Eq. (18) becomes 0 as well. Combining with the initial condition $x_i = t\, y_i + (1-t)\, z_i$, we have $(s - t)\, y_i = (s - t + 1)\, z_i$, i.e., the sequence y_i and z_i (and hence b_i and c_i) must be linearly dependent to ensure that $A = B = 0$, resulting in $\phi''(t) = 0$. Hence Lemma 2 is proven and that consequently proves Theorem 3. □

The following theorem spells the condition for equality in Theorem 3 to hold.

Theorem 4 *In Theorem 3, $\sigma_f = \sigma_g + \sigma_h$ holds if and only if f is uniform and $f(x) = \mathcal{U}(x|a, b)$, $g(x) = \mathcal{U}(x|a, c)$, $h(x) = \mathcal{U}(x|c, b)$ where $a < c < b$.*

Proof To ensure that $\sigma_f = \sigma_g + \sigma_h$, identities must hold in Eqs. (15) and (16). In Eq. (15), identity in the first inequality is achievable only if $\mu_{v_1} = \cdots = \mu_{v_n}$. Similarly, we must have $\mu_{w_1} = \cdots = \mu_{w_n}$. As for the second inequality in Eq. (15), identity holds if and only if $v_i(x) = \mathcal{U}(x| - k_i, K_i)$ and $w_i = \mathcal{U}(x|K_i, k_i)$ for all i. When this occurs, we have $|\mu_{v_i} - \mu_{w_i}| = l_i + m_i = k_i$ (or $\mu_{v_i} - \mu_{w_i} = \pm k_i$) for all i. The only possible solution is $k_1 = \cdots = k_n$ and $K_1 = \cdots = K_n$. Hence, the necessary condition is that f is uniform with the prescribed decomposition. The sufficient condition is trivial. □

The results in this section can be summarized as follows: "The uniform density is $\mathcal{M}$-undecomposable. All other symmetric unimodal densities with finite second moments are strictly $\mathcal{M}$-undecomposable."

5 Concept demonstrator of clustering multimodal densities

Example 3 (Multimodal density with 3 peaks) Let p be a probability density function with the following functional form:

$$p(x) = \frac{1}{3}\mathcal{U}(x|-3, -2) + \frac{1}{3}\mathcal{U}(x|-0.5, 0.5) + \frac{1}{3}\mathcal{U}(x|2, 3).$$

The standard deviation of p is computed as $\sigma_p = \sqrt{4.25} > 2.061$. One possible decomposition pair $\{q, r\}$ is $q(x) = \mathcal{U}(x|-3, -2)$ and $r(x) = 0.5\mathcal{U}(x|-0.5, 0.5) + 0.5\mathcal{U}(x|2, 3)$. Computing standard deviations, we obtain $\sigma_q = 1/\sqrt{12} < 0.289$ and $\sigma_r = \sqrt{79/48} < 1.283$, yielding $\sigma_q + \sigma_r < 1.572 < \sigma_p$. Therefore, p is $\mathcal{M}$-decomposable. Figure 4 depicts the densities of p, q and r.

It is not necessary to stop here. One of the components, r, is again $\mathcal{M}$-decomposable. One possible decomposition pair $\{r_1, r_2\}$ for r is $r_1(x) \equiv \mathcal{U}(x|-0.5, 0.5)$ and $r_2(x) \equiv \mathcal{U}(x|2, 3)$. Computing standard deviations, we have $\sigma_{r_1} + \sigma_{r_2} = 1/\sqrt{3} < \sigma_r$. Therefore, the original density p can be expressed as a mixture of three densities in the form of

$$p(x) = \frac{1}{3}q(r) + \frac{1}{3}r_1(x) + \frac{1}{3}r_2(x).$$

Here, all three modes are recovered simply by rewriting $\mathcal{M}$-decomposable densities as mixtures of "simpler" densities. This "toy" example demonstrates the possibility of applying $\mathcal{M}$-decomposability to mode-finding, which forms a natural criterion to cluster analysis.

