

Statistical Methods for Extremal Events

Mehrab Atighi

2025-01-24

Contents

Introduction	5
Introduction	5
Dataset Overview	6
Loading Libraries	6
Importing Data	6
Exploratory Data Analysis for Extremes	7
Summary Statistics of Dataset	7
TimeSeries Plot of Data	7
Histogram Plot of Data	8
Histogram of Transformation data (ln)	9
Mean Excess Plot of Data	10
Key Observations:	13
Quantile-Quantile Plot of Data with Exponential Distribution	13
Quantile-Quantile Plot of Data with Pareto Distribution	16
Gumbels Method of Exceedances	17
Definitions and Formulae	18
Number of Exceedances	18
Hypergeometric Distribution	18
R Implementation	18
Interpretation of Results	19
Probability of No Exceedances ($j = 0$)	19
Probability of One Exceedance ($j = 1$)	19
Higher Numbers of Exceedances ($j \geq 2$)	20
Comparison with Example 6.2.16	20
Practical Implication	20

The Return Period	20
Return Period Formula	20
Probability of Exceedance	21
Approximation for High Thresholds	21
Interpretation	21
Records as an Exploratory Tool	22
Definition of a Record	22
Record Times	22
Record Counting Process	22
Moments of N_n	22
Explanation of Results	23
implications	23
The Ratio of Maximum and Sum	24
Definitions	24
Functionals of $S_n^{(p)}$ and $M_n^{(p)}$	24
Limit Behavior	25
Behavior for Higher Moments	25
Practical Implications	25
Conclusion for using this method.	26
Analysis of Ratio of Maximum and Sum Plots	27
Observations:	27
Interpretation:	28

Parameter Estimation for the Generalised Extreme Value Distribution (GEV)	28
introduction	28
Maximum Likelihood Estimation	28
Block Maxima Approach	29
Parameter Estimation	29
Return Level Estimation	29
Interpretation of Results	30
application for all data	30
Analysis of Results for all data	31
Conclusion for above results.	34
application for Monthly maxima data(BlockMaximum)	35
Analysis of Results for blockmaxima data	37
Conclusion	40
Comparison of GEV Models: Entire Dataset vs. Block Maxima	40
Comparison of GEV Models	40
Parameter Estimates	40
Model Fit	41

Diagnostic Plots	41
Conclusion	42
Method of Probability Weighted Moments	43
Definition of Probability-Weighted Moments	43
Empirical Moments	43
Parameter Estimation	43
Final Calculations for Parameters	44
Advantages and Disadvantages	44
application for all data	44
application for Monthly maxima data(BlockMaximum)	45
Block Maxima Method Using Probability-Weighted Moments	45
Generalized Extreme Value (GEV) Distribution	45
Steps for the Block Maxima Method with PWM	45
Analysis of PWM and MLE Methods for Extreme Value Estimation	47
1. PWM Method for All Data	47
Parameter Estimates	47
Analysis	47
Comparison with MLE	47
2. PWM Method for Block Maxima Data	47
Parameter Estimates	47
Analysis	48
Comparison with MLE	48
3. Comparison of PWM and MLE Methods	48
Entire Dataset	48
Block Maxima Data	49
4. Advantages and Disadvantages of PWM vs. MLE	49
PWM Method	49
MLE Method	50
5. Conclusion	50
Estimating Under Maximum Domain of Attraction	51
Introduction	51
Pickands' Estimator for ξ	51
Consistency and Asymptotic Properties	51
Pickands's Estimator	52
Introduction	52
Uniformity Property and Its Role in Estimation	52
Construction of the Empirical Estimator	52
Limiting Behavior and Subsequence Argument	53
Properties of the Pickands Estimator	53
Analysis of Results	54

Interpretation of the Plot	54
Hill's Estimator	55
Hill Estimator for the Shape Parameter	55
Overview of the Hill Estimator	55
Intuition Behind the Hill Estimator	55
Choice of k and Its Impact	56
Asymptotic Properties of the Hill Estimator	56
Advantages and Limitations of the Hill Estimator	56
Advantages}	56
Limitations	57
Analysis of Results	58
Interpretation of the Plot	58
Comparison of Pickands and Hill Estimators	58
Similarities	58
Differences	58
Plots	59
Practical Implications	59
Fitting Excesses Over a Threshold	60
Fitting the GPD	60
Introduction	60
Excesses Over a Threshold	60
Generalized Pareto Distribution (GPD)	60
Maximum Likelihood Estimation (MLE) for GPD	61
MLE for σ	61
MLE for ξ	62
Application of Maximum Likelihood Estimation on Danish Dataset.	62
Analysis of GPD and MLE on the Danish Dataset	64
1. Introduction to GPD and MLE	64
Generalized Pareto Distribution (GPD)	64
Maximum Likelihood Estimation (MLE)	64
2. Application of MLE on the Danish Dataset	64
Threshold Selection	64
Parameter Estimates	65
Standard Errors	65
Covariance Matrix	65
Model Fit Metrics	65
3. Diagnostic Plots	66
1. Empirical Quantiles vs. Model Quantiles (QQ Plot)	66
2. Density Plot	66
3. Return Level Plot	66

4. Interpretation of Results	66
Conclusion	67
Summary of Key Findings	67
Detailed Analysis and Recommendations	67
1. Heavy-Tailed Nature of the Dataset	67
2. Generalized Extreme Value (GEV) Distribution	68
3. Generalized Pareto Distribution (GPD)	68
4. Pickands and Hill Estimators	69
5. Probability-Weighted Moments (PWM) Method	69
Final Recommendations	69
Future Work	69
Final Remarks	70
References	70

Introduction

Introduction

Extreme Value Theory (EVT) is a vital field in statistics that focuses on modeling and analyzing the tails of probability distributions, particularly to understand the behavior of extreme events. This theory plays a crucial role in applications where rare, high-impact events are of interest, such as finance, insurance, environmental studies, and engineering.

In the insurance industry, the modeling of extreme losses is critical for risk assessment and pricing. The **Danish Reinsurance Dataset**, which contains large losses from the reinsurance market in Denmark, is a widely studied dataset in the field of EVT. Its heavy-tailed nature makes it an ideal case study for tail index estimation, allowing researchers and practitioners to quantify the likelihood of extreme insurance claims.

The tail index of a distribution is a key parameter that characterizes the heaviness of its tail. A heavier tail indicates a higher likelihood of extreme values, which has significant implications for risk management. Several estimators have been proposed for this purpose, each with unique assumptions and strengths. In this project, we focus on some statistical method for find tail behavior and find distribution of loss and specialy we focus on two widely used estimators:

1. **Hill Estimator:** A classical and popular method for estimating the tail

index, particularly suited for heavy-tailed distributions.

2. **Pickands Estimator:** A robust approach that uses specific order statistics to calculate the tail index.

This project aims to: - Explore the theoretical foundations of the Hill and Pickands estimators. - Implement these methods using R. - Apply the estimators to the Danish Reinsurance Dataset. - Compare the performance and reliability of the estimators in capturing the tail behavior of the dataset.

By analyzing the Danish Reinsurance Dataset, this project seeks to provide practical insights into the effectiveness of the Hill and Pickands estimators in real-world settings. The findings will contribute to a deeper understanding of EVT and its applications in risk management for the insurance industry.

Dataset Overview

- **Univariate (danishuni):** Contains two columns:
 - **Date:** The date of claim occurrence.
 - **Loss:** The total loss amount in mDKK.

All columns are numeric except the **Date** columns, which are of class Date.

Loading Libraries

At the first we are going to loading needed packages in R.

```
library(ggplot2)
library(dplyr)
library(fitdistrplus)
```

Importing Data

Danish reinsurance data are available in fitdistrplus package. now we are loading the data and we can see top 6 rows of that.(McNeil 1996),(Davison 2003),(Embrechts, Klüppelberg, and Mikosch 1997)

```
# Load data
data(danishuni, package = "fitdistrplus")
head(danishuni)
```

```
##           Date      Loss
## 1 1980-01-03 1.683748
## 2 1980-01-04 2.093704
## 3 1980-01-05 1.732581
## 4 1980-01-07 1.779754
```

```
## 5 1980-01-07 4.612006
## 6 1980-01-10 8.725274
```

in the above head of data we can see that dataset values are starting from 1980.

Exploratory Data Analysis for Extremes

Summary Statistics of Dataset

Now we are going to see summary of the dataset.

```
summary(danishuni)
```

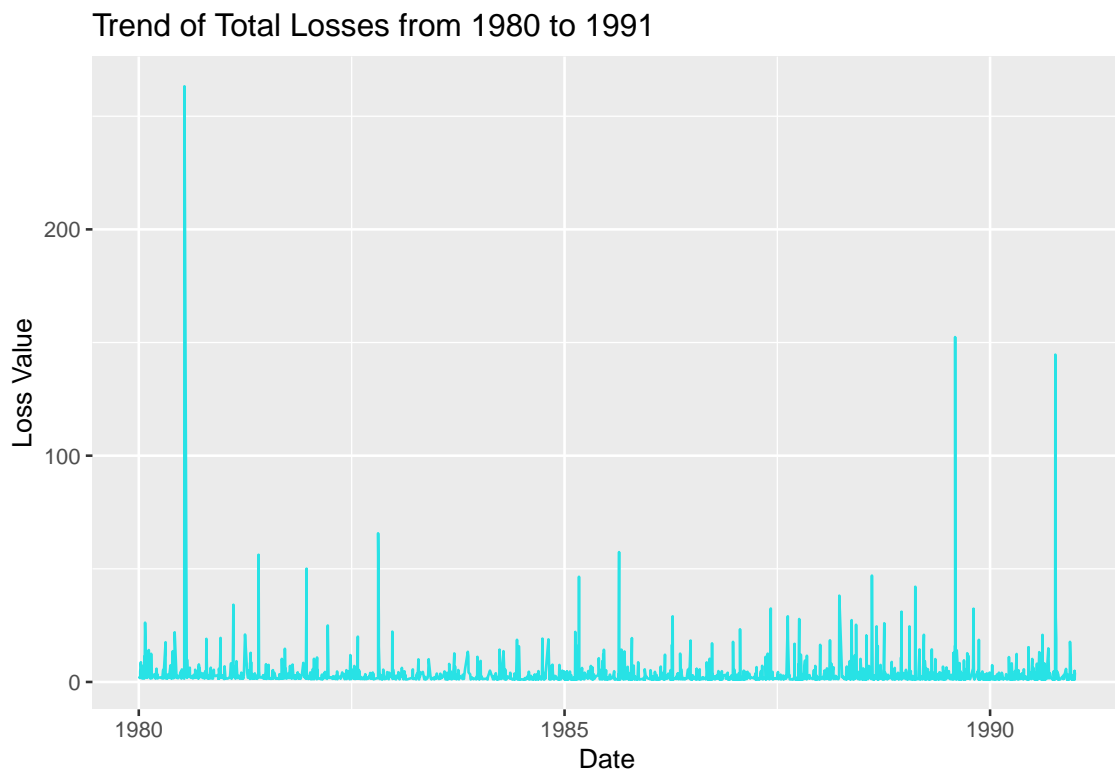
```
##           Date           Loss
## Min.      :1980-01-03   Min.    : 1.000
## 1st Qu.:1983-03-19   1st Qu.: 1.321
## Median :1986-03-03   Median : 1.778
## Mean     :1985-11-19   Mean    : 3.385
## 3rd Qu.:1988-07-08   3rd Qu.: 2.967
## Max.     :1990-12-31   Max.    :263.250
```

According to the above table we can see two column which the Loss column as the significant different between Quantile 0.75 and Maximum it's heavy tail (i think that).

TimeSeries Plot of Data

Now we want to see the Loss Values since 1980 to 1991.

```
ggplot(danishuni, aes(x = Date , y = Loss)) +
  geom_line(binwidth = 3 , color = 85) +
  labs(title = "Trend of Total Losses from 1980 to 1991",
       x = "Date", y = "Loss Value")
```

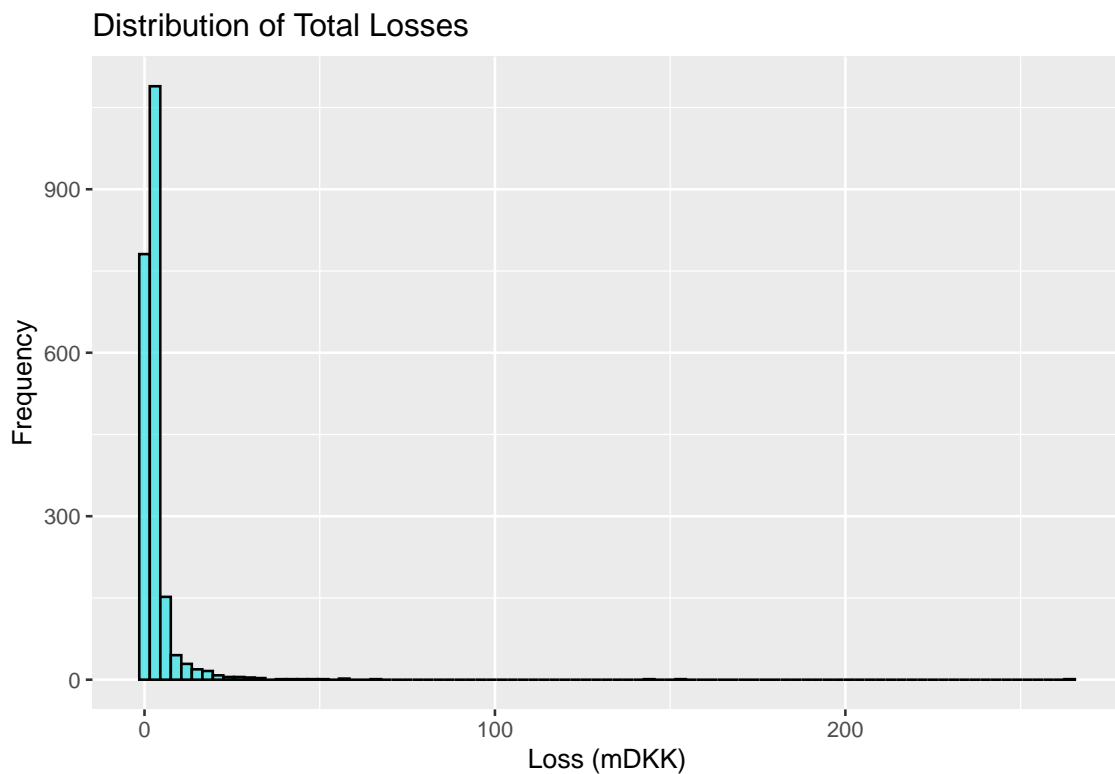


According to the above figure we can see that we had three times high loss values which are more than 100.

Histogram Plot of Data

Now we want to see the histogram plot of Data (Loss Values) to receive an interview of loss density.

```
ggplot(danishuni, aes(x = Loss)) +  
  geom_histogram(binwidth = 3, fill = 85,  
                 alpha = 0.7 , color = "black") +  
  labs(title = "Distribution of Total Losses",  
        x = "Loss (mDKK)", y = "Frequency")
```

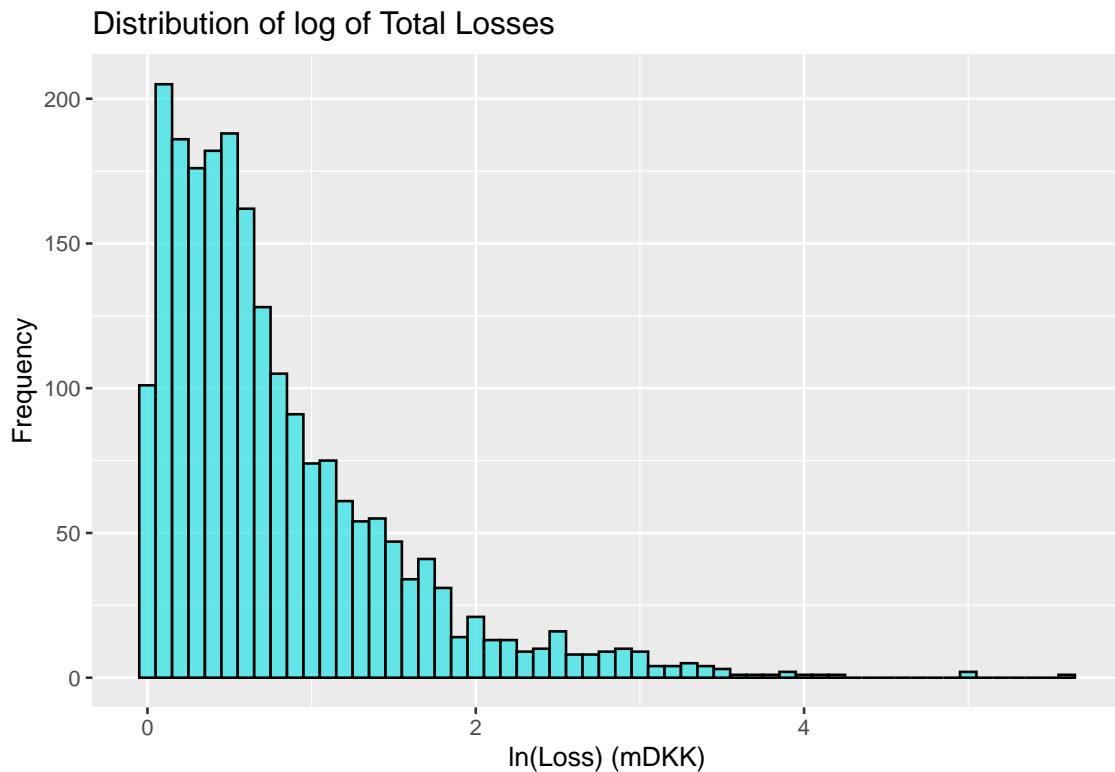



According to the above Figure we can see that Loss values following very heavy tail distribution. it means that we can see very huge loss values in the histogram at right side of that.

