

Article

A Posterior p -Value for Homogeneity Testing of the Three-Sample Problem

Yufan Wang¹ and Xingzhong Xu^{2,*}

¹ School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081, China; 3120170684@bit.edu.cn

² School of Mathematical Science, Shenzhen University, Shenzhen 518060, China

* Correspondence: xuxz@szu.edu.cn

Abstract: In this paper, we study a special kind of finite mixture model. The sample drawn from the model consists of three parts. The first two parts are drawn from specified density functions, f_1 and f_2 , while the third one is drawn from the mixture. A problem of interest is whether the two functions, f_1 and f_2 , are the same. To test this hypothesis, we first define the regular location and scale family of distributions and assume that f_1 and f_2 are regular density functions. Then the hypothesis transforms to the equalities of the location and scale parameters, respectively. To utilize the information in the sample, we use Bayes' theorem to obtain the posterior distribution and give the sampling method. We then propose the posterior p -value to test the hypothesis. The simulation studies show that our posterior p -value largely improves the power in both normal and logistic cases and nicely controls the Type-I error. A real halibut dataset is used to illustrate the validity of our method.

Keywords: three-sample problem; homogeneity test; Bernstein-von Mises theorem; posterior p -value

MSC: 62F03; 62F05; 62F15

1. Introduction

In this paper, we focus on the model proposed by Hosmer [1], which is used to study the halibut data. There are two different sources of halibut data. One is from the research cruises, where the sex, age and length of the halibut are available, while another comes from the commercial catch where only age and length can be obtained since the fish have been cleaned before the boats returned to the port. The length distribution of an age class of halibut is closely approximated by a mixture of two normal distributions, which are

$$\begin{aligned} X_i &\stackrel{iid}{\sim} f_1(y), \quad i = 1, \dots, n_1 \\ Y_j &\stackrel{iid}{\sim} f_2(y), \quad j = 1, \dots, n_2 \\ Z_k &\stackrel{iid}{\sim} \lambda f_1(y) + (1 - \lambda) f_2(y), \quad k = 1, \dots, n_3, \end{aligned} \quad (1)$$

where f_1 and f_2 are the probability density functions of the normal distributions and λ is the proportion of the male halibut in the commercial catches. Hosmer [1] estimated the parameters of the two normal distributions using the iterative maximum likelihood estimate method. Murray and Titterington [2] generalized the problem to higher dimensions and summarized a variety of possible techniques, such as maximum likelihood estimation and Bayesian analysis. Anderson [3] proposed a semiparametric modeling assumption known as the exponential tilt mixture model, where the estimating of the proportion is performed by a general method based on direct estimation of the likelihood ratio. This semiparametric model is further studied by Qin [4], who extended Owen's [5] empirical likelihood to the semiparametric model and gave the asymptotic variance formula for the maximum semiparametric likelihood estimation. However, empirical likelihood may suffer from some



Citation: Wang, Y.; Xu, X. A Posterior p -Value for Homogeneity Testing of the Three-Sample Problem. *Mathematics* **2023**, *11*, 3849. <https://doi.org/10.3390/math11183849>

Academic Editor: Alicia Nieto-Reyes

Received: 15 August 2023

Revised: 6 September 2023

Accepted: 6 September 2023

Published: 8 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

computational difficulties. Therefore, Zou et al. [6] proposed the use of partial likelihood and showed that the asymptotic null distribution of the log partial likelihood ratio is chi-square. To estimate the mixing proportion, an EM algorithm is given by Zhang [7]. It is shown that the sequence of proposed EM iterates, irrespective of the starting value, converges to the maximum semiparametric likelihood estimator of the parameters in the mixture model. Furthermore, Inagaki and Komaki [8] and Tan [9] respectively modified the profile likelihood function to provide better estimators for the parameters.

Except for the estimation of parameters, another important issue is to test the homogeneity of the model. Thus, the null hypothesis is

$$H_0 : f_1 = f_2.$$

To test the null hypothesis, the classical results on the likelihood ratio test (LRT) may be invalid. This is caused by the lack of identifiability of some nuisance parameters. To solve this problem, Liang and Rathouz [10] proposed a score test and applied it to genetic linkage analysis. They showed that the score test has a simple asymptotic distribution under the null hypothesis and maintains adequate power in detecting the alternatives. This idea is further generalized by Duan et al. [11] and Fu et al. [12]. On the other hand, Chen et al. [13,14] proposed modified likelihood functions to make the LRT available. They gave the asymptotic theory of the modified LRT and showed that the asymptotic null distribution is a mixture of χ -type distributions and is asymptotically most powerful under local alternatives. Furthermore, Chen and Li [15] designed an EM-test for finite normal mixture models, which performed promisingly in their simulation study. To solve the problem of degeneration of the Fisher information, Li et al. [16] used a high-order expansion to establish a nonstandard convergence rate, $N^{-1/4}$, for the odds ratio parameter estimator. The methods mentioned above have been applied successfully in many real applications. For example, genetic imprinting and quantitative trait locus mapping; see Li et al. [17] and Liu et al. [18].

Most of the mixture models described above mainly consider the case when f_1 and f_2 are normal density functions or have an exponential tilt. In this paper, we want to extend the conclusion to more general cases. A similar question has been researched by Ren et al. [19]. In their paper, a two-block Gibbs sampling method is proposed to obtain the samples of the generalized pivot quantities of the parameters. They studied both cases when f_1 and f_2 are normal and logistic density functions. In our paper, we assume that f_1 and f_2 are in a specified location-scale family with location parameter μ and scale parameter σ . We propose a posterior p -value based on the posterior distribution to test the homogeneity. We aim to give a p -value under the posterior distribution, such that it has the same frequentist properties as the classical p -value. This means that the Bayesian p -value under proper definition can play the same role as the classical one. To sample from the posterior distributions, we propose to use the approximate Bayesian computation (ABC) method for the case when f_1 and f_2 are normal density functions, which is different from the cases when f_1 and f_2 are general. This is because the posterior distribution of the normal case can be regarded as using the information contained in the first two samples as prior distribution and updating it via the third one without loss of information. We find in our simulation that this method is promising and efficient, even though we use the simplest reject-sampling. For the general case, since the ABC method is no longer available, we use Markov Chain Monte Carlo (MCMC) methods, such as the Metropolis–Hastings sampling method proposed by Hannig et al. [20] and the two-block Gibbs sampling proposed by Ren et al. [19] to sample from the posterior distribution.

The paper is organized as follow. In Section 2, we first define the regular location-scale family and give some properties of the family. We then propose our posterior p -value for testing the homogeneity. We further introduce the sampling method for different cases. Real data of the halibut is studied in Section 3 to illustrate the validity of our method. The simulation study is given in Section 4, while the conclusion is given in Section 5.

2. Test Procedure

In this section, we consider model (1), where the distributions are in a certain regular location-scale family. Thus, we first give the definition in the following subsection.

2.1. Regular Location-Scale Family

In this section, we first give the definition of the regular location-scale family.

Definition 1 (regular location-scale family). *Let $f(x)$ be a probability density function. If $f(x)$ satisfies*

- (1) $f(x) > 0, -\infty < x < \infty$;
- (2) $f''(x)$ is continuous;
- (3) $\lim_{x \rightarrow -\infty} x^2 f'(x) = \lim_{x \rightarrow +\infty} x^2 f'(x) = 0$;
- (4) $\int_{-\infty}^{+\infty} x^2 \frac{[f'(x)]^2}{f(x)} dx < \infty$.

Then $f(x)$ is defined as a regular density function, and

$$\mathcal{R}_f = \left\{ \frac{1}{\sigma} f\left(\frac{x - \mu}{\sigma}\right); \mu \in (-\infty, +\infty), \sigma \in (0, +\infty) \right\}$$

is defined as the regular location-scale family.

It is easy to verify that many families of distributions are regular location-scale families. For example, let

$$f_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad f_2(x) = \frac{e^{-x}}{(1 + e^{-x})^2}.$$

Then $f_1(x)$ and $f_2(x)$ are regular density functions. The families of distributions that are constructed by $f_1(x)$ and $f_2(x)$ are regular, and they are the families of normal distributions and logistic distributions, respectively. The two families of distributions are included later in the paper.

The following lemma highlights some properties of this family.

Lemma 1. *If $f(x)$ is a regular density function, then we have*

- (1) $\lim_{x \rightarrow -\infty} x f(x) = \lim_{x \rightarrow +\infty} x f(x) = 0$;
- (2) $\int_{-\infty}^{+\infty} f'(x) dx = 0$;
- (3) $\int_{-\infty}^{+\infty} x f'(x) dx = -1$;
- (4) $\int_{-\infty}^{+\infty} f''(x) dx = 0$;
- (5) $\int_{-\infty}^{+\infty} x f''(x) dx = 0$;
- (6) $\int_{-\infty}^{+\infty} x^2 f''(x) dx = 0$.

The proof of this lemma is given in Appendix A.

We further calculate the Fisher information matrix of the regular location-scale family with the following proposition.

Proposition 1. *Assume that $f(x; \xi) = \frac{1}{\sigma} f\left(\frac{x - \mu}{\sigma}\right)$ is in the regular location-scale family, where $\xi = (\mu, \sigma)^\top$. The parameter space is $\Omega = \{(\mu, \sigma) : -\infty < \mu < \infty, \sigma > 0\}$. Let $l(X; \xi)$ be $\log f(X; \xi)$. Then*

- (1) *The score function satisfies*

$$E_{\xi} \left[\frac{\partial l(X; \xi)}{\partial \xi} \right] = 0,$$

where

$$\frac{\partial l(X; \xi)}{\partial \xi} = \left(\frac{\partial l(X; \xi)}{\partial \mu}, \frac{\partial l(X; \xi)}{\partial \sigma} \right)^\top$$

is a two-dimensional vector. E_{ξ} denotes the expectation under the distribution of parameters $\xi = (\mu, \sigma)^\top$.

(2) The Fisher information matrix satisfies

$$0 < I_f(\xi) = E_{\xi} \left[\left(\frac{\partial l(X; \xi)}{\partial \xi} \right) \left(\frac{\partial l(X; \xi)}{\partial \xi} \right)^\top \right] = \frac{1}{\sigma^2} \mathbf{C}(f) = \frac{1}{\sigma^2} \begin{pmatrix} C_{11}(f) & C_{12}(f) \\ C_{21}(f) & C_{22}(f) \end{pmatrix} < \infty,$$

where

$$C_{11}(f) = \int_{-\infty}^{\infty} \frac{[f'(y)]^2}{f(y)} dy$$

$$C_{12}(f) = C_{21}(f) = \int_{-\infty}^{\infty} y \frac{[f'(y)]^2}{f(y)} dy$$

$$C_{22}(f) = \int_{-\infty}^{\infty} \left(1 + y \frac{f'(y)}{f(y)} \right)^2 f(y) dy = \int_{-\infty}^{\infty} y^2 \frac{[f'(y)]^2}{f(y)} dy - 1$$

(3) The Fisher information matrix is given by

$$I_f(\xi) = -E_{\xi} \left[\frac{\partial^2 f(X, \xi)}{\partial \xi \partial \xi^\top} \right].$$

The proof is given in Appendix A.

Proposition 2. Assume that $0 < \lambda_0 < 1$ and $f(\cdot)$ is regular. Then $\{g(x; \theta) : \theta \in \Omega\}$ given by

$$g(x, \theta) = \frac{\lambda_0}{\sigma_1} f\left(\frac{x - \mu_1}{\sigma_1}\right) + \frac{1 - \lambda_0}{\sigma_2} f\left(\frac{x - \mu_2}{\sigma_2}\right), \tag{2}$$

where $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2)^\top$ and $\Omega = R^2 \times R^{+2}$ has the following properties.

(1)

$$E_{\theta} \left[\frac{\partial \log g(x, \theta)}{\partial \theta} \right] = 0;$$

(2)

$$I(\theta) = E \left[\left(\frac{\partial \log g(x, \theta)}{\partial \theta} \right) \left(\frac{\partial \log g(x, \theta)}{\partial \theta} \right)^\top \right] < \infty;$$

(3)

$$I(\theta) = -E_{\theta} \left[\frac{\partial^2 \log g(x; \theta)}{\partial \theta \partial \theta^\top} \right].$$

The proof is given in Appendix A.

