

**ANALISIS KEJADIAN CURAH HUJAN EKSTREM  
DI KOTA MAKASSAR MENGGUNAKAN DISTRIBUSI  
*GENERALIZED EXTREME VALUE* DAN *GENERALIZED PARETO***

*ANALYSIS OF EXTREME RAINFALL EVENT IN MAKASSAR  
CITY USING DISTRIBUTIONS OF GENERALIZED EXTREME  
VALUE AND GENERALIZED PARETO*

**NURUL AZIZAH MUZAKIR**



**PROGRAM STUDI MAGISTER STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR  
2023**

**ANALISIS KEJADIAN CURAH HUJAN EKSTREM  
DI KOTA MAKASSAR MENGGUNAKAN DISTRIBUSI  
*GENERALIZED EXTREME VALUE* DAN *GENERALIZED PARETO***

Tesis

sebagai salah satu syarat untuk mencapai gelar magister

Program Studi Magister Statistika

Disusun dan diajukan oleh

Nurul Azizah Muzakir

H062211001

Kepada

**PROGRAM STUDI MAGISTER STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR  
2023**

**TESIS**

**ANALISIS KEJADIAN CURAH HUJAN EKSTREM DI KOTA  
MAKASSAR MENGGUNAKAN DISTRIBUSI *GENERALIZED EXTREME  
VALUE* DAN *GENERALIZED PARETO***

**NURUL AZIZAH MUZAKIR**

H062211001

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka  
Penyelesaian Program Studi Magister Statistika Fakultas Matematika dan Ilmu  
Pengetahuan Alam Universitas Hasanuddin  
pada tanggal 7 Agustus 2023  
dan dinyatakan telah memenuhi syarat kelulusan

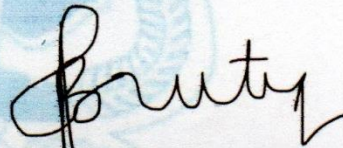
Menyetujui,

Pembimbing Utama,

Pembimbing Pendamping



**Dr. Nurtiti Sunusi, S.Si., M.Si.**  
NIP. 19720117 199703 2 002



**Sri Astuti Thamrin, S.Si., M.Stat., Ph.D.**  
NIP. 19740713 199903 2 001

Ketua Program Studi  
Magister Statistika,

Dekan Fakultas Matematika dan  
Ilmu Pengetahuan Alam  
Universitas Hasanuddin



**Dr. Dr. Georgina M. Tinungki, M.Si.**  
NIP. 19620926 198702 2 001



**Dr. Eng. Amiruddin, S.Si., M.Si.**  
NIP. 19720515 199702 1 002



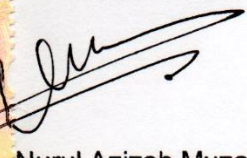
## PERNYATAAN KEASLIAN TESIS DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, tesis berjudul "Analisis Kejadian Curah Hujan Ekstrem di Kota Makassar Menggunakan Distribusi *Generalized Extreme Value* dan *Generalized Pareto*" adalah benar karya saya dengan arahan dari komisi pembimbing (Dr. Nurtiti Sunusi, S.Si., M.Si. dan Sri Astuti Thamrin, S.Si., M.Stat., Ph.D.). Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apapun kepada perguruan tinggi manapun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka tesis ini. Sebagian dari isi tesis ini telah dipublikasikan di Jurnal (*International Journal of Research Publications*, Volume 129, Halaman 259–275, dan DOI: <https://doi.org/10.47119/IJRP1001291720235256>) sebagai artikel dengan judul "Modeling Extreme Rainfall Events in the Tropical Monsoon Region: A Generalized Pareto Distribution Approach".

Dengan ini saya melimpahkan hak cipta dari karya tulis saya berupa tesis ini kepada Universitas Hasanuddin.

Makassar, 7 Agustus 2023



  
Nurul Azizah Muzakir  
NIM. H062211001

## UCAPAN TERIMA KASIH

Puji dan syukur ke hadirat Allah karena atas pertolongan-Nyalah sehingga tesis ini dapat diselesaikan sebagai salah satu syarat untuk menyelesaikan studi pada program Magister di Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin. Tidak lupa pula penulis mengirimkan shalawat serta salam kepada bimbingan kita, Rasulullah Muhammad ﷺ yang telah memberikan banyak pedoman kehidupan bagi manusia.

Dalam menyelesaikan tesis ini, penulis tentunya tidak luput dari berbagai kekeliruan dan hambatan. Namun, pada akhirnya tesis ini dapat diselesaikan berkat adanya kontribusi dari berbagai pihak yang selalu memberikan motivasi, bantuan, dan bimbingan bagi penulis. Oleh karena itu, disamping rasa syukur ke hadirat Allah, penulis juga mengucapkan terima kasih dan rasa hormat kepada kedua orang tua penulis, Ayahanda Abdul Muzakir dan Ibunda Dra. Sapinah, M.Si. yang telah menjadi orang terhebat dalam membesarkan dan mendidik penulis. Hanya doalah balasan terbaik yang dapat penulis sampaikan. Selanjutnya, ucapan terima kasih juga penulis sampaikan kepada:

1. Bapak Prof. Dr. Ir. Jamaluddin Jompa, M.Sc. selaku rektor Universitas Hasanudin.
2. Bapak Dr. Eng. Amiruddin, S.Si., M.Si. selaku dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin.
3. Ibu Dr. Anna Islamiyati, S.Si., M.Si. selaku ketua Departemen Statistika sekaligus dosen pada program studi magister Statistika.
4. Ibu Dr. Dr. Georgina Maria Tinungki, M.Si. selaku ketua program studi Magister Statistika sekaligus selaku tim penguji tesis penulis. Terima kasih atas perhatian yang diberikan kepada kami mahasiswa pascasarjana, juga atas saran untuk perbaikan tesis penulis.
5. Ibu Dr. Nurtiti Sunusi, S.Si., M.Si. selaku pembimbing utama penulis. Terima kasih atas segala kebaikan dan perhatian yang ditujukan kepada penulis, juga kesediaan dan kesabaran meluangkan waktu, tenaga dan pikiran untuk membimbing penulis sehingga tesis ini dapat diselesaikan.
6. Ibu Sri Astuti Thamrin, S.Si., M.Stat., Ph.D. selaku pembimbing pertama penulis. Terima kasih atas kesediaan dan kesabaran meluangkan waktu untuk membimbing penulis, juga arahan untuk perbaikan tesis penulis.

7. Ibu Dr. Erna Tri Herdiani, S.Si., M.Si. selaku tim penguji tesis penulis. Terima kasih atas saran dan arahan yang diberikan kepada penulis dalam menyelesaikan tesis ini.
8. Bapak Dr. Nirwan, M.Si. selaku tim penguji tesis penulis. Terima kasih atas saran dan arahan yang diberikan kepada penulis dalam menyelesaikan tesis ini.
9. Staff di Departemen Statistika FMIPA Unhas. Terima kasih atas bantuan terkait dengan persuratan yang diberikan kepada kami mahasiswa pascasarjana.
10. Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) yang menyediakan data secara *online* sehingga memudahkan penulis untuk menyelesaikan tesis ini.
11. Teman-teman mahasiswa program Magister Statistika angkatan empat: Kak Zarkawi, Kak Widia, Kak Muammar, dan Kak Nunu. Sebuah kehormatan bisa diberikan kesempatan untuk belajar bersama.
12. Teman-teman mahasiswa program Magister Statistika lainnya: Amel, Arya, Mila, Samsir, Kak Misbah, Kak Andis, Kak Gio, Kak Imma, Kak Tisen, Kak Awi, Kak Gery, Kak Nunu, Kak Mumun, Kak Nurul, Kak Mita, Kak Alfi, Kak Ardi, dan lainnya yang penulis tidak bisa sebutkan satu persatu.

Kepada seluruh pihak yang tidak sempat disebutkan atas keikutsertaan membantu menyelesaikan tesis ini, penulis ucapkan terima kasih. Semoga bantuan yang diberikan kepada penulis mendapatkan balasan dari Allah berupa kebaikan. Semoga tesis ini dapat bermanfaat bagi pembaca.

Penulis,

Nurul Azizah Muzakir

## ABSTRAK

NURUL AZIZAH MUZAKIR. **Analisis Kejadian Curah Hujan Ekstrem di Kota Makassar Menggunakan Distribusi Generalized Extreme Value dan Generalized Pareto** (dibimbing oleh Nurtiti Sunusi dan Sri Astuti Thamrin).

Kejadian curah hujan ekstrem adalah salah satu fenomena cuaca yang merusak karena sering diikuti oleh banjir. Informasi mengenai *return level*, yaitu curah hujan yang diharapkan terjadi dalam suatu periode tertentu, sangat penting untuk perencanaan mitigasi bencana banjir yang efektif. Salah satu pendekatan statistika yang dapat digunakan untuk menganalisis fenomena alam yang ekstrem adalah distribusi *Generalized Extreme Value* (GEV) dan *Generalized Pareto* (GP). Penelitian ini bertujuan untuk mengkaji estimasi parameter distribusi GEV dan GP, kemudian mengaplikasikannya untuk mengestimasi *return level* curah hujan ekstrem. Untuk mengestimasi parameter digunakan metode *Maximum Likelihood Estimation* (MLE). Data curah hujan harian di kota Makassar dari stasiun hujan Hasanuddin periode 1980–2022 digunakan dalam penelitian ini. Metode *Block Maxima* (BM) dan *Peaks Over Threshold* (POT) digunakan untuk mengidentifikasi kejadian curah hujan ekstrem. Berdasarkan hasil kajian, estimasi parameter distribusi GEV dan GP sulit diselesaikan secara analitik, sehingga dilanjutkan menggunakan metode Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton. Selanjutnya, estimasi *return level* curah hujan pada periode ulang 2, 3, 4, dan 5 tahun dilakukan menggunakan distribusi GEV dan GP. Estimasi ini digunakan untuk memprediksi kejadian curah hujan ekstrem pada periode mendatang. Hasil estimasi menunjukkan bahwa curah hujan maksimum di kota Makassar pada tahun 2023 dan tahun-tahun berikutnya diperkirakan sama dengan atau melebihi 144.68 mm/hari dengan peluang kejadian sebesar 50%, 164.6 mm/hari dengan peluang kejadian sebesar 33%, 178.17 mm/hari dengan peluang kejadian sebesar 25%, 188.64 mm/hari dengan peluang kejadian sebesar 20%, 194.1 mm/hari dengan peluang kejadian sebesar 6.8%, dan 203.13 mm/hari dengan peluang kejadian sebesar 5.5%.

Kata Kunci: *generalized extreme value*, *generalized pareto*, *peaks over threshold*, *block maxima*, *return level*, curah hujan

## ABSTRACT

**NURUL AZIZAH MUZAKIR. Analysis of Extreme Rainfall Event in Makassar City Using Distributions of Generalized Extreme Value and Generalized Pareto** (supervised by Nurtiti Sunusi and Sri Astuti Thamrin).

The occurrence of extreme rainfall events is a damaging weather phenomenon often accompanied by floods. Information about return levels, which represent the expected rainfall occurring within a specific period, is crucial for effective flood disaster mitigation planning. One of the statistical approaches that can be used to analyze extreme natural phenomena is the Generalized Extreme Value (GEV) and Generalized Pareto (GP) distributions. This research aims to examine the parameter estimation of the GEV and GP distributions and apply them to estimate the return levels of extreme rainfall. The Maximum Likelihood Estimation (MLE) method was used to estimate the parameters. Daily rainfall data from the Hasanuddin rain gauge station in Makassar covering the period from 1980 to 2022 was utilized for this study. The Block Maxima (BM) and Peaks Over Threshold (POT) methods were employed to identify extreme rainfall events. The results reveal that the parameter estimation of the GEV and GP distributions cannot be analytically solved. Hence, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton method was applied. Furthermore, the estimated return levels for 2, 3, 4, and 5-year return periods were determined using the GEV and GP distributions. These estimates were then utilized to predict future extreme rainfall events. The findings indicate that the maximum rainfall in Makassar in 2023 and subsequent years is projected to equal or exceed 144.68 mm/day with a 50% probability, 164.6 mm/day with a 33% probability, 178.17 mm/day with a 25% probability, 188.64 mm/day with a 20% probability, 194.1 mm/day with a 6.8% probability, and 203.13 mm/day with a 5.5% probability.

