Media Statistika

p-ISSN 1979 – 3693 e-ISSN 2477 – 0647
MEDIA STATISTIKA 13(1) 2020: 13-24
http://ejournal.undip.ac.id/index.php/media_statistika

PENGELOMPOKAN RUMAH TANGGA DI INDONESIA BERDASARKAN PENDAPATAN PER KAPITA DENGAN MODEL FINITE MIXTURE

Irwan Susanto, Sri Sulistijowati Handajani
Program Studi Statistika FMIPA Universitas Sebelas Maret Surakarta

e-mail: irwansusanto@staff.uns.ac.id

DOI: 10.14710/medstat.13.1.13-24

Article Info:
Received: 26 April 2019
Accepted: 22 June 2020
Available Online: 26 June 2020

Keywords:
Finite Mixture, Income Distribution, EM Algorithm, AIC, BIC.
.

Abstract: In the statistical modeling framework, the form of the income distribution can be approaching based on certain statistical distributions. The use of the finite mixture model is relatively flexible in the modeling of the income distribution that has a multimodal pattern. The multimodal pattern can be indicated as the existence of different cluster on the data. The different clusters which can reflect the economic homogeneity of income are represented by the mixture components of the finite mixture model. In this paper, the finite mixture model is implemented for modeling the distribution of household income per capita in Indonesia based on The Fifth Wave of the Indonesia Family Life Survey (IFLS5) 2014-2015. The mixture components of the finite mixture model have been build based on the heavy-tailed statistical distributions, i.e., gamma, lognormal, and Weibull distributions. The estimation of the fitting finite mixture model was conducted using the maximum-likelihood estimation method through the expectation-maximization (EM) algorithm. The suitable finite mixture models were verified with the bootstrap likelihood ratio statistics test, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Based on the results, the distribution of household income per capita in Indonesia can be modeled by the four components-lognormal mixture model.

1. PENDAHULUAN

Pemodelan distribusi pendapatan dengan pendekatan parametrik menggunakan distribusi probabilitas untuk mengestimasi model distribusi pendapatan pada suatu populasi. Estimasi model distribusi pendapatan tersebut dipergunakan sebagai bagian dari proses analisis terhadap distribusi pendapatan di suatu daerah yang berkaitan dengan ketimpangan ekonomi dan kemiskinan yang terjadi di daerah tersebut (Chotikapanich dan Griffiths, 2008). Distribusi *finite mixture* relatif fleksibel dalam memodelkan distribusi pendapatan yang mempunyai sub-sub populasi berbeda, dimana sub-sub populasi tersebut dapat merefleksikan kelompok dengan homogenitas secara ekonomi (Pittau *et al.*, 2016). Kelompok atau sub-sub populasi tersebut direpresentasikan dalam model *finite mixture* sebagai suatu komponen *mixture*.

Distribusi pendapatan mempunyai karakteristik data bertipe kontinu, positif dan *heavy-tailed*. Sehingga dalam penyusunan model *finite mixture*, penggunaan distribusi probabilitas yang berkarakteristik *heavy-tailed*, dalam hal ini termasuk distribusi gamma, lognormal dan Weibull, lebih tepat untuk digunakan sebagai komponen *mixture* (Cowell dan Flachaire, 2015). Beberapa penelitian telah menerapkan model *finite mixture* untuk pemodelan distribusi pendapatan, diantaranya model *finite mixture* dengan distribusi Weibull untuk merepresentasikan kelompok negara kaya dan negara miskin yang diestimasi melalui metode estimasi maksimum likelihood (Paap dan van Dijk, 1998) dan model *finite mixture* dengan distribusi lognormal dan gamma sebagai komponennya melalui metode estimasi GMM (Griffiths dan Hajargasht, 2012).

Griffiths dan Hajargasht (2012) meneliti pemodelan *finite mixture* untuk distribusi pendapatan di negara-negara China, India, Pakistan, Rusia, Afrika Selatan, Brazil, dan Indonesia. Model *finite mixture* dengan semua komponen berdistribusi gamma dan lognormal diestimasi melalui pendekatan *Generalized Method of Moments* (GMM). Khusus untuk Indonesia, penelitian dari Griffiths dan Hajargasht tersebut menggunakan data pengeluaran rumah tangga sebagai *proxy* untuk data pendapatan rumah tangga. Data pendapatan rumah tangga mulai digunakan pada penelitian oleh Susanto *et al.*(2018). Melalui pendekatan Bayesian, distribusi pendapatan rumah tangga dapat dimodelkan melalui model *finite mixture* dengan dua atau tiga komponen berdistribusi gamma (Susanto *et al.*, 2019). Namun demikian dalam dua penelitian terdahulu tersebut, data pendapatan rumah tangga diukur secara total keseluruhan dalam tiap rumah tangga. Hal tersebut dapat memberikan kesimpulan yang keliru berkaitan dengan kesejahteraan ekonomi, karena tidak mempertimbangkan banyaknya anggota rumah tangga yang menjadi tanggungan dalam rumah tangga tersebut.