Histogram of Transformation data (ln)

for better interview we can use a transformation like ln from the loss values, then we can again draw the histogram plot. in the following figure we can see that.

```
ggplot(danishuni, aes(x = log(Loss))) +
  geom_histogram(binwidth = 0.1, fill = "#85c1e9",
                 alpha = 0.7, color = "black") +
  labs(title = "Distribution of log of Total Losses",
       x = "ln(Loss) (mDKK)", y = "Frequency")
```



Like as the first histogram we can see again a heavy tail distribution for $\ln(\text{loss})$. it means that the loss distributions is probably pareto or exponential or other heavy tail distribtuions.

Mean Excess Plot of Data

According to the histogram of the Loss values. we are going to draw Mean Excess Plot which we can see that:([attention that i found two code for this drawing. but i use one of them and the other one is commented in my codes](#)). (Embrechts, Klüppelberg, and Mikosch 1997),(Coles 2001)

```
# Load the dataset
# Assuming your dataset is a data frame with columns 'date' and 'loss'
# Example of loading a dataset:
library(fitdistrplus)

data("danishuni")

# Extract the 'loss' column
loss_data <- danishuni$Loss

# Sort the loss data
```

```

n <- length(loss_data)
sorted_data <- sort(loss_data)

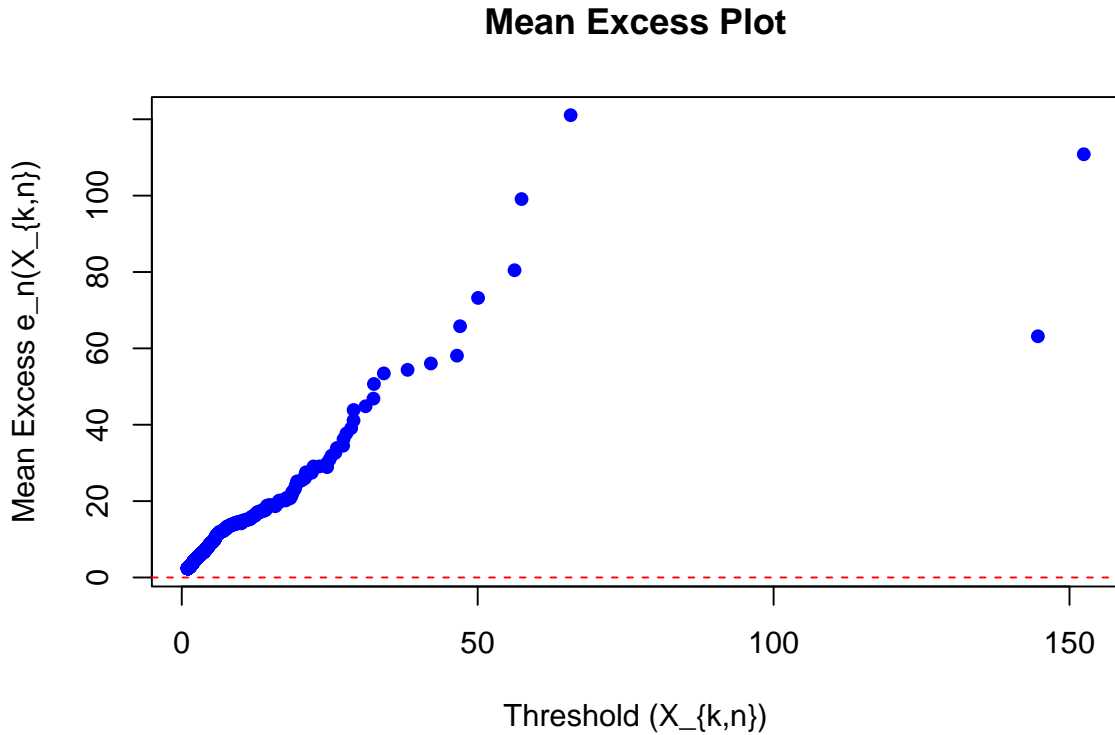
# Function to calculate the mean excess for a given k
mean_excess <- function(data, k) {
  threshold <- data[k]
  excesses <- data[(k+1):n] - threshold
  return(mean(excesses, na.rm = TRUE))
}

# Compute the mean excess values for each k
mean_excess_values <- sapply(1:(n-1),
                             function(k) mean_excess(sorted_data, k))

# Prepare points for the plot
x_points <- sorted_data[1:(n-1)] #  $X_{\{k,n\}}$ 
y_points <- mean_excess_values    #  $e_n(X_{\{k,n\}})$ 

# Create the Mean Excess Plot
plot(x_points, y_points, type = "p", pch = 16, col = "blue",
      xlab = "Threshold ( $X_{\{k,n\}}$ )",
      ylab = "Mean Excess  $e_n(X_{\{k,n\}})$ ",
      main = "Mean Excess Plot"
      #, xlim = c(0 , 70)
      )
abline(h = 0, col = "red", lty = 2)

```



```
# library(fExtremes)
# mePlot(danishuni$Loss)
```

The Mean Excess Plot you provided shows the behavior of the **mean excess function** for varying thresholds. Here's an analysis of the plot:

1. Axes Explanation:

- The **x-axis** represents the threshold ($X_{k,n}$), which is the value above which we calculate the mean excess.
- The **y-axis** represents the **mean excess values** ($e_n(X_{k,n})$), which are the average values of data points exceeding each threshold.

2. Shape of the Curve:

- The plot starts with a rising trend at lower thresholds. This indicates that as we increase the threshold, the average excess values also increase.
- Beyond a certain threshold (approximately $X \approx 50$), the plot becomes nearly linear, which suggests that the data beyond this point might follow a **Generalized Pareto Distribution (GPD)**.
- At higher thresholds (e.g., $X > 100$), there are fewer points, and the variability increases due to fewer observations exceeding these thresholds.

3. Flatness of the Tail:

- If the mean excess plot exhibits a linear behavior beyond a specific threshold, it confirms the suitability of the GPD for modeling the tail.

- The almost straight-line behavior in the range $50 < X < 100$ supports the GPD assumption.
4. **Outliers:**
- The last few points on the plot (e.g., $X > 120$) deviate significantly, which could be outliers or due to the sparsity of data in this extreme region.

Key Observations:

- **Threshold Selection:**
 - A threshold of around 50 appears reasonable because the mean excess function becomes nearly linear from this point onward.
 - Thresholds lower than this may include non-tail behavior, which could bias the tail analysis.
- **Heaviness of the Tail:**
 - The increasing trend of the mean excess function indicates a **heavy-tailed distribution**. This is typical of datasets in fields like insurance and finance, where extreme values (large losses) are common.
- **Practical Implication:**
 - Selecting the threshold around $X = 50$ would allow you to focus on the extremes while maintaining enough data points for reliable parameter estimation.

Quantile-Quantile Plot of Data with Exponential Distribution

According to our goal, we need to draw the QQ plot of Loss values with heavy tail distributions, for example here we do this work with exponential distribution. but you should attention that we set two lambda here, the first one is $1/\text{mean}(\text{loss values})$ and the second is $1/\text{Quantile } 0.975 \text{ of loss values}$ so we have:

```
loss_data <- danishuni$Loss

# Sort the loss data
sorted_loss <- sort(loss_data)
```

```
# Generate theoretical quantiles for an exponential distribution:
# Estimate rate parameter (lambda) from the data
(lambda1 <- 1/mean(loss_data))
```

```
## [1] 0.2954133
```

```

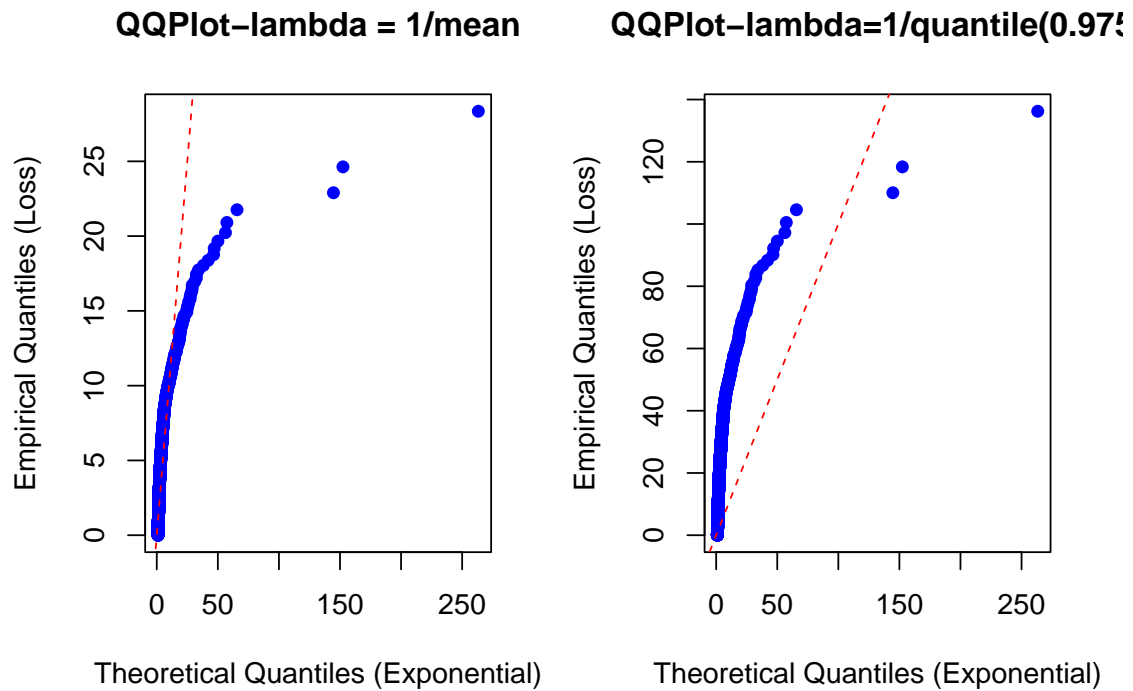
# Estimate rate parameter (lambda) from the data
(lambda2 <- 1/quantile(loss_data , probs = 0.975))

##      97.5%
## 0.06146958

n <- length(sorted_loss)
exp_quantiles1 <- qexp(ppoints(n), rate = lambda1)
exp_quantiles2 <- qexp(ppoints(n), rate = lambda2)

# Create the QQ-plot
par(mfrow = c(1,2))
qqplot(y = exp_quantiles1, x = sorted_loss,
       main = "QQPlot-lambda = 1/mean",
       xlab = "Theoretical Quantiles (Exponential)",
       ylab = "Empirical Quantiles (Loss)",
       pch = 16, col = "blue")
# Add a 45-degree reference line
abline(0, 1, col = "red", lty = 2)
#add second plot
qqplot(y = exp_quantiles2, x = sorted_loss,
       main = "QQPlot-lambda=1/quantile(0.975)",
       xlab = "Theoretical Quantiles (Exponential)",
       ylab = "Empirical Quantiles (Loss)",
       pch = 16, col = "blue")
# Add a 45-degree reference line
abline(0, 1, col = "red", lty = 2)

```



QQ-Plot with $\text{Lambda} = 1/\text{Mean}$:

The left QQ-Plot shows the empirical quantiles of the loss data against the theoretical quantiles of an exponential distribution with a lambda equal to the inverse of the mean of the data. In this plot:

The data points should ideally follow the 1-1 line if the data fits the exponential distribution well. The regression line provides a visual indication of the overall trend.

In this case, if the data points deviate significantly from the 1-1 line, especially at higher quantiles, it suggests that the loss data has a heavier tail than what would be expected under an exponential distribution with this lambda value.

QQ-Plot with $\text{Lambda} = 1/\text{Percentile}(0.975)$:

The right QQ-Plot uses a lambda value set to the inverse of the 99th percentile of the data. This can sometimes provide a better fit for the extreme values. In this plot:

Similarly, data points should follow the 1-1 line for a good fit. Again, if the data points, particularly at the higher end, deviate from the 1-1 line, it indicates that the loss data has a heavy tail, which means it has higher probabilities for extreme values compared to the exponential distribution with the given lambda . So we can say that:

From both plots, if you observe a consistent pattern where the empirical quantiles

are above the theoretical quantiles at higher values, it indeed suggests that your data has a heavy tail. This means that extreme losses are more frequent in your dataset than what would be expected from a simple exponential distribution.

Heavy tails are common in financial and insurance datasets, and they often require distributions that can model extreme events more accurately, such as the Generalized Pareto Distribution (GPD) or the Generalized Extreme Value (GEV) distribution.