**Keywords:** generalized extreme value, generalized pareto, peaks over threshold, block maxima, return level, rainfall



## DAFTAR ISI

|                                                                  |      |
|------------------------------------------------------------------|------|
| UCAPAN TERIMA KASIH.....                                         | v    |
| ABSTRAK.....                                                     | vii  |
| ABSTRACT .....                                                   | viii |
| DAFTAR ISI.....                                                  | ix   |
| DAFTAR TABEL .....                                               | xi   |
| DAFTAR GAMBAR .....                                              | xii  |
| DAFTAR LAMPIRAN .....                                            | xiii |
| BAB I PENDAHULUAN.....                                           | 1    |
| 1.1 Latar Belakang .....                                         | 1    |
| 1.2 Rumusan Masalah .....                                        | 5    |
| 1.3 Tujuan Penelitian.....                                       | 5    |
| 1.4 Batasan Masalah.....                                         | 5    |
| 1.5 Manfaat Penelitian.....                                      | 6    |
| BAB II TINJAUAN PUSTAKA .....                                    | 7    |
| 2.1 <i>Extreme Value Theory</i> .....                            | 7    |
| 2.2 Metode <i>Block Maxima</i> .....                             | 8    |
| 2.3 <i>Generalized Extreme Value</i> .....                       | 9    |
| 2.3.1 Distribusi <i>extreme value</i> .....                      | 9    |
| 2.3.2 Distribusi <i>generalized extreme value</i> .....          | 10   |
| 2.4 Metode <i>Peaks Over Threshold</i> .....                     | 12   |
| 2.5 Distribusi <i>Generalized Pareto</i> .....                   | 12   |
| 2.6 Pemilihan <i>Threshold</i> .....                             | 13   |
| 2.6.1 <i>Mean residual life plot</i> .....                       | 14   |
| 2.6.2 <i>Parameter stability plot</i> .....                      | 15   |
| 2.7 Estimasi Parameter .....                                     | 15   |
| 2.8 Optimasi Numerik.....                                        | 17   |
| 2.8.1 Optimasi tanpa kendala.....                                | 17   |
| 2.8.2 Metode Newton.....                                         | 18   |
| 2.8.3 Metode Broyden-Fletcher-Goldfarb-Shanno quasi-Newton ..... | 19   |
| 2.9 Uji Kesesuaian Distribusi.....                               | 20   |
| 2.10 Uji Stasioneritas.....                                      | 21   |
| 2.11 <i>Return Level</i> .....                                   | 23   |
| 2.12 Curah Hujan .....                                           | 24   |
| 2.13 Penelitian Terdahulu .....                                  | 26   |
| 2.14 Kerangka Konseptual .....                                   | 27   |
| BAB III METODE PENELITIAN .....                                  | 29   |

|                                   |                                                                                           |    |
|-----------------------------------|-------------------------------------------------------------------------------------------|----|
| 3.1                               | Data dan Sumber Data.....                                                                 | 29 |
| 3.2                               | Variabel Penelitian .....                                                                 | 29 |
| 3.3                               | Tahapan Penelitian .....                                                                  | 29 |
| BAB IV HASIL DAN PEMBAHASAN ..... |                                                                                           | 33 |
| 4.1                               | Estimasi Parameter Distribusi <i>Generalized Extreme Value</i> .....                      | 33 |
| 4.2                               | Estimasi Parameter Distribusi <i>Generalized Pareto</i> .....                             | 43 |
| 4.3                               | Identifikasi Kejadian Curah Hujan Ekstrem di Kota Makassar .....                          | 47 |
| 4.3.1                             | Karakteristik data curah hujan harian .....                                               | 47 |
| 4.3.2                             | Metode <i>block maxima</i> .....                                                          | 52 |
| 4.3.3                             | Metode <i>peaks over threshold</i> .....                                                  | 53 |
| 4.3.4                             | Pengujian hipotesis .....                                                                 | 56 |
| 4.4                               | Estimasi <i>Return Level</i> Curah Hujan Ekstrem di Kota Makassar .....                   | 59 |
| 4.4.1                             | Estimasi parameter distribusi GEV dan GP .....                                            | 59 |
| 4.4.2                             | <i>Return level</i> menggunakan distribusi GEV .....                                      | 60 |
| 4.4.3                             | <i>Return level</i> menggunakan distribusi GP.....                                        | 61 |
| 4.4.4                             | Perbandingan estimasi <i>return level</i> distribusi GEV dan GP dengan data historis..... | 62 |
| 4.5                               | Prediksi Curah Hujan Ekstrem di Kota Makassar Menggunakan Distribusi GEV dan GP.....      | 66 |
| BAB V KESIMPULAN DAN SARAN.....   |                                                                                           | 67 |
| 5.1                               | Kesimpulan.....                                                                           | 67 |
| 5.2                               | Saran .....                                                                               | 68 |
| DAFTAR PUSTAKA.....               |                                                                                           | 69 |
| LAMPIRAN.....                     |                                                                                           | 73 |

## DAFTAR TABEL

|                                                                                                   |    |
|---------------------------------------------------------------------------------------------------|----|
| Tabel 1. Struktur data .....                                                                      | 29 |
| Tabel 2. Ringkasan statistik curah hujan bulanan di Kota Makassar periode 1980–2022.....          | 49 |
| Tabel 3. Hasil uji Kolmogorov-Smirnov pada data curah hujan ekstrem di kota Makassar .....        | 57 |
| Tabel 4. Hasil uji Mann-Kendall pada data curah hujan ekstrem di kota Makassar .....              | 59 |
| Tabel 5. Estimasi parameter distribusi GEV dan GP menggunakan metode MLE .....                    | 60 |
| Tabel 6. Return level curah hujan ekstrem di kota Makassar menggunakan model distribusi GEV ..... | 61 |
| Tabel 7. Return level curah hujan ekstrem di kota Makassar menggunakan model distribusi GP .....  | 62 |
| Tabel 8. Banyaknya kejadian curah hujan berdasarkan estimasi return level 2 tahun.....            | 63 |
| Tabel 9. Banyaknya kejadian curah hujan berdasarkan estimasi return level 3 tahun.....            | 64 |
| Tabel 10. Banyaknya kejadian curah hujan berdasarkan estimasi return level 4 tahun.....           | 65 |
| Tabel 11. Banyaknya kejadian curah hujan berdasarkan estimasi return level 5 tahun.....           | 65 |
| Tabel 12. Peluang kejadian curah hujan ekstrem di kota Makassar periode 2023 dan seterusnya ..... | 66 |

## DAFTAR GAMBAR

|                                                                                                                                    |    |
|------------------------------------------------------------------------------------------------------------------------------------|----|
| Gambar 1. Ilustrasi metode block maxima .....                                                                                      | 9  |
| Gambar 2. Kurva PDF distribusi Gumbel, Frechet, dan Weibull .....                                                                  | 11 |
| Gambar 3. Ilustrasi metode POT (sumber: Susan (2015)) .....                                                                        | 12 |
| Gambar 4. Kurva PDF distribusi GP .....                                                                                            | 13 |
| Gambar 5. Bagan kerangka konseptual .....                                                                                          | 28 |
| Gambar 6. Plot data curah hujan harian di Makassar periode 1980 – 2022 .....                                                       | 48 |
| Gambar 7. Plot rata-rata curah hujan setiap bulan di kota Makassar periode 1980–2022 .....                                         | 49 |
| Gambar 8. Persentase hari hujan setiap bulan di kota Makassar periode 1980–2022 .....                                              | 50 |
| Gambar 9. Sebaran data curah hujan harian di Kota Makassar periode 1980–2022 .....                                                 | 51 |
| Gambar 10. Box plot data curah hujan harian setiap bulan di kota Makassar periode 1980-2022 .....                                  | 51 |
| Gambar 11. Plot curah hujan ekstrem periode 1980–2022 di kota Makassar berdasarkan metode BM .....                                 | 52 |
| Gambar 12. Frekuensi kejadian curah hujan ekstrem di kota Makassar periode 1980 – 2022 berdasarkan metode BM .....                 | 53 |
| Gambar 13. Mean residual life plot .....                                                                                           | 54 |
| Gambar 14. Plot curah hujan ekstrem periode 1980 – 2022 di kota Makassar berdasarkan metode POT dengan threshold 100 mm/hari ..... | 55 |
| Gambar 15. Frekuensi kejadian curah hujan ekstrem di kota Makassar periode 1980 – 2022 berdasarkan metode POT .....                | 56 |
| Gambar 16. Plot kurva density distribusi data ekstrem berdasarkan metode BM vs distribusi GEV .....                                | 58 |
| Gambar 17. Plot kurva density distribusi data ekstrem berdasarkan metode POT vs distribusi GP .....                                | 58 |

## DAFTAR LAMPIRAN

|                                                                                                                                         |    |
|-----------------------------------------------------------------------------------------------------------------------------------------|----|
| Lampiran 1. Kejadian curah hujan ekstrem periode 1980 – 2022 di kota Makassar berdasarkan metode BM .....                               | 73 |
| Lampiran 2. Kejadian curah hujan ekstrem periode 1980 – 2022 di kota Makassar berdasarkan metode POT dengan threshold 100 mm/hari ..... | 74 |
| Lampiran 3. Tabel nilai kritis uji Kolmogorov-Smirnov .....                                                                             | 76 |
| Lampiran 4. Uji Kolmogorov-Smirnov pada data ekstrem berdasarkan metode BM .....                                                        | 77 |
| Lampiran 5. Uji Kolmogorov-Smirnov pada data ekstrem berdasarkan metode POT .....                                                       | 78 |
| Lampiran 6. Tabel distribusi normal baku .....                                                                                          | 79 |
| Lampiran 7. Kode program untuk estimasi parameter .....                                                                                 | 80 |

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Pemanasan global dan perubahan iklim merupakan isu yang mempengaruhi lingkungan tropis yang saat ini banyak dibahas oleh para peneliti. Perubahan iklim merupakan fenomena alam yang tidak dapat dihindari dan bersifat terus-menerus (Loo et al., 2015). Indonesia sebagai negara kepulauan dan memiliki iklim tropis merupakan salah satu negara yang rentan terhadap dampak negatif perubahan iklim. Akibat perubahan iklim global, wilayah Indonesia diproyeksikan akan mengalami perubahan pola dan intensitas curah hujan (Sunusi et al., 2017). Curah hujan merupakan variabel penting dalam bidang hidrologi. Curah hujan adalah ketinggian air hujan yang terkumpul dalam penakar hujan pada tempat yang datar, tidak menyerap, tidak meresap, dan tidak mengalir (BMKG, 2017).

Kejadian curah hujan ekstrem adalah salah satu fenomena cuaca yang paling merusak karena sering diikuti oleh banjir (Supari et al., 2012; Trenberth, 2011). Banjir disebabkan oleh hujan lebat yang melebihi daya dukung sistem drainase air (Musliadi et al., 2021). Banjir berdampak luas terhadap berbagai sektor kehidupan masyarakat, mulai dari sektor pertanian, infrastruktur, kesehatan, dan transportasi. Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) sebagai pemegang otoritas informasi meteorologi dan klimatologi di Indonesia mengategorikan curah hujan 50–100 mm/ hari sebagai hujan lebat, curah hujan 100–150 mm/hari sebagai hujan sangat lebat, dan curah hujan di atas 150 mm/hari sebagai hujan ekstrem (BMKG, 2022). Curah hujan ekstrem merupakan salah satu faktor yang berpotensi menimbulkan bencana alam di wilayah tropis (Fadholi, 2020).

Sulawesi Selatan merupakan salah satu provinsi di Indonesia dengan ibu kota Makassar. Makassar adalah daerah tropis dengan iklim muson tropis. Hal ini ditunjukkan dengan kontrasnya musim hujan dan musim kemarau dalam satu tahun. Sebagai kota yang sebagian besar wilayahnya merupakan daerah dataran rendah, Makassar memiliki potensi besar untuk mengalami banjir. Puncak banjir biasanya terjadi pada bulan Januari hingga Februari (Musliadi et al., 2021). Oleh



karena itu, penelitian tentang estimasi curah hujan ekstrem pada musim hujan menjadi sangat penting untuk mengantisipasi dampak banjir.

