

Pemodelan Besar Klaim Asuransi Kendaraan Bermotor Indonesia Menggunakan Model Komposit Log-Logistik-Generalized Pareto

ACENG KOMARUDIN MUTAQIN¹, RESYI PANDINI SAFITRI²

Program Studi Statistika, Universitas Islam Bandung
Jl. Ranggagading No. 8 Bandung, 40116
email: aceng.k.mutaqin@gmail.com

ABSTRAK

Dalam artikel ini dibahas pemodelan komposit distribusi log-logistik-Generalized Pareto pada data besar klaim asuransi kendaraan bermotor di Indonesia. Dalam model komposit terdapat nilai ambang batas yang dihitung menggunakan metode heuristik aturan akar kuadrat. Pengujian kecocokan distribusi dilakukan menggunakan uji kecocokan Kolmogorov-Smirnov. Penaksiran parameter untuk masing-masing distribusinya menggunakan metode penaksiran kemungkinan maksimum melalui metode numerik Newton-Raphson. Data yang digunakan adalah data yang diperoleh dari perusahaan asuransi umum PT. XYZ. Data tersebut berisi data besar klaim partial loss pemegang polis produk asuransi comprehensive untuk kendaraan bermotor kategori 6 (jenis kendaraan Truk & Pickup) wilayah 2 yang mencakup DKI Jakarta, Jawa Barat, dan Banten. Berdasarkan hasil penerapan, model komposit log-logistik-Generalized Pareto cocok untuk memodelkan data besar klaim di atas, dengan nilai ambang batas 65.575.000.

Kata Kunci: Model Komposit, Metode Heuristik, Distribusi Log-logistik, Distribusi Generalized Pareto, Newton-Raphson, Kolmogorov-Smirnov.

ABSTRACT

This article discusses the composite modeling of logistic-logistic-Generalized Pareto distribution on the big data of motor vehicle insurance claims in Indonesia. In the composite model, there is a threshold value calculated using the heuristic method of the square root rule. Distribution suitability testing was performed using the Kolmogorov-Smirnov fit test. The estimation of parameters for each distribution uses the maximum likelihood method using the Newton-Raphson numerical method. The data used is data obtained from the general insurance company PT. XYZ. The data contains big data on partial loss claims for policy holders of comprehensive insurance products for motor vehicles category 6 (types of trucks & pickups) region 2 which includes DKI Jakarta, West Java and Banten. Based on the results of the application, the log-logistic-Generalized Pareto composite model is suitable for modeling the claim size data above, with a threshold value of 65.575.000.

Keywords: Composite Model, Heuristic Method, Log-logistic Distribution, Generalized Pareto Distribution, Newton-Raphson, Kolmogorov-Smirnov.

1. PENDAHULUAN

Pemodelan besar klaim merupakan cara untuk memprediksi besar klaim yang dapat digunakan untuk memperkirakan penetapan harga premi asuransi di masa datang. Cadangan biaya klaim dari perusahaan asuransi cenderung lebih besar dipengaruhi oleh besar klaim dari pada frekuensi klaim. Kerugian yang besar merupakan salah satu bagian utama yang harus diperhatikan dalam mengganti biaya yang dikeluarkan perusahaan asuransi. Kerugian besar sering disebut kerugian ekstrim yang mana bisa terjadi sesekali atau beberapa kali. Peristiwa ekstrim sangat menarik bagi aktuaris karena berperan penting dalam prosedur pengambilan keputusan untuk asuransi, misalnya untuk menghitung premi.

Memilih model atau distribusi untuk besar klaim sangatlah penting, terutama jika terdapat data besar klaim yang ekstrim. Besar klaim dalam asuransi kerugian merupakan peubah acak positif, yang cenderung mengikuti distribusi miring dengan ekor kanannya tebal. Beberapa

distribusi dengan ekor kanan tebal telah digunakan dalam asuransi kerugian, diantaranya adalah distribusi gamma, log-normal, log-gamma, Weibull dan Pareto (Kleiber dan Kotz, 2003; Klugman dkk., 2012). Distribusi-distribusi tersebut mungkin cocok digunakan untuk data besar klaim yang nilainya kecil dan menengah. Namun demikian, distribusi-distribusi di atas kemungkinan tidak cocok ketika dalam data besar klaim memuat data besar klaim yang ekstrim. Untuk itu sangat penting untuk memiliki distribusi atau model yang baik untuk data besar klaim yang kecil, menengah dan besar atau ekstrim.

