Silesian J. Pure Appl. Math.
vol. 6, is. 1 (2016), 59–67

Grzegorz SITEK

Department of Statistics, University of Economics in Katowice, Poland

THE MODES OF A MIXTURE OF TWO NORMAL DISTRIBUTIONS

Abstract. Mixture distributions arise naturally where a statistical population contains two or more subpopulations. Finite mixture distributions refer to composite distributions constructed by mixing a number K of component distributions. The first account of mixture data being analyzed was documented by Pearson in 1894. We consider the distribution of a mixture of two normal distributions and investigate the conditions for which the distribution is bimodal. This paper presents a procedure for answering the question of whether a mixture of two normal distributions which five known parameters μ_1, μ_2, σ_1, σ_2, p is unimodal or not. For finding the modes, a simple iterative procedure is given. This article presents the possibility of estimation of modes using biaverage.

1. Introduction

Mixture models have been widely used in econometrics and social science, and the theories for mixture models have been well studied Lindsay (see [5]). The importance of the research for unimodality or bimodality in statistics have been described by Murphy (see [6]). Consider

$$f(x,p) = pf_1(x) + (1-p)f_2(x) \tag{1}$$

2010 Mathematics Subject Classification: 60E05, 62E99.
Keywords: mixture of normal distributions, bimodal distributions, biaverage.
Corresponding author: G. Sitek (grzegors12@wp.pl).
Received: 25.07.2016.

where, for $i = 1, 2$,

$$f_i(x) = \frac{1}{\sigma_i\sqrt{2\pi}} e^{-(x-\mu_i)^2/2\sigma_i^2}$$

with $0 < p < 1$. The function $f(x, p)$ is the probability density function of a mixture of two normal distributions. Here we are concerned with the study of the modes of the mixtures (1). The density $f(x, p)$ may have more then one mode, but, except in a very special case, there is no simple rule to know whether the mixture is unimodal or bimodal.

2. Theoretical discussion

The separation of the components of a two-component Gaussian mixture can be expressed by the difference between the component means, which is

$$\Delta = \mu_2 - \mu_1.$$

Eisenberger (see [4]) gave the sufficient condition that a mixture is unimodal if

$$\Delta^2 < \frac{27\sigma_1^2\sigma_2^2}{4(\sigma_1^2 + \sigma_2^2)}.$$

Accordingly to this condition, a mixture with $\sigma_1 = \sigma_2 = 1$ is unimodal for $\Delta < 1.84$. Behboodian (see [3]) considered this problem, too, and derived the following sufficient condition for a mixture of two Gaussian distributions to be unimodal

$$\Delta \leq 2\min\{\sigma_1, \sigma_2\}.$$

Since $\sigma_1 = \sigma_2 = 1$ is assumed, his classification corresponds to that one chosen in this work,which is $\Delta < 2$.

We consider the following conditions:

1) If $\mu_1 = \mu_2$ $f(x, p)$ is unimodal for all p, $0 < p < 1$.

$$f'(x, p) = -\frac{p(x-\mu_1)}{\sqrt{2\pi}\sigma_1^3} \exp\left[\frac{-(x-\mu_1)^2}{2\sigma_1^2}\right] - \\ - \frac{(1-p)(x-\mu_2)}{\sqrt{2\pi}\sigma_2^3} \exp\left[\frac{-(x-\mu_2)^2}{2\sigma_2^2}\right] = 0. \quad (2)$$

Equation (2) has one root $x = \mu_1 = \mu_2$.

2) If $\mu_1 \neq \mu_2$ and $\sigma_1 = \sigma_2 = \sigma$ and $p = \frac{1}{2}$.