We then give the Fisher information matrix of the normal and logistic distribution. For the normal distribution, we have

$$C_{11}(f) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} y^2 dy = 1$$

$$C_{12}(f) = C_{21}(f) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} y^3 dy = 0$$

$$C_{22}(f) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} y^4 dy - 1 = 2$$

Thus, the Fisher information matrix of normal distribution is

$$I_f(\xi) = \frac{1}{\sigma^2} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

Similarly, for the logistic distribution,

$$C_{11}(f) = \int_{-\infty}^{\infty} \frac{e^{-y}(e^{-y} - 1)^2}{(1 + e^{-y})^4} dy = 1/3$$

$$C_{12}(f) = C_{21}(f) = \int_{-\infty}^{\infty} \frac{e^{-y}(e^{-y} - 1)^2}{(1 + e^{-y})^4} dy = 0$$

$$C_{22}(f) = \int_{-\infty}^{\infty} \frac{e^{-y}(e^{-y} - 1)^2}{(1 + e^{-y})^4} dy = \frac{1}{3} + \frac{\pi^2}{9}$$

Thus, the Fisher information matrix of logistic distribution is

$$I_f(\xi) = \frac{1}{\sigma^2} \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{3} + \frac{\pi^2}{9} \end{pmatrix}$$

2.2. A Posterior p-Value

We now consider testing the homogeneity of model (1), where f_1 and f_2 are in \mathcal{R}_f ,

$$f_1 = \frac{1}{\sigma_1} f\left(\frac{x - \mu_1}{\sigma_1}\right), \quad f_2 = \frac{1}{\sigma_2} f\left(\frac{x - \mu_2}{\sigma_2}\right).$$

This is equivalent to testing the equality of the parameters of the two density functions, that is,

$$H_0 : \mu_1 = \mu_2, \sigma_1 = \sigma_2 \quad v.s. \quad H_1 : \mu_1 \neq \mu_2 \text{ or } \sigma_1 \neq \sigma_2. \tag{3}$$

Consider the density function

$$g(x; \theta) = \frac{\lambda}{\sigma_1} f\left(\frac{x - \mu_1}{\sigma_1}\right) + \frac{\lambda}{\sigma_2} f\left(\frac{x - \mu_2}{\sigma_2}\right),$$

where $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, \lambda)^\top$ is the unknown parameter. When $f(x)$ is the regular density function, then the Fisher information matrix is

$$I_g(\theta) = E_\theta \left[\frac{\partial \log g(x; \theta)}{\partial \theta} \right] \left[\frac{\partial \log g(x; \theta)}{\partial \theta} \right]^\top,$$

where

$$\frac{\partial \log g(x; \theta)}{\partial \theta} = \begin{pmatrix} \frac{1}{g(x; \theta)} \left[-\frac{\lambda}{\sigma_1^2} f' \left(\frac{x - \mu_1}{\sigma_1} \right) \right] \\ \frac{1}{g(x; \theta)} \left[-\frac{1 - \lambda}{\sigma_2^2} f' \left(\frac{x - \mu_2}{\sigma_2} \right) \right] \\ \frac{1}{g(x; \theta)} \left[-\frac{\lambda}{\sigma_1^2} f' \left(\frac{x - \mu_1}{\sigma_1} \right) - \frac{\lambda}{\sigma_1^3} f' \left(\frac{x - \mu_1}{\sigma_1} \right) \right] \\ \frac{1}{g(x; \theta)} \left[-\frac{1 - \lambda}{\sigma_2^2} f' \left(\frac{x - \mu_2}{\sigma_2} \right) - \frac{1 - \lambda}{\sigma_2^3} f' \left(\frac{x - \mu_2}{\sigma_2} \right) \right] \\ \frac{1}{g(x; \theta)} \left[\frac{1}{\sigma_1} f \left(\frac{x - \mu_1}{\sigma_1} \right) - \frac{1}{\sigma_2} f \left(\frac{x - \mu_2}{\sigma_2} \right) \right] \end{pmatrix}$$

When $\mu_1 = \mu_2, \sigma_1 = \sigma_2$,

$$\frac{\partial \log g(x; \theta)}{\partial \lambda} = 0,$$

the last row and column of $I_g(\theta)$ is zero, which means that $|I_g(\theta)| = 0$ and is non-definite. Thus, we may encounter some difficulties when using some traditional test methods, such as the likelihood ratio test.

We suggest a solution here. First we assume that $\lambda = \lambda_0$ is known. There are then four parameters, and we still denote them by $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2)^\top$. We use the estimate of λ instead since λ is unknown. This is because when the homogeneity hypothesis holds, the distribution of the population is irrelative to λ , so the level of the test is irrelative to the

estimate of λ . We then give the inference on θ below. For the first two samples, the fiducial density of (μ_1, σ_1) and (μ_2, σ_2) are

$$(\mu_1, \sigma_1) \propto \left[\prod_{i=1}^{n_1} \frac{1}{\sigma_1} f\left(\frac{x_{1i} - \mu_1}{\sigma_1}\right) \right] \frac{1}{\sigma_1}, (\mu_2, \sigma_2) \propto \left[\prod_{j=1}^{n_2} \frac{1}{\sigma_2} f\left(\frac{x_{2j} - \mu_2}{\sigma_2}\right) \right] \frac{1}{\sigma_2}, \tag{4}$$

where “ \propto ” denotes “proportion to”; see example 3 of Hannig et al. [20]. To combine (4) with the third sample, we regard (4) as the prior distribution. By the Bayes’ theorem

$$\theta \propto \prod_{k=1}^{n_3} \left[\frac{\lambda_0}{\sigma_1} f\left(\frac{x_{3k} - \mu_1}{\sigma_1}\right) + \frac{1 - \lambda_0}{\sigma_2} f\left(\frac{x_{3k} - \mu_2}{\sigma_2}\right) \right] \cdot \left[\prod_{i=1}^{n_1} \frac{1}{\sigma_1} f\left(\frac{x_{1i} - \mu_1}{\sigma_1}\right) \right] \frac{1}{\sigma_1} \cdot \left[\prod_{j=1}^{n_2} \frac{1}{\sigma_2} f\left(\frac{x_{2j} - \mu_2}{\sigma_2}\right) \right] \frac{1}{\sigma_2} \tag{5}$$

Denote the probability measure on the parameter space determined by (5) by $P^{\Theta|x}$, where $x = (x_1^\top, x_2^\top, x_3^\top)$, $x_1 = (x_{11}, x_{12}, \dots, x_{1n_1})^\top$, $x_2 = (x_{21}, x_{22}, \dots, x_{2n_2})^\top$, $x_3 = (x_{31}, x_{32}, \dots, x_{3n_3})^\top$. Θ denotes the random variable. We can see from expression (5) that $P^{\Theta|x}$ is the posterior distribution under the prior distribution

$$d\theta = \frac{1}{\sigma_1 \sigma_2} d\mu_1 d\mu_2 d\sigma_1 d\sigma_2.$$

Let

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \quad \mathbf{b} = [0, 0]^\top.$$

Then, hypotheses (3) is equivalent to

$$H_0 : \mathbf{A}\theta = \mathbf{b} \quad v.s. \quad H_1 : \mathbf{A}\theta \neq \mathbf{b}. \tag{6}$$

where $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2)^\top$.

To establish Bernstein-von Mises theorem for multiple samples, we first introduce some necessary assumptions below. Let $l_i(\theta)$ be the log-likelihood function of the i th sample, where $i = 1, 2, 3$.

Assumption 1. Given any $\varepsilon > 0$, there exists $\delta > 0$, such that in the expansion

$$l_i(\theta) = l_i(\theta_0) + (\theta - \theta_0)^\top l'_i(\theta_0) - \frac{1}{2}(\theta - \theta_0)^\top [nI_i(\theta_0) + R_{n_i}(\theta)](\theta - \theta_0), \quad i = 1, 2, 3,$$

where θ_0 is the true value of the parameter. $I_i(\theta_0)$ is the Fisher Information matrix. The probability of the following event

$$\sup \left\{ \frac{1}{n} \lambda_{\max}[R_{n_i}(\theta)] : \|\theta - \theta_0\| \leq \delta \right\} \geq \varepsilon$$

tends to 0 as $n \rightarrow \infty$, where $\|\cdot\|$ is the Euclidean norm and $\lambda_{\max}(\mathbf{A})$ denotes the largest absolute eigenvalues of a square matrix, \mathbf{A} .

Assumption 2. For any $\delta > 0$, there exists $\varepsilon > 0$, such that the probability of the event

$$\sup \left\{ \frac{1}{n_i} [l_i(\theta) - l_i(\theta_0)] : \|\theta - \theta_0\| \geq \delta \right\} \leq -\varepsilon$$

tends to 1 as $n \rightarrow \infty$.

Assumption 3. Under the prior π , there exist k_0 , such that the integral of $\|\theta\|$ below exists,

$$\int \|\theta\|^2 \prod_{i=1}^{k_0} \left[\frac{1}{\sigma_1} f\left(\frac{x_{1i} - \mu_1}{\sigma_1}\right) \right] \prod_{j=1}^{k_0} \left[\frac{1}{\sigma_2} f\left(\frac{x_{2j} - \mu_2}{\sigma_2}\right) \right] \pi(\theta) d\theta < \infty.$$

Assumption 4. When $n = n_1 + n_2 + n_3 \rightarrow \infty$,

$$\frac{n_i}{n} \rightarrow r_i \in (0, 1), \quad i = 1, 2, 3.$$

We then give the Bernstein-von Mises theorem for multiple samples as follows.

Theorem 1. Denote the posterior density of $t = \sqrt{n}(\theta - T_n)$ by $\pi^*(t|x)$, where

$$T_n = \theta_0 + \frac{1}{n} I^{-1}(\theta_0) l'(\theta_0).$$

If Assumptions 1, 2 and 4 hold, then

$$\int_{\Omega} \left| \pi^*(t|x) - (2\pi)^{-\frac{k}{2}} |I(\theta_0)|^{\frac{1}{2}} e^{-\frac{1}{2} t^T I(\theta_0) t} \right| dt \xrightarrow{P} 0.$$

Furthermore, if Assumption 3 holds, then

$$\int_{\Omega} \left(1 + \|t\|^2 \right) \left| \pi^*(t|x) - (2\pi)^{-\frac{k}{2}} |I(\theta_0)|^{\frac{1}{2}} e^{-\frac{1}{2} t^T I(\theta_0) t} \right| dt \xrightarrow{P} 0.$$

We can then define the posterior p -value as follows

Definition 2. Let

$$\begin{aligned} p(x) &= P^{\Theta|x} \left((\Theta - \theta_B)^T A^T (A \Sigma_B A^T)^{-1} A (\Theta - \theta_B) \right. \\ &\quad \left. \geq (\mathbf{b} - A \theta_B)^T (A \Sigma_B A^T)^{-1} (\mathbf{b} - A \theta_B) \right), \end{aligned} \tag{7}$$

where $P^{\Theta|x}(\cdot)$ is the probability under the posterior distribution. θ_B is the posterior mean and Σ_B is the posterior covariance matrix. We call $p(x)$ a posterior p -value.

It should be noted that $p(x)$ is defined under the posterior distribution, which is the distribution of parameters given the observation $X = x$. However, when studying the properties of $p(x)$, we regard it as a random variable and denote it by $p(X)$. The theorem below guarantees the validity of the posterior p -value.

Theorem 2. Under the assumption of Theorem 1, if the null hypothesis in (3) is true, that is, $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$, then the p -value defined by (7) satisfies

$$p(X) \xrightarrow{d} U(0, 1).$$

where “ \xrightarrow{d} ” is the convergence in distribution and $U(0, 1)$ is the uniform distribution on the interval $(0, 1)$.

The proof is given in the Appendix A. For a given significance level, α , we may reject the null hypothesis if the p -value is less than α .

2.3. Sampling Method

The posterior mean, θ_B , and the posterior variance, Σ_B , in Equation (7) can be estimated by the sample mean and variance, respectively. Now the remain problem is how to sample

from the posterior distribution. When λ is unknown, we first propose an EM algorithm to estimate λ , then we sample from the posterior distribution where λ is fixed to the estimate of λ . The Markov Chain Monte Carlo (MCMC) method is commonly used. However, as we have mentioned earlier, the MCMC method needs to discard a large number of samples in the burn-in period to guarantee the samples are accepted sufficiently close to the ones from the real distribution. Fortunately, when f_1 and f_2 are normal density functions, we find that the posterior distribution can be transformed and sampled by using the approximate Bayesian computation (ABC) method. However, when f_1 and f_2 are more general, such as the logistic density functions, the two-block Gibbs sampling proposed by Ren et al. [19] can be an appropriate substitution. We will discuss the details in the following subsection.

