



Article On the Reversible Jump Markov Chain Monte Carlo (RJMCMC) Algorithm for Extreme Value Mixture Distribution as a Location-Scale Transformation of the Weibull Distribution

Dwi Rantini ^{1,2}, Nur Iriawan ^{1,*} and Irhamah ¹

- ¹ Department of Statistics, Faculty of Science and Data Analytics, Institute Teknologi Sepuluh Nopember, Jl. Arif Rahman Hakim, Surabaya 60111, Indonesia; rantinidwi2@gmail.com (D.R.); irhamah@statistika.its.ac.id (I.)
- ² Department of Data Science Technology, Faculty of Advanced Technology and Multidiscipline, Universitas Airlangga, Surabaya 60115, Indonesia

Abstract: Data with a multimodal pattern can be analyzed using a mixture model. In a mixture

* Correspondence: nur_i@statistika.its.ac.id

check for updates

Citation: Rantini, D.; Iriawan, N.; Irhamah On the Reversible Jump Markov Chain Monte Carlo (RJMCMC) Algorithm for Extreme Value Mixture Distribution as a Location-Scale Transformation of the Weibull Distribution. *Appl. Sci.* 2021, *11*, 7343. https://doi.org/10.3390/ app11167343

Academic Editor: Antonio López-Quílez

Received: 9 July 2021 Accepted: 6 August 2021 Published: 10 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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/). model, the most important step is the determination of the number of mixture components, because finding the correct number of mixture components will reduce the error of the resulting model. In a Bayesian analysis, one method that can be used to determine the number of mixture components is the reversible jump Markov chain Monte Carlo (RJMCMC). The RJMCMC is used for distributions that have location and scale parameters or location-scale distribution, such as the Gaussian distribution family. In this research, we added an important step before beginning to use the RJMCMC method, namely the modification of the analyzed distribution into location-scale distribution. We called this the non-Gaussian RJMCMC (NG-RJMCMC) algorithm. The following steps are the same as for the RJMCMC. In this study, we applied it to the Weibull distribution. This will help many researchers in the field of survival analysis since most of the survival time distribution is Weibull. We transformed the Weibull distribution into a location-scale distribution, which is the extreme value (EV) type 1 (Gumbel-type for minima) distribution. Thus, for the mixture analysis, we call this EV-I mixture distribution. Based on the simulation results, we can conclude that the accuracy level is at minimum 95%. We also applied the EV-I mixture distribution and compared it with the Gaussian mixture distribution for enzyme, acidity, and galaxy datasets. Based on the Kullback-Leibler divergence (KLD) and visual observation, the EV-I mixture distribution has higher coverage than the Gaussian mixture distribution. We also applied it to our dengue hemorrhagic fever (DHF) data from eastern Surabaya, East Java, Indonesia. The estimation results show that the number of mixture components in the data is four; we also obtained the estimation results of the other parameters and labels for each observation. Based on the Kullback-Leibler divergence (KLD) and visual observation, for our data, the EV-I mixture distribution offers better coverage than the Gaussian mixture distribution.

Keywords: Reversible Jump Markov Chain Monte Carlo (RJMCMC); location-scale distribution; mixture distribution; Bayesian analysis

1. Introduction

Understanding the type of distribution of data is the first step in a data-driven statistical analysis, especially in a Bayesian analysis. This is very important because the distribution that we use must be as great as possible to cover the data we have. By knowing the distribution of the data, the error in the model can be minimized. However, it is not rare for the identified data to have a multimodal pattern. A model with data that has a multimodal pattern becomes imprecise when it is analyzed using a single mode distribution. This type of data is best modeled using mixture analysis. The most important thing in the mixture analysis is to determine the number of mixture components. If we know the correct number of mixture components, then the error in the resulting model can be minimized. In this way, our model will describe the real situation, because it is data-driven.

There are several methods for determining the number of mixture components in a dataset. Roeder [1] used a graphical technique to determine the number of mixture components in the Gaussian distribution. The Expectation-Maximization (EM) method was carried out by Carreira-Perpinán and Williams [2], while the Greedy EM method was carried out by Vlassis and Likas [3]. The likelihood ratio test (LRT) method has been carried out by Jeffries, Lo et al., and Kasahara and Shimotsu [4–6]. The LRT method was also carried out by McLachlan [7], but to the assessment null distribution, a bootstrapping approach was employed. A comparison between EM and LRT methods to determine the number of mixture components was carried out by Soromenho [8]. The determination of the number of mixture components using the inverse-Fisher information matrix was carried out by Bozdogan [9], and modification of the information matrix was carried out by Polymenis and Titterington [10]. Baudry et al. [11] used the Bayesian information criterion (BIC) method for clustering. In a comparison of several methods carried out by Lukočiene and Vermunt [12], these methods are the Akaike information criterion (AIC), BIC, AIC3, consistent AIC (CAIC), the information theoretic measure of complexity (ICOMP), and log-likelihood. Miller and Harrison [13] and Fearnhead [14] used the Dirichlet mixture process (DPM). Research conducted by McLachlan and Rathnayake [15] investigated several methods, including the LRT, resampling, information matrix, Clest method, and BIC methods. The methods mentioned above are used in the Gaussian distribution.

For complex computations, there is a method from the Bayesian perspective to determine the number of mixture components of the distribution, namely the reversible jump Markov chain Monte Carlo (RJMCMC). This method was initially introduced by Richardson and Green [16]. They used this method to determine the number of mixture components in the Gaussian distribution. This method is very flexible since it allows one to identify the number of components in data with a known or unknown number of components [17]. RJMCMC is also able to move between subspace parameters depending on the model, though the number of mixture components is different. Mathematically, the RJMCMC is simply derived in its random scan form; when the available moves are scanned systematically, the RJMCMC will be as valid as the idea from Metropolis-Hastings methods [18]. In real life, the RJMCMC can be applied to highly dimensional data [19]. Several studies have used the RJMCMC method for the Gaussian distribution [17,18,20–28].

Based on the advantages of the RJMCMC method as well as the previous studies mentioned above, this method will be powerful if it can be used not only in the Gaussian mixture distribution, because the Gaussian mixture is certainly not always the best approximation in some cases, although it may provide a reasonable approximation to many real-word distribution situations [29]. For example, if this method can be used on data with a domain more than zero, this would be very helpful to researchers in the fields of survival, reliability, etc. In most cases, survival data follow the Weibull distribution. Several studies of survival data using the Weibull distribution have been conducted [30–37]. Studies using the Weibull distribution for reliability were carried out by Villa-Covarrubias et al. and Zamora-Antuñano et al. [38,39]. In addition, some studies used Weibull mixture distribution for survival data [40–51]. According to the above studies that used the Weibull mixture distribution, the determination of the number of mixture components is a consideration. Determining the number of mixture components is an important first step in the mixture model. Therefore, we created a new algorithm that is a modification of the RJMCMC method for non-Gaussian distributions; we called it the non-Gaussian RJMCMC (NG-RJMCMC) algorithm. What we have done is to convert the original distribution of the data into a location-scale distribution, so that it has the same parameters as the Gaussian distribution, and finally the RJMCMC method can be applied. Thus, the number of mixture components can be determined before further analysis. In particular, our algorithm can help researchers in the field of survival by utilizing the Weibull mixture distribution. In

general, our algorithm can be applied to any mixture distribution, converting the original distribution into a location-scale family.

Several studies using the RJMCMC method on the Weibull distribution have been carried out. First, Newcombe et al. [52] used the RJMCMC method to implement a Bayesian variable selection for the Weibull regression model for breast cancer survival cases. The study conducted by Denis and Molinari [53] also used the RJMCMC method as covariate selection for the Weibull distribution in two datasets, namely Stanford heart transplant and lung cancer survival data. Mallet et al. [54] used the RJMCMC method to search for the best configuration of functions for Lidar waveforms. In their library of modeling functions, there are generalized Gaussian, Weibull, Nakagami, and Burr distributions. With their analysis, the Lidar waveform is a combination of these distributions. In our research, we have used the RJMCMC method on the Weibull distribution but from a different perspective, namely identifying multimodal data by determining the number of components, then determining the membership of each mixture component and determining the estimation results.

This paper is organized as follows. Section 2 introduces the basic formulation for the Bayesian mixture model and hierarchical model in general. Section 3 describes the location-scale distributions and NG-RJMCMC algorithm. Section 4 describes the transformation from the Weibull distribution to the location-scale distribution, determining the prior distributions, and explaining the move types in the RJMCMC method. Section 5 contains the simulation study. Section 6 provides misspecification cases as the opposite of a simulation study to strengthen the proposed method. Section 7 provides analysis results for applications of enzyme, acidity, and galaxy datasets, as well as our data, namely dengue hemorrhagic fever (DBD) in eastern Surabaya, East Java, Indonesia. The conclusions are given in Section 8.

2. Bayesian Model for Mixtures

2.1. Basic Formulation

For independent scalar or vector observations t_i , the basic mixture model can be written as in Equation (1):

$$t_i \sim \sum_{j=1}^k w_j f(\cdot | \theta_j)$$
 independently for $i = 1, 2, \cdots, n,$ (1)

where $f(\cdot|\theta)$ is a given parametric family of densities indexed by a scalar or vector parameter θ [16]. The purpose of this analysis is to infer to the unknown: the number of components (k), the weights of components (w_j), and the parameters of components (θ_j). Suppose a heterogeneous population consists of groups $j = 1, 2, \dots, k$ in proportion to w_j . The identity or label of the group is unknown for every observation. In this context, it is natural to create a group label z_i for the *i*-th observation as a latent allocation variable. The unobserved vector $\mathbf{z} = (z_1, z_2, \dots, z_n)$ is usually known as the "membership vector" of the mixture model [29]. Then, z_i is assumed to be drawn independently of the distributions

$$\Pr(z_i = j) = w_i$$
 for $j = 1, 2, \cdots, k$. (2)

2.2. Hierarchical Model in General

As explained in the previous subsection, there are three unknown components, k, w, and θ . In the Bayesian framework, these three components are drawn from the appropriate prior distribution [16]. The joint distribution of all variables can be written as in Equation (3):

$$\Pr(k, w, z, \theta, t) = \Pr(k)\Pr(w|k)\Pr(z|w, k)\Pr(\theta|z, w, k)\Pr(t|\theta, z, w, k),$$
(3)

where $w = (w_j)_{j=1,2,\dots,k'} z = (z_i)_{i=1,2,\dots,n'}$, $\theta = (\theta_j)_{j=1,2,\dots,k'}$, $t = (t_i)_{i=1,2,\dots,n'}$, and $\Pr(\cdot|\cdot)$ are the generic conditional distributions [16]. Then, Equation (3) may be forced naturally, so that $\Pr(\theta|z, w, k) = \Pr(\theta|k)$ and $\Pr(t|\theta, z, w, k) = \Pr(t|\theta, z)$ [16]. Therefore, for the

Bayesian hierarchical model, the joint distribution in Equation (3) can be simplified into Equation (4):

$$\Pr(k, w, z, \theta, t) = \Pr(k)\Pr(w|k)\Pr(z|w, k)\Pr(\theta|k)\Pr(t|\theta, z).$$
(4)

Next, we add an additional layer to the hierarchy, namely hyperparameters γ , δ , and ξ for k, w, and θ , respectively. These hyperparameters will be drawn from independent hyperpriors. The joint distribution of all variables is then given in Equation (5):

$$\Pr(\gamma, \delta, \xi, k, w, z, \theta, t) = \Pr(\gamma) \Pr(\delta) \Pr(\xi) \Pr(k|\gamma) \Pr(w|k, \delta) \Pr(z|w, k) \Pr(\theta|k, \xi) \Pr(t|\theta, z)$$

3. Non-Gaussian Reversible Jump Markov Chain Monte Carlo (NG-RJMCMC) Algorithm for Mixture Model

3.1. The Family of Location-Scale Distributions

A random variable *T* is defined as belonging to the location-scale family when its cumulative distribution function (CDF)

$$F_T(t|\mu,\sigma) = \Pr(T \le t|\mu,\sigma)$$

is a function only of $\frac{t-\mu}{\sigma}$, as in Equation (6) [55]:

$$F_T(t|\mu,\sigma) = F\left(\frac{t-\mu}{\sigma}\right); \mu \in \mathbb{R}, \sigma > 0,$$
(6)

where $F(\cdot)$ is a distribution without other parameters. The two-dimensional parameter (μ, σ) is called the location-scale parameter, with μ being the location parameter and σ being the scale parameter. For fixed $\sigma = 1$, we have a subfamily that is a location family with a parameter μ , and for fixed $\mu = 0$, we have a scale family with a parameter σ . If *T* is continuous with the probability density function (p.d.f.)

$$f_T(t|\mu,\sigma) = rac{dF_T(t|\mu,\sigma)}{dt}$$

then (μ, σ) is a location-scale parameter for *T* if (and only if)

$$f_T(t|\mu,\sigma) = \frac{1}{\sigma}g\left(\frac{t-\mu}{\sigma}\right),\tag{7}$$

where the functional form *g* is completely specified, but the location and scale parameters, μ and σ , of $f_T(t|\mu, \sigma)$ are unknown, and $g(\cdot)$ is the standard form of the density $f_T(t|\mu, \sigma)$ [56,57].

3.2. NG-RJMCMC Algorithm

In the mixture model analysis, one method is known, namely the reversible jump Markov chain Monte Carlo (RJMCMC). The RJMCMC method can be used to estimate an unknown quantity, such as the number of mixture components, the weights of mixture components, and the distribution parameters of mixture components. For its usefulness in determining the number of mixture components in the indicated multimodal data, RJMCMC has been extensively used. For the Gaussian distribution, this was initially carried out by Richardson and Green [16]. Over time, RJMCMC can be used for distributions other than Gaussian. RJMCMC for general beta distribution was carried out in two studies by Bouguila and Elguebaly [29,58]. In their research, the general beta distribution consists of four parameters, namely the lower limit, the upper limit, and two shape parameters. Then, to use the RJMCMC algorithm, they obtained the location-scale parameterization for this distribution. Another study followed the location-scale parameterization for the RJMCMC method, namely research on the symmetry gamma distribution [59].