The mixture density

$$f(x,p) = \frac{0.5}{\sqrt{2\pi}\sigma} \exp\left[\frac{-(x-\mu_1)^2}{2\sigma^2}\right] + \frac{0.5}{\sqrt{2\pi}\sigma} \exp\left[\frac{-(x-\mu_2)^2}{2\sigma^2}\right]$$

of two normal probability density functions with the same standard deviation, σ, but with different means, μ_1 and μ_2 respectively, is bimodal if and only if

$$|\mu_2 - \mu_1| > 2\sigma.$$

Depending on the distance between μ_1 and μ_2, the mixture density will have either a maximum at $x_0 = \frac{1}{2}(\mu_1 + \mu_2)$ (the unimodal case) or a local minimum at $x_0 = \frac{1}{2}(\mu_1 + \mu_2)$ (the bimodal case). Indeed, x_0 is a stationary point because

$$\begin{aligned} f'(x_0) = {} & \frac{1}{2\sqrt{2\pi}\sigma} \frac{-(x_0-\mu_1)}{\sigma^2} \exp\left[\frac{-(x_0-\mu_1)^2}{2\sigma^2}\right] + \\ & + \frac{1}{2\sqrt{2\pi}\sigma} \frac{-(x_0-\mu_2)}{\sigma^2} \exp\left[\frac{-(x_0-\mu_2)^2}{2\sigma^2}\right] = \\ = {} & \frac{1}{2\sqrt{2\pi}\sigma} \frac{\frac{-(\mu_2-\mu_1)}{2}}{\sigma^2} \exp\left[\frac{-(\frac{\mu_2-\mu_1}{2})^2}{2\sigma^2}\right] + \frac{1}{2\sqrt{2\pi}\sigma} \frac{\frac{-(\mu_1-\mu_2)}{2}}{\sigma^2} \exp\left[\frac{-(\frac{\mu_2-\mu_1}{2})^2}{2\sigma^2}\right] = 0. \end{aligned}$$

Now we must check the second derivative to see whether a maximum or a minimum occurs.

$$\begin{aligned} f''(x_0) = {} & \frac{1}{2\sqrt{2\pi}\sigma^3}\Bigg(- \exp\left[\frac{-(x_0-\mu_1)^2}{2\sigma^2}\right] + \frac{(x_0-\mu_1)^2}{\sigma^2} \exp\left[\frac{-(x_0-\mu_1)^2}{2\sigma^2}\right] - \\ & - \exp\left[\frac{-(x_0-\mu_2)^2}{2\sigma^2}\right] + \frac{(x_0-\mu_2)^2}{\sigma^2} \exp\left[\frac{-(x_0-\mu_2)^2}{2\sigma^2}\right]\Bigg) = \\ = {} & \frac{1}{2\sqrt{2\pi}\sigma^3}\Bigg(- \exp\left[\frac{\left(\frac{-(\mu_2-\mu_1)}{2}\right)^2}{2\sigma^2}\right] + \frac{\left(\frac{-(\mu_2-\mu_1)}{2}\right)^2}{\sigma^2} \exp\left[\frac{\left(\frac{-(\mu_2-\mu_1)}{2}\right)^2}{2\sigma^2}\right] - \\ & - \exp\left[\frac{\left(\frac{-(\mu_1-\mu_2)}{2}\right)^2}{2\sigma^2}\right] + \frac{\left(\frac{-(\mu_1-\mu_2)}{2}\right)^2}{\sigma^2} \exp\left[\frac{\left(\frac{-(\mu_1-\mu_2)}{2}\right)^2}{2\sigma^2}\right]\Bigg) = \\ = {} & \frac{1}{2\sqrt{2\pi}\sigma^3}\Bigg(- \exp\left[\frac{\left(\frac{-(\mu_2-\mu_1)}{2}\right)^2}{2\sigma^2}\right]\Bigg)\Bigg(-1 + \frac{\left(\frac{-(\mu_2-\mu_1)}{2}\right)^2}{\sigma^2}\Bigg) > 0 \end{aligned}$$

if $(\mu_2 - \mu_1)^2 > 4\sigma^2$ or, equivalently, if $|\mu_2 - \mu_1| > 2\sigma$. Thus, a minimum occurs only if the distance between the two means exceeds two standard deviations.

3) If $\mu_1 \neq \mu_2$ and $\sigma_1 \neq \sigma_2$. A sufficient condition that there exists values of p, $0 < p < 1$ for which $f(x,p)$ is bimodal is that

$$(\mu_2 - \mu_1)^2 > \frac{8\sigma_1^2\sigma_2^2}{(\sigma_1^2 + \sigma_2^2)}.$$

For every set of values μ_1, μ_2, σ_1, σ_2 exist p, $0 < p < 1$ for which $f(x,p)$ is unimodal.