Artikel penelitian ini membahas pemodelan distribusi pendapatan rumah tangga per kapita di Indonesia dengan model *finite mixture* menggunakan metode estimasi maksimum likelihood melalui algoritma *Expectation-Maximization* (EM). Tiga distribusi probabilitas statistika yaitu gamma, lognormal, dan Weibull yang mempunyai karakteristik kontinu, positif dan *heavy-tailed* menjadi materi utama kajian penelitian sebagai komponen *mixture* dari model *finite mixture*.

2. TINJAUAN PUSTAKA

2.1. Model *Finite Mixture*

Suatu vektor variabel random $\mathbf{x} = [x_1, x_2, ..., x_n]^T$ yang bertipe diskrit atau kontinu dikatakan berasal dari suatu distribusi *finite mixture* $g(x_i)$, jika fungsi densitas probabilitas $g(x_i)$ dapat didefinisikan sebagai

$$g(x_i) = w_1 f_1(x_i) + ... + w_K f_K(x_i) \tag{1}$$

yang berlaku untuk semua $x_i, i = 1, 2, ..., n$ dengan $f_k(x_i)$ adalah fungsi densitas probabilitas untuk semua $k = 1, 2, \dots, K$. $f_k(x_i)$ disebut sebagai komponen *mixture* dan k merupakan banyaknya komponen *mixture*. Parameter $w_1, ..., w_K$ disebut sebagai parameter *weight* atau bobot dan vektor $\mathbf{w} = [w_1,\ ..., w_K]^T$ disebut sebagai vektor parameter bobot dari distribusi *finite mixture*. Nilai–nilai dalam $\mathbf{w}$ harus memenuhi

$$0 \le w_k \le 1 \text{ dan } w_1 + ... + w_K = 1$$

Pada beberapa kasus penerapan, jika diasumsikan semua komponen dari distribusi *finite mixture* berasal dari suatu distribusi probabilitas yang mempunyai vektor parameter $\boldsymbol{\theta}$, maka

$$g(x_i \mid \boldsymbol{\psi}) = w_1 f_1(x_i \mid \boldsymbol{\theta_1}) + ... + w_K f_K(x_i \mid \boldsymbol{\theta_K}) = \sum_{k=1}^{K} w_k f_k(x_i \mid \boldsymbol{\theta}_k) \tag{2}$$

dengan $\boldsymbol{\psi} = [\mathbf{w}, \boldsymbol{\theta}]^T$, $\boldsymbol{\theta} = [\boldsymbol{\theta_1}, ... , \boldsymbol{\theta_K}]^T$ (Frühwirth-Schnatter, 2006). Model statistika yang mengimplementasikan konsep distribusi *finite mixture* dalam pemodelannya biasa disebut sebagai model *finite mixture*.

2.2. Metode Estimasi Maksimum Likelihood

Jika diberikan observasi $x_i,\ i = 1, 2,\ ..., n$ saling independen dan $\boldsymbol{\psi} = [\mathbf{w}, \boldsymbol{\theta}]^T$, maka fungsi likelihood dari model *finite mixture* (2) didefinisikan dengan

$$L(\boldsymbol{\psi}) = \prod_{i=1}^{n} \left[\sum_{k=1}^{K} w_k f_k(x_i \mid \boldsymbol{\theta}_k) \right] \tag{3}$$

Pada pendekatan metode estimasi maksimum likelihood untuk estimasi parameter model *finite mixture*, penduga maksimum likelihood $\hat{\boldsymbol{\psi}}$ merupakan penyelesaian dari persamaan

$$\frac{\partial \ell(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} = 0 \tag{4}$$

dengan

$$\ell(\boldsymbol{\psi}) = \ln L(\boldsymbol{\psi}) = \sum_{i=1}^{n} \ln \left[\sum_{k=1}^{K} w_k f_k(x_i \mid \boldsymbol{\theta}_k) \right]$$

merupakan fungsi log-likelihood observasi. Proses penyelesaian persamaan (4) tidak mudah dilakukan, baik secara analitik maupun secara pendekatan numerik melalui metode Newton Raphson, hal ini disebabkan karena fungsi likelihood (3) tidak dalam bentuk *closed-form*. Salah satu langkah untuk mengatasi masalah adalah dengan menggunakan algoritma *Expectation-Maximization* (EM) yang diterapkan pertama kali untuk pemodelan *finite mixture* oleh Dempster, *et al.*(1977).

