Modeling extreme values with a GEV mixture probability distributions

Standard Gumbel distribution

Pascal Alain Dkengne Sielenou

2023-10-28

```r
# Load useful functions
```

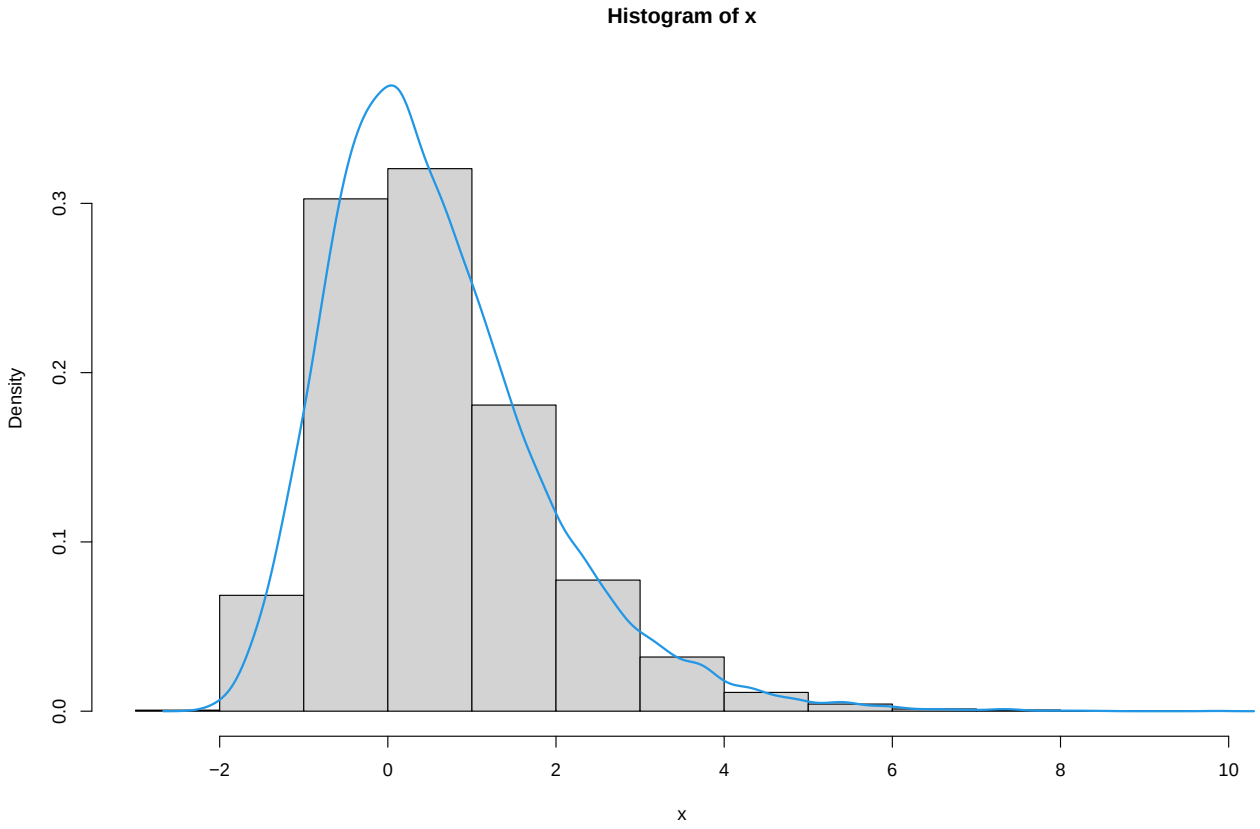
```r
path <- ".."