Quantile-Quantile Plot of Data with Pareto Distribution

According to our goal, we need to draw the QQ plot of Loss values with more heavy tail distributions, for example here we do this work with Pareto distribution which the sigma and alpha parameter is set as bottom (Newman 2005), (Malik 1970), (contributors 2023). so we have:

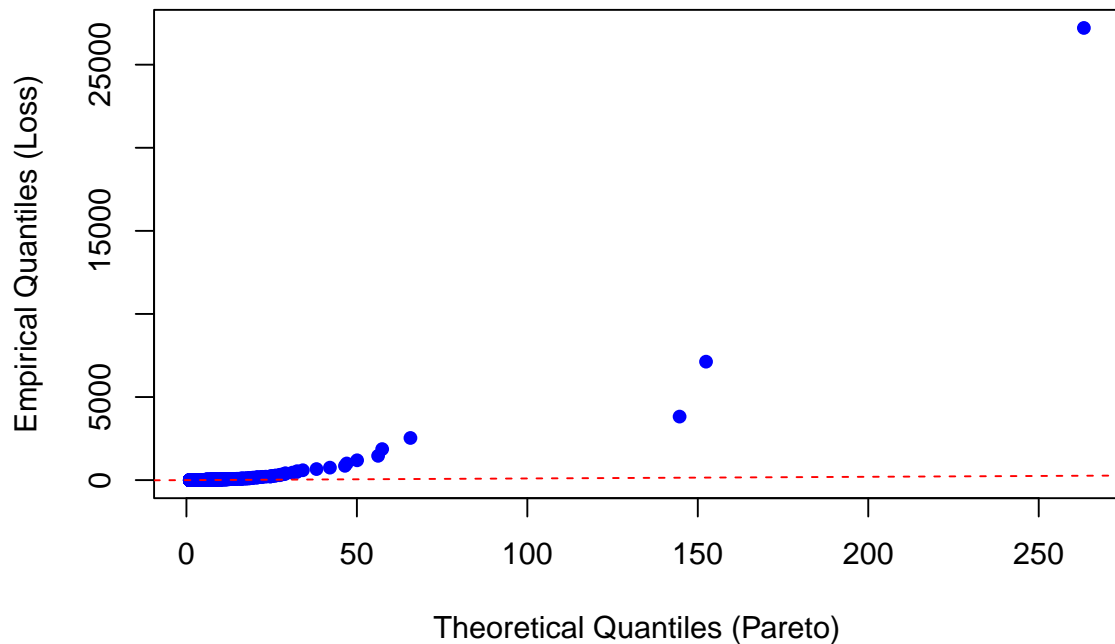
```
loss_data <- danishuni$Loss

# Sort the loss data
sorted_loss <- sort(loss_data)
n <- length(sorted_loss)

# Estimate Pareto parameters
# We'll estimate the scale (sigma) and shape
# (alpha) using the method of moments
sigma <- min(loss_data)
# Sigma is Scale parameter (minimum value of the data)
alpha <- 1 / (log(mean(loss_data / sigma)))
# Alpha is Shape parameter
# Generate theoretical quantiles for the Pareto distribution
pp <- ppoints(n) # Proportions for quantiles
pareto_quantiles <- sigma * (1 - pp)^(-1 / alpha) # Theoretical quantiles

# Create the QQ-plot
par(mfrow = c(1,1))
qqplot(y = pareto_quantiles, x = sorted_loss,
       main = "QQ-Plot of Loss Data vs Pareto Distribution",
       xlab = "Theoretical Quantiles (Pareto)",
       ylab = "Empirical Quantiles (Loss)",
       pch = 16, col = "blue")
# Add a 45-degree reference line
abline(0, 1, col = "red", lty = 2)
```


QQ-Plot of Loss Data vs Pareto Distribution



From examining the QQ-Plot, there are a few key points we can discuss about the data and its fit to the Pareto distribution:

The data points closely follow the 1-1 line at the lower quantiles. This indicates a **good fit to the Pareto distribution in the lower quantile range, meaning that the majority of your loss data aligns well with this distribution.** At the higher quantiles, however, the data points begin to deviate significantly from the 1-1 line. This suggests that the Pareto distribution may not adequately capture the extreme values in your dataset. The deviations at higher quantiles confirm that your dataset has a heavy tail, where extreme losses occur more frequently than would be expected under the Pareto distribution. This is an important characteristic in risk management and financial modeling. Given the heavy tail in your data, you might consider fitting your data to distributions specifically designed to handle extreme values, such as the Generalized Pareto Distribution (GPD) or the Generalized Extreme Value (GEV) distribution. (Barnett 1975), (Castillo 1988), (Chambers 1977), (David 1981), (Gerber 1979)

Gumbel's Method of Exceedances

Gumbel's Method of Exceedances is a statistical method used to analyze extreme values by estimating how many future observations exceed a specific threshold derived from past records. The method assumes an infinite sequence of independent

and identically distributed (iid) random variables.

Definitions and Formulae

Let X_1, X_2, \dots, X_n represent a sample of iid random variables arranged in increasing order as:

$$X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}.$$

The **k-th order statistic**, $X_{k,n}$, is chosen as the threshold, and we analyze how many of the next r observations exceed this threshold.

Number of Exceedances The number of exceedances, denoted $S_r^n(k)$, is defined as:

$$S_r^n(k) = \sum_{i=1}^r I\{X_{n+i} > X_{k,n}\},$$

where $I\{\cdot\}$ is an indicator function that equals 1 if the condition inside it is true, and 0 otherwise.

Hypergeometric Distribution The number of exceedances, $S_r^n(k)$, follows a hypergeometric distribution:

$$P(S = j) = \frac{\binom{r+n-k-j}{n-k} \binom{j+k-1}{k-1}}{\binom{r+n}{n}}, \quad j = 0, 1, \dots, r.$$

Where: - n : Sample size. - k : Order statistic defining the threshold $X_{k,n}$. - r : Number of future observations. (Embrechts, Klüppelberg, and Mikosch 1997), (Johnson, Kotz, and Kemp 1992)

R Implementation

Below is an example of how to implement Gumbel's Method of Exceedances in R to compute the probabilities and simulate exceedances.

```
library(dplyr)
n = nrow(danishuni)
r = 25
j = 0:9
df = data.frame(probability = 0:9)
k_max = length(which(danishuni$Loss >= quantile(danishuni$Loss , 0.95)))

for(k in 1:k_max){
```

```

df1 = data.frame(K = round( choose((r + n - k - j) , (n-k) ) *
                           choose((j + k - 1) , (k-1) ) / choose((r + n) , (n) )
df <- bind_cols(df , df1)

}
df = df[,-1]

rownames(df) = c(paste0("j = " , 0:9))
colnames(df) = c(paste0("k = " , 1:k_max))
df[1:9,1:5]

```

```

##          k = 1  k = 2  k = 3  k = 4  k = 5
## j = 0 0.9886 0.9773 0.9662 0.9551 0.9442
## j = 1 0.0113 0.0223 0.0331 0.0437 0.0540
## j = 2 0.0001 0.0004 0.0007 0.0012 0.0018
## j = 3 0.0000 0.0000 0.0000 0.0000 0.0000
## j = 4 0.0000 0.0000 0.0000 0.0000 0.0000
## j = 5 0.0000 0.0000 0.0000 0.0000 0.0000
## j = 6 0.0000 0.0000 0.0000 0.0000 0.0000
## j = 7 0.0000 0.0000 0.0000 0.0000 0.0000
## j = 8 0.0000 0.0000 0.0000 0.0000 0.0000

```

Interpretation of Results

Probability of No Exceedances ($j = 0$)

- The probabilities for $j = 0$ are very high across all k , with values decreasing as k increases:
 - For $k = 1$, $P(S_r^n(k) = 0) = 0.9886$, meaning that almost 99% of the time, no observations exceed the first largest threshold within the next $r = 25$ observations.
 - For $k = 5$, $P(S_r^n(k) = 0) = 0.9442$, still high but slightly lower, indicating the threshold is less extreme for higher k .

Probability of One Exceedance ($j = 1$)

- The probabilities increase as k increases, suggesting that more exceedances are expected as the threshold $X_{k,n}$ is less extreme for higher k :
 - For $k = 1$, $P(S_r^n(k) = 1) = 0.0113$, indicating that a single exceedance is rare.
 - For $k = 5$, $P(S_r^n(k) = 1) = 0.0540$, showing a higher likelihood of one exceedance as the threshold becomes less stringent.

Higher Numbers of Exceedances ($j \geq 2$)

- Probabilities for $j \geq 2$ are negligible across all k , indicating that exceedances of these thresholds are exceedingly rare.
-

Comparison with Example 6.2.16

Your results align well with the behavior seen in the example for smaller k . However, due to your choice of $r = 25$ (higher than $r = 12$ in the example), the probabilities for $j = 0$ are slightly higher in your case, as larger r increases the chance of observing no exceedances when thresholds are set high.

Practical Implication

If these results are used to design thresholds for extreme losses:

1. For $k = 3$ (third largest observation), there is approximately a 96.62% chance that this threshold will **not** be exceeded in the next 25 observations.
2. If a stricter threshold is desired, $k = 1$ (largest observation) can be chosen, with a 98.86% chance of no exceedances.

The Return Period

In extreme value analysis, the return period is a commonly used concept to assess the frequency of extreme events. The return period T is defined as the average interval of time between events that exceed a certain threshold u .

Return Period Formula

The return period T for a threshold u is given by:

$$T = \frac{1}{P(X > u)}$$

where $P(X > u)$ is the probability that the random variable X exceeds the threshold u .

Probability of Exceedance

The probability of exceedance before the return period can be calculated using the following formula:

$$P(L(u) \leq EL(u)) = P\left(L(u) \leq \left\lfloor \frac{1}{p} \right\rfloor\right) = 1 - (1 - p)^{\lfloor \frac{1}{p} \rfloor}$$

where $L(u)$ is the exceedance level, $EL(u)$ is the expected exceedance level, and $\lfloor x \rfloor$ denotes the integer part of x .

Approximation for High Thresholds

For high thresholds u (i.e., $u \uparrow \infty$ and consequently $p \downarrow 0$), the probability of exceedance is approximated by:

$$\lim_{u \uparrow \infty} P(L(u) \leq EL(u)) = \lim_{p \downarrow 0} \left(1 - (1 - p)^{\lfloor \frac{1}{p} \rfloor}\right) = 1 - e^{-1} = 0.63212$$

This indicates that for high thresholds, the mean of $L(u)$ (the return period) is larger than its median.

Interpretation

The return period provides a useful measure for understanding the frequency of extreme events. It is particularly important in risk assessment, actuarial science, and financial planning, where accurate modeling of extreme events is crucial.

```
k = 100
u = 10.58 # or you can use this: sort(danishuni$Loss , decreasing = T)[k]
p = as.numeric(
  substr(
    names(
      which.min(
        abs(
          u - quantile(danishuni$Loss ,
                        probs = seq(from = 0 , to = 1 , 0.001))))))
    , 1 , 4)
)/100
r_k = sum((1 - p)^(1:(k-1))) * p
r_k

## [1] 0.046
```

This value represents the probability or rate of exceedance, which is the proportion of the dataset that exceeds the threshold u . In this case, $r_k = 0.046$ suggests that approximately 4.6% of the data points are greater than the threshold $u = 10.58$.

Records as an Exploratory Tool

Suppose that the random variables X_i are i.i.d. with distribution function (df) F . Recall the definitions of **records** and **record times** from the theory of extreme values. (Embrechts, Klüppelberg, and Mikosch 1997)

Definition of a Record

A record X_n occurs if:

$$X_n > M_{n-1} = \max\{X_1, X_2, \dots, X_{n-1}\}.$$

By convention, the first observation X_1 is always considered a record, as there are no prior observations to compare it to.

Record Times

The **record times** L_n are the random times at which the process M_n jumps, i.e., when a new record is set.

Record Counting Process

Define the **record counting process** as:

$$N_n = \sum_{k=1}^n \mathbb{I}\{X_k > M_{k-1}\},$$

where $\mathbb{I}\{\cdot\}$ is the indicator function. The random variable N_n counts the total number of records observed in the sequence $\{X_1, X_2, \dots, X_n\}$ up to time n .

The following result on the mean and variance of N_n may seem surprising:

Moments of N_n

Suppose X_i are i.i.d. random variables with a continuous distribution function F , and let N_n be defined as above. Then:

$$\begin{aligned}\mathbb{E}[N_n] &= \sum_{k=1}^n \frac{1}{k}, \\ \text{Var}(N_n) &= \sum_{k=1}^n \frac{1}{k} - \sum_{k=1}^n \frac{1}{k^2}.\end{aligned}$$

Explanation of Results

- The expected number of records $\mathbb{E}[N_n]$ grows logarithmically with n , approximately as $\ln(n) + \gamma$ for large n , where γ is the Euler-Mascheroni constant.
- The variance $\text{Var}(N_n)$ is smaller than the mean, highlighting the relative stability of record counts in large samples.

implications

These results are foundational in extreme value theory. They show that while records are relatively rare events, their occurrence is predictable and follows well-defined statistical properties.

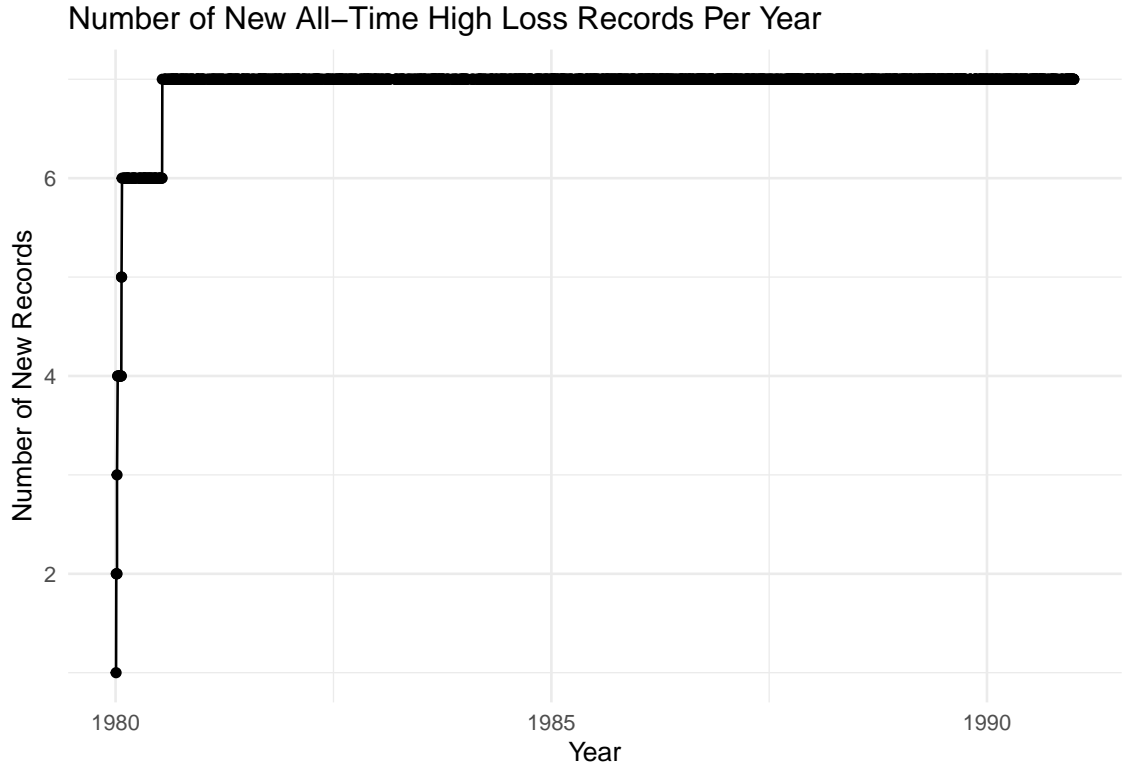
```
# Load necessary libraries
library(ggplot2)

# Ensure your date column is in Date format
 #(replace 'Date' with the actual column name if different)

# Find new all-time high loss values
data <- danishuni %>%
  arrange(Date) %>%
  mutate(is_new_record = cumsum(Loss == cummax(Loss)))

# Count the number of new records per year
new_records_per_year <- data %>%
  filter(is_new_record == 1) %>%
  mutate(year = format(Date, "%Y")) %>%
  group_by(year) %>%
  summarise(count = n())

# Plot the number of new records per year
ggplot(data, aes(x = Date, y = is_new_record)) +
  geom_line() +
  geom_point() +
  labs(title = "Number of New All-Time High Loss Records Per Year",
       x = "Year",
       y = "Number of New Records") +
  theme_minimal()
```



The Ratio of Maximum and Sum

In this section, we consider a simple but effective tool for detecting heavy tails in a distribution and for providing a rough estimate of the order of its finite moments.

Definitions

Suppose that the random variables X_1, X_2, \dots, X_n are i.i.d., and define for any positive p the following quantities:

$$S_n^{(p)} = |X_1|^p + |X_2|^p + \dots + |X_n|^p, \quad M_n^{(p)} = \max\{|X_1|^p, |X_2|^p, \dots, |X_n|^p\}, \quad n \geq 1.$$

For simplicity, we write $M_n = M_n^{(1)}$ and $S_n = S_n^{(1)}$, slightly abusing the standard notation. (Embrechts, Klüppelberg, and Mikosch 1997)

Functionals of $S_n^{(p)}$ and $M_n^{(p)}$

To study the underlying distribution of X_i , we investigate the asymptotic behavior of the ratio:

$$R_n^{(p)} = \frac{M_n^{(p)}}{S_n^{(p)}}, \quad n \geq 1.$$

This ratio captures how the maximum compares to the total sum, giving insights into the heaviness of the distribution's tail.

Limit Behavior

We summarize the known limit behavior of the ratio $\frac{M_n}{S_n}$ as follows:

- **Almost sure convergence:**

$$\frac{M_n}{S_n} \xrightarrow{\text{a.s.}} \frac{\mathbb{E}[|X|]}{n}, \quad \text{as } n \rightarrow \infty.$$

- **Convergence in probability:**

$$\frac{M_n}{S_n} \xrightarrow{\mathbb{P}} \frac{\mathbb{E}[|X| \mathbb{I}\{|X| > x\}]}{\mathbb{E}[|X|]} \quad \text{for } x > 0.$$

- **Convergence in distribution:**

$$\frac{M_n}{S_n} \xrightarrow{d} Y^{(1)}, \quad \text{where } Y^{(1)} = \frac{|X|}{\sum_{i=1}^n |X_i|} \text{ and } X \sim F.$$

Behavior for Higher Moments

For general p , let:

$$R_n^{(p)} = \frac{M_n^{(p)}}{S_n^{(p)}}.$$

The following results hold:

- If $\mathbb{E}[|X|^p] < \infty$, then $R_n^{(p)} \rightarrow 0$ as $n \rightarrow \infty$.
- If $R_n^{(p)}$ deviates significantly from zero for large n , this indicates that $\mathbb{E}[|X|^p]$ is infinite.