Penerapan konsep EM untuk model *finite mixture* memandang bahwa observasi x_i dianggap sebagai data yang tidak lengkap dan suatu variabel alokasi-komponen z_i merupakan bagian dari pelengkap data yang tidak terobservasi atau data hilang tersebut. Nilai $z_{ik} = 1$, jika x_i masuk dalam komponen *mixture* *k* dan $z_{ik} = 0$, jika x_i tidak masuk dalam komponen *mixture* *k* tersebut. Hal ini berarti setiap observasi x_i berasal dari salah satu komponen dari model *finite mixture*. Sehingga fungsi log-likehood untuk data lengkap diberikan oleh

$$\ell_c(\boldsymbol{\psi}) = \sum_{i=1}^{n}\sum_{k=1}^{K} z_{ik} \ln\left[w_k f_k(x_i \mid \boldsymbol{\theta}_k)\right] \tag{5}$$

Prosedur algoritma EM berjalan dalam dua tahap yaitu tahap *Expectation* (E) dan tahap *Maximization* (M). Diberikan nilai awal $\boldsymbol{\psi}^{(0)}$, algoritma EM untuk estimasi model *finite mixture* secara umum dapat disusun sebagai berikut,

1. Tahap E.
   Pada tahap *expectation*, saat iterasi ke-*s*, dihitung nilai harapan fungsi log-likelihood untuk data lengkap yang didefinisikan sebagai

   $$\begin{aligned} Q(\boldsymbol{\psi};\boldsymbol{\psi}^{(s)}) &= E(\ell_c(\boldsymbol{\psi}) \mid x_i, \boldsymbol{\psi}^{(s)}) \\ &= \sum_{i=1}^{n}\sum_{k=1}^{K} z_{ik}{}^{(s)} \left\{\ln(w_k{}^{(s)}) + \ln(f_k(x_i \mid \boldsymbol{\theta}_k{}^{(s)})\right\} . \end{aligned}$$

   $z_{ik}{}^{(s)}$ merupakan probabilitas observasi x_i $i = 1, 2, ..., n$ untuk menjadi anggota atau bagian dalam komponen *mixture* k, $k = 1, 2, ..., K$, pada saat iterasi ke-*s*. Sehingga pada tahap ini penduga $\hat{z}_{ik}$ diproses melalui

   $$\hat{z}_{ik}{}^{(s)} = \frac{w_k{}^{(s)} f_k(x_i \mid \boldsymbol{\theta}_k{}^{(s)})}{\sum_{h=1}^{k} w_h{}^{(s)} f_h(x_i \mid \boldsymbol{\theta}_h{}^{(s)})} \tag{6}$$

   dengan $i = 1, 2, ..., n$ dan $k = 1, 2, ..., K$. Penduga $\hat{z}_{ik}{}^{(s)}$ dari tahap E selanjutnya digunakan untuk tahap M pada saat iterasi ke-(*s+1*).

2. Tahap M
   Pada iterasi ke-(*s+1*), parameter bobot w_k diduga dengan

   $$\hat{w}_k{}^{(s+1)} = \frac{\sum_{i=1}^{n} \hat{z}_{ik}{}^{(s)}}{n} \tag{7}$$

   dan penduga $\hat{\boldsymbol{\theta}}_k{}^{(s+1)}$ merupakan penyelesaian dari

   $$\frac{\partial}{\partial \boldsymbol{\theta}} \left\{ \sum_{i=1}^{n}\sum_{k=1}^{K} \hat{z}_{ik}{}^{(s)} \ln f_k(x_i \mid \boldsymbol{\theta}_k) \right\} = 0 \tag{8}$$

   Tahap E dan tahap M dilakukan bergantian secara iteratif. Prosedur akan berhenti pada saat fungsi likelihood, $L(\boldsymbol{\psi})$, pada persamaan (3) bernilai tidak turun setelah proses iterasi dari algoritma EM. Hal ini berarti setelah proses iterasi ke-(*s+1*) berlaku

   $$L(\hat{\boldsymbol{\psi}}^{(s+1)}) \geq L(\hat{\boldsymbol{\psi}}^{(s)}) \tag{9}$$

   untuk $s = 0, 1, 2, ...$ dengan $\hat{\boldsymbol{\psi}} = [\hat{\mathbf{w}}, \hat{\boldsymbol{\theta}}]^T$ (Dempster dkk, 1977).

   Algoritma EM diterapkan untuk estimasi parameter dari tiga model *finite mixture* yaitu model *finite mixture* gamma, model *finite mixture* lognormal dan model *finite mixture* Weibull. Pada estimasi parameter model *finite mixture* gamma, $f(x_i \mid \boldsymbol{\theta}_k)$ pada persamaan (2) adalah distribusi gamma yang didefinisikan sebagai

$$f(x_i \mid \boldsymbol{\theta}_k) = f(x_i \mid \alpha_k, \beta_k) = \beta_k^{\alpha_k} \left(\Gamma(\alpha_k)\right)^{-1} x_i^{\alpha_k - 1} e^{-\beta_k x_i}$$

dengan $\boldsymbol{\theta}_k = [\alpha_k, \beta_k]^T$, $\alpha_k > 0$, $\beta_k > 0$, dan $f(x_i \mid \boldsymbol{\theta}_k) \sim Gam(\alpha_k, \beta_k)$. Sementara pada estimasi parameter model *finite mixture* lognormal, $f(x_i \mid \boldsymbol{\theta}_k)$ merupakan distribusi lognormal, $f(x_i \mid \boldsymbol{\theta}_k) \sim Logn(\mu_k, \sigma_k^2)$,