Cebriaan (2003) menyatakan bahwa sifat-sifat *Generalized Pareto Distribution* (GDP) memungkinkan untuk membangun sebuah model komposit yang terdiri dari dua bagian, yaitu (1) model untuk klaim-klaim di bawah ambang batas (*threshold*), dan (2) model untuk klaim-klaim di atas ambang batas. Metode untuk membangun model komposit dikenal sebagai pendekatan *Excess Over Threshold* (EOT) atau pendekatan *Peaks Over Threshold* (POT). Salah satu perhatian utama dari pendekatan EOT adalah pemilihan ambang batas yang tepat untuk satu data set yang diberikan. Ada beberapa pendekatan untuk menaksir ambang batas. Metode penaksiran ambang batas yang tradisional adalah menggunakan grafik untuk memilih ambang batas secara visual. Plot dari *mean residual life* atau plot dari *mean excess* adalah alat yang paling umum di bidang asuransi. Selain itu ada juga plot Gerstengarbe (Cebriaan dkk., 2003). Ada juga metode heuristik untuk memilih ambang batas, diantaranya adalah aturan 10% teratas (DuMouchel, 1983), aturan akar kuadrat (Ferreira dkk., 2003), dan aturan empirik (Loretan dan Phillips, 1994). Pendekatan-pendekatan di atas tidak memiliki justifikasi teoritis, tetapi mudah untuk diimplementasikan dan sering digunakan dalam praktik. Alternatif lainnya adalah pemilihan ambang batas otomatis dengan tujuan menyeimbangkan bias dan varians. Metode yang paling umum digunakan didasarkan pada meminimumkan taksiran Mean Squared Error atau MSE (Caeiro dan Gomes, 2015). Metode bootstrap digunakan oleh Hall (1990) dan Gomes dan Oliveira (2001) untuk memilih ambang batas. Sementara itu Guillou dan Hall (2001) menggunakan prosedur reduksi bias untuk memilih ambang batas optimal.

Wang, dkk. (2019) melakukan studi simulasi untuk mengetahui metode yang optimal dalam pemodelan komposit untuk data asuransi kerugian. Hasil studi simulasi menunjukkan bahwa ketika datanya mencukupi, metode heuristik dengan aturan akar kuadrat memiliki kinerja yang terbaik dalam menaksir cadangan klaim dibandingkan dengan metode yang lainnya. Dalam artikel ini dibahas penerapan model komposit distribusi log-logistik-Generalized Pareto untuk data besar klaim asuransi kerugian di Indonesia dimana ambang batasnya dipilih menggunakan metode heuristik aturan akar kuadrat.

2. DISTRIBUSI LOG-LOGISTIK

Misalkan X merupakan peubah acak yang berdistribusi Log-logistik dengan parameter γ dan η maka fungsi densitas peluangnya adalah sebagai berikut:

$$f(x) = \frac{\left(\frac{\gamma}{\eta}\right)\left(\frac{x}{\eta}\right)^{\gamma-1}}{\left(1 + \left(\frac{x}{\eta}\right)^{\gamma}\right)^2}, \quad x > 0$$

Fungsi distribusi kumulatif dari distribusi Log-logistik adalah:

$$F(x) = \left(1 + \left(\frac{\eta}{x}\right)^{\gamma}\right)^{-1}, \quad x > 0$$

Ekspektasi dan variansi dari distribusi Log-logistik masing-masing adalah:

$$E(X) = \eta B\left(\frac{\gamma+1}{\gamma}, \frac{\gamma-1}{\gamma}\right)$$

$$V(X) = \eta^2 B\left(\left(\frac{\gamma+2}{\gamma}, \frac{\gamma-2}{\gamma}\right)\right) - \left(\eta B\left(\frac{\gamma+1}{\gamma}, \frac{\gamma-1}{\gamma}\right)\right)^2$$

dimana $B(\cdot, \cdot)$ menyatakan fungsi beta.

