Jurnal Diferensial, Vol. 02, No. 01, April 2020

ESTIMASI PARAMETER DISTRIBUSI PARETO

Maria Fransiska Muda[1*], Maria A. Kleden[1], Keristina Br. Ginting[1]

1. Program studi Matematika,Fakultas Sains dan Teknik, Universitas Nusa Cendana
Jl. Adisucipto, Penfui Kupang

*Penulis korespondensi: riskamuda1@gmail.com

ABSTRAK

Distribusi pareto (*Generalized Pareto Distribution/GPD)* diperkenalkan pertama kali oleh Pickands pada tahun 1975. Distribusi pareto lebih sering digunakan untuk pemodelan data ekstrim seperti hidrologi, klimatologi, dan udara. Distribusi Pareto mempunyai dua parameter yaitu parameter skala (σ) dan parameter bentuk (ξ). Hasil estimasi parameter menggunakan metode *maximum likelihood* (MLE) adalah $\hat{\xi}_0 = \dfrac{n^2 s - \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i - n\sum_{i=1}^{n} x_i}$ dan $\hat{\sigma} = \overline{x}$.

Kata kunci: ditribusi Pareto, Metode *maximum likelihood*

1. PENDAHULUAN

Distribusi pareto (*Generalized Pareto Distribution/GPD)* diperkenalkan pertama kali oleh Pickands pada tahun 1975. Distribusi pareto lebih sering digunakan untuk pemodelan data ekstrim seperti hidrologi, klimatologi, dan udara. Fungsi distribusi kumulatif dari distribusi GPD didekati menggunakan metode Blok Maksimal. Distribusi pareto (*Generalized Pareto Distribution/GPD)* mempunyai dua parameter yaitu parameter skala (σ) dan parameter bentuk (ξ).

Fungsi kepadatan peluang dari distribusi pareto adalah sebagai berikut:

$$f(\xi,\sigma|x) = \begin{cases} \frac{1}{\sigma}(1+\frac{\xi x}{\sigma})^{\frac{-1}{\xi}-1} & ,\xi\neq 0 \\ \frac{1}{\sigma}\exp\left(\frac{-x}{\sigma}\right) & ,\xi=0 \end{cases}$$

Parameter didefinisikan sebagai hasil pengukuran yang menggambarkan karakteristik dari suatu populasi. Disisi lain karakteristik sampel didefinisikan sebagai statistik. Sebagai contoh adalah rata-rata populasi *(population mean)* μ, variansi populasi (*population variance)* σ^2, dan koefisien korelasi populasi *(population correlation coefficient*) ρ. Dalam distribusi Pareto, semakin besar nilai parameter ξ mengakibatkan semakin panjang semakin panjang ekor. Hal ini berdampak pada peluang terjadinya nilai ekstrim semakin besar[1].

Parameter biasanya tidak diketahui dan dengan statistik maka harga-harga parameter itu bisa diduga atau diestimasi. Sebagai contoh adalah rata-rata sampel $\bar{x}$ digunakan untuk menaksir rata-rata populasi μ yang tidak diketahui dari pengambilan sampel suatu populasi. Dalam statistik non-parametrik, parameter yang cukup menarik untuk dikaji adalah median populasi. Parameter ini sering digunakan untuk analisis statistik non-parametrik untuk mengganti rata-rata populasi sebagai ukuran untuk lokasi yang lebih disukai [2]

Menduga (*estimasi*) adalah proses yang menggunakan sampel statistik untuk menaksir hubungan atau nilai parameter populasi yang tidak diketahui. Pendugaan merupakan suatu pernyataan mengenai parameter populasi yang diketahui berdasarkan populasi dari sampel, dalam hal ini sampel random, yang diambil dari populasi yang bersangkutan, jadi dengan penduga-penduga ini keadaan populasi dapat diketahui.

Misalkan terdapat sebuah peubah acak X yang mengikuti sebaran tertentu dengan nilai yang diamati $x, x_2, x_3, \dots, x_n$ jika nilai-nilai pengamatan mempunyai peluang yang sama untuk diperoleh, maka nilai tengahnya :

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$= \frac{\sum_{i=1}^{n} x_i}{n}$$

$\bar{X}$ merupakan suatu penduga titik (point estimate) dan nilai tengah populasi μ Penduga titik ini dinotasikan dengan $\hat{\mu}$ karena merupakan penduga dari μ.