(5)

Not only in the RJMCMC method but also in other methods, the location-scale parameterization is carried out in the mixture analysis. First, research on the exponential and Gaussian distribution using the Dirichlet process mixture was carried out by Jo et al. [60]. Secondly, research on the asymmetric Laplace error distribution using the likelihood-based approach was carried out by Kobayashi and Kozumi [61]. Finally, research on the exponential distribution using the Gibbs sampler was carried out by Gruet et al. [62]. Based on the studies mentioned above, in the mixture analysis using any method, it will be easier if the distribution used in the research follows the location-scale (family) parameterization. Thus, we given Algorithm 1 as a modification of the RJMCMC algorithm.

Algorithm 1. NG-RJMCMC Algorithm

- 1. Modify the distribution to be analyzed into a member of the location-scale family, determine:
 - (a) form of transformation, and
 - (b) location-scale parameter
- 2. Determine the appropriate priors for:
 - (a) component weights (*w*),
 - (b) location-scale parameter (μ , σ), and
 - (c) latent allocation variable (z)
- 3. Do all six types of RJMCMC moves by Richardson and Green [16]. A sweep is defined as a complete run on these six moves [16]:
 - (a) updating the component weights (*w*),
 - (b) updating the location-scale parameter (μ, σ) ,
 - (c) updating the latent allocation variable (z),
 - (d) updating the hyperparameters (γ , δ , and ξ). Moves (a), (b), (c), and (d) can therefore be performed in parallel
 - (e) do the splitting or combining mixture component
 - (f) do the birth or death of an empty component.

Letting Δ denote the state variable (in this study, Δ be the complete set of unknowns (μ, σ, k, w, z)), and $p(\Delta)$ be the target probability measure (the posterior distribution), we consider a countable family of move types, indexed by $m = 1, 2, \cdots$. When the current state is Δ , a move type m and destination Δ^* are proposed, with joint distribution given by $q_m(\Delta, \Delta^*)$. The move is accepted with probability

$$\alpha_m(\Delta, \Delta^*) = \min\left\{1, \frac{p(\Delta^*)q_m(\Delta^*, \Delta)}{p(\Delta)q_m(\Delta, \Delta^*)}\right\}.$$
(8)

If Δ^* has a higher dimensional space than Δ , it is possible to create a vector of continuous random variables **u**, independent of Δ [16]. Then, the new state Δ^* is set by an invertible deterministic function of Δ and **u**: $f(\Delta, \mathbf{u})$. Then, the acceptance probability in Equation (8) can be rewritten as in Equation (9):

$$\alpha_m(\Delta, \Delta^*) = \min\left\{1, \frac{p(\Delta^*)r_m(\Delta^*)}{p(\Delta)r_m(\Delta)q(\mathbf{u})} \left| \frac{\partial \Delta^*}{\partial(\Delta, \mathbf{u})} \right|\right\},\tag{9}$$

where $r_m(\Delta)$ is the probability of choosing move type *m* when in the state Δ , and $q(\mathbf{u})$ is the p.d.f. of \mathbf{u} . The last term, $\left|\frac{\partial \Delta^*}{\partial(\Delta, \mathbf{u})}\right|$, is the determinant of Jacobian matrix resulting from modifying the variable from (Δ, \mathbf{u}) to Δ^* .

4. Bayesian Analysis of Weibull Mixture Distribution Using NG-RJMCMC Algorithm

4.1. Change the Weibull Distribution into a Member of the Location-Scale Family

4.1.1. Form of Transformation and Location-Scale Parameter of Weibull Distribution

If the random variable *T* follows the Weibull distribution with the shape parameter $\eta > 0$ and the scale parameter $\lambda > 0$, $T \sim \text{Weibull}(\eta, \lambda)$, then the p.d.f. is given by Equation (10) [63]:

$$f_T(t|\eta,\lambda) = \begin{cases} \frac{\eta}{\lambda} \left(\frac{t}{\lambda}\right)^{\eta-1} e^{-\left(\frac{t}{\lambda}\right)^{\eta}} & ,t \ge 0\\ 0 & ,t < 0 \end{cases}$$
(10)

and CDF is given by Equation (11) [64]:

$$F_T(t|\eta,\lambda) = \begin{cases} 1 - e^{-(\frac{t}{\lambda})^{\eta}} & ,t \ge 0\\ 0 & ,t < 0 \end{cases}.$$
 (11)

Equation (11) can be rewritten as Equation (12):

$$F_T(t|\eta,\lambda) = \begin{cases} 1 - e^{-\left(\frac{t-0}{\lambda}\right)^{\eta}} , t \ge 0\\ 0 , t < 0\\ = F_T\left(\frac{t-\mu}{\lambda}\right)_{\mu=0}. \end{cases}$$
(12)

Based on Equation (12) and the explanation in Section 3.1, it can be concluded that the Weibull distribution is a member of the scale family. To facilitate the analysis of the Weibull distribution, it is necessary to transform it into a location-scale distribution.

If the random variable *T* is transformed into a new variable, $Y = \ln T$, then *Y* has an extreme value (EV) type 1 (Gumbel-type for minima) distribution with the location parameter $\mu = \ln \lambda$ and the scale parameter $\sigma = \frac{1}{\eta}$ [64,65] (this explanation can be seen in Appendix A). Therefore, the p.d.f. and CDF for $Y \sim \text{EV} - I(\mu = \ln \lambda, \sigma = \frac{1}{\eta})$ are given by Equations (13) and (14), respectively [56,66,67]:

$$f_{Y}(y|\mu,\sigma) = \frac{1}{\sigma} \exp\left[\left(\frac{y-\mu}{\sigma}\right) - \exp\left(\frac{y-\mu}{\sigma}\right)\right]$$
(13)

$$F_{Y}(y|\mu,\sigma) = 1 - \exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right],$$
(14)

where $-\infty < y < \infty$, $-\infty < \mu < \infty$, and $0 < \sigma < \infty$.

4.1.2. Finite EV-I Mixture Distribution

Based on the explanation provided in the previous subsection, $T \sim \text{Weibull}(\eta, \lambda)$ as a member-scale family can be transformed into a member of the location-scale family $Y \sim \text{EV} - \text{I}\left(\mu = \ln \lambda, \sigma = \frac{1}{\eta}\right)$, where $Y = \ln T$. It is well-known that $T \sim \text{Weibull}(\eta, \lambda)$ and $Y \sim \text{EV} - \text{I}\left(\mu = \ln \lambda, \sigma = \frac{1}{\eta}\right)$ are equivalent models [68]. As $Y \sim \text{EV} - \text{I}\left(\mu = \ln \lambda, \sigma = \frac{1}{\eta}\right)$ belongs to the location-scale family, it is sometimes easier to work with $Y \sim \text{EV} - \text{I}\left(\mu = \ln \lambda, \sigma = \frac{1}{\eta}\right)$ rather than $T \sim \text{Weibull}(\eta, \lambda)$ [68], especially in the analysis of the mixture model. Consequently, the next analysis will use the variable Y. The EV-I mixture distribution with k components is defined as in Equation (15):

$$f(y|\Theta) = \sum_{j=1}^{k} w_j f(y|\mu_j, \sigma_j), \qquad (15)$$

where $\Theta = (\theta, w)$ refers to the complete set of parameters to be estimated, where $\theta = (\mu_1, \sigma_1; \mu_2, \sigma_2; \cdots; \mu_k, \sigma_k)$.

4.2. Determine the Appropriate Priors

4.2.1. Hierarchical Model

According to the general form of the hierarchical model in Section 2.2, a hierarchical model for the EV-I mixture distribution can be written as in Equation (16):

$$\Pr(\gamma, \delta, \xi, k, w, z, \theta, y) = \Pr(\gamma) \Pr(\delta) \Pr(\xi) \Pr(k|\gamma) \Pr(w|k, \delta) \Pr(z|w, k) \Pr(\theta|k, \xi) \Pr(y|\theta, z),$$
(16)

where *k* is the number of components, $w = (w_j)_{j=1,2,\cdots,k}$ is the weights of the components, $z = (z_i)_{i=1,2,\cdots,n}$ is the latent allocation variable, $\theta = (\theta_j) = (\mu_j, \sigma_j)_{j=1,2,\cdots,k}$ is the location-scale parameter, $y = (y_i)_{i=1,2,\cdots,n}$ is the data with $Y \sim \text{EV} - I(\mu, \sigma)$, and γ , δ , and ξ are the hyperparameters for *k*, *w*, and θ , respectively. If we condition on *z*, the distribution of y_i is given by the z_i -th component in the mixture, so that $\Pr(y|\theta, z) = \prod_{i=1}^n \Pr(y_i|\theta_{z_i})$ [29]. Then, the final form of the joint distribution can be found via Equation (17):

$$\Pr(\gamma, \delta, \xi, k, w, z, \theta, y) = \Pr(\gamma) \Pr(\delta) \Pr(\xi) \Pr(k|\gamma) \Pr(w|k, \delta) \Pr(z|w, k) \Pr(\theta|k, \xi) \prod_{i=1} \Pr(y_i|\theta_{z_i}).$$
(17)

4.2.2. Priors and Posteriors

In this section, we define the priors. In the hierarchical model in Equation (16), for each parameter, we assume that the priors are drawn independently. Based on research conducted by Yoon et al. [69], Coles and Tawn [70] and Tancredi et al. [71], the priors for the location and scale parameters in the extreme value distribution are flat. In research by Yoon et al. [69], they chose the adoption of near-flat priors for the location and scale parameters. In research by Coles and Tawn [70], the location and scale parameter priors are almost noninformative: the prior for μ is extremely flat, while that for σ resembles $\frac{1}{\sigma}$. Based on research by Tancredi et al. [71], they have chosen a uniform distribution for the location and scale parameters. Therefore, in this study, the Gaussian distribution (with the large variance) with mean ε and variance ζ^2 was selected as a prior for location parameter μ . Thus, μ_i for each component is given by

$$f\left(\mu_{j}\left|\varepsilon,\zeta^{2}\right.\right) = \frac{1}{\sqrt{2\pi\zeta^{2}}} \exp\left\{-\frac{1}{2}\left(\frac{\mu_{j}-\varepsilon}{\zeta}\right)^{2}\right\}.$$
(18)

Since the scale parameter σ controls the dispersion of the distribution, an appropriate prior is an inverse gamma distribution with the shape and scale parameters are ϑ and ω , respectively. This prior selection is supported by Richardson and Green [16], and Bouguila and Elguebaly [29]. Thus, σ_i for each component is given by

$$f(\sigma_{j}|\vartheta,\varpi) = \frac{\varpi^{\vartheta} \exp\left(-\frac{\varpi}{\sigma_{j}}\right)}{\Gamma(\vartheta)\sigma_{j}^{\vartheta+1}},$$
(19)

where $\omega \sim \text{Gamma}(g, h)$. Used Equations (18) and (19), we get

$$f(\theta|k,\xi) = \prod_{j=1}^{k} f(\mu_{j}|\varepsilon,\zeta^{2}) f(\sigma_{j}|\theta,\omega)$$

$$= \prod_{j=1}^{k} \left[\frac{1}{\sqrt{2\pi\zeta^{2}}} \exp\left\{ -\frac{1}{2} \left(\frac{\mu_{j}-\varepsilon}{\zeta} \right)^{2} \right\} \right] \left[\frac{\omega^{\theta} \exp\left(-\frac{\omega}{\sigma_{j}} \right)}{\Gamma(\theta)\sigma_{j}^{\theta+1}} \right]$$

$$= \left(\frac{\omega^{\theta}}{\Gamma(\theta)\sqrt{2\pi\zeta^{2}}} \right)^{k} \prod_{j=1}^{k} \frac{\exp\left\{ -\frac{1}{2} \left(\frac{\mu_{j}-\varepsilon}{\zeta} \right)^{2} \right\} \exp\left(-\frac{\omega}{\sigma_{j}} \right)}{\sigma_{j}^{\theta+1}}$$

$$= \frac{\omega^{\theta k}}{\Gamma(\theta)^{k}(2\pi)^{\frac{k}{2}} \zeta^{k}} \prod_{j=1}^{k} \frac{\exp\left\{ -\frac{1}{2} \left(\frac{\mu_{j}-\varepsilon}{\zeta} \right)^{2} - \frac{\omega}{\sigma_{j}} \right\}}{\sigma_{j}^{\theta+1}}$$
(20)

Therefore, the hyperparameter ξ in Equation (17) is actually (ε , ζ^2 , ϑ , ω). Thus, according to Equation (20) and the joint distribution in Equation (17), the fully conditional posterior distribution for μ_i and σ_j are

$$f(\mu_{j}|\cdots) \propto \prod_{j=1}^{\kappa} f(\mu_{j}|\varepsilon,\zeta^{2}) f(\sigma_{j}|\vartheta,\varpi) \prod_{i=1}^{n} f(y_{i}|\vartheta_{z_{i}}) \propto f(\mu_{j}|\varepsilon,\zeta^{2}) \prod_{i=1}^{n} f(y_{i}|\vartheta_{z_{i}}) \propto \left[\frac{1}{\sqrt{2\pi\zeta^{2}}} \exp\left\{-\frac{1}{2}\left(\frac{\mu_{j}-\varepsilon}{\zeta}\right)^{2}\right\}\right] \times \prod_{i=1}^{n} \frac{1}{\sigma_{j}} \exp\left\{\left(\frac{y_{i}-\mu_{j}}{\sigma_{j}}\right) - \exp\left(\frac{y_{i}-\mu_{j}}{\sigma_{j}}\right)\right\} \propto \left[\frac{1}{\sqrt{2\pi\zeta^{2}}} \exp\left\{-\frac{1}{2}\left(\frac{\mu_{j}-\varepsilon}{\zeta}\right)^{2}\right\}\right] \times \left(\frac{1}{\sigma_{j}}\right)^{n_{j}} \prod_{z_{i}=j} \exp\left\{\left(\frac{y_{i}-\mu_{j}}{\sigma_{j}}\right) - \exp\left(\frac{y_{i}-\mu_{j}}{\sigma_{j}}\right)\right\}$$
(21)

and

$$f(\sigma_{j}|\cdots) \propto \prod_{j=1}^{k} f(\mu_{j}|\varepsilon,\zeta^{2}) f(\sigma_{j}|\vartheta,\varpi) \prod_{i=1}^{n} f(y_{i}|\theta_{z_{i}})$$

$$\propto f(\sigma_{j}|\vartheta,\varpi) \prod_{i=1}^{n} f(y_{i}|\theta_{z_{i}})$$

$$\propto \left[\frac{\omega^{\vartheta} \exp\left(-\frac{\omega}{\sigma_{j}}\right)}{\Gamma(\vartheta)\sigma_{j}^{\vartheta+1}}\right] \times \prod_{i=1}^{n} \frac{1}{\sigma_{j}} \exp\left\{\left(\frac{y_{i}-\mu_{j}}{\sigma_{j}}\right) - \exp\left(\frac{y_{i}-\mu_{j}}{\sigma_{j}}\right)\right\}$$

$$\propto \left[\frac{\omega^{\vartheta} \exp\left(-\frac{\omega}{\sigma_{j}}\right)}{\Gamma(\vartheta)\sigma_{j}^{\vartheta+1}}\right] \times \left(\frac{1}{\sigma_{j}}\right)^{n_{j}} \prod_{z_{i}=j} \exp\left\{\left(\frac{y_{i}-\mu_{j}}{\sigma_{j}}\right) - \exp\left(\frac{y_{i}-\mu_{j}}{\sigma_{j}}\right)\right\},$$
(22)

where $n_j = \#\{i : z_i = j\}$ represents the number of vectors in the cluster *j*, and we use '|··· ' to designate conditioning on all other variables.