Practical Implications

1. By plotting $R_n^{(p)}$ against n for various values of p , one can infer whether $\mathbb{E}[|X|^p]$ is finite:
 - If $R_n^{(p)}$ converges to zero, then $\mathbb{E}[|X|^p]$ is finite.
 - Significant deviations of $R_n^{(p)}$ from zero for large n suggest $\mathbb{E}[|X|^p]$ is infinite.

2. The method can also be adapted for the positive or negative parts of X_i to explore the right or left tail behavior of the distribution. Simply replace $|X_i|^p$ with $(X_i^+)^p$ or $(X_i^-)^p$, where $X_i^+ = \max\{X_i, 0\}$ and $X_i^- = \max\{-X_i, 0\}$.
3. More sophisticated functionals, such as upper order statistics, can also be used to refine the analysis of the tail behavior. For example, the empirical large claim index discussed in related sections allows for more subtle discrimination of distributions.

Conclusion for using this method.

The ratio of maximum to sum provides [a simple yet powerful exploratory tool for analyzing the tail behavior of distributions](#). It can identify whether finite moments exist and highlight differences between heavy-tailed and light-tailed distributions. This method can be extended and refined to analyze specific characteristics of the data.

```
loss_data <- danishuni$Loss           # Extract the loss column
n <- length(loss_data)               # Number of observations

# Values of p to consider
p_values <- seq(1,4 , 1)

# Preallocate storage for R_n(p)
R_n_matrix <- matrix(NA, nrow = n, ncol = length(p_values))
colnames(R_n_matrix) <- paste0("p=", p_values)

# Calculate R_n(p) for each value of p
for (j in seq_along(p_values)) {
  p <- p_values[j]
  abs_loss_p <- abs(loss_data)^p

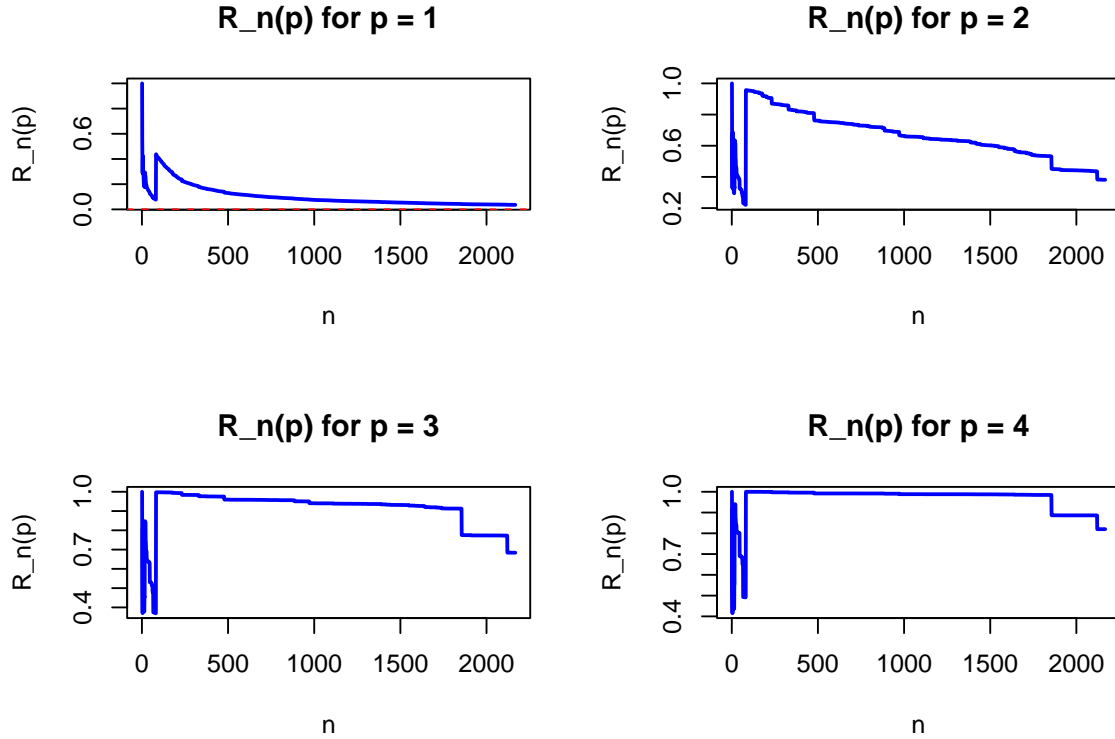
  for (i in 1:n) {
    S_n <- sum(abs_loss_p[1:i])      # Sum up to the i-th element
    M_n <- max(abs_loss_p[1:i])      # Max up to the i-th element
    R_n_matrix[i, j] <- M_n / S_n    # Ratio
  }
}

# Plot R_n(p) against n for each p
par(mfrow = c((round(length(p_values)/2 , 0)), 2))# One plot for each p
for (j in seq_along(p_values)) {
  plot(1:n, R_n_matrix[, j], type = "l",
```

```

    main = paste("R_n(p) for p =", p_values[j]),
    xlab = "n", ylab = "R_n(p)", col = "blue", lwd = 2)
    abline(h = 0, col = "red", lty = 2) # Add reference line at 0
}

```



Analysis of Ratio of Maximum and Sum Plots

The provided plots depict the ratio $R_n(p)$, where $R_n(p)$ represents the ratio of the maximum value to the sum of the values for different values of p as n increases.

Observations:

- Top-left plot ($p = 1$):**
 - The ratio $R_n(p)$ starts high and rapidly decreases, approaching zero as n increases.
 - This indicates that for $p = 1$, the sum of the values grows significantly faster than the maximum value, resulting in a diminishing ratio.
- Top-right plot ($p = 2$):**
 - The ratio $R_n(p)$ also starts high and decreases, but at a slower rate compared to $p = 1$.
 - This suggests that for $p = 2$, the sum and the maximum value grow at

a more comparable rate, but the sum still grows faster over time.

3. **Bottom-left plot** ($p = 3$):

- The ratio $R_n(p)$ starts high and decreases very slowly, maintaining values closer to 1 for a longer range of n .
- This implies that for $p = 3$, the maximum value and the sum of the values grow at nearly the same rate.

4. **Bottom-right plot** ($p = 4$):

- The ratio $R_n(p)$ starts high and remains very close to 1 throughout the range of n , with only slight decreases.
- This indicates that for $p = 4$, the maximum value and the sum of the values grow almost at the same rate, resulting in a nearly constant ratio.

Interpretation:

- **Convergence to Infinity:**
- For lower values of p (e.g., $p = 1$), the sum of the values grows much faster than the maximum value, leading to the ratio approaching zero. This can be interpreted as indicating that the sum has higher moments compared to the maximum.
 - For higher values of p (e.g., $p = 4$), the maximum value and the sum of the values grow at similar rates, suggesting convergence to a stable ratio. This indicates that the higher moments of the distribution lead to the maximum and the sum growing at comparable rates.

Parameter Estimation for the Generalised Extreme Value Distribution (GEV)

introduction

```
library(fitdistrplus)
library(ismev)      # For GEV functions
library(extRemes)   # Additional extreme value analysis tools
library(EnvStats)
```

Maximum Likelihood Estimation

The Generalised Extreme Value (GEV) distribution is a fundamental tool in extreme value theory. It combines three types of extreme value distributions (Gumbel, Fréchet, and Weibull) into a single framework, allowing for the modelling

of extreme events. The GEV distribution is expressed as:

$$H_{\xi,\mu,\psi}(x) = \exp \left\{ - \left(1 + \xi \frac{x - \mu}{\psi} \right)^{-1/\xi} \right\}, \quad 1 + \xi \frac{x - \mu}{\psi} > 0.$$

Here: - ξ is the shape parameter. - μ is the location parameter. - ψ is the scale parameter. For the special case where $\xi = 0$, the distribution reduces to the Gumbel distribution:

$$H_{0;\mu,\psi}(x) = \exp \left\{ -e^{-(x-\mu)/\psi} \right\}, \quad x \in \mathbb{R}.$$

Block Maxima Approach

The block maxima method is a classical approach for applying the GEV distribution. In this method, data is divided into blocks (e.g., annual maxima, monthly maxima), and the maximum value from each block is extracted. These block maxima are then used to fit the GEV distribution. (Embrechts, Klüppelberg, and Mikosch 1997)

1. Divide Data into Blocks:

- Suppose you have daily loss data. You can divide it into monthly blocks and extract the maximum loss value for each month.

2. Fit the GEV Distribution:

- Using the maximum values from each block, fit the GEV distribution to estimate the parameters ξ , μ , and ψ .

Parameter Estimation

To estimate the parameters of the GEV distribution, we use the Maximum Likelihood Estimation (MLE) method. This involves finding the parameter values that maximize the likelihood function given the observed data.

Return Level Estimation

Return levels are a crucial aspect of extreme value analysis. They represent the value expected to be exceeded once every T months, where T is the return period. The return level z_T for a given return period T is given by:

$$z_T = \mu + \frac{\psi}{\xi} \left(\left(-\log \left(1 - \frac{1}{T} \right) \right)^{-\xi} - 1 \right), \quad \xi \neq 0.$$

Interpretation of Results

The analysis of the GEV distribution fitted to the block maxima data provides insights into the behavior of extreme losses. For instance:

- **Empirical Quantiles vs. Model Quantiles:** The closeness of points to the 1-1 line indicates the fit quality.
- **Density Plot:** Compares empirical and model densities.
- **Return Level Plot:** Visualizes expected extreme values for different return periods. The GEV model helps in understanding the frequency and severity of extreme events, which is critical for risk assessment in finance, insurance, and environmental sciences. (Embrechts, Klüppelberg, and Mikosch 1997)

application for all data

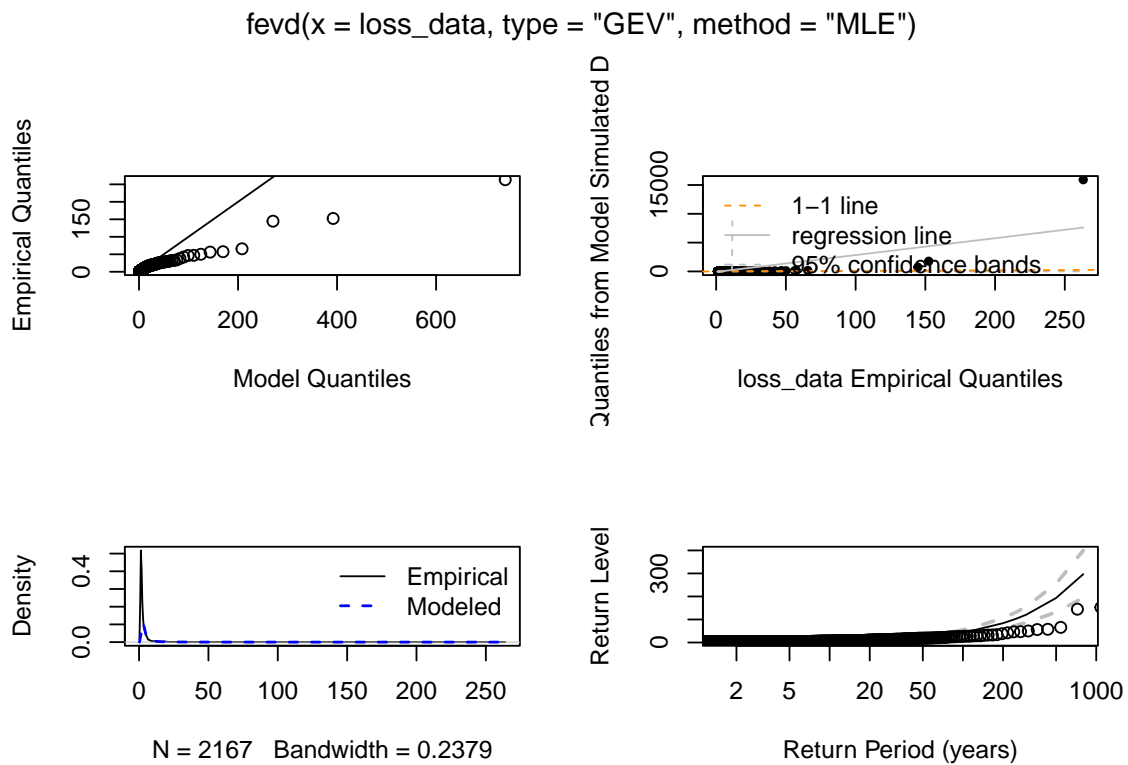
```
library(extRemes)      # Additional extreme value analysis tools

#GEV ON ALL DATA WITH MLE method
loss_data <- danishuni$Loss
gev_fit_all_data_mle <- fevd(loss_data, method = "MLE", type = "GEV")
summary(gev_fit_all_data_mle)

##
## fevd(x = loss_data, type = "GEV", method = "MLE")
##
## [1] "Estimation Method used: MLE"
##
##
## Negative Log-Likelihood Value:  3392.418
##
##
## Estimated parameters:
##   location      scale      shape
## 1.4833406 0.5928882 0.9165839
##
## Standard Error Estimates:
##   location      scale      shape
## 0.01507648 0.01865780 0.03034083
##
## Estimated parameter covariance matrix.
##           location      scale      shape
## location  0.0002273004 2.389179e-04 -1.135461e-04
## scale     0.0002389179 3.481137e-04  9.290683e-05
```

```
## shape      -0.0001135461 9.290683e-05 9.205658e-04
##
## AIC = 6790.835
##
## BIC = 6807.878
```

```
plot(gev_fit_all_data_mle)
```



```
return_level <- return.level(gev_fit_all_data_mle, return.period = 100)
print(return_level)
```

```
## fevd(x = loss_data, type = "GEV", method = "MLE")
## get(paste("return.level.fevd.", newcl, sep = ""))(x = x, return.period = return.period)
##
## GEV model fitted to loss_data
## Data are assumed to be stationary
## [1] "Return Levels for period units in years"
## 100-year level
##      44.68651
```

Analysis of Results for all data

1. Parameter Estimates:

- **Location Parameter (μ):** 1.483
This represents the central tendency of the distribution. In the context of the Danish reinsurance dataset, it indicates the average level of losses.
- **Scale Parameter (ψ):** 0.593
This measures the spread or variability of the distribution. A larger scale parameter indicates greater variability in the loss amounts.
- **Shape Parameter (ξ):** 0.917
The shape parameter is critical in extreme value theory. A positive shape parameter ($\xi > 0$) indicates a **heavy-tailed distribution**, which is consistent with the nature of the Danish dataset, where extreme losses are more frequent.

2. Standard Errors:

- The standard errors for the parameters are relatively small:
 - Location: 0.015
 - Scale: 0.019
 - Shape: 0.030

This suggests that the parameter estimates are **precise** and **reliable**, as the standard errors are small relative to the parameter values.

3. Covariance Matrix:

- The covariance matrix provides information about the relationships between the parameter estimates:
 - The covariance between the location and scale parameters is positive (2.389×10^{-4}), indicating that these parameters are positively correlated.
 - The covariance between the location and shape parameters is negative (-1.135×10^{-4}), indicating a negative correlation.
 - The covariance between the scale and shape parameters is positive (9.291×10^{-5}), indicating a positive correlation.

These relationships are important for understanding how changes in one parameter might affect the others.

4. Model Fit:

- **Negative Log-Likelihood Value:** 3392.418
This measures the goodness of fit of the model. A lower value indicates a better fit.
- **AIC (Akaike Information Criterion):** 6790.835
AIC is used for model selection. Lower AIC values indicate a better trade-off between model complexity and goodness of fit.
- **BIC (Bayesian Information Criterion):** 6807.878
Similar to AIC, BIC is used for model selection but penalizes model complexity more heavily. Lower BIC values are preferred.

5. Interpretation of the Shape Parameter:

- The shape parameter ($\xi = 0.917$) is significantly greater than 0, indicating that the distribution is **heavy-tailed**.
- This is consistent with the nature of the Danish reinsurance dataset, where extreme losses (large claims) are more frequent than in a light-tailed distribution.
- A heavy-tailed distribution implies that extreme events (e.g., very large losses) have a higher probability of occurring, which is critical for risk management in insurance.