Now suppose $\mu_2 > \mu_1$. Since $x = \mu_1$ is not a root of $f'(x,p) = 0$, one can divide (2) by the first term of $f'(x,p)$. After rearranging, one obtains

$$g(x) = \frac{\mu_2 - x}{x - \mu_1} h(x) = \frac{\sigma_2^3 p}{\sigma_1^3(1-p)},$$

where

$$h(x) = \exp\left[-\frac{(x-\mu_2)^2}{2\sigma_2^2} + \frac{(x-\mu_1)^2}{2\sigma_1^2}\right].$$

Since

$$\frac{\sigma_2^3 p}{\sigma_1^3(1-p)} > 0$$

and this term takes on all finite positive values exactly once on the interval $0 < p < 1$ for all fixed values σ_1 and σ_1, each value x for which $g(x) > 0$ there is a root of equation

$$g(x) = \frac{\sigma_2^3 p}{\sigma_1^3(1-p)}$$

for some unique p, and hence is a root of $f'(x,p) = 0$ for exactly one value of p. For $x \geqslant \mu_2$ and $x < \mu_1$, $g(x) \leqslant 0$, so that one is interested only in values of x on the interval $\mu_1 < x < \mu_2$. In this interval $g(x) > 0, g(x) \to \infty$ as $x \to \mu_1$ and $g(\mu_2) = 0$.

Therefore, since $g(x)$ is continuous on $\mu_1 < x < \mu_2$, $g(x)$ takes on all positive finite values at least once in the interval. Moreover, if $g(x)$ is monotone decreasing in this interval, all positive values will be attained exactly once so that since there will then exist a one-to-one correspondence between the values of $g(x)$ for $\mu_1 < x < \mu_2$ and

$$\frac{\sigma_2^3 p}{\sigma_1^3(1-p)}$$

for $0 < p < 1$, $f(x,p)$ will have a single maximum for all p and will be unimodal. Since decreasing monotonicity is implied by $g'(x) < 0$, on $\mu_1 < x < \mu_2$, condition for which this relation is satisfied will now be investigated.

For $\mu_1 < x < \mu_2$:

$$g'(x) = \frac{h(x)}{\sigma_1^2\sigma_2^2(x-\mu_1)^2}\Big[\sigma_1^2(x-\mu_1)(\mu_2-x)^2 + \\ + \sigma_2^2(x-\mu_1)^2(\mu_2-x) - \sigma_2^2\sigma_1^2(\mu_2-\mu_1)\Big] < \\ < \frac{h(x)}{\sigma_1^2\sigma_2^2(x-\mu_1)^2}\Big[\frac{27}{4}(\sigma_1^2+\sigma_2^2)(\mu_2-\mu_1)^3 - \sigma_2^2\sigma_1^2(\mu_2-\mu_1)\Big] < 0$$

if

$$(\mu_2-\mu_1)^2 < \frac{27\sigma_1^2\sigma_2^2}{4(\sigma_1^2+\sigma_2^2)}. \tag{3}$$

Thus for values of $\mu_1, \mu_2, \sigma_1, \sigma_1$, satisfying the inequality (3), $g(x)$ decreases monotonically on $\mu_1 < x < \mu_2$. Then, for each value of p, $0 < p < 1$, there exists only one value of x for which $f'(x,p) = 0$. This must be a maximum since $f(x,p) \to 0$ as $x \to \pm\infty$.

However, for $x = \frac{(\mu_1+\mu_2)}{2}$:

$$g'\Big(\frac{\mu_1+\mu_2}{2}\Big) = \frac{4h(\frac{\mu_1+\mu_2}{2})}{\sigma_1^2\sigma_2^2(\mu_2-\mu_1)^2}\Big[\frac{1}{8}(\sigma_1^2+\sigma_2^2)(\mu_2-\mu_1)^3 - \sigma_2^2\sigma_1^2(\mu_2-\mu_1)\Big] > 0$$

if

$$(\mu_2-\mu_1)^2 > \frac{8\sigma_1^2\sigma_2^2}{(\sigma_1^2+\sigma_2^2)}.$$

3. Biaverage and modes of a mixture of two normal distributions

We consider a mixture of two normal distributions

$$f(x,p) = \frac{p}{\sqrt{2\pi}\sigma_1}\exp\left[\frac{-(x-\mu_1)^2}{2\sigma_1^2}\right] + \frac{1-p}{\sqrt{2\pi}\sigma_2}\exp\left[\frac{-(x-\mu_2)^2}{2\sigma_2^2}\right].$$