3. GENERALIZED PARETO DISTRIBUTION (GPD)

Misalkan X merupakan peubah acak yang berdistribusi GPD dengan parameter a adalah parameter bentuk, β adalah parameter skala dan b adalah parameter lokasi. Fungsi densitas peluangnya adalah sebagai berikut:

$$f(x) = \frac{1}{\beta} \left(1 + \alpha \frac{(x-b)}{\beta}\right)^{-\left(1+\frac{1}{\alpha}\right)}, \quad x > b$$

Fungsi distribusi kumulatif dari GPD adalah:

$$F(x) = 1 - \left(1 + \alpha \frac{(x-b)}{\beta}\right)^{-\frac{1}{\alpha}}, \quad x > b$$

Ekspektasi dan variansi dari GPD masing-masing adalah:

$$E(X) = b + \frac{\beta}{1 - \alpha}$$

$$V(X) = \frac{\beta^2}{(1 - \alpha)^2(1 - 2\alpha)}$$

4. MODEL KOMPOSIT

Model komposit biasa digunakan untuk pemodelan data klaim yang memuat nilai ekstrim, dimana nilai klaim yang berada di bawah ambang batas (b), diasumsikan “normal” (Wang dkk., 2019), sehingga mengikuti distribusi besar klaim yang umum, terpancung di kanan pada b ; sedangkan nilai klaim yang berada di atas ambang batas diasumsikan nilai ekstrim. Misalkan f_1 dan f_2 adalah dua fungsi densitas peluang pada $[0, \infty)$ dan $b > 0$. Kemudian F_1 dan F_2 masing-masing menyatakan fungsi distribusi kumulatifnya. Definisikan

$$c_1 = \int_0^b f_1(z) dz = F_1(b)$$

$$c_2 = \int_b^{\infty} f_2(z) dz = 1 - F_2(b)$$

Kemudian misalkan $r \in [0,1]$ dan definisikan fungsi densitas peluang baru yaitu:

$$f(z) = I(z \leq b)r c_1^{-1}f_1(z) + I(z > b)(1 - r)c_2^{-1}f_2(z),$$

dimana:

$$I(z \leq b) = \begin{cases} 1, & z \leq b \\ 0, & z > b \end{cases}$$

Asumsikan secara khusus bahwa $f_2(z) = 0$ untuk $z < b$. Dengan demikian, diperoleh

$$c_2 = \int_b^{\infty} f_2(z) dz = \int_0^{\infty} f_2(z) dz = 1.$$

Dengan demikian diperoleh

$$f(z) = I(z \leq b)r \frac{f_1(z)}{F_1(b)} + I(z > b)(1 - r)f_2(z).$$

Asumsikan bahwa $r = c_1 = F_1(b)$. Sehingga

$$\begin{aligned} f(z) &= I(z \leq b)f_1(z) + I(z > b)(1 - c_1)f_2(z) \\ &= I(z \leq b)f_1(z) + I(z > b)(1 - F_1(b))f_2(z). \end{aligned}$$

Misalkan $f_1 = f_{<b}$ dan $f_2 = f_{>b}$, masing-masing adalah fungsi densitas peluang dari distribusi di bawah dan di atas ambang batas dan $f_{\leq b|z \leq b}$ dan $f_{>b|z > b}$ adalah fungsi densitas peluang bersyarat dari distribusi di bawah dan di atas ambang batas masing-masing bersyarat pada $Z \leq b$ dan $Z > b$. Selanjutnya fungsi densitas peluang dari model komposit dapat dituliskan sebagai

$$\begin{aligned} f(z) &= I(z \leq b)f_{\leq b}(z) + (1 - I(z \leq b))(1 - F_{\leq b}(b))f_{>b}(z) \\ &= I(z \leq b)f_{\leq b|z \leq b}(z | z \leq b)F_{\leq b}(b) \\ &\quad + (1 - I(z \leq b))f_{>b|z > b}(z | z > b)(1 - F_{\leq b}(b)), \quad z > 0 \end{aligned}$$

Pickands (1975) dan Balkema dan de Haan (1974) telah menunjukkan bahwa $f_{>b|z > b}$ untuk peubah yang digeser $[Z - b | Z > b]$, dapat didekati oleh distribusi GPD untuk b yang cukup besar. Disini dibatasi untuk distribusi Pareto tipe II yang merupakan kasus khusus dari GPD dan dalam praktik sangat umum digunakan untuk klaim asuransi yang ekstrim. Ini berarti bahwa

$$f_{>b|z > b}(z | z > b) = \frac{\alpha/\beta}{\left(1 + \frac{z-b}{\beta}\right)^{\alpha+1}}, \quad z > b,$$

dimana α adalah parameter bentuk dan β adalah parameter skala.

