II. LANDASAN TEORI

Dalam proses penelitian pendugaan parameter dari suatu distribusi diperlukan beberapa konsep dan teori yang mendukung dari ilmu statistika. Berikut akan dijelaskan beberapa konsep dan teori yang berkaitan dengan pendugaan parameter GLD dengan Metode Kemungkinan Maksimum menggunakan *software* R.

2.1 *Generalized Lambda Distribution* (GLD)

Keluarga distribusi Lambda Tukey didefinisikan oleh fungsi persentil $Q(u)$ yang berasal dari distribusi lambda satu parameter yang diusulkan oleh John Tukey (1960).

$$Q(u) = \begin{cases} \dfrac{u^{\lambda} - (1-u)^{\lambda}}{\lambda} & ,\lambda \neq 0 \\ \dfrac{\log(u)}{1-u} & ,\lambda = 0 \end{cases} \quad \text{dimana } 0 \leq u \leq 1$$

Distribusi Lambda Tukey digeneralisasi dengan tujuan untuk membangkitkan peubah acak dalam pembelajaran simulasi Monte Carlo ke dalam empat parameter GLD oleh Ramberg dan Schmeiser (1972-1974), dan Mykytka (1979). (Aljazar, 2005).

Generalized Lambda Distribution (GLD) dengan parameter $\lambda_1, \lambda_2, \lambda_3$, dan λ_4, GLD $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, dengan fungsi persentilnya (invers dari fungsi distribusinya F(x)),

$$F^{-1}(x) = Q(y) = Q(y; \lambda_1, \lambda_2, \lambda_3, \lambda_4) = \lambda_1 + \frac{y^{\lambda_3} - (1-y)^{\lambda_4}}{\lambda_2} \tag{2.1}$$

dengan $0 \leq y \leq 1$

Parameter λ_1 dan λ_2 menunjukkan parameter lokasi dan parameter skala (*scale parameter*), serta λ_3 dan λ_4 menunjukkan kemiringan (*skewness*) dan keruncingan (*kurtosis*) dari GLD $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$. (Karian dan Dudewicz, 2000).

Dalam menduga parameter GLD diperlukan fungsi kepekatan peluang GLD. Fungsi kepekatan peluang GLD akan dijelaskan pada Subbab 2.2.

2.2. Fungsi Kepekatan Peluang GLD

Untuk GLD $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, fungsi kepekatan peluangnya adalah

$$f(x; \lambda_1, \lambda_2, \lambda_3, \lambda_4) = \frac{\lambda_2}{\lambda_3 y^{\lambda_3 - 1} + \lambda_4 (1-y)^{\lambda_4 - 1}} \quad ; \; x = Q(y) \tag{2.2}$$

Bukti :

Jika $x = Q(y)$, maka kita memiliki $y = F(x)$. Diturunkan terhadap x, maka diperoleh

$$\frac{dy}{dx} = f(x)$$

Atau

$$f(x) = \frac{dy}{d\big(Q(y)\big)} = \frac{1}{\frac{d(Q(y))}{dy}}$$

Karena bentuk dari $Q(y)$ pada fungsi peluang dari GLD sudah diketahui, maka :

$$\frac{d\,(Q(y))}{dy} = \frac{d}{dy}\left(\lambda_1 + \frac{y^{\lambda_3} - (1-y)^{\lambda_4}}{\lambda_2}\right)$$

$$= \frac{\lambda_3 y^{\lambda_3 - 1} + \lambda_4 (1-y)^{\lambda_4 - 1}}{\lambda_2}$$

Sehingga,

$$f(x) = \frac{1}{\frac{d(Q(y))}{dy}}$$

$$= \frac{1}{\frac{\lambda_3 y^{\lambda_3 - 1} + \lambda_4 (1-y)^{\lambda_4 - 1}}{\lambda_2}}$$

$$= \frac{\lambda_2}{\lambda_3 y^{\lambda_3 - 1} + \lambda_4 (1-y)^{\lambda_4 - 1}}$$

Jadi terbukti bahwa :

$$f(x) = \frac{\lambda_2}{\lambda_3 y^{\lambda_3 - 1} + \lambda_4 (1-y)^{\lambda_4 - 1}}$$

(Karian dan Dudewicz, 2000).