2.3.1. EM Algorithm for λ

In this subsection, we propose the EM algorithm for estimating λ . The log-likelihood function of the model is

$$L(x; \theta, \lambda) = \sum_{i=1}^{n_1} \log[f_1(x_{1i}; \theta)] + \sum_{j=1}^{n_2} \log[f_2(x_{2j}; \theta)] + \sum_{k=1}^{n_3} \log[p(x_{3k}; \theta, \lambda)],$$

where f_1 and f_2 are in the same regular location-scale family, \mathcal{R}_f , with parameters (μ_1, σ_1) and (μ_2, σ_2) , respectively. In the log-likelihood function of the third sample, $p(x_{3k}; \theta, \lambda)$ is

$$p(x_{3k}; \theta, \lambda) = \lambda f_1(x_{3k}; \theta) + (1 - \lambda) f_2(x_{3k}; \theta).$$

The EM algorithm was first proposed by Dempster et al. [21] and broadly applied to a wide variety of parametric models; see McLachlan and Krishnan [22] for a better review.

Assume that we have obtained the estimate of the parameters after m times of iterative, denote them by $\theta^{(m)} = (\mu_1^{(m)}, \sigma_1^{(m)}, \mu_2^{(m)}, \sigma_2^{(m)})^\top$ and $\lambda^{(m)}$. We introduce the latent variable $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{n_3})^\top$; the component γ_k indicates which distribution the sample x_{3k} is drawn from. $\gamma_k = 1$ when it is drawn from the first distribution $f_1(x_{3k}; \theta)$, otherwise, $\gamma_k = 0$. We then have

$$P(\gamma_k = 1) = \lambda, \quad P(\gamma_k = 0) = 1 - \lambda, \quad k = 1, 2, \dots, n_3$$

The density of the joint distribution of X_1, X_2, X_3, γ is

$$\prod_{i=1}^{n_1} f_1(x_{1i}; \theta) \prod_{j=1}^{n_2} f_2(x_{2j}; \theta) \prod_{k=1}^{n_3} [\lambda f_1(x_{3k}; \theta)]^{\gamma_k} [(1 - \lambda) f_2(x_{3k}; \theta)]^{1 - \gamma_k}.$$

Given $X_1 = x_1, X_2 = x_2, X_3 = x_3$, the conditional distribution of γ_k is

$$\left[\frac{\lambda f_1(x_{3k}; \theta)}{\lambda f_1(x_{3k}; \theta) + (1 - \lambda) f_2(x_{3k}; \theta)} \right]^{\gamma_k} \left[\frac{(1 - \lambda) f_2(x_{3k}; \theta)}{\lambda f_1(x_{3k}; \theta) + (1 - \lambda) f_2(x_{3k}; \theta)} \right]^{1 - \gamma_k},$$

where $\gamma_k = 0, 1, k = 1, 2, \dots, n_3$. Thus, the conditional expectation of γ_k is

$$E_{(\theta, \lambda)} \gamma_k = P_{(\theta, \lambda)}(\gamma_k = 1) = \frac{\lambda f_1(x_{3k}; \theta)}{\lambda f_1(x_{3k}; \theta) + (1 - \lambda) f_2(x_{3k}; \theta)}.$$

When $\theta = \theta^{(m)}$ and $\lambda = \lambda^{(m)}$, the conditional expectation of γ_k can be the estimate of γ_k .

$$\hat{\gamma}_k(\theta^{(m)}, \lambda^{(m)}) = \frac{\lambda^{(m)} f_1(x_{3k}; \theta^{(m)})}{\lambda^{(m)} f_1(x_{3k}; \theta^{(m)}) + (1 - \lambda^{(m)}) f_2(x_{3k}; \theta^{(m)})},$$

$k = 1, 2, \dots, n_3$.

The log likelihood function is

$$\sum_{i=1}^{n_1} \log f_1(x_{1i}; \theta) + \sum_{j=1}^{n_2} \log f_2(x_{2j}; \theta) + \sum_{k=1}^{n_3} \gamma_k \log f_1(x_{3k}; \theta,)$$

$$+ \sum_{k=1}^{n_3} (1 - \gamma_k) \log f_2(x_{3k}; \theta,) + \left(\sum_{k=1}^{n_3} \gamma_k \right) \log \lambda + \left(n_3 - \sum_{k=1}^{n_3} \gamma_k \right) \log(1 - \lambda).$$

Since the latent variable is unknown, we use its conditional expectation. Furthermore, the MLE of λ is

$$\lambda^{(m+1)} = \frac{\sum_{k=1}^{n_3} \hat{\gamma}_k(\theta^{(m)}, \lambda^{(m)})}{n_3}.$$

Then, in the E-step, we calculate the expectation of new parameters conditional on $(\theta^{(m)}, \lambda^{(m)})$,

$$Q(\theta, \lambda | \theta^{(m)}, \lambda^{(m)}) = \sum_{i=1}^{n_1} \log f_1(x_{1i}; \theta) + \sum_{j=1}^{n_2} \log f_2(x_{2j}; \theta)$$

$$+ \sum_{k=1}^{n_3} \hat{\gamma}_k((\theta^{(m)}, \lambda^{(m)})) \log f_1(x_{3k}; \theta,)$$

$$+ \sum_{k=1}^{n_3} (1 - \hat{\gamma}_k((\theta^{(m)}, \lambda^{(m)}))) \log f_2(x_{3k}; \theta,).$$

Let $\hat{\gamma}_k = \hat{\gamma}_k(\theta^{(m)}, \lambda^{(m)})$, then

$$Q(\theta, \lambda | \theta^{(m)}, \lambda^{(m)}) = \sum_{i=1}^{n_1} \left[-\log \sigma_1 + \log f \left(\frac{x_{1i} - \mu_1}{\sigma_1} \right) \right]$$

$$+ \sum_{j=1}^{n_2} \left[-\log \sigma_2 + \log f \left(\frac{x_{2j} - \mu_2}{\sigma_2} \right) \right]$$

$$+ \sum_{k=1}^{n_3} \hat{\gamma}_k \left[-\log \sigma_1 + \log f \left(\frac{x_{3k} - \mu_1}{\sigma_1} \right) \right]$$

$$+ \sum_{k=1}^{n_3} (1 - \hat{\gamma}_k) \left[-\log \sigma_2 + \log f \left(\frac{x_{3k} - \mu_2}{\sigma_2} \right) \right].$$

In the M-step, we compute the simultaneous equations below to maximize $Q(\theta, \lambda | \theta^{(m)}, \lambda^{(m)})$. The solutions are the new parameters $(\theta^{(m+1)}, \lambda^{(m+1)})$. We give the equations of (μ_1, σ_1) ; similarly, we can obtain (μ_1, σ_1) .

$$\begin{cases} 0 = \sum_{k=1}^{n_3} \frac{\gamma_k}{\lambda f_1 + (1-\lambda) f_2} f' \left(\frac{x_{3k} - \mu_1}{\sigma_1} \right) + \sum_{i=1}^{n_1} \left[\frac{1}{f_1} f' \left(\frac{x_{1i} - \mu_1}{\sigma_1} \right) \right] \\ 0 = \sum_{k=1}^{n_3} \frac{\gamma_k}{\lambda f_1 + (1-\lambda) f_2} \left[\sigma_1^2 f_1 + f' \left(\frac{x_{3k} - \mu_1}{\sigma_1} \right) (x_{3k} - \mu_1) \right] + \sum_{i=1}^{n_1} \left[\sigma_1^2 + (x_{1i} - \mu_1) f' \left(\frac{x_{1i} - \mu_1}{\sigma_1} \right) \frac{1}{f_1} \right] \\ \lambda = \frac{1}{n_3} \sum_{k=1}^{n_3} \gamma_k \end{cases}$$

In the simulation study, we consider the normal and logistic cases. The maximization step of the normal case can be simplified as

$$\mu_1 = \frac{\sum_{k=1}^{n_3} \gamma_k x_{3k} + \sum_{i=1}^{n_1} x_{1i}}{\sum_{k=1}^{n_3} \gamma_k + n_1},$$

$$\sigma_1^2 = \frac{\sum_{k=1}^{n_3} \gamma_k (x_{3k} - \mu_1)^2 + \sum_{i=1}^{n_1} (x_{1i} - \mu_1)^2}{\sum_{k=1}^{n_3} \gamma_k + n_1},$$

while that of the logistic case is

$$\begin{cases} 0 = \sum_{k=1}^{n_3} \gamma_k \left(\frac{e^{\frac{x_{3k}-\mu_1}{\sigma_1}} - 1}{e^{\frac{x_{3k}-\mu_1}{\sigma_1}} + 1} \right) + \sum_{i=1}^{n_1} \frac{e^{\frac{x_{1i}-\mu_1}{\sigma_1}} - 1}{e^{\frac{x_{1i}-\mu_1}{\sigma_1}} + 1} \\ 0 = \sum_{k=1}^{n_3} \left[\gamma_k \left(\frac{x_{3k}-\mu_1}{\sigma_1} \frac{e^{\frac{x_{3k}-\mu_1}{\sigma_1}} - 1}{e^{\frac{x_{3k}-\mu_1}{\sigma_1}} + 1} - 1 \right) \right] + \sum_{i=1}^{n_1} \left[\frac{x_{1i}-\mu_1}{\sigma_1} \left(\frac{e^{\frac{x_{1i}-\mu_1}{\sigma_1}} - 1}{e^{\frac{x_{1i}-\mu_1}{\sigma_1}} + 1} \right) - 1 \right]. \end{cases}$$

The two steps are repeated sufficiently to guarantee the convergence. We can then obtain the MLE of the parameters.

2.3.2. Normal Case

When the estimate of λ is obtained, the posterior distribution (5) can be rewritten as

$$\begin{aligned} \pi(\theta|X_1, X_2, X_3) &= \left(\frac{1}{\sigma_1} \prod_{i=1}^{n_1} f(X_{1i}; \mu_1, \sigma_1) \right) \times \left(\frac{1}{\sigma_2} \prod_{j=1}^{n_2} f(X_{2j}; \mu_2, \sigma_2) \right) \\ &\times \prod_{k=1}^{n_3} \left[\hat{\lambda} f(X_{3k}; \mu_1, \sigma_1) + (1 - \hat{\lambda}) f(X_{3k}; \mu_2, \sigma_2) \right]. \end{aligned} \tag{8}$$

This means that the posterior distribution is equivalent to using the first two terms on the right side of the equation as the “prior distribution” and the third term as the likelihood function. For the first term, we have

$$\frac{1}{\sigma_1} \prod_{i=1}^{n_1} f(X_i; \mu_1, \sigma_1) = \frac{1}{\sigma_1} \left(\frac{1}{\sqrt{2\pi}\sigma_1} \right)^n \exp \left[- \frac{\sum_{i=1}^{n_1} (X_i - \mu_1)^2}{2\sigma_1^2} \right].$$

By denoting the sample mean and variance by \bar{X} and S_1^2 , respectively, we have

$$\bar{X}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} X_{1i}, \quad S_1^2 = \frac{1}{n-1} \sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^2,$$

which follows a normal and $\chi^2(n_1 - 1)$ distribution, respectively; that is,

$$\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right), \quad \frac{(n_1 - 1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1 - 1).$$

Let $U \sim N(0, 1)$ and $V \sim \chi^2(n_1 - 1)$ be two independent random variables. Then

$$\bar{X}_1 = \mu_1 + \frac{\sigma_1}{\sqrt{n_1}}U, \quad (n_1 - 1)S_1^2 = \sigma_1^2V.$$

Given $\bar{X}_1 = \bar{x}_1$ and $S_1^2 = s_1^2$, then μ_1 and σ_1 can be regarded as the functions of U and V

$$\mu_1 = \bar{x}_1 - \frac{\sigma_1}{\sqrt{n_1}}U, \quad \sigma_1^2 = \frac{(n_1 - 1)s_1^2}{V}.$$

