STUDI SIMULASI PARAMETER DISTRIBUSI *GENERALIZED EXTREME VALUE* (GEV) DENGAN PENDEKATAN L-*MOMENTS* DAN MLE

Inayatus Sholichah*, Heri Kuswanto, Brodjol Sutijo

Abstrak

Iklim ekstrim seringkali memberikan dampak negatif bagi kehidupan. Sehingga mengetahui pola perubahannya dapat mengurangi resiko dampak yang ditimbulkan iklim ekstrim. Mencari nilai estimasi yang tepat dari data frekuensi iklim ekstrim bisa dijadikan acuan untuk mengetahui keadaan iklim yang akan datang. Motede analisis data ekstrim yang biasa digunakan kalangan hidrologi adalah Exterme Value Theory (EVT) dengan pendekatan Linier Moments (L-Moments) dan Maximum Likelihood Estimation (MLE). Pada penelitian ini dilakukan pengambilan sampel dengan Block Maxima mengikuti distribusi Generalized Extreme Value (GEV). Perbandingan performa metode pendekatan estimasi parameter dengan dilakukan studi simulasi. Hasil penelitian ini yaitu persamaan estimasi parameter dengan metode MLE membentuk fungsi yang tidak closed form sedangkan metode L-Moments membentuk fungsi yang closed form. Hasil studi simulasi untuk distribusi Gumbel, nilai bias estimator metode MLE dan L-Moments memiliki nilai yang hampir sama. Untuk data sampel besar dari distribusi Gumbel lebih sesuai menggunakan pendekatan estimasi dengan metode MLE karena nilai bias estimatornya lebih kecil dari nilai bias estimator L-Moments.

Kata-kata kunci: GEV,MLE,L-Moments, Simulasi.

Pendahuluan

Perubahan iklim ekstrim seringkali memberikan dampak negatif bagi kehidupan. Salah satu unsur utama iklim adalah curah hujan. Mengetahui perubahan iklim menjadi hal yang penting karena merupakan bagian dari informasi tren perubahan cuaca yang akan terjadi. Banyak akibat yang ditimbulkan dari perubahan iklim ekstrim, seperti menaikkan resiko terjadinya banjir, merusak area pangan, dll. Maka dari itu, perlu mencari nilai estimasi yang tepat dari data iklim series yang memiliki parameter distribusi betsifat independen dan identik (IID) pada suatu percobaan. Sehingga hasil yang didapatkan bisa dijadikan acuan untuk keadaan iklim yang akan datang.

Metode statistika yang dikembangkan berkaitan dengan analisis kejadian ekstrim adalah *Extreme Value Theory* (EVT) [1]. Pendekatan yang digunakan untuk mengidentifikasi pergerakan nilai ekstrim dalam EVT yaitu *Block Maxima* (BM) dari distribusi *Generalized Extreme Value* (GEV). Metode *Block Maxima* merupakan metode yang mengidentifikasi nilai ekstrim berdasarkan nilai tertinggi data observasi yang sudah dikelompokkan berdasarkan periode tertentu. Pendekatan estimasi curah hujan ekstrim biasanya menggunakan data series maksimum tahunan. Metode *Linier Moments* (L-*Moments*) banyak digunakan oleh kalangan hydrology untuk mencari kesesuaian distribusi *Extreme Value (*EV) untuk estimasi frekuensi banjir dan sudah dijadikan acuan oleh *Flood Estimation Handbook* (FEH) [2]. Namun, dalam perkembangan komputasi, muncul metode *Maximum Likelihood Estimate* (MLE) untuk mencari estimasi parameter dari suatu distribusi.

Sehingga muncul penelitian yang dilakukan oleh Collier [3] mengestimasi parameter dengan membandingkan dua teknik pendekatan, yaitu *Maksimum Likelihood Estimate* (MLE) dan *Linier Moments* (L-*Moments*). Untuk membandingkan MLE dan L-*Moment*s, data sintetis dibentuk dari hasil simulasi digunakan untuk menganalisis satu lokasi dan beberapa lokasi titik pengambilan data curah hujan. Simulasi yang digunakan adalah *Monte Carlo*, merupakan metode simulasi yang digunakan untuk membandingkan performa dari suatu teknik pendekatan parameter. Berdasarkan uraian tersebut, maka dalam paper ini melakukan studi simulasi data curah hujan ekstrim dengan pendekatan distribusi *Generalized Extreme Value* (GEV) dengan membandingkan dua teknik pendekatan estimasi parameter, yaitu *Maksimum Likelihood Estimate* (MLE) dan *Linier Moments* (L-*Moments*). Penelitian ini juga membandingkan teknik pendekatan estimasi parameter terhadap distribusi GEV dengan simulasi data dari macam-macam distribusi *Generalized Extreme Value* (GEV), yaitu distribusi Gumbel, Frechet, dan Weibull.

