A Bayesian Test for the Number of Modes in a Gaussian Mixture

Article *in* Asian Journal of Earth Sciences · July 2021

Sreenivasa Rao Jammalamadaka
University of California, Santa Barbara

Asian Journal of Statistical Sciences
Vol. 1, No. 1, (2021), pp. 09-22

ARF INDIA
Academic Open Access Publishing
Journal homepage link:
http://arfjournals.com/ajss

A Bayesian Test for the Number of Modes in a Gaussian Mixture

Sreenivasa Rao Jammalamadaka[a] and Qianyu Jin [b]

[a] *Department of Statistics and Applied Probability, University of California, Santa Barbara* [b] *JPMorgan Chase & Co.*

ARTICLE HISTORY
Compiled June 20, 2021

Rceived 20 March 2021; Accepted 13 June 2021

ABSTRACT
This paper provides a Bayesian framework for testing the number of modes in a two-component Gaussian mixture. The test is done by first setting up conjugate priors and computing the corresponding posteriors, which are then integrated over the restricted subspace of unimodal parameter space for the mixture distribution, thus obtaining the prior and posterior probabilities of unimodality. Monte Carlo and Gibbs sampling methods are employed to numerically compute these probabilities due to the difficulty in finding analytical solutions. A conclusion on unimodality for the given data is arrived at based on the Bayes factor. Effectiveness of the proposed Bayes test is demonstrated via simulations, and applied to a practical data set on adult human heights in order to answer the question whether the combined height data for men and women is bimodal.

KEYWORDS
Gaussian mixture and unimodal/bimodal distribution and Bayesian test and Monte Carlo method and Gibbs sampling and Bayes factor and human heights

1. Introduction

In this paper we provide a test as to when a mixture of two Gaussian distributions becomes unimodal or bimodal. Certain subspace of the parameter space leads to unimodality of the mixture, and using Bayesian arguments, we find the posterior probability as well as the Bayes Factor for unimodality for a given data set.

2. Mixture of Two Gaussian Distributions

Consider a two-component mixture of Gaussian distributions, say $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ with the mixing parameter $0 < p < 1$, of the form $pN(\mu_1, \sigma_1^2) + (1 - p)N(\mu_2, \sigma_2^2)$, with the probability density function (PDF)

$$f(x|\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p) = pf_1(x|\mu_1, \sigma_1^2) + (1 - p)f_2(x|\mu_2, \sigma_2^2),$$

CONTACT Author[b]. Email: qianyu.homu@gmail.com

where f_1 and f_2 represent the PDFs of $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ respectively. The 5-dimensional parameter vector $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p)$ lies in the parameter space, $\Omega = \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}^+ \times (0, 1)$. Depending on different parameter values, such a Gaussian mixture can be either unimodal or bimodal, so that the parameter space can be partitioned into two disjoint subsets: an Ω_0 where the mixture distribution is unimodal, and an Ω_1 where it is bimodal.

Surprisingly, necessary and sufficient conditions for determining whether a certain parameter vector gives unimodal or bimodal distribution for the general 2-component Gaussian mixture with unequal variances has not yet been fully explored, except for some incomplete discussions (see e.g. [1]). However, for the special case where $\sigma_1^2 = \sigma_2^2 = \sigma^2$, the conditions have been discussed in [3], which we state in the following

Theorem 2.1 (Behboodian (1970) [3]). *The 2-component Gaussian mixture with equal variances $pN(\mu_1, \sigma^2) + (1-p)N(\mu_2, \sigma^2)$ is unimodal if and only if either of the following conditions is satisfied:*

(a) $D^2 \leq 1$,
(b) $D^2 > 1$ *and* $|\log \frac{p}{1-p}| \geq 2\log(D - \sqrt{D^2 - 1}) + 2D\sqrt{D^2 - 1}$,

where $D = \frac{|\mu_1 - \mu_2|}{2\sigma}$.

Let $\theta = (\mu_1, \mu_2, \sigma^2, p)$ denote the parameter vector, and Ω_0 the subspace corresponding to unimodality while its complement Ω_1 where the mixture density becomes bimodal. A visual illustration of the boundary between Ω_0 and Ω_1 is given in Figure 1. For any given σ value, Ω_1 is the region on the right of the colored line (borderline) while Ω_0 is the region on the left.

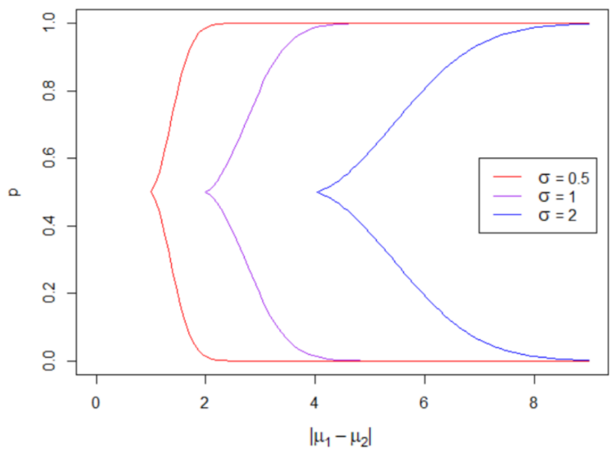


Figure 1. The borderline separating Ω_0 and Ω_1, the unimodal and bimodal parameter space for $pN(\mu_1, \sigma^2) + (1-p)N(\mu_2, \sigma^2)$.

3. A Bayesian Test Procedure

For a sample from a two-component Gaussian mixture with equal variances with unknown parameters, suppose we want to test for the unimodality of the underlying distribution. A parametric likelihood ratio test was proposed and studied in [4]. We propose here an alternative procedure based on Bayesian arguments, along the lines of [2] who discuss testing unimodality of a two-component von Mises mixture.