As we know that the weights of components $w = (w_j)_{j=1,2,\dots,k}$ are defined on the simplex $\left\{ (w_1, w_2, \dots, w_k) : \sum_{j=1}^{k-1} w_j < 1 \right\}$, the appropriate prior for the weights of components is a Dirichlet distribution with parameters $\delta = (\delta_1, \delta_2, \dots, \delta_k)$ [72], with the p.d.f. as in Equation (23):

$$f(w|k,\delta) = \frac{1}{B(\delta)} \prod_{j=1}^{k} w_j^{\delta_j - 1}$$

$$= \frac{\Gamma\left(\sum_{j=1}^{k} \delta_j\right)}{\prod_{j=1}^{k} \Gamma(\delta_j)} \prod_{j=1}^{k} w_j^{\delta_j - 1},$$
(23)

where $B(\delta) = \frac{\prod_{j=1}^{k} \Gamma(\delta_j)}{\Gamma\left(\sum_{j=1}^{k} \delta_j\right)}$. According to Equation (2), we also have

$$f(z|w,k) = \prod_{j=1}^{k} w_j^{n_j}.$$
(24)

Using Equations (23) and (24), and our joint distribution from Equation (17), we obtained

$$f(w|\dots) \propto f(w|k,\delta)f(z|w,k)$$

$$\propto \left[\frac{\Gamma\left(\sum_{j=1}^{k}\delta_{j}\right)}{\prod_{j=1}^{k}\Gamma(\delta_{j})}\prod_{j=1}^{k}w_{j}^{\delta_{j}-1}\right]\left[\prod_{j=1}^{k}w_{j}^{n_{j}}\right]$$

$$\propto \prod_{j=1}^{k}w_{j}^{n_{j}+\delta_{j}-1},$$
(25)

where $\frac{\Gamma\left(\sum_{j=1}^{k} \delta_{j}\right)}{\prod_{i=1}^{k} \Gamma(\delta_{j})}$ is a constant. This is in fact proportional to a Dirichlet distribution with pa-

rameters $(\delta_1 + n_1, \delta_2 + n_2, \dots, \delta_k + n_k)$. Using Equations (2) and (17), we get the posterior for the allocation variables

$$f(z_i = j | \dots) \propto w_j \frac{1}{\sigma_j} \exp\left[\left(\frac{y_i - \mu_j}{\sigma_j}\right) - \exp\left(\frac{y_i - \mu_j}{\sigma_j}\right)\right].$$
 (26)

The last, proper prior for *k* is the Poisson distribution with hyperparameter γ [16], then the p.d.f. for *k* can be seen as in Equation (27):

$$f(k|\gamma) = \frac{\gamma^k e^{-\gamma}}{k!}.$$
(27)

Our hierarchical model can be displayed as a directed acyclic graph (DAG), as shown in Figure 1.



Figure 1. Graphical representation of the Bayesian hierarchical finite EV-I mixture distribution. Nodes in this graph represent random variables, green boxes are fixed hyperparameters, boxes indicate repetition (with the number of repetitions in the upper left), arrows describe conditional dependencies between variables, and the red arrow describes a variable transformation.

4.3. RJMCMC Move Types for EV-I Mixture Distribution

It was mentioned in Section 3.2 that the Algorithm 1, which moves (a), (b), (c), and (d), can be run in parallel. This section will explain in more detail moves (e) and (f), namely split and combine moves, and birth and death moves.

4.3.1. Split and Combine Moves

For move (e), we choose between split or combine, with the probabilities b_k and $d_k = 1 - b_k$, respectively, depending on k. Note that $d_1 = 0$ and $b_{k_{\text{max}}} = 0$, where k_{max} is the maximum value for k; otherwise, we choose $b_k = d_k = 0.5$, for $k = 1, 2, \dots, k_{\text{max}} - 1$. The combining proposal works as follows: choose two components j_1 and j_2 , where $\mu_1 < \mu_2$ with no other $\mu_j \in [\mu_1, \mu_2]$. If these components are combined, we reduce k by 1, which forms a new component j^* containing all the observation previously allocated to j_1 and j_2 , and then creates values for w_{j^*} , μ_{j^*} , and σ_{j^*} by preserving the first two moments, as follows:

$$\begin{array}{c}
w_{j^{*}} = w_{j_{1}} + w_{j_{2}} \\
w_{j^{*}} \mu_{j^{*}} = w_{j_{1}} \mu_{j_{1}} + w_{j_{2}} \mu_{j_{2}} \\
w_{j^{*}} \left(\mu_{j^{*}}^{2} + \sigma_{j^{*}}\right) = w_{j_{1}} \left(\mu_{j_{1}}^{2} + \sigma_{j_{1}}\right) + w_{j_{2}} \left(\mu_{j_{2}}^{2} + \sigma_{j_{2}}\right).
\end{array}$$
(28)

The splitting proposal works as follows: a random j^* component is selected then split into two new components, j_1 and j_2 , with the weights and parameters $(w_{j_1}, \mu_{j_1}, \sigma_{j_1})$ and $(w_{j_2}, \mu_{j_2}, \sigma_{j_2})$, respectively, conforming to Equation (28). Based on this information, we have three degrees of freedom, so we generate three random numbers $\mathbf{u} = (u_1, u_2, u_3)$, where $u_1 \sim \text{Beta}(2, 2)$, $u_2 \sim \text{Beta}(2, 2)$, and $u_3 \sim \text{Beta}(1, 1)$ [16]. Then, split transformations are defined as follows:

$$w_{j_{1}} = w_{j^{*}}u_{1}, w_{j_{2}} = w_{j^{*}}(1 - u_{1})$$

$$\mu_{j_{1}} = \mu_{j^{*}} - u_{2}\sqrt{\sigma_{j^{*}}}\sqrt{\frac{w_{j_{2}}}{w_{j_{1}}}}, \mu_{j_{2}} = \mu_{j^{*}} + u_{2}\sqrt{\sigma_{j^{*}}}\sqrt{\frac{w_{j_{1}}}{w_{j_{2}}}}$$

$$\sigma_{j_{1}} = u_{3}(1 - u_{2}^{2})\sigma_{j^{*}}\frac{w_{j^{*}}}{w_{j_{1}}}, \sigma_{j_{2}} = (1 - u_{3})(1 - u_{2}^{2})\sigma_{j^{*}}\frac{w_{j^{*}}}{w_{j_{2}}}.$$

$$(29)$$

Then, we compute the acceptance probabilities of split and combine moves: $\min\{1, A\}$ and $\min\{1, A^{-1}\}$, respectively. According to Equation (9), we obtain *A* as in Equation (30):

$$A = \frac{\Pr(z,w,k+1,\theta,\varepsilon,\zeta,\theta,\omega|y)d_{k+1}}{\Pr(z,w,k,\theta,\varepsilon,\zeta,\theta,\omega|y,b_k} p_{alloc}q(\mathbf{u})} \left| \frac{\partial \Delta^*}{\partial (\Delta,\mathbf{u})} \right|$$

= likelihood ratio $\left[\frac{f(k+1)}{f(k)} \right] (k+1) \frac{w_{j_1}^{\delta+l_1-1} w_{j_2}^{\delta+l_2-1}}{w_{j^*}^{\delta+l_1+l_2-1} B(\delta,k\delta)}$
 $\times \left(\frac{1}{2\pi\zeta^2} \right)^{\frac{1}{2}} e^{-\frac{1}{2\zeta^2} [(\mu_{j_1}-\varepsilon)^2 + (\mu_{j_2}-\varepsilon)^2 - (\mu_{j^*}-\varepsilon)^2]}$
 $\times \frac{\omega^{\theta}}{\Gamma(\theta)} \left(\frac{\sigma_{j^*}}{\sigma_{j_1}\sigma_{j_2}} \right)^{\theta+1} \exp\left(-\omega\left(\frac{1}{\sigma_{j_1}} + \frac{1}{\sigma_{j_2}} - \frac{1}{\sigma_{j^*}} \right) \right)$
 $\times \frac{d_{k+1}}{b_k p_{alloc}} \frac{g_{2,2}(u_1)g_{2,2}(u_2)g_{1,1}(u_3)}{g_{2,2}(u_1)g_{2,1}(u_{3})\sigma_{j^*}}$ (30)

where *k* is the number of components before the split, l_1 and l_2 are the numbers of observations proposed to be assigned to j_1 and j_2 , $B(\cdot, \cdot)$ is the beta function, p_{alloc} is the probability that this particular allocation is made, $g_{p,q}$ is the beta (p,q) density, the (k + 1)-factor in the second line is the ratio $\frac{(k+1)!}{k!}$ from the order statistics densities for the location-scale parameters (μ, σ) [16], and the other terms have been fully explained in Appendices B and C.

4.3.2. Birth and Death Moves

The following is an explanation of move (f), namely birth and death moves. These moves are simpler than split and combine moves [16]. The first step consists of making a random choice between birth and death, with the same probabilities b_k and d_k as stated above. For birth, the proposed new component has parameters μ_{j^*} and $\sigma_{j^*}^2$, which are generated from the associated prior distributions shown in Equations (18) and (19), respectively. The weight of the new component w_{j^*} follows a beta distribution, $w_{j^*} \sim \text{Beta}(1, k)$. To remain valid for the constraint $\sum_{j=1}^k w_j + w_{j^*} = 1$, the previous weights w_j for $j = 1, 2, \dots, k$

must be rescaled by multiplying all by $(1 - w_{j^*})$. Therefore, $(1 - w_{j^*})^k$ is the determinant of Jacobian matrix corresponding to the birth move. For the opposite move, namely the death move, we randomly choose any empty component to remove. This step always

death move, we randomly choose any empty component to remove. This step always considers the constraint that the remaining weights are rescaled to sum 1. The acceptance probabilities of the birth and death moves: min{1, A} and min{1, A^{-1} }, respectively. According to Equation (9), we obtain A as in Equation (31):

$$A = \frac{\Pr(k+1)}{\Pr(k)} \frac{1}{B(k\delta,\delta)} w_{j^*}^{\delta-1} (1-w_{j^*})^{n+k\delta-k} (k+1) \frac{d_{k+1}}{b_k(k_0+1)} \frac{1}{g_{1,k}(w_{j^*})} (1-w_{j^*})^k, \quad (31)$$

where k_0 is the number of empty components before birth, and $B(\cdot, \cdot)$ is the Beta function.

5. Simulation Study

In this section, we have 16 scenarios, namely Weibull mixture distribution with two components, three components, four components, and five components, each of which is generated with a sample of 125, 250, 500, and 1000 per component. Detailed descriptions of each scenario are given in Table 1, where the "Parameter of EV-I distribution" column is transformed from the "Parameter of Weibull distribution" column.

Table 1. Sixteen scenarios of the Weibull mixture distribution and their transformation into the EV-I mixture distribution.

Scenario	Number of	Component	Number of	Parameter Distr	r of Weibull ibution	Paramete Distri	er of EV-I bution
	Components		Generated Data	η	λ	μ	σ
1	2	1st	125	1	exp(-2)	-2	1
1	2	2nd	125	1/0.5	exp(0)	0	0.5
2	2	1st	250	1	exp(-2)	-2	1
2	2	2nd	250	1/0.5	exp(0)	0	0.5
2	2	1st	500	1	exp(-2)	-2	1
3	2	2nd	500	1/0.5	exp(0)	0	0.5
4	2	1st	1000	1	exp(-2)	-2	1
4	2	2nd	1000	1/0.5	exp(0)	0	0.5
		1st	125	1	exp(-2)	-2	1
5	3	2nd	125	1/0.5	exp(0)	0	0.5
		3rd	125	1/1.5	exp(3)	3	1.5
		1st	250	1	exp(-2)	-2	1
6	3	2nd	250	1/0.5	exp(0)	0	0.5
		3rd	250	1/1.5	exp(3)	3	1.5
		1st	500	1	exp(-2)	-2	1
7	3	2nd	500	1/0.5	exp(0)	0	0.5
		3rd	500	1/1.5	exp(3)	3	1.5
		1st	1000	1	exp(-2)	-2	1
8	3	2nd	1000	1/0.5	exp(0)	0	0.5
		3rd	1000	1/1.5	exp(3)	3	1.5
		1st	125	1	exp(-2)	-2	1
9	4	2nd	125	1/0.5	exp(0)	0	0.5
2	1	3rd	125	1/1.5	exp(3)	3	1.5
		4th	125	1/0.2	exp(5)	5	0.2
		1st	250	1	exp(-2)	-2	1
10	4	2nd	250	1/0.5	exp(0)	0	0.5
10	1	3rd	250	1/1.5	exp(3)	3	1.5
_		4th	250	1/0.2	exp(5)	5	0.2
		1st	500	1	exp(-2)	-2	1
11	4	2nd	500	1/0.5	exp(0)	0	0.5
11	Т	3rd	500	1/1.5	exp(3)	3	1.5
	-	4th	500	1/0.2	exp(5)	5	0.2

Scenario	Number of	Component	Number of	Parameter Distr	r of Weibull ibution	Paramete Distri	er of EV-I bution
	Components		Generated Data	η	λ	μ	σ
		1st	1000	1	exp(-2)	-2	1
12	4	2nd	1000	1/0.5	exp(0)	0	0.5
12	Ĩ	3rd	1000	1/1.5	exp(3)	3	1.5
		4th	1000	1/0.2	exp(5)	5	0.2
		1st	125	1	exp(-2)	-2	1
		2nd	125	1/0.5	exp(0)	0	0.5
13	5	3rd	125	1/1.5	exp(3)	3	1.5
		4th	125	1/0.2	exp(5)	5	0.2
		5th	125	1	exp(7)	7	1
		1st	250	1	exp(-2)	-2	1
		2nd	250	1/0.5	exp(0)	0	0.5
14	5	3rd	250	1/1.5	exp(3)	3	1.5
		4th	250	1/0.2	exp(5)	5	0.2
		5th	250	1	exp(7)	7	1
		1st	500	1	exp(-2)	-2	1
		2nd	500	1/0.5	exp(0)	0	0.5
15	5	3rd	500	1/1.5	exp(3)	3	1.5
		4th	500	1/0.2	exp(5)	5	0.2
		5th	500	1	exp(7)	7	1
		1st	1000	1	exp(-2)	-2	1
		2nd	1000	1/0.5	exp(0)	0	0.5
16	5	3rd	1000	1/1.5	exp(3)	3	1.5
		4th	1000	1/0.2	exp(5)	5	0.2
		5th	1000	1	exp(7)	7	1

Table 1. Cont.