$$f(x_i \mid \boldsymbol{\theta}_k) = f(x_i \mid \mu_k, \sigma_k^2) = \frac{1}{x_i \sigma_k \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln x_i - \mu_k}{\sigma_k}\right)^2\right]$$

dengan $\boldsymbol{\theta}_k = [\mu_k, \sigma_k^2]^T$, $\mu_k > 0$, dan $\sigma_k > 0$. Sedangkan estimasi parameter model *finite mixture* Weibull, $f(x_i \mid \boldsymbol{\theta}_k)$ berdistribusi Weibull, $f(x_i \mid \boldsymbol{\theta}_k) \sim Wei(\delta_k, \lambda_k)$,

$$f(x_i \mid \boldsymbol{\theta}_k) = f(x_i \mid \delta_k, \lambda_k) = \frac{\delta_k}{\lambda_k}\left(\frac{x_i}{\lambda_k}\right)^{\delta_k - 1} \exp\left(-x_i / \lambda_k\right)^{\delta_k}$$

dengan $\boldsymbol{\theta}_k = [\delta_k, \lambda_k]^T$, $\delta_k \geq 0$, dan $\lambda_k \geq 0$.

2.3. Uji Identifikasi Distribusi Data

Uji Anderson Darling digunakan untuk mengidentifikasi distribusi probabilitas dari suatu data. Penerapan uji Anderson Darling diperlukan terutama saat ekor dari distribusi data menjadi bagian yang penting (Ang dan Tang, 2007). Oleh karenanya, uji ini lebih sesuai diterapkan untuk distribusi data yang mempunyai karakteristik *heavy-tailed*. Langkah-langkah pengujian dilakukan sebagai berikut:

1. Menentukan pernyataan uji hipotesis.

   H_0 : data terdistribusi probabilitas *unimodal* tertentu

   H_1 : data tidak terdistribusi probabilitas *unimodal* tertentu
2. Menentukan tingkat signifikansi α.
3. Menghitung statistik uji Anderson Darling

   $$AD = -n - \frac{1}{n}\sum_{i=1}^{n}\left[2i - 1\right]\left[\ln\left(F(x_i)\right) + \ln\left(1 - F(x_{n+1-i})\right)\right]$$

   dengan $F(x_i)$ merupakan fungsi distribusi kumulatif untuk x_i

   Dihitung nilai kritis c_α pada tingkat signifikansi α

   $$c_\alpha = a_\alpha\left(1 - \frac{b_\alpha}{n} - \frac{d_\alpha}{n^2}\right)$$

   dengan $a_\alpha, b_\alpha, d_\alpha$ ditunjukan dalam tabel nilai kritis Anderson-Darling.
4. Menentukan daerah kritis

   H_0 ditolak pada tingkat signifikansi α, apabila $A^* > c_\alpha$ atau *p-value* < α .
5. Menentukan kesimpulan

2.4. Uji Signifikansi Model *Finite Mixture*

Uji signifikansi untuk model *finite mixture* digunakan untuk mengetahui model *finite mixture* yang sesuai untuk memodelkan data. Uji disusun berbasis *bootstrap likelihood ratio statistics* (Feng dan McCulloch, 1996). Pernyataan hipotesis dalam proses pengujian menggunakan banyaknya komponen *mixture* K yang sesuai untuk model. Langkah-langkah uji hipotesis diberikan sebagai berikut (Yu, 2018)

1. Menentukan pernyataan uji hipotesis

   $H_0 : K = K_0$

   $H_1 : K = K_1 = K_0 + 1.$

   Hipotesis null (H_0) menyatakan model *finite mixture* mempunyai K_0 komponen *mixture* sedangkan pada hipotesis alternatif (H_1) banyaknya komponen *mixture* dalam model ditambahkan satu komponen dari banyaknya komponen *mixture* dalam hipotesis null, $K_1 = K_0 + 1$.
2. Menentukan tingkat signifikansi α
3. Menghitung statistik uji
   (i) Diberikan vektor observasi $\mathbf{x}$, penduga parameter $\hat{\boldsymbol{\psi}}_0$ untuk model *finite mixture* dari hipotesis null dan penduga parameter $\hat{\boldsymbol{\psi}}_1$ untuk model *finite mixture* dari hipotesis alternatif. Selanjutnya dihitung fungsi log-likelihood observasi berdasar penduga parameter $\hat{\boldsymbol{\psi}}_0, \ell(\hat{\boldsymbol{\psi}}_0)$ dan berdasar penduga parameter $\hat{\boldsymbol{\psi}}_1, \ell(\hat{\boldsymbol{\psi}}_1)$ yang digunakan untuk membentuk *likelihood ratio statistic* (lrs_0) berikut

   $$lrs_0 = -2(\ell(\hat{\boldsymbol{\psi}}_0) - \ell(\hat{\boldsymbol{\psi}}_1))$$