The joint distribution of (U, V) is

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \frac{v^{\frac{n_1-1}{2}-1}}{\Gamma(\frac{n_1-1}{2}) 2^{\frac{n_1-1}{2}}} e^{-\frac{v}{2}}.$$

Then the joint distribution of (μ_1, σ_1) can be calculated as

$$\pi(\mu_1, \sigma_1 | x_{obs}) = \frac{\sqrt{n_1}}{\sqrt{2\pi}\sigma_1} e^{-\frac{n_1(\bar{x}_1 - \mu_1)^2}{2\sigma_1^2}} \frac{(s_1^2)^{\frac{n_1-1}{2}} (n_1 - 1)^{\frac{n_1-1}{2}}}{\Gamma(\frac{n_1-1}{2}) 2^{\frac{n_1-1}{2}}} \left(\frac{1}{\sigma_1^2}\right)^{\frac{n_1-1}{2} - 1 + \frac{3}{2}} e^{-\frac{(n_1-1)s_1^2}{2\sigma_1^2}}, \tag{9}$$

where $x_{1obs} = (x_1, x_2, \dots, x_{n_1})^\top$. This coincides with the joint fiducial density proposed by Fisher [23,23], which means that the fiducial distribution of (μ_1, σ_1) is

$$\mu_1 | \sigma_1^2 \sim N\left(\bar{x}_1, \frac{\sigma_1^2}{n_1}\right), \quad \frac{1}{\sigma_1^2} \sim \frac{\chi^2(n_1 - 1)}{(n_1 - 1)s_1^2}. \tag{10}$$

Similarly, can we obtain

$$\mu_2 | \sigma_2^2 \sim N\left(\bar{x}_2, \frac{\sigma_2^2}{n_2}\right), \quad \frac{1}{\sigma_2^2} \sim \frac{\chi^2(n_2 - 1)}{(n_2 - 1)s_2^2}, \tag{11}$$

where \bar{x}_2 and s_2^2 are the sample mean and variance of the second sample and $x_{2obs} = (x_{21}, x_{22}, \dots, x_{2n_2})^\top$.

With the conclusion above, sampling from the posterior distribution (5) can be conducted by sampling first from the fiducial distribution of the parameters and then combine the information with the likelihood function of the third sample from the mixture model (1). This can be carried out simply using the approximate Bayesian computation (ABC) method. In this case, we regard the fiducial distributions of $(\mu_1, \sigma_1, \mu_2, \sigma_2)^\top$ as the prior distribution. After we have drawn samples of parameters from (10) and (11), denoted by $(\mu_1^*, \sigma_1^*, \mu_2^*, \sigma_2^*)^\top$, we generate simulations from the model below and denote them by $x_{3sim} = (x_{31}, x_{32}, \dots, x_{3n_3})^\top$,

$$\hat{\lambda}f(x; \mu_1^*, \sigma_1^*) + (1 - \hat{\lambda})f(x; \mu_2^*, \sigma_2^*)$$

where $\hat{\lambda}$ is the MLE of λ , estimated beforehand using the EM algorithm proposed in the last subsection. We then calculate the distance between the simulations and the observation and accept those whose distance is below a given threshold, ε . The algorithm is given below.

1. Compute the sample mean and variance of the first two samples and denote them by \bar{x} , s_1^2 , \bar{y} and s_2^2 . Calculate the MLE of λ using the EM algorithm and denote it by $\hat{\lambda}$.
2. Sample U_1 and U_2 from the standard normal distribution, V_1 from the $\chi^2(n_1 - 1)$ distribution and V_2 from $\chi^2(n_2 - 1)$, respectively. To sample from the fiducial distributions of the parameters, we calculate μ_1 , σ_1 , μ_2 and σ_2 using

$$\begin{aligned} \mu_1 &= \bar{x}_1 - \frac{U_1}{V_1/\sqrt{n_1-1}} \frac{s_1}{\sqrt{n_1}}, & \sigma_1^2 &= \frac{(n_1 - 1)s_1^2}{V_1^2}, \\ \mu_2 &= \bar{x}_2 - \frac{U_2}{V_2/\sqrt{n_2-1}} \frac{s_2}{\sqrt{n_2}}, & \sigma_2^2 &= \frac{(n_2 - 1)s_2^2}{V_2^2}. \end{aligned}$$

3. We denote the samples of the parameters by $\theta^* = (\mu_1^*, \mu_2^*, \sigma_1^*, \sigma_2^*)^\top$. Generate a simulation of size n_3 from

$$\hat{\lambda}f(x; \mu_1^*, \sigma_1^*) + (1 - \hat{\lambda})f(x; \mu_2^*, \sigma_2^*).$$

The simulation is represented by $x_{3sim} = (x_{31}^*, x_{32}^*, \dots, x_{3n_3}^*)^\top$.

4. Calculate the Euclidean distance between the order statistics of the observation Z_1, Z_2, \dots, Z_{n_3} and the simulation $z_1^*, z_2^*, \dots, z_{n_3}^*$. We accept the parameters if the distance is below a given threshold, ε . Otherwise, we reject the parameters.
5. The procedure is repeated until we accept a certain number of parameters.

A remark that should be noted in this algorithm is that the samples we receive are an approximation of the posterior distribution (5). We actually receive samples from

$$\pi(\theta, \varepsilon | x_{1obs}, x_{2obs}, x) \propto \pi(\theta | x_{1obs}, x_{2obs}, x_{3obs}) \mathbb{I}(\|x - x_{3obs}\| \leq \varepsilon), \tag{12}$$

where \mathbb{I} is the indicator function. ε controls the proximity of (12) to (5) and can be adjusted to balance the accuracy and computational cost.

2.3.3. General Case

When f_1 and f_2 are not normal, to sample from the posterior (8), it is natural to use the Markov chain Monte Carlo (MCMC) method. The Metropolis–Hastings (MH) sampling method and Gibbs sampling method are commonly used. An early version of the MH algorithm was given by Metropolis et al. [24] in a statistical physics context, with subsequent generalization by Hastings [25], who focused on statistical problems. Some computational problem and solutions can be further seen in Owen and Glynn [26].

The initial values of the parameters can be determined by the EM algorithm mentioned above. For the proposal distribution, we choose

$$\begin{aligned} q(\mu_k | \mu_k^{(\tau)}) &= N(\cdot; \mu_k^{(\tau)}, 1) \\ q(\sigma_k | \sigma_k^{(\tau)}) &= Ga(\cdot; \sigma_k^{(\tau)}, 1) \end{aligned} \tag{13}$$

where $Ga(\cdot)$ and $N(\cdot)$ denote the gamma distribution and normal distribution, respectively. $k = 1, 2$ and $\mu_k^{(\tau)}, \sigma_k^{(\tau)}$ denotes the parameters accepted in the τ th loop. After we obtain $\theta^{(\tau)} = (\mu_1^{(\tau)}, \sigma_1^{(\tau)}, \mu_2^{(\tau)}, \sigma_2^{(\tau)})^\top$, we can further obtain $\theta^{(\tau+1)}$ via the following two-step algorithm.

1. Sample $(\mu_1^*, \sigma_1^*, \mu_2^*, \sigma_2^*)^\top$ respectively from the proposal distribution (13). Compute

$$\begin{aligned} \log Q(\theta^* | \theta^\tau) &= \log q(\mu_1^* | \mu_1^{(\tau)}) + \log q(\mu_2^* | \mu_2^{(\tau)}) + \log q(\sigma_1^* | \sigma_1^{(\tau)}) + \log q(\sigma_2^* | \sigma_2^{(\tau)}) \\ &\quad + \sum_{i=1}^{n_1} \log f(x_i; \mu_1^*, \sigma_1^*) + \sum_{i=1}^{n_2} \log f(y_i; \mu_2^*, \sigma_2^*) \\ &\quad + \sum_{i=1}^{n_3} [\hat{\lambda} f(z_i; \mu_1^*, \sigma_1^*) + (1 - \hat{\lambda}) f(z_i; \mu_2^*, \sigma_2^*)] - \log(\sigma_1^* \sigma_2^*). \end{aligned}$$

2. Accept θ^* with probability

$$P(\theta^*, \theta^{(\tau)}) = \exp\left\{ \min\left[0, \log Q(\theta^* | \theta^\tau) - \log Q(\theta^{(\tau)} | \theta^*)\right] \right\}.$$

and let $\theta^{(\tau+1)} = \theta^*$. Otherwise, we reject the parameters and return to the first step.

The algorithm should be repeated sufficiently before obtaining the samples from the posterior distribution. This costs much more time compared with the ABC algorithm for the normal case. What is more, in our simulation we found that the MH algorithm may be too conservative. A better substitution could be the two-block Gibbs sampling proposed by Ren et al. [19]. In this sampling method, λ is first estimated using the EM algorithm, then for each loop, the parameters are updated by the conditional generalized pivotal quantities.

3. Real Data Example

In this section, we apply the proposed posterior p -value to the real halibut dataset studied by Hosmer [1], which was provided by the International Halibut Commission in Seattle, Washington. This dataset consists of the lengths of 208 halibut caught on one of their research cruises, in which 134 are female while the rest 74 are male. The data is summarized by Karunamuni and Wu [27]. We follow their method and randomly select 14 males and 26 females from the samples and regard them as the first and second sample of

the mixture model (1). The remaining male proportion of 60/168 is approximately identical to the original male proportion of 74/208, which is 0.3558. One hundred replications are generated with the same procedure. Hosmer [1] pointed out that the component for the dataset can be fitted by the normal distribution. A problem of interest is whether the sex effects the length of the halibut.

To test the homogeneity, for each replication we first use the EM algorithm to estimate λ , then we use the reject-ABC method to generate 8000 samples. We choose a moderate threshold, ϵ , to balance the accuracy and the computational cost. For the 100 replications, the mean estimate of the male proportion, λ , is 0.3381, with the mean squared error of 0.0045, which illustrates the accuracy of our EM algorithm. The estimates of the location and scale parameter of the male halibut are $\hat{\mu}_1 = 96.655$ and $\hat{\sigma}_1 = 12.983$, while those of the female ones are $\hat{\mu}_2 = 118.806$ and $\hat{\sigma}_2 = 9.077$. This is close to the estimates of Ren et al. [19]. As with the hypothesis testing $f_1 = f_2$, we calculate the posterior p -value of the 100 replications. Given the significance level, $\alpha = 0.05$, all the p -values are less than α . Thus, the null hypothesis is rejected, which indicates that there exists an association between the sex and length of the halibut.

4. Simulation Study

In this section, we present the simulation study of the cases discussed above. We compare the results of the posterior p -value (7) using different sampling methods and the generalized fiducial method proposed by Ren et al. [19]. As we can see from the simulations, the posterior p -value we proposed largely improves the testing of homogeneity. R programming language is used for our calculation and simulations.

4.1. Normal Case

When f_1 and f_2 are normal density functions, we compare the results of three different tests. The first two are the posterior p -value we proposed, but using the two-block Gibbs sampling and reject-ABC sampling methods, respectively. The last one is the generalized fiducial method proposed by [19]. In the following tables, the first two are denoted by “ T_G ” and “ T_R ”, while the last is denoted by “ G ”. We fix f_1 to $N(0, 1)$ while f_2 is set to be $N(0, 1)$, $N(1, 1)$, $N(0, 1.5^2)$ and $N(1, 1.5^2)$. For each f_1 and f_2 we consider $\lambda = 0.3, 0.5, 0.7$ and different sample sizes for n_1, n_2 and n_3 . We simulate $N = 10000$ repetitions for each case. For the Gibbs sampling, we accept 3000 samples after burning in the first 2000. For the reject-ABC sampling method, we first calculate the estimate of λ and accept 4000 parameters with ϵ set to $\sqrt{n_3}/2$. We then calculate the posterior p -value using the samples. We set the significance level to $\alpha = 0.05$ and reject the null hypothesis when the posterior p -value is below α . The results are shown in Tables 1–4. We further provide the QQ-plot of T_R in Figure 1, which indicates the correctness of Theorem 2. The first rows are the cases of $(n_1, n_2, n_3) = (10, 10, 10), (20, 20, 20)$ and $(30, 30, 30)$, while the last rows are the cases of $(10, 20, 30), (30, 20, 10)$ and $(15, 25, 150)$.