```

```r
xfun::in_dir(dir = path, expr = source("./src/generate_gev_sample.R"))
xfun::in_dir(dir = path, expr = source("./src/calculate_gev_inverse_cdf.R"))
xfun::in_dir(dir = path, expr = source("./src/estimate_gev_mixture_model_parameters.R"))
xfun::in_dir(dir = path, expr = source("./src/plot_gev_mixture_model_pdf.R"))
xfun::in_dir(dir = path, expr = source("./src/plot_gev_mixture_model_cdf.R"))
xfun::in_dir(dir = path, expr = source("./src/estimate_gev_mixture_model_quantile.R"))


# Generate a random sample
```

```r
n <- 20000
```

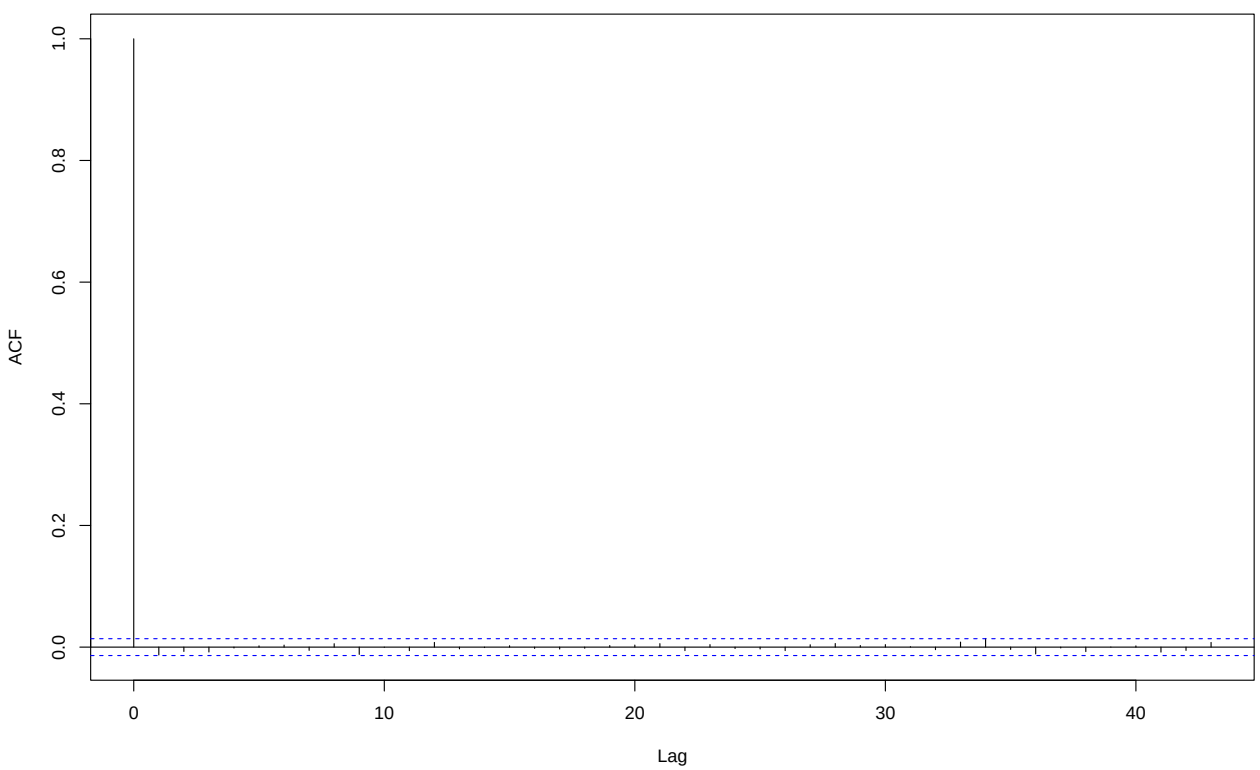
```r
loc <- 0
scale <- 1
shape <- 0

set.seed(1122)
x <- generate_gev_sample(n = n, loc = loc, scale = scale, shape = shape)

# Histogram of all data
```

```r
dens_x <- density(x)
hist(x, prob = TRUE, ylim = range(dens_x$y))
lines(dens_x, lwd = 2, col = 4)
```



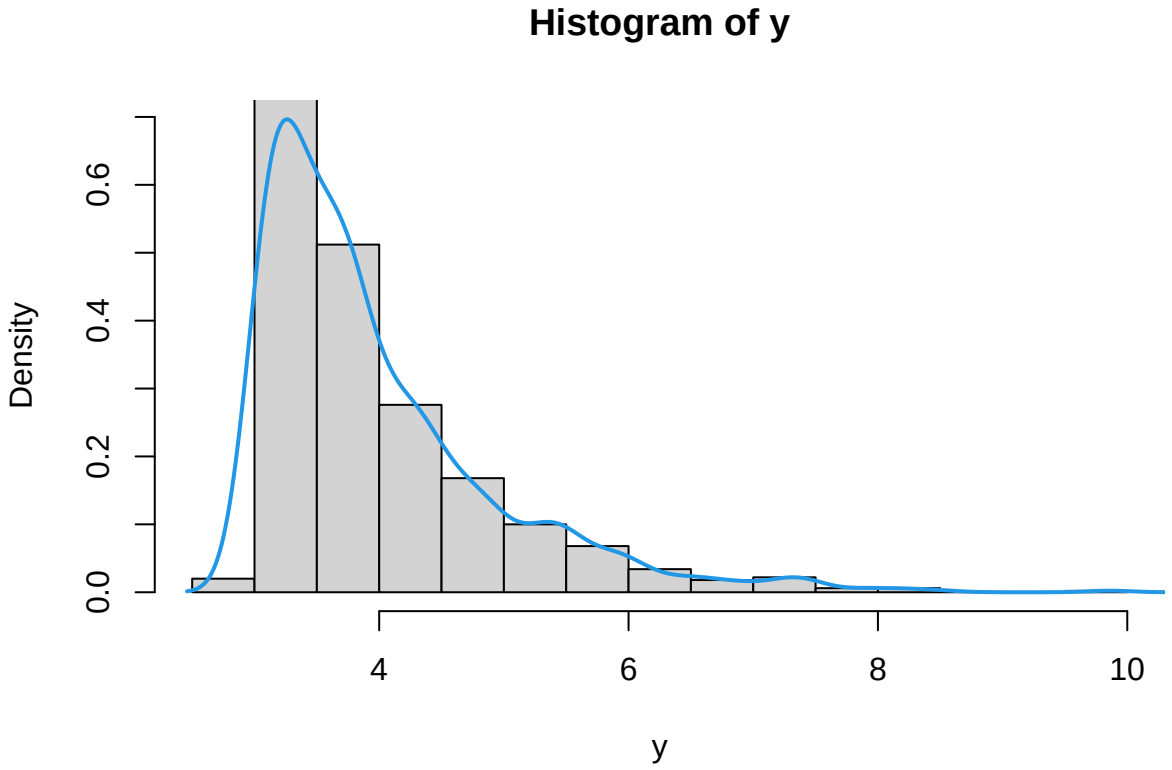

```
# Autocorrelation function of all data
```

```
acf(x)
```



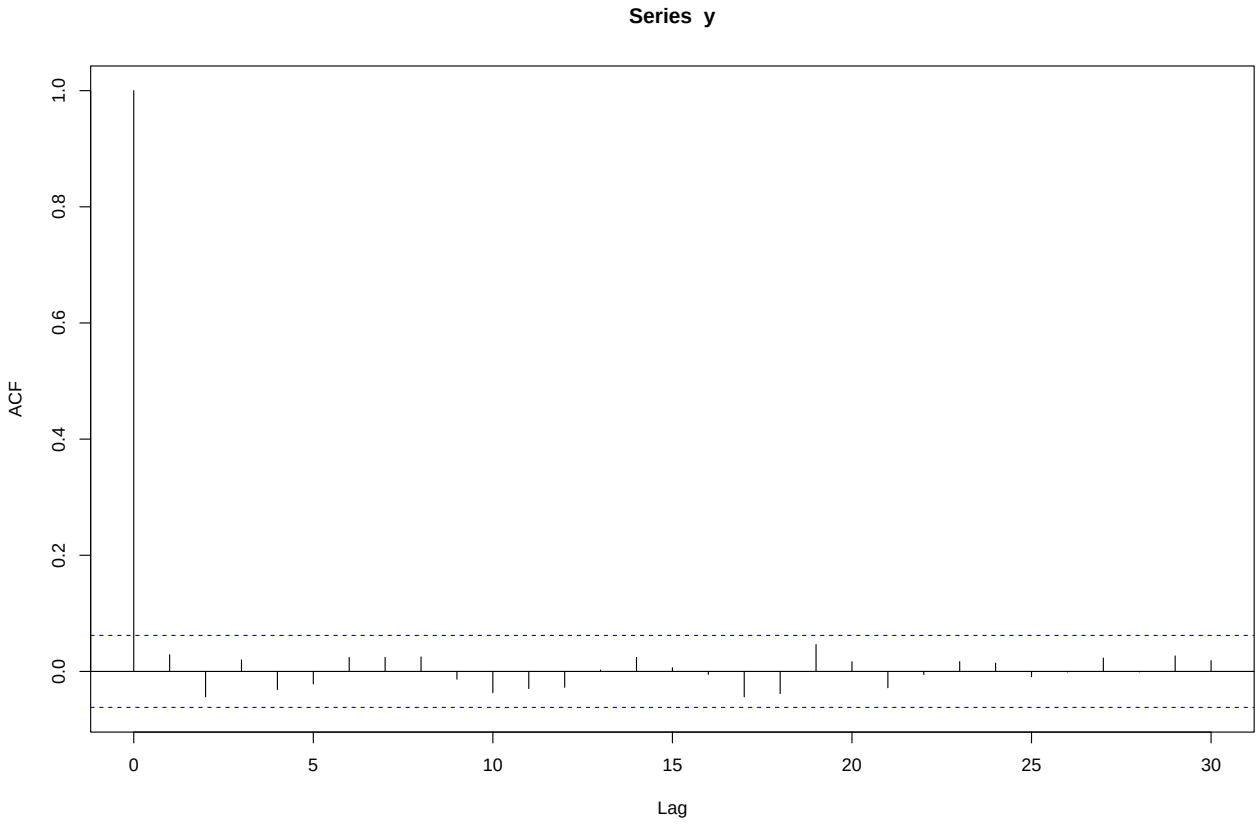

```
# Histogram of the largest data
```

```
nlargest <- 1000
y <- extract_nlargest_sample(x, n = nlargest)
dens_y <- density(y)
hist(y, prob = TRUE, ylim = range(dens_y$y))
lines(density(y), lwd = 2, col = 4)
```




```
# Autocorrelation function of the largest data
```

```
acf(y)
```



```r
# Estimation of gev mixture models
```

```r