Let $\mathbf{X} = (X_1, X_2, ..., X_n)$ be i.i.d. random variables from the Gaussian mixture with PDF

$$f(x|\theta) = pf_1(x|\mu_1, \sigma^2) + (1-p)f_2(x|\mu_2, \sigma^2), \tag{1}$$

where f_1, f_2 are the PDFs of $N(\mu_1, \sigma^2), N(\mu_1, \sigma^2)$, and $\mathbf{x} = (x_1, x_2, ..., x_n)$ is a random sample for $\mathbf{X}$. We wish to consider a Bayesian test for testing

$$\mathcal{H}_0 : \theta \in \Omega_0 \quad \text{versus} \quad \mathcal{H}_1 : \theta \in \Omega_1.$$

Let $\pi(\theta)$ denote the joint prior on the vector parameter θ. We compute the joint posterior of θ as $\pi(\theta|\mathbf{x})$ and proceed with the following steps.

(1) Set up a prior $\pi(\theta)$ for the parameter vector $\theta = (\mu_1, \mu_2, \sigma, p)$, and compute the posterior $\pi(\theta|\mathbf{x})$
(2) Calculate prior probability of unimodality

$$\mathbb{P}(\mathcal{H}_0) = \int_{\Omega_0} \pi(\theta) d\theta$$

using Monte Carlo methods
(3) Calculate posterior probability of unimodality

$$\mathbb{P}(\mathcal{H}_0|\mathbf{x}) = \int_{\Omega_0} \pi(\theta|\mathbf{x}) d\theta$$

using Gibbs sampling and Monte Carlo methods
(4) Then the "Bayes factor" is computed as

$$B_{10} = \frac{\mathbb{P}(\mathcal{H}_1|\mathbf{x})\mathbb{P}(\mathcal{H}_0)}{\mathbb{P}(\mathcal{H}_0|\mathbf{x})\mathbb{P}(\mathcal{H}_1)}.$$

A justifiable conclusion may be reached by comparing this Bayes Factor with what is suggested for instance, in [5].

4. Monte Carlo Method for Sampling from Prior and Posterior Distributions

4.1. Computation of the Prior and Posterior Distributions

To perform a Bayes test, we need to start with a prior for the parameter vector and compute the needed conditional posteriors. It would be convenient to use conjugate priors so that the conditional posteriors are in the same family as the priors and have nice analytical forms, as we demonstrate below. This task is made possible by considering the following framework called indicator Gaussian mixture model.

Suppose $\mathbf{x} = (x_1, ..., x_n)$ is an i.i.d random sample of size n from the Gaussian mixture (1). Each observation can be interpreted as being drawn from f_1 with probability

p and being drawn from f_2 with probability $1-p$. Define the indicators

$$z_i = I\{x_i \text{ is drawn from } f_1\},$$

for $i = 1, ..., n$, then there are

$$1 - z_i = I\{x_i \text{ is drawn from } f_2\},$$

In general, $\mathbf{z} = (z_1, ..., z_n)$ is not observable. However, conditionally given $\mathbf{z}$, the conditional distribution of $\mathbf{x}$ on $\mathbf{z}$ is

$$f(\mathbf{x}|\mathbf{z}, \theta) = \prod_{i=1}^{n} [f_1(x_i)]^{z_i} \cdot [f_2(x_i)]^{1-z_i}.$$

Using the fact that the mixture parameter is p, $z_i|\theta \sim Bernoulli(p)$ for $i = 1, ..., n$, and the joint distribution of $\mathbf{x}, \mathbf{z}$ is

$$f(\mathbf{x}, \mathbf{z}|\theta) = \prod_{i=1}^{n} [pf_1(x_i)]^{z_i} \cdot [(1-p)f_2(x_i)]^{1-z_i}. \tag{2}$$

This setup expresses the joint distribution as a product, and we chose appropriate conjugate priors as follows:

$$\sigma^2 \sim InverseGamma(\frac{\nu}{2}, \frac{s^2}{2}), \tag{3}$$
$$\mu_1|\sigma^2 \sim N(\xi_1, \frac{\sigma^2}{m_1}), \tag{4}$$
$$\mu_2|\sigma^2 \sim N(\xi_2, \frac{\sigma^2}{m_2}), \tag{5}$$
$$p \sim Uniform(0, 1), \tag{6}$$

where p is independent of μ_1, μ_2, or σ^2; and $m_1, m_2, \xi_1, \xi_2, \nu, s^2$ are pre-selected hyperparameters.

Theorem 4.1. *Given the sample joint distribution 2 and the priors defined in (3)–(6), the conditional posterior distributions are given by*

$$\sigma^2|\mathbf{x}, \mathbf{z} \sim InvGamma\left(\frac{n+\nu}{2}, \frac{s^2 + \sum_{i=1}^n x_i^2 + m_1\xi_1^2 + m_2\xi_2^2 - C_1 - C_2}{2}\right), \tag{7}$$
$$\mu_1|\sigma^2, \mathbf{x}, \mathbf{z} \sim N\left(\frac{\sum_{i=1}^n z_i x_i + m_1\xi_1}{\sum_{i=1}^n z_i + m_1}, \frac{\sigma^2}{\sum_{i=1}^n z_i + m_1}\right), \tag{8}$$
$$\mu_2|\sigma^2, \mathbf{x}, \mathbf{z} \sim N\left(\frac{\sum_{i=1}^n (1-z_i) x_i + m_2\xi_2}{n - \sum_{i=1}^n z_i + m_2}, \frac{\sigma^2}{n - \sum_{i=1}^n z_i + m_2}\right), \tag{9}$$
$$p|\mathbf{x}, \mathbf{z} \sim Beta\left(\sum_{i=1}^n z_i + 1, n - \sum_{i=1}^n z_i + 1\right). \tag{10}$$

where

$$C_1 = \frac{(\sum_{i=1}^{n} z_i x_i + m_1\xi_1)^2}{m_1 + \sum_{i=1}^{n} z_i},$$
$$C_2 = \frac{(\sum_{i=1}^{n}(1 - z_i)x_i + m_2\xi_2)^2}{m_2 + n - \sum_{i=1}^{n} z_i}.$$