In these scenarios, our specific choices for the hyperparameters were $\zeta = R$, $\vartheta = 2$, g = 0.2, $h = \frac{10}{R^2}$, $\delta = 1$, and $k_{\text{max}} = 30$, where *R* and ε are the length and midpoint (median) of the observed data, respectively (see Richardson and Green [16]). Based on the selection of the hyperparameters, we performed an analysis with 200,000 sweeps. With these 200,000 sweeps, we got a value of *k* as high as 200,000. From this, we took the most frequently occurring (mode) of *k*. Then, we replicated this step 500 times. Thus, we already had one mode *k* for each replication. Finally, we had 500 *k* and we calculated them as a percentage. The results of grouping the mixture components can be seen in Table 2, while the parameter estimation results can be seen in Table 3. Based on Table 2, each scenario provides a grouping with an accuracy level of at least 95%; the accuracy level is not 100% only when the sample size is 125 per component. Based on Table 3, it can be seen that the estimated parameters are close to their real parameters for all scenarios. Note: details of the computer and time required for the running simulation study are given in Appendix D.

Scenario	Number of Components	Sample Size (per Component)	k=1(%)	k = 2(%)	k = 3(%)	k = 4(%)	k = 5(%)
1	2	125	0	100	0	0	0
2	2	250	0	100	0	0	0
3	2	500	0	100	0	0	0
4	2	1000	0	100	0	0	0
5	3	125	0.8	4.2	95	0	0
6	3	250	0	0	100	0	0
7	3	500	0	0	100	0	0
8	3	1000	0	0	100	0	0
9	4	125	0	0	3.2	96.8	0
10	4	250	0	0	0	100	0
11	4	500	0	0	0	100	0
12	4	1000	0	0	0	100	0
13	5	125	0	0	0	2.4	97.6
14	5	250	0	0	0	0	100
15	5	500	0	0	0	0	100
16	5	1000	0	0	0	0	100

Table 2. Summary of the results of grouping the EV-I mixture distribution with 200,000 sweeps and replicated 500 times.

Table 3. Parameter estimation of 16 scenarios, where μ , σ , and w are the real parameters and $\hat{\mu}$, $\hat{\sigma}$, and \hat{w} are the estimated parameters.

				Par. I	ameter of Distributio	EV-I on	Estimate	ed Paramete Distribution	r of EV-I
Scenario	Number of Components	Component	Number of Generated Data	μ	σ	w	û	ô	ŵ
		1st	125	-2	1	0.5	-1.9483	1.019	0.498
1	2	2nd	125	0	0.5	0.5	-0.2771	0.5077	0.502
	_	1st	250	-2	1	0.5	-2.0034	0.9911	0.4978
2	2	2nd	250	0	0.5	0.5	0.0266	0.4911	0.5022
	2	1st	500	-2	1	0.5	-2.0009	0.9959	0.5024
3	2	2nd	500	0	0.5	0.5	-0.037	0.4942	0.4976
	2	1st	1000	-2	1	0.5	-2	1.0023	0.4998
4	Z	2nd	1000	0	0.5	0.5	-0.002	0.5082	0.5002
		1st	125	-2	1	0.3333	-1.9736	1.004	0.336
5	3	2nd	125	0	0.5	0.3333	0.005	0.5009	0.328
		3rd	125	3	1.5	0.3333	2.9974	1.4818	0.336
		1st	250	-2	1	0.3333	-1.9734	1.0019	0.3376
6	3	2nd	250	0	0.5	0.3333	0.0179	0.5115	0.3296
		3rd	250	3	1.5	0.3333	3.038	1.5129	0.3328
		1st	500	-2	1	0.3333	-1.9598	0.9935	0.332
7	3	2nd	500	0	0.5	0.3333	-0.0049	0.5162	0.3295
		3rd	500	3	1.5	0.3333	3.021	1.5046	0.3385
		1st	1000	-2	1	0.3333	-2.0062	1	0.3346
8	3	2nd	1000	0	0.5	0.3333	0.0324	0.4943	0.3331
		3rd	1000	3	1.5	0.3333	3.0292	1.5013	0.3323
		1st	125	-2	1	0.25	-1.9909	1.0021	0.2568
9	4	2nd	125	0	0.5	0.25	-0.0563	0.4958	0.2461
		3rd	125	3	1.5	0.25	3.0346	1.497	0.2452
		4th	125	5	0.2	0.25	5.0297	0.205	0.2519

	Number of Number of			Par I	ameter of Distributio	EV-I on	Estimate	ed Paramete Distribution	r of EV-I
Scenario	Number of Components	Component	Number of Generated Data					^	
				μ	σ	w	μ	σ	ŵ
		1st	250	-2	1	0.25	-2.028	0.9977	0.246
10	4	2nd	250	0	0.5	0.25	-0.0213	0.4975	0.2484
10	1	3rd	250	3	1.5	0.25	2.9958	1.496	0.2486
		4th	250	5	0.2	0.25	5.0453	0.2054	0.257
		1st	500	-2	1	0.25	-2.0375	1.0029	0.2499
11	4	2nd	500	0	0.5	0.25	-0.0127	0.4979	0.2489
11	1	3rd	500	3	1.5	0.25	2.987	1.499	0.249
		4th	500	5	0.2	0.25	5.0461	0.2035	0.2522
		1st	1000	-2	1	0.25	-2.023	1.0001	0.2506
12	4	2nd	1000	0	0.5	0.25	0.0266	0.5048	0.2507
12	т	3rd	1000	3	1.5	0.25	2.9932	1.4988	0.2496
		4th	1000	5	0.2	0.25	5.0021	0.1985	0.2491
		1st	125	-2	1	0.2	-1.9904	0.9984	0.1953
		2nd	125	0	0.5	0.2	-0.0613	0.4978	0.1948
13	5	3rd	125	3	1.5	0.2	3.0027	1.4953	0.1909
		4th	125	5	0.2	0.2	5.0289	0.2024	0.1926
		5th	125	7	1	0.2	6.9631	0.9953	0.2264
		1st	250	-2	1	0.2	-2.0178	0.9991	0.1952
		2nd	250	0	0.5	0.2	0.0599	0.498	0.1969
14	5	3rd	250	3	1.5	0.2	3.0547	1.4995	0.2026
		4th	250	5	0.2	0.2	4.9164	0.2005	0.1997
		5th	250	7	1	0.2	6.952	0.9979	0.2056
		1st	500	-2	1	0.2	-2.0065	0.996	0.1992
		2nd	500	0	0.5	0.2	0.0085	0.4998	0.2033
15	5	3rd	500	3	1.5	0.2	3.0289	1.5015	0.1974
		4th	500	5	0.2	0.2	5.0036	0.1965	0.1983
		5th	500	7	1	0.2	7.057	0.9978	0.2018
		1st	1000	-2	1	0.2	-2.0132	1.0048	0.2
		2nd	1000	0	0.5	0.2	-0.0128	0.5025	0.2009
16	5	3rd	1000	3	1.5	0.2	2.9889	1.5011	0.2002
		4th	1000	5	0.2	0.2	4.9884	0.1999	0.2004
		5th	1000	7	1	0.2	7.0045	0.9996	0.1985

Table 3. Cont.

Besides displaying the results of grouping in Table 2 and parameter estimation results in Table 3, we also provide an overview of the histogram and predictive density for each scenario. Histograms and predictive densities for the first to fourth scenarios can be seen in Figure 2a–d, the fifth to eighth scenarios in Figure 3a–d, the ninth to twelfth scenarios in Figure 4a–d, and the thirteenth to sixteenth scenarios in Figure 5a–d.



Figure 2. Histograms and predictive densities for the two-component EV-I mixture distribution with (**a**) 125, (**b**) 250, (**c**) 500, and (**d**) 1000 samples per component.



Figure 3. Histograms and predictive densities for the three-component EV-I mixture distribution with (**a**) 125, (**b**) 250, (**c**) 500, and (**d**) 1000 samples per component.



Figure 4. Histograms and predictive densities for the four-component EV-I mixture distribution with (**a**) 125, (**b**) 250, (**c**) 500, and (**d**) 1000 samples per component.



Figure 5. Histograms and predictive densities for the five-component EV-I mixture distribution with (**a**) 125, (**b**) 250, (**c**) 500, and (**d**) 1000 samples per component.

6. Misspecification Cases

For the simulation study, we provided 16 scenarios to validate our proposed algorithm. In this section, we provide the opposite—i.e., we intentionally generate data that are not derived from the Weibull distribution and then analyze them using our proposed algorithm. We generate two datasets taken from different distributions. The first dataset is taken from a double-exponential distribution with location and scale parameters of 0 and 1, respectively, and the second dataset is taken from a logistic distribution with location-scale parameters of 2 and 0.4, respectively. Each of these datasets has as many as 1000 data points from the distribution.

We used the EV-I mixture and Gaussian mixture distributions to analyze the data described above. We used the same hyperparameters as in the simulation study section because the double-exponential and logistic distributions both have location and scale parameters. To compare the performance between the EV-I mixture and Gaussian mixture distributions, we used the Kullback–Leibler divergence; a complete explanation and formula for the Kullback–Leibler divergence (KLD) can be found in Van Erven and Harremos [73]. In the simple case, the KLD value of 0 indicates that the true values with fitted densities have identical quantities of information. Thus, the smaller the KLD value, the more identical the true and fitted densities are.

The posterior distribution of k for the misspecification cases data can be seen in Table 4. Based on Table 4, the data that we generate are detected with multimodal data, even though the data we generate are unimodal. Then, in Table 5 can be seen the comparison of KLD for EV-I mixture distribution and Gaussian mixture distribution. Based on Table 5, it can be concluded that the EV-I mixture distribution covers more than the Gaussian mixture distribution for these data.

Table 4. Posterior distribution of *k* for misspecification cases data based on mixture model using the EV-I distribution.

Distribution of Random Data	п			$\Pr(k \mid y)$		
Double-exponential (0,1)	1000	Pr(1) = 0 Pr(6) = 0.0789	$\begin{array}{l} Pr(2) = 0.2519 \\ Pr(7) = 0.0515 \end{array}$	$\Pr(3) = 0.2631$ $\sum_{\geq 8} \Pr(k y) = 0.0661$	$\Pr(4) = 0.1703$	$\Pr(5) = 0.1182$
Logistic (2,0.4)	1000	$\begin{array}{l} Pr(1) = 0.1492 \\ Pr(6) = 0.0122 \end{array}$	$\begin{array}{l} Pr(2) = 0.4960 \\ Pr(7) = 0.0055 \end{array}$	$\Pr(3) = 0.2175$ $\sum_{\geq 10} \Pr(k y) = 0.0035$	Pr(4) = 0.0843	Pr(5) = 0.0318

 Table 5. Comparison of Kullback–Leibler divergence using EV-I mixture and Gaussian mixture distributions for misspecification cases data.

Distribution of		Kullback-	Leibler
Random Data	Number of Components	EV-I Mixture Distribution	Gaussian Mixture Distribution
	2	0.0935	0.5311
	3	0.0789	0.5924
Double-exponential (0,1)	4	0.0902	0.6840
-	5	0.0873	0.3892
	6	0.0789	0.2908
	7	0.0677	0.2471
	2	0.0629	0.4920
	3	0.0632	0.3288
Logistic (2,0.4)	4	0.0760	0.2981
-	5	0.0754	0.2765
	6	0.0707	0.2587
	7	0.0633	0.2563

7. Application

7.1. Enzyme, Acidity, and Galaxy Datasets

In this section, we analyze enzyme, acidity, and galaxy datasets such as those of Richardson and Green [16] (these three datasets can be obtained from https://people. maths.bris.ac.uk/~mapjg/mixdata, accessed on 22 March 2021). We analyzed the datasets using EV-I mixture distribution with the hyperparameters for enzyme data being R = 2.86, $\varepsilon = 1.45$, $\zeta = 2.86$, $\vartheta = 2$, g = 0.2, h = 1.22, $\delta = 1$; for acidity data, R = 4.18, $\varepsilon = 5.02$, $\zeta = 4.18$, $\vartheta = 2$, g = 0.2, h = 0.573, $\delta = 1$; and for galaxy data, R = 25.11, $\varepsilon = 21.73$, $\zeta = 25.11$, $\vartheta = 2$, g = 0.2, h = 0.016, $\delta = 1$; for these three datasets, we used $k_{max} = 30$. The provisions for selecting these hyperparameters are explained in the Section 5. The posterior distribution of k for all three datasets can be seen in Table 6. Then, we compared the predictive densities of the enzyme, acidity, and galaxy datasets using the EV-I mixture and the Gaussian mixture distributions, which can be seen in Figures 6–8. Visually, based on Figures 6–8, it can be seen that the EV-I mixture distribution has better coverage than the Gaussian mixture distribution. Then, by using the KLD, it can be seen in Table 7, that the EV-I mixture distribution.