gev_mixture_model <- estimate_gev_mixture_model_parameters(x = x,
                                                           block_sizes = NULL,
                                                           minimum_nblocks = 50,
                                                           threshold = NULL,
                                                           nlargest = nlargest,
                                                           confidence_level = 0.95,
                                                           use_extremal_index = TRUE,
                                                           use_lower_threshold = FALSE,
                                                           maximum_iterations = 1500,
                                                           log_mv = TRUE,
                                                           log_pw = TRUE,
                                                           trace = FALSE,
                                                           method = "MLE")
```

```
##   Successful convergence.
##   Successful convergence.
```

```r
gev_mixture_model$extremal_indexes
```

```
##           13           14           15           16           17           18
## 0.8920296849 0.9412167819 0.9109104376 0.9663117674 1.0000000000 1.0000000000
##           19           20
## 0.9153737685 1.0000000000
```

```r
gev_mixture_model$normalized_gev_parameters_object
```

```
##       loc_star   scale_star     shape_star
## 13 2.738730574 1.2972204786 -0.11094214484
## 14 3.173500173 1.0537637674 -0.06604485185
## 15 2.962237100 1.1338401059 -0.07956577765
## 16 2.943030605 1.1978828474 -0.09487103167
## 17 2.640938537 1.3340390618 -0.11772435182
## 18 3.339054187 0.9284680982 -0.02948815858
## 19 2.200619920 1.5943004373 -0.15367097316
## 20 3.387462307 0.9167513605 -0.02821859821
```

```r
gev_mixture_model$full_normalized_gev_parameters_object
```

```
##       loc_star   scale_star     shape_star
## 13 2.589572170 1.3137684319 -0.11094214484
## 14 3.109533392 1.0579884440 -0.06604485185
## 15 2.856043969 1.1422894450 -0.07956577765
## 16 2.901913848 1.2017836365 -0.09487103167
## 17 2.640938537 1.3340390618 -0.11772435182
## 18 3.339054187 0.9284680982 -0.02948815858
## 19 2.058685279 1.6161116717 -0.15367097316
## 20 3.387462307 0.9167513605 -0.02821859821
```

```r
gev_mixture_model$automatic_weights_pw_shape
```

```
##           13           14           15           16           17           18
## 0.1340587202 0.1178963789 0.1227637377 0.1282733987 0.1365001850 0.1057314779
##           19           20
## 0.1494401683 0.1053359333
```

```
gev_mixture_model$automatic_weights_pw_scale
```

```