Metode Maksimum Likelihood

1. Fungsi Likelihood

Fungsi Likelihood dari n variabel acak $x_1, x_2, \dots\dots\dots, x_n$ didefenisikan sebagai fungsi kepadatan bersama dari n variabel random. Fungsi kepadatan bersama $f_{x_{1,\dots,}x_n}(x_1, \dots, x_n;\theta)$, yang mempertimbangkan fungsi dari θ. Jika $x_1, x_2, \dots\dots\dots, x_n$ merupakan sampel acak berukuran n dari fungsi kepadatan peluang $f(x;\theta)$ maka fungsi Likelihoodnya adalah $f(x_1;\theta)\, f(x_2;\theta), \dots\dots, f(x_n;\theta)$.

2. Estimasi Maksimum Likelihood

Suatu penduga bersifat tak bias,efisien dan konsisten dapat diketahui dengan menggunakan suatu metode yaitu metode *maksimum likelihood.* Metode ini sering memberikan hasil (penaksir) yang baik.

Definisi :

Misalkan $X_1, X_2, \dots, X_n$ peubah acak dengan fungsi distribusi $(F(x_1, x_2, \dots, x_n|\theta)$ dengan $\theta\epsilon\Omega$ yang tidak diketahui,maka fungsi likelihoodnya adalah :

$L(\theta) = \{ f(x_1, x_2, \dots, x_n|\theta \text{ ; } Jika\ F\ mempunyai\ fungsi\ padat\ f \}$

Untuk setiap $\hat{\theta}$=$\hat{\theta}_n$ $(x_1, x_2, \dots, x_n)\epsilon\Omega$,sehingga $L(\hat{\theta})$ =$(L(\theta): \theta\epsilon\Omega)$ disebut maksimum likelihood estimation.

2. METODE PENELITIAN

Penelitian ini merupakan penelitian kajian pustaka. Melalui kajian dari berbagai referensi untuk mengestimasi parameter menggunakan metode maksimum likelihood.. Salah satu metode untuk mengestimasi parameter *Generalized Pareto Distriutioni* (GPD) adalah *Maximum Likelihood Estimation* (MLE). Cara kerja metode ini adalah memaksimumkan fungsi *Likelihood* yang merupakan fungsi peluang bersama dari $X_1, X_2, \dots X_n$ yang mempunyai fungsi kepadatan peluang sebagai berikut:

$$f(\xi, \sigma | x) = \begin{cases} \frac{1}{\sigma}(1+\frac{\xi x}{\sigma})^{\frac{-1}{\xi}-1} & ,\xi \neq 0 \\ \frac{1}{\sigma}\exp\left(\frac{-x}{\sigma}\right) & ,\xi = 0 \end{cases}$$

Tahapan metode maksimum Likelihood adalah membentuk fungsi likelihood; membentuk fungsi log likelihood; menurunkan fungsi log likelihood; dan menyelesaikan turunan pertama sama dengan nol.

3. HASIL DAN PEMBAHASAN

Salah satu metode untuk mengestimasi parameter *Generalized Pareto Distriutioni* (GPD) adalah *Maximum Likelihood Estimation* (MLE). Cara kerja metode ini adalah memaksimumkan fungsi *Likelihood* yang merupakan fungsi peluang bersama dari $X_1, X_2, \dots X_n$

$$f(\xi, \sigma | x) = \begin{cases} \frac{1}{\sigma}(1+\frac{\xi x}{\sigma})^{\frac{-1}{\xi}-1} & ,\xi \neq 0 \\ \frac{1}{\sigma}\exp\left(\frac{-x}{\sigma}\right) & ,\xi = 0 \end{cases}$$

Fungsi *Likelihood* dari *probability density Generalized Pareto Distriutioni* (GPD) untuk $\xi \neq 0$ adalah sebagai berikut :

$$L(\xi,\sigma \mid x_1,x_2,...x_n)=\prod_{i=1}^{n} f(x)$$

$$L(\xi,\sigma \mid x_1,x_2,...x_n)=\prod_{i=1}^{n}\frac{1}{\sigma}\left(1+\frac{\xi x}{\sigma}\right)^{-\left(\frac{1}{\xi}+1\right)}$$

$$L(\xi,\sigma \mid x_1,x_2,...x_n)=\sigma^{-n}\prod_{i=1}^{n}\left(1+\frac{\xi x_i}{\sigma}\right)^{-\left(\frac{1}{\xi}+1\right)}$$