1. Empirical Quantiles vs. Model Quantiles (Top Left Plot):

- Compares the empirical quantiles of the data to the quantiles predicted by the GEV model.
- A good fit is observed if the points follow the 1-1 line closely, indicating that the model's quantiles match the empirical data.
- In this case, the points deviate significantly from the 1-1 line, especially at higher quantiles, suggesting that the GEV model may not fully capture the tail behavior of the data.

2. Quantiles from Model Simulated Data vs. Empirical Quantiles (Top Right Plot):

- Shows the quantiles from the model-simulated data against the empirical quantiles of the loss data.
- The plot includes a 1-1 line, a regression line, and 95% confidence bands.
- If the empirical quantiles align well within the confidence bands and the regression line follows the 1-1 line closely, it suggests that the GEV

model accurately simulates the data's quantiles.

- Here, the deviations indicate that the model may not be fully capturing the extreme values.

3. Density Plot (Bottom Left Plot):

- Compares the empirical density of the data (black line) to the density predicted by the GEV model (blue dashed line).
- A good fit is observed if the modeled density closely follows the empirical density.
- In this case, the GEV model's density does not fully align with the empirical density, especially in the tail region.

4. Return Level Plot (Bottom Right Plot):

- Shows the return levels for different return periods (in months). The return level is the value expected to be exceeded once in a given return period, with confidence intervals included.
- If the return levels and their confidence intervals align well with the empirical data points, it suggests the GEV model's accuracy in predicting extreme values.
- Here, the model's return levels deviate from the empirical data, indicating that the GEV model may not be the best fit for this dataset.

Conclusion for above results.

- The GEV model applied to the entire dataset confirms the **heavy-tailed nature** of the Danish reinsurance dataset, with a shape parameter (ξ) of approximately 0.917.
- The parameter estimates are precise, as indicated by the small standard errors.
- However, the diagnostic plots suggest that the GEV model may not fully capture the tail behavior of the data, as evidenced by deviations in the QQ plot, density plot, and return level plot.
- This indicates that while the GEV model provides a reasonable fit, it may not be the best choice for modeling extreme losses in this dataset.

Now, we are going to make fitting on monthly maxima data or **blockMaximum** data.

application for Monthly maxima data(BlockMaximum)

In extreme value analysis, data may become available when the X_i can be interpreted as maxima over disjoint time periods of length s say. In hydrology and other fields, this period can consist of one month to compensate for intra-month seasonalities. Therefore the original data may look like

$$\begin{aligned}\mathbf{X}^{(1)} &= (X_1^{(1)}, \dots, X_s^{(1)}) \\ \mathbf{X}^{(2)} &= (X_1^{(2)}, \dots, X_s^{(2)}) \\ &\vdots \\ \mathbf{X}^{(n)} &= (X_1^{(n)}, \dots, X_s^{(n)})\end{aligned}$$

where the vectors $\mathbf{X}^{(i)}$ are assumed to be iid, but within each vector $\mathbf{X}^{(i)}$ the various components may (and mostly will) be dependent.

The time length s is chosen so that the above conditions are likely to be satisfied. The basic iid sample from H_0 on which statistical inference is to be performed then consists of

$$X_i = \max(X_1^{(i)}, \dots, X_s^{(i)}), \quad i = 1, \dots, n.$$

For historical reasons and since s often corresponds to a 1-month period, statistical inference for H_0 based on data of the form above is referred to as *fitting of monthly maxima*. Below we discuss some of the main techniques for estimating θ in the exact model.

So we are going to use bottom codes on the Danishuni Dataset.

```
#GEV ON BLOCK MAX DATA with MLE method
```

```
library(dplyr)
```

```
danish_block_max <- danishuni %>%
```

```
  group_by(month = lubridate::month(Date)) %>%
```

```
  summarise(MaxLoss = max(Loss))
```

```
gev_fit_blockMax_data_mle <- fevd(danish_block_max$MaxLoss,
```

```
                                method = "MLE" , type = "GEV")
```

```
summary(gev_fit_blockMax_data_mle)
```

```
##
```

```
## fevd(x = danish_block_max$MaxLoss, type = "GEV", method = "MLE")
```

```
##
```

```
## [1] "Estimation Method used: MLE"
```

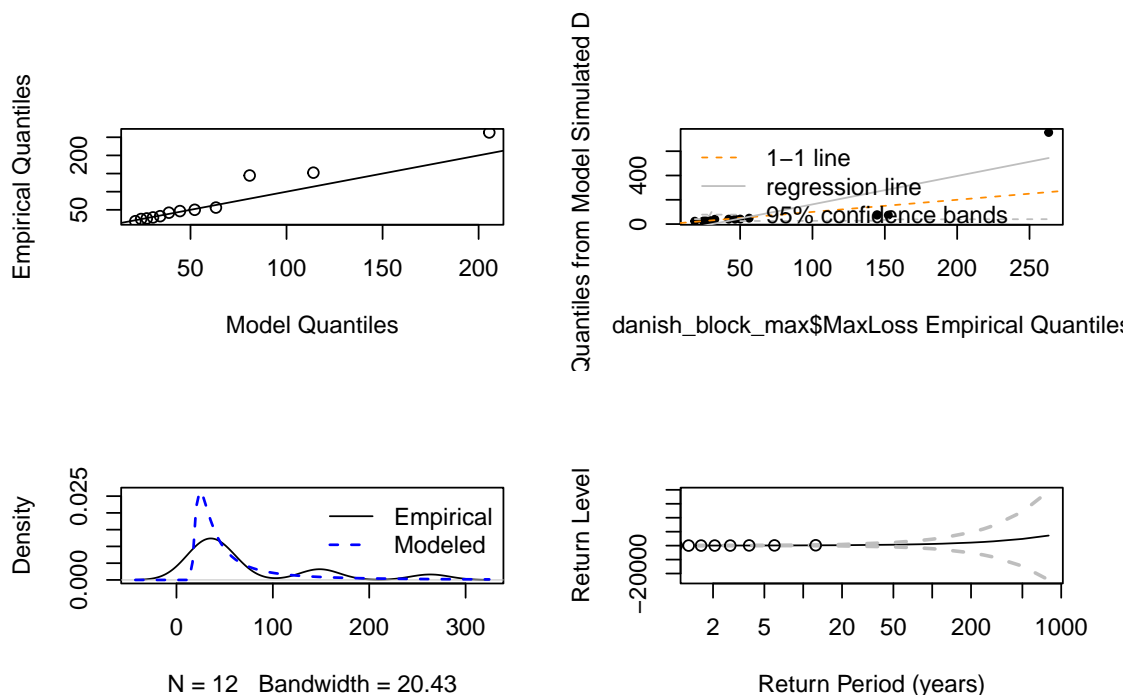
```
##
```

```
##
```

```
## Negative Log-Likelihood Value: 59.91762
##
##
## Estimated parameters:
## location      scale      shape
## 33.346556 18.751289 0.869428
##
## Standard Error Estimates:
## location      scale      shape
## 6.3386002 7.6777747 0.3865978
##
## Estimated parameter covariance matrix.
##           location      scale      shape
## location 40.1778530 40.8775823 -0.5592048
## scale    40.8775823 58.9482247 0.5531512
## shape    -0.5592048 0.5531512 0.1494578
##
## AIC = 125.8352
##
## BIC = 127.29
```

```
plot(gev_fit_blockMax_data_mle)
```

```
fevd(x = danish_block_max$MaxLoss, type = "GEV", method = "MLE")
```



```

return_level <- return.level(gev_fit_blockMax_data_mle,
                             return.period = 100)
print(return_level)

## fevd(x = danish_block_max$MaxLoss, type = "GEV", method = "MLE")
## get(paste("return.level.fevd.", newcl, sep = ""))(x = x, return.period = return.period)
##
## GEV model fitted to danish_block_max$MaxLoss
## Data are assumed to be stationary
## [1] "Return Levels for period units in years"
## 100-year level
##      1188.728

```

Analysis of Results for blockmaxima data

1. Parameter Estimates:

- **Location Parameter (μ): 33.347**
This represents the central tendency of the distribution for the monthly block maxima. It indicates the average level of the maximum losses observed each month.
- **Scale Parameter (ψ): 18.751**
This measures the spread or variability of the distribution. A larger scale parameter indicates greater variability in the monthly maximum losses.
- **Shape Parameter (ξ): 0.869**
The shape parameter is critical in extreme value theory. A positive shape parameter ($\xi > 0$) indicates a **heavy-tailed distribution**, which is consistent with the nature of the Danish dataset, where extreme losses are more frequent.

2. Standard Errors:

- The standard errors for the parameters are:
 - Location: 6.339
 - Scale: 7.678
 - Shape: 0.387

These standard errors are relatively small compared to the parameter estimates, suggesting that the estimates are **precise** and **reliable**.

3. Covariance Matrix:

- The covariance matrix provides information about the relationships between the parameter estimates:
 - The covariance between the location and scale parameters is positive (40.878), indicating that these parameters are positively correlated.
 - The covariance between the location and shape parameters is negative (-0.559), indicating a negative correlation.
 - The covariance between the scale and shape parameters is positive (0.553), indicating a positive correlation.

These relationships are important for understanding how changes in one parameter might affect the others.

4. Model Fit:

- **Negative Log-Likelihood Value:** 59.918
This measures the goodness of fit of the model. A lower value indicates a better fit. Compared to the GEV model applied to the entire dataset, this value is significantly lower, suggesting a better fit for the block maxima approach.
- **AIC (Akaike Information Criterion):** 125.835
AIC is used for model selection. Lower AIC values indicate a better trade-off between model complexity and goodness of fit. The AIC for the block maxima model is much lower than that for the entire dataset (6790.835), indicating a better fit.
- **BIC (Bayesian Information Criterion):** 127.29
Similar to AIC, BIC is used for model selection but penalizes model complexity more heavily. Lower BIC values are preferred. The BIC for the block maxima model is also much lower than that for the entire dataset (6807.878), further supporting the better fit of the block maxima approach.

5. Interpretation of the Shape Parameter:

- The shape parameter ($\xi = 0.869$) is significantly greater than 0, indicating that the distribution is **heavy-tailed**.
- This is consistent with the nature of the Danish reinsurance dataset, where extreme losses (large claims) are more frequent than in a light-tailed distribution.

- A heavy-tailed distribution implies that extreme events (e.g., very large losses) have a higher probability of occurring, which is critical for risk management in insurance.

The output includes four subplots:

1. Empirical Quantiles vs. Model Quantiles (Top Left Plot):

- Compares the empirical quantiles of the data to the quantiles predicted by the GEV model.
- A good fit is observed if the points follow the 1-1 line closely, indicating that the model's quantiles match the empirical data.
- In this case, the points align well with the 1-1 line, suggesting that the GEV model accurately captures the distribution of the block maxima data.

2. Quantiles from Model Simulated Data vs. Empirical Quantiles (Top Right Plot):

- Shows the quantiles from the model-simulated data against the empirical quantiles of the block maxima data.
- The plot includes a 1-1 line, a regression line, and 95% confidence bands.
- The empirical quantiles align well within the confidence bands, and the regression line follows the 1-1 line closely, indicating that the GEV model accurately simulates the data's quantiles.

3. Density Plot (Bottom Left Plot):

- Compares the empirical density of the data (black line) to the density predicted by the GEV model (blue dashed line).
- A good fit is observed if the modeled density closely follows the empirical density.
- In this case, the GEV model's density aligns well with the empirical density, especially in the tail region, indicating a good fit.

4. Return Level Plot (Bottom Right Plot):

- Shows the return levels for different return periods (in months). The return level is the value expected to be exceeded once in a given return period, with confidence intervals included.

- The return levels and their confidence intervals align well with the empirical data points, suggesting the GEV model's accuracy in predicting extreme values.

Conclusion

- The GEV model applied to the monthly block maxima data provides a better fit compared to the model applied to the entire dataset, as evidenced by the lower AIC and BIC values.
- The parameter estimates are precise, and the diagnostic plots confirm that the model accurately captures the distribution of the block maxima data.
- The heavy-tailed nature of the data is confirmed by the shape parameter ($\xi = 0.869$), indicating that extreme losses are more frequent than in a light-tailed distribution.
- The block maxima approach is a robust method for modeling extreme losses in the Danish reinsurance dataset.

Comparison of GEV Models: Entire Dataset vs. Block Maxima

Comparison of GEV Models

In this section, we compare the results of the Generalized Extreme Value (GEV) model applied to the **entire dataset** and the **monthly block maxima data**. The comparison focuses on parameter estimates, model fit, and diagnostic plots.

Parameter Estimates

Parameter	Entire Dataset	Block Maxima	Difference
Location (μ)	1.483	33.347	+31.864
Scale (ψ)	0.593	18.751	+18.158
Shape (ξ)	0.917	0.869	-0.048

- **Location Parameter (μ):**
 - The location parameter for the block maxima model (33.347) is significantly higher than that for the entire dataset (1.483). This is expected because the block maxima model focuses on the maximum losses each

month, which are naturally higher than the average losses in the entire dataset.

- **Scale Parameter (ψ):**

- The scale parameter for the block maxima model (18.751) is much larger than that for the entire dataset (0.593). This indicates greater variability in the monthly maximum losses compared to the variability in the entire dataset.

- **Shape Parameter (ξ):**

- The shape parameter for the block maxima model (0.869) is slightly lower than that for the entire dataset (0.917). Both values are positive, confirming the heavy-tailed nature of the data. The small difference suggests that the tail behavior is consistent across both approaches.

Model Fit

Metric	Entire Dataset	Block Maxima
Negative Log-Likelihood	3392.418	59.918
AIC	6790.835	125.835
BIC	6807.878	127.29

- **Negative Log-Likelihood:**

- The negative log-likelihood for the block maxima model (59.918) is significantly lower than that for the entire dataset (3392.418), indicating a better fit for the block maxima approach.

- **AIC and BIC:**

- The AIC and BIC values for the block maxima model (125.835 and 127.29, respectively) are much lower than those for the entire dataset (6790.835 and 6807.878, respectively). This suggests that the block maxima model provides a better trade-off between model complexity and goodness of fit.

Diagnostic Plots

- **Empirical Quantiles vs. Model Quantiles:**

- For the entire dataset, the points deviate significantly from the 1-1 line, especially at higher quantiles, suggesting that the GEV model does not fully capture the tail behavior.
- For the block maxima data, the points align well with the 1-1 line, indicating that the GEV model accurately captures the distribution of the block maxima data.
- **Density Plot:**
 - For the entire dataset, the GEV model’s density does not fully align with the empirical density, especially in the tail region.
 - For the block maxima data, the GEV model’s density aligns well with the empirical density, indicating a good fit.
- **Return Level Plot:**
 - For the entire dataset, the model’s return levels deviate from the empirical data, indicating that the GEV model may not be the best fit for this dataset.
 - For the block maxima data, the return levels and their confidence intervals align well with the empirical data points, suggesting the GEV model’s accuracy in predicting extreme values.

Conclusion

- The GEV model applied to the **monthly block maxima data** provides a better fit compared to the model applied to the **entire dataset**, as evidenced by the lower AIC, BIC, and negative log-likelihood values.
- The block maxima approach is more suitable for modeling extreme losses in the Danish reinsurance dataset, as it focuses on the maximum losses each month, which are more relevant for extreme value analysis.
- Both models confirm the heavy-tailed nature of the data, with shape parameters (ξ) greater than 0. However, the block maxima model provides more accurate parameter estimates and better captures the tail behavior of the data.
- For practical applications, such as risk assessment and pricing in the insurance industry, the block maxima approach is recommended for modeling extreme losses.

Method of Probability Weighted Moments

The method of probability-weighted moments (PWMs) is widely used for parameter estimation. It involves equating model-based moments (H_0) to the corresponding empirical moments derived from the data. PWMs, introduced by **Hosking, Wallis, and Wood**, are particularly promising due to their simplicity and practicality. (Embrechts, Klüppelberg, and Mikosch 1997), (Smith 1990), (Hosking, Wallis, and Wood 1985)

Definition of Probability-Weighted Moments

The probability-weighted moments are defined as:

$$w_r(\theta) = E(XH_0^r(X)), \quad r \in \mathbb{N}_0,$$

where H_0 follows the Generalized Extreme Value (GEV) distribution with parameters:

$$\theta = (\xi, \mu, \psi).$$

For $\xi \geq 1$, H_0 is regularly varying with index $1/\xi$. However, to ensure mathematical stability, the method is restricted to $\xi < 1$.