Table 6. Posterior distribution of *k* for enzyme, acidity, and galaxy datasets based on mixture model using the EV-I distribution.

Dataset	п	$\Pr(k \mid y)$
Enzyme	155	$\begin{array}{lll} \Pr(1) = 0 & \Pr(2) = 0.0181 & \Pr(3) = 0.2444 & \Pr(4) = 0.3232 & \Pr(5) = 0.2232 \\ \Pr(6) = 0.1135 & \Pr(7) = 0.0484 & \sum_{\geq 8} \Pr(k y) = 0.0292 \end{array}$
Acidity	245	$\begin{array}{lll} \Pr(1) = 0 & \Pr(2) = 0.0663 & \Pr(3) = 0.2470 & \Pr(4) = 0.2486 & \Pr(5) = 0.1890 \\ \Pr(6) = 0.1197 & \Pr(7) = 0.0685 & \sum_{\geq 8} \Pr(k y) = 0.0609 \end{array}$
Galaxy	82	$\begin{array}{ll} \Pr(1) = 5 \times 10^{-6} & \Pr(2) = 0.0004 & \Pr(3) = 0.0649 & \Pr(4) = 0.1410 & \Pr(5) = 0.1993 \\ \Pr(6) = 0.2095 & \Pr(7) = 0.1708 & \sum_{\geq 8} \Pr(k y) = 0.2141 & \end{array}$



Figure 6. Cont.



Figure 6. Predictive densities for enzyme data using the EV-I mixture and the Gaussian mixture distributions with (**a**) two components, (**b**) three components, (**c**) four components, (**d**) five components, (**e**) six components, and (**f**) seven components.



Figure 7. Cont.



Figure 7. Predictive densities for acidity data using the EV-I mixture and the Gaussian mixture distributions with (**a**) two components, (**b**) three components, (**c**) four components, (**d**) five components, (**e**) six components, and (**f**) seven components.



Figure 8. Cont.



Figure 8. Predictive densities for galaxy data using the EV-I mixture and the Gaussian mixture distributions with (**a**) two components, (**b**) three components, (**c**) four components, (**d**) five components, (**e**) six components, and (**f**) seven components.

7.2. Dengue Hemorrhagic Fever (DHF) in Eastern Surabaya, East Java, Indonesia

In this section, we apply the EV-I mixture distribution using RJMCMC to a real dataset. These data are the time until patient recovery from dengue hemorrhagic fever (DHF). We obtained the secondary data from medical records from Dr. Soetomo Hospital, Surabaya, East Java, Indonesia. The data concern patients in eastern Surabaya, which consists of seven subdistricts. Our data consist of 21 cases, with each case widespread over each subdistrict. The histogram of the spread of DHF in each subdistrict can be seen in the research conducted by Rantini et al. [36]. It was explained in their study that the data have a Weibull distribution. Whether the data are multimodal or not is unknown. The histogram of our original data is shown in Figure 9a.

	Number of	Kullba	ck–Leibler	
Dataset	Components	EV-I Mixture Distribution	Gaussian Mixture Distribution	
	2	0.3699	0.7173	
	3	0.2670	0.5673	
Enzyme	4	0.6710	0.7992	
5	5	0.7657	0.9251	
	6	2.0578	2.5590	
	7	3.4676	3.6022	
	2	0.6508	1.5108	
	3	0.5785	1.2076	
Acidity	4	1.0884	1.4265	
ý	5	1.2894	1.4491	
	6	1.2604	1.3877	
	7	1.2102	1.3409	
	2	0.1692	0.2071	
	3	0.0880	0.1445	
Galaxy	4	0.0771	0.1542	
2	5	0.1469	0.1556	
	6	0.2005	0.2266	
	7	0.2648	0.4460	

Table 7. Comparison of Kullback–Leibler divergence using EV-I mixture and Gaussian mixture distributions for enzyme, acidity, and galaxy datasets.



Figure 9. Histograms of DHF data in seven subdistricts in eastern Surabaya: (**a**) original data with Weibull distribution and (**b**) transformation of the original data into EV-I distribution.

To determine the number of mixture components in our data, we applied the NG-RJMCMC algorithm. Of course, the first step was to transform the original data into a location-scale family, which can be seen in Figure 9b. Then, for our transformed data, we used the hyperparameters R = 1.9459, $\varepsilon = 1.3863$, $\zeta = 1.9459$, $\vartheta = 2$, g = 0.2, h = 2.6409, $\delta = 1$, and $k_{\text{max}} = 30$. Then, we did all six moves type on the data with 200,000 sweeps. The results of the grouping are shown in Table 8. Based on Table 8, the DHF data in

eastern Surabaya have a multimodal pattern with the highest probability of having four components.

Table 8. Summary of the results of grouping for DHF data in eastern Surabaya using the EV-I mixture distribution.

n			$\Pr(k \mid y)$		
21	$\Pr(1) = 0.0055$ $\Pr(6) = 0.1618$	Pr(2) = 0.0095 Pr(7) = 0.0845	Pr(3) = 0.0068 Pr(8) = 0.0454	$\Pr(4) = 0.3487$ $\Pr(9) = 0.0255$	$\Pr(5) = 0.2778$ $\sum \Pr(k u) = 0.0245$
	11(0) = 0.1018	11(7) = 0.0045	11(0) = 0.0434	11(9) = 0.0233	$\sum_{\geq 10} \Gamma(\kappa y) = 0.0343$

Using the EV-I mixture distribution with four components, the results of the parameter estimation for each component are shown in Table 9. Then, the membership label of each observation in each mixture component is shown in Figure 10. Finally, the analysis was compared using the four-component EV-I mixture distribution and the four-component Gaussian mixture distribution, as shown in Figure 11. According to Table 10 and Figure 11, it can be seen that our data are better covered by using the four-components EV-I mixture distribution.

Table 9. The result of parameter estimation for each component of the EV-I mixture distribution on the DHF data in eastern Surabaya.

Component	w	μ	σ
1st	0.0805	0.0006	0.0053
2nd	0.1614	0.6957	0.0048
3rd	0.2408	1.1022	0.0048
4th	0.5174	1.6265	0.1927

Table 10. Kullback–Leibler divergence using EV-I mixture and Gaussian mixture distributions for the DHF data in eastern Surabaya.

	Kullback-	Leibler
Number of Components	EV-I Mixture Distribution	Gaussian Mixture Distribution
4	0.0711	0.2131



Figure 10. Membership label of each observation on the DHF data in eastern Surabaya using the EV-I mixture distribution with four components.



Figure 11. Comparison of predictive densities for DHF data in eastern Surabaya using the fourcomponent EV-I mixture distribution and the four-component Gaussian mixture distribution.

8. Conclusions

We provided an algorithm in the Bayesian mixture analysis. We called it non-Gaussian reversible jump Markov chain Monte Carlo (NG-RJMCMC). Our algorithm is a modification of RJMCMC, where there is a difference in the initial steps, namely changing the original distribution into a location-scale family. This step facilitates the grouping of each observation into the mixture components. Our algorithm allows researchers to easily analyze data that are not from the Gaussian family. In our study, we used Weibull distribution, then transformed it into the EV-I distribution.

To validate our algorithm, we performed 16 scenarios for the EV-I mixture distribution simulation study. The first to fourth scenarios had two components, the fifth to eighth scenarios had three components, the ninth to twelfth scenarios had four components, and the thirteenth to sixteenth scenarios had five components. We generated data in different sizes, ranging from 125 to 1000 samples per mixture component. Next, we analyzed them using a Bayesian analysis with the appropriate prior distributions. We used 200,000 sweeps per scenario and replicated them 500 times. The results of this simulation indicate that each scenario provides a minimum level of accuracy of 95%. Moreover, the estimated parameters come close to the real parameters for all scenarios.

To strengthen the proposed method, we provided misspecification cases. We deliberately generated unimodal data with double-exponential and logistic distributions, then estimated them using the EV-I mixture distribution and Gaussian mixture distribution. The results indicated that the data we generated are multimodally detected. Based on the KLD, the EV-I mixture distribution has better coverage than the Gaussian mixture distribution.

We also implemented our algorithm for real datasets, namely enzyme, acidity, and galaxy datasets. We compared the EV-I mixture distribution with the Gaussian mixture distribution for all three datasets. Based on the KLD, we found that the EV-I mixture distribution has better coverage than the Gaussian mixture distribution. Then, visually, the results also show that the EV-I mixture distribution has better coverage. We also compared the EV-I mixture distribution for the DHF data in eastern Surabaya. In our previous research, we analyzed the data using the Weibull distribution. We do not know whether the data were identified as multimodal or not. Using our algorithm, we found that the data are multimodal with four components. We also compared the EV-I mixture distribution and the Gaussian mixture distribution. Again, the EV-I mixture distribution indicated better coverage, seen both through the KLD and visually.

Author Contributions: D.R., N.I. and I. designed the research; D.R. collected and analyzed the data and drafted the paper. All authors have critically read and revised the draft and approved the final paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Ministry of Education, Culture, Research, and Technology Indonesia, which gave the scholarship in Program Magister Menuju Doktor Untuk Sarjana Unggul (PMDSU).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available in a publicly accessible repository.

Acknowledgments: The authors thank the referees for their helpful comments.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

 $T \sim \text{Weibull}(\eta, \lambda)$ with the CDF

$$F_T(t|\eta,\lambda) = \begin{cases} 1 - e^{-\left(\frac{t}{\lambda}\right)^{\eta}} & ,t \ge 0\\ 0 & ,t < 0 \,. \end{cases}$$

Define a new variable, $Y = \ln T$, its CDF is

$$F_Y(y) = \Pr(Y \le y) = \Pr(\ln T \le y) = \Pr(T \le e^y) = F_T(e^y) = 1 - \exp\left(-\left(\frac{e^y}{\lambda}\right)^\eta\right)$$

note that: $e^{\ln(\cdot)} = (\cdot)$

so that

$$F_{Y}(y) = 1 - \exp\left(-\left(\frac{e^{y}}{\lambda}\right)^{\eta}\right) = 1 - \exp\left(-\exp\left(\ln\left(\left(\frac{e^{y}}{\lambda}\right)^{\eta}\right)\right)\right) = 1 - \exp\left(-\exp\left(\eta\ln\left(\frac{e^{y}}{\lambda}\right)\right)\right)$$

$$= 1 - \exp(-\exp(\eta(y - \ln\lambda))) = 1 - \exp\left(-\exp\left(\frac{(y - \ln\lambda)}{\frac{1}{\eta}}\right)\right)$$

and its p.d.f is

$$f_{Y}(y) = \frac{dF_{Y}(y)}{dy} = \frac{d\left(1 - \exp\left(-\exp\left(\frac{(y - \ln\lambda)}{\frac{1}{\eta}}\right)\right)\right)}{dy} = \eta \exp\left[(y - \ln\lambda)\eta\right] \exp\left\{-\exp\left[(y - \ln\lambda)\eta\right]\right\} = \frac{1}{\frac{1}{\eta}} \exp\left(\frac{y - \ln\lambda}{\frac{1}{\eta}}\right) \exp\left(-\exp\left(\frac{y - \ln\lambda}{\frac{1}{\eta}}\right)\right).$$

Based on its CDF and p.d.f, it can be seen that $Y \sim EV - I(\mu, \sigma)$ where $\mu = \ln \lambda$ and $\sigma = \frac{1}{\eta}$. Then the appropriate support is as follows:

- $t \ge 0 \rightarrow y = \ln t \in (-\infty, \infty)$
- $\lambda > 0 \rightarrow \mu = \ln \lambda \in (-\infty, \infty)$
- $\eta > 0 \rightarrow \sigma = \frac{1}{\eta} \in (0, \infty)$

Appendix **B**

 $A = \text{posterior ratio} \times \text{proposal ratio}$

- = (likelihood ratio \times prior ratio) \times proposal ratio
- = (likelihood ratio × prior ratio) × ratio between two states × $\frac{1}{q(\mathbf{u})}$ × |Jacobian|
- = (likelihood ratio × prior ratio) × $\frac{d_{k+1}}{b_k p_{alloc}}$ × $\frac{1}{q(\mathbf{u})}$ × |Jacobian|