   (ii) Dengan menggunakan penduga parameter $\hat{\boldsymbol{\psi}}_0$ untuk model *finite mixture* dari hipotesis null, dibangkitkan data sampel observasi baru $\mathbf{x}^*$ berdasar model *finite mixture* tersebut. Selanjutnya $\mathbf{x}^*$ digunakan untuk membentuk *likelihood ratio statistic* (*lrs*), lrs_1,

   $$lrs_1 = -2(\ell(\hat{\boldsymbol{\psi}}_0; \mathbf{x}^*) - \ell(\hat{\boldsymbol{\psi}}_1; \mathbf{x}^*))$$

   (iii) Dilakukan proses *bootstrap* sebanyak B kali untuk membentuk vektor $lrs_1^{(1)}$, ..., $lrs_1^{(B)}$ yang diperlukan untuk proses penghitungan *p-value* secara empirik,

   $$p - value_{(empirik)} = \frac{1}{B} \sum_{b=1}^{B} I\left(lrs_1^{(b)} > lrs_0\right)$$

   dengan I fungsi indikator.
4. Menentukan daerah kritis

   H_0 ditolak pada tingkat signifikansi α, apabila $p - value_{(empirik)} < \alpha$.
5. Menentukan kesimpulan.

2.5. Pemilihan Model

Pemilihan model *finite mixture* berkaitan dengan penentuan banyaknya komponen *mixture* yang sesuai untuk mewakili pola pengelompokkan dalam distribusi data. Metode berbasis kriteria informasi, *Akaike Information Criterion* (AIC) dan *Bayesian Information Criterion* (BIC) merupakan dua ukuran yang sering digunakan untuk menentukan banyaknya komponen *mixture* dalam model (Celeux *et al.*, 2018).

Proses seleksi model *finite mixture* berdasar AIC, model dipilih sedemikian sehingga meminimumkan nilai

$$AIC = -2\ln L(\hat{\boldsymbol{\psi}}) + 2p \tag{10}$$

dengan $L(\hat{\boldsymbol{\psi}})$ merupakan fungsi likelihood dari penduga maksimum likelihood $\hat{\boldsymbol{\psi}}$ pada persamaan (3) dan p banyaknya parameter dalam model *finite mixture*. Sedangkan berdasar ukuran BIC yang didefinisikan sebagai

$$BIC = -2\ln L(\hat{\boldsymbol{\psi}}) + p\ln(n) \tag{11}$$

dengan n banyaknya data observasi, model *finite mixture* yang dipilih model yang mempunyai nilai BIC terkecil.

3. METODE PENELITIAN

3.1. Sumber Data

Data pendapatan rumah tangga per kapita diolah berdasar data survei dari *The Fifth Wave of the Indonesia Family Life Survey (IFLS5)* 2014-2015 yang meneliti tentang kehidupan rumah tangga di Indonesia (Strauss *et al.*, 2016). Berdasar data IFLS5, diteliti sumber-sumber pendapatan yang menentukan pendapatan dari setiap anggota rumah tangga dalam tiap rumah tangga. Selanjutnya dengan membagi besarnya total pendapatan pada setiap rumah tangga dengan banyaknya anggota rumah tangga yang menjadi tanggungan, dapat dihitung pendapatan rumah tangga per kapita per tahun. Penelitian ini menggunakan jumlah sampel 10.680 rumah tangga.

3.2. Pemodelan *Finite Mixture*

Pemodelan *finite mixture* yang dilakukan dalam penelitian ini meliputi tahapan-tahapan berikut:

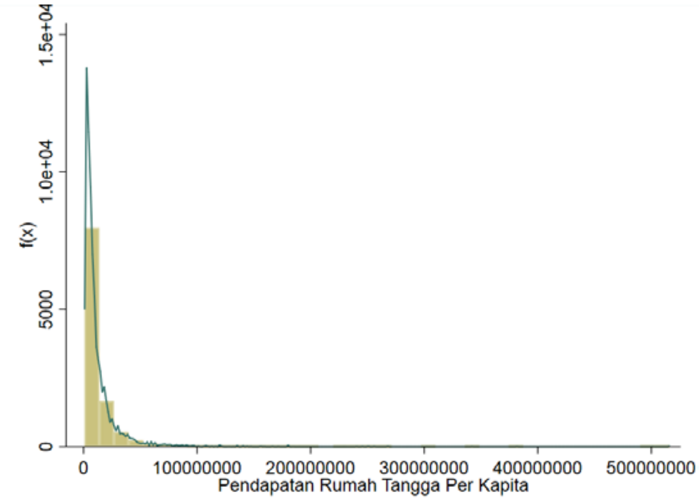
1. Identifikasi distribusi pendapatan rumah tangga per kapita dengan memperhatikan pola bentuk distribusi dan melalui uji Anderson Darling.
2. Estimasi parameter model *finite mixture* dengan metode estimasi maksimum likelihood melalui algoritma *Expectation-Maximization*.
3. Uji signifikansi model *finite mixture* dengan *bootstrap likelihood ratio statistics test.*
4. Seleksi banyaknya komponen *mixture* dalam model *finite mixture* berdasarkan ukuran AIC dan BIC.

Komputasi dari proses estimasi memanfaatkan perangkat lunak R (R Core Team, 2018).