Menurut Coles (2001), metode statistika yang dikembangkan berkaitan dengan kejadian ekstrem adalah *Extreme Value Theory* (EVT) atau teori nilai ekstrem. *Extreme Value Theory* berkembang dengan pesat sejak diperkenalkan oleh Fisher dan Tippett pada tahun 1928 dan diformalkan oleh Gnedenko pada tahun 1943. Metode ini dapat melihat karakteristik nilai ekstrem karena berfokus pada perilaku ekor (*tail*) distribusi dalam menentukan probabilitas nilai ekstremnya (Fisher & Tippett, 1928; Gnedenko, 1943). *Extreme Value Theory* dapat memprediksi terjadinya kejadian ekstrem pada data berekor gemuk (*heavy tail*), yaitu ekor distribusi turun secara lambat apabila dibandingkan dengan distribusi normal. Akibatnya, peluang terjadinya nilai ekstrem akan lebih besar daripada pemodelan dengan distribusi normal (Hartono & Sutikno, 2020). *Extreme Value Theory* saat ini telah diterapkan di berbagai disiplin ilmu seperti hidrologi, *engineering*, klimatologi, keuangan dan finansial, maupun manajemen risiko. Terdapat dua metode yang digunakan untuk mengidentifikasi nilai ekstrem, yaitu metode *Block Maxima* (BM) dan metode *Peaks Over Threshold* (POT) (Rahayu, 2013).

Metode *Block Maxima* adalah metode yang mengidentifikasi nilai ekstrem berdasarkan nilai maksimum data pengamatan yang dikelompokkan berdasarkan blok atau periode tertentu. *Block Maxima* lebih sering dan lebih efektif digunakan dalam bidang hidrologi (Abad et al., 2014). Menurut Bisono (2012), kekurangan metode BM adalah terlalu banyak data pengamatan yang terbuang, terutama ketika menggunakan blok satu tahun. Selain itu, ada kemungkinan dua atau lebih data terbesar berada dalam blok atau periode yang sama, padahal dalam metode BM hanya satu nilai maksimum yang diambil, sehingga ada risiko kehilangan informasi penting. Nilai ekstrem yang diperoleh menggunakan metode BM mengikuti distribusi *Generalized Extreme Value* (GEV).

Distribusi *Generalized Extreme Value* digunakan untuk memodelkan kejadian ekstrem dalam ilmu lingkungan dan bidang ilmu lainnya. Distribusi *Generalized Extreme Value* pertama kali diperkenalkan ke bidang meteorologi oleh Jenkinson (1955) dan digunakan untuk memodelkan fenomena alam ekstrem. Beberapa peneliti telah menggunakan distribusi GEV untuk memodelkan kejadian ekstrem, seperti curah hujan ekstrem (Boudrissa et al., 2017; Iliopoulou et al., 2018; Musara et al., 2022), suhu ekstrem (Hasan et al., 2013), dan kecepatan

angin (Ben Alaya et al., 2021). Selain metode BM dengan pendekatan distribusi GEV, identifikasi nilai ekstrem juga dapat dilakukan menggunakan metode *Peaks Over Threshold* (POT).

Metode *Peaks Over Threshold* mengidentifikasi nilai ekstrem berdasarkan nilai ambang yang disebut *threshold*. Data pengamatan yang melebihi *threshold* tertentu merupakan nilai ekstrem pada metode POT. Metode *Peaks Over Threshold* telah digunakan di berbagai bidang untuk mengidentifikasi kejadian ekstrem, seperti ketinggian gelombang, banjir, kecepatan angin, dan klaim asuransi. Metode *Peaks Over Threshold* membutuhkan penentuan nilai ambang (*threshold*) yang tidak terlalu tinggi juga tidak terlalu rendah. Nilai *threshold* yang terlalu tinggi akan memperkecil bias namun variansinya besar untuk penduga parameter karena lebih sedikit data untuk estimasi parameter. Sebaliknya jika nilai *threshold* terlalu rendah akan memperbesar bias namun variansi estimatornya kecil karena lebih banyak data yang digunakan untuk estimasi parameter (Bommier, 2014). Terlepas dari kesulitan pemilihan nilai *threshold* yang tepat, metode POT menggunakan data pengamatan lebih efisien. Efisiensi data diperlukan pada EVT karena kejadian ekstrem merupakan suatu kejadian yang jarang terjadi, sehingga data pengamatan ekstrem yang tersedia sangat terbatas (Fawcett & Walshaw, 2007). Distribusi peluang dari nilai ekstrem yang diambil menggunakan metode POT mengikuti distribusi *Generalized Pareto* (GP). Distribusi *Generalized Pareto* pertama kali diperkenalkan oleh Pickands pada tahun 1975 (Pickands, 1975). Beberapa peneliti telah menggunakan distribusi GP untuk memodelkan kejadian ekstrem, seperti ketinggian gelombang air laut (Teixeira et al., 2018), suhu harian maksimum (Salleh & Hasan, 2018), dan curah hujan (Martins et al., 2020).

Beberapa penelitian menunjukkan bahwa distribusi GEV atau GP merupakan distribusi peluang yang paling cocok untuk memodelkan data curah hujan ekstrem. Sebagai contoh, Zalina et al. (2002) membandingkan delapan kandidat distribusi dalam memberikan perkiraan curah hujan maksimum yang akurat untuk Malaysia. Distribusi yang dipertimbangkan adalah distribusi Gamma, Generalized Normal, GP, GEV, Gumbel, Log Pearson tipe III, Pearson tipe III, dan Wakeby. Hasil penelitian menyimpulkan bahwa distribusi GEV merupakan distribusi yang paling tepat untuk mendeskripsikan deret curah hujan maksimum tahunan di Malaysia. Kemudian penelitian yang dilakukan oleh Shaharudin et al. (2020) menggunakan beberapa distribusi peluang untuk mencocokkan data curah

hujan harian di Malaysia. Distribusi yang dipertimbangkan adalah distribusi GP, Lognormal, dan Gamma. Hasil penelitian menunjukkan bahwa distribusi GP merupakan distribusi yang paling cocok untuk menggambarkan data curah hujan harian. Penelitian lain yang dilakukan oleh Rahman et al. (2013) mempertimbangkan lima belas distribusi peluang yang berbeda pada data banjir maksimum tahunan di Australia. Distribusi Log Pearson tipe III, GEV, dan GP diidentifikasi sebagai tiga distribusi yang paling sesuai. Karena distribusi GEV atau GP cocok untuk memodelkan data curah hujan ekstrem, sehingga kedua jenis distribusi peluang tersebut paling sering digunakan untuk menganalisis frekuensi curah hujan ekstrem.

Estimasi parameter distribusi GEV dan GP dari sampel ekstrem adalah hal yang penting. Estimasi parameter yang ideal harus memenuhi sifat tidak bias (estimator dikatakan tidak bias apabila parameter yang diestimasi mendekati dengan nilai parameter asli), efisien (estimator dikatakan efisien apabila estimator tersebut memiliki *mean square error* yang minimum), dan konsisten (estimator dikatakan konsisten apabila fungsinya *well converged*) (Kahal Musakkal et al., 2017). Salah satu metode yang dapat digunakan untuk estimasi parameter distribusi GEV dan distribusi GP adalah *Maximum Likelihood Estimation* (MLE). *Maximum Likelihood Estimation* merupakan metode klasik dan paling sering digunakan karena kesederhanaannya. Salah satu sifat dari MLE adalah asimtotik yang baik seperti konsisten dan efisien. *Maximum Likelihood Estimation* dapat digunakan pada model kompleks seperti model non-stasioner, *temporal dependence*, dan model yang memiliki efek kovariat (Coles & Dixon, 1999). Nilai hasil estimasi parameter distribusi GEV dan GP dapat digunakan untuk mengestimasi *return level*.

*Return level* adalah ukuran yang digunakan dalam analisis hidrologi atau analisis risiko bencana untuk mengevaluasi intensitas atau keparahan suatu peristiwa ekstrem. Secara umum, *return level* menunjukkan estimasi nilai maksimum, dalam hal ini adalah curah hujan maksimum, yang mungkin terjadi dalam interval waktu tertentu berdasarkan data historis atau model statistik.

Berdasarkan studi eksplorasi data curah hujan harian di kota Makassar periode Januari 1980 sampai Desember 2022, diperoleh bahwa terdapat curah hujan yang berdasarkan BMKG termasuk dalam kategori curah hujan lebat yaitu 50 – 100 mm per hari, curah hujan sangat lebat yaitu 100 mm – 150 mm per hari, dan curah hujan ekstrem yaitu curah hujan di atas 150 mm per hari (BMKG, 2022).

Pola sebaran data curah hujan di kota Makassar memiliki ekor distribusi yang gemuk (*heavy tail*). Distribusi *heavy tailed* cenderung memiliki banyak pencilan (*outlier*) dengan nilai yang sangat tinggi. Hal tersebut menandakan adanya kejadian ekstrem. Karena data curah hujan harian di Makassar memiliki ekor distribusi yang gemuk (*heavy tail*), maka dapat diterapkan metode EVT untuk menganalisis curah hujan ekstrem di kota Makassar.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana estimasi parameter distribusi GEV menggunakan metode MLE?
2. Bagaimana estimasi parameter distribusi GP menggunakan metode MLE?
3. Bagaimana mengidentifikasi kejadian curah hujan ekstrem di kota Makassar menggunakan metode BM dan POT?
4. Bagaimana *return level* curah hujan ekstrem di kota Makassar menggunakan distribusi GEV dan GP?
5. Bagaimana prediksi curah hujan ekstrem di kota Makassar pada periode mendatang menggunakan distribusi GEV dan GP?

## 1.3 Tujuan Penelitian

Tujuan yang ingin dicapai pada penelitian ini adalah untuk:

1. Memperoleh estimator parameter distribusi GEV menggunakan metode MLE.
2. Memperoleh estimator parameter distribusi GP menggunakan metode MLE.
3. Memperoleh kejadian curah hujan ekstrem di kota Makassar menggunakan metode BM dan POT.
4. Memperoleh *return level* curah hujan ekstrem di Kota Makassar menggunakan distribusi GEV dan GP.
5. Memperoleh prediksi curah hujan ekstrem di kota Makassar pada periode mendatang menggunakan distribusi GEV dan GP.

## 1.4 Batasan Masalah

Penelitian ini dibatasi pada data curah hujan harian di kota Makassar periode Januari 1980 sampai Desember 2022. Pengambilan sampel nilai ekstrem menggunakan metode BM dengan pendekatan distribusi GEV dan metode POT

dengan pendekatan distribusi GP. Estimasi parameter distribusi GEV dan GP menggunakan metode MLE.

### **1.5 Manfaat Penelitian**

Manfaat yang diharapkan dari penelitian ini adalah:

1. Hasil penelitian ini diharapkan dapat dijadikan acuan serta bahan referensi mengenai metode statistika yaitu *Extreme Value Theory* (EVT) dalam menganalisis suatu kejadian yang bersifat ekstrem.
2. Informasi mengenai estimasi curah hujan ekstrem di kota Makassar diharapkan dapat dimanfaatkan oleh pemerintah setempat untuk merencanakan langkah-langkah mitigasi bencana banjir yang efektif.

## BAB II

### TINJAUAN PUSTAKA

Pada bagian ini, dibahas tentang konsep teoritis yang digunakan pada bagian hasil dan pembahasan.

#### 2.1 *Extreme Value Theory*

*Extreme Value Theory* (EVT) atau teori nilai ekstrem adalah cabang statistika yang berkaitan dengan perilaku stokastik maksimum dan minimum variabel acak yang independen dan identik. Nilai ekstrem adalah nilai yang memiliki penyimpangan besar dari median distribusi probabilitas (Kotz & Nadarajah, 2000). *Extreme Value Theory* digunakan untuk menganalisis suatu kejadian yang bersifat ekstrem, dimana kejadian ini jarang terjadi dan berlangsung dalam waktu singkat namun memberikan dampak yang cukup besar. *Extreme Value Theory* digunakan untuk kasus univariat. Pengaplikasian EVT sudah dimulai lebih dari 50 tahun lalu dalam berbagai bidang, seperti hidrologi, klimatologi, dan teori reliabilitas.