Table 1. Type-I errors (%) of the three methods in normal cases with nominal level $\alpha = 0.05$.

n_1, n_2, n_3	G	T_G	T_R
(10, 10, 10)	3.89	3.92	4.47
(20, 20, 20)	4.63	4.84	4.72
(30, 30, 30)	4.63	4.75	5.42
(10, 20, 30)	4.36	4.85	4.36
(30, 20, 10)	4.46	4.46	4.79
(10, 10, 100)	3.92	4.64	4.77
(15, 25, 150)	4.64	5.65	4.51

Table 2. Power comparison (%) of the cases when $f_1 = N(0, 1)$ and $f_2 = N(1, 1)$.

	0.3			0.5			0.7		
	G	T_G	T_R	G	T_G	T_R	G	T_G	T_R
(10, 10, 10)	32.2	33.9	36.8	32.2	33.9	36.5	33.3	35.9	36.5
(20, 20, 20)	70.6	74.2	76.3	70.3	74.0	76.0	72.3	76.2	75.6
(30, 30, 30)	90.2	93.3	92.1	89.2	91.2	92.2	90.7	92.2	92.0
(10, 20, 30)	46.3	50.9	46.4	51.3	56.1	45.6	56.8	63.2	45.6
(30, 20, 10)	84.3	85.8	85.3	81.3	83.3	85.4	82.8	84.3	85.2
(10, 10, 100)	35.4	43.0	41.8	33.1	41.9	41.7	34.3	44.5	41.2
(15, 25, 150)	63.5	70.7	71.5	66.8	75.1	74.0	75.8	82.3	72.6

Table 3. Power comparison (%) of the cases when $f_1 = N(0, 1)$ and $f_2 = N(0, 1.5^2)$.

	0.3			0.5			0.7		
	G	T_G	T_R	G	T_G	T_R	G	T_G	T_R
(10, 10, 10)	11.8	20.8	22.5	13.9	22.3	22.8	14.0	24.3	22.4
(20, 20, 20)	28.4	39.9	43.8	29.7	41.6	44.6	32.8	44.5	43.4
(30, 30, 30)	42.2	57.1	58.5	44.1	56.3	58.3	45.5	57.4	58.5
(10, 20, 30)	14.3	23.8	26.3	15.5	26.8	26.5	20.9	33.5	26.8
(30, 20, 10)	38.4	50.5	50.1	38.6	49.8	51.3	38.9	49.2	50.7
(10, 10, 100)	9.4	18.0	25.2	14.4	22.2	23.7	18.5	27.6	25.4
(15, 25, 150)	20.9	35.4	35.6	28.2	42.5	37.7	35.3	48.8	38.4

Table 4. Power comparison (%) of the cases when $f_1 = N(0, 1)$ and $f_2 = N(1, 1.5^2)$.

	0.3			0.5			0.7		
	G	T_G	T_R	G	T_G	T_R	G	T_G	T_R
(10, 10, 10)	38.9	42.6	46.6	38.2	42.1	41.7	42.0	46.2	42.6
(20, 20, 20)	77.3	79.7	82.3	75.2	78.0	80.6	80.0	82.6	80.8
(30, 30, 30)	93.6	94.2	95.5	93.8	94.0	93.3	93.9	94.5	94.4
(10, 20, 30)	47.2	50.0	55.2	55.4	59.6	57.2	62.1	66.0	56.1
(30, 20, 10)	88.4	89.0	88.9	85.8	87.2	86.5	86.3	87.6	85.3
(10, 10, 100)	46.0	48.9	49.1	53.8	56.9	45.9	59.1	64.6	48.6
(15, 25, 150)	77.1	78.6	78.1	83.9	84.8	74.4	89.6	91.5	75.1

We can see from the results that the posterior p -value largely improves the testing of homogeneity in normal cases. The Type-I error is controlled as well as the generalized fiducial methods. Moreover, our method significantly improves the power of testing homogeneity, especially when σ is different. The reject-ABC sampling method has the advantage of lower computational cost, compared with the two-block Gibbs sampling method. However, the power of using the reject-ABC sampling method is smaller than using two-block Gibbs sampling when n_3 is much larger than n_1 and n_2 . Thus, we can use the reject-ABC sampling method when the sample size is small or moderate and two-block Gibbs sampling when the sample size is large.

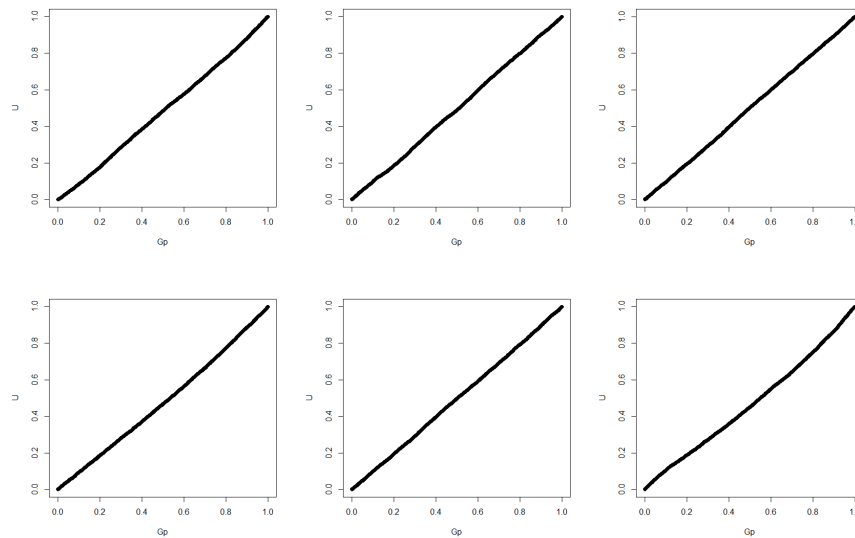


Figure 1. QQ-plot of the normal cases.

4.2. General Case

For the general case, we assume that f_1 and f_2 are logistic density functions. The location and scale parameters of f_1 and f_2 are set the same as that of the normal case. We simulate 10,000 repetitions for each sample size. We propose three methods in this simulation. The first two are the generalized fiducial method proposed by Ren et al. [19] and our posterior p -value using two-blocks Gibbs sampling. They are denoted by “ G ” and “ T_G ”, as in the last simulation. The last one is the posterior p -value using the M–H algorithm, which is denoted by “ T_M ”. First, we calculate the MLE of λ using the EM algorithm. We then propose the Metropolis–Hastings algorithm to obtain 12,000 samples after the first burn-in of 8000 ones. To avoid the dependency between the samples, we choose the first one in every three samples, which leaves us 4000 samples. We then use these samples to calculate the posterior p -value. The algorithm is natural and seems to be feasible. However, from Table 5 we can see that with this sampling method the results are rather conservative. Given the significance level $\alpha = 0.05$, the type-I error of T_M is always much smaller than 0.05, which makes the power of T_M also smaller than the other two when $f_1 \neq f_2$. However, We find that the two-block Gibbs sampling method can successfully solve the problem. It can be seen that the type-I error of “ T_G ” can be controlled well, while the power is largely improved compared with the generalized fiducial method. The results are shown in Tables 6–8. We also provide the QQ-plot of “ T_G ” in Figure 2.

Table 5. Type-I errors (%) of the three methods in logistic cases with nominal level $\alpha = 0.05$.

n_1, n_2, n_3	G	T_G	T_M
(10, 10, 10)	4.15	3.34	2.98
(20, 20, 20)	4.81	4.83	2.81
(10, 20, 30)	4.18	3.85	2.67
(30, 20, 10)	4.73	4.61	2.31
(30, 30, 30)	4.72	4.92	3.41
(10, 10, 100)	4.16	4.92	2.58
(15, 25, 150)	5.06	5.96	2.91

Table 6. Power comparison (%) of the cases when $f_1 = \text{Logis}(0,1)$ and $f_2 = \text{Logis}(1,1)$.

	0.3			0.5			0.7		
	G	T_G	T_M	G	T_G	T_M	G	T_G	T_M
(10, 10, 10)	12.2	12.0	13.2	12.6	13.3	14.5	13.2	13.0	13.7
(20, 20, 20)	29.4	30.3	26.2	28.3	29.9	26.4	28.9	30.2	26.0
(30, 30, 30)	43.2	45.4	36.6	43.1	45.2	35.9	43.9	46.5	37.3
(10, 20, 30)	17.5	18.1	15.1	19.3	20.2	16.9	22.2	23.2	18.7
(30, 20, 10)	35.4	36.6	31.8	33.9	35.8	32.4	33.5	35.1	30.4
(10, 10, 100)	13.8	17.8	12.7	13.8	17.1	11.8	13.9	16.6	11.9
(15, 25, 150)	24.4	29.3	20.1	25.3	30.0	22.7	30.9	38.1	24.1

Table 7. Power comparison (%) of the cases when $f_1 = \text{Logis}(0,1)$ and $f_2 = \text{Logis}(0,1.5)$.

	0.3			0.5			0.7		
	G	T_G	T_M	G	T_G	T_M	G	T_G	T_M
(10, 10, 10)	10.3	13.5	16.2	10.5	14.7	16.3	11.2	15.9	15.9
(20, 20, 20)	20.8	32.4	23.6	21.7	33.3	23.7	23.0	33.4	24.1
(30, 30, 30)	32.4	44.7	33.3	33.8	44.9	33.2	34.9	46.2	34.1
(10, 20, 30)	9.6	15.3	12.1	11.1	18.0	15.2	14.2	21.9	17.9
(30, 20, 10)	27.4	37.1	21.9	27.1	36.7	22.4	27.1	36.3	22.3
(10, 10, 100)	8.9	14.8	10.6	10.2	16.3	12.2	12.7	18.4	16.8
(15, 25, 150)	15.1	26.2	20.6	21.1	32.0	20.7	24.2	35.6	22.0

Table 8. Power comparison (%) of the cases when $f_1 = \text{Logis}(0,1)$ and $f_2 = \text{Logis}(1,1.5)$.

	0.3			0.5			0.7		
	G	T_G	T_M	G	T_G	T_M	G	T_G	T_M
(10, 10, 10)	17.4	17.9	18.1	18.3	21.7	18.8	18.5	22.6	19.6
(20, 20, 20)	43.7	49.4	36.3	43.7	49.3	36.2	45.2	51.3	37.7
(30, 30, 30)	62.9	67.3	58.1	62.2	67.2	57.9	63.6	68.4	59.6
(10, 20, 30)	20.6	24.2	21.3	24.1	29.4	26.2	30.1	35.8	35.8
(30, 20, 10)	52.7	57.0	37.6	51.7	55.9	39.7	51.5	56.0	38.8
(10, 10, 100)	19.9	24.1	18.7	23.1	27.3	19.8	23.5	30.2	19.7
(15, 25, 150)	36.7	43.1	40.4	44.6	51.5	40.7	51.0	57.8	40.1

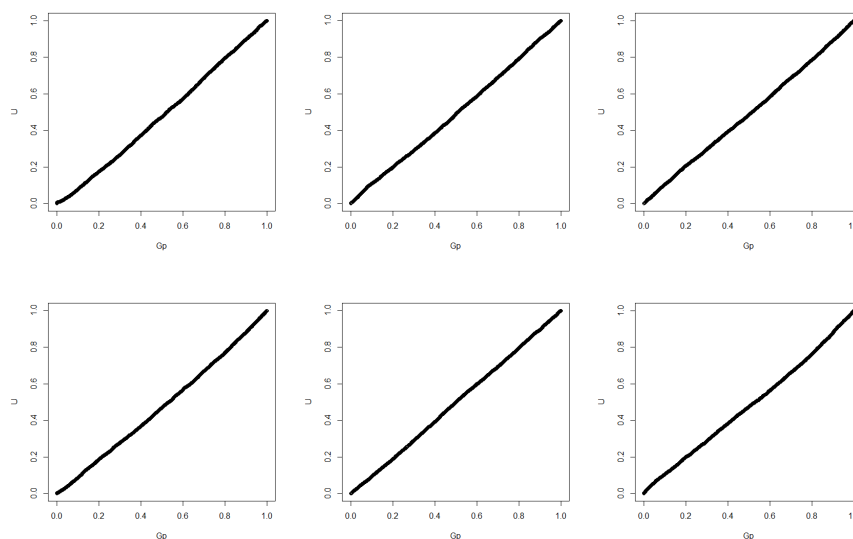


Figure 2. QQ-plot of the logistic cases.