Proof. Using the notation $\pi(\cdot)$ to denote the joint prior and $\pi(\cdot|\mathbf{x}, \mathbf{z})$ to denote the joint posterior. The joint posterior is given by

$$\pi(\theta|\mathbf{x}, \mathbf{z}) = \frac{f(\mathbf{x}, \mathbf{z}|\theta)\pi(\theta)}{\int_\Omega f(\mathbf{x}, \mathbf{z}|\theta)\pi(\theta)d\theta} \tag{11}$$

where $\pi(\theta)$ is the joint prior. Further decomposition of the joint prior and joint posterior distributions are given by

$$\pi(\theta) = \pi(\mu_1|\sigma^2) \cdot \pi(\mu_2|\sigma^2) \cdot \pi(\sigma^2) \cdot \pi(p),$$
$$\pi(\theta|\mathbf{x}, \mathbf{z}) = \pi(\mu_1|\sigma^2, \mathbf{x}, \mathbf{z}) \cdot \pi(\mu_2|\sigma^2, \mathbf{x}, \mathbf{z}) \cdot \pi(\sigma^2|\mathbf{x}, \mathbf{z}) \cdot \pi(p|\mathbf{x}, \mathbf{z}).$$

Therefore, the conditional posterior of a single parameter can be obtained by integrating the joint posterior with respect to other parameters over the appropriate regions. For example,

$$\begin{aligned}
&\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{0}^{+\infty} \pi(\theta|\mathbf{x}, \mathbf{z})d\mu_1 d\mu_2 d(\sigma^2) \\
=&\int_{-\infty}^{+\infty} \pi(\mu_1|\sigma^2, \mathbf{x}, \mathbf{z})d\mu_1 \cdot \int_{-\infty}^{+\infty} \pi(\mu_2|\sigma^2, \mathbf{x}, \mathbf{z})d\mu_2 \\
&\cdot \int_{0}^{+\infty} \pi(\sigma^2|\mathbf{x}, \mathbf{z})d(\sigma^2) \cdot \pi(p|\mathbf{x}, \mathbf{z}) \\
=&\pi(p|\mathbf{x}, \mathbf{z})
\end{aligned}$$

Since the denominator of (11) serves only as a normalizing constant, we have

$$\pi(\mu_1|\sigma^2, \mathbf{x}, \mathbf{z}) \cdot \pi(\mu_2|\sigma^2, \mathbf{x}, \mathbf{z}) \cdot \pi(\sigma^2|\mathbf{x}, \mathbf{z}) \cdot \pi(p|\mathbf{x}, \mathbf{z}) \propto f(\mathbf{x}, \mathbf{z}|\theta)\pi(\theta),$$

meaning that the conditional posterior distributions can be found by decomposing $f(\mathbf{x}, \mathbf{z}|\theta)\pi(\theta)$ into a product of kernels for $\mu_1, \mu_2, \sigma^2, p$, each kernel corresponding to a family of distributions. The details of this operation are shown below:

$$
\begin{aligned}
f(\mathbf{x},\mathbf{z}|\theta)\pi(\theta) &= f(\mathbf{x},\mathbf{z}|\theta)\cdot\pi(\mu_1|\sigma^2)\cdot\pi(\mu_2|\sigma^2)\cdot\pi(\sigma^2)\cdot\pi(p)\\
&= p^{\sum_{i=1}^n z_i}\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{\sum_{i=1}^n z_i}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n z_i(x_i-\mu_1)^2\right\}\\
&\quad\cdot(1-p)^{\sum_{i=1}^n(1-z_i)}\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{\sum_{i=1}^n(1-z_i)}\\
&\quad\cdot\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n(1-z_i)(x_i-\mu_2)^2\right\}\\
&\quad\cdot\frac{\sqrt{m_1}}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{m_1}{2\sigma^2}(\mu_1-\xi_1)^2\right\}\cdot\frac{\sqrt{m_2}}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{m_2}{2\sigma^2}(\mu_2-\xi_2)^2\right\}\\
&\quad\cdot\frac{(s^2/2)^{\nu/2}}{\Gamma(\nu/2)}(\sigma^2)^{-\nu/2-1}\exp\left\{-\frac{s^2}{2\sigma^2}\right\}\cdot 1\\
&\propto T_1(\mu_1,\sigma^2)\cdot T_2(\mu_2,\sigma^2)\cdot T_3(\sigma^2)\cdot T_4(p),
\end{aligned}
$$

where T_1 through T_4 are the kernels for each parameter and have the following forms:

$$
\begin{aligned}
T_1(\mu_1,\sigma^2) &= \frac{1}{\sigma}\exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n z_i(x_i-\mu_1)^2+m_1(\mu_1-\xi_1)^2\right]\right\}\\
&\propto\frac{1}{\sigma}\exp\left\{-\frac{1}{2\sigma^2}\left[\left(\sum_{i=1}^n z_i+m_1\right)\mu_1^2-2\left(\sum_{i=1}^n z_ix_i+m_1\xi_1\right)\mu_1+C_1\right]\right\}\\
&\propto N\left(\frac{\sum_{i=1}^n z_ix_i+m_1\xi_1}{\sum_{i=1}^n z_i+m_1},\frac{\sigma^2}{\sum_{i=1}^n z_i+m_1}\right),\\
T_2(\mu_2,\sigma^2) &= \frac{1}{\sigma}\exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n(1-z_i)(x_i-\mu_2)^2+m_2(\mu_2-\xi_2)^2\right]\right\}\\
&\propto\frac{1}{\sigma}\exp\left\{-\frac{1}{2\sigma^2}\left[\left(n-\sum_{i=1}^n z_i+m_2\right)\mu_2^2\right.\right.\\
&\quad\left.\left.-2\left(\sum_{i=1}^n(1-z_i)x_i+m_2\xi_2\right)\mu_2+C_2\right]\right\}\\
&\propto N\left(\frac{\sum_{i=1}^n(1-z_i)x_i+m_2\xi_2}{n-\sum_{i=1}^n z_i+m_2},\frac{\sigma^2}{n-\sum_{i=1}^n z_i+m_2}\right),\\
T_3(\sigma^2) &= (\sigma^2)^{-\frac{n+\nu}{2}-1}\exp\left\{-\frac{s^2+\sum_{i=1}^n x_i^2+m_1\xi_1^2+m_2\xi_2^2-C_1-C_2}{2\sigma^2}\right\}\\
&\propto InvGamma\left(\frac{n+\nu}{2},\frac{s^2+\sum_{i=1}^n x_i^2+m_1\xi_1^2+m_2\xi_2^2-C_1-C_2}{2}\right),\\
T_4(p) &= p^{\sum_{i=1}^n z_i}(1-p)^{\sum_{i=1}^n(1-z_i)}\\
&\propto Beta\left(\sum_{i=1}^n z_i+1,n-\sum_{i=1}^n z_i+1\right).
\end{aligned}
$$

where

$$C_1 = \frac{(\sum_{i=1}^n z_i x_i + m_1\xi_1)^2}{m_1 + \sum_{i=1}^n z_i},$$
$$C_2 = \frac{(\sum_{i=1}^n (1-z_i) x_i + m_2\xi_2)^2}{m_2 + n - \sum_{i=1}^n z_i}.$$

The distributions that correspond to the 4 kernels are the conditional posterior distributions for each parameter, which correspond to Eqns. (7)–(10), as stated. □

4.2. Computation of the Prior Probabilities of Unimodality and Bimodality

After setting up the priors and finding the conditional posteriors, the next step is to calculate the prior probability of unimodality and bimodality. To calculate the prior probability of unimodality, viz.

$$\mathbb{P}(\mathcal{H}_0) = \int_{\Omega_0} \pi(\theta) d\theta,$$

one needs to integrate the joint priors over the unimodal parameter subspace, Ω_0. Since the criterion that determines the boundary of Ω_0 is complex as can be seen from Figure 1, it is difficult to find an analytical expression for the needed integral. Instead, a Monte Carlo method is employed to get its value numerically.

Algorithm 1 (Monte Carlo method for priors)**.** To numerically compute the prior probability of unimodality and bimodality, denoted by $\mathbb{P}(\mathcal{H}_0)$ and $\mathbb{P}(\mathcal{H}_1)$, these steps are followed.

(1) Determine the values for prior hyperparameters $m_1, m_2, \xi_1, \xi_2, \nu, s^2$.
(2) Generate N parameter vectors $\theta^{(1)}, ..., \theta^{(N)}$ from prior distributions (3) through (6). The order of parameter generation is $\sigma^2 \to \mu_1 \to \mu_2 \to p$.
(3) For each $\theta^{(i)}, i = 1, ..., N$, check if it belongs in Ω_0 using the conditions in Theorem 2.1 and obtain the values of indicators $I\left\{\theta^{(i)} \in \Omega_0\right\}$.
(4) Compute $\mathbb{P}(\mathcal{H}_0) = \frac{1}{N}\sum_{i=1}^N I\left\{\theta^{(i)} \in \Omega_0\right\}$ and $\mathbb{P}(\mathcal{H}_1) = 1 - \mathbb{P}(\mathcal{H}_0)$.

4.3. Computation of the Posterior Probabilities of Unimodality and Bimodality

Similar to the computation of $\mathbb{P}(\mathcal{H}_0)$, the posterior probability of unimodality

$$\mathbb{P}(\mathcal{H}_0|\mathbf{x}) = \int_{\Omega_0} \pi(\theta|\mathbf{x}) d\theta$$

also needs to be computed numerically using a Monte Carlo approach. The conditional posteriors given in Theorem 4.1 are conditional on latent variables $\mathbf{z}$ that are not observable from the sample. To mitigate this problem, we would like to treat $\mathbf{z}$ as unknown parameters and use Gibbs sampling method to generate samples from θ and $\mathbf{z}$ simultaneously.

Algorithm 2 (Gibbs sampling)**.** To generate θ and $\mathbf{z}$ from conditional distributions, these steps are followed.

(1) Set up initial values $\theta^{(0)}$. Then, at the k-th iteration,
(2) Generate $\mathbf{z}^{(k)}$ from $f(\mathbf{z}^{(k)}|\theta^{(i-1)}, \mathbf{x})$.
(3) Generate $\theta^{(k)}$ from $\pi(\theta^{(k)}|\mathbf{x}, \mathbf{z}^{(k)})$.
(4) Repeat Step 2 and 3 for a total of N times until convergence.