*Extreme Value Theory* dapat meramalkan terjadinya kejadian ekstrem pada data *heavy tail* yang tidak dapat dilakukan dengan pendekatan standar. Metode ini mampu menjelaskan kerugian kejadian esktrm yang tidak dapat dimodelkan dengan pendekatan biasa. Tujuan metode ini adalah untuk menentukan estimasi peluang kejadian ekstrem dengan memperhatikan ekor (*tail*) fungsi distribusi berdasarkan nilai-nilai ekstrem yang diperoleh. Metode ini sangat mirip dengan *Central Limit Theorem* (CLT). *Extreme Value Theory* dan *Central Limit Theorem* melibatkan perilaku limit distribusi dari variabel acak yang independen dan berdistribusi identik. Tetapi terdapat perbedaan dari EVT dan CLT, yaitu CLT berkaitan dengan perilaku distribusi variabel acak sedangkan EVT hanya berkaitan dengan perilaku ekor distribusi tersebut.

Untuk menjelaskan ide dasar dari EVT, Coles (2001) mengembangkan model yang mewakili landasan teori nilai ekstrem. Model berfokus pada perilaku statistik dari

$$M_n = \max\{X_1, \dots, X_n\},$$

dimana  $X_1, \dots, X_n$  adalah deretan variabel acak dengan fungsi distribusi  $F$ . Dalam aplikasinya,  $X_i$  biasanya mewakili nilai dari proses pengukuran skala waktu, seperti



pengukuran permukaan laut setiap jam, atau suhu rata-rata harian, sehingga  $M_n$  mewakili maksimum dari proses selama  $n$  unit waktu pengamatan. Jika  $n$  adalah banyaknya pengamatan dalam satu tahun, maka  $M_n$  adalah maksimum tahunan (*annual maximum*).

Distribusi  $M_n$  dapat diturunkan untuk semua nilai  $n$ :

$$\begin{aligned}\Pr\{M_n \leq z\} &= \Pr\{X_1 \leq z_1, \dots, X_n \leq z_n\} \\ &= \Pr\{X_1 \leq z_1\} \times \dots \times \Pr\{X_n \leq z_n\} \\ &= \{F(z)\}^n\end{aligned}\quad (2.1)$$

Namun, hal ini tidak langsung membantu dalam praktik, karena fungsi distribusi  $F$  tidak diketahui. Satu kemungkinan adalah dengan menggunakan teknik statistik standar untuk mengestimasi  $F$  dari data pengamatan, kemudian melakukan substitusi estimasi tersebut ke persamaan (2.1). Tetapi, perbedaan yang sangat kecil dalam mengestimasi  $F$  dapat menyebabkan perbedaan substansial untuk  $F^n$ .

Pendekatan alternatif adalah menerima bahwa  $F$  tidak diketahui dan mencari perkiraan keluarga model untuk  $F^n$  yang dapat diestimasi hanya berdasarkan data ekstrem. Hal tersebut serupa dengan memperkirakan distribusi rata-rata sampel dengan distribusi normal, sebagaimana dibenarkan oleh teorema limit pusat.

Selanjutnya melihat perilaku  $F^n$  dengan  $n \rightarrow \infty$ . Tetapi tidak cukup untuk sebarang  $z < z_+$ , dimana  $z_+$  adalah *upper endpoint* dari  $F$  ( $z_+$  adalah nilai terkecil dari  $z$  sehingga  $F(z) = 1$ ),  $F^n(z) \rightarrow 0$  saat  $n \rightarrow \infty$ , sehingga distribusi  $M_n$  berdegenerasi (*degenerates*) menjadi massa titik (*point mass*) di  $z_+$ . Kesulitan tersebut dihindari dengan melakukan renormalisasi linier variabel  $M_n$ :

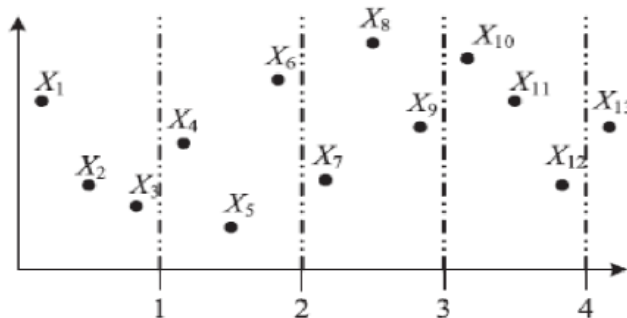
$$M_n^* = \frac{M_n - b_n}{a_n},$$

untuk deretan konstanta  $\{a_n > 0\}$  dan  $\{b_n\}$ . Pemilihan  $\{a_n\}$  dan  $\{b_n\}$  yang tepat dapat menstabilkan lokasi dan skala dari  $M_n^*$  saat  $n$  meningkat.

## 2.2 Metode *Block Maxima*

Metode *Block Maxima* (BM) atau yang dikenal *Annual Maxima* adalah metode fundamental dalam EVT dan telah banyak digunakan dalam hidrologi dan penerapan lingkungan lainnya saat menganalisis nilai ekstrem. Metode ini membagi data pengamatan menjadi beberapa blok kemudian setiap blok

ditentukan nilai yang paling tinggi (maksimum) yang disebut sebagai nilai ekstrem. Sebagai contoh  $X_1, X_2, \dots, X_{13}$  adalah deretan variabel acak seperti yang ditunjukkan pada Gambar 1.



Gambar 1. Ilustrasi metode block maxima

Gambar 1 menunjukkan ilustrasi pengambilan sampel ekstrem dengan metode BM, dimana data pengamatan dibagi menjadi empat blok. Nilai observasi maksimum pada blok pertama adalah  $X_1$ , sehingga data yang digunakan sebagai sampel untuk blok pertama adalah  $X_1$ . Untuk blok kedua, nilai observasi maksimum adalah  $X_6$  sehingga  $X_6$  dijadikan sebagai sampel untuk blok kedua. Untuk blok ketiga, nilai observasi maksimum adalah  $X_8$  sehingga  $X_8$  dijadikan sebagai sampel untuk blok ketiga. Untuk blok keempat, nilai observasi maksimum adalah  $X_{10}$  sehingga  $X_{10}$  dijadikan sebagai sampel untuk blok keempat.

## 2.3 Generalized Extreme Value

### 2.3.1 Distribusi extreme value

Seluruh rentang distribusi limit yang mungkin untuk  $M_n^*$  pada bagian 2.1 diberikan oleh Teorema 2.1, yaitu teorema tipe ekstrem.

#### Teorema 2.1 (Teorema Tipe Ekstrem)

Jika terdapat deretan konstanta  $\{a_n > 0\}$  dan  $\{b_n\}$  sehingga

$$Pr \left\{ \frac{M_n - b_n}{a_n} \leq z \right\} \rightarrow G(z) \text{ saat } n \rightarrow \infty,$$

dimana  $G$  adalah fungsi distribusi *non-degenerate*, maka  $G$  adalah bagian dari salah satu keluarga distribusi berikut:

$$\begin{aligned} \text{I: } G(z) &= e^{-e^{\left(\frac{z-b}{a}\right)}}, \quad -\infty < z < \infty \\ \text{II: } G(z) &= \begin{cases} 0 & \text{jika } z \leq b \\ e^{-\left(\frac{z-b}{a}\right)^{-\alpha}} & \text{jika } z > b \end{cases} \end{aligned}$$

$$\text{III : } G(z) = \begin{cases} e^{-\left(\frac{z-b}{a}\right)^\alpha} & \text{jika } z < b \\ 1 & \text{jika } z \geq b \end{cases}$$

dengan parameter  $a > 0$ ,  $b$ , dan  $\alpha > 0$  untuk tipe II dan III.

(Coles, 2001)

Teorema 2.1 menyatakan bahwa sampel maksimum  $\frac{M_n - b_n}{a_n}$  yang diskala ulang konvergen dalam distribusi ke variabel yang memiliki distribusi dalam salah satu dari tiga tipe distribusi. Ketiga tipe distribusi ini disebut distribusi nilai ekstrem, dengan tipe I, II, dan III dikenal dengan keluarga Gumbel, Frechet dan Weibull. Setiap keluarga distribusi memiliki parameter lokasi  $a$ , skala  $b$ , dan bentuk  $\alpha$  untuk distribusi Frechet dan Weibull.

### 2.3.2 Distribusi *generalized extreme value*

Tiga tipe distribusi pada Teorema 2.1 memiliki bentuk perilaku yang berbeda, sesuai dengan perilaku ekor fungsi distribusi  $F$ . Dalam aplikasi awal teori nilai ekstrem, biasanya mengadopsi salah satu dari tiga tipe distribusi, kemudian memperkirakan parameter yang relevan dari distribusi itu. Tetapi ada dua kelemahan: pertama, diperlukan teknik untuk memilih mana dari tiga tipe distribusi yang paling sesuai untuk data yang ada; kedua, setelah keputusan seperti itu dibuat, kesimpulan selanjutnya menganggap pilihan tersebut benar.

Analisis yang lebih baik ditawarkan dengan me-reformulasi model pada Teorema 2.1. Keluarga distribusi Gumbel, Frechet dan Weibull dapat digabungkan menjadi satu keluarga model yang memiliki fungsi distribusi seperti pada persamaan (2.2)

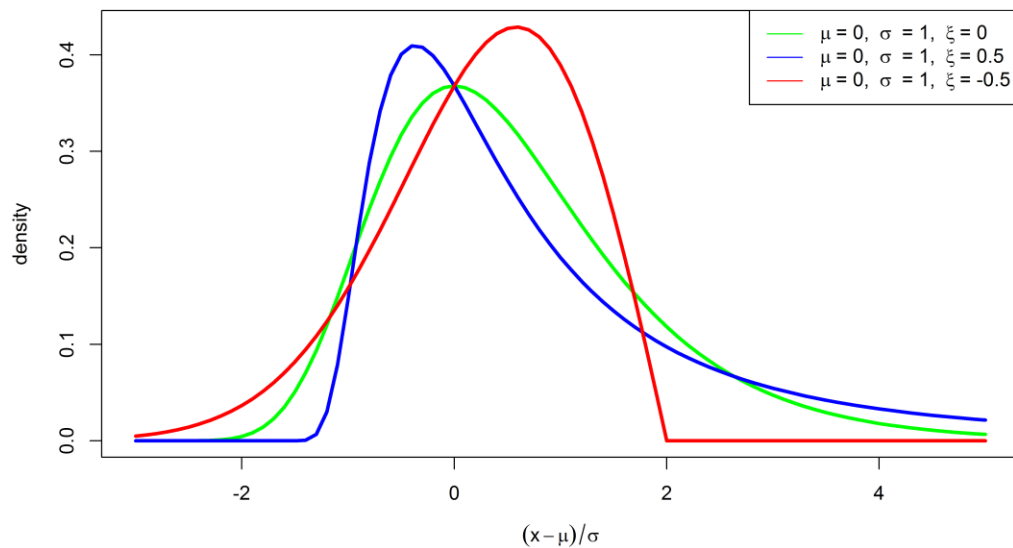
$$G(z) = e^{-\left(1 + \xi \left(\frac{z - \mu}{\sigma}\right)\right)^{-1/\xi}} \quad (2.2)$$

didefinisikan pada himpunan  $\left\{z: 1 + \xi \left(\frac{z - \mu}{\sigma}\right) > 0\right\}$ , dengan parameter  $-\infty < \mu < \infty$ ,  $\sigma > 0$  dan  $-\infty < \xi < \infty$ . Ini adalah keluarga distribusi Generalized Extreme Value (GEV). Model distribusi GEV memiliki tiga parameter: parameter lokasi (*location*)  $\mu$ , yang menentukan lokasi pusat distribusi; parameter skala (*scale*)  $\sigma$ , yang menentukan ukuran penyimpangan di sekitar parameter lokasi; dan parameter bentuk (*shape*)  $\xi$ , yang menentukan tipe distribusi GEV. Tipe distribusi GEV akan mengarah pada distribusi Frechet untuk  $\xi > 0$  dan distribusi Weibull untuk  $\xi < 0$ . Subset dari keluarga GEV dengan  $\xi = 0$  diinterpretasikan sebagai

limit dari persamaan (2.2) saat  $\xi \rightarrow 0$ , mengarah pada distribusi Gumbel dengan fungsi distribusi seperti pada persamaan (2.3).

$$G(z) = e^{-e\left(-\frac{z-\mu}{\sigma}\right)} \quad (2.3)$$

Semakin besar nilai  $\xi$ , maka distribusi memiliki ekor yang semakin berat (*heavy tail*) sehingga akan berdampak pada peluang terjadinya nilai ekstrem yang semakin besar. Untuk parameter bentuk dengan  $\xi = 0$  dikatakan "*medium tail*", untuk  $\xi > 0$  dikatakan "*long tail*", dan untuk  $\xi < 0$  dikatakan "*short tail*". Ketiga tipe distribusi GEV menunjukkan bahwa distribusi yang memiliki ekor paling berat (*heavy tail*) adalah distribusi Frechet. Gambar 2 menunjukkan kurva ketiga distribusi *extreme value*.