5. Conclusions

In this paper, we propose a new posterior p -value for testing the homogeneity of the three-sample problem. We define the regular location-scale family and assume that both f_1 and f_2 are in the same family. Therefore, testing the homogeneity is equivalent to testing the equality of the location and scale parameters. We use the Bayes’ theorem to obtain the posterior distribution of the parameters and propose the Bernstein-von Mises theorem for multiple samples. We then propose the posterior p -value for testing the equality of the parameters. To sample from the posterior distribution, we compare different sampling methods. The simulation studies illustrate that the reject-ABC sampling method may be a good choice for the normal case while the two-block Gibbs sampling is better for the general ones. It should be noted that we transform the hypotheses of homogeneity to hypotheses (6). Then, with a different matrix, \mathbf{A} , we can generate our method to a variety of hypotheses.

Author Contributions: Conceptualization, X.X.; methodology, X.X.; software, Y.W.; validation, Y.W. and X.X.; formal analysis, Y.W. and X.X.; writing—original draft preparation, Y.W.; writing—review and editing, Y.W.; visualization, Y.W.; supervision, X.X.; project administration, X.X.; funding acquisition, X.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant No. 11471030 and No. 11471035.

Institutional Review Board Statement: The study did not require ethical approval.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The author is very grateful to the referees and to the assistant editor for their kind and professional remarks.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Proof of Lemma 1.

(1) First we show that

$$\lim_{x \rightarrow -\infty} f(x) = \lim_{x \rightarrow +\infty} f(x) = 0. \tag{A1}$$

By the third condition in Definition 1, there exists an $M > 0$, such that when $|x| > M$, $x^2|f'(x)| < 1$. Let $\{x_n, n = 1, 2, \dots\}$ be a sequence that satisfies $\lim_{n \rightarrow \infty} x_n = -\infty$ and let $a_n = f(x_n)$. Then, for sufficiently large m and n , such that $|x_n| > M$ and $|x_m| > M$, we have

$$|a_m - a_n| = |f(x_m) - f(x_n)| = \left| \int_{x_m}^{x_n} f'(x) dx \right| \leq \left| \int_{x_n}^{x_m} \frac{1}{x^2} dx \right| = \left| \frac{1}{x_m} - \frac{1}{x_n} \right|.$$

This indicates that when $m, n \rightarrow \infty$, $|a_m - a_n| \rightarrow 0$. Thus, $\{a_n, n = 1, 2, \dots\}$ is a Cauchy sequence, which must be convergent. Since $\{x_n, n = 1, 2, \dots\}$ is an arbitrary sequence, then the limit $\lim_{x \rightarrow -\infty} f(x)$ exists. Notice that $f(x)$ is a continuous density function, so

$$\lim_{x \rightarrow -\infty} f(x) = 0.$$

Similarly, we can show that

$$\lim_{x \rightarrow +\infty} f(x) = 0.$$

By condition (3) in Definition 1, for arbitrary $\varepsilon > 0$, there exists a number $B > 0$ such that when $|x| > B$, $x^2|f'(x)| < \varepsilon$. Then, by Equation (A1), if $x < -B$, we have

$$|xf(x)| = |x \int_{-\infty}^x f'(t)dt| \leq |x| \int_{-\infty}^x |f'(t)|dt \leq |x| \int_{-\infty}^x \frac{\varepsilon}{t^2}dt = \varepsilon.$$

This means that $\lim_{x \rightarrow -\infty} xf(x) = 0$. If $x > B$, we have

$$|xf(x)| = |x \int_x^{+\infty} f'(t)dt| \leq |x| \int_x^{+\infty} |f'(t)|dt \leq |x| \int_x^{+\infty} \frac{\varepsilon}{t^2}dt = \varepsilon.$$

This means that $\lim_{x \rightarrow +\infty} xf(x) = 0$.

(2) From (A1) we can get

$$\int_{-\infty}^{\infty} f'(x)dx = f(x) \Big|_{-\infty}^{\infty} = 0.$$

(3) As we can see

$$\int_{-\infty}^{\infty} xf'(x)dx = xf(x) \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f(x)dx.$$

Then, by the lemma, we just proved the fact that $f(x)$ is a density function,

$$\int_{-\infty}^{\infty} xf'(x)dx = -1. \tag{A2}$$

(4) Since

$$\lim_{x \rightarrow -\infty} f'(x) = \lim_{x \rightarrow +\infty} f(x) = 0.$$

Then it is easy to get

$$\int_{-\infty}^{\infty} f''(x)dx = f'(x) \Big|_{-\infty}^{\infty} = 0.$$

(5)

$$\lim_{x \rightarrow -\infty} xf'(x) = \lim_{x \rightarrow +\infty} xf'(x) = 0$$

$$\int_{-\infty}^{+\infty} xf''(x)dx = xf'(x) \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f'(x)dx = 0 - f(x) \Big|_{-\infty}^{\infty}$$

Then, by Equation (A1),

$$\int_{-\infty}^{\infty} xf''(x)dx = 0.$$

(6)

$$\int_{-\infty}^{\infty} x^2 f''(x)dx = x^2 f'(x) \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} 2xf'(x)dx = -2 \int_{-\infty}^{\infty} xf'(x)dx$$

Then by, Equation (A2), we have

$$\int_{-\infty}^{\infty} x^2 f''(x)dx = 2$$

□

Proof of Proposition 1.

(1) The log likelihood function, $l(\xi, x)$, is

$$l(\xi, x) = \log f(x, \xi) = -\log \sigma + \log f\left(\frac{x - \mu}{\sigma}\right).$$

We can then get the derivatives as below

$$\begin{aligned} \frac{\partial l(\xi, x)}{\partial \mu} &= -\frac{1}{\sigma} f' \left(\frac{x - \mu}{\sigma} \right) / f \left(\frac{x - \mu}{\sigma} \right) \\ \frac{\partial l(\xi, x)}{\partial \sigma} &= -\frac{1}{\sigma} - \frac{1}{\sigma} \frac{x - \mu}{\sigma} f' \left(\frac{x - \mu}{\sigma} \right) / f \left(\frac{x - \mu}{\sigma} \right) \end{aligned}$$

Using the second term in Lemma 1, we can get the expectation of the first derivatives

$$\begin{aligned} E_{\xi} \left[\frac{\partial l(\xi, x)}{\partial \mu} \right] &= -\frac{1}{\sigma^2} \int_{-\infty}^{\infty} f' \left(\frac{x - \mu}{\sigma} \right) dx = -\frac{1}{\sigma} \int_{-\infty}^{\infty} f'(y) dy = 0 \\ E_{\xi} \left[\frac{\partial l(\xi, x)}{\partial \sigma} \right] &= 0. \end{aligned}$$

(2) The elements of the Fisher information matrix are computed as

$$\begin{aligned} E_{\xi} \left[\frac{\partial l(\xi, x)}{\partial \mu} \right]^2 &= \int_{-\infty}^{\infty} \frac{1}{\sigma^2} \frac{\left[f' \left(\frac{x - \mu}{\sigma} \right) \right]^2}{f^2 \left(\frac{x - \mu}{\sigma} \right)} \cdot \frac{1}{\sigma} f \left(\frac{x - \mu}{\sigma} \right) dx \\ &= \frac{1}{\sigma^2} \int_{-\infty}^{\infty} \frac{[f'(y)]^2}{f(y)} dy = \frac{C_{11}(f)}{\sigma^2}, \\ E_{\xi} \left[\frac{\partial l(\xi, x)}{\partial \mu} \frac{\partial l(\theta, x)}{\partial \sigma} \right] &= \int_{-\infty}^{\infty} \frac{1}{\sigma} \frac{f' \left(\frac{x - \mu}{\sigma} \right)}{f \left(\frac{x - \mu}{\sigma} \right)} \left[\frac{1}{\sigma} + \frac{1}{\sigma} \frac{x - \mu}{\sigma} \frac{f' \left(\frac{x - \mu}{\sigma} \right)}{f \left(\frac{x - \mu}{\sigma} \right)} \right] \frac{1}{\sigma} f \left(\frac{x - \mu}{\sigma} \right) dx \\ &= \frac{1}{\sigma^2} \int_{-\infty}^{\infty} [f'(y) + y[f'(y)]^2 / f(y)] dy \\ &= \frac{1}{\sigma^2} \int_{-\infty}^{\infty} y \frac{[f'(y)]^2}{f(y)} dy = \frac{C_{12}(f)}{\sigma^2}, \\ E_{\xi} \left[\frac{\partial l(\xi, x)}{\partial \sigma} \right]^2 &= \frac{1}{\sigma^2} \int_{-\infty}^{\infty} \left[1 + \frac{x - \mu}{\sigma} \frac{f' \left(\frac{x - \mu}{\sigma} \right)}{f \left(\frac{x - \mu}{\sigma} \right)} \right]^2 \frac{1}{\sigma} f \left(\frac{x - \mu}{\sigma} \right) dx \\ &= \frac{1}{\sigma^2} \int_{-\infty}^{\infty} \left[1 + y \frac{f'(y)}{f(y)} \right]^2 f(y) dy \\ &= \frac{1}{\sigma^2} \left[1 + 2 \int_{-\infty}^{\infty} y f'(y) dy + \int_{-\infty}^{\infty} y^2 \frac{[f'(y)]^2}{f(y)} dy \right] \\ &= \frac{1}{\sigma^2} \left[\int_{-\infty}^{\infty} y^2 \frac{[f'(y)]^2}{f(y)} dy - 1 \right] = \frac{C_{22}(f)}{\sigma^2}. \end{aligned}$$

So the equation holds. By the fourth condition in Definition 1, we can prove that

$$I_f(\xi) = \frac{1}{\sigma^2} \mathbf{C}(f) < \infty.$$

Now, we show that $\mathbf{C}(f) > 0$. Suppose that $|\mathbf{C}(f)| = 0$, then there exists a nonzero vector $\mathbf{a} = (a_1, a_2)^\top$, such that $\mathbf{a}^\top \mathbf{C}\mathbf{a} = 0$, which also means that

$$\mathbf{a}^\top \frac{\partial l(\xi, x)}{\partial \theta} \Big|_{\xi=(0,1)^\top} = 0, \quad a.e. \quad f(x).$$

Since $f(x) > 0$, we have

$$a^\top \frac{\partial l(\xi, x)}{\partial \theta} \Big|_{\xi=(0,1)^\top} = -a_1 \frac{f'(x)}{f(x)} + a_2 \left(-1 - x \frac{f'(x)}{f(x)} \right) = 0, \quad a.e. \quad L$$

where L is the Lebesgue measure. Because \mathbf{a} is nonzero, so $a_2 \neq 0$, then

$$(x + b) \frac{f'(x)}{f(x)} + 1 = 0,$$

where $b = a_1/a_2$. When $x > -b$,

$$\begin{aligned} (\log f(x))' &= \frac{f'(x)}{f(x)} = -\frac{1}{x + b}, \\ \log f(x) &= -\ln(x + b) + D, \end{aligned}$$

where D is a constant. Then, $f(x) = e^D/(x + b)$, this contradicts the first equation in Lemma 1. Thus, the assumption of $|\mathbf{C}(f)| = 0$ is not true, then $I(\theta) > 0$.

