The POT package

By Avraham Adler

This paper is produced mechanically as part of FAViR.
See `http://www.favir.net` for more information.

Abstract

This paper is intended to briefly demonstrate the **POT** package for use in Extreme Value Theory.

1 Introduction

Extreme Value Theory (EVT) is one of the methods used by actuaries to estimate the tails of loss severity distributions. McNeil [3] discusses how the Generalized Pareto distribution (GPD) can be used to to model the tails of extreme events. [1] There exist a number of packages for the 'R' statistical platform which may be used to investigate data in this framework. One of them is called **'POT'**, or the **P**eaks **O**ver **T**hreshold package [4]. The package does more than mere Generalized Pareto fitting, but lends itself nicely to such. This brief paper assumes a basic knowledge of EVT, and is focused on demonstrating the use of the POT package.

2 Example

Using `actuar` [1] we can create a dataset for investigation. We will set a specific seed so that the results are reproducible.

```
> set.seed(254)
> test.data <- rpareto(n = 1000, shape = 1.5, scale = 100000)
```

Let's take a look at this data. Percentiles are below and histograms are in figures 1 and 2.

[1] A distinction must be made between the Generalized Pareto of Extreme Value Theory and the Generalized Pareto of actuarial literature as defined in Klugman, Panjer, and Wilmot (KPW) [2]. The GPD of EVT is a two-parameter pareto distribution where the shape and scale factors are less correlated than the classic two-parameter pareto. The GPD of KPW is a three-parameter distribution.

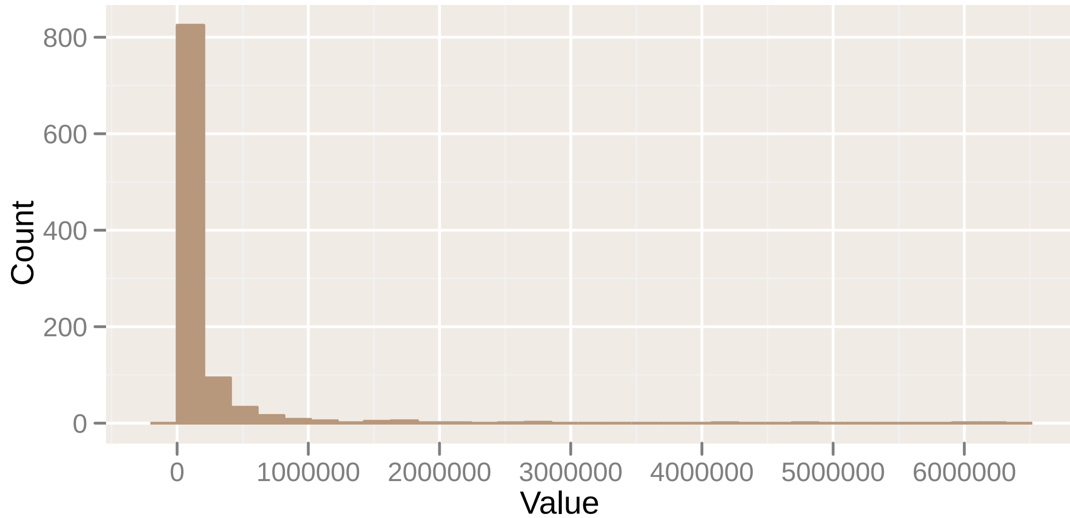


Figure 1: Basic Histogram

```
> summary(test.data)

    Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
    40.2   18700.0   59400.0  164000.0  155000.0 6110000.0
```

In EVT analysis, one often wants to identify the threshold over which the tail exhibits Pareto behavior. One of the primary tools used is the "Sample Mean Excess" or "Mean Residual Life" plot. Where this plot begins to appears linear is often a decent estimate of an appropriate threshold. The POT package has a function to display such a plot: `mrlplot`.

```
> par(mfrow = c(1, 2))
> mrlplot(test.data)
> mrlplot(test.data, xlim = c(0, 1000000))
```