Distribusi untuk di bawah ambang batas diberikan oleh

$$f_{\leq b|z \leq b}(z | z \leq b) = \frac{f_{\leq b}(z)}{P(Z \leq b)} = \frac{f_{\leq b}(z)}{F_{<b}(b)}, \quad 0 < z \leq b, \quad (1)$$

dimana $f_{<b}$ bisa distribusi apa saja yang cocok untuk besar klaim yang bernilai kecil dan moderat, seperti gamma, Weibull, lognormal atau log-logistik. Bentuk distribusi yang ada pada Persamaan (1) dikenal sebagai distribusi terpancung kanan.

5. METODE HEURISTIK UNTUK MEMILIH AMBANG BATAS

Metode heuristik merupakan suatu aturan praktis yang sangat mudah untuk diterapkan sehingga banyak digunakan dalam praktik di lapangan. Misalkan $z_{(1)} \leq \dots \leq z_{(n)}$ adalah data yang sudah diurutkan dari kecil ke besar, dan k adalah bilangan riil dalam interval $[0, n]$. Metode heuristik menaksir ambang batas, b , berdasarkan persamaan $\hat{b} = z_{([n-k])}$, dimana $[n-k]$ menyatakan bilangan integer diantara $1, \dots, n$ yang paling dekat ke $n - k$. Ada beberapa pilihan metode untuk menentukan nilai k , salah satunya adalah aturan akar kuadrat yang diusulkan oleh Ferreira dkk. (2003), yaitu $k = \sqrt{n}$.

6. PENAKSIRAN PARAMETER DISTRIBUSI DI BAWAH AMBANG BATAS

Fungsi log-likelihood dari distribusi di bawah ambang batas yang berasal dari distribusi log-logistik terpancung kanan adalah sebagai berikut:

$$l(\gamma; \eta) = n \ln \gamma - \gamma n \ln \eta + (\gamma - 1) \sum_{i=1}^n \ln(z_i) - 2 \sum_{i=1}^n \ln\left(1 + \left(\frac{z_i}{\eta}\right)^\gamma\right) + \sum_{i=1}^n \ln\left(1 + \left(\frac{\eta}{b}\right)^\gamma\right)$$

Taksiran kemungkinan maksimum untuk parameter γ dan η adalah solusi dari dua persamaan berikut:

$$\frac{n}{\gamma} - n \ln \eta + \sum_{i=1}^n \ln(z_i) - 2 \sum_{i=1}^n \frac{\left(\frac{z_i}{\eta}\right)^\gamma \ln\left(\frac{z_i}{\eta}\right)}{\left(1 + \left(\frac{z_i}{\eta}\right)^\gamma\right)} + \sum_{i=1}^n \frac{\left(\frac{\eta}{b}\right)^\gamma \ln\left(\frac{\eta}{b}\right)}{\left(1 + \left(\frac{\eta}{b}\right)^\gamma\right)} = 0$$

$$-\frac{n\gamma}{\eta} - 2 \sum_{i=1}^n \frac{\gamma \left(\frac{z_i}{\eta}\right)^{\gamma-1} \left(-\frac{z_i}{\eta^2}\right)}{\left(1 + \left(\frac{z_i}{\eta}\right)^\gamma\right)} + \sum_{i=1}^n \frac{\gamma \left(\frac{\eta}{b}\right)^{\gamma-1} \left(\frac{1}{b}\right)}{\left(1 + \left(\frac{\eta}{b}\right)^\gamma\right)} = 0$$

7. PENAKSIRAN PARAMETER DISTRIBUSI DI ATAS AMBANG BATAS

Fungsi log-likelihood dari distribusi di atas ambang batas yang berasal dari *generalized pareto distribution* adalah sebagai berikut:

$$l(\alpha; \beta) = n \ln \alpha - (\alpha + 1) \sum_{i=1}^n \ln(\beta + z_i - b) + n \alpha \ln \beta$$

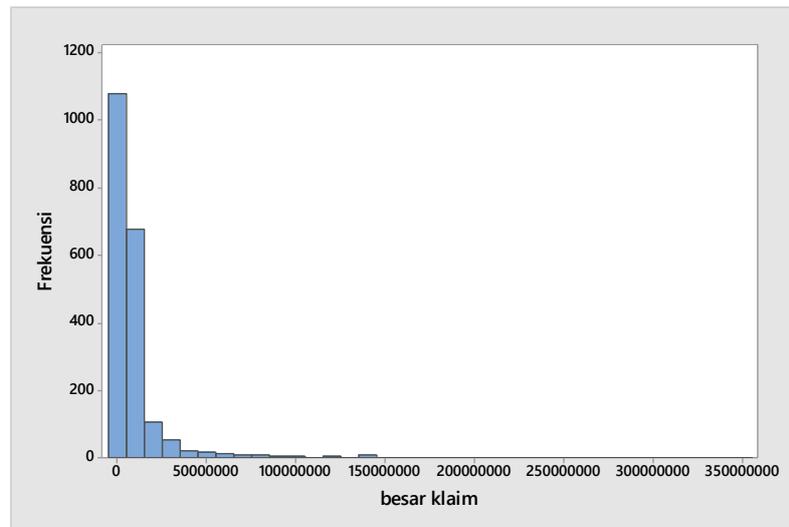
Taksiran kemungkinan maksimum untuk parameter α dan β adalah solusi dari dua persamaan berikut:

$$\frac{n}{\alpha} - \sum_{i=1}^n \ln(\beta + z_i - b) + n \ln \beta = 0$$

$$-(\alpha + 1) \sum_{i=1}^n \frac{1}{(\beta + z_i - b)} + \frac{n\alpha}{\beta} = 0$$

8. APLIKASI

Sebagai bahan aplikasi adalah data besar klaim *partial loss* pemegang polis asuransi umum PT. XYZ tahun 2014 produk asuransi kendaraan bermotor *comprehensive* untuk kategori jenis kendaraan Truk & Pickup di wilayah DKI Jakarta, Jawa Barat, dan Banten. Datanya terdiri dari 2.001 pengamatan besar klaim *partial loss*. Nilai minimum, nilai maksimum dan rata-ratanya masing-masing adalah Rp23.350, Rp348.500.000 dan Rp9.966.640. Histogram untuk data besar klaim *partial loss* di atas disajikan dalam Gambar 1.



Gambar 1. Histogram untuk data klaim *partial loss*

Taksiran nilai ambang batasnya adalah **65.575.000**. Dengan menggunakan uji kecocokan distribusi Kolmogorov-Smirnov melalui bantuan perangkat lunak EasyFit 5.0. diperoleh kecocokan distribusi log-logistik untuk data besar klaim *partial loss* yang ada di bawah ambang batas. Nilai taksiran parameter distribusi log-logistik standar berdasarkan perangkat lunak EasyFit 5.0 adalah $\hat{\gamma} = 1,6809$ dan $\hat{\eta} = 4.337.800$. Distribusi log-logistik dan taksiran parameternya menjadi distribusi dan taksiran awal bagi distribusi komposit untuk data yang berada di bawah ambang batas.

Hasil pengujian kecocokan distribusi menggunakan uji kecocokan Kolmogorov-Smirnov menunjukkan bahwa distribusi GPD cocok untuk data di atas ambang batas. Nilai taksiran parameter GPD standar berdasarkan perangkat lunak EasyFit 5.0 adalah $\hat{\alpha} = 0,2180$ dan $\hat{\beta} = 46.673.000$ akan menjadi nilai taksiran awal bagi distribusi komposit untuk data yang berada di atas ambang batas.

Nilai taksiran parameter dari distribusi log-logistik terpancung kanan (untuk data di bawah ambang batas) masing-masing adalah $\hat{\gamma} = 1,6871$ dan $\hat{\eta} = 4.337.811$. Sehingga taksiran fungsi densitas dari distribusi di bawah ambang batasnya adalah:

$$f_{\leq b|z \leq b}(z | z \leq 65.575.000) = \frac{4,3043 \cdot 10^{12}}{0,9899z^{2,6871}}; z \leq 65.575.000.$$

Nilai taksiran parameter dari distribusi GPD (untuk data di atas ambang batas) masing-masing adalah $\hat{\alpha} = 1,5401$ dan $\hat{\beta} = 46.673.020$. Dengan demikian taksiran fungsi densitas dari distribusi di atas ambang batas adalah

$$f_{> b|z > b}(z | z > 65.575.000) = \frac{1,0922 \cdot 10^{13}}{z^{2,5401} - 3,2881 \cdot 10^{19}}; z > 65.575.000.$$

Dengan demikian dapat dituliskan model komposit log-logistik-generalized Pareto sebagai berikut

$$f(z) = I(z \leq 65.575.000) \frac{4,3043 \cdot 10^{12}}{0,9899z^{2,6871}} + (1 - I(z \leq 65.575.000)) \frac{1,0922 \cdot 10^{13}}{z^{2,5401} - 3,2881 \cdot 10^{19}} 0,0101; \text{ untuk } z > 0,$$

dimana

$$I(z \leq 65.575.000) = \begin{cases} 1, & z \leq 65.575.000 \\ 0, & z > 65.575.000 \end{cases}.$$