To find the conditional distribution of step 2 in Algorithm 2, we have the following results.

Lemma 4.2. *The conditional distribution of each z_i in $\mathbf{z}$ is given by*

$$z_i|\theta, x_i \sim Bernoulli\left(\frac{pf_1(x_i|\mu_1, \sigma^2)}{pf_1(x_i|\mu_1, \sigma^2) + (1-p)f_2(x_i|\mu_2, \sigma^2)}\right).$$

Proof. From (1) and (2) and since x_i's are i.i.d.,

$$\begin{aligned} f(\mathbf{z}|\theta, \mathbf{x}) &= \frac{f(\mathbf{x}, \mathbf{z}|\theta)}{f(\mathbf{x}|\theta)} \\ &= \frac{\prod_{i=1}^{n} [pf_1(x_i|\mu_1, \sigma^2)]^{z_i} \cdot [(1-p)f_2(x_i|\mu_1, \sigma^2)]^{1-z_i}}{\prod_{i=1}^{n} [pf_1(x_i|\mu_1, \sigma^2) + (1-p)f_2(x_i|\mu_2, \sigma^2)]} \\ &= \prod_{i=1}^{n} P_i^{z_i}(1-P_i)^{1-z_i} \end{aligned}$$

where each

$$P_i = \frac{pf_1(x_i|\mu_1, \sigma^2)}{pf_1(x_i|\mu_1, \sigma^2) + (1-p)f_2(x_i|\mu_2, \sigma^2)}.$$

Since each z_i is only dependent on the x_i, and the x_i's are independent, z_i's are independent of each other as well. It is easy to see

$$f(z_i|\theta, x_i) = P_i^{z_i}(1-P_i)^{1-z_i}$$

for $z_i \in \{0, 1\}, i = 1, ..., n$. This is the PMF of $Bernoulli(P_i)$. □

Given the Gibbs sampling procedure, the Monte Carlo calculation is outlined below.

Algorithm 3 (Monte Carlo method for posteriors)**.** To numerically compute the posterior probability of unimodality and bimodality, denoted by $\mathbb{P}(\mathcal{H}_0|\mathbf{x})$ and $\mathbb{P}(\mathcal{H}_1|\mathbf{x})$, these steps are followed.

(1) Generate N parameter vectors $\theta^{(1)}, ..., \theta^{(N)}$ for conditional posterior distributions (7) - (10) by utilizing Gibbs sampling procedure in Algorithm 2. The order of parameter generation is $\sigma^2 \to \mu_1 \to \mu_2 \to p$.
(2) Make sure the Markov chain generated in Step 1 is convergent after a burn-in period of length K.
(3) For each $\theta^{(i)}, i = K+1, ..., N$, check if it belongs in Ω_0 using the conditions in Theorem 2.1 and obtain the values of indicators $I\left\{\theta^{(i)} \in \Omega_0\right\}$

(4) Compute

$$\mathbb{P}(\mathcal{H}_0|\mathbf{x}) = \frac{1}{N-K} \sum_{i=K+1}^{N} I\left\{\theta^{(i)} \in \Omega_0\right\}$$

and

$$\mathbb{P}(\mathcal{H}_1|\mathbf{x}) = 1 - \mathbb{P}(\mathcal{H}_0|\mathbf{x}).$$

Remark: A mixture distribution has two equivalent parameter representations, namely $(\mu_1, \mu_2, \sigma^2, p)$ and $(\mu_2, \mu_1, \sigma^2, 1-p)$. To differentiate between these two parameterizations and uniquely define the mixture, we restrict $p \in (0, 0.5]$, so that the prior of p becomes

$$p \sim Uniform(0, 0.5] \tag{12}$$

and the conditional posterior of p becomes

$$p|\mathbf{x}, \mathbf{z} \sim Beta\left(\sum_{i=1}^{n} z_i + 1, n - \sum_{i=1}^{n} z_i + 1\right) \text{ restricted on } (0, 0.5]. \tag{13}$$

Note that when $p = 0.5$ the two parameterizations become the same.

4.4. Judging the Bayes Factor

[5] discuss the Bayes factor for testing $\mathcal{H}_1$ against $\mathcal{H}_0$ which is defined as

$$B_{10} = \frac{\text{posterior odds}}{\text{prior odds}} = \frac{\mathbb{P}(\mathcal{H}_1|\mathbf{x})\mathbb{P}(\mathcal{H}_0)}{\mathbb{P}(\mathcal{H}_0|\mathbf{x})\mathbb{P}(\mathcal{H}_1)}. \tag{14}$$

It is used as a summary of the evidence provided by data in favor of $\mathcal{H}_1$ and against $\mathcal{H}_0$. In general, larger values of Bayes factor indicate a stronger evidence in favor of $\mathcal{H}_1$. [5] suggest using Table 1 as a reasonable scale for interpreting B_{10} and $\log_{10}(B_{10})$ values. Although not shown in the Table, it should be noted that $0 < B_{10} < 1$ (or equivalently $\log_{10} B_{10} < 0$) indicates almost no evidence against $\mathcal{H}_0$.