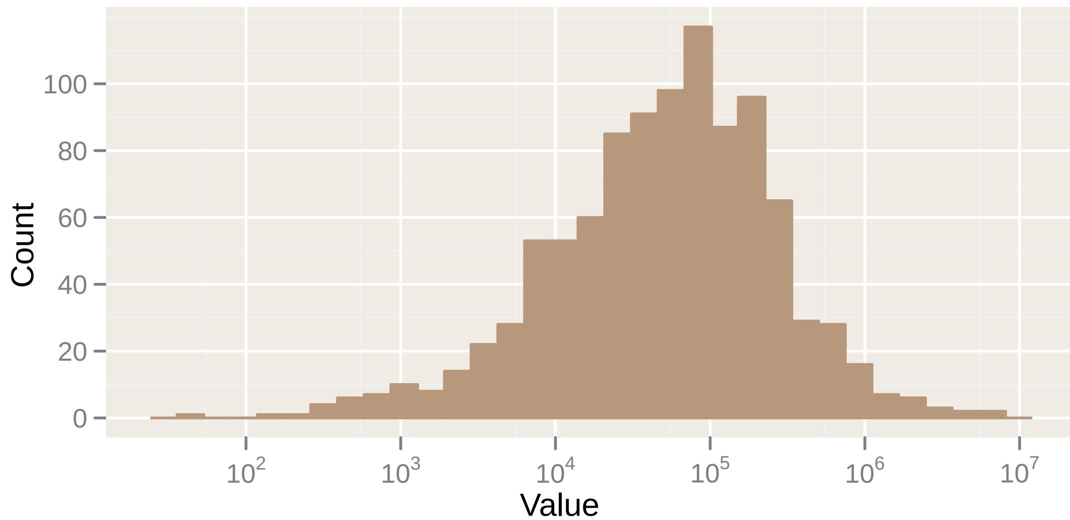
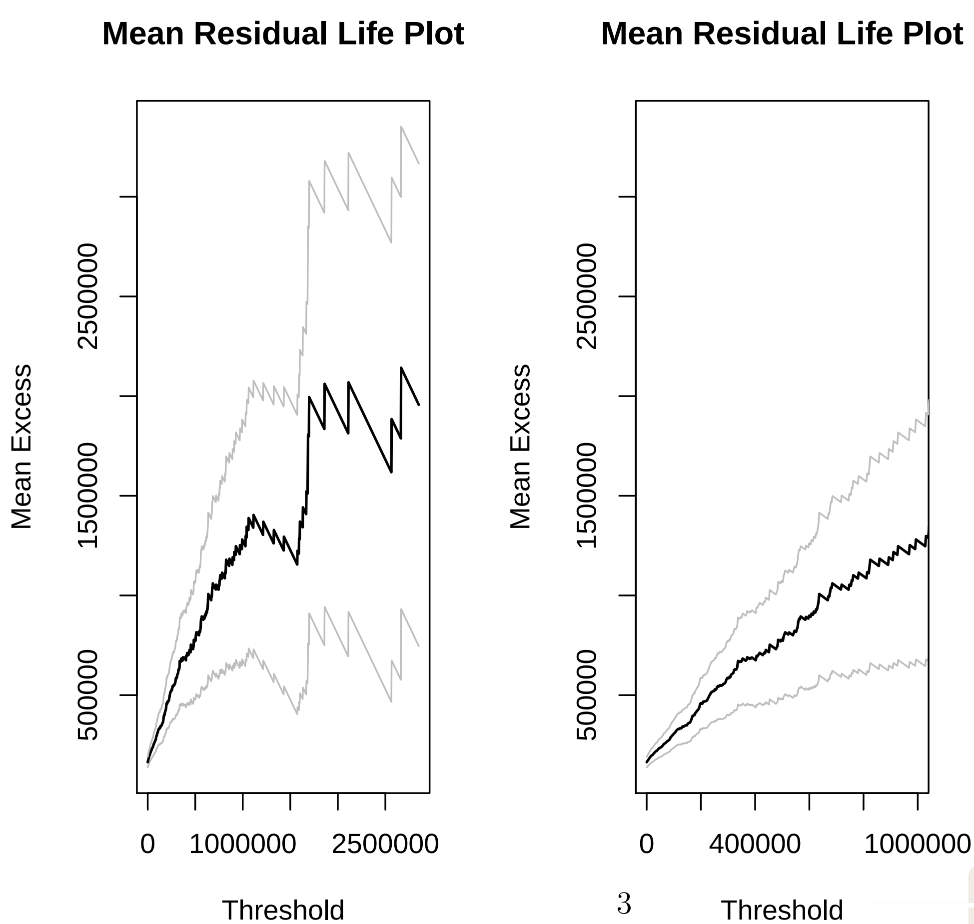