Selanjutnya fungsi Likelihood dilogaritmakan dengan rumus

$$\ln(\xi,\sigma \mid x_1,x_2,....x_n)=\ln\left(\sigma^{-n}\prod_{i=1}^{n}\left(1+\frac{\xi x_i}{\sigma}\right)^{-\left(\frac{1}{\xi}+1\right)}\right)$$

$$=-n\ln\sigma-\left(\frac{1}{\xi}+1\right)\sum_{i=1}^{n}\ln\left(1+\frac{\xi x_i}{\sigma}\right)$$

Setelah diperoleh fungsi log likelihood, selanjutnya mencari nilai maksimum dari $\ln(\xi,\sigma \mid x_1,x_2,....x_n)$ terhadap masing-masing fungsi dengan menurunkan fungsi paremeter.

Turunan $\ln(\xi,\sigma \mid x_1,x_2,....x_n)$ terhadap ξ:

$$\frac{\partial \ln L}{\partial \xi}=\frac{1}{\xi^2}\sum_{i=1}^{n}\ln\left(1+\frac{\varepsilon x_i}{\sigma}\right)-\left(\frac{1}{\varepsilon}+1\right)\sum_{i=1}^{n}\frac{x_i}{\sigma+\varepsilon x_i}$$

$$\frac{\partial \ln L}{\partial \xi}=0$$

$$0=\frac{1}{\xi^2}\sum_{i=1}^{n}\ln\left(1+\frac{\xi x_i}{\sigma}\right)-\left(\frac{1}{\xi}+1\right)\sum_{i=1}^{n}\frac{x_i}{\sigma+\varepsilon x_i}$$

$$\xi+\xi^2=\frac{\sum_{i=1}^{n}\ln\left(1+\frac{\xi x_i}{\sigma}\right)}{\sum_{i=1}^{n}\frac{x_i}{\sigma+\xi x_i}}$$

$$\xi(1+\xi)=\frac{\sum_{i=1}^{n}\ln\left(1+\frac{\xi x_i}{\sigma}\right)}{\sum_{i=1}^{n}\frac{x_i}{\sigma+\xi x_i}}$$

$$\hat{\xi}=\frac{\sum_{i=1}^{n}\ln\left(1+\frac{\hat{\xi} x_i}{\hat{\sigma}}\right)}{(1+\hat{\xi})\sum_{i=1}^{n}\frac{x_i}{\hat{\sigma}+\hat{\xi} x_i}}$$

Turunan $\ln(\xi,\sigma \mid x_1,x_2,....x_n)$ terhadap σ:

$$\frac{\partial \ln L}{\partial \sigma} = \sigma^{-1}\left(-n+(1+\xi)\sum_{i=1}^{n}\frac{x_i}{\sigma+\xi x_i}\right)$$

$$\frac{\partial \ln L}{\partial \sigma} = 0$$

$$0 = \sigma^{-1}\left(-n+(1+\xi)\sum_{i=1}^{n}\frac{x_i}{\sigma+\xi x_i}\right)$$

$$\hat{\sigma} = \frac{\left(1+\hat{\xi}-n\hat{\xi}\right)\sum_{i=1}^{n}x_i}{n^2}$$

Persamaan belum dapat memberikan penyelesaian karena bentuknya yang tidak *closed from. Closed from* artinya persamaan yang tidak lagi mengandung parameter didalamnya. Terlihat bahawa persamaan $\hat{\sigma} = \frac{\left(1+\hat{\xi}-n\hat{\xi}\right)\sum_{i=1}^{n}x_i}{n^2}$ masih terdapat parameter ξ di dalam persamaannya. Untuk itu perlu dilakukan analisi lebih lanjut untuk mendapatkan persamaan yang *closed from* . Salah satu penyelesaian untuk persamaan yang tidak *closed from* adalah metode Newton Raphson.