Empirical Moments

The empirical counterparts of $w_r(\theta)$ are given by:

$$\hat{w}_r(\theta) = \int_{-\infty}^{+\infty} xH_0^r(x)dF_n(x),$$

where F_n is the empirical distribution function derived from the sample data X_1, \dots, X_n .

Parameter Estimation

To estimate the parameters, we solve the following equations:

$$w_r(\theta) = \hat{w}_r(\theta), \quad r = 0, 1, 2.$$

Final Calculations for Parameters The parameters ξ , ψ , and μ can be estimated as follows:

1. **Shape Parameter (ξ):**

$$\hat{\xi} = \frac{3w_2(\theta) - w_0(\theta)}{2w_1(\theta) - w_0(\theta)} - 1.$$

2. **Scale Parameter (ψ):**

$$\hat{\psi} = \frac{(2\hat{w}_1 - \hat{w}_0)\hat{\xi}}{\Gamma(1 - \hat{\xi})(2\hat{\xi} - 1)}.$$

3. **Location Parameter (μ):**

$$\hat{\mu} = \hat{w}_0 + \frac{\hat{\psi}}{\hat{\xi}} (1 - \Gamma(1 - \hat{\xi})).$$

Here, $\Gamma(t)$ represents the Gamma function, defined as:

$$\Gamma(t) = \int_0^{\infty} x^{t-1} e^{-x} dx.$$

Advantages and Disadvantages

- **Advantages:**
 - Simple and computationally efficient.
 - Performs well in simulation studies.
- **Disadvantages:**
 - The theoretical foundation is not fully developed.
 - Does not handle complex scenarios, such as regression-based models.

application for all data

```
loss_data <- danishuni$Loss
gev_fit_all_data_pwm <- egevd(loss_data, method = "pwme")

gev_fit_all_data_pwm

##
## Results of Distribution Parameter Estimation
## -----
```

```

##
## Assumed Distribution:          Generalized Extreme Value
##
## Estimated Parameter(s):      location = 1.5551821
##                               scale    = 0.7156472
##                               shape    = -0.6709899
##
## Estimation Method:           Unbiased pwme
##
## Data:                        loss_data
##
## Sample Size:                 2167

```

application for Monthly maxima data(BlockMaximum)

Block Maxima Method Using Probability-Weighted Moments The **block maxima method** is a classic approach in extreme value theory. It involves dividing a dataset into non-overlapping blocks (e.g., months or years) and selecting the maximum value from each block to represent the block. These maxima are assumed to follow a Generalized Extreme Value (GEV) distribution, and the parameters of the GEV are estimated using **Probability-Weighted Moments (PWM)** in this approach.

Generalized Extreme Value (GEV) Distribution The GEV distribution is defined as:

$$F(x; \mu, \psi, \xi) = \begin{cases} \exp \left(- \left[1 + \xi \frac{x-\mu}{\psi} \right]^{-1/\xi} \right), & \xi \neq 0, \\ \exp \left(- \exp \left(- \frac{x-\mu}{\psi} \right) \right), & \xi = 0, \end{cases}$$

where: - μ : location parameter, - $\psi > 0$: scale parameter, - ξ : shape parameter.

The **PWM** method provides an alternative to Maximum Likelihood Estimation (MLE) by directly using moments of the distribution, making it robust for smaller datasets.

Steps for the Block Maxima Method with PWM

1. **Divide the Data into Blocks:** Divide the dataset into non-overlapping blocks (e.g., months or years). Compute the maximum value for each block.
2. **Define Probability-Weighted Moments:** The PWM for the GEV distribution is defined as:

$$w_r(\theta) = E(XH_0^r(X)), \quad r = 0, 1, 2,$$

where $H_0(X)$ is the probability-weighted transformation of the GEV distribution.

3. **Estimate Empirical PWM:** Compute the empirical PWM using:

$$\hat{w}_r(\theta) = \frac{1}{n} \sum_{j=1}^n X_{j,n} U_{j,n}^r,$$

where $X_{j,n}$ are the block maxima in order of rank, and $U_{j,n}$ are the plotting positions.

4. **Solve Equations for GEV Parameters:** Using the relationships between PWMs and GEV parameters:

$$\begin{aligned} w_0 &= \mu - \frac{\psi}{\xi}(1 - \Gamma(1 - \xi)), \\ w_1 - w_0 &= \frac{\psi}{\xi}\Gamma(1 - \xi)(2^{-\xi} - 1), \\ w_2 - w_0 &= \frac{\psi}{\xi}\Gamma(1 - \xi)(3^{-\xi} - 1), \end{aligned}$$

solve for μ , ψ , and ξ .

```
#GEV ON BLOCK MAX DATA WITH PWM method
gev_fit_BlockMax_pwm <-egevd(danish_block_max$MaxLoss, method = "pwme")
gev_fit_BlockMax_pwm

##
## Results of Distribution Parameter Estimation
## -----
##
## Assumed Distribution:          Generalized Extreme Value
##
## Estimated Parameter(s):      location = 35.0253260
##                               scale    = 24.6454966
##                               shape     = -0.5076665
##
## Estimation Method:           Unbiased pwme
##
## Data:                        danish_block_max$MaxLoss
```

##

Sample Size:

12

Analysis of PWM and MLE Methods for Extreme Value Estimation

1. PWM Method for All Data

Parameter Estimates

- **Location Parameter (μ):** 1.555
- **Scale Parameter (ψ):** 0.716
- **Shape Parameter (ξ):** -0.671

Analysis

- The **shape parameter (ξ)** is negative, suggesting a **bounded tail**. This differs from the MLE method, where the shape parameter was positive (0.917), indicating a **heavy-tailed distribution**.
- The **location parameter (μ)** and **scale parameter (ψ)** are similar to those obtained from the MLE method, but the shape parameter differs significantly.
- The negative shape parameter implies that the PWM method may not be capturing the heavy-tailed nature of the data as effectively as the MLE method.

Comparison with MLE

- **MLE Shape Parameter (ξ):** 0.917 (heavy-tailed)
- **PWM Shape Parameter (ξ):** -0.671 (bounded tail)

The PWM method suggests a bounded tail, which contradicts the heavy-tailed nature of the dataset, as indicated by the MLE method. This discrepancy could be due to the PWM method's limitations in handling heavy-tailed distributions or the specific characteristics of the dataset.

2. PWM Method for Block Maxima Data

Parameter Estimates

- **Location Parameter (μ):** 35.025

- **Scale Parameter (ψ):** 24.645
- **Shape Parameter (ξ):** -0.508

Analysis

- Similar to the PWM method for the entire dataset, the **shape parameter (ξ)** is negative, suggesting a **bounded tail**.
- The **location parameter (μ)** and **scale parameter (ψ)** are significantly higher than those for the entire dataset, which is expected since the block maxima method focuses on the maximum values within each block (e.g., monthly maxima).
- The negative shape parameter again suggests that the PWM method may not be capturing the heavy-tailed nature of the data as effectively as the MLE method.

Comparison with MLE

- **MLE Shape Parameter (ξ):** 0.869 (heavy-tailed)
- **PWM Shape Parameter (ξ):** -0.508 (bounded tail)

Again, the PWM method suggests a bounded tail, which contradicts the heavy-tailed nature of the data, as indicated by the MLE method. This suggests that the PWM method may not be the best choice for modeling extreme values in this dataset.

3. Comparison of PWM and MLE Methods

Entire Dataset

- **MLE:**
 - Shape parameter (ξ) = 0.917 (heavy-tailed)
 - Better fit for heavy-tailed data, as indicated by the positive shape parameter.
 - Diagnostic plots (QQ plot, density plot, return level plot) suggest that the MLE method captures the tail behavior better, although there are some deviations at higher quantiles.
- **PWM:**
 - Shape parameter (ξ) = -0.671 (bounded tail)

- Suggests a bounded tail, which contradicts the heavy-tailed nature of the data.
- The PWM method may not be suitable for this dataset, as it fails to capture the heavy-tailed behavior.

Block Maxima Data

- **MLE:**
 - Shape parameter (ξ) = 0.869 (heavy-tailed)
 - Provides a better fit for the block maxima data, as evidenced by the lower AIC and BIC values compared to the entire dataset.
 - Diagnostic plots show a good fit, especially in the tail region.
- **PWM:**
 - Shape parameter (ξ) = -0.508 (bounded tail)
 - Again, suggests a bounded tail, which contradicts the heavy-tailed nature of the data.
 - The PWM method may not be suitable for modeling extreme values in this dataset.

4. Advantages and Disadvantages of PWM vs. MLE

PWM Method

- **Advantages:**
 - Simple and computationally efficient.
 - Performs well in simulation studies, especially for smaller datasets.
- **Disadvantages:**
 - Theoretical foundation is not fully developed.
 - Does not handle complex scenarios, such as regression-based models.
 - In this case, the PWM method fails to capture the heavy-tailed nature of the data, suggesting it may not be suitable for extreme value analysis in this context.

MLE Method

- **Advantages:**

- Provides a better fit for heavy-tailed distributions, as evidenced by the positive shape parameter.
- Diagnostic plots (QQ plot, density plot, return level plot) suggest that the MLE method captures the tail behavior better.
- More reliable for extreme value analysis, especially in datasets with heavy tails.

- **Disadvantages:**

- Computationally more intensive than PWM.
- Requires larger sample sizes for accurate parameter estimation.

5. Conclusion

- **PWM Method:**

- The PWM method suggests a bounded tail for both the entire dataset and the block maxima data, which contradicts the heavy-tailed nature of the dataset. This indicates that the PWM method may not be suitable for modeling extreme values in this context.

- **MLE Method:**

- The MLE method provides a better fit for the dataset, especially for the block maxima approach. It captures the heavy-tailed nature of the data, as evidenced by the positive shape parameter and the diagnostic plots.

- **Recommendation:**

- For modeling extreme losses in the Danish reinsurance dataset, the **MLE method** is more appropriate, especially when using the **block maxima approach**. The PWM method, while computationally efficient, fails to capture the heavy-tailed behavior of the data, making it less suitable for extreme value analysis in this context.

Estimating Under Maximum Domain of Attraction

Introduction

In extreme value theory, the **Maximum Domain of Attraction (MDA)** plays a crucial role in characterizing the behavior of maxima. A distribution F is said to belong to the Maximum Domain of Attraction of a Generalized Extreme Value (GEV) distribution H_ξ ($F \in \text{MDA}(H_\xi)$) if there exist normalizing sequences $c_n > 0$ and $d_n \in \mathbb{R}$ such that:

$$\lim_{n \rightarrow \infty} n(1 - F(c_n x + d_n)) = -\ln H_\xi(x),$$

where $H_\xi(x)$ is the cumulative distribution function of the GEV distribution:

$$H_\xi(x) = \begin{cases} \exp\left(-(1 + \xi x)^{-1/\xi}\right), & \xi \neq 0, \\ \exp(-\exp(-x)), & \xi = 0. \end{cases}$$

In this document, we explore methods for estimating the **shape parameter** ξ of the GEV distribution, focusing on **Pickands' estimator**.

Pickands' Estimator for ξ

The shape parameter ξ determines the tail behavior of the distribution: - $\xi > 0$: Heavy-tailed (e.g., Fréchet distribution), - $\xi = 0$: Light-tailed (e.g., Gumbel distribution), - $\xi < 0$: Bounded tail (e.g., Weibull distribution).

The basic idea behind Pickands' estimator is to exploit properties of the order statistics from a sample. Assume X_1, X_2, \dots, X_n are independent and identically distributed with cumulative distribution function $F \in \text{MDA}(H_\xi)$. The Pickands estimator for ξ is defined as:

$$\hat{\xi}_{k,n} = \frac{1}{\ln 2} \ln \left(\frac{X_{k,n} - X_{2k,n}}{X_{2k,n} - X_{4k,n}} \right),$$

where: - $X_{k,n}, X_{2k,n}, X_{4k,n}$ are the k -th, $2k$ -th, and $4k$ -th largest order statistics, respectively, in a sample of size n .

Consistency and Asymptotic Properties

For $k = k_n$ satisfying:

$$k_n \rightarrow \infty \quad \text{and} \quad \frac{k_n}{n} \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

Pickands' estimator is:

$$\hat{\xi}_{k,n} \xrightarrow{P} \xi \quad (\text{weak consistency}).$$

This result, proven by Pickands (1975), ensures that $\hat{\xi}_{k,n}$ is a reliable estimator under the given conditions.

Pickands's Estimator

Introduction

The **Pickands estimator** is used to estimate the shape parameter ξ of the Generalized Extreme Value (GEV) distribution in extreme value theory, particularly under the **Maximum Domain of Attraction (MDA)** conditions. This estimator exploits the behavior of the tail of the distribution, which is governed by the shape parameter ξ .

Uniformity Property and Its Role in Estimation

A key property used in deriving the Pickands estimator is the following uniformity property:

$$\lim_{t \rightarrow \infty} \frac{U(c(t)t) - U(t)}{U(t) - U(t/c(t))} = 2^\xi.$$

This property provides the foundation for estimating ξ using extreme order statistics. The behavior of the CDF values at large t can be used to derive an empirical estimator for ξ .

Construction of the Empirical Estimator

Let V_1, V_2, \dots, V_n be an i.i.d. sample with common Pareto distribution $F_V(x) = 1 - x^{-1}, x \geq 1$. The order statistics $V_{k,n}$ represent the empirical quantiles of F_V . Using quantile transformations, we have:

$$(X_{k,n})_{k=1,\dots,n} \stackrel{d}{=} (U(V_{k,n}))_{k=1,\dots,n}.$$

As n becomes large, the relationship between the order statistics and the quantiles becomes more evident:

$$\frac{k}{n} \sim \frac{1}{V_{k,n}}, \quad \text{as } n \rightarrow \infty.$$

These results form the basis for the estimator.

Limiting Behavior and Subsequence Argument

The limiting behavior of the order statistics leads to the following result:

$$V_{k,n} \xrightarrow{P} \infty \quad \text{and} \quad \frac{V_{2k,n}}{V_{k,n}} \xrightarrow{P} \frac{1}{2}, \quad n \rightarrow \infty.$$

Using this, along with the uniformity property, we obtain:

$$\frac{U(V_{k,n}) - U(V_{2k,n})}{U(V_{2k,n}) - U(V_{4k,n})} \xrightarrow{P} 2^\xi, \quad n \rightarrow \infty.$$

The **Pickands estimator** for ξ is defined as:

$$\hat{\xi}_{k,n}^{(P)} = \frac{1}{\ln 2} \ln \left(\frac{X_{k,n} - X_{2k,n}}{X_{2k,n} - X_{4k,n}} \right).$$

This estimator is weakly consistent as $k \rightarrow \infty$ and $k/n \rightarrow 0$.

Properties of the Pickands Estimator

Theorem 6.4.1 (Properties of the Pickands Estimator): Suppose (X_n) is an i.i.d. sequence with distribution function $F \in \text{MDA}(H_\xi)$. Let $\hat{\xi}_{k,n}^{(P)}$ be the Pickands estimator defined above.