Figure 2: Log-scale Histogram



Looking at the plot, a reasonable selection for the threshold would be 300,000. Once the threashold is selected, POT uses the `fitgpd` command to fit a GPD with the selected threshold.

```
> GPD1 <- fitgpd(test.data, threshold = 300000)
> GPD1

Estimator: MLE
 Deviance: 3212.4
      AIC: 3216.4

Varying Threshold: FALSE

  Threshold Call: 300000
    Number Above: 113
Proportion Above: 0.113

Estimates
     scale       shape
586836.724       0.247

Standard Error Type: expected

Standard Errors
    scale      shape
87174.571      0.117

Asymptotic Variance Covariance
       scale      shape
scale   7.60e+09  -6.47e+03
shape  -6.47e+03   1.38e-02

Optimization Information
  Convergence: successful
  Function Evaluations: 14
  Gradient Evaluations: 6
```

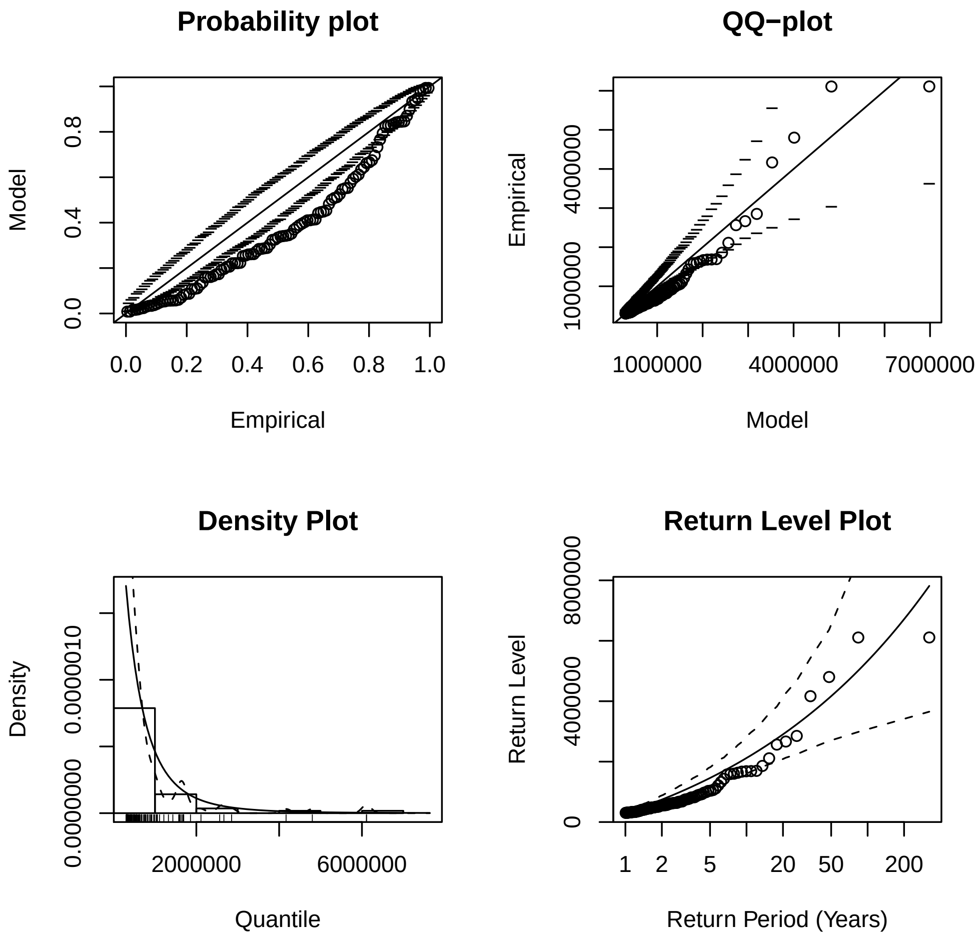
The default parameters that `fitgpd` passes to `optim` often prevent good convergence, so it pays to re-run the optimization passing a vector of parameter scales.

```
> GPD2 <- fitgpd(test.data, threshold = 300000, control = list(parscale = c(100000,
+     0.1)))
> GPD2
```