6 Concluding remarks

In this paper, the concept of $\mathcal{M}$-decomposability is introduced. A probability density function f is defined to be $\mathcal{M}$-decomposable if one can rewrite f as a two-component

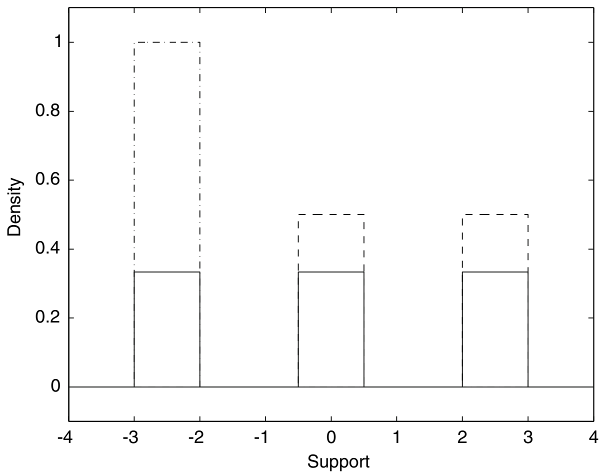

Fig. 4 Densities p, q, r shown in Example 3; denoted by *solid*, *dotted* and *broken lines* respectively

mixture such that the sum of the standard deviations of the components is less than the standard deviation of f. Otherwise f is $\mathcal{M}$-undecomposable. In general, multimodal densities are $\mathcal{M}$-decomposable (see Example 3) and unimodal densities are $\mathcal{M}$-undecomposable.

The main theoretical result of this paper (Theorem 3) says that all symmetric unimodal densities with finite variances are $\mathcal{M}$-undecomposable. This result encompasses a wide class of densities, including Gaussian, logistic, Laplace, Von Mises, Student-t, uniform and many other artificial densities. Theorem 3 implies the existence of a certain inherent minimality in symmetric unimodal densities in statistics, making it a theoretically justificable basis for mode-finding and cluster analysis.

Intuitively, one expects all unimodal densities to be $\mathcal{M}$-undecomposable. However, a counterexample given in Example 2 shows that unimodality alone is not sufficient to imply $\mathcal{M}$-undecomposability. To be rigorous, symmetric unimodality in Theorem 3 provides a sufficient but not necessary condition for $\mathcal{M}$-undecomposability. For practical purposes, densities which are approximately symmetric unimodal are probably $\mathcal{M}$-undecomposable, though a rigorous proof may be difficult.

As mentioned earlier, one possible usage of the results in this paper is clustering and mode-finding. By imposing a weak assumption of approximate symmetric unimodality on clusters or modes, one can use $\mathcal{M}$-undecomposability as a criterion for cluster analysis. This is because if f is $\mathcal{M}$-undecomposable, then from Theorem 3, f cannot be symmetric unimodal. Furthermore, it is possible to express f as a two-component mixture, each component having a much smaller standard deviation. Example 3 demonstrates the concept of non-parametric clustering or mode-finding. One is able to find the modes of a trimodal density without assumptions of the functional forms of modes. Future work in this area should include the extension of $\mathcal{M}$-decomposability to d-dimensions ($d > 1$) and application of $\mathcal{M}$-decomposability in scientific and engineering fields, e.g., clustering and machine learning, among others.

Acknowledgments The authors are indebted to Katuomi Hirano for considerate and kind comments. Two anonymous referees, whose useful and helpful insights tremendously improve the quality and readability of the paper, are gratefully acknowledged. Nicholas Chia dedicates this paper to Taichi Morichika and wishes him all the best in his adventures in life.

References

Anderson, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proceedings of the American Mathematical Society*, *6*, 170–176.

Dharmadhikari, S. W., Joag-Dev, K. (1987). *Unimodality, convexity and applications*. New York: Academic

Hardy, G., Littlewood, J. E., Pòlya, G. (1988). *Inequalities* (2nd ed.), Cambridge: Cambridge University Press.

Ibragimov, I. A. (1956, in Russian). On the composition of unimodal distributions. *Theory of Probability and its Applications*, *1*, 255–260.

Kotz, S., Read, C. B., Balakrishnan, N., Vidakovic, B. (2005). *Encyclopedia of statistical sciences* (16 Volume Set, 2nd ed.), New York: John Wiley and Sons.

Pòlya, G., Szegö, G. (1972). *Problems and theorems in analysis I* (English ed.), Berlin: Springer.