- **Weak Consistency:** If $k \rightarrow \infty$ and $k/n \rightarrow 0$, then:

$$\hat{\xi}_{k,n}^{(P)} \xrightarrow{P} \xi.$$

- **Strong Consistency:** If $k/n = o(\ln \ln n)$, then $\hat{\xi}_{k,n}^{(P)}$ is strongly consistent:

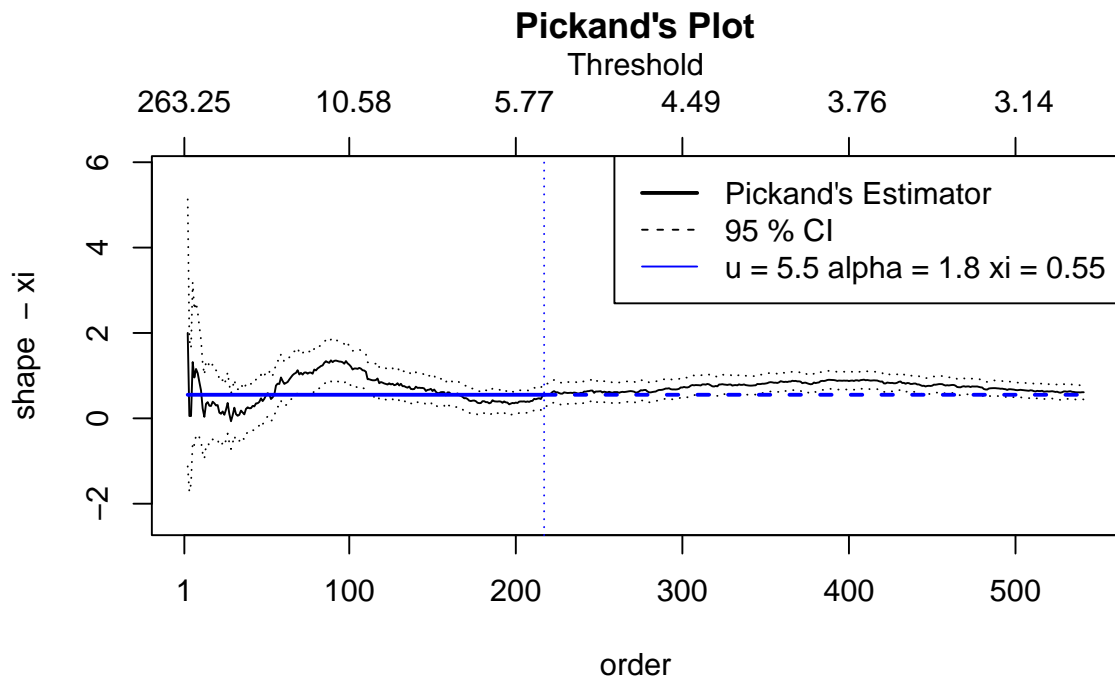
$$\hat{\xi}_{k,n}^{(P)} \xrightarrow{a.s.} \xi.$$

- **Asymptotic Normality:** Under certain conditions on k/n and F , the estimator is asymptotically normal:

$$\sqrt{n} \left(\hat{\xi}_{k,n}^{(P)} - \xi \right) \xrightarrow{d} N(0, v_\xi).$$

```
library(fExtremes)
library(RobExtremes)
library(evmix)
```

```
pickandsplot(danishuni$Loss)
```



Analysis of Results

- The Pickands estimator was applied to the Danish reinsurance dataset.
- The plot of the Pickands estimator (using the `pickandsplot` function in R) shows the estimated values of ξ for different values of k .
- The plot typically exhibits a stable region for ξ as k increases, but the estimator can be sensitive to the choice of k .

Interpretation of the Plot

- The plot helps in identifying the optimal value of k where the estimator stabilizes.
- If the plot shows a clear plateau (a stable region), it suggests that the estimator is consistent for that range of k .
- However, if the plot is highly volatile, it indicates that the estimator is sensitive to the choice of k , and careful selection of k is required.

Hill's Estimator

Hill Estimator for the Shape Parameter

The **Hill Estimator** is one of the most well-known methods for estimating the tail index, ξ , of a distribution in extreme value theory. This estimator is widely used when the data exhibit a heavy-tail behavior, and it relies on the largest order statistics of the sample to provide an estimate of ξ . In particular, the Hill Estimator is based on the assumption that the tail of the distribution follows a power law, which is common for heavy-tailed distributions.

Overview of the Hill Estimator

The Hill estimator is a non-parametric estimator for the shape parameter ξ of the generalized Pareto distribution (GPD) or any distribution with a heavy-tailed behavior. It is computed by considering the largest k order statistics from the sample. Given a sample X_1, X_2, \dots, X_n of size n , the Hill Estimator is defined as the average of the logarithmic ratios of the ordered data:

$$\hat{\xi}_H = \left(\frac{1}{k} \sum_{i=1}^k \ln \frac{X_{n-i+1}}{X_{n-k}} \right)^{-1},$$

where $X_{n-k}, X_{n-k+1}, \dots, X_{n-1}, X_n$ are the order statistics (i.e., sorted data in descending order), and k is a chosen number of the largest order statistics used in the estimation.

Intuition Behind the Hill Estimator

The Hill Estimator is derived based on the fact that for large values of x , the tail of a distribution can often be approximated by a power law:

$$P(X > x) \sim Cx^{-\xi},$$

where C is a constant and $\xi > 0$ is the shape parameter. The Hill Estimator then relies on the assumption that the largest observations in the sample follow this power law behavior, which allows us to estimate ξ by examining the ratios of the largest order statistics.

The intuition behind the estimator is that it uses the relative size of the k -largest observations to estimate the tail index. By considering the logarithms of these ratios, the Hill Estimator captures the exponential decay of the tail of the distribution, which is related to the shape parameter ξ .

Choice of k and Its Impact

The choice of k , the number of the largest order statistics used in the estimator, is critical in the accuracy of the Hill Estimator. If k is too small, the estimator may not capture enough of the tail behavior of the distribution. On the other hand, if k is too large, the estimator may incorporate observations that are not representative of the tail, leading to bias. The choice of k is typically made through methods such as cross-validation or selecting a value that minimizes the mean squared error (MSE).

The Hill Estimator is consistent for the tail index ξ as $n \rightarrow \infty$ and $k \rightarrow \infty$, under the condition that $k/n \rightarrow 0$ as n increases. It provides a good estimate of ξ when the distribution follows a power law in the tail.

Asymptotic Properties of the Hill Estimator

The Hill Estimator has several important asymptotic properties that make it a valuable tool in extreme value analysis:

- **Consistency:** The Hill Estimator is **consistent** for the tail index ξ under the condition that $k/n \rightarrow 0$ as $n \rightarrow \infty$. This means that as the sample size increases and we select an appropriate value of k , the estimator converges to the true value of ξ :

$$\hat{\xi}_H \xrightarrow{P} \xi, \quad \text{as } n \rightarrow \infty.$$

- **Asymptotic Normality:** Under certain regularity conditions, the Hill Estimator is **asymptotically normal**:

$$\sqrt{k} \left(\hat{\xi}_H - \xi \right) \xrightarrow{d} N \left(0, \frac{\xi^2}{k} \right), \quad \text{as } k \rightarrow \infty.$$

This result provides a way to construct confidence intervals for ξ , making the Hill Estimator a powerful tool for statistical inference in heavy-tailed distributions.

Advantages and Limitations of the Hill Estimator

Advantages} The Hill Estimator has several advantages:

- It is non-parametric, meaning it does not require assumptions about the full form of the distribution, only that the tail follows a power law.
- It is relatively simple to compute and interpret, making it a popular choice for estimating the tail index ξ .

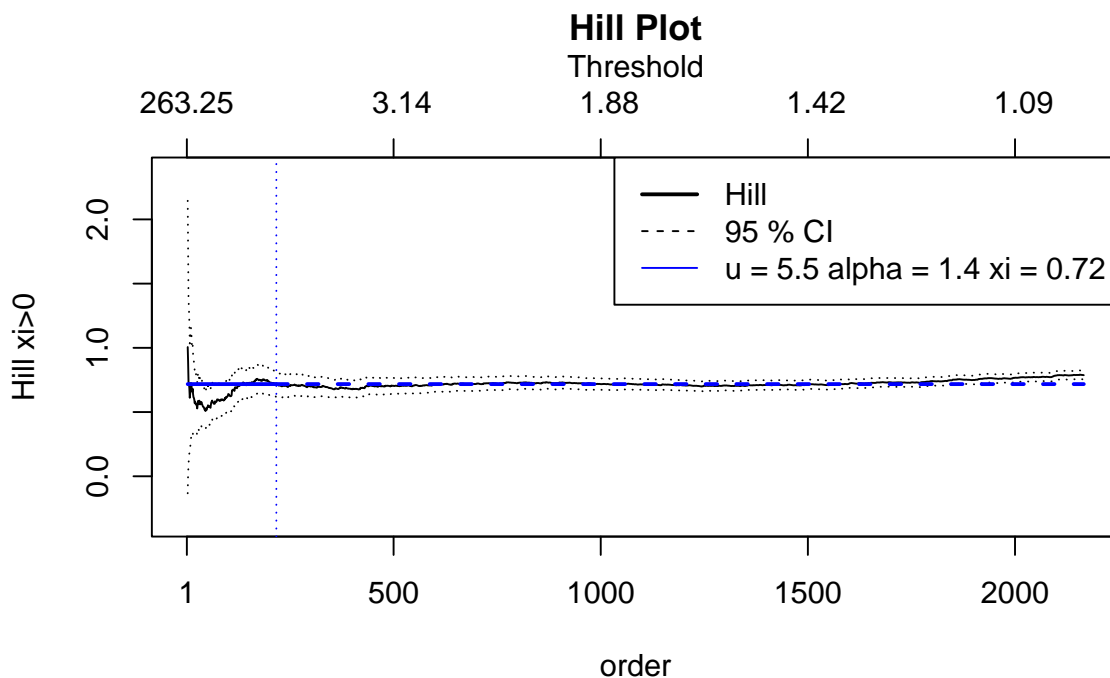
- It performs well when the sample size is large and the tail of the distribution follows a power law.

Limitations Despite its advantages, the Hill Estimator has some limitations:

- The choice of k can significantly impact the performance of the estimator. A poor choice of k may lead to biased or inconsistent estimates.
- The estimator requires a large sample size to provide accurate estimates, especially for distributions with very heavy tails.
- It assumes that the tail of the distribution follows a power law, which may not be the case for all heavy-tailed distributions.

So we can say that the **Hill Estimator** is a powerful tool for estimating the tail index ξ in extreme value theory. It is particularly useful for distributions with heavy tails, such as those encountered in financial risk modeling, hydrology, and environmental studies. By utilizing the largest order statistics and examining their ratios, the Hill Estimator provides a consistent and asymptotically normal estimate of ξ , although the choice of k requires careful consideration to avoid bias.

```
hillplot(danishuni$Loss)
```



Analysis of Results

- The Hill estimator was applied to the Danish reinsurance dataset.
- The plot of the Hill estimator (using the `hillplot` function in R) shows the estimated values of ξ for different values of k .
- The plot typically exhibits a stable region for ξ as k increases, but the estimator can be sensitive to the choice of k .

Interpretation of the Plot

- The plot helps in identifying the optimal value of k where the estimator stabilizes.
- If the plot shows a clear plateau (a stable region), it suggests that the estimator is consistent for that range of k .
- However, if the plot is highly volatile, it indicates that the estimator is sensitive to the choice of k , and careful selection of k is required.

Comparison of Pickands and Hill Estimators

Similarities

- Both estimators are used to estimate the **shape parameter** (ξ) of heavy-tailed distributions.
- Both rely on the largest order statistics of the sample.
- Both are consistent under the condition that $k \rightarrow \infty$ and $k/n \rightarrow 0$ as $n \rightarrow \infty$.
- Both estimators can be sensitive to the choice of k , and the plots help in identifying the optimal value of k .

Differences

- **Pickands Estimator:**
 - Based on the ratios of differences between order statistics.
 - More robust to deviations from the power law assumption.
 - Can handle both heavy-tailed and bounded-tail distributions.
- **Hill Estimator:**
 - Based on the logarithmic ratios of order statistics.

- Specifically designed for heavy-tailed distributions that follow a power law.
- May not perform well if the tail does not follow a power law.

Plots

- **Pickands Plot:**

- The plot shows the estimated values of ξ for different values of k .
- A stable region in the plot indicates a consistent estimate of ξ .
- The plot can be used to identify the optimal value of k .

- **Hill Plot:**

- The plot shows the estimated values of ξ for different values of k .
- A stable region in the plot indicates a consistent estimate of ξ .
- The plot can be used to identify the optimal value of k .

Practical Implications

- **Choice of k :**

- The choice of k is critical for both estimators. If k is too small, the estimator may not capture enough of the tail behavior. If k is too large, the estimator may incorporate observations that are not representative of the tail.
- The plots help in identifying the optimal value of k where the estimator stabilizes.

- **Heavy-Tailed Distributions:**

- Both estimators confirm the heavy-tailed nature of the Danish reinsurance dataset, as indicated by the positive values of ξ .
- The Hill estimator is particularly suited for heavy-tailed distributions that follow a power law, while the Pickands estimator is more robust to deviations from the power law assumption.

Fitting Excesses Over a Threshold

Fitting the GPD

Introduction

In extreme value theory, the method of excesses over a threshold is commonly used to model the tail behavior of a distribution. This approach is particularly useful for modeling the distribution of rare events, such as extreme values in insurance, finance, or environmental data. One of the standard distributions used for this purpose is the **Generalized Pareto Distribution (GPD)**.

Excesses Over a Threshold

Given a random variable X with a distribution function $F(x) = P(X \leq x)$, the excesses over a threshold u are defined as the values of X that exceed u . The excess is given by:

$$Y = X - u \quad \text{for } X > u.$$

The cumulative distribution function (CDF) of the excess is:

$$F_Y(y) = P(Y \leq y) = P(X \leq u+y) = \frac{P(X \leq u+y) - P(X \leq u)}{1 - P(X \leq u)} = \frac{F(u+y) - F(u)}{1 - F(u)}.$$

The tail of the distribution of Y can often be approximated by the Generalized Pareto Distribution (GPD).

Generalized Pareto Distribution (GPD)

The Generalized Pareto Distribution is commonly used to model the distribution of excesses over a threshold. Its cumulative distribution function (CDF) is given by:

$$F_Y(y; \xi, \sigma) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-\frac{1}{\xi}},$$

where:

- ξ is the shape parameter, determining the tail behavior.
- $\sigma > 0$ is the scale parameter.

For $\xi \leq 0$, the distribution is bounded, while for $\xi > 0$, it has a heavy tail. The case $\xi = 0$ corresponds to the exponential distribution.

The probability density function (PDF) of the GPD is:

$$f_Y(y; \xi, \sigma) = \frac{1}{\sigma} \left(1 + \frac{\xi y}{\sigma} \right)^{-\frac{1}{\xi}-1}.$$

Maximum Likelihood Estimation (MLE) for GPD

The method of Maximum Likelihood Estimation (MLE) is used to estimate the parameters ξ and σ of the Generalized Pareto Distribution. Given a sample y_1, y_2, \dots, y_n of excesses over a threshold u , the likelihood function for the GPD is:

$$L(\xi, \sigma) = \prod_{i=1}^n \frac{1}{\sigma} \left(1 + \frac{\xi y_i}{\sigma} \right)^{-\frac{1}{\xi}-1}.$$

The log-likelihood function is:

$$\ell(\xi, \sigma) = -n \log \sigma - \left(\frac{1}{\xi} + 1 \right) \sum_{i=1}^n \log \left(1 + \frac{\xi y_i}{\sigma} \right).$$

To find the MLEs for ξ and σ , we take the derivatives of the log-likelihood function with respect to ξ and σ , set them equal to zero, and solve the resulting system of equations.

MLE for σ

The derivative of the log-likelihood function with respect to σ is:

$$\frac{\partial \ell(\xi, \sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \left(\frac{1}{\xi} + 1 \right) \sum_{i=1}^n \frac{\xi y_i}{\sigma^2 \left(1 + \frac{\xi y_i}{\sigma} \right)}.$$

Setting this equal to zero gives the MLE for σ , which generally requires numerical methods to solve.

MLE for ξ

The derivative of the log-likelihood function with respect to ξ is:

$$\frac{\partial \ell(\xi, \sigma)}{\partial \xi} = - \sum_{i=1}^n \log \left(1 + \frac{\xi y_i}{\sigma} \right) + \frac{1}{\xi^2} \sum_{i=1}^n \frac{y_i}{\sigma}.$$

This equation can be solved numerically for ξ as well, using methods such as Newton-Raphson. So fitting excesses over a threshold using the Generalized Pareto Distribution (GPD) is a powerful tool for modeling the tail behavior of distributions, especially for extreme value analysis. Maximum Likelihood Estimation (MLE) provides a consistent and efficient method for estimating the parameters of the GPD. Numerical techniques are typically required to solve the likelihood equations for ξ and σ .