In this section, we have explained the above equation term by term

likelihood ratio term

likelihood ratio =
$$\frac{\text{new}}{\text{old}} = \frac{\prod_{i=1, z_i=j_1}^N p(y_i | \mu_{j_1}, \sigma_{j_1}) \prod_{i=1, z_i=j_2}^N p(y_i | \mu_{j_2}, \sigma_{j_2})}{\prod_{i=1, z_i=j^*}^N p(y_i | \mu_{j^*}, \sigma_{j^*})}$$

prior ratio term

$$\begin{aligned} \text{ratio of } z &= \frac{\text{new}}{\text{old}} = \frac{\left\{ w_{j_{1}} | \cdots \sim \text{Dirichlet}(\delta+l_{1}) \right\} | \times \left\{ w_{j_{2}} | \cdots \sim \text{Dirichlet}(\delta+l_{2}) \right\}}{w_{j^{*}} | \cdots \sim \text{Dirichlet}(\delta+l_{1}+l_{2})} \\ &= \frac{\frac{1}{B(\delta,k\delta)} w_{j^{*}}^{\delta+l_{1}-1} \frac{1}{B(\delta,k\delta)} w_{j^{*}}^{\delta+l_{2}-1}}{\frac{1}{B(\delta,k\delta)} w_{j^{*}}^{\delta+l_{1}-1} w_{j^{*}}^{\delta+l_{1}-1} w_{j^{*}}^{\delta+l_{2}-1}}{w_{j^{*}}^{\delta+l_{1}+l_{2}-1}} = \frac{\frac{1}{B(\delta,k\delta)} w_{j^{*}}^{\delta+l_{1}-1} w_{j^{*}}^{\delta+l_{2}-1}}{w_{j^{*}}^{\delta+l_{1}+l_{2}-1}} = \frac{w_{j^{*}}^{\delta+l_{1}-1} w_{j^{*}}^{\delta+l_{1}-1} w_{j^{*}}^{\delta+l_{2}-1}}{w_{j^{*}}^{\delta+l_{1}+l_{2}-1}} \\ \text{ratio of } \mu &= \frac{\text{new}}{\text{old}} = \frac{f(\mu_{j_{1}})f(\mu_{j_{2}})}{f(\mu_{j^{*}})} = \frac{\left\{ \frac{1}{\sqrt{2\pi\zeta^{2}}} e^{-\frac{1}{2}\left(\frac{\mu_{j_{1}}-e^{j^{2}}}{\zeta^{2}}\right)} \right\} \left\{ \frac{1}{\sqrt{2\pi\zeta^{2}}} e^{-\frac{1}{2}\left(\frac{\mu_{j_{2}}-e^{j^{2}}}{\zeta^{2}}\right)} \\ &= \frac{1}{\sqrt{2\pi\zeta^{2}}} e^{-\frac{1}{2}\left(\frac{\mu_{j_{1}}-e^{j^{2}}}{\zeta^{2}}-\frac{1}{2}\left(\frac{\mu_{j_{2}}-e^{j^{2}}}{\zeta^{2}}\right)^{2}}{\zeta^{2}} = \sqrt{\frac{1}{2\pi\zeta^{2}}} e^{-\frac{1}{2\zeta^{2}}\left[(\mu_{j_{1}}-e)^{2}-(\mu_{j^{*}}-e)^{2}\right]} \\ &= \left(\frac{1}{\sqrt{2\pi\zeta^{2}}}\right)^{\frac{1}{2}} e^{-\frac{1}{2\zeta^{2}}\left[(\mu_{j_{1}}-e)^{2}+(\mu_{j_{2}}-e)^{2}-(\mu_{j^{*}}-e)^{2}\right]}{\zeta^{2}}} \\ &= \left(\frac{1}{2\pi\zeta^{2}}\right)^{\frac{1}{2}} e^{-\frac{1}{2\zeta^{2}}\left[(\mu_{j_{1}}-e)^{2}+(\mu_{j_{2}}-e)^{2}-(\mu_{j^{*}}-e)^{2}\right]} \\ &= \left(\frac{1}{2\pi\zeta^{2}}\right)^{\frac{1}{2}} e^{-\frac{1}{2\zeta^{2}}\left[(\mu_{j_{1}}-e)^{2}+(\mu_{j_{2}}-e)^{2}-(\mu_{j^{*}}-e)^{2}\right]} \\ &= \left(\frac{1}{2\pi\zeta^{2}}\right)^{\frac{1}{2}} e^{-\frac{1}{2\zeta^{2}}\left[(\mu_{j_{1}}-e)^{2}+(\mu_{j_{2}}-e)^{2}-(\mu_{j^{*}}-e)^{2}\right]} \\ &= \frac{e^{\sigma}}{\log} = \frac{f(\sigma_{j_{1}})f(\sigma_{j_{2}})}{f(\sigma_{j^{*}})} = \frac{\left\{\frac{\omega^{\sigma}\exp\left(-\frac{\sigma}{\sigma_{j_{1}}}\right)}{e^{\sigma}}\right\} \times \left\{\frac{\omega^{\sigma}\exp\left(-\frac{\sigma}{\sigma_{j_{2}}}\right)}{e^{\sigma}}\right\}} \\ &= \frac{\omega^{\sigma}}{e^{\sigma}} \frac{\sigma^{\sigma}}_{j^{n+1}} \sigma_{j^{n+1}}^{\sigma}}{\sigma_{j^{n+1}}^{\sigma}} \exp\left(-\frac{\omega}{\sigma_{j_{1}}} - \frac{\omega}{\sigma_{j_{2}}} + \frac{\omega}{\sigma_{j^{*}}}\right) = \frac{\omega^{\sigma}}{e^{\sigma}} \left(\frac{\sigma}{\sigma_{j}} + \frac{\sigma}{\sigma_{j}}}\right) \left(\frac{\omega^{\sigma}}{\sigma_{j^{*}}} + \frac{\omega^{\sigma}}{\sigma_{j^{*}}} - \frac{\omega^{\sigma}}{\sigma_{j^{*}}}\right)}{e^{\sigma}} = \frac{\omega^{\sigma}}{e^{\sigma}} \left(\frac{\omega^{\sigma}}{\sigma_{j^{*}}} + \frac{\omega^{\sigma}}{\sigma_{j^{*}}} + \frac{\omega^{\sigma}}}{\sigma_{j^{*}}}\right) = \frac{\omega^{\sigma}}{e^{\sigma}} \left(\frac{\omega^{\sigma}}{\sigma_{j}} + \frac{\omega^{\sigma}}{\sigma_{j^{*}}}\right) = \frac{\omega^{\sigma}}{e^{\sigma}} \left(\frac{\omega^{\sigma}}{\sigma_{j^{*}}} + \frac{\omega^{\sigma}}}{\sigma_{j^{*}}} + \frac{\omega^{$$

• $\frac{d_{k+1}}{b_k p_{alloc}}$ term

$$p_{alloc} = \prod_{z_i=j_1} \frac{w_{j_1} f(y_i | \mu_{j_1}, \sigma_{j_1})}{w_{j_1} f(y_i | \mu_{j_1}, \sigma_{j_1}) + w_{j_2} f(y_i | \mu_{j_2}, \sigma_{j_2})} \times \prod_{z_i=j_2} \frac{w_{j_2} f(y_i | \mu_{j_2}, \sigma_{j_2})}{w_{j_1} f(y_i | \mu_{j_1}, \sigma_{j_1}) + w_{j_2} f(y_i | \mu_{j_2}, \sigma_{j_2})}$$

- $\frac{1}{q(\mathbf{u})\oplus}$ term $q(\mathbf{u}) = f(u_1)f(u_2)f(u_3)$, where $u_1 \sim \text{Beta}(2,2)$, $u_2 \sim \text{Beta}(2,2)$, and $u_3 \sim \text{Beta}(1,1)$, so $q(\mathbf{u}) = g_{2,2}(u_1)g_{2,2}(u_2)g_{1,1}(u_3)$
- |Jacobian| term

Because this part has a deeper explanation than others, so we write separately in Appendix C.

Appendix C

Based on Equation (29), we have partial derivatives for each variable as follows

$$\begin{array}{l} \frac{\partial w_{l_{1}}}{\partial w_{l_{1}}} = \frac{\partial \left(w_{l_{1}} u_{1}\right)}{\partial u_{l_{1}}} = 0 \\ \frac{\partial w_{l_{1}}}{\partial \mu_{l_{1}}} = \frac{\partial \left(w_{l_{1}} u_{1}\right)}{\partial \mu_{l_{1}}} = 0 \\ \frac{\partial w_{l_{1}}}{\partial \mu_{l_{1}}} = \frac{\partial \left(w_{l_{1}} u_{1}\right)}{\partial \mu_{l_{1}}} = 0 \\ \frac{\partial w_{l_{1}}}{\partial u_{1}} = \frac{\partial \left(w_{l_{1}} u_{1}\right)}{\partial u_{1}} = w_{l^{*}} \\ \frac{\partial w_{l_{1}}}{\partial u_{2}} = \frac{\partial \left(w_{l_{1}} u_{1}\right)}{\partial u_{3}} = 0 \\ \frac{\partial w_{l_{1}}}{\partial u_{2}} = \frac{\partial \left(w_{l_{1}} u_{1}\right)}{\partial u_{3}} = 0 \\ \frac{\partial u_{l_{1}}}{\partial u_{2}} = \frac{\partial \left(w_{l_{1}} u_{1}\right)}{\partial u_{3}} = 0 \\ \frac{\partial u_{l_{1}}}{\partial u_{2}} = \frac{\partial \left(w_{l_{1}} - u_{2}\sqrt{\sigma_{l^{*}}}\sqrt{\frac{w_{l_{2}}}{w_{l_{1}}}}\right)}{\partial u_{l_{1}}} = 0 \\ \frac{\partial u_{l_{1}}}{\partial u_{l_{1}}} = \frac{\partial \left(u_{l^{*}} - u_{2}\sqrt{\sigma_{l^{*}}}\sqrt{\frac{w_{l_{2}}}{w_{l_{1}}}}\right)}{\partial u_{1}} = 0 \\ \frac{\partial u_{l_{1}}}{\partial u_{1}} = \frac{\partial \left(u_{l^{*}} - u_{2}\sqrt{\sigma_{l^{*}}}\sqrt{\frac{w_{l_{2}}}{w_{l_{1}}}}\right)}{\partial u_{1}} = 0 \\ \frac{\partial u_{l_{1}}}{\partial u_{1}} = \frac{\partial \left(u_{l^{*}} - u_{2}\sqrt{\sigma_{l^{*}}}\sqrt{\frac{w_{l_{2}}}{w_{l_{1}}}}\right)}{\partial u_{1}} = 0 \\ \frac{\partial u_{l_{1}}}{\partial u_{2}} = \frac{\partial \left(u_{l^{*}} - u_{2}\sqrt{\sigma_{l^{*}}}\sqrt{\frac{w_{l_{2}}}{w_{l_{1}}}}\right)}{\partial u_{2}} = -\sqrt{\sigma_{l^{*}}}}\sqrt{\frac{w_{l_{2}}}{w_{l_{1}}}} \\ \frac{\partial u_{l_{1}}}{\partial u_{2}} = \frac{\partial \left(u_{l^{*}} - u_{2}\sqrt{\sigma_{l^{*}}}\sqrt{\frac{w_{l_{2}}}{w_{l_{1}}}}\right)}{\partial u_{2}} = 0 \\ \frac{\partial u_{l_{1}}}{\partial u_{2}} = \frac{\partial \left(u_{l^{*}} - u_{2}\sqrt{\sigma_{l^{*}}}\sqrt{\frac{w_{l_{2}}}{w_{l_{1}}}}\right)}{\partial u_{2}} = 0 \\ \frac{\partial u_{l_{1}}}{\partial u_{2}} = \frac{\partial \left(u_{l^{*}} - u_{2}\sqrt{\sigma_{l^{*}}}\sqrt{\frac{w_{l_{2}}}{w_{l_{1}}}}\right)}{\partial u_{2}}} = 0 \\ \frac{\partial u_{l_{1}}}}{\partial u_{2}} = \frac{\partial \left(u_{l^{*}} - u_{2}\sqrt{\sigma_{l^{*}}}\sqrt{\frac{w_{l_{2}}}{w_{l_{1}}}}\right)}{\partial u_{2}}} = 0 \\ \frac{\partial u_{l_{1}}}}{\partial u_{2}} = \frac{\partial \left(u_{l^{*}} - u_{2}\sqrt{\sigma_{l^{*}}}\sqrt{\frac{w_{l_{1}}}{w_{l_{1}}}}\right)}{\partial u_{2}}} = 0 \\ \frac{\partial u_{l_{1}}}}{\partial u_{l_{1}}}} = \frac{\partial \left(u_{l^{*}} - u_{2}\sqrt{\sigma_{l^{*}}}\sqrt{\frac{w_{l_{1}}}{w_{l_{1}}}}\right)}{\partial u_{2}}} = 0 \\ \frac{\partial u_{l_{1}}}}{\partial u_{l_{1}}}} = \frac{\partial \left(u_{l^{*}} - u_{l^{*}}^{*}/\sqrt{\frac{w_{l_{1}}}{w_{l_{1}}}}\right)}{\partial u_{l_{1}}}} = 0 \\ \frac{\partial u_{l_{1}}}}{\partial u_{l_{1}}}} = \frac{\partial \left(u_{l^{*}} - u_{l^{*}}^{*}/\sqrt{\frac{w_{l_{1}}}{w_{l_{1}}}}\right)}{\partial u_{l_{1}}}} = 0 \\ \frac{\partial u_{l_{1}}}}{\partial u_{l_{1}}}} \\ \frac{\partial u_{l_{1}}}}{\partial u_{l_{1}}} = \frac{\partial \left(u_{l^{*}} - u_{l^{*}}^$$

$$\begin{split} \frac{\partial \mu_{j_2}}{\partial w_{j^*}} &= \frac{\partial \left(\mu_{j^*} + u_2 \sqrt{\sigma_{j^*}} \sqrt{\frac{w_{j_1}}{w_{j_2}}} \right)}{\partial w_{j^*}} = 0 \\ \frac{\partial \mu_{j_2}}{\partial \mu_{j^*}} &= \frac{\partial \left(\mu_{j^*} + u_2 \sqrt{\sigma_{j^*}} \sqrt{\frac{w_{j_1}}{w_{j_2}}} \right)}{\partial \mu_{j^*}} = 1 \\ \frac{\partial \mu_{j_2}}{\partial \sigma_{j^*}} &= \frac{\partial \left(\mu_{j^*} + u_2 \sqrt{\sigma_{j^*}} \sqrt{\frac{w_{j_1}}{w_{j_2}}} \right)}{\partial \sigma_{j^*}} = 0 + \frac{1}{2} u_2 \sqrt{\frac{w_{j_1}}{w_{j_2}}} \sigma_{j^*}^{\frac{1}{2} - 1} = \frac{1}{2} u_2 \sqrt{\frac{w_{j_1}}{w_{j_2}}} \sigma_{j^*}^{-\frac{1}{2}} \\ \frac{\partial \mu_{j_2}}{\partial u_1} &= \frac{\partial \left(\mu_{j^*} + u_2 \sqrt{\sigma_{j^*}} \sqrt{\frac{w_{j_1}}{w_{j_2}}} \right)}{\partial u_1} = 0 \\ \frac{\partial \mu_{j_2}}{\partial u_2} &= \frac{\partial \left((\mu_{j^*} + u_2 \sqrt{\sigma_{j^*}} \sqrt{\frac{w_{j_1}}{w_{j_2}}} \right)}{\partial u_2} = \sqrt{\sigma_{j^*}} \sqrt{\frac{w_{j_1}}{w_{j_2}}} \\ \frac{\partial \mu_{j_2}}{\partial u_2} &= \frac{\partial \left((1 - u_3) \left(1 - u_2^2 \right) \sigma_{j^*} \frac{w_{j^*}}{w_{j_2}} \right)}{\partial u_3} = 0 \\ \frac{\partial \sigma_{j_2}}{\partial \omega_{j^*}} &= \frac{\partial \left((1 - u_3) \left(1 - u_2^2 \right) \sigma_{j^*} \frac{w_{j^*}}{w_{j_2}} \right)}{\partial \mu_{j^*}} = 0 \\ \frac{\partial \sigma_{j_2}}{\partial \mu_{j^*}} &= \frac{\partial \left((1 - u_3) \left(1 - u_2^2 \right) \sigma_{j^*} \frac{w_{j^*}}{w_{j_2}} \right)}{\partial \mu_{j^*}}} = 0 \\ \frac{\partial \sigma_{j_2}}{\partial \sigma_{j^*}} &= \frac{\partial \left((1 - u_3) \left(1 - u_2^2 \right) \sigma_{j^*} \frac{w_{j^*}}{w_{j_2}} \right)}{\partial \mu_{j^*}}} = 0 \\ \frac{\partial \sigma_{j_2}}}{\partial \sigma_{j^*}} &= \frac{\partial \left((1 - u_3) \left(1 - u_2^2 \right) \sigma_{j^*} \frac{w_{j^*}}{w_{j_2}} \right)}{\partial \mu_{j^*}}} = 0 \\ \frac{\partial \sigma_{j_2}}}{\partial u_2} &= \frac{\partial \left((1 - u_3) \left(1 - u_2^2 \right) \sigma_{j^*} \frac{w_{j^*}}{w_{j_2}} \right)}{\partial u_2}} = (1 - u_3) \left(1 - u_2^2 \right) \frac{w_{j^*}}{w_{j_2}}} \\ \frac{\partial \sigma_{j_2}}}{\partial u_2} &= \frac{\partial \left((1 - u_3) \left(1 - u_2^2 \right) \sigma_{j^*} \frac{w_{j^*}}{w_{j_2}} \right)}{\partial u_3}} = (1 - u_3) \sigma_{j^*} \frac{w_{j^*}}{w_{j^*}} \left(0 - 2u_2 \right) = -2(1 - u_3) u_2 \sigma_{j^*} \frac{w_{j^*}}{w_{j_2}}} \\ \frac{\partial \sigma_{j_2}}}{\partial u_3} &= \frac{\partial \left((1 - u_3) \left(1 - u_2^2 \right) \sigma_{j^*} \frac{w_{j^*}}{w_{j_2}} \right)}{\partial u_3}} = (-1) \left(1 - u_2^2 \right) \sigma_{j^*} \frac{w_{j^*}}}{w_{j_2}} = -(1 - u_2^2) \sigma_{j^*} \frac{w_{j^*}}}{w_{j_2}}} \end{split}$$