4. HASIL DAN PEMBAHASAN

4.1. Identifikasi Distribusi Pendapatan

Diberikan observasi x_i sebagai pendapatan rumah tangga per kapita per tahun, selanjutnya distribusi pendapatan diteliti pola distribusinya untuk identifikasi adanya pola *multimodal*. Adanya *multimodal* dalam distribusi pendapatan memberikan arti terjadinya pengelompokkan data yang konvergen secara lokal dan hal ini dapat dianalisis melalui pendekatan pemodelan *finite mixture* (Pittau *et al.*, 2016). Terdapat keterkaitan antara suatu distribusi pendapatan yang mempunyai karaketristik *multimodal* dengan banyaknya komponen dalam distribusi *finite mixture*. Penentuan banyaknya kelompok pendapatan dapat ditentukan dengan mengidentifikasi banyaknya komponen *mixture* dalam model *finite mixture* (Vollmer *et al.*, 2013). Identifikasi pola distribusi pendapatan rumah tangga per kapita per tahun dilakukan melalui plot kernel histogram yang ditunjukkan pada Gambar 1.



Gambar 1 Plot Kernel Histogram Distribusi Pendapatan Rumah Tangga Per Kapita Per Tahun Di Indonesia

Grafik distribusi pendapatan rumah tangga per kapita pada Gambar 1 menunjukkan adanya beberapa puncak dari distribusi atau bersifat *multimodal* dan disamping itu bersifat *heavy-tailed*. Mengingat karakteristik data pendapatan rumah tangga yang bernilai positif dan cenderung *heavy-tailed*, maka distribusi yang sesuai adalah distribusi Lognormal, Gamma, ataupun Weibull. Kemudian berdasarkan uji Anderson Darling yang dipaparkan pada Tabel 1, terlihat bahwa data tidak mengikuti distribusi *unimodal* lognormal, gamma dan Weibull. Hasil uji ini menjadi salah satu indikasi bahwa data terdistribusi *finite mixture*.

Tabel 1 Uji Hipotesis Anderson Darling

Distribusi Probabilitas pada H_0	*AD*	*p-value*	Keputusan*
Gamma	222,963	0,005	H_0 ditolak
Lognormal	16,380	0,005	H_0 ditolak
Weibull	213,396	0,010	H_0 ditolak

Catatan: *pada tingkat signifikansi $\alpha = 0,05$

Pada tahapan berikutnya dilakukan uji signifikansi berbasis *bootstrap likelihood ratio statistics* untuk menguji lebih lanjut penerapan model *finite mixture* untuk distribusi data tersebut.

4.2. Uji Signifikansi Model

Hipotesis null pada uji signifikansi berbasis *bootstrap likelihood ratio statistics* menyatakan bahwa data tidak mengikuti model *finite mixture*, dalam hal ini berarti data mengikuti distribusi probabilitas *unimodal*. Sedangkan hipotesis alternatif menyatakan bahwa data mengikuti model *finite mixture*, dimana dalam hal ini banyaknya komponen *mixture* terkecil yaitu dua komponen *mixture*. Sehingga dalam hipotesis null nilai $K = 1$, sedangkan hipotesis alternatif nilai $K = 2$ yang berarti model *finite mixture* dengan dua komponen. Pernyataan uji hipotesis diberikan sebagai berikut

$$H_0 : K = 1$$
$$H_1 : K = 2$$

Proses pengujian diberlakukan untuk setiap model *finite mixture* gamma, lognormal dan Weibull. Hasil komputasi uji signifikansi model dipaparkan dalam Tabel 2. Pada tabel tersebut nilai *p-value* < α yang berarti data lebih sesuai dimodelkan dengan model *finite mixture* dengan komponen *mixture* yang berdistribusi Gamma, Lognormal, ataupun Weibull.

Tabel 2 Uji Hipotesis Signifikansi Model *Finite Mixture*

Distribusi Komponen *Mixture*	*p-value*	Keputusan*	Kesimpulan
Gamma	0,00	H_0 ditolak	*mixture*
Lognormal	0,00	H_0 ditolak	*mixture*
Weibull	0,00	H_0 ditolak	*mixture*

Catatan: *pada tingkat signifikansi $\alpha = 0,05$

Selanjutnya dilakukan proses seleksi model melalui ukuran AIC dan BIC untuk mengetahui distribusi probabilitas yang sesuai untuk data pada tiap komponen *mixture* dan banyaknya komponen *mixture* yang sesuai untuk model.

4.3. Seleksi Model

Hasil perhitungan AIC dan BIC berdasar persamaan (10) dan (11) untuk setiap distribusi probabilitas dan banyaknya komponen disajikan dalam Tabel 2. Banyaknya komponen yang dihitung maksimal hingga sebanyak 6 komponen *mixture*. Hal ini disebabkan karena pada banyaknya komponen sebanyak 7 dan komponen yang lebih besar dari 7, perhitungan ukuran AIC dan BIC tidak konvergen, sehingga tidak memberikan hasil perhitungan.