Gambar 2. Kurva PDF distribusi Gumbel, Frechet, dan Weibull

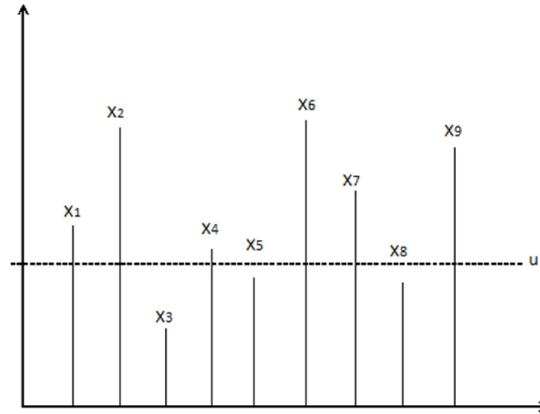
Pada Gambar 2, kurva berwarna hijau merupakan kurva distribusi Gumbel, kurva berwarna biru merupakan kurva distribusi Frechet, dan kurva berwarna merah merupakan kurva distribusi Weibull. Distribusi Gumbel dan Frechet memiliki ujung distribusi yang tidak terbatas, sedangkan distribusi Weibull memiliki ujung distribusi yang terbatas.

Penyatuan tiga tipe dari distribusi nilai ekstrem menjadi satu keluarga sangat menyederhanakan implementasi statistik. Melalui inferensi, data sendiri yang menentukan tipe yang paling tepat dari perilaku ekor. *Probability Density Function* (PDF) untuk distribusi GEV diberikan oleh persamaan (2.4).

$$g(z) = \begin{cases} \frac{1}{\sigma} \left[ 1 + \xi \left( \frac{z-\mu}{\sigma} \right) \right]^{-\left(\frac{1}{\xi}\right)-1} e^{-\left(1+\xi\left(\frac{z-\mu}{\sigma}\right)\right)^{-1/\xi}} & , \xi \neq 0, 1 + \xi \left( \frac{z-\mu}{\sigma} \right) > 0 \\ \frac{1}{\sigma} e^{-\left(\frac{z-\mu}{\sigma}\right)} e^{-e^{-\left(\frac{z-\mu}{\sigma}\right)}} & , \xi = 0 \end{cases} \quad (2.4)$$

## 2.4 Metode Peaks Over Threshold

Metode *Peaks Over Threshold* (POT) adalah metode pada EVT yang dalam mengidentifikasi nilai ekstrem berdasarkan pada nilai ambang batas yang disebut dengan *threshold* ( $u$ ). Data yang melebihi nilai *threshold* akan diidentifikasi sebagai nilai ekstrem. Gambar 3 menunjukkan ilustrasi pemilihan nilai ekstrem dengan menggunakan *threshold*.



Gambar 3. Ilustrasi metode POT (sumber: Susan (2015))

Misalkan  $X_1, X_2, \dots, X_n$  merupakan deretan variabel acak dengan fungsi distribusi  $F$  dan *right endpoint*  $x_F = \sup\{x; F(x) < 1\}$ . Untuk setiap  $u < x_F$ , fungsi

$$F_u(x) = \Pr\{X - u \leq x | X > u\}, \quad x \geq 0 \quad (2.5)$$

disebut fungsi distribusi dari nilai-nilai melebihi *threshold*  $u$ . Dengan probabilitas bersyarat,  $F_u$  juga dapat didefinisikan sebagai

$$F_u(x) = \begin{cases} \frac{F(u+x) - F(u)}{1 - F(u)}, & x \geq 0 \\ 0, & x \text{ yang lain} \end{cases} \quad (2.6)$$

Misalkan  $Y = X - u$  untuk  $X > u$  dan untuk  $n$  variabel pengamatan  $X_1, X_2, \dots, X_n$ , dapat ditulis  $Y_j = X_i - u$  sehingga  $i$  adalah indeks pelampauan ke  $j$ ,  $j = 1, \dots, n_u$ . Distribusi dari pelampauan  $(Y_{1, \dots, Y_{n_u}})$  dapat didekati dengan distribusi *Generalized Pareto* (Pickands, 1975).

## 2.5 Distribusi Generalized Pareto

Distribusi *Generalized Pareto* (GP) adalah keluarga distribusi peluang kontinu, sering digunakan untuk memodelkan ekor distribusi. Sama halnya dengan distribusi GEV, distribusi GP memiliki tiga parameter, yaitu lokasi  $\mu$ , skala  $\sigma$ , dan bentuk  $\xi$ . Namun terkadang distribusi GP ditentukan hanya dengan parameter

skala dan bentuk (Hosking & Wallis, 1987). Distribusi *Generalized Pareto* memiliki fungsi distribusi seperti pada persamaan (2.7)

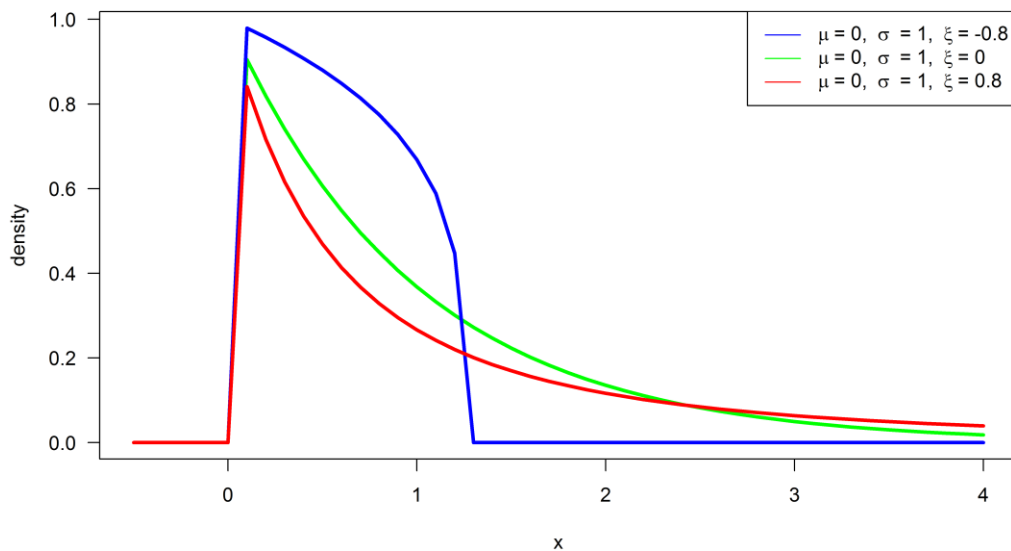
$$F(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - e\left(-\frac{y}{\sigma}\right), & \xi = 0 \end{cases} \quad (2.7)$$

dimana  $y \geq 0$  jika  $\xi \geq 0$  dan  $0 \leq y \leq u - \frac{\sigma}{\xi}$  jika  $\xi < 0$ .

*Probability Density Function* (PDF) distribusi GP adalah sebagai berikut:

$$f(x - u) = \begin{cases} \frac{1}{\sigma} \left[1 + \xi \frac{(x-u)}{\sigma}\right]^{-\frac{1}{\xi}-1}, & \xi \neq 0 \\ \frac{1}{\sigma} e\left(-\frac{(x-u)}{\sigma}\right), & \xi = 0 \end{cases} \quad (2.8)$$

Berdasarkan nilai dari  $\xi$ , distribusi GP dapat dibedakan menjadi tiga tipe. Jika  $\xi = 0$ , maka distribusi GP memiliki ekor tipe I dan ekuivalen dengan distribusi eksponensial dengan mean  $\sigma$ . Jika  $\xi > 0$ , distribusi GP memiliki ekor tipe II dan mengarah ke distribusi Pareto. Jika  $\xi < 0$ , memiliki ekor tipe III dan memiliki kasus khusus dari distribusi Beta (Caires, 2016). Gambar 4 menampilkan kurva PDF dari tiga distribusi GP dengan parameter  $\xi$  yang berbeda.



Gambar 4. Kurva PDF distribusi GP

## 2.6 Pemilihan *Threshold*

Pemilihan *threshold* pada metode POT adalah hal yang sangat penting. Nilai *threshold* perlu dipilih dengan hati-hati karena nilai *threshold* yang terlalu kecil dapat mengakibatkan parameter yang bias dan apabila terlalu tinggi maka jumlah



observasi semakin sedikit dan variansinya menjadi tinggi. Ada beberapa cara dalam menentukan *threshold*, yaitu dengan *mean residual life plot* (atau dikenal dengan *mean excess plot*) dan *parameter stability plot* (Bommier, 2014).

### 2.6.1 Mean residual life plot

Davison dan Smith (1990) mengusulkan metode grafis untuk pemilihan *threshold*. Metode ini berdasarkan pada *mean* dari distribusi GP. Nilai ekspektasi dari distribusi GP dengan parameter  $\sigma$  dan  $\xi$  adalah

$$E(Y) = \frac{\sigma}{1-\xi} \text{ untuk } \xi < 1 \quad (2.9)$$

Ketika  $\xi \geq 1$ , *mean* tidak terhingga.

Misalkan distribusi GP adalah model yang valid untuk nilai-nilai yang melebihi *threshold*  $u_0$  diperoleh dari deretan  $X_1, X_2, \dots, X_n$ . Dari persamaan (2.9),

$$E(X - u_0 | X > u_0) = \frac{\sigma_{u_0}}{1-\xi} \quad (2.10)$$

untuk  $\xi < 1$  dan  $\sigma_{u_0}$  adalah parameter skala distribusi GP dengan *threshold*  $u_0$ .

Sifat stabilitas *threshold* dari distribusi GP berarti bahwa jika distribusi GP adalah model yang valid untuk nilai-nilai yang melebihi *threshold*  $u_0$ , maka berlaku pula untuk kelebihan atas semua *threshold*  $u > u_0$ . Jadi, untuk  $u > u_0$ ,

$$E(X - u | X > u) = \frac{\sigma_u}{1-\xi} = \frac{\sigma_{u_0} + \xi u}{1-\xi} \quad (2.11)$$

Dengan demikian, untuk semua  $u > u_0$ ,  $E(X - u | X > u)$  adalah fungsi linier dari  $u$ . Lebih jauh,  $E(X - u | X > u)$  adalah *mean* dari kelebihan ambang batas (*threshold*)  $u$  dan dapat diestimasi dengan *mean* sampel. Hal ini mengarah pada *mean residual life plot* yang ditentukan oleh tempat kedudukan titik

$$\left\{ \left( u, \frac{1}{n_u} \sum_{i=1}^{n_u} (x_{(i)} - u) \right); u < x_{max} \right\}$$

dimana  $x_{(1)}, \dots, x_{(n_u)}$  terdiri dari pengamatan  $n_u$  yang melebihi  $u$ , dan  $x_{max}$  adalah nilai maksimum dari  $X_i$ .

Hal ini berarti bahwa untuk rentang *threshold*  $u$ , diidentifikasi *mean threshold* yang bersesuaian, kemudian buat plot *mean threshold excess* terhadap  $u$ , dan cari nilai  $u_0$  dimana dapat dilihat kelinieran dalam plot. Kelemahan dari plot ini adalah dibutuhkan interpretasi dan hal tersebut bisa jadi menantang.

## 2.6.2 Parameter stability plot

Metode grafis lain yang sering digunakan untuk menentukan *threshold*  $u$  adalah *parameter stability plot*. Ide dari plot ini adalah jika nilai-nilai yang melebihi sebuah *threshold* tinggi  $u_0$  mengikuti distribusi GP dengan parameter  $\xi$  dan  $\sigma_{u_0}$ , maka untuk sebarang *threshold*  $u$  sedemikian sehingga  $u > u_0$ , nilai-nilai yang terlampaui masih mengikuti distribusi GP dengan parameter bentuk  $\xi_u = \xi$  dan parameter skala  $\sigma_u = \sigma_{u_0} + \xi(u - u_0)$ . Misalkan

$$\sigma^* = \sigma_u - \xi_u u \quad (2.12)$$

Parameterisasi baru ini tidak lagi bergantung pada  $u$ , mengingat  $u_0$  adalah *threshold* yang cukup tinggi. *Parameter stability plot* ditentukan oleh tempat kedudukan titik (*locus of points*)

$$\{(u, \sigma^*); u < x_{max}\} \text{ dan } \{(u, \xi_u); u < x_{max}\} \quad (2.13)$$

Dengan demikian, estimasi dari  $\sigma^*$  dan  $\xi_u$  adalah konstan untuk setiap  $u > u_0$  jika  $u_0$  adalah *threshold* yang cocok untuk pendekatan asimtotik. *Threshold* dipilih pada nilai dimana parameter skala dan bentuk tetap konstan.