9. KESIMPULAN

Dalam artikel ini telah diterapkan model komposit log-logistik-generalized Pareto untuk data besar klaim partial loss asuransi kendaraan bermotor di Indonesia untuk jenis kendaraan Truk & Pickup di wilayah DKI Jakarta, Jawa Barat, dan Banten dimana ambang batasnya dipilih menggunakan metode heuristik aturan akar kuadrat. Hasil perhitungan menunjukkan bahwa nilai ambang batasnya adalah 65.575.000 dan model komposit log-logistik-generalized Pareto cocok untuk data besar klaim di atas.

DAFTAR PUSTAKA

- Balkema, A., de Haan, L. (1974). Residual Life Time At Great Age. *Annals of Probability*, 2:782–804.
- Caeiro, F., Gomes, M. (2005). *Threshold Selection In Extreme Value Analysis. Extreme Value Modeling And Risk Analysis: Methods And Applications*. Wiley.
- Cebriaan, A. C., Denuit, M., Lambert, P. (2003). Generalized Pareto Fit To The Society Ofactuaries' Large Claims Database. *North American Actuarial Journal*, 7:146–154.
- DuMouchel, W. H. (1983). Estimating The Stable Indexain Order To Measure Tail Thickness: A Critique. *The Annals of Statistics*, 11(4):1019–1031
- Edwin, J., Purcell, & Varberg, D. (1987). *Kalkulus dan Geometri Analisis*. Edisi Kelima, Erlangga, Bandung.
- Evans, M., Hastings, N., & Peacock, B. (2000), *Statistical Distributions* (3rd ed.), John Wiley, New York.
- Ferreira, A., de Haan, L., Peng, L. (2003). On Optimising The Estimation Of High Quantiles of A Probability Distribution. *Statistics*, 37(5):401–434.
- Gomes, M. I., Oliveira, O. (2001) The Bootstrap Methodology In Statistics Of Extremes–Choice Of The Optimal Sample Fraction. *Extremes*, 4(4):331–358.
- Guillou, A., Hall, P. (2001). A Diagnostic For Selecting The Threshold In Extreme Value Analysis. *Journal of the Royal Statistical Society Series B*, 63(2):293–305.
- Hall, P.(1990). Using The Bootstrap To Estimate Mean Squared Errorand Select Smooth-Ing Parameter In Nonparametric Problems. *Journal of Multivariate Analysis*, 32(2):177–203.
- Ignatius, R. Y. S. (2018). *Pengantar Asuransi*. ACA Asuransi, Jakarta.
- Kitab Undang-Undang Hukum Dagang (KUHD), Bab IX Tentang Asuransi atau Pertanggunganan pada Umumnya, Pasal 246.
- Kleiber, C., Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. Wiley, New York.
- Klugman, S. A., Panjer, H. H., dan Wilmot, G. (2012). *Loss Models. From Data to Decisions*, Willey Interscience, New York.
- Loretan, M., Phillips, P. C. (1994). Testing The Covariance Stationarity Of Heavy-Tailed Timeseries: An Overview Of The Theory With Applications To Several Financial Datasets. *Journal of Empirical Finance*, 1(2):211–248.
- Nazmi, N. (2019). *Pemodelan Distribusi Binomial Negatif Poisson-Lindley Diboboti pada Data Frekuensi Klaim Asuransi Kendaraan Bermotor Di Indonesia*. Skripsi S1 Jurusan Statistika Universitas Islam Bandung.
- Otoritas Jasa Keuangan. (2017). *Surat Edaran Otoritas Jasa Keuangan Nomor 6/SEOJK.05/2017: Tentang Penetapan Tarif Premi atau Kontribusi pada Lini Usaha Asuransi Harta Benda dan Asuransi Kendaraan Bermotor tahun 2017*, OJK, Jakarta.
- Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, 3(1):119–131.
- Pigeon, M., & Denuit, M. (2011). Composite Lognormal–Pareto model with random threshold. *Scandinavian Actuarial Journal*, 2011(3):177–192.
- Sudjana. (1996). *Metode Statistika*. Tarsito, Bandung.
- Wang, Y., Haff, I. H. & Huseby, A. (2019). *Modelling Extreme Claims Via Composite Models And Threshold Selection Methods*. Department of Mathematics, University of Oslo.