Penggunaan metode Newton Raphson dilakukan dengan menggunakan iterasi-iterasi hingga didapatkan hasil yang konvergen. Persamaan umum Newton Raphson sebagai berikut:

$$\theta_{l+1} = \theta_l - g(\theta_l)H^{-1}(\theta_l)$$

$$g(\theta) = \begin{bmatrix} \frac{\partial \ln L}{\partial \xi} & \frac{\partial \ln L}{\partial \sigma} \end{bmatrix}$$

$$H(\theta) = \begin{bmatrix} \frac{\partial^2 \ln L}{\partial \xi^2} & \frac{\partial^2 \ln L}{\partial \xi \partial \sigma} \\ \frac{\partial^2 \ln L}{\partial \xi \partial \sigma} & \frac{\partial^2 \ln L}{\partial \sigma^2} \end{bmatrix}$$

Turunan kedua dari fungsi ln *likelihood* sebagai berikut:

$$\frac{\partial^2 \ln L}{\partial^2 \xi} = 2\xi^{-3}\left[\xi\sum_{i=1}^{n}\frac{x_i}{\sigma+\xi x_i} - \sum_{i=1}^{n}\ln\left(1+\frac{\xi x_i}{\sigma}\right) + \left(1+\frac{1}{\xi}\right)\sum_{i=1}^{n}\frac{x_i^{\,2}}{(\sigma+\xi x_i)^2}\right]$$

$$\frac{\partial^2 \ln L}{\partial \sigma^2} = \sigma^{-2}\left[n-(1+\xi)\sum_{i=1}^{n}\frac{x_i(2\sigma+\xi x_i)}{(\sigma+\xi x_i)^2}\right]$$

$$\frac{\partial^2 \ln L}{\partial \xi \partial \sigma} = \xi^{-1}\left[(1+\xi)\sum_{i=1}^{n}\frac{x_i}{(\sigma+\xi x_i)^2} - \sigma^{-1}\sum_{i=1}^{n}\frac{x_i}{\sigma+\xi x_i}\right]$$

$$\hat{\xi}_0 = \frac{n^2 s - \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i - n\sum_{i=1}^{n} x_i}$$

Selanjutnya dengan cara yang sama untuk fungsi

$$f(\xi, \sigma \mid x) = \frac{1}{\sigma}\exp\left(\frac{-x}{\sigma}\right), \quad \xi = 0$$

$$f(\xi, \sigma \left| x_1, x_2, ...x_n \right.) = \frac{1}{\sigma}\exp\left(\frac{-x}{\sigma}\right)$$

$$L\left(\xi, \sigma \left| x_1, x_2, ...x_n \right.\right) = \prod_{i=1}^{n} \frac{1}{\sigma}\exp\left(\frac{-x}{\sigma}\right)$$

$$L\left(\xi, \sigma \left| x_1, x_2, ...x_n \right.\right) = \sigma^{-n} e^{\sum_{i=1}^{n} \frac{x_i}{\sigma}}$$

Selanjutnya logaritmakan perasamaan $L\left(\xi, \sigma \left| x_1, x_2, ...x_n \right.\right) = \sigma^{-n} e^{\sum_{i=1}^{n} \frac{x_i}{\sigma}}$.

Diperoleh :

$$\ln L = -n\ln\sigma - \frac{1}{\sigma}\sum_{i=1}^{n} x_i$$

Seperti proses pada Metode Maksimum Likelihood, dicari turunan pertama dari fungsi log likelihood yaitu:

$\frac{\partial \ln L}{\partial \sigma} = \frac{-n}{\sigma} + \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i$ Ag $\frac{\partial \ln L}{\partial \sigma} = 0$ maka diperoleh:

$$0 = \frac{-n}{\sigma} + \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i$$

$$\frac{n}{\sigma} = \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i$$

$$\frac{\frac{1}{\sigma}}{\frac{1}{\sigma^2}} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\hat{\sigma} = \overline{x}$$

4 SIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan maka diperoleh kesimpulan sebagai berikut Estimasi parameter distribusi Pareto menggunakan metode *maximum likelihood* untuk $\xi \neq 0$ diperoleh parameter bentuk:

$$\hat{\xi}_0 = \frac{n^2 s - \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i - n\sum_{i=1}^{n} x_i}$$ untuk $\xi = 0$ mendapatkan parameter skala: $\hat{\sigma} = \overline{x}$

DAFTAR PUSTAKA

[1] P. Ringganis, A. Rusgiono, and A. Prahutama, "Aplikasi Metode Momen Probabilitas Terboboti Untuk Estimasi Parameter Distribusi Pareto Terampat Pada Data Curah Hujan (Studi Kasus : Data Curah Hujan di Kota Semarang Tahun 2004-2013)," vol. 3, no. 4, pp. 821–830, 2014.

[2] S. Jaffarus, S. Setiawan, and S. Sutikno, "Pengukuran Risiko pada Klaim Asuransi 'X' dengan Menggunakan Metode Generalized Extreme Value dan Generalized Pareto Distribution," vol. 1, no. 1, pp. 75–80, 2012.