Tabel 3 Nilai AIC dan BIC dari Estimasi Model *Finite Mixture*

Distribusi	Banyak Komponen	AIC	BIC
Gamma	2	368024	368061
	3	367553	367611
	4	367520	367600
	5	367288	367389
	6	367464	367588
Lognormal	2	367321	367358
	3	367314	367372
	4	367185*	367265*
	5	367188	367290
	6	367304	367427
Weibull	2	368395	368432
	3	367867	367925
	4	367555	367635
	5	367438	367540
	6	367310	367434

Berdasarkan Tabel 3, nilai AIC dan BIC terendah terjadi untuk model *finite mixture* lognormal dengan empat komponen, dimana AIC = 367185 dan BIC = 367265. Sehingga dapat disimpulkan bahwa estimasi model *finite mixture* yang sesuai untuk distribusi pendapatan rumah tangga per kapita di Indonesia adalah model *finite mixture* lognormal dengan empat komponen. Hasil komputasi estimasi model *finite mixture* diberikan oleh

$$g(x_i|\Psi) = \hat{w}_1 Logn(\hat{\mu}_1, \hat{\sigma}_1{}^2) + \hat{w}_2 Logn(\hat{\mu}_2, \hat{\sigma}_2{}^2) + \hat{w}_3 Logn(\hat{\mu}_3, \hat{\sigma}_3{}^2) + \hat{w}_4 Logn(\hat{\mu}_4, \hat{\sigma}_4{}^2) \quad (12)$$

dengan parameter bobot $\hat{w}_1 = 0,1020$, $\hat{w}_2 = 0,4694$, $\hat{w}_3 = 0,4280$ dan $\hat{w}_4 = 0,0001$. Sedangkan parameter-parameter dari distribusi lognormal pada komponen pertama $\hat{\mu}_1 = 2015487$ dan $\hat{\sigma}_1 = 683854$, pada komponen kedua $\hat{\mu}_2 = 6676115$ dan $\hat{\sigma}_2 = 4561228$, pada komponen ketiga $\hat{\mu}_3 = 20947417$ dan $\hat{\sigma}_3 = 22505321$, dan pada komponen keempat $\hat{\mu}_4 = 508009924$ dan $\hat{\sigma}_4 = 8001307$.

Model (12) memberikan arti terdapat empat kelompok pendapatan yang masing-masing berdistribusi lognormal dengan parameter bobot 0,1020 untuk kelompok pertama, 0,4694 untuk kelompok kedua, 0,4280 untuk kelompok ketiga dan 0,0001 untuk kelompok keempat. Parameter bobot pada model tersebut dapat diinterpretasikan sebagai besarnya prosentase kelompok pendapatan tersebut dalam populasi.

Dengan memperhatikan parameter untuk distribusi lognormal pada masing-masing komponen *mixture*, maka dapat diketahui deskripsi dari masing-masing kelompok pendapatan. Kelompok pertama dengan prosentase 10,2 % dari populasi atau 1.088 rumah tangga mempunyai rata-rata pendapatan rumah tangga per kapita sebesar Rp. 2.015.487 per tahun. Kelompok kedua dengan prosentase 46,94 % dari populasi mempunyai rata-rata pendapatan rumah tangga per kapita Rp. 6.676.115 per tahun. Terdapat 5.012 rumah pada kelompok kedua. Kelompok ketiga mempunyai rata-rata pendapatan rumah tangga per kapita Rp. 20.947.417 per tahun. Sebanyak 4.570 rumah tangga atau 42,8 % dari populasi yang termasuk dalam kelompok ketiga. Sedangkan kelompok keempat hanya terdapat 10 rumah tangga atau 0,01 % dari populasi yang mempunyai rata-rata pendapatan rumah tangga per kapita sebesar Rp. 508.009.924 per tahun.

Hasil penelitian yang menunjukan kesesuaian model *finite mixture* lognormal ini selaras dengan hasil penelitian sebelumnya yang telah dilakukan oleh Griffiths dan Hajargasht (2012). Meskipun demikian terdapat perbedaan pada jumlah komponen *mixture* dari model, Griffiths dan Hajargasht menemukan tiga komponen *mixture* sementara pada penelitian ini menggunakan empat komponen *mixture*. Hal ini dapat terjadi mengingat Griffiths dan Hajargasht memanfaatkan data pengeluaran rumah tangga berdasar tahun 2008 sebagai *proxy* untuk data pendapatan rumah tangga, sehingga dimungkinkan kondisi ekonomi rumah tangga saat ini telah mengalami perubahan. Apabila dibandingkan dengan hasil penelitian sebelumnya oleh Susanto *et al*., (2018) dan Susanto *et al*., (2019), maka terdapat perbedaan pada distribusi yang digunakan sebagai komponen *mixture*. Pada kedua penelitian tersebut ditemukan bahwa model *finite mixture* gamma sesuai untuk memodelkan distribusi pendapatan rumah tangga. Namun demikian dalam penelitian tersebut data pendapatan rumah tangga tidak mempertimbangkan banyaknya anggota rumah tangga, sehingga hal tesebut dapat menyebabkan pengambilan kesimpulan yang tidak tepat berkaitan dengan kondisi ekonomi suatu rumah tangga.