The modes of this mixture is determined from the condition

$$pf_1'(x,\mu_1,\sigma_1) + (1-p)f_2'(x,\mu_2,\sigma_2) = 0.$$

This formula can be written as [2]:

$$x = \frac{\frac{p\mu_1}{\sigma_1^3}\exp\left[\frac{-(x-\mu_1)^2}{2\sigma_1^2}\right] + \frac{(1-p)\mu_2}{\sigma_2^3}\exp\left[\frac{-(x-\mu_2)^2}{2\sigma_2^2}\right]}{\frac{p}{\sigma_1^3}\exp\left[\frac{-(x-\mu_1)^2}{2\sigma_1^2}\right] + \frac{(1-p)}{\sigma_2^3}\exp\left[\frac{-(x-\mu_2)^2}{2\sigma_2^2}\right]}. \tag{4}$$

The above equation can be solved iteratively.

Suppose there are two numbers $\overline{m}$ and $\underline{m}$ such that

$$\min_{a,b} = E((X-a)(X-b))^2 = E((X-\overline{m})(X-\underline{m}))^2, \tag{5}$$

where

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx.$$

The numbers $\underline{m}$ and $\overline{m}$ then call biaverage.Biaverage is a two-dimensional vector

$$(\underline{m}, \overline{m}).$$

If a random variable X has four first moments and variance different from zero, the equation (5) has the following solution

$$\underline{m} = \frac{1}{2}\left(P - \sqrt{P^2 + 4Q}\right), \tag{6}$$

$$\overline{m} = \frac{1}{2}\left(P + \sqrt{P^2 + 4Q}\right), \tag{7}$$

$$P = \frac{E(X^3) - E(X^2)E(X)}{E(X^2) - E^2(X)}, \tag{8}$$

$$Q = \frac{E^2(X^2) - E(X^3)E(X)}{E(X^2) - E^2(X)}. \tag{9}$$

The dispersion of the random variable around the biaverage can be calculated as

$$V_0 = E\left((X-\underline{m})(X-\bar{m})\right)^2. \tag{10}$$

Then, the standard deviation of the biaverage has the following form:

$$\sigma_0 = \sqrt[4]{V_0}. \tag{11}$$

The value of biaverage can be evaluated using the random sample

$$X_1, X_2, \dots, X_n$$

chosen from a bimodal population. It is shown (Antoniewicz [1]) that if sample moments are good estimators of population moments, the biaverage is a good estimator of modes and specifies concentration of two probability masses.

4. Examples

Example 1. We consider a mixture of normal distributions with the following parameters

$$\mu_1 = 0,\ \mu_2 = 3,\ \sigma_1 = \sigma_2 = 1,\ p = \frac{2}{3}.$$

Using the formula (4) we calculate iteratively modes: $M_{01} = 0.0175$ and $M_{02} = 2.917$. We calculate the first three raw moments

$$E(X) = p\mu_1 + (1-p)\mu_2 = 1,$$

$$E(X^2) = p(\mu_1^2 + \sigma_1^2) + (1-p)(\mu_2^2 + \sigma_2^2) = 4,$$

$$E(X^3) = p(\mu_1^3 + 3\mu_1\sigma_1^2) + (1-p)(\mu_2^3 + 3\mu_2\sigma_2^2) = 12.$$

Based on the formulas (8) and (9) we are setting the parameters P, Q and biaverage

$$P = \frac{8}{3},\quad Q = \frac{4}{3},\quad \underline{m} = -0.43\ \ \overline{m} = 3.09.$$

Example 2. Some examples of two component Gaussian mixtures are illustrated in Figures 1–4. The figures show mixtures with standard deviations $\sigma_1 = \sigma_2 = 1$, mixing proportions $p = 0.5$ and different component means, starting with $\Delta = \mu_2 - \mu_1 = 1$, ending with $\Delta = 3$.