so that the determinant of the Jacobian can be written as

$$\begin{split} \left|\frac{\partial \Delta^{*}}{\partial(\Delta,\mathbf{u})}\right| &= \left|\frac{\partial \left(w_{j_{1}},\mu_{j_{1}},\sigma_{j_{1}},w_{j_{2}},\mu_{j_{2}},\sigma_{j_{2}}\right)}{\partial \left(w_{j^{*}},\mu_{j^{*}},\sigma_{j^{*}}^{*},\mu_{1},u_{2},u_{3}\right)}\right| = \begin{vmatrix} \frac{\partial w_{j_{1}}}{\partial w_{j^{*}}} & \frac{\partial w_{j_{1}}}{\partial \mu_{j^{*}}} & \frac{\partial w_{j_{1}}}}{\partial \mu_{j^{*}}} & \frac{\partial w_{j_{1}}}{\partial \mu_{j^{*}}} & \frac{\partial w_{j_{1}}}{\partial \mu_{j^{*}}} & \frac{\partial w_{j_{1}}}{\partial \mu_{j^{*}}} & \frac{\partial w_{j_{1}}}}{\partial \mu_{j^{*}}} & \frac{\partial w_{j^{*}}}}{\partial \mu_{j^{*}}} & \frac{\partial w_{j^{*}}}}{\partial \mu_{j^{*}}} & \frac{\partial w_{j^{*}}}}{\partial \mu_{$$

Manually, calculating the determinant of a 6×6 matrix is not easy, so we can use software for calculations. In this calculation, we used Maple software, and we obtained

$$\left|\frac{\partial \Delta^{*}}{\partial(\Delta,\mathbf{u})}\right| = \frac{w_{j^{*}}^{3}\sigma_{j^{*}}^{\frac{3}{2}}(u_{2}^{2}-1)\left(\sqrt{\frac{w_{j_{1}}}{w_{j_{2}}}}+\sqrt{\frac{w_{j_{2}}}{w_{j_{1}}}}\right)}{w_{j_{1}}w_{j_{2}}}$$

Mathematically, the equation above can be rewritten as

$$\begin{split} \frac{w_{j^*}^3 \sigma_{j^*} \frac{3}{2} (u_2^2 - 1) \left(\sqrt{\frac{w_{j_1}}{w_{j_2}}} + \sqrt{\frac{w_{j_2}}{w_{j_1}}} \right)}{w_{j_1} w_{j_2}} \\ &= \frac{w_{j^*}^2 w_{j^*} \sigma_{j^*} \frac{3}{2} + \frac{1}{2} - \frac{1}{2} \left(-(1 - u_2^2) \right) \frac{(1 - u_2^2)}{(1 - u_2^2)} \frac{(1 - u_3)}{(1 - u_3)} \frac{u_3}{u_3} \frac{u_2}{u_2} \frac{\sigma_{j^*}}{\sigma_{j^*}} \left(\sqrt{\frac{w_{j_1}}{w_{j_2}}} + \sqrt{\frac{w_{j_2}}{w_{j_1}}} \right)}{w_{j_1} w_{j_2}} \\ &= \left(\frac{w_{j^*}^2 \sigma_{j^*}^2 (1 - u_2^2)^2 (1 - u_3) u_3}{w_{j_1} w_{j_2}} \right) \times \left(-\sigma_{j^*} - \frac{1}{2} + 1 u_2 \left(\sqrt{\frac{w_{j_1}}{w_{j_2}}} + \sqrt{\frac{w_{j_2}}{w_{j_1}}} \right) \right) \times \left(\frac{w_{j^*}}{(1 - u_2^2) (1 - u_3) u_3 u_2 \sigma_{j^*}} \right) \\ &= \left(\frac{w_{j^*}^2 \sigma_{j^*}^2 (1 - u_2^2)^2 (1 - u_3) u_3}{w_{j_1} w_{j_2}} \right) \times \left(-\sigma_{j^*} \frac{1}{2} u_2 \left(\sqrt{\frac{w_{j_1}}{w_{j_2}}} + \sqrt{\frac{w_{j_2}}{w_{j_1}}} \right) \right) \times \left(\frac{w_{j^*}}{(1 - u_2^2) (1 - u_3) u_3 u_2 \sigma_{j^*}} \right) \\ &= \sigma_{j_1} \sigma_{j_2} \times |\mu_{j_1} - \mu_{j_2}| \times \left(\frac{w_{j^*}}{(1 - u_2^2) (1 - u_3) u_3 u_2 \sigma_{j^*}} \right) \\ &= \frac{w_{j^*} |\mu_{j_1} - \mu_{j_2}| \sigma_{j_1} \sigma_{j_2}}{u_2 (1 - u_2^2) u_3 (1 - u_3) \sigma_{j^*}} \end{split}$$

Two things must be considered mathematically for this proof:

i. For
$$|\mu_{j_1} - \mu_{j_2}|$$

$$\begin{aligned} |\mu_{j_{1}} - \mu_{j_{2}}| &= \left| \left(\mu_{j^{*}} - u_{2} \sqrt{\sigma_{j^{*}}} \sqrt{\frac{w_{j_{2}}}{w_{j_{1}}}} \right) - \left(\mu_{j^{*}} + u_{2} \sqrt{\sigma_{j^{*}}} \sqrt{\frac{w_{j_{1}}}{w_{j_{2}}}} \right) \right| \\ &= \left| \mu_{j^{*}} - u_{2} \sqrt{\sigma_{j^{*}}} \sqrt{\frac{w_{j_{2}}}{w_{j_{1}}}} - \mu_{j^{*}} - u_{2} \sqrt{\sigma_{j^{*}}} \sqrt{\frac{w_{j_{1}}}{w_{j_{2}}}} \right| \\ &= \left| -u_{2} \sqrt{\sigma_{j^{*}}} \left(\sqrt{\frac{w_{j_{2}}}{w_{j_{1}}}} + \sqrt{\frac{w_{j_{1}}}{w_{j_{2}}}} \right) \right| \\ &= \left| -u_{2} \sigma_{j^{*}} \frac{1}{2} \left(\sqrt{\frac{w_{j_{2}}}{w_{j_{1}}}} + \sqrt{\frac{w_{j_{1}}}{w_{j_{2}}}} \right) \right| \\ &= -u_{2} \sigma_{j^{*}} \frac{1}{2} \left(\sqrt{\frac{w_{j_{2}}}{w_{j_{1}}}} + \sqrt{\frac{w_{j_{1}}}{w_{j_{2}}}} \right) \\ &\left(\sqrt{\frac{w_{j_{2}}}{w_{j_{1}}}} + \sqrt{\frac{w_{j_{1}}}{w_{j_{2}}}} \right) &= \frac{|\mu_{j_{1}} - \mu_{j_{2}}|}{-u_{2} \sigma_{j^{*}} \frac{1}{2}} \end{aligned}$$

ii. For $\sigma_{j_1}^2 \sigma_{j_2}^2$

$$\begin{aligned} \sigma_{j_1}\sigma_{j_2} &= \left(u_3\left(1-u_2^2\right)\sigma_{j^*}\frac{w_{j^*}}{w_{j_1}}\right)\left((1-u_3)\left(1-u_2^2\right)\sigma_{j^*}\frac{w_{j^*}}{w_{j_2}}\right) \\ &= u_3\left(1-u_2^2\right)^2(1-u_3)\sigma_{j^*}\frac{w_{j^*}}{w_{j_2}}\end{aligned}$$

Appendix D

The simulation study ran well when using a computer with processor Intel Core i7, 32GB RAM, 447GB SSD. We have used the R software. Then, we calculated the CPU time by using the "proc.time()" function in R. The CPU time can be seen in Table A1. The "user" is the CPU time charged for the execution of user instructions of the calling process, while the "system" is the CPU time charged for execution by the system on behalf of the calling process. This CPU time is for one replication, so the total CPU time required for the sixteen scenarios is to multiply each time in the Table A1 by 500.

Scenario	Number of Components	Sample Size (per Component)	CPU Time (in Second)		
			User	System	Elapsed
1	2	125	181.65	3.06	262.43
2	2	250	221.25	2.56	288.73
3	2	500	302.20	2.78	542.76
4	2	1000	760.76	4.32	1270.12
5	3	125	204.40	2.73	1111.32
6	3	250	274.46	2.95	321.70
7	3	500	404.71	3.34	607.15
8	3	1000	656.23	3.40	708.84
9	4	125	253.71	3.01	566.96
10	4	250	347.84	3.09	364.46
11	4	500	551.57	3.10	1197.01
12	4	1000	973.84	3.53	1403.56
13	5	125	299.40	3.26	2535.14
14	5	250	446.78	3.03	627.54
15	5	500	740.00	3.18	1197.18
16	5	1000	1324.29	3.84	3526.37

Table A1. Report the CPU time for simulation study.

References

- 1. Roeder, K. A Graphical Technique for Determining the Number of Components in a Mixture of Normals. J. Am. Stat. Assoc. 1994, 89, 487–495. [CrossRef]
- Carreira-Perpinán, M.A.; Williams, C.K.I. On the Number of Modes of a Gaussian Mixture. In Proceedings of the International Conference on Scale-Space Theories in Computer Vision, Isle of Skye, UK, 10–12 June 2003; pp. 625–640. [CrossRef]
- 3. Vlassis, N.; Likas, A. A Greedy EM Algorithm for Gaussian Mixture Learning. Neural Process. Lett. 2002, 15, 77–87. [CrossRef]
- 4. Jeffries, N.O. A Note on "Testing the Number of Components in a Normal Mixture". Biometrika 2003, 90, 991–994. [CrossRef]
- 5. Lo, Y.; Mendell, N.R.; Rubin, D.B. Testing the Number of Components in a Normal Mixture. *Biometrika* 2001, *88*, 767–778. [CrossRef]
- Kasahara, H.; Shimotsu, K. Testing the Number of Components in Normal Mixture Regression Models. J. Am. Stat. Assoc. 2015, 110, 1632–1645. [CrossRef]
- McLachlan, G.J. On Bootstrapping the Likelihood Ratio Test Statistic for the Number of Components in a Normal Mixture. J. R. Stat. Soc. Ser. C Appl. Stat. 1987, 36, 318–324. [CrossRef]
- 8. Soromenho, G. Comparing Approaches for Testing the Number of Components in a Finite Mixture Model. *Comput. Stat.* **1994**, *9*, 65–78.
- Bozdogan, H. Choosing the Number of Component Clusters in the Mixture-Model Using a New Informational Complexity Criterion of the Inverse-Fisher Information Matrix. In *Information and Classification*; Springer: Berlin/Heidelberg, Germany, 1993; pp. 40–54. [CrossRef]
- 10. Polymenis, A.; Titterington, D.M. On the Determination of the Number of Components in a Mixture. *Stat. Probab. Lett.* **1998**, *38*, 295–298. [CrossRef]
- 11. Baudry, J.P.; Raftery, A.E.; Celeux, G.; Lo, K.; Gottardo, R. Combining Mixture Components for Clustering. *J. Comput. Graph. Stat.* **2010**, *19*, 332–353. [CrossRef]
- 12. Lukočiene, O.; Vermunt, J.K. Determining the Number of Components in Mixture Models for Hierarchical Data. In *Studies in Classification, Data Analysis, and Knowledge Organization*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 241–249. [CrossRef]
- 13. Miller, J.W.; Harrison, M.T. Mixture Models with a Prior on the Number of Components. *J. Am. Stat. Assoc.* **2018**, *113*, 340–356. [CrossRef] [PubMed]
- 14. Fearnhead, P. Particle Filters for Mixture Models with an Unknown Number of Components. *Stat. Comput.* **2004**, *14*, 11–21. [CrossRef]
- 15. Mclachlan, G.J.; Rathnayake, S. On the Number of Components in a Gaussian Mixture Model. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2014**, *4*, 341–355. [CrossRef]
- 16. Richardson, S.; Green, P.J. On Bayesian Analysis of Mixtures with an Unknown Number of Components. J. R. Stat. Soc. Ser. B Stat. Methodol. 1997, 59, 731–792. [CrossRef]
- 17. Astuti, A.B.; Iriawan, N.; Irhamah; Kuswanto, H. Development of Reversible Jump Markov Chain Monte Carlo Algorithm in the Bayesian Mixture Modeling for Microarray Data in Indonesia. In *AIP Conference Proceedings*; AIP Publishing LLC: Melville, NY, USA, 2017; Volume 1913, p. 20033. [CrossRef]
- Liu, R.-Y.; Tao, J.; Shi, N.-Z.; He, X. Bayesian Analysis of the Patterns of Biological Susceptibility via Reversible Jump MCMC Sampling. *Comput. Stat. Data Anal.* 2011, 55, 1498–1508. [CrossRef]