Teorema 2.1 Peubah Acak GLD $X + \lambda_1$

Jika peubah acak X adalah GLD $(0, \lambda_2, \lambda_3, \lambda_4)$, maka peubah acak $X + \lambda_1$ merupakan GLD $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$,

Bukti :

Jika X adalah GLD $(0, \lambda_2, \lambda_3, \lambda_4)$, maka dari Persamaan (2.1) dapat diperoleh

$$Q(y) = \frac{y^{\lambda_3} - (1-y)^{\lambda_4}}{\lambda_2}$$

Sehingga,

$$F_{X+\lambda_1}(x) = P[X + \lambda_1 \leq x] = P[X \leq x - \lambda_1] = F_X(x - \lambda_1\ )$$

Oleh karena itu $F_X(x - \lambda_1) = y$ yang mengakibatkan $F_{X+\lambda_1}(x) = y$

Menghasilkan

$$x - \lambda_1 = Q_X(y)$$

$$x = \lambda_1 + Q_X(y)$$

$$x = \lambda_1 + \frac{y^{\lambda_3} - (1-y)^{\lambda_4}}{\lambda_2}$$

Ini membuktikan bahwa peubah acak $X + \lambda_1$ merupakan GLD $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$. (Karian dan Dudewicz, 2000).

Teorema 2.2 Peubah Acak GLD $\boldsymbol{X - \lambda_1}$

Jika X adalah suatu peubah acak dari GLD $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, maka $Z = X - \lambda_1$ merupakan GLD $(0, \lambda_2, \lambda_3, \lambda_4)$

Bukti :

Jika X adalah GLD $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, maka

$$Q_X(y) = \lambda_1 + \frac{y^{\lambda_3} - (1-y)^{\lambda_4}}{\lambda_2}$$

dan

$$F_{X-\lambda_1}(x) = P[X - \lambda_1 \leq x] = P[X \leq x + \lambda_1] = F_X(x + \lambda_1\ )$$

Jika ditentukan $F_X(x + \lambda_1) = y$, maka diperoleh

$$x + \lambda_1 = Q_X(y) = \lambda_1 + \frac{y^{\lambda_3} - (1-y)^{\lambda_4}}{\lambda_2}\ , x = Q_{X-\lambda_1}(y)$$

Selain itu $F_{X-\lambda_1}(x) = y$ dimana

$$Q_{X-\lambda_1}(y) = x = \frac{y^{\lambda_3} - (1-y)^{\lambda_4}}{\lambda_2}$$

Ini membuktikan bahwa $X - \lambda_1$ merupakan GLD $(0, \lambda_2, \lambda_3, \lambda_4)$.

(Karian dan Dudewicz, 2000).

Statistika inferensia terdiri dari pengujian hipotesis dan pendugaan. Pada penelitian ini akan dilakukan pendugaan parameter. Pendugaan parameter dilakukan untuk menduga ukuran dari suatu populasi yang belum diketahui. Definisi pendugaan parameter akan dijelaskan pada Subbab 2.3.

2.3 Pendugaan Parameter

Dalam statistika inferensia dibutuhkan pemahaman mengenai kaidah-kaidah pengambilan kesimpulan tentang suatu parameter populasi berdasarkan karakteristik sampel. Hal ini membangun apa yang disebut dengan pendugaan titik dari suatu fungsi kepekatan peluang parameter yang tidak diketahui.

Definisi 2.1

Misal suatu peubah acak X memiliki fungsi kepekatan peluang yang bergantung pada suatu parameter tak diketahui θ dengan sebarang nilai dalam suatu himpunan ruang parameter Ω, maka dinotasikan dengan

$f(x;\theta),\ \theta \in \Omega.$

Definisi 2.2

Misal $X_1, X_2, \dots, X_n$ berdistribusi bebas stokastik identik dengan fungsi kepekatan peluang $f(x;\theta)$, $\theta \in \Omega$. Suatu statistik $U(X_1, X_2, \dots, X_n) = U(X)$ yang digunakan untuk menduga $g(\theta)$ disebut sebagai penduga bagi $g(\theta)$.