(3) We first calculate the second derivatives of the parameters.

$$\begin{aligned} \frac{\partial^2 l(\xi, x)}{\partial \mu^2} &= \frac{1}{\sigma^2} \left[\frac{f''\left(\frac{x-\mu}{\sigma}\right)}{f\left(\frac{x-\mu}{\sigma}\right)} - \frac{\left(f'\left(\frac{x-\mu}{\sigma}\right)\right)^2}{f^2\left(\frac{x-\mu}{\sigma}\right)} \right], \\ \frac{\partial^2 l(\xi, x)}{\partial \mu \partial \sigma} &= \frac{1}{\sigma^2} \left[\frac{f'\left(\frac{x-\mu}{\sigma}\right)}{f\left(\frac{x-\mu}{\sigma}\right)} + \frac{x-\mu}{\sigma} \frac{f''\left(\frac{x-\mu}{\sigma}\right)}{f\left(\frac{x-\mu}{\sigma}\right)} - \frac{x-\mu}{\sigma} \frac{\left(f'\left(\frac{x-\mu}{\sigma}\right)\right)^2}{f^2\left(\frac{x-\mu}{\sigma}\right)} \right], \\ \frac{\partial^2 l(\xi, x)}{\partial \sigma^2} &= \frac{1}{\sigma^2} + \frac{2(x-\mu)}{\sigma^3} \cdot \frac{f'\left(\frac{x-\mu}{\sigma}\right)}{f\left(\frac{x-\mu}{\sigma}\right)} + \frac{(x-\mu)^2}{\sigma^4} \frac{f''\left(\frac{x-\mu}{\sigma}\right)}{f\left(\frac{x-\mu}{\sigma}\right)} - \frac{(x-\mu)^2}{\sigma^4} \frac{\left(f'\left(\frac{x-\mu}{\sigma}\right)\right)^2}{\left(f\left(\frac{x-\mu}{\sigma}\right)\right)^2}. \end{aligned}$$

Then, by Lemma 1, we have

$$\begin{aligned} E_{\xi} \left[\frac{\partial^2 l(\xi, x)}{\partial \mu^2} \right] &= -\frac{C_{11}(f)}{\sigma^2}, \\ E_{\xi} \left[\frac{\partial^2 l(\xi, x)}{\partial \mu \partial \sigma} \right] &= -\frac{C_{12}(f)}{\sigma^2}, \\ E_{\xi} \left[\frac{\partial^2 l(\xi, x)}{\partial \sigma^2} \right] &= -\frac{C_{22}(f)}{\sigma^2}. \end{aligned}$$

□

Proof of Proposition 2.

(1) First, we calculate the derivatives as follows.

$$\begin{aligned} \frac{\partial \log g(x, \theta)}{\partial \mu_1} &= -\frac{\lambda_0}{\sigma_1^2} f'\left(\frac{x-\mu_1}{\sigma_1}\right) / g(x; \theta); \\ \frac{\partial \log g(x, \theta)}{\partial \mu_2} &= -\frac{1-\lambda_0}{\sigma_2^2} f'\left(\frac{x-\mu_2}{\sigma_2}\right) / g(x; \theta); \\ \frac{\partial \log g(x, \theta)}{\partial \sigma_1} &= \left[-\frac{\lambda_0}{\sigma_1^2} f\left(\frac{x-\mu_1}{\sigma_1}\right) - \frac{\lambda_0}{\sigma_1} \frac{x-\mu_1}{\sigma_1^2} f'\left(\frac{x-\mu_1}{\sigma_1}\right) \right] / g(x; \theta); \\ \frac{\partial \log g(x, \theta)}{\partial \sigma_2} &= \left[-\frac{1-\lambda_0}{\sigma_2^2} f\left(\frac{x-\mu_2}{\sigma_2}\right) - \frac{1-\lambda_0}{\sigma_2} \frac{x-\mu_2}{\sigma_2^2} f'\left(\frac{x-\mu_2}{\sigma_2}\right) \right] / g(x; \theta). \end{aligned}$$

Then, by the second equation in Lemma 1, we have

$$E_{\theta} \left[\frac{\partial \log g(x; \theta)}{\partial \mu_1} \right] = -\frac{\lambda_0}{\sigma_1^2} \int_{-\infty}^{\infty} f'\left(\frac{x-\mu_1}{\sigma}\right) dx = -\frac{\lambda_0}{\sigma} \int_{-\infty}^{\infty} f(y) dy = 0$$

By Lemma 1(3),

$$\begin{aligned} E_{\theta} \left[\frac{\partial \log g(x; \theta)}{\partial \sigma_1} \right] &= \int_{-\infty}^{+\infty} \left[-\frac{\lambda_0}{\sigma_1^2} f\left(\frac{x - \mu_1}{\sigma_1}\right) - \frac{\lambda_0}{\sigma_1} \frac{x - \mu_1}{\sigma_1^2} f'\left(\frac{x - \mu_1}{\sigma_1}\right) \right] dx \\ &= \int_{-\infty}^{\infty} \left[-\frac{\lambda_0}{\sigma_1} f(y) - \frac{\lambda_0}{\sigma_1} y f(y) \right] dy \\ &= -\frac{\lambda_0}{\sigma_1} \left[1 + \int_{-\infty}^{\infty} y f(y) dy \right] = 0. \end{aligned}$$

Similarly, can we prove that

$$E_{\theta} \left[\frac{\partial \log g(x; \theta)}{\partial \sigma_2} \right] = 0.$$

(2) First, we calculate the derivatives on the location parameter as

$$\begin{aligned} E_{\theta} \left[\frac{\partial \log g(x; \theta)}{\partial \mu_1} \right]^2 &= \int_{-\infty}^{\infty} \frac{\lambda_0^2}{\sigma_1^4} \left[f'\left(\frac{x - \mu_1}{\sigma_1}\right) \right]^2 / g(x; \theta) dx \\ &\leq \int_{-\infty}^{\infty} \frac{\lambda_0^4}{\sigma_1^2} \left[f'\left(\frac{x - \mu_1}{\sigma_1}\right) \right]^2 / \left[\frac{\lambda_0}{\sigma_1} f\left(\frac{x - \mu_1}{\sigma_1}\right) \right] dx \\ &= \frac{\lambda_0}{\sigma_1^2} \int_{-\infty}^{\infty} \frac{(f'(y))^2}{f(y)} dy \\ &= \frac{\lambda_0}{\sigma_1^2} C_{11}(f) < \infty. \end{aligned}$$

Similarly, we have

$$E_{\theta} \left[\frac{\partial \log g(x; \theta)}{\partial \mu_2} \right]^2 \leq \frac{1 - \lambda_0}{\sigma_2^2} C_{11}(f) < \infty$$

Then, as with the scale parameter, we have

$$\begin{aligned} E_{\theta} \left[\frac{\partial \log g(x; \theta)}{\partial \sigma_1} \right]^2 &= \frac{\lambda_0^2}{\sigma_1^2} \int_{-\infty}^{\infty} \left[\frac{1}{\sigma_1} f\left(\frac{x - \mu_1}{\sigma_1}\right) + \frac{x - \mu_1}{\sigma_1^2} f'\left(\frac{x - \mu_1}{\sigma_1}\right) \right]^2 / g(x; \theta) dx \\ &\leq \frac{\lambda_0^2}{\sigma_1^2} \int_{-\infty}^{\infty} \left[\frac{1}{\sigma_1} f\left(\frac{x - \mu_1}{\sigma_1}\right) + \frac{x - \mu_1}{\sigma_1^2} f'\left(\frac{x - \mu_1}{\sigma_1}\right) \right]^2 / \frac{\lambda_0}{\sigma_1} f\left(\frac{x - \mu_1}{\sigma_1}\right) dx \\ &= \frac{\lambda_0}{\sigma_1^2} \int_{-\infty}^{\infty} \left[1 + y \frac{(f'(y))^2}{f(y)} \right]^2 f(y) dy \\ &= \frac{\lambda_0}{\sigma_1^2} C_{22}(f) < \infty \end{aligned}$$

$$E_{\theta} \left[\frac{\partial \log g(x; \theta)}{\partial \sigma_2} \right]^2 = \frac{1 - \lambda_0}{\sigma_2^2} C_{22}(f) < \infty.$$

Then $I(\theta) < \infty$.

(3)

$$\frac{\partial^2 \log g(x; \theta)}{\partial \mu_1^2} = \frac{\frac{\lambda_0}{\sigma_1^3} f''\left(\frac{x - \mu_1}{\sigma_1}\right)}{g(x; \theta)} - \frac{\frac{\lambda_0^2}{\sigma_1^4} \left(f'\left(\frac{x - \mu_1}{\sigma_1}\right) \right)^2}{(g(x; \theta))^2}$$

Then, by the fourth equation in Lemma 1 can we get

$$E_{\theta} \frac{\partial^2 \log g(x; \theta)}{\partial \mu_1^2} = -E_{\theta} \frac{\frac{\lambda_0^2}{\sigma_1^4} \left(f'\left(\frac{x - \mu_1}{\sigma_1}\right) \right)^2}{g(x; \theta)^2} = -E_{\theta} \left(\frac{\partial \log g(x; \theta)}{\partial \mu_1} \right)^2$$

The same procedure can be applied to the other nine equations to show that the conclusion holds. \square

Proof of Theorem 1. First, we provide the Bernstein-von Mises theorem for multiple samples; see Theorem 2 in Long and Xu [28]. Besides Assumptions 1 to 4 in the context, there are some other assumptions below

Assumption A1. For all $i = 1, 2, \dots, k$, the density function, $f_i(x|\theta)$, of the population G_i satisfies the following conditions:

(a) The parameter space of θ contains an open subset, $\omega \subset \Omega$, in which the true value is included.

(b) The set $A_i = \{x : f_i(x|\theta) > 0\}$ is independent of θ .

(c) For almost all $x \in A_i$, $f_i(x|\theta)$ as a function of θ admits continuous second derivatives $\frac{\partial^2}{\partial\theta_j\partial\theta_h} f_i(x|\theta)$, $j, h = 1, 2, \dots, d$, for all $\theta \in \omega$.

(d) Denote by $I^{(i)}(\theta)$ the Fisher's information matrix of $f_i(x|\theta)$. The first and second derivatives of the logarithm of $f_i(x|\theta)$ satisfy the equations

$$E_\theta \left[\frac{\partial}{\partial\theta_j} \log f_i(x|\theta) \right] = 0, \quad j = 1, \dots, d,$$

$$I_{jh}^{(i)}(\theta) = E_\theta \left[\frac{\partial}{\partial\theta_j} \log f_i(x|\theta) \cdot \frac{\partial}{\partial\theta_h} \log f_i(x|\theta) \right]$$

$$= E_\theta \left[-\frac{\partial^2}{\partial\theta_j\partial\theta_h} \log f_i(x|\theta) \right], \quad j, h = 1, 2, \dots, d.$$

(e) Suppose the sample size n_i of G_i satisfies when $n \rightarrow \infty$, $n_i/n \rightarrow r_i \in (0, 1)$. Let

$$I(\theta) = \sum_{i=1}^k r_i I^{(i)}(\theta).$$

We assume that all entries of $I(\theta)$ are finite, and $I(\theta)$ is positive definite.

Then, by Definition 1, Propositions 1 and 2 and Assumptions 1–A1, Theorem 1 holds. It should be noted that since the prior is $\pi(\theta) = 1/\sigma_1\sigma_2$, its second moment does not exist. Therefore, we draw k_0 samples from the first two density functions and combine them with $\pi(\theta)$; thus, we get the new prior. This is a trick in the research of big data. \square

Proof of Theorem 2. First, we present two conclusions

$$\sqrt{n}(\theta_B - T_n) \xrightarrow{P} 0, \quad n\Sigma_B \xrightarrow{P} I^{-1}(\theta_0). \tag{A3}$$

Let $E_P g(\theta)$ be the expectation of $g(\theta)$ under distribution P . Then

$$\sqrt{n}(\theta_B - T_n) = \sqrt{n}(E_\pi\theta - T_n) = E_\pi[\sqrt{n}(\theta - T_n)] = E_{\pi^*}\theta - E_{N(0, I^{-1}(\theta_0))}\theta.$$

$$\begin{aligned} \|\sqrt{n}(\theta_B - T_n)\| &= \|E_{\pi^*}\theta - E_{N(0, I^{-1}(\theta_0))}\theta\| \\ &\leq \int \|\theta\| \left| \pi^*(\theta|x) - (2\pi)^{-\frac{d}{2}} |I(\theta_0)|^{\frac{1}{2}} e^{-\theta^\top I(\theta_0)\theta} \right| d\theta. \end{aligned}$$

By Theorem 1, the above equation converges in probability to 0.

$$\begin{aligned}
 n\Sigma_B &= nE_\pi(\theta - \theta_B)(\theta - \theta_B)^\top \\
 &= nE_\pi(\theta - T_n + T_n - \theta_B)(\theta - T_n + T_n - \theta_B)^\top \\
 &= nE_\pi(\theta - T_n)(\theta - T_n)^\top + nE_\pi(\theta - T_n)(T_n - \theta_B)^\top \\
 &\quad + n(T_n - \theta_B)E_\pi(\theta - T_n) + n(T_n - \theta_B)(T_n - \theta_B)^\top \\
 &= E_{\pi^*}\theta\theta^\top + E_{\pi^*}\theta\sqrt{n}(T_n - \theta_B)^\top + \sqrt{n}(T_n - \theta_B)E_{\pi^*}\theta + [\sqrt{n}(T_n - \theta_B)][\sqrt{n}(T_n - \theta_B)]^\top.
 \end{aligned}$$