Table 1. Suggested scale for interpreting B_{10} and $\log_{10}(B_{10})$ values

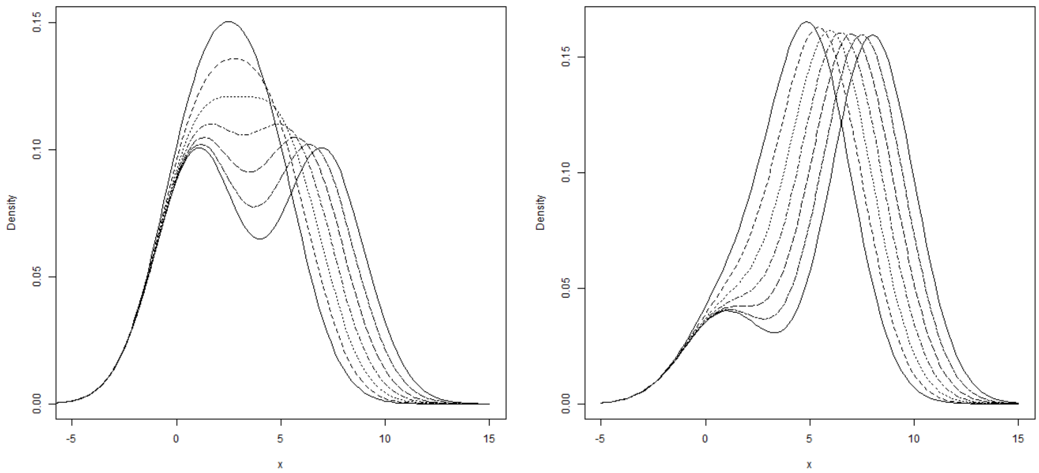
$\log_{10} B_{10}$	B_{10}	Evidence against $\mathcal{H}_0$
0 to 1/2	1 to 3.2	Not worth more than a bare mention
1/2 to 1	3.2 to 10	Substantial
1 to 2	10 to 100	Strong
>2	>100	Decisive

5. A Simulation Study

To examine the proposed Bayesian test via simulation studies, we first select 2 sets of Gaussian mixture distributions to generate the data. Each set of mixtures consists of 7 mixture combinations, among which μ_1, σ^2 and p are fixed while μ_2 are different, resulting in parameter combinations from both the unimodal and bimodal parameter space. The actual parameter values and their corresponding modality are presented in Table 2. The mixing probabilities are even ($p = 0.5$) for distributions from the first set and uneven for distributions from the second set. Figure 2 shows the density plots for mixture distribution within each set. It is clear that as μ_2 moves away from μ_1, the mixture density curve gradually changes from being unimodal to being bimodal.

Table 2. Choice of parameters for simulated data.

	μ_1	σ^2	p	μ_2	modality
Mixture Set 1	1.0	4	0.5	4.0, 4.5, 5.0	unimodal
				5.5, 6.0, 6.5, 7.0	bimodal
Mixture Set 2	1.0	4	0.2	5.0, 5.5, 6.0, 6.5	unimodal
				7.0, 7.5, 8.0	bimodal



(a) Mixture Set 1 (b) Mixture Set 2

Figure 2. Density plots for Gaussian mixtures in Mixture Set 1 and 2 .

For each distribution in the two Mixture sets, we generate 50 samples of size $n = 25$ and 50 samples of size $n = 50$, then conduct Bayesian test for each sample assuming the same prior hyper-parameter values, namely ($\nu = 4, \xi_1 = 0, \xi_2 = 10, s^2 = 30, m_1 = 10$ and $m_2 = 10$). The Bayes factors are obtained by running Algorithms 1 and 3 with $N = 100,000$ and burn-in period of $20,000$ and plugging their results into (14). Restricted prior (12) and conditional posterior (13) for p are used to avoid identifiability issues in the simulation of θ. The initial values of the parameters are $\mu_1 = 0, \mu_2 = 10, \sigma^2 = 100, p = 0.5$ in posterior simulation for all samples.

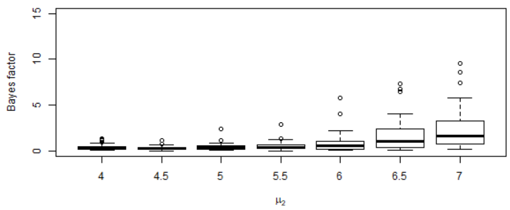
Figure 3 displays the Bayes factors in boxplots for each of the μ_2 values. The plots in the same row correspond to the same sample sizes and the plots in the same column correspond to the same distribution set. Table 3 and 4 show the proportion of times that the Bayes factor exceeds 3.2. From these figures and tables, we see that (i) within each distribution set, the Bayes factor is more likely to have larger values when μ_2 is larger; (ii) given a certain bimodal distribution, the Bayes factor is more likely to have larger values when n is larger. Both these observations (i) and (ii) show that the Bayes factor is more likely to have larger values when there is a stronger evidence that the sample came from a bimodal distribution, showing that our procedure works well in detecting unimodality.

Table 3. Proportion of times that Bayes factors > 3.2 for Mixture Set 1.

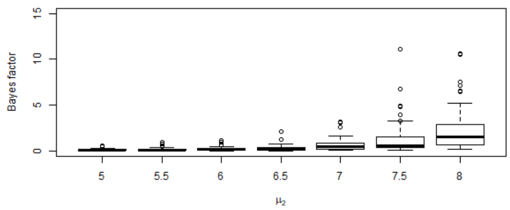
μ_2	4.0	4.5	5.0	5.5	6.0	6.5	7.0
Mixture Set 1, $n = 25$	0.00	0.00	0.00	0.00	0.06	0.12	0.18
Mixture Set 1, $n = 50$	0.00	0.02	0.00	0.00	0.06	0.20	0.46

Table 4. Proportion of times that Bayes factors > 3.2 for Mixture Set 2.