Berkaitan dengan pendugaan parameter akan dijelaskan beberapa sifat penduga yang baik sebagai berikut:

1. Tak Bias

Penduga $U(X)$ dikatakan sebagai penduga tak bias bagi $g(\theta)$ jika

$E\big(U(X)\big) = g(\theta), \forall\theta \in \Omega$

2. Varians Minimum

Misal T menyatakan suatu penduga tak bias $g(\theta)$, maka T disebut penduga varians minimum jika

$$Var\ (T) \geq \frac{(\frac{\partial}{\partial\boldsymbol{\theta}} g(\theta))^2}{n \cdot E\left[\frac{\partial}{\partial\boldsymbol{\theta}} ln\ f(X;\theta)\right]^2}$$

3. Konsisten

Penduga $U(X)$dikatakan sebagai penduga konsisten bagi $g(\theta)$ jika $U(X) \xrightarrow{p} g(\theta)$ untuk $n \to \infty, \forall\theta \in \Omega$ yaitu bila

$P\{|U(X) - g(\theta)| \geq \varepsilon\} = 0$

(Hoog & Craig, 1995)

Untuk menduga parameter dari suatu distribusi dapat dilakukan dengan beberapa metode. Dalam penelitian ini pendugaan parameter GLD akan dilakukan dengan menggunakan Metode Kemungkinan Maksimum. Definisi Metode Kemungkinan Maksimum akan dijelaskan pada Subbab 2.4.

2.4 Metode Pendugaan Kemungkinan Maksimum (*Maximum Likelihood Estimation Method)*

Definisi 2.3

Misalkan $X_1, X_2, \dots, X_n$ adalah sampel acak berukuran n yang saling bebas stokastik identik dari suatu distribusi yang mempunyai fungsi kepekatan peluang $f(x;\theta), \theta \in \Omega$. Fungsi kepekatan peluang bersama dari $X_1, X_2, \dots, X_n$ adalah $f\ (x_1;\theta)\, f(x_2;\theta) \dots f(x_n;\theta)$ yang merupakan fungsi kemungkinan (*Likelihood Function).* Untuk $x_1, x_2, \dots, x_n$ tetap, fungsi kemungkinan merupakan fungsi dari θ dan dilambangkan dengan $L\ (\theta)$ dan dinotasikan sebagai berikut:

$$
\begin{aligned}
L\ (\theta) &= f\ (\tilde{x}\ ;\theta) \\
&= f(x_1, x_2, \dots, x_n;\theta) \\
&= f\ (x_1;\theta)\, f(x_2;\theta) \dots f(x_n;\theta), \qquad \theta \in \Omega \\
&= \prod_{i=1}^{n} f(x_i;\theta)
\end{aligned}
$$

Definisi 2.4

$L\ (\theta) = f(x_1, x_2, \dots, x_n;\theta), \theta \in \Omega$ merupakan fungsi kepekatan peluang dari $x_1, x_2, \dots, x_n$. Untuk hasil pengamatan $x_1, x_2, \dots, x_n$, nilai $\hat{\theta}$ berada dalam Ω $\left(\hat{\theta} \in \Omega\right)$, dimana $L\ (\theta)$ maksimum yang disebut sebagai *Maximum Likelihood Estimation* (MLE) dari θ. Jadi, $\hat{\theta}$ merupakan penduga bagi θ.

Jika $f\left(x_1, x_2, \dots, x_n;\hat{\theta}\right) = max\ f(x_1, x_2, \dots, x_n;\theta),\ \ \theta \in \Omega$

Maka untuk memaksimumkan $L\ (\theta)$ dapat diperoleh dengan mencari turunan dari $L\ (\theta)$ terhadap parameternya. Biasanya mencari turunan dari $L\ (\theta)$ terhadap parameternya relatif sulit, sehingga dalam penyelesaiannya dapat diatasi dengan menggunakan logaritma atau fungsi *ln* dari $L\ (\theta)$ yaitu:

$ln\ L(\theta) = \sum_{i=1}^{n} ln\ f(x_i; \theta)$

Karena fungsi ln merupakan fungsi monoton naik, maka memaksimumkan $ln\ L(\theta)$ setara dengan memaksimumkan $L\ (\theta)$. Untuk memaksimumkan $ln\ L(\theta)$ adalah dengan mencari turunan dari $ln\ L(\theta)$ terhadap parameternya, dimana hasil turunannya disamadengankan nol.

$$\frac{\partial\ ln\ L(\theta)}{\partial\theta} = 0$$

(Hoog & Craig, 1995)

Dalam menduga parameter dari suatu distribusi ada penduga parameter yang tidak dapat diselesaikan secara analitik, sehingga perlu diselesaikan dengan cara numerik. Salah satu cara yang digunakan adalah dengan teknik iteratif yaitu Metode Newton Raphson. Metode Newton Raphson sering digunakan karena metode ini lebih sederhana dan mempunyai konvergensi yang cepat. Subbab 2.5 akan menjelaskan tentang definisi Metode Newton Raphson.