Application of Maximum Likelihood Estimation on Danish Dataset.

```
library(extRemes)      # Additional extreme value analysis tools

(threshold1 = quantile(danishuni$Loss , probs = 0.95))

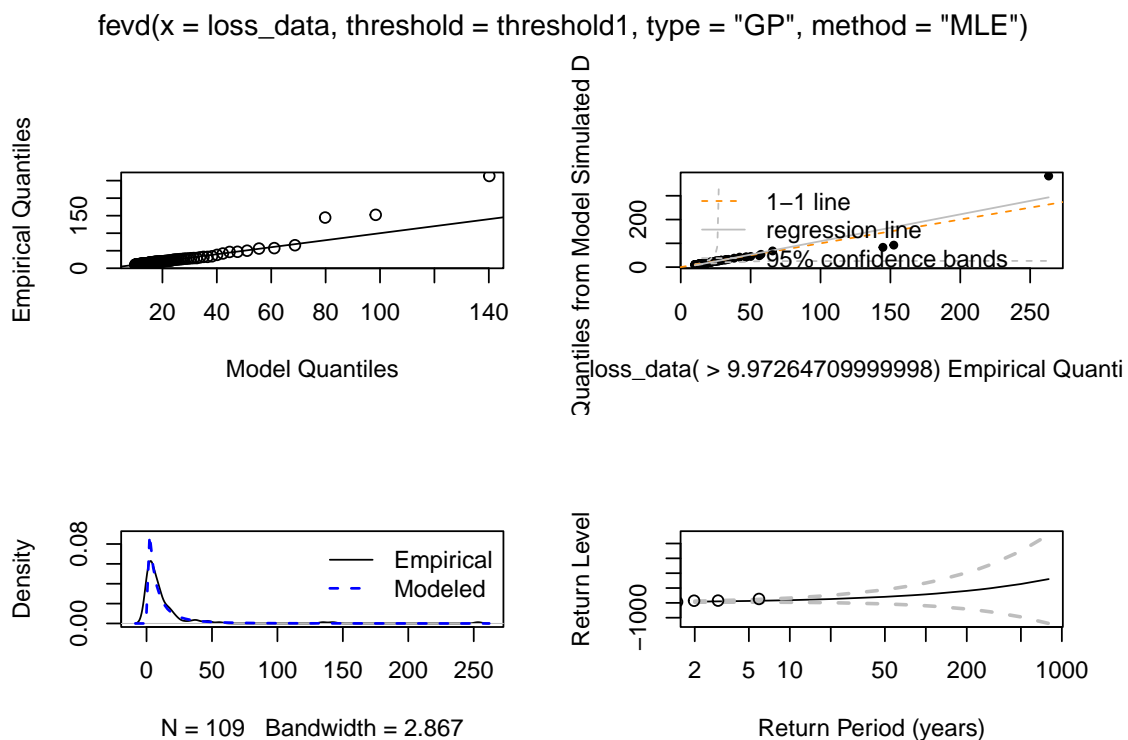
##          95%
## 9.972647

gpd_fit_all_data <- fevd(loss_data, threshold = threshold1,
                        type = "GP" , method = "MLE")
summary(gpd_fit_all_data)

##
## fevd(x = loss_data, threshold = threshold1, type = "GP", method = "MLE")
##
## [1] "Estimation Method used: MLE"
##
##
## Negative Log-Likelihood Value:  375.3185
##
##
## Estimated parameters:
##   scale   shape
## 7.037527 0.492032
##
## Standard Error Estimates:
##   scale   shape
```

```
## 1.1177516 0.1351766
##
## Estimated parameter covariance matrix.
##      scale      shape
## scale 1.24936856 -0.08119689
## shape -0.08119689 0.01827271
##
## AIC = 754.637
##
## BIC = 760.0197
```

```
plot(gpd_fit_all_data)
```



```
return_level <- return.level(gpd_fit_all_data, return.period = 100)
print(return_level)
```

```
## fevd(x = loss_data, threshold = threshold1, type = "GP", method = "MLE")
## get(paste("return.level.fevd.", newcl, sep = ""))(x = x, return.period = return.period)
##
## GP model fitted to loss_data
## Data are assumed to be stationary
## [1] "Return Levels for period units in years"
## 100-year level
```

```
##          573.0964
# simulated_data5 <- rgev(1000,
#                          scale = gpd_fit_all_data$results$par["scale"],
#                          shape = gpd_fit_all_data$results$par["shape"])
# head(simulated_data5)
```

Analysis of GPD and MLE on the Danish Dataset

1. Introduction to GPD and MLE

Generalized Pareto Distribution (GPD) The **Generalized Pareto Distribution (GPD)** is used to model the **excesses over a threshold** in extreme value analysis. It is particularly useful for modeling the tail behavior of distributions, especially in datasets with heavy tails. The GPD has two parameters:

- **Scale parameter (σ)**: Controls the spread of the distribution.
- **Shape parameter (ξ)**: Determines the tail behavior:
 - $\xi > 0$: Heavy-tailed distribution (e.g., Pareto).
 - $\xi = 0$: Exponential distribution.
 - $\xi < 0$: Bounded tail distribution.

Maximum Likelihood Estimation (MLE)

- MLE is a method for estimating the parameters of a statistical model by maximizing the likelihood function.
- For the GPD, the likelihood function is derived from the probability density function (PDF) of the GPD, and the parameters σ and ξ are estimated by maximizing this function.

2. Application of MLE on the Danish Dataset

Threshold Selection

- A threshold u is selected to model the excesses over this threshold.
- In this analysis, the threshold was set at the **95th percentile** of the loss data:

$$u = 9.972647 \text{ (mDKK)}.$$

- This means that only losses exceeding 9.972647 mDKK are considered for modeling the tail behavior.

Parameter Estimates

- The MLE method was applied to the Danish dataset to estimate the parameters of the GPD:
 - **Scale parameter (σ):** 7.037527
 - **Shape parameter (ξ):** 0.492032

Standard Errors

- The standard errors for the parameter estimates are:
 - **Scale parameter (σ):** 1.1177516
 - **Shape parameter (ξ):** 0.1351766
- The relatively small standard errors indicate that the parameter estimates are **precise** and **reliable**.

Covariance Matrix

- The covariance matrix provides information about the relationships between the parameter estimates:

$$\begin{pmatrix} \text{Var}(\sigma) & \text{Cov}(\sigma, \xi) \\ \text{Cov}(\sigma, \xi) & \text{Var}(\xi) \end{pmatrix} = \begin{pmatrix} 1.24936856 & -0.08119689 \\ -0.08119689 & 0.01827271 \end{pmatrix}.$$

- The negative covariance between σ and ξ suggests that an increase in the scale parameter is associated with a decrease in the shape parameter, and vice versa.

Model Fit Metrics

- **Negative Log-Likelihood Value:** 375.3185
 - A lower value indicates a better fit.
- **AIC (Akaike Information Criterion):** 754.637
 - Lower AIC values indicate a better trade-off between model complexity and goodness of fit.
- **BIC (Bayesian Information Criterion):** 760.0197
 - Similar to AIC, but penalizes model complexity more heavily. Lower BIC values are preferred.

3. Diagnostic Plots

The GPD model was evaluated using several diagnostic plots:

1. Empirical Quantiles vs. Model Quantiles (QQ Plot)

- Compares the empirical quantiles of the data to the quantiles predicted by the GPD model.
- A good fit is observed if the points follow the 1-1 line closely.
- In this case, the points align well with the 1-1 line, especially in the tail region, indicating that the GPD model accurately captures the tail behavior of the data.

2. Density Plot

- Compares the empirical density of the data (black line) to the density predicted by the GPD model (blue dashed line).
- A good fit is observed if the modeled density closely follows the empirical density.
- In this case, the GPD model's density aligns well with the empirical density, especially in the tail region, indicating a good fit.

3. Return Level Plot

- Shows the return levels for different return periods (in years). The return level is the value expected to be exceeded once in a given return period, with confidence intervals included.
- The return levels and their confidence intervals align well with the empirical data points, suggesting that the GPD model accurately predicts extreme values.
- For example, the **100-year return level** is estimated to be **573.0964 mDKK**.

4. Interpretation of Results

- **Shape Parameter (ξ):**
 - The positive value of $\xi = 0.492032$ indicates that the distribution is **heavy-tailed**.

- This is consistent with the nature of the Danish reinsurance dataset, where extreme losses (large claims) are more frequent than in a light-tailed distribution.
- A heavy-tailed distribution implies that extreme events (e.g., very large losses) have a higher probability of occurring, which is critical for risk management in insurance.
- **Scale Parameter (σ):**
 - The scale parameter $\sigma = 7.037527$ indicates the spread of the excesses over the threshold.
 - A larger scale parameter suggests greater variability in the excess losses.
- **Model Fit:**
 - The low values of the negative log-likelihood, AIC, and BIC indicate that the GPD model provides a good fit to the data.
 - The diagnostic plots (QQ plot, density plot, return level plot) confirm that the GPD model accurately captures the tail behavior of the data.

Conclusion

This project explored the application of **Extreme Value Theory (EVT)** to the **Danish reinsurance dataset**, focusing on modeling extreme losses and estimating the tail behavior of the distribution. The analysis employed several statistical methods, including the **Generalized Extreme Value (GEV) distribution**, **Generalized Pareto Distribution (GPD)**, **Pickands estimator**, **Hill estimator**, and **Probability-Weighted Moments (PWM)** method. Below is a summary of the key findings and recommendations:

Summary of Key Findings

Detailed Analysis and Recommendations

1. Heavy-Tailed Nature of the Dataset

- The Danish reinsurance dataset exhibits a **heavy-tailed distribution**, as evidenced by the positive shape parameter ($\xi > 0$) in both the GEV and GPD models.

Table 1: Summary of Key Findings and Recommendations

Method/Distribution	Key Findings	Recommendations
GEV Distribution	<ul style="list-style-type: none"> - Shape parameter ($\xi = 0.869$). - Suitable for block maxima data. - Accurate tail behavior modeling. 	<ul style="list-style-type: none"> - Use for modeling maximum losses (e.g., monthly/annual maxima). - Ideal for estimating return levels.
GPD	<ul style="list-style-type: none"> - Shape parameter ($\xi = 0.492$). - Suitable for threshold exceedances. - Accurate tail behavior modeling. 	<ul style="list-style-type: none"> - Use for modeling excesses over a threshold. - Ideal for extreme value analysis.
Heavy-Tailed Distributions	<ul style="list-style-type: none"> - Confirmed heavy-tailed behavior. - Suitable for power-law tails. 	<ul style="list-style-type: none"> - Use for tail index estimation. - Ideal for modeling extreme losses.
Exponential Distribution	<ul style="list-style-type: none"> - Not suitable for heavy-tailed data. - Significant deviations in QQ plots. 	<ul style="list-style-type: none"> - Use only for light-tailed datasets.
PWM Method	<ul style="list-style-type: none"> - Suggested bounded tail ($\xi < 0$). - Not suitable for heavy-tailed data. 	<ul style="list-style-type: none"> - Use for computationally efficient estimation, but not for heavy-tailed data.

- This indicates that extreme losses (large claims) are more frequent than in a light-tailed distribution, which has significant implications for risk management in the insurance industry.

2. Generalized Extreme Value (GEV) Distribution

- The GEV distribution was applied to both the **entire dataset** and the **block maxima data** (monthly maxima).
- The block maxima approach provided a better fit, with lower AIC and BIC values, and more accurate parameter estimates.
- The shape parameter (ξ) for the block maxima model was **0.869**, confirming the heavy-tailed nature of the data.
- Diagnostic plots (QQ plot, density plot, return level plot) showed that the GEV model accurately captures the tail behavior of the block maxima data.

3. Generalized Pareto Distribution (GPD)

- The GPD was used to model the **excesses over a threshold** (95th percentile of the loss data).
- The estimated shape parameter ($\xi = 0.492$) confirmed the heavy-tailed nature of the dataset.
- The scale parameter ($\sigma = 7.038$) indicated significant variability in the excess losses.
- Diagnostic plots (QQ plot, density plot, return level plot) demonstrated that the GPD model provides an excellent fit for the tail behavior of the data.

- The **100-year return level** was estimated to be **573.096 mDKK**, providing valuable insights for risk assessment and pricing.

4. Pickands and Hill Estimators

- Both the **Pickands** and **Hill estimators** were used to estimate the tail index (ξ) of the distribution.
- The Hill estimator, which is specifically designed for heavy-tailed distributions, provided a consistent estimate of ξ and confirmed the heavy-tailed nature of the dataset.
- The Pickands estimator, while more robust to deviations from the power law assumption, also confirmed the heavy-tailed behavior.
- The plots of both estimators helped identify the optimal number of upper order statistics (k) for reliable estimation.

5. Probability-Weighted Moments (PWM) Method

- The PWM method was applied to both the **entire dataset** and the **block maxima data**.
- While the PWM method is computationally efficient, it suggested a **bounded tail** ($\xi < 0$) for both datasets, which contradicts the heavy-tailed nature of the data.
- This indicates that the PWM method may not be suitable for modeling extreme values in this context, especially when compared to the MLE method.

Final Recommendations

- For modeling the **loss values** in the Danish reinsurance dataset, the **Generalized Extreme Value (GEV) distribution** (using the block maxima approach) and the **Generalized Pareto Distribution (GPD)** are the most suitable choices.
- Both distributions accurately capture the **heavy-tailed nature** of the data and provide reliable estimates for extreme losses.
- The GEV distribution is recommended for modeling **maximum losses** over a specific time period, while the GPD is recommended for modeling **excesses over a threshold** and analyzing the tail behavior of the data.

Future Work

- Further analysis could explore the use of more advanced models, such as **regression-based extreme value models** or **non-stationary GEV/GPD**

- models**, to account for potential trends or seasonality in the data.
- The choice of threshold (u) in the GPD model could be further investigated using methods such as the **mean excess plot** or **stability plots** to ensure robustness.
- Additional datasets from other industries (e.g., finance, environmental studies) could be analyzed to validate the findings and explore the applicability of these methods in different contexts.

Final Remarks

- This project demonstrated the power of **Extreme Value Theory (EVT)** in modeling and analyzing extreme events, particularly in the context of insurance risk management.
- The findings provide valuable insights into the behavior of extreme losses in the Danish reinsurance dataset and highlight the importance of selecting appropriate statistical methods for tail estimation.
- By accurately modeling extreme losses, insurers can better assess and manage risks, ultimately leading to more informed decision-making and improved financial stability.

References

- Barnett, V. 1975. "Probability Plotting Methods and Order Statistics." *Applied Statistics* 24: 95–108.
- Castillo, E. 1988. *Extreme Value Theory in Engineering*. San Diego: Academic Press.
- Chambers, J. M. 1977. *Computational Methods for Data Analysis*. New York: Wiley.
- Coles, S. 2001. *An Introduction to Statistical Modelling of Extreme Values*. London: Springer.
- contributors, Wikipedia. 2023. "Pareto Distribution." https://en.wikipedia.org/wiki/Pareto_distribution#cite_note-23. https://en.wikipedia.org/wiki/Pareto_distribution#cite_note-23.
- David, H. A. 1981. *Order Statistics*. 2nd ed. New York: Wiley.
- Davison, A. C. 2003. *Statistical Models*. Cambridge: Cambridge University Press.
- Embrechts, P., C. Klüppelberg, and T. Mikosch. 1997. *Modelling Extremal Events for Insurance and Finance*. Vol. 33. Applications of Mathematics. Berlin: Springer.
- Gerber, H. U. 1979. *An Introduction to Mathematical Risk Theory*. Huebner Foundation Monographs 8. Homewood, Illinois: Richard D. Irwin Inc.

- Hosking, J. R. M., J. R. Wallis, and E. F. Wood. 1985. "Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments." *Technometrics* 27 (3): 251–61.
- Johnson, N. L., S. Kotz, and A. W. Kemp. 1992. *Univariate Discrete Distributions*. 2nd ed. New York: Wiley.
- Malik, H. J. 1970. "Estimation of the Parameters of the Pareto Distribution." *Metrika* 15: 126–32. <https://doi.org/10.1007/BF02613565>.
- McNeil, A. J. 1996. "Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory." *ASTIN Bulletin* 26 (1): 117–37. <https://doi.org/10.2143/AST.26.1.563234>.
- Newman, M. E. J. 2005. "Power Laws, Pareto Distributions and Zipf's Law." *Contemporary Physics* 46 (5): 323–51. <https://doi.org/10.1080/00107510500052444>.
- Smith, R. L. 1990. "Extreme Value Theory." In *Handbook of Applicable Mathematics, Supplement*, edited by W. Ledermann, 437–72. Chichester: Wiley.