Teori

Metode yang digunakan adalah *Extreme Value Theory* (EVT) dengan pengambilan sampel menggunakan *Block Maxima* dari Distribusi *Generalized Extreme Value* (GEV).

ISBN: 978-602-19655-8-0

Berikut ini adalah *Comulative Distribution Function* (CDF) dari *Generalized Extreme Value* (GEV).

$$F_*(x) = \begin{cases} \exp\left(-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{\frac{-1}{\xi}}\right) \\ \exp\left(-exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right) \end{cases} \quad (1)$$

dimana μ = parameter lokasi , σ = parameter skala dan ξ = parameter bentuk *tail* index. Berdasarkan nilai parameter bentuk *tail* index (ξ), distribusi GEV dibedakan menjadi 3 tipe yaitu Distribusi Gumbel ($\xi = 0$), Distribusi Frechet ($\xi > 0$) dan Distribusi Weibull ($\xi < 0$).

Penaksir parameter metode *Generalized Extreme Value* (GEV) dapat ditaksir dengan menggunakan metode *Maximum Likelihood Estimation* (MLE). MLE merupakan salah satu metode estimasi parameter dari suatu distribusi dengan memaksimumkan fungsi likelihood. Selanjutnya dari persamaan ln likelihood yang diperoleh kemudian diturunkan terhadap parameter yang akan ditaksir dan disamakan dengan nol.

Adapun untuk metode L- *Moments*, yaitu metode pendekatan parameter dengan kombinasi linier dari nilai ekspektasi orde statistik variabel dan hasil estimasi dari sampel yang menggunakan rata-rata pembobot dari orde statistik. Diketahui X_1, X_2, ..., X_r adalah sampel acak berukuran r, dan $X_{1:r} \leq X_{2:r} \leq \cdots \leq X_{r:r}$ berupa orde statistik. Maka persamaan L-*Moments* (r^{th}) adalah sebagai berikut.

$$\lambda_r = \frac{1}{r}\sum_{n=0}^{r-1}(-1)^n\binom{r-1}{n}E(X_{r-n:r}) \quad (2)$$

nilai $E(X_{r-n:r})$ bergantung dengan jenis distribusinya. Dimana nilai $E(X_{i-r})$ seperti pada persamaan 3.

$$E(X_{i:r}) = \frac{r!}{(i-1)!(r-i)!}\int_0^1 x(F)F^{i-1}(1-F)^{r-1}dF \quad (3)$$

dimana nilai *X(F)* merupakan fungsi quantile dari suatu distribusi. Untuk distribusi GEV memiliki fungsi quantile sebagai berikut.

$$X(F) = \begin{cases} \hat{\mu} + \hat{\sigma}\left[1-(-\log F)^{\hat{\xi}}\right]/\hat{\xi} & jika\ \xi \neq 0 \\ \hat{\mu} - \hat{\sigma}(-\log F) & jika\ \xi = 0 \end{cases} \quad (4)$$

Sehingga L-*Moments* yang terbentuk secara berurutan $\lambda_1, \lambda_2, \lambda_3$ dan λ_4 adalah ukuran populasi dari lokasi, skala, *skewness*, dan *kurtosis* dari suatu distribusi. Selanjutnya mencari nilai estimasi parameter distribusi dengan mensubtitusi dan mengeliminasi bentuk persamaan fungsi $\lambda_1, \lambda_2, \lambda_3$ dan λ_4.