- Bourouis, S.; Al-Osaimi, F.R.; Bouguila, N.; Sallay, H.; Aldosari, F.; Al Mashrgy, M. Bayesian Inference by Reversible Jump MCMC for Clustering Based on Finite Generalized Inverted Dirichlet Mixtures. *Soft Comput.* 2019, 23, 5799–5813. [CrossRef]
- 20. Green, P.J. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika* **1995**, *82*, 711–732. [CrossRef]
- Sanquer, M.; Chatelain, F.; El-Guedri, M.; Martin, N. A Reversible Jump MCMC Algorithm for Bayesian Curve Fitting by Using Smooth Transition Regression Models. In Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, 22–27 May 2011; pp. 3960–3963. [CrossRef]
- 22. Wang, Y.; Zhou, X.; Wang, H.; Li, K.; Yao, L.; Wong, S.T.C. Reversible Jump MCMC Approach for Peak Identification for Stroke SELDI Mass Spectrometry Using Mixture Model. *Bioinformatics* **2008**, *24*, i407–i413. [CrossRef]
- 23. Razul, S.G.; Fitzgerald, W.J.; Andrieu, C. Bayesian Model Selection and Parameter Estimation of Nuclear Emission Spectra Using RJMCMC. *Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrometers Detect. Assoc. Equip.* **2003**, 497, 492–510. [CrossRef]
- 24. Karakuş, O.; Kuruoğlu, E.E.; Altınkaya, M.A. Bayesian Volterra System Identification Using Reversible Jump MCMC Algorithm. *Signal Process.* **2017**, *141*, 125–136. [CrossRef]
- 25. Nasserinejad, K.; van Rosmalen, J.; de Kort, W.; Lesaffre, E. Comparison of Criteria for Choosing the Number of Classes in Bayesian Finite Mixture Models. *PLoS ONE* **2017**, *12*, e0168838. [CrossRef]
- Zhang, Z.; Chan, K.L.; Wu, Y.; Chen, C. Learning a Multivariate Gaussian Mixture Model with the Reversible Jump MCMC Algorithm. *Stat. Comput.* 2004, 14, 343–355. [CrossRef]
- 27. Kato, Z. Segmentation of Color Images via Reversible Jump MCMC Sampling. Image Vis. Comput. 2008, 26, 361–371. [CrossRef]
- Lunn, D.J.; Best, N.; Whittaker, J.C. Generic Reversible Jump MCMC Using Graphical Models. *Stat. Comput.* 2009, 19, 395. [CrossRef]
- 29. Bouguila, N.; Elguebaly, T. A Fully Bayesian Model Based on Reversible Jump MCMC and Finite Beta Mixtures for Clustering. *Expert Syst. Appl.* **2012**, *39*, 5946–5959. [CrossRef]
- 30. Chen, M.H.; Ibrahim, J.G.; Sinha, D. A New Bayesian Model for Survival Data with a Surviving Fraction. *J. Am. Stat. Assoc.* **1999**, 94, 909–919. [CrossRef]
- 31. Banerjee, S.; Carlin, B. Hierarchical Multivariate CAR Models for Spatio-Temporally Correlated Survival Data. *Bayesian Stat.* 2003, 7, 45–63.
- 32. Darmofal, D. Bayesian Spatial Survival Models for Political Event Processes. Am. J. Pol. Sci. 2009, 53, 241–257. [CrossRef]
- Motarjem, K.; Mohammadzadeh, M.; Abyar, A. Bayesian Analysis of Spatial Survival Model with Non-Gaussian Random Effect. J. Math. Sci. 2019, 237, 692–701. [CrossRef]
- 34. Thamrin, S.A.; McGree, J.M.; Mengersen, K.L. Bayesian Weibull Survival Model for Gene Expression Data. *Case Stud. Bayesian Stat. Model. Anal.* 2013, *1*, 171–185. [CrossRef]
- 35. Iriawan, N.; Astutik, S.; Prastyo, D.D. Markov Chain Monte Carlo—Based Approaches for Modeling the Spatial Survival with Conditional Autoregressive (CAR) Frailty. *IJCSNS Int. J. Comput. Sci. Netw. Secur.* **2010**, *10*, 211–217.
- 36. Rantini, D.; Abdullah, M.N.; Iriawan, N.; Irhamah; Rusli, M. On the Computational Bayesian Survival Spatial Dengue Hemorrhagic Fever (DHF) Modeling with Double-Exponential CAR Frailty. *J. Phys. Conf. Ser.* **2021**, 1722, 012042. [CrossRef]
- Rantini, D.; Candrawengi, N.L.P.I.; Iriawan, N.; Irhamah; Rusli, M. On the Computational Bayesian Survival Spatial DHF Modelling with CAR Frailty. In *AIP Conference Proceedings*; AIP Publishing LLC: Melville, NY, USA, 2021; Volume 2329, p. 60028. [CrossRef]
- Villa-Covarrubias, B.; Piña-Monarrez, M.R.; Barraza-Contreras, J.M.; Baro-Tijerina, M. Stress-Based Weibull Method to Select a Ball Bearing and Determine Its Actual Reliability. *Appl. Sci.* 2020, 10, 8100. [CrossRef]
- Zamora-Antuñano, M.A.; Mendoza-Herbert, O.; Culebro-Pérez, M.; Rodríguez-Morales, A.; Rodríguez-Reséndiz, J.; Gonzalez-Duran, J.E.E.; Mendez-Lozano, N.; Gonzalez-Gutierrez, C.A. Reliable Method to Detect Alloy Soldering Fractures under Accelerated Life Test. *Appl. Sci.* 2019, *9*, 3208. [CrossRef]
- 40. Tsionas, E.G. Bayesian Analysis of Finite Mixtures of Weibull Distributions. *Commun. Stat. Theory Methods* **2002**, *31*, 37–48. [CrossRef]
- 41. Marín, J.M.; Rodríguez-Bernal, M.T.; Wiper, M.P. Using Weibull Mixture Distributions to Model Heterogeneous Survival Data. *Commun. Stat. Simul. Comput.* **2005**, *34*, 673–684. [CrossRef]
- 42. Greenhouse, J.B.; Wolfe, R.A. A Competing Risks Derivation of a Mixture Model for the Analysis of Survival Data. *Commun. Stat. Methods* **1984**, *13*, 3133–3154. [CrossRef]
- 43. Liao, J.J.Z.; Liu, G.F. A Flexible Parametric Survival Model for Fitting Time to Event Data in Clinical Trials. *Pharm. Stat.* **2019**, *18*, 555–567. [CrossRef]
- 44. Zhang, Q.; Hua, C.; Xu, G. A Mixture Weibull Proportional Hazard Model for Mechanical System Failure Prediction Utilising Lifetime and Monitoring Data. *Mech. Syst. Signal Process* **2014**, *43*, 103–112. [CrossRef]
- 45. Elmahdy, E.E. A New Approach for Weibull Modeling for Reliability Life Data Analysis. *Appl. Math. Comput.* **2015**, 250, 708–720. [CrossRef]
- 46. Farcomeni, A.; Nardi, A. A Two-Component Weibull Mixture to Model Early and Late Mortality in a Bayesian Framework. *Comput. Stat. Data Anal.* **2010**, *54*, 416–428. [CrossRef]
- Phillips, N.; Coldman, A.; McBride, M.L. Estimating Cancer Prevalence Using Mixture Models for Cancer Survival. *Stat. Med.* 2002, 21, 1257–1270. [CrossRef]

- Lambert, P.C.; Dickman, P.W.; Weston, C.L.; Thompson, J.R. Estimating the Cure Fraction in Population-based Cancer Studies by Using Finite Mixture Models. J. R. Stat. Soc. Ser. C Appl. Stat. 2010, 59, 35–55. [CrossRef]
- 49. Sy, J.P.; Taylor, J.M.G. Estimation in a Cox Proportional Hazards Cure Model. Biometrics 2000, 56, 227–236. [CrossRef] [PubMed]
- 50. Franco, M.; Balakrishnan, N.; Kundu, D.; Vivo, J.-M. Generalized Mixtures of Weibull Components. *Test* **2014**, *23*, 515–535. [CrossRef]
- 51. Bučar, T.; Nagode, M.; Fajdiga, M. Reliability Approximation Using Finite Weibull Mixture Distributions. *Reliab. Eng. Syst. Saf.* **2004**, *84*, 241–251. [CrossRef]
- 52. Newcombe, P.J.; Raza Ali, H.; Blows, F.M.; Provenzano, E.; Pharoah, P.D.; Caldas, C.; Richardson, S. Weibull Regression with Bayesian Variable Selection to Identify Prognostic Tumour Markers of Breast Cancer Survival. *Stat. Methods Med. Res.* **2017**, *26*, 414–436. [CrossRef]
- 53. Denis, M.; Molinari, N. Free Knot Splines with RJMCMC in Survival Data Analysis. *Commun. Stat. Theory Methods* **2010**, *39*, 2617–2629. [CrossRef]
- 54. Mallet, C.; Lafarge, F.; Bretar, F.; Soergel, U.; Heipke, C. Lidar Waveform Modeling Using a Marked Point Process. In Proceedings of the Conference on Image Processing, ICIP, Cairo, Egypt, 7–10 November 2009; pp. 1713–1716. [CrossRef]
- 55. Mitra, D.; Balakrishnan, N. Statistical Inference Based on Left Truncated and Interval Censored Data from Log-Location-Scale Family of Distributions. *Commun. Stat. Simul. Comput.* **2019**, *50*, 1073–1093. [CrossRef]
- 56. Balakrishnan, N.; Ng, H.K.T.; Kannan, N. Goodness-of-Fit Tests Based on Spacings for Progressively Type-II Censored Data from a General Location-Scale Distribution. *IEEE Trans. Reliab.* **2004**, *53*, 349–356. [CrossRef]
- 57. Castro-Kuriss, C. On a Goodness-of-Fit Test for Censored Data from a Location-Scale Distribution with Applications. *Chil. J. Stat.* **2011**, *2*, 115–136.
- 58. Bouguila, N.; Elguebaly, T. A Bayesian Approach for Texture Images Classification and Retrieval. In Proceedings of the International Conference on Multimedia Computing and Systems, Ouarzazate, Morocco, 7–9 April 2011; pp. 1–6. [CrossRef]
- 59. Naulet, Z.; Barat, É. Some Aspects of Symmetric Gamma Process Mixtures. Bayesian Anal. 2018, 13, 703–720. [CrossRef]
- 60. Jo, S.; Roh, T.; Choi, T. Bayesian Spectral Analysis Models for Quantile Regression with Dirichlet Process Mixtures. J. Nonparametr. Stat. 2016, 28, 177–206. [CrossRef]
- 61. Kobayashi, G.; Kozumi, H. Bayesian Analysis of Quantile Regression for Censored Dynamic Panel Data. *Comput. Stat.* **2012**, 27, 359–380. [CrossRef]
- 62. Gruet, M.A.; Philppe, A.; Robert, C.P. Mcmc Control Spreadsheets for Exponential Mixture Estimation? *J. Comput. Graph. Stat.* **1999**, *8*, 298–317. [CrossRef]
- 63. Ulrich, W.; Nakadai, R.; Matthews, T.J.; Kubota, Y. The Two-Parameter Weibull Distribution as a Universal Tool to Model the Variation in Species Relative Abundances. *Ecol. Complex.* **2018**, *36*, 110–116. [CrossRef]
- 64. Scholz, F.W. Inference for the Weibull Distribution: A Tutorial. Quant. Methods Psychol. 2015, 11, 148–173. [CrossRef]
- 65. Zhang, J.; Ng, H.K.T.; Balakrishnan, N. Statistical Inference of Component Lifetimes with Location-Scale Distributions from Censored System Failure Data with Known Signature. *IEEE Trans. Reliab.* **2015**, *64*, 613–626. [CrossRef]
- Park, H.W.; Sohn, H. Parameter Estimation of the Generalized Extreme Value Distribution for Structural Health Monitoring. Probabilistic Eng. Mech. 2006, 21, 366–376. [CrossRef]
- 67. Loaiciga, H.A.; Leipnik, R.B. Analysis of Extreme Hydrologic Events with Gumbel Distributions: Marginal and Additive Cases. *Stoch. Environ. Res. Risk Assess.* **1999**, *13*, 251–259. [CrossRef]
- Banerjee, A.; Kundu, D. Inference Based on Type-II Hybrid Censored Data from a Weibull Distribution. *IEEE Trans. Reliab.* 2008, 57, 369–378. [CrossRef]
- 69. Yoon, S.; Cho, W.; Heo, J.H.; Kim, C.E. A Full Bayesian Approach to Generalized Maximum Likelihood Estimation of Generalized Extreme Value Distribution. *Stoch. Environ. Res. Risk Assess.* **2010**, *24*, 761–770. [CrossRef]
- 70. Coles, S.G.; Tawn, J.A. A Bayesian Analysis of Extreme Rainfall Data. Appl. Stat. 1996, 45, 463. [CrossRef]
- Tancredi, A.; Anderson, C.; O'Hagan, A. Accounting for Threshold Uncertainty in Extreme Value Estimation. *Extremes* 2006, 9, 87–106. [CrossRef]
- 72. Robert, C.P. *The Bayesian Choice: From Decision—Theoretic Foundations to Computational Implementation;* Springer Science & Business Media: New York, NY, USA, 2007; Volume 91.
- 73. Van Erven, T.; Harremos, P. Rényi Divergence and Kullback-Leibler Divergence. *IEEE Trans. Inf. Theory* **2014**, *60*, 3797–3820. [CrossRef]