## 2.7 Estimasi Parameter

*Maximum likelihood estimation* (MLE) merupakan salah satu teknik untuk menaksir parameter populasi yang tidak diketahui. Metode *maximum likelihood* sejauh ini merupakan teknik yang paling sering digunakan untuk menurunkan estimator. Jika  $X_1, \dots, X_n$  adalah peubah acak yang saling bebas dari populasi dengan PDF  $f(x|\theta_1, \dots, \theta_k)$ , maka fungsi *likelihood* didefinisikan sebagai fungsi kepadatan peluang bersama (*joint PDF*) dan dinyatakan seperti pada persamaan (2.14) (Casella & Berger, 2002).

$$L(\theta|x) = L(\theta_1, \dots, \theta_k|x_1, \dots, x_n) = \prod_{i=1}^n f(x_i|\theta_1, \dots, \theta_k) \quad (2.14)$$

Dalam prosesnya, metode ini berupaya menemukan nilai estimator bagi parameter yang dapat memaksimalkan fungsi *likelihood*. Dari persamaan (2.14),  $L(\theta|x)$  didefinisikan sebagai hasil kali (*product*) sebanyak  $n$  bentuk, yang tidak mudah untuk dimaksimumkan. Memaksimumkan  $L(\theta|x)$  ekuivalen dengan memaksimumkan  $\log L(\theta|x)$  karena log adalah fungsi monoton naik. Fungsi *log-likelihood* dapat ditulis dalam bentuk

$$l(\theta|x) = \log L(\theta|x) = \log \prod_{i=1}^n f(x_i|\theta_1, \dots, \theta_k) = \sum_{i=1}^n \log f(x_i|\theta_1, \dots, \theta_k) \quad (2.15)$$

Jika fungsi *log-likelihood* terdiferensialkan di  $\theta_i$ , kandidat yang mungkin untuk MLE adalah nilai dari  $(\theta_1, \dots, \theta_k)$  yang memenuhi

$$\frac{\partial}{\partial \theta_i} \log L(\theta|x) = 0, \quad i = 1, \dots, k \quad (2.16)$$

Perhatikan bahwa solusi persamaan (2.16) hanya kandidat yang mungkin dari MLE karena turunan pertama sama dengan nol hanya syarat perlu untuk maksimum, bukan syarat cukup.

Untuk beberapa kasus, mungkin untuk mendapatkan estimator secara eksplisit dengan menyelesaikan persamaan (2.16) terhadap masing-masing parameter. Namun, sulit untuk mendapatkan solusi secara analitik ketika model melibatkan banyak parameter dan PDF nonlinier. Sehingga estimasi parameter harus dicari secara numerik menggunakan algoritma optimasi nonlinier.

### Definisi 2.1 (Turunan Parsial)

Misalkan  $f$  adalah fungsi dua variabel  $x$  dan  $y$ . Jika  $y$  dipertahankan konstan, katakan  $y = y_0$ , maka  $f(x, y_0)$  adalah fungsi dengan variabel tunggal  $x$ . Turunannya di  $x = x_0$  disebut turunan parsial fungsi  $f$  terhadap  $x$  di  $(x_0, y_0)$  dan dinotasikan dengan  $f_x(x_0, y_0)$ . Dengan demikian,

$$f_x(x_0, y_0) = \lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x, y_0) - f(x_0, y_0)}{\Delta x}$$

Hal yang sama dengan turunan parsial fungsi  $f$  terhadap  $y$  di  $(x_0, y_0)$  dinotasikan dengan  $f_y(x_0, y_0)$  dan dinyatakan sebagai

$$f_y(x_0, y_0) = \lim_{\Delta y \rightarrow 0} \frac{f(x_0, y_0 + \Delta y) - f(x_0, y_0)}{\Delta y}$$

Poin pentingnya adalah aturan untuk turunan fungsi satu variabel juga berlaku untuk menemukan turunan parsial. Simbol  $\partial$  biasanya digunakan sebagai tanda turunan parsial. Jika  $z = f(x, y)$ , dapat digunakan notasi alternatif

$$f_x(x, y) = \frac{\partial z}{\partial x} = \frac{\partial f(x, y)}{\partial x} \quad f_y(x, y) = \frac{\partial z}{\partial y} = \frac{\partial f(x, y)}{\partial y}$$

(Varberg et al., 2007)

### Teorema 2.2 (Aturan Hasil Kali)

Jika  $f$  dan  $g$  adalah fungsi yang terdiferensiasi, maka

$$(f \cdot g)'(x) = f(x)g'(x) + g(x)f'(x)$$

Dengan kata lain,

$$D_x[f(x)g(x)] = f(x)D_x g(x) + g(x)D_x f(x)$$

(Varberg et al., 2007)

### Teorema 2.3 (Aturan Rantai)

Misalkan  $y = f(u)$  dan  $u = g(x)$ . Jika  $g$  terdiferensialkan di  $x$  dan  $f$  terdiferensialkan di  $u = g(x)$ , maka fungsi komposisi  $f \circ g$ , didefinisikan sebagai  $(f \circ g)(x) = f(g(x))$ , terdiferensialkan di  $x$  dan

$$(f \circ g)'(x) = f'(g(x))g'(x)$$

Dengan kata lain,

$$D_x(f(g(x))) = f'(g(x))g'(x)$$

atau

$$\frac{dy}{dx} = \frac{dy}{du} \frac{du}{dx}$$

(Varberg et al., 2007)

## 2.8 Optimasi Numerik

### 2.8.1 Optimasi tanpa kendala

Perhatikan masalah matematik untuk menemukan fungsi  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  dari beberapa variabel argumen yang sesuai dengan nilai fungsi minimal

$$x^* = \operatorname{argmin} f(x), x \in \mathbb{R}^n$$

Fungsi  $f$  disebut fungsi objektif (*objective function*), dan  $x^*$  adalah solusi. Fungsi  $f$  diasumsikan terdiferensiasi dua kali. Solusi  $x^* \in X$  adalah minimum global dari  $f$  jika  $f(x^*) \leq f(x) \forall x \in X$ , dimana  $X$  adalah wilayah layak atau himpunan kendala (*constraint set*).

Masalah memaksimumkan dapat dengan mudah dirumuskan kembali menjadi masalah meminimumkan dengan mengubah tanda dari objektif. Masalah optimasi berasal dari berbagai kondisi, dan kriteria berikut dapat digunakan untuk klasifikasi: banyaknya variabel, banyaknya kendala dan sifat fungsi objektif (linier, kuadrat, nonlinier, konveks).

Berikut ini didefinisikan gradien dari  $f(x)$ ,  $\nabla f(x)$ ,  $x \in \mathbb{R}^n$  dan matriks Hessian  $\{\nabla^2 f(x)\}_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$  seperti pada persamaan (2.17) (Gilli et al., 2019).

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_b} \end{bmatrix} \text{ dan } \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} \quad (2.17)$$

**Definisi 2.2 (Matriks Simetrik)**

Suatu matriks bujursangkar  $A$  dikatakan simetrik jika  $A = A^T$ .

(Anton & Rorres, 2004)

**Contoh 2.1:**

Misalkan matriks  $A = \begin{bmatrix} 1 & -3 & 2 \\ -3 & 4 & 6 \\ 2 & 6 & 1 \end{bmatrix}$ . Karena matriks  $A = A^T$ , maka matriks  $A$  disebut matriks simetrik.

**Definisi 2.3 (Submatriks Utama yang Pertama)**

Diberikan suatu matriks  $A$  berukuran  $n \times n$ . Misalkan  $A_r$  adalah matriks yang terbentuk dengan menghilangkan  $n - r$  baris terakhir dan  $n - r$  kolom terakhir dari  $A$ , maka  $A_r$  disebut submatriks utama yang pertama (*leading principal submatrix*) dari  $A$  dengan ukuran  $r \times r$  ( $1 \leq r \leq n$ )

(Leon, 2001)

**Contoh 2.2:**

Dari contoh 2.1, submatriks utama yang pertama dari matriks  $A$  adalah:

$$A_1 = [1]; A_2 = \begin{bmatrix} 1 & -3 \\ -3 & 4 \end{bmatrix}; A_3 = \begin{bmatrix} 1 & -3 & 2 \\ -3 & 4 & 6 \\ 2 & 6 & 1 \end{bmatrix}.$$

**Teorema 2.4 (Matriks Definit Positif)**

Misalkan  $A$  matriks simetrik berukuran  $n \times n$  dan  $A_i$  adalah submatriks utama yang pertama dari matriks  $A$  dengan  $i = 1, 2, \dots, n$ . Matriks  $A$  disebut matriks definit positif jika dan hanya jika  $\det(A_i) > 0$ .

(Horn & Johnson, 2012)

**2.8.2 Metode Newton**

Metode Newton merupakan salah satu metode yang digunakan untuk menyelesaikan persamaan non-linier secara numerik. Metode Newton dapat digunakan untuk kasus optimasi satu dimensi (*one-dimensional optimization*) dan masalah  $n$  dimensi. Metode Newton diturunkan dari ekspansi Taylor sebuah fungsi  $f$  di persekitaran  $x$ ,

$$f(x + \Delta_x) = f(x) + \Delta_x f'(x) + R(x) \quad (2.18)$$

Dengan mengabaikan sisa  $R(x)$ , dicari langkah  $\Delta_x$  yang mana  $f(x + \Delta_x) = 0$

$$f(x_{k+1}) \approx f(x_k) + (x_{k+1} - x_k) f'(x) = 0 \quad (2.19)$$

Sehingga dapat dihitung langkah  $s$  dari  $x_k$  ke  $x_{k+1}$  dengan  $x_{k+1}$  adalah sebagai berikut

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \quad (2.20)$$

Algoritma metode Newton untuk optimasi tanpa kendala fungsi satu dimensi adalah sebagai berikut (Gilli et al., 2019):

1. Inisialisasi  $k = 0, x = x^{(0)}$ ;
2. Hitung  $f'(x^{(0)})$  dan  $f''(x^{(0)})$ ;
3. Update solusi  $x^{(k+1)} = x^{(k)} - \frac{f'(x)}{f''(x)}$ ;
4. Ulangi langkah 2 dengan  $k = k + 1$  hingga konvergen.

Algoritma metode Newton untuk optimasi tanpa kendala dalam  $n$  dimensi adalah sebagai berikut (Gilli et al., 2019):

1. Inisialisasi  $k = 0, x = x^{(0)}$ ;
2. Hitung  $\nabla f(x^{(0)})$  dan  $\nabla^2 f(x^{(0)})$ ;
3. Selesaikan  $\nabla^2 f(x^{(0)})s^{(0)} = -\nabla f(x^{(0)})$ ;
4. Update solusi  $x^{(k+1)} = x^{(k)} + s^{(k)}$ ;
5. Ulangi langkah 2 dengan  $k = k + 1$  hingga konvergen.

Dalam metode Newton seringkali sulit untuk mendapatkan matriks Hessian untuk masalah yang diberikan karena perhitungan turunan kedua dari fungsi tujuan. Dengan demikian, metode Newton jarang digunakan dalam bentuk klasiknya untuk masalah pemrograman nonlinier. Untuk mengatasi kesulitan ini, metode quasi-Newton telah digunakan dalam banyak masalah pemrograman nonlinier.