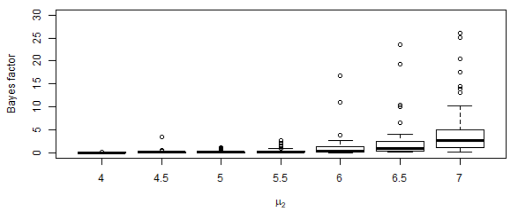
μ_2	5.0	5.5	6.0	6.5	7.0	7.5	8.0
Mixture Set 2, $n = 25$	0.00	0.00	0.00	0.02	0.02	0.14	0.20
Mixture Set 2, $n = 50$	0.00	0.00	0.00	0.00	0.06	0.16	0.44



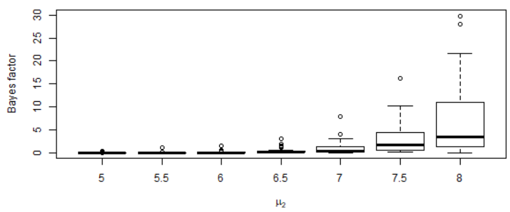
(a) Bayes factor for Mixture Set 1, $n = 25$



(b) Bayes factor for Mixture Set 2, $n = 50$



(c) Bayes factor for Mixture Set 1, $n = 25$



(d) Bayes factor for Mixture Set 2, $n = 50$

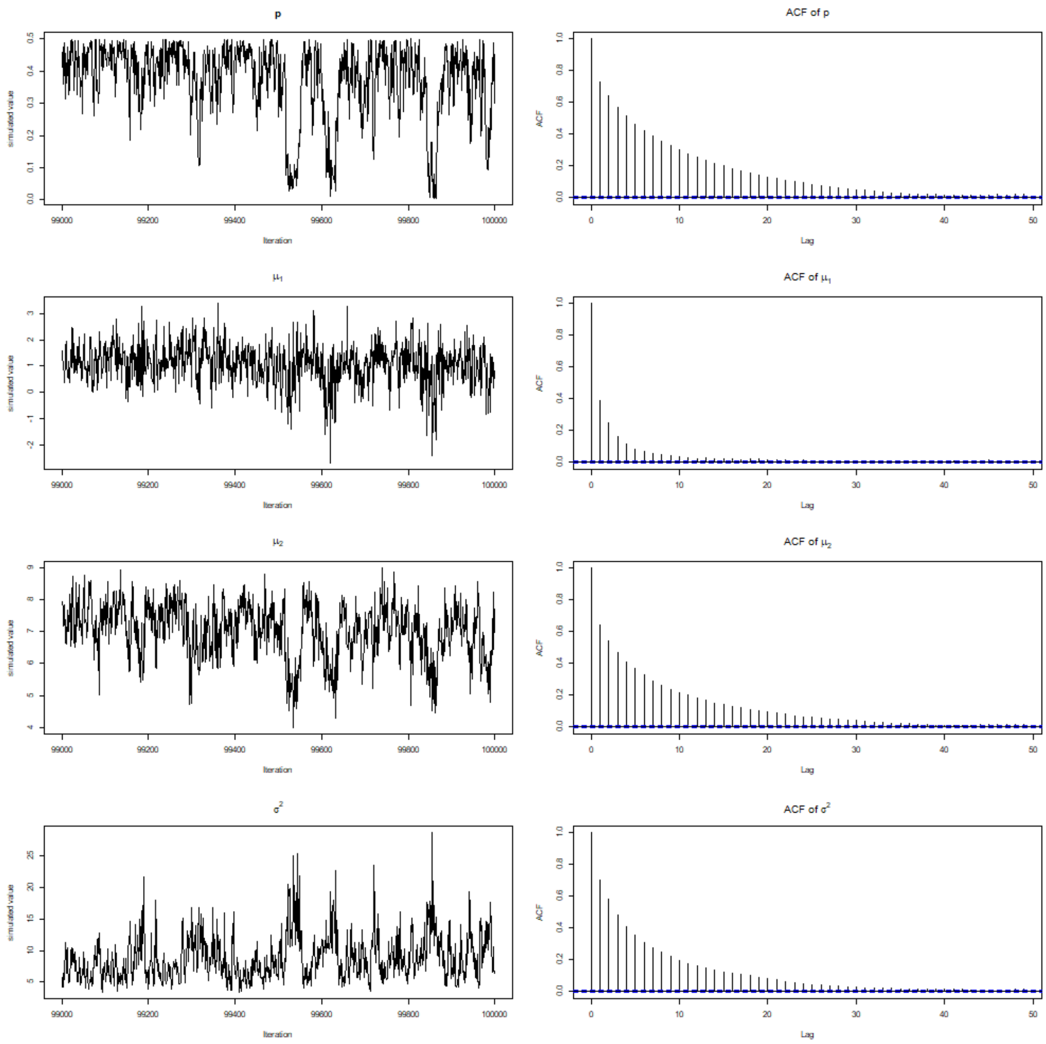
Figure 3. Boxplots of Bayes factors for Mixture Set 1 and 2 and sample size $n = 25, 50$.

To verify the properties of the Markov chain generated in the simulation of the posterior distributions, an example Markov chain that corresponds to a sample from Mixture Set 1, $\mu_2 = 8, n = 50$ is selected, and its path plot and ACF plot for each parameter are shown in Figure 4. Note that only the last 10,000 iterations are plotted in the path plot and only the iterations after the burn-in period (> 20,000) are used to compute the ACFs. It can be seen that the paths of p and σ^2 are able to traverse a wide range within their parameter space and there is no sign of identifiability issue between μ_1 and μ_2. The number of lags required for ACFs to diminish is pretty large (≈ 30) for all parameters.