5. KESIMPULAN

Distribusi pendapatan rumah tangga per kapita di Indonesia mempunyai pola distribusi *multimodal* dan *heavy-tailed*. Hasil penelitian menunjukan bahwa model *finite mixture* lebih tepat untuk merepresentasikan distribusi pendapatan tersebut, sementara distribusi probabilitas yang berkarakteristik *unimodal* tidak sesuai untuk digunakan. Estimasi model melalui metode EM memperlihatkan bahwa model *finite mixture* dengan komponen *mixture* yang berdistribusi lognormal, gamma atau Weibull sesuai untuk diimplementasikan. Berdasarkan proses seleksi model dengan ukuran AIC dan BIC dapat disimpulkan bahwa model *finite mixture* lognormal dengan empat komponen *mixture* merupakan model yang terbaik. Empat komponen *mixture* dalam model *finite mixture* lognormal tersebut mendeskripsikan sebagai empat kelompok pendapatan. Kelompok pertama dengan rata-rata pendapatan rumah tangga per kapita sebesar Rp. 2.015.487 per tahun. Kelompok kedua mempunyai rata-rata pendapatan rumah tangga per kapita Rp. 6.676.115 per tahun. Kelompok ketiga dengan rata-rata pendapatan rumah tangga per kapita Rp. 20.947.417 per tahun dan kelompok keempat rata-rata pendapatan rumah tangga per kapita sebesar Rp. 508.009.924 per tahun.

UCAPAN TERIMA KASIH

Artikel ini merupakan bagian dari program kegiatan Penelitian Mandiri Aktif UNS tahun 2019. Penulis mengucapkan terima kasih kepada Lembaga Penelitian dan Pengabdian Kepada Masyarakat (LPPM) Universitas Sebelas Maret atas dukungannya terhadap keberlangsungan program kegiatan penelitian ini.

DAFTAR PUSTAKA

Ang, A. H. -S. dan Tang, W. H. 2007. *Probability ConceptsiIn Engineering : Emphasis on Applications in Civil & Environmental Engineering*. Wiley.

Celeux, G., Fruewirth-Schnatter, S., dan Robert, C. P. 2018. Model Selection for Mixture Models - Perspectives and Strategies, *Handbook of Mixture Analysis*. CRC Press, <hal-01961077>.

Chotikapanich, D. dan Griffiths, W. E. 2008. Estimating Income Distributions Using a Mixture of Gamma Densities, *Modeling Income Distributions and Lorenz Curves*. New York, NY: Springer New York, hal 285-302.

Cowell, F. A. dan Flachaire, E. 2015. Statistical Methods for Distributional Analysis, *Handbook of Income Distribution*. Elsevier, hal 359-465.

Dempster, A. P., Laird, N. M., dan Rubin, D. B. 1977. Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 39, No. 1, hal 1-22.

Feng, Z. D. dan McCulloch, C. E. 1996. Using Bootstrap Likelihood Ratios in Finite Mixture Models. *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 58, No. 3, hal 609-617.

Frühwirth-Schnatter, S. 2006. *Finite Mixture and Markov Switching Models*. New York: Springer.

Griffiths, W. E. dan Hajargasht, G. 2012. *GMM Estimation of Mixtures from Grouped Data : An Application to Income Distributions*. Vol. 1148. Department of Economics Working Papers.

Paap, R. dan van Dijk, H. K. 1998. Distribution and Mobility of Wealth of Nations. *European Economic Review*. Vol. 42, No. 7, hal 1269-1293.

Pittau, M. G., Zelli, R. dan Massari, R. 2016. Evidence of Convergence Clubs Using Mixture Models. *Econometric Reviews*, Vol. 35, No. 7, hal 1317-1342.

R Core Team 2018. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Strauss, J., Witoelar, F. dan Sikoki, B. 2016. *The Fifth Wave of the Indonesia Family Life Survey: Overview and Field Report: Volume 1*. RAND Corporation.

Susanto, I., Iriawan, N., Kuswanto, H., dan Suhartono 2019. Bayesian Inference for the Finite Gamma Mixture Model of Income Distribution. *Journal of Physics: Conference Series*, Vol. 1217, hal 012077.

Susanto, I., Iriawan, N., Kuswanto, H., Suhartono, Fithriasari, K., Ulama, B. S. S., Suryaningtyas, W., dan Pravitasari, A. A. 2018. On The Markov Chain Monte Carlo Convergence Diagnostic of Bayesian Finite Mixture Model for Income Distribution. *Journal of Physics: Conference Series*, Vol. 1090, hal 012014.

Vollmer, S., Holzmann, H., dan Schwaiger, F. 2013. Peaks vs Components. *Review of Development Economics*, Vol. 17, No. 2, hal 352-364.

Yu, Y. 2018. *mixR: Finite Mixture Modeling for Raw and Binned Data* [Daring]. R Package Version 0.1.1. Available from: https://cran.r-project.org/package=mixR.