### 2.8.3 Metode Broyden-Fletcher-Goldfarb-Shanno quasi-Newton

Metode quasi-Newton merupakan generalisasi dari metode secant untuk masalah satu dimensi. Kondisi secant pada masalah multi dimensi adalah sebagai berikut

$$\nabla^2 f(x^{(k+1)}) \cdot (x^{(k+1)} - x^{(k)}) \approx \nabla f(x^{(k+1)}) - \nabla f(x^{(k)}) \quad (2.21)$$

Pada pendekatan quasi-Newton, matriks Hessian  $\nabla^2 f(x^{(k+1)})$  digantikan dengan matriks  $H^{(k+1)}$ , yang menghasilkan biaya komputasi yang lebih rendah untuk menghitung matriks Hessian. Perhatikan persamaan

$$H^{(k+1)}d^k = y^k \quad (2.22)$$

dimana vektor  $d^k$  dan  $y^k$  didefinisikan sebagai

$$d^k = x^{(k+1)} - x^{(k)} \quad (2.23)$$

$$y^k = \nabla f(x^{(k+1)}) - \nabla f(x^{(k)}) = g^{(k+1)} - g^{(k)}. \quad (2.24)$$

Metode quasi-Newton menghitung matriks Hessian  $H^{(k+1)}$  dari matriks Hessian sebelumnya  $H^{(k)}$  secara iteratif selama proses pemrograman. Metode Broyden-Fletcher-Goldfarb-Shanno (BFGS) adalah metode iteratif dengan formula BFGS diberikan pada persamaan (2.25), diusulkan untuk memperbarui matriks Hessian.

$$H^{(k+1)} = H^{(k)} + \frac{g^{(k)}g^{(k)T}}{g^{(k)T}d^k} + \frac{y^k y^{kT}}{y^{kT}d^k} \quad (2.25)$$

Biasanya, masalahnya disederhanakan lebih lanjut dengan menggunakan invers matriks  $A^{(k)} = H^{(-k)}$ . Metode iteratif BFGS dalam menghitung  $A^{(k+1)}$  diberikan oleh

$$A^{(k+1)} = \left( I - \frac{d^{(k)}y^{(k)T}}{d^{(k)T}y^k} \right) A^{(k)} \left( I - \frac{y^{(k)}d^{(k)T}}{d^{(k)T}y^k} \right) + \frac{d^{(k)}d^{(k)T}}{d^{(k)T}y^k} \quad (2.26)$$

Algoritma metode BFGS quasi-Newton adalah sebagai berikut (Gilli et al., 2019):

1. Inisialisasi  $x_0 \in \mathfrak{R}^n$  dan sebuah matriks Hessian  $H_0$ ;
2. Hitung  $d_k = -H_0 g_0$ , dimana  $g_0 = \nabla f(x_0)$ ;
3. Perbarui solusi  $x_{k+1} = x_k + \alpha_k d_k$   
 $\alpha_k$  adalah ukuran langkah (skalar) yang memenuhi kondisi  $f(x_k + \alpha_k d_k) \leq f(x_k) + b \alpha_k g_k^T d_k$  dengan  $b \in (0,1)$ ;
4. Hitung  $\rho_k = x_{k+1} - x_k$ ;
5. Hitung  $w_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ ;
6. Perbarui  $H_{k+1} = H_k - \frac{(H_k \rho_k)(H_k \rho_k)^T}{\rho_k^T H_k \rho_k} + \frac{w_k w_k^T}{w_k^T \rho_k}$ ;
7. Periksa konvergensi: jika  $\|\nabla f(x_k)\| < \varepsilon$ , maka iteasi berhenti, jika tidak kembali ke langkah 3 dengan menetapkan  $k = k + 1$ .

## 2.9 Uji Kesesuaian Distribusi

Uji kesesuaian distribusi bertujuan untuk mengidentifikasi apakah pola sebaran suatu data sesuai dengan pola sebaran teoritis. Salah satu metode yang dapat digunakan untuk menguji kesesuaian atau *goodness of fit* fungsi distribusi probabilitas teoritis terhadap fungsi probabilitas empiris adalah uji Kolmogorov Smirnov. Uji Kolmogorov Smirnov adalah uji nonparametrik dari persamaan

kontinu, distribusi probabilitas satu dimensi yang dapat digunakan untuk membandingkan sebuah sampel dengan distribusi probabilitas referensi (Siregar, 2015). Uji Kolmogorov Smirnov menetapkan suatu titik dengan kedua distribusi memiliki perbedaan terbesar. Misalkan  $X$  adalah variabel acak yang berasal dari populasi yang mengikuti suatu distribusi teoritis tertentu.  $F_0(x)$  adalah fungsi distribusi frekuensi kumulatif teoritis dan  $S_n(x)$  adalah fungsi distribusi frekuensi kumulatif yang diamati dari suatu sampel acak dengan  $n$  observasi.

Hipotesis yang digunakan pada uji Kolmogorov Smirnov adalah sebagai berikut.

$H_0$  :  $X$  berasal dari populasi yang mengikuti suatu distribusi teoritis tertentu atau

$$S_n(x) = F_0(x)$$

$H_1$  :  $X$  tidak berasal dari populasi yang mengikuti suatu distribusi teoritis tertentu

$$\text{atau } S_n(x) \neq F_0(x)$$

Statistik uji yang digunakan adalah:

$$D_{hitung} = \max |S_n(x) - F_0(x)| \quad (2.27)$$

Daerah kritis:

Tolak  $H_0$  jika  $D_{hitung} > D_\alpha$  pada tabel Kolmogorov Smirnov satu sampel dengan taraf signifikansi  $\alpha$ .

## 2.10 Uji Stasioneritas

Untuk memenuhi asumsi stasioneritas distribusi GEV dan GP, salah satu teknik yang dapat digunakan adalah uji Mann-Kendall. Uji Mann-Kendall merupakan uji statistika nonparametrik yang digunakan untuk melihat ada tidaknya kecenderungan pada suatu seri data yang didasarkan atas rangking relatif dari data runtun waktu (Indarto et al., 2011). Uji Mann-Kendall adalah salah satu uji coba bebas distribusi yang banyak digunakan dalam data runtun waktu. Uji Mann-Kendall secara umum banyak digunakan untuk mendeteksi kecenderungan dalam runtun waktu misalnya untuk fluktuasi yang terjadi di alam. Kelebihan metode Mann-Kendall adalah dapat mendeteksi kecenderungan data runtun waktu, baik kecenderungan linier maupun non-linier dan data tidak harus berdistribusi secara normal.

Adapun asumsi uji Mann-Kendall adalah sebaga berikut:

- a. Pengukuran (data pengamatan) yang diperoleh dari waktu ke waktu bersifat independen (pengamatan tidak berkorelasi secara serial dari waktu ke waktu) dan berdistribusi secara identik.



- b. Pengamatan yang diperoleh dari waktu ke waktu mewakili kondisi sebenarnya pada waktu sampling.
- c. Metode pengumpulan, pengamatan, dan pengukuran sampel memberikan pengamatan yang tidak bias dan representatif terhadap populasi yang mendasari dari waktu ke waktu.

Rumusan hipotesis dari uji Mann-Kendall adalah sebagai berikut.

- a. Pengujian dua pihak

$H_0: S = 0$  (tidak ada kecenderungan pada data runtun waktu)

$H_1: S \neq 0$  (terdapat kecenderungan pada data runtun waktu)

- b. Pengujian satu pihak

- 1. Uji pihak kanan

$H_0: S = 0$  (tidak ada kecenderungan pada data runtun waktu)

$H_1: S > 0$  (terdapat kecenderungan pada data runtun waktu)

- 2. Uji pihak kiri

$H_0: S = 0$  (tidak ada kecenderungan pada data runtun waktu)

$H_1: S < 0$  (terdapat kecenderungan pada data runtun waktu)

Jika diketahui data runtun waktu dengan panjang data  $n (x_1, x_2, \dots, x_n)$ , maka statistik uji diberikan oleh persamaan (2.28) berikut.

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n a_{ij} \quad (2.28)$$

dimana

$$a_{ij} = \text{sign}(x_j - x_i) = \text{sign}(R_j - R_i) = \begin{cases} 1 & \text{jika } x_j - x_i > 0 \\ 0 & \text{jika } x_j - x_i = 0 \\ -1 & \text{jika } x_j - x_i < 0 \end{cases} \quad (2.29)$$

$R_i$  dan  $R_j$  adalah urutan pengamatan  $x_i$  dan  $x_j$  dari runtun waktu dengan asumsi bahwa data tersebut independen dan berdistribusi secara identik. Fungsi  $\text{sign}$  nilainya 1, 0, atau -1 dan nilai tersebut tergantung dari  $(x_j - x_i)$  yang positif, nol, atau negatif. Jika nilai  $S$  positif, maka menunjukkan adanya kenaikan kecenderungan. Jika nilai  $S$  sama dengan nol, maka tidak terdapat kecenderungan. Sedangkan jika nilai  $S$  negatif, maka menunjukkan penurunan kecenderungan. Distribusi nilai  $S$  dapat menggunakan pendekatan distribusi normal untuk ukuran sampel yang besar dengan rata-rata  $E(S)$  dan variansi  $\text{Var}(S)$  sebagai berikut

$$E(S) = 0 \quad (2.30)$$

$$Var(S) = \frac{n(n-1)(2n+5)}{18} \quad (2.31)$$

dimana  $n$  adalah jumlah observasi. Keberadaan barisan terikat (observasi yang sama) dalam data menghasilkan pengurangan variansi  $S$  menjadi

$$Var(S) = n(n-1)(2n+5)/18 - \sum_{j=1}^m t_j(t_j-1)(2t_j+5)/18 \quad (2.32)$$

dimana  $m$  adalah kelompok barisan terikat masing-masing dengan observasi  $t_j$ .

Pengujian hipotesis menggunakan uji Z seperti pada persamaan (2.33) berikut

$$Z = \begin{cases} \frac{(S-1)}{\sqrt{Var(S)}} & \text{jika } S > 0 \\ 0 & \text{jika } S = 0 \\ \frac{(S+1)}{\sqrt{Var(S)}} & \text{jika } S < 0 \end{cases} \quad (2.33)$$

Nilai Z merupakan nilai standar sebaran normal dan  $\alpha$  adalah tingkat kepercayaan. Tingkat kepercayaan yang digunakan dalam penelitian ini adalah 95%, dengan nilai signifikan Z dan  $\alpha = 5\%$ .

Kriteria pengujian hipotesis uji Mann-Kendall adalah sebagai berikut.

a. Pengujian dua pihak

$H_0$  diterima jika  $-Z_{(\alpha/2)} \leq Z_{hitung} \leq Z_{(\alpha/2)}$  dan  $H_0$  ditolak jika  $Z_{hitung} < -Z_{(\alpha/2)}$  atau  $Z_{hitung} > Z_{(\alpha/2)}$ .

b. Pengujian satu pihak

1. Uji pihak kanan

$H_0$  diterima jika  $Z \leq Z_{(\alpha)}$  dan  $H_0$  ditolak jika  $Z > Z_{(\alpha/2)}$ .

2. Uji pihak Kiri

$H_0$  diterima jika  $Z \geq Z_{(\alpha)}$  dan  $H_0$  ditolak jika  $Z < Z_{(\alpha/2)}$ .

## 2.11 Return Level

Secara umum estimasi  $p$ -year return level yaitu  $\hat{z}_p$  adalah nilai yang akan dilampaui sekali setiap  $p$  year dengan probabilitas  $\frac{1}{p}$ . Persamaan  $p$  year return level untuk distribusi GEV adalah sebagai berikut (Coles, 2001):

$$\hat{z}_p = \begin{cases} \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}} \left[ (-\log(1-p^{-1}))^{-\hat{\xi}} - 1 \right], & \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\sigma} \log(-\log(1-p^{-1})), & \hat{\xi} = 0 \end{cases} \quad (2.34)$$

Penentuan *return level* pada distribusi GP melibatkan parameter  $\xi, \sigma$ , dan nilai *threshold*  $u$ . Persamaan *p year return level* untuk distribusi GP adalah sebagai berikut:

$$\hat{z}_p = \begin{cases} \hat{u} + \frac{\hat{\sigma}}{\xi} \left[ (pn_x \hat{\lambda}_u)^\xi - 1 \right], & \xi \neq 0 \\ \hat{u} + \hat{\sigma} \log(pn_x \hat{\lambda}_u), & \xi = 0 \end{cases} \quad (2.35)$$

dimana  $p$  adalah *return period* (tahun),  $n_x$  adalah banyaknya pengamatan setiap tahun, dan  $\hat{\lambda}_u$  adalah peluang pengamatan individu melebihi *threshold*  $u$  dan dapat diestimasi menggunakan persamaan (2.36).

$$\hat{\lambda}_u = \frac{k}{n} \quad (2.36)$$

dimana  $k$  adalah banyaknya nilai yang melewati *threshold*  $u$  dan  $n$  adalah banyaknya pengamatan.