Monte Carlo merupakan metode simulasi yang digunakan untuk membandingkan performa dari suatu teknik pendekatan parameter. Untuk distribusi GEV dengan simulasi *Monte Carlo,* nilai parameter skala (σ) dan lokasi (μ) yang digunakan adalah antara 0 dan 1 [4] dengan bentuk parameter bentuk *tail* (ξ) antara (-0.4 , 0.4) [5]. Perbandingan performa teknik pendekatan parameter dilihat dari *Root Mean Square Error* (RMSE) untuk ukuran sampel (*N*) dan bentuk parameter bentuk *tail* (ξ) yang berbeda. Persamaan untuk RMSE untuk kriteria pemilihan metode yang sesuai dengan persamaannya sebagai berikut.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2} \quad (5)$$

dimana x_i = Nilai aktual, $\hat{x}_i$ = Nilai dugaan dan N = Banyaknya nilai yang diduga. Metode yang terpilih adalah metode dengan nilai bias estimator yang terkecil.

Hasil dan diskusi

Untuk kajian teori MLE distribusi GEV dengan $\xi = 0$ dengan fungsi likelihood-nya adalah

$$\begin{aligned} L(\mu,\sigma) &= \prod_{i=1}^{n} f(x_i) \\ &= \prod_{i=1}^{n}\frac{1}{\sigma}\exp\left[-\left(\frac{x_i-\mu}{\sigma}\right)\right]\exp\left(-\exp\left[-\left(\frac{x_i-\mu}{\sigma}\right)\right]\right) \\ &= \left(\frac{1}{\sigma}\right)^n \exp\left[-\prod_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)\right]\exp\left(-\prod_{i=1}^{n}\exp\left[-\left(\frac{x_i-\mu}{\sigma}\right)\right]\right) \end{aligned}$$

sehingga fungsi ln likelihood nya menjadi

$$\ln L(\mu,\sigma) = -n\ln(\sigma) - \sum_{i=1}^{n}\left(\frac{x-\mu}{\sigma}\right) - \sum_{i=1}^{n}\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]$$

Selanjutnya dari persamaan ln likelihood yang diperoleh kemudian diturunkan terhadap parameter yang akan ditaksir dan disamakan dengan nol. Berdasarkan persamaan yang terbentuk, diperoleh persamaan seperti berikut

$$\frac{\partial \ln L(\mu,\sigma)}{\partial\mu} = \frac{n}{\sigma} - \frac{1}{\sigma}\sum_{i=1}^{n}\exp\left(-\frac{x_i-\mu}{\sigma}\right) = 0$$

$$\frac{\partial \ln L(\mu,\sigma)}{\partial\sigma} = \frac{n}{\sigma} + \sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma^2}\right) + \sum_{i=1}^{n}\left(-\frac{x_i-\mu}{\sigma^2}\right)\exp\left(-\frac{x_i-\mu}{\sigma}\right) = 0$$

Dari persamaan yang terbentuk, diperoleh persamaan yang tidak *closed form* maka dibutuhkan analisis numerik lebih lanjut dengan cara iterasi untuk memaksimumkan fungsi ln likelihood. Salah satunya dengan analisis numerik misalnya metode *Newton Raphson*.

ISBN: 978-602-19655-8-0

Untuk kajian L-*Moments* distribusi GEV dengan $\xi = 0$, didapatkan persamaan berikut.

$$\lambda_1 = E(X_{1:1})$$
$$\lambda_2 = \frac{1}{2}E(X_{2:2} - X_{1:2})$$

Dengan memasukan fungsi quantile dari distribusi GEV dengan $\xi = 0$, persamaan L-*Moments* menjadi sebagai berikut

$$\lambda_1 = \hat{\mu} + \gamma\,\hat{\sigma}$$
$$\lambda_2 = \hat{\sigma}\log 2$$

dengan γ merupakan bilangan Euler yaitu 0,5772. Untuk L-*skewness* dan L-*kurtosis* dari distribusi GEV dengan $\xi = 0$ sebagai berikut

$$\tau_3 = \frac{\hat{\sigma}\log(\frac{9}{8})}{\hat{\sigma}\log 2} = 0.1699$$
$$\tau_4 = \frac{\hat{\sigma}(16\log 2 - 10\log 3)}{\hat{\sigma}\log 2} = 0.1504$$

selanjutnya nilai $\hat{\mu}$ dan $\hat{\sigma}$ dapat diestimasi dari fungsi λ_1 dan λ_2, sehingga fungsi estimasi parameternya adalah

$$\hat{\mu} = \lambda_1 - \gamma\,\hat{\sigma}$$
$$\hat{\sigma} = \lambda_2/\log 2$$

Dari persamaan estimasi parameter yang terbentuk, diperoleh persamaan yang *closed form* untuk mengestimasi parameter distribusi GEV dengan pendekatan L-*Moments* untuk distribusi GEV $(\xi = 0)$.