6. Real Data Application—Adult Height Data

This section presents an application of the modality test on adult human heights. There have been several discussions on whether the combined data on heights of men and women will give a unimodal or bimodal distribution; see e.g. [6]. To revisit this problem, we consider data from 2013–2014 National Health and Nutrition Examination Survey (NHANES)[1]. This is a program of studies to assess the health and nutritional

[1] https://www.cdc.gov/nchs/nhanes/



(a) Path plots (simulated values) (b) ACF plots

Figure 4. Path plots and ACF plots for each parameter in a Markov chain in posterior simulation. The sample is from Mixture Set 1, $\mu_2 = 8, n = 50$.

status of adults and children in the United States, and it is conducted by National Center for Health Statistics (NCHS), which is a part of the Centers for Disease Control and Prevention (CDC). In selecting and screening the data for our test, we took the variables of gender and age from the demographic variable,s and sample heights dataset on the measurement of standing height in centimeters from the Body Measures dataset, matched the records on the unique identifier (respondent sequence number), and filtered out the records when the age is less than 18 or have any other missing values. An overview of the filtered dataset is given in Table 5 and the histogram is shown in Figure 5. Although male and female adult heights seem to have slightly different standard deviations, we will assume they are the same when applying our Bayesian test.

Table 5. Overview of NHANE 2013–2014 adult standing height data

	# of records	% of records	Average height (cm)	SD of height (cm)
Overall	5857	100%	167.09	7.508 (pooled)
Male	2795	48.07%	173.13	7.834
Female	3062	51.9%	159.67	7.194

Table 6. Prior parameters and Bayes factor results for NHANE 2013–2014 adult height data

Prior hyperparameters						Prior bimodal prob.	Bayes factor	Conclusion
ξ_1	ξ_2	σ^2	m_1	m_2	ν			
150	180	400	10	10	8	0.1663	2.24×10^{-5}	Unimodal
					5	0.4175	1.79×10^{-5}	Unimodal
					3	0.6129	7.71×10^{-5}	Unimodal
					1.5	0.8710	2.53×10^{-4}	Unimodal

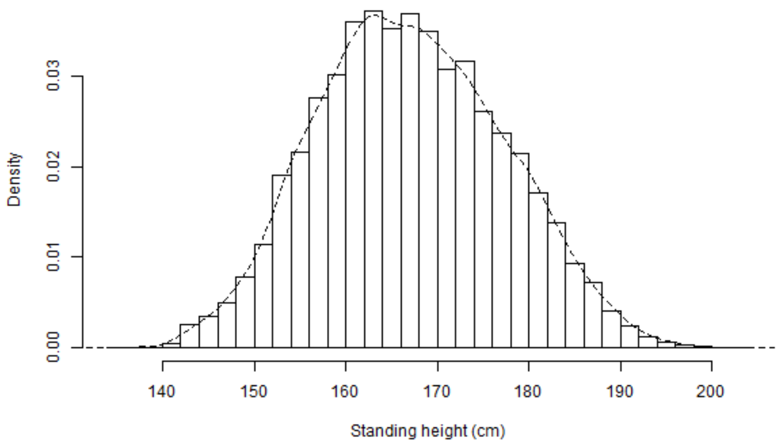


Figure 5. Histogram of NHANE 2013–2014 adult standing height data. The dashed line represents kernel density estimation of the histogram

The data indicates that the sample distribution of adult heights is unimodal when modeled by a Gaussian mixture model $\hat{p}N(\hat{\mu}_1, \hat{\sigma}^2) + (1-\hat{p})N(\hat{\mu}_2, \hat{\sigma}^2)$ where the mixing components are estimated from male and female heights data. The histogram also supports this claim. However, are we going to reach the same conclusion without the information about respondents' gender? To check this, we performed the Bayesian test with $N = 100,000$ Markov chain iterations and a burn-in period of $20,000$ under several different choices of prior distribution parameters. Table 6 shows that the Bayes factor is close to 0 in all cases, strongly suggesting the underlying distribution is unimodal.

7. Concluding Remarks

This paper introduces a Bayesian framework for testing the number of modes in a two-component Gaussian mixture. The first step involves setting up an appropriate prior distribution for the parameters involved and calculating the corresponding posterior. Then, the prior and posterior probabilities are obtained for unimodality by integrating these over the restricted subspace of unimodal parameter space for the mixture distribution. Given the complex structure of this subspace, these are numerically computed by using Monte Carlo and Gibbs Sampling methods. Finally, the conclusion on modality is made based on the Bayes factor. The testing framework is shown to work successfully in a simulation study, and is eventually applied to a data set on adult human heights to investigate whether the combined data on heights for men and women is bimodal. The **R**-code for performing the Bayesian test is available by requesting the authors.

References

[1] Aprausheva, N. N. and Sorokin, S. V. (2019). Recognition of unimodality and bimodality of a two-component gaussian mixture with different variances. *Software and Hardware for Pattern Recognition and Image analysis*, 29(2):252–257.

[2] Basu, S. and Jammalamadaka, S. R. (2002). *Unimodality in Circular Data: A Bayes Test*, volume Advances on Methodological and Applied Aspects of Probability and Statistics, (Ed. Balakrishnan, N. et al), pages 141–153. Taylor & Francis.

[3] Behboodian, J. (1970). On a mixture of normal distributions. *Biometrika*, 57(1):215–217.

[4] Holzmann, H. and Vollmer, S. (2008). A likelihood ratio test for bimodality in two-component mixtures with application to regional income distribution in the eu. *AStA Advances in Statistical Analysis*, 92(1):57–69.

[5] Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of American Statistician Association*, 90(430):773–795.

[6] Schilling, M. F., Watkins, A. E., and Watkins, W. (2002). Is human height bimodal? *The American Statistician*, 56(3):223–229.