Example 3. Figure 5 illustrates the dependency of the bimodality property on the parameter p. For both cases

$$(\mu_2 - \mu_1)^2 = 4 > \frac{8\sigma_1^2\sigma_2^2}{(\sigma_1^2 + \sigma_2^2)} = 1.6$$

yet $f(x, 0.85)$ is unimodal although $f(x, 0.4)$ is bimodal.

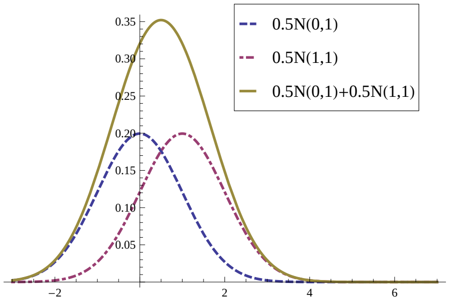

Fig. 1. Mixture of two normal distributions with $\Delta = 1$; the unimodal case

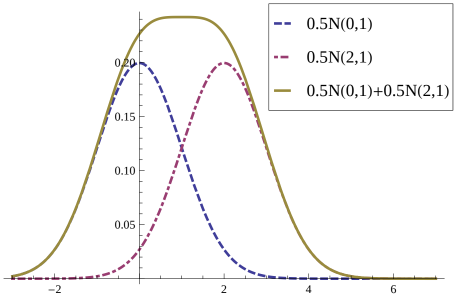

Fig. 2. Mixture of two normal distributions with $\Delta = 2$; the unimodal case

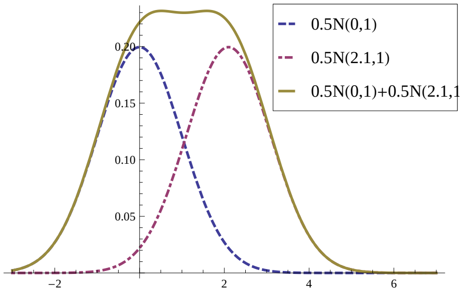

Fig. 3. Mixture of two normal distributions with $\Delta = 2.1$; the bimodal case

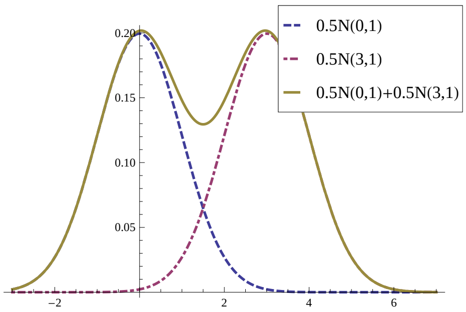

Fig. 4. Mixture of two normal distributions with $\Delta = 3$; the bimodal case

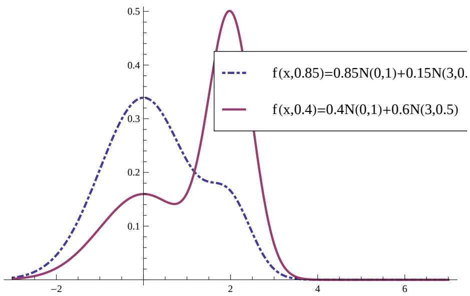

Fig. 5. Mixture of two normal distributions; dependency on p of the bimodality property

References

1. Antoniewicz R.: *About Average and Mean*. Publisher University of Economics, Wrocław 2005 (in Polish).
2. Antoniewicz R., Misztal A.: *Biaverage*. Statist. Rev. **47**, no. 3-4 (2001), 269–274 (in Polish).
3. Behboodian J.: *On the modes of a mixture of two normal distributions*. Technometrics **12**, no. 1 (1970), 131–139.
4. Eisenberger I.: *Genesis of bimodal distributions*. Technometrics **6**, no. 4 (1964), 357–363.
5. Lindsay B.G.: *Mixture Models: Theory, Geometry and Applications*. Institute of Mathematical Statistics, Hayward 1995.
6. Murphy A.: *One cause? Many causes? The argument from the bimodal distribution*. J. Chronic Diseases **6**, no. 4 (1964), 301–324.