## 2.12 Curah Hujan

Curah hujan merupakan ketinggian air hujan yang terkumpul dalam tempat yang datar, tidak menguap, tidak meresap, dan tidak mengalir. Di Indonesia satuan curah hujan adalah milimeter (mm). Curah hujan 1 (satu) milimeter artinya dalam luasan satu meter persegi pada tempat yang datar tertampung air sebanyak satu liter (Maulidani et al., 2015). Menurut Supari et al. (2012), jumlah hari hujan ditentukan berdasarkan jumlah hari dengan batas curah hujan 1 milimeter.

Menurut BMKG (2022), ambang batas nilai yang digunakan untuk menentukan intensitas curah hujan sebagai berikut:

- a. 0 mm/hari : berawan
- b. 0.5 – 20 mm/ hari : hujan ringan
- c. 20 – 50 mm/hari : hujan sedang
- d. 50 – 100 mm/ hari : hujan lebat
- e. 100 – 150 mm/ hari : hujan sangat lebat
- f. >150 mm/ hari : hujan ekstrem

Berdasarkan terjadinya, hujan dibedakan menjadi lima jenis, yaitu (Hermawan, 2015):

- a. Hujan siklonal

Hujan siklonal adalah hujan yang terjadi karena udara panas yang naik disertai dengan angin berputar.

b. Hujan zenithal

Hujan zenithal adalah hujan yang sering terjadi di daerah sekitar ekuator akibat pertemuan angin pasat timur laut dengan angin pasat tenggara. Setelah itu, angin tersebut naik dan membentuk gumpalan-gumpalan awan di sekitar ekuator yang berakibat awan menjadi jenuh dan hujan turun.

c. Hujan orografis

Hujan orografis adalah hujan yang terjadi karena angin yang mengandung uap air yang bergerak horisontal. Angin tersebut naik menuju pegunungan, kemudian suhu udara menjadi dingin sehingga terjadi kondensasi.

d. Hujan frontal

Hujan frontal adalah hujan yang terjadi apabila massa udara yang dingin bertemu dengan massa udara yang panas. Tempat pertemuan antara kedua massa tersebut disebut bidang *front*. Karena lebih berat, massa udara dingin lebih berada di bawah. Di sekitar bidang *front* inilah sering terjadi hujan lebat yang disebut hujan frontal.

e. Hujan monsun

Hujan monsun adalah hujan yang terjadi karena angin musim (angin monsun). Penyebab terjadinya angin monsun karena ada pergeseran semu tahunan matahari antara garis balik utara dan garis balik selatan. Di Indonesia, secara teoritis hujan monsun terjadi bulan Oktober sampai April, sedangkan di kawasan Asia Timur terjadi di bulan Mei sampai Agustus.

Berdasarkan pola umum terjadinya, curah hujan di Indonesia dapat dibedakan menjadi tiga tipe, yaitu tipe ekuatorial, tipe monsun, dan tipe lokal (Tukudin, 2010).

a. Tipe ekuatorial

Pola ekuatorial dicirikan oleh tipe curah hujan dengan bentuk bimodal (dua puncak musim hujan) yang biasanya terjadi sekitar bulan Maret dan Oktober atau pada saat terjadi ekuinoks, yaitu waktu atau peristiwa matahari berada dalam bidang katulistiwa bumi dimana peristiwa ini terjadi dua kali dalam setahun.

b. Tipe monsun

Tipe hujan monsun di Indonesia dicirikan oleh adanya perbedaan yang jelas antara periode musim hujan dengan musim kemarau dalam satu tahun. Curah hujan dipengaruhi oleh tiupan angin monsun dan bersifat unimodal (satu puncak musim hujan). Periode Desember-Januari-Februari (DJF) merupakan periode musim hujan, Maret-April-Mei (MAM) merupakan periode transisi dari

musim hujan menuju musim kemarau, Juni-Juli-Agustus (JJA) merupakan periode musim kemarau, dan September-Oktober-November (SON) merupakan periode transisi dari musim kemarau menuju musim hujan.

c. Tipe lokal

Curah hujan dipengaruhi oleh kondisi lingkungan setempat, yakni adanya perairan sebagai sumber penguapan dan pegunungan sebagai daerah tangkapan hujan. Pola curah hujan lokal memiliki distribusi hujan bulanan, kebalikan dengan pola monsun, dicirikan oleh bentuk pola hujan unimodal (satu puncak hujan), tetapi bentuknya berlawanan dengan tipe hujan monsun.

### 2.13 Penelitian Terdahulu

Beberapa penelitian terdahulu yang dijadikan rujukan dan landasan dari penelitian ini adalah sebagai berikut:

1. Boudrissa, Cheraitia, dan Halimi (2017)

Memodelkan data curah hujan harian maksimum tahunan mulai 1936 sampai 2009 menggunakan distribusi GEV di 18 stasiun di Aljazair utara yang dikelompokkan dalam tiga kelas. Setiap kelas dipilih satu stasiun perwakilan, stasiun tersebut adalah Miliana, Algiers, dan Oran. Estimasi parameter menggunakan *maximum likelihood*. Distribusi *Generalized Extreme Value* tipe I atau Gumbel sesuai untuk stasiun Algiers dan Miliana, sementara distribusi GEV tipe Frechet yang paling sesuai untuk stasiun Oran. Kemudian menghitung *return level* untuk beberapa periode ulang (*return period*) yaitu, 2, 20, 50, dan 100 tahun. Sekitar 50 tahun (*return period*  $T=50$ ) harus berlalu untuk mencatat level curah hujan 153 mm per hari di Miliana, 147.9 mm di Algiers, dan 93.01 mm di Oran.

2. Martins *et al.* (2020)

Menerapkan distribusi GP pada data curah hujan di kota Uruguaiana, Brazil. Data yang digunakan mulai Januari 1961 sampai April 2019, kemudian data dikelompokkan dalam periode bulanan. Pemilihan *threshold* menggunakan *mean residual life plot* dan *parameter stability plot*. Estimasi parameter menggunakan MLE. Hasilnya menunjukkan bahwa distribusi GP cocok untuk semua bulan. Estimasi curah hujan maksimum dari Januari sampai Desember dihitung untuk periode ulang 2, 5, 10, 30, 50, dan 100 tahun. Perkiraan curah hujan tertinggi diamati pada bulan April dengan curah hujan diatas 170 mm setiap 100 tahun dan perkiraan curah hujan terendah terjadi di bulan Juli dengan curah hujan dekat dengan 90 mm setiap 100 tahun.

### 3. Musara, Nadarajah, dan Wiegand (2022)

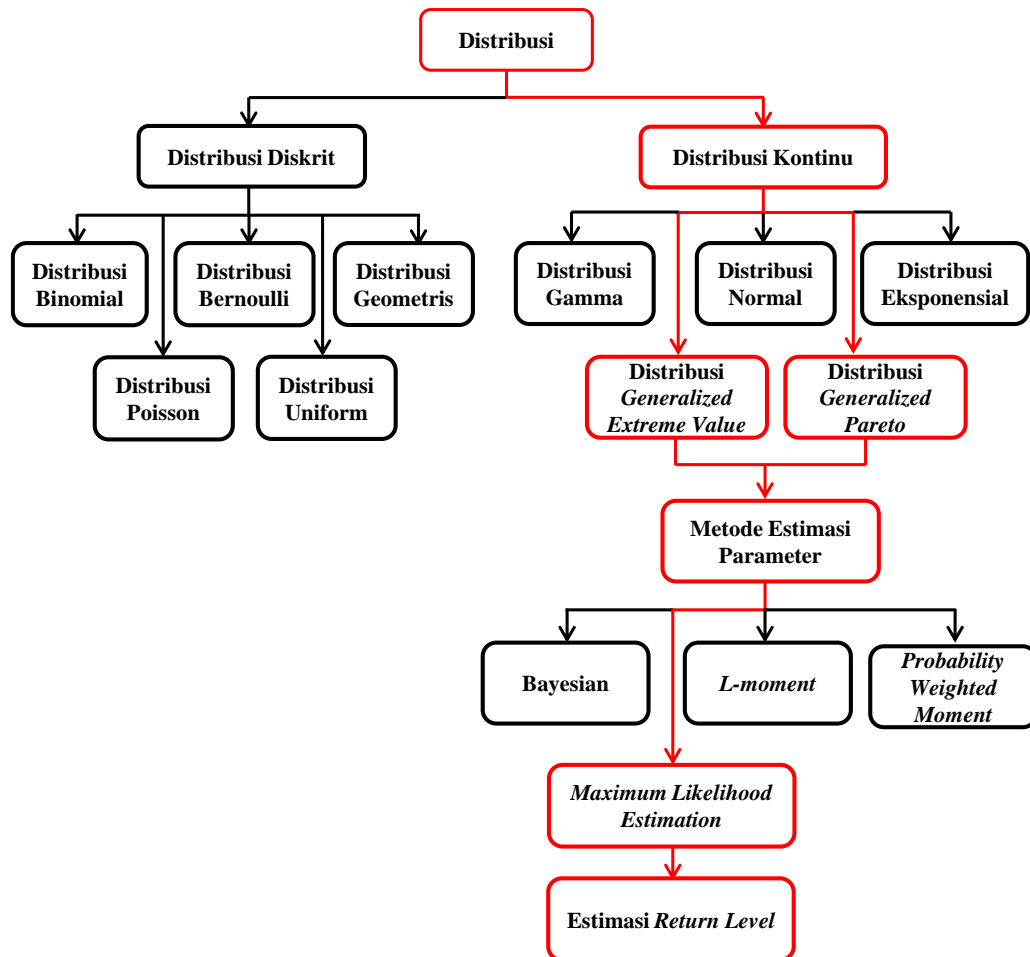
Melakukan analisis curah hujan maksimum di Zimbabwe menggunakan data curah hujan maksimum bulanan dari 103 stasiun. Distribusi GEV terbukti cocok untuk data setiap stasiun dengan melihat *probability plots*, *quantile plots*, dan uji Kolmogorov-Smirnov. Sebagian besar stasiun tidak menunjukkan tren curah hujan yang signifikan. Dua belas stasiun menunjukkan adanya tren negatif dan tiga stasiun menunjukkan tren positif. Kemudian menghitung estimasi *return level* selama 2, 5, 10, 20, 50, dan 100 tahun.

## 2.14 Kerangka Konseptual

Dalam teori peluang dan statistika, distribusi peluang menunjukkan besarnya peluang dari setiap hasil (*outcome*) yang muncul dalam suatu percobaan acak. Distribusi peluang terbagi menjadi dua jenis, yaitu distribusi peluang diskrit dan distribusi peluang kontinu. Distribusi peluang diskrit adalah distribusi peluang terjadinya setiap nilai variabel acak diskrit. Variabel acak adalah transformasi yang memasangkan setiap anggota di ruang sampel ke suatu nilai numerik. Sedangkan variabel acak diskrit adalah variabel acak yang memiliki nilai yang dapat dihitung atau berhingga. Jenis distribusi peluang diskrit adalah distribusi binomial, distribusi hipergeometris, distribusi Bernoulli, distribusi binom negatif, distribusi geometris, distribusi trinom, distribusi uniform (seragam), distribusi Poisson, dan distribusi multinom (Tiro et al., 2008).

Distribusi peluang kontinu adalah distribusi peluang dimana semesta variabel acaknya tak terhingga jumlahnya. Jenis distribusi peluang kontinu adalah distribusi gamma, distribusi normal, distribusi eksponensial, distribusi beta, distribusi chi-kuadrat, distribusi student-t, distribusi Cauchy, distribusi *Generalized Extreme Value* (GEV) dan distribusi *Generalized Pareto* (GP).

Penelitian ini berfokus pada dua distribusi peluang kontinu, yaitu distribusi GEV dan GP. Estimasi parameter distribusi GEV dan GP dapat dilakukan dengan beberapa pendekatan, seperti Bayesian, metode *moment* (*L-moment*, *Probability Weighted Moment*), dan metode *likelihood* (*Maximum Likelihood*). Pada penelitian ini, estimasi parameter distribusi GEV dan GP dilakukan dengan menggunakan metode MLE. Nilai estimasi parameter distribusi GEV dan GP digunakan untuk mengestimasi *return level*. *Return level* merupakan kuantil dari distribusi GEV dan GP. Secara singkat, kerangka konseptual dari penelitian ini dijelaskan pada Gambar 5.



Gambar 5. Bagan kerangka konseptual