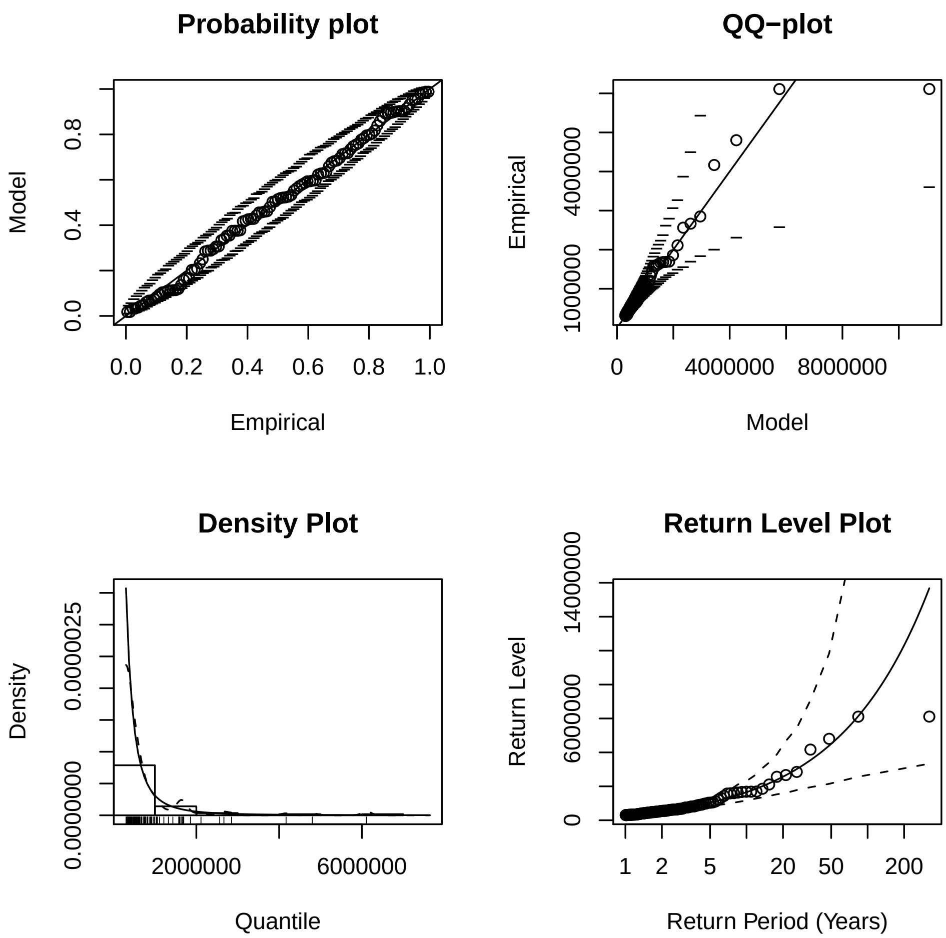
```
Estimator: MLE
 Deviance: 3192
      AIC: 3196

Varying Threshold: FALSE

  Threshold Call: 300000
    Number Above: 113
Proportion Above: 0.113

Estimates
     scale       shape
279868.290       0.582

Standard Error Type: observed

Standard Errors
    scale      shape
48091.411      0.154

Asymptotic Variance Covariance
       scale      shape
scale   2.31e+09  -4.34e+03
shape  -4.34e+03   2.37e-02

Optimization Information
  Convergence: successful
  Function Evaluations: 18
  Gradient Evaluations: 11
```

Note how the fit is now significantly better. Lastly, POT comes with built-in plotting methods, so fits can be analyzed and compared. Below, the two GPD fits will be plotted using default methods.

```
> par(mfrow = c(2, 2))
> plot(GPD1)
```


Probability plot
Model
Empirical
0.0
0.4
0.8
0.0 0.2 0.4 0.6 0.8 1.0
QQ−plot
Empirical
Model
1000000
4000000
1000000 4000000 7000000
Density Plot
Density
Quantile
0.0000000
0.0000010
2000000 6000000
Return Level Plot
Return Level
Return Period (Years)
0
4000000
8000000
1 2 5 20 50 200

```
> par(mfrow = c(2, 2))
> plot(GPD2)
```



The POT package contains much more functionality than Generalized Pareto fitting and there are other EVT packages which can be found on CRAN such as **evir, evd**, etc.

3 Bibliography

1. Christophe Dutang, Vincent Goulet, and Mathieu Pigeon. “actuar: An r package for actuarial science”. *Journal of Statistical Software*

2. Stuart A. Klugman, Harry H. Panjer, and Gordon E. Willmot. *Loss models: from data to decisions* Wiley series in probability and statistics, New York, NY, 1998.

3. Alexander J. McNeil. “Estimating the tails of loss severity distributions using extreme value theory.” *ASTIN Bulletin*, 27(1):117–138, May 1997.

4. Mathieu Ribatet. *POT: Generalized Pareto Distribution and Peaks Over Threshold*, 2009. R package version 1.1-0.

4 Legal