From the conclusion above we have

$$\sqrt{n}(T_n - \theta_B) \xrightarrow{P} 0, \quad E_{\pi^*}\theta \xrightarrow{P} 0,$$

then, by Theorem 1,

$$E_{\pi^*}(\theta\theta^\top) \rightarrow I^{-1}(\theta_0).$$

thus,

$$n\Sigma_B \xrightarrow{P} I^{-1}(\theta_0).$$

Then we can get

$$\begin{aligned}
 &(\theta - \theta_B)^\top \mathbf{A}^\top (\mathbf{A}\Sigma_B\mathbf{A}^\top)^{-1} \mathbf{A}(\theta - \theta_B) \\
 &= [\sqrt{n}(\theta - \theta_B)]^\top \mathbf{A}^\top [\mathbf{A}(n\Sigma_B)\mathbf{A}^\top]^{-1} \mathbf{A}[\sqrt{n}(\theta - \theta_B)] \\
 &= [\sqrt{n}(\theta - T_n) - \sqrt{n}(\theta_B - T_n)]^\top \mathbf{A}^\top [\mathbf{A}(n\Sigma_B)\mathbf{A}^\top]^{-1} \mathbf{A}[\sqrt{n}(\theta - T_n) - \lambda_n(\theta_B - T_n)] \\
 &= t^\top \mathbf{A}^\top [\mathbf{A}(n\Sigma_B)\mathbf{A}^\top]^{-1} \mathbf{A}t - 2t^\top \mathbf{A}^\top [\mathbf{A}(n\Sigma_B)\mathbf{A}^\top]^{-1} \mathbf{A}[\sqrt{n}(\theta_B - T_n)] \\
 &\quad + [\sqrt{n}(\theta_B - T_n)]^\top \mathbf{A}^\top [\mathbf{A}(n\Sigma_B)\mathbf{A}^\top]^{-1} \mathbf{A}[\sqrt{n}(\theta_B - T_n)].
 \end{aligned}$$

The expression above should have the same asymptotic distribution as $t^\top \mathbf{A}^\top [\mathbf{A}(n\Sigma_B)\mathbf{A}^\top]^{-1} \mathbf{A}t$, where $t \sim N_p(0, I^{-1}(\theta_0))$ and $\mathbf{A}t \sim N_k(0, \mathbf{A}I^{-1}(\theta_0)\mathbf{A}^\top)$. From the conclusion above, we have $\mathbf{A}(n\Sigma_B)\mathbf{A}^\top \xrightarrow{P} \mathbf{A}I^{-1}(\theta_0)\mathbf{A}^\top$; thus, we can get

$$t^\top \mathbf{A}^\top [\mathbf{A}(n\Sigma_B)\mathbf{A}^\top]^{-1} \mathbf{A}t \rightarrow \chi^2(k),$$

where k is the degree of freedom and also the rows of matrix \mathbf{A} . Thus

$$(\theta - \theta_B)^\top \mathbf{A}^\top (\mathbf{A}\Sigma_B\mathbf{A}^\top)^{-1} \mathbf{A}(\theta - \theta_B) \xrightarrow{d} \chi^2(k)$$

Under the null hypothesis,

$$\begin{aligned}
 &(b - \mathbf{A}\theta_B)^\top (\mathbf{A}\Sigma_B\mathbf{A}^\top)^{-1} (b - \mathbf{A}\theta_B) \\
 &= [b - \mathbf{A}T_n - \mathbf{A}(\theta_B - T_n)]^\top (\mathbf{A}\Sigma_B\mathbf{A}^\top)^{-1} [b - \mathbf{A}T_n - \mathbf{A}(\theta_B - T_n)] \\
 &= (b - \mathbf{A}T_n)^\top (\mathbf{A}\Sigma_B\mathbf{A}^\top)^{-1} (b - \mathbf{A}T_n) + (\theta_B - T_n)^\top \mathbf{A}^\top (\mathbf{A}\Sigma_B\mathbf{A}^\top)^{-1} \mathbf{A}(\theta_B - T_n) \\
 &\quad - 2(b - \mathbf{A}T_n)^\top (\mathbf{A}\Sigma_B\mathbf{A}^\top)^{-1} \mathbf{A}(\theta_B - T_n).
 \end{aligned}$$

Since $b = \mathbf{A}\theta_0$, the expression above is equivalent to

$$\begin{aligned} & \left(\frac{1}{n}I^{-1}(\theta_0)l'(\theta_0)\right)^\top \mathbf{A}^\top (\mathbf{A}\Sigma_B\mathbf{A}^\top)^{-1} \mathbf{A} \left(\frac{1}{n}I''(\theta_0)l'(\theta_0)\right) \\ & + [\sqrt{n}(\theta_B - T_n)]^{-1} \mathbf{A}^\top [\mathbf{A}(n\Sigma_B)\mathbf{A}^\top]^{-1} \mathbf{A} [\sqrt{n}(\theta_B - T_n)] \\ & + 2\left(\frac{1}{\sqrt{n}}I^{-1}(\theta_0)l^1(\theta_0)\right)^\top \mathbf{A}^\top [\mathbf{A}(n\Sigma_B)\mathbf{A}^\top]^{-1} \mathbf{A} [\sqrt{n}(\theta_B - T_n)]. \end{aligned}$$

The first term can be rewritten as

$$\left(\frac{1}{\sqrt{n}}I^{-1}(\theta_0)l'(\theta_0)\right)^\top \mathbf{A}^\top [\mathbf{A}(n\Sigma_B)\mathbf{A}^\top]^{-1} \mathbf{A} \left(\frac{1}{\sqrt{n}}I^{-1}(\theta_0)l'(\theta_0)\right).$$

This asymptotically follows the $\chi^2(k)$ distribution. The second and third terms tend to 0 in probability by Equation (A3). Thus

$$p(x) - F_k \left[1 - \left(\frac{1}{\sqrt{n}}I^{-1}(\theta_0)l'(\theta_0)\right)^\top \mathbf{A}^\top [\mathbf{A}(n\Sigma_B)\mathbf{A}^\top]^{-1} \mathbf{A} \left(\frac{1}{n}I^{-1}(\theta_0)l'(\theta_0)\right) \right] \xrightarrow{P} 0.$$

where F_k is the cumulative distribution function of $\chi^2(k)$. Then, by the asymptotic property, we have

$$p(x) \xrightarrow{d} U(0,1).$$

□

References

- Hosmer, D.W. A Comparison of Iterative Maximum Likelihood Estimates of the Parameters of a Mixture of Two Normal Distributions Under Three Different Types of Sample. *Biometrics* **1973**, *29*, 761–770. [\[CrossRef\]](#)
- Murray, G.D.; Titterton, D.M. Estimation Problems with Data from a Mixture. *J. R. Stat. Soc. Ser. (Appl. Stat.)* **1978**, *27*, 325–334. [\[CrossRef\]](#)
- Anderson, J.A. Multivariate logistic compounds. *Biometrika* **1979**, *66*, 17–26. [\[CrossRef\]](#)
- Qin, J. Empirical likelihood ratio based confidence intervals for mixture proportions. *Ann. Stat.* **1999**, *27*, 1368–1384. [\[CrossRef\]](#)
- Owen, A. Empirical Likelihood Ratio Confidence Regions. *Ann. Stat.* **1990**, *18*, 90–120. [\[CrossRef\]](#)
- Zou, F.; Fine, J.P.; Yandell, B.S. On empirical likelihood for a semiparametric mixture model. *Biometrika* **2002**, *89*, 61–75. [\[CrossRef\]](#)
- Zhang, B. Assessing goodness-of-fit of generalized logit models based on case-control data. *J. Multivar. Anal.* **2002**, *82*, 17–38. [\[CrossRef\]](#)
- Inagaki, K.; Komaki, F. A modification of profile empirical likelihood for the exponential-tilt model. *Stat. Probab. Lett.* **2010**, *80*, 997–1004. [\[CrossRef\]](#)
- Tan, Z. A note on profile likelihood for exponential tilt mixture models. *Biometrika* **2009**, *96*, 229–236. [\[CrossRef\]](#)
- Liang, K.Y.; Rathouz, P.J. Hypothesis Testing Under Mixture Models: Application to Genetic Linkage Analysis. *Biometrics* **1999**, *55*, 65–74. [\[CrossRef\]](#)
- Duan, R.; Ning, Y.; Wang, S.; Lindsay, B.G.; Carroll, R.J.; Chen, Y. A fast score test for generalized mixture models. *Biometrics* **2019**, *76*, 811–820. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fu, Y.; Chen, J.; Kalbfleisch, J.D. Testing for homogeneity in genetic linkage analysis. *Stat. Sin.* **2006**, *16*, 805–823.
- Chen, H.; Chen, J.; Kalbfleisch, J.D. A Modified Likelihood Ratio Test for Homogeneity in Finite Mixture Models. *J. R. Stat. Soc. Ser. Stat. Methodol.* **2002**, *63*, 19–29. [\[CrossRef\]](#)
- Chen, H.; Chen, J.; Kalbfleisch, J.D. Testing for a Finite Mixture Model with Two Components. *J. R. Stat. Soc. Ser. Stat. Methodol.* **2003**, *66*, 95–115. [\[CrossRef\]](#)
- Chen, J.; Li, P. Hypothesis test for normal mixture models: The EM approach. *Ann. Stat.* **2009**, *37*, 2523–2542. [\[CrossRef\]](#)
- Li, P.; Liu, Y.; Qin, J. Semiparametric Inference in a Genetic Mixture Model. *J. Am. Stat. Assoc.* **2017**, *112*, 1250–1260. [\[CrossRef\]](#)
- Li, S.; Chen, J.; Guo, J.; Jing, B.Y.; Tsang, S.Y.; Xue, H. Likelihood Ratio Test for Multi-Sample Mixture Model and Its Application to Genetic Imprinting. *J. Am. Stat. Assoc.* **2015**, *110*, 867–877. [\[CrossRef\]](#)
- Liu, G.; Li, P.; Liu, Y.; Pu, X. Hypothesis testing for quantitative trait locus effects in both location and scale in genetic backcross studies. *Scand. J. Stat.* **2020**, *47*, 1064–1089. [\[CrossRef\]](#)
- Ren, P.; Liu, G.; Pu, X. Generalized fiducial methods for testing the homogeneity of a three-sample problem with a mixture structure. *J. Appl. Stat.* **2023**, *50*, 1094–1114. [\[CrossRef\]](#)

20. Hannig, J.; Iyer, H.; Lai, R.C.S.; Lee, T.C.M. Generalized Fiducial Inference: A Review and New Results. *J. Am. Stat. Assoc.* **2016**, *111*, 1346–1361. [[CrossRef](#)]
21. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm; With discussion. *J. R. Stat. Soc. Ser. B* **1977**, *39*, 1–22.
22. McLachlan, G.J.; Krishnan, T. *The EM Algorithm and Extensions*; Wiley Series in Probability and Statistics: Applied Probability and Statistics; A Wiley-Interscience Publication; John Wiley & Sons, Inc.: New York, NY, USA, 1997; pp. xviii+274.
23. Fisher, R.A. The fiducial argument in statistical inference. *Ann. Eugen.* **1935**, *6*, 391–398. [[CrossRef](#)]
24. Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092. [[CrossRef](#)]
25. Hastings, W.K. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* **1970**, *57*, 97–109. [[CrossRef](#)]
26. Owen, A.B.; Glynn, P.W. (Eds.) *Monte Carlo and Quasi-Monte Carlo Methods*; Springer International Publishing: Cham, Switzerland, 2018.
27. Karunamuni, R.; Wu, J. Minimum Hellinger distance estimation in a nonparametric mixture model. *J. Stat. Plan. Inference* **2009**, *139*, 1118–1133. [[CrossRef](#)]
28. Long, Y.; Xu, X. Bayesian decision rules to classification problems. *Aust. N. Z. J. Stat.* **2021**, *63*, 394–415. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.