Berdasarkan persamaan estimasi metode MLE dan L-*Moments* untuk distribusi GEV, dilakukan simulasi data pembangkitan distribusi GEV $(\xi = 0)$ yaitu Distribusi Gumbel. Simulasi dilakukan untuk membangkitkan data distribusi GEV dengan ukuran sampel 50, 100 dan 200 dengan 1000 perulangan.

Dari hasil simulasi dengan tipe distribusi Gumbel, didapatkan nilai bias estimator yang secara rinci pada Tabel 1.

Tabel 1. Bias Estimator Simulasi Gumbel

Simulasi	N	MLE		L-*Moments*	
		μ	σ	μ	σ
μ = 0.4, σ = 0.4	50	0,0515	0,0382	0,0472	0,0417
	100	0,0358	0,0266	0,0335	0,0283
	200	0,0247	0,0186	0,0236	0,0211
μ = 0.6, σ = 0.6	50	0,0780	0,0573	0,0699	0,0613
	100	0,0537	0,0389	0,0485	0,0446
	200	0,0375	0,0280	0,0368	0,0304
μ = 0.8, σ = 0.8	50	0,1048	0,0760	0,0951	0,0797
	100	0,0752	0,0528	0,0675	0,0585
	200	0,0501	0,0362	0,0493	0,0412

Terlihat dari hasil bias estimator simulasi Gumbel, metode MLE dan L-*Moments* memiliki nilai bias estimator yang mendekati satu sama lain. Tetapi semakin banyak sampel data, nilai bias estimator MLE lebih kecil daripada L-*Moments*. Sehingga untuk data sampel besar lebih sesuai menggunakan metode MLE.

Kesimpulan

Dari pembahasan mengenai kajian teori MLE dan L-*Moments* didapatkan estimasi parameter dengan metode MLE membentuk fungsi yang tidak *closed form* sedangkan metode L-*Moments* membentuk fungsi yang *closed form*. Untuk studi simulasi distribusi GEV didapatkan dari hasil studi simulasi untuk distribusi Gumbel, nilai bias estimator metode MLE dan L-*Moments* memiliki nilai yang hampir sama. Untuk data sampel besar dari distibusi Gumbel lebih sesuai menggunakan metode MLE karena nilai bias estimatornya lebih kecil dari nilai bias estimator L-*Moments*.

Ucapan terima kasih

Penulis mengucapkan terima kasih atas bimbingan dari dosen-dosen Statistika ITS dalam dukungannya untuk keikutsertaan dalam kegiatan ilmiah ini.

Referensi

[1] Coles, S. (2001). An Introduction to Statistical Modelling of Extreme Values. London: Spinger-Verlag.
[2] Hosking, J. R. M. (1990). L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics Journal of the Royal Statistical Society, Series B, 52, p105–124.
[3] Collier, A. J (2010).Extreme Value Analysis of Non-Stationary Processes – a Study of Extreme Rainfall Under Changing Climate. Tesis: Civil Engineering, University of Newcastle.
[4] Shabri, A. & Jemain, A.A. (2007). LQ-Moments for Statistical Analysis of Extreme Events. Journal of Modern Applied Statistical Methods, Vol. 6, No. 1, 228-238.
[5] Wang, Q.J. (1990). Estimation of the GEV Distribution from Censored Samples By Method of Partial Probability Weighted Moments. Journal of Hydrology, 120 103-114.

Inayatus Sholichah*
Mahasiswa Pascasarjana Statiska
Institut Teknologi Sepuluh Nopember
Inayah.sholichah@gmail.com

Heri Kuswanto
Dosen Statistika
Institut Teknologi Sepuluh Nopember
kuswanto.its@gmail.com

Brodjol Sutijo
Dosen Statistika
Institut Teknologi Sepuluh Nopember
brodjol.su@ gmail.com

*Corresponding author