2.5 Metode Newton Raphson

Apabila dalam proses pendugaan parameter di dapat persamaan akhir yang non linear maka tidak mudah memperoleh pendugaan parameter tersebut, sehingga diperlukan suatu metode numerik untuk memecahkan persamaan non linear tersebut. Salah satu metode yang digunakan untuk memecahkan sistem persamaan non linear adalah Metode Newton Raphson. Metode Newton Raphson adalah metode untuk menyelesaikan persamaan non linear secara iteratif .

Jika $\theta_0$merupakan nilai awal (inisialisasi) dari θ atau θ_0 merupakan nilai ke-1 dari θ, maka dapat dimisalkan $\theta_0 = \theta_i$ dan $\theta = \theta_{i+1}$ dengan i awal = 0.

Metode ini dapat diperluas untuk menyelesaikan sistem persamaan dengan lebih dari satu parameter. Misal $\theta_1, \theta_2, \dots, \theta_p$ maka iterasinya sebagai berikut:

$\theta_{i+1} = \theta_i - [H^{-1}g]$

Dimana $\hat{\theta}_{i+1} = \begin{bmatrix} \hat{\theta}_{1_{i+1}} \\ \vdots \\ \hat{\theta}_{p_{i+1}} \end{bmatrix}$ dan $\hat{\theta}_i = \begin{bmatrix} \hat{\theta}_{1_i} \\ \vdots \\ \hat{\theta}_{p_i} \end{bmatrix}$

Vektor gradien atau vektor turunan pertama terhadap parameternya dan dilambangkan dengan $g(\theta)$yaitu :

$$g(\theta) = \frac{\partial\, In\, L(\theta)}{\partial \theta} = \begin{bmatrix} \frac{\partial\, In\, L(\theta)}{\partial \theta_1} \\ \frac{\partial\, In\, L(\theta)}{\partial \theta_2} \\ \vdots \\ \frac{\partial\, In\, L(\theta)}{\partial \theta_n} \end{bmatrix}$$

Matriks Hessian atau matriks turunan kedua terhadap parameternya, dilambangkan dengan $\mathrm{H}(\theta)$ yaitu :

$$H(\theta) = \frac{\partial^2 In\, L(\theta)}{\partial \theta \partial \theta'} = \begin{bmatrix} \frac{\partial^2\, In\, L(\theta)}{\partial {\theta_1}^2} & \frac{\partial^2\, In\, L(\theta)}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2\, In\, L(\theta)}{\partial \theta_1 \partial \theta_n} \\ \frac{\partial^2\, In\, L(\theta)}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2\, In\, L(\theta)}{\partial {\theta_2}^2} & \cdots & \frac{\partial^2\, In\, L(\theta)}{\partial \theta_2 \partial \theta_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2\, In\, L(\theta)}{\partial \theta_n \partial \theta_1} & \frac{\partial^2\, In\, L(\theta)}{\partial \theta_n \partial \theta_2} & \cdots & \frac{\partial^2\, In\, L(\theta)}{\partial {\theta_n}^2} \end{bmatrix}$$

(Seber and Wild, 2003).

Pada penelitian ini untuk memudahkan melakukan proses iterasi dengan Metode Newton Raphson peneliti menggunakan *software* R. Penjelasan mengenai *software* R akan dijelaskan pada Subbab 2.6.

2.6 Program *R*

R adalah perangkat lunak bebas untuk komputasi statistik dan grafik. Merupakan proyek GNU *General Public License Free Software Foundation* yang mirip dengan bahasa S yang dikembangkan di Bell Laboratories oleh Jhon Chambers dan rekan. R menyediakan berbagai statistik seperti linear dan nonlinear *modeling*, pengujian analisis klasik, analisis *time-series*, klasifikasi dan lainnya. Sebuah rangkaian perangkat lunak yang digunakan untuk manipulasi data, perhitungan, dan tampilan grafik yang mencakup antara lain sebagai berikut :

a. Penanganan data yang efektif dan penyimpanan data.
b. Rangkaian operator untuk perhitungan array dalam matriks tertentu.
c. Fasilitas grafik untuk analisis data dan menampilkan baik pada layar maupun *hardcopy*.
d. Bahasa pemprograman yang sederhana, berkembang dengan baik dan efektif.

(r-project.org)