##           13           14           15           16           17           18
## 0.1168023571 0.1315041849 0.1253806495 0.1224047623 0.1157881200 0.1427175822
##           19           20
## 0.1016704514 0.1437318925
```

```
gev_mixture_model$automatic_weights_pw_loc
```

```
##           13           14           15           16           17           18
## 0.1161439230 0.1315013058 0.1229014897 0.1244577977 0.1169475723 0.1392863559
##           19           20
## 0.1078334539 0.1409281017
```

```
gev_mixture_model$weighted_normalized_gev_parameters_object[3, ]
```

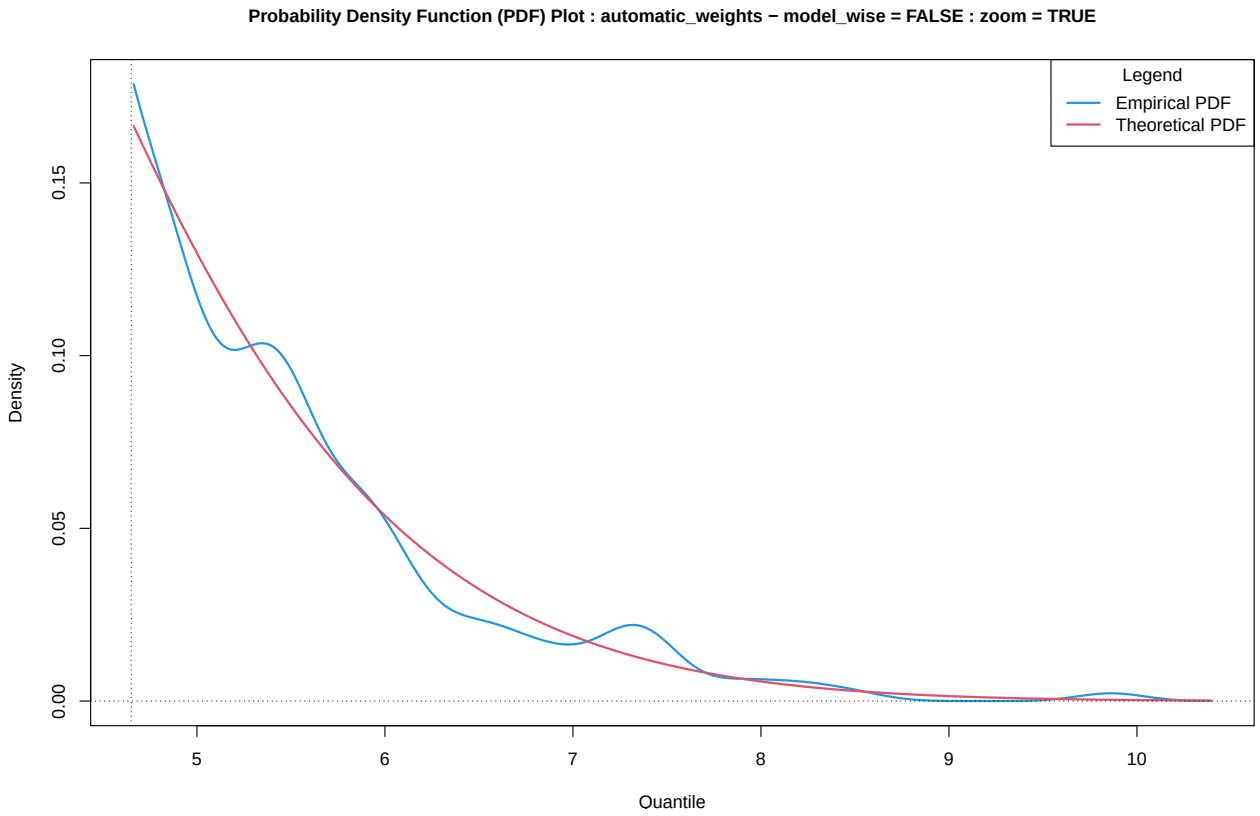
```
##                     loc_star  scale_star     shape_star
## automatic_weights 2.95437791 1.159715179 -0.08972070362
```

```
gev_mixture_model$automatic_weights_mw
```

```
##            13            14            15            16            17
## 0.14900807916 0.30984800794 0.33393783784 0.16664609783 0.04055997724
##            18            19            20
## 0.00000000000 0.00000000000 0.00000000000
```

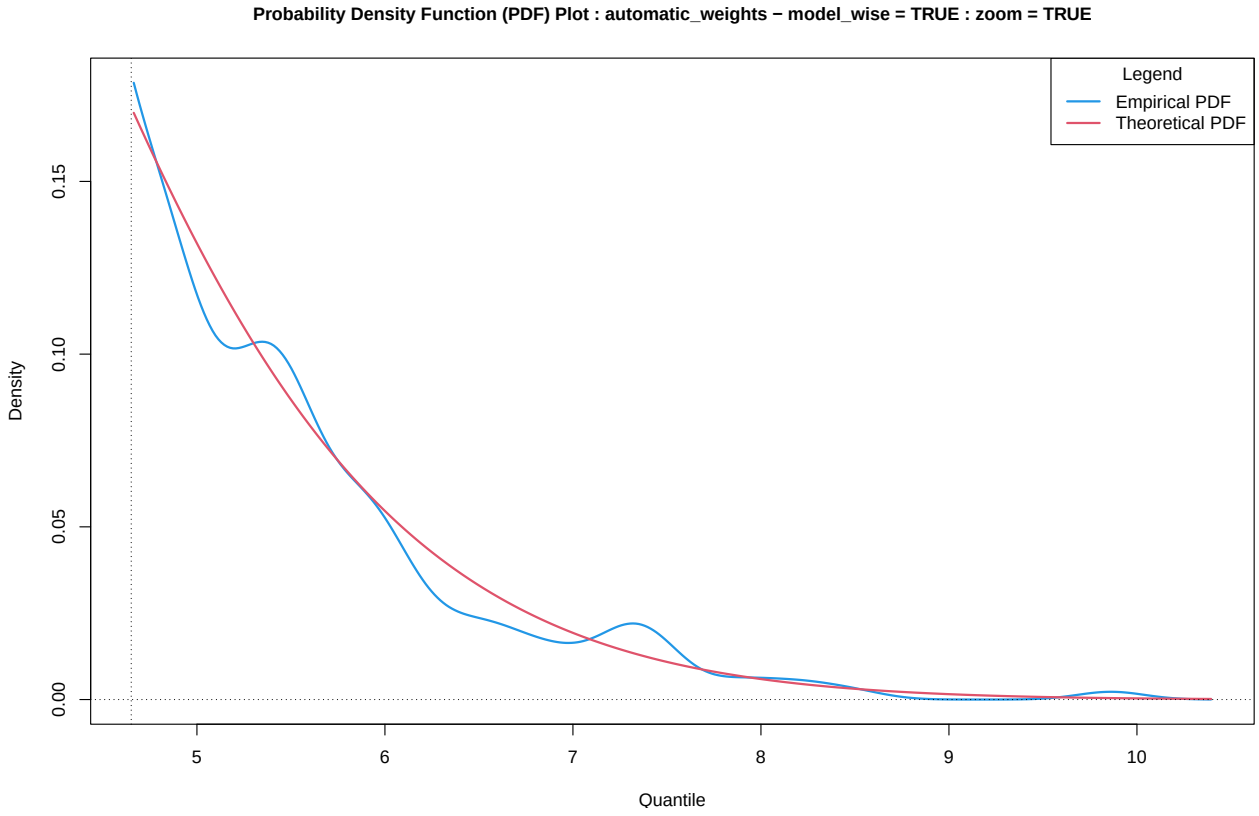
```
# Model diagnostics

## GEV mixture model with respect to parameters
```

```
plot_gev_mixture_model_pdf(gev_mixture_model,
                           type = "automatic_weights",
                           model_wise = FALSE,
                           zoom = TRUE,
                           xlab = "Quantile",
                           ylab = "Density",
                           main = "Probability Density Function (PDF) Plot")
```




```
## GEV mixture model with respect to distribution functions
```

```
plot_gev_mixture_model_pdf(gev_mixture_model,
                           type = "automatic_weights",
                           model_wise = TRUE,
                           zoom = TRUE,
                           xlab = "Quantile",
                           ylab = "Density",
                           main = "Probability Density Function (PDF) Plot")
```




```
# Estimation of an extreme quantile
```

```
estimator_types <- c("automatic_weights_mw",
                     "pessimistic_weights_mw",
                     "identic_weights_mw",
                     "automatic_weights_pw",
                     "pessimistic_weights_pw",
                     "identic_weights_pw",
                     "model_wise",
                     "parameter_wise",
                     "empirical")
```

```
alpha <- 10^(-14)

## Quantile from the true distribution
```

```
true_rl <- calculate_gev_inverse_cdf(p = 1 - alpha,
                                     loc = loc,
                                     scale = scale,
                                     shape = shape)

true_rl
```

```
## [1] 32.2369909
```

```
## Quantile from GEV mixture model with respect to parameters
```

```
rl_pw <- estimate_gev_mixture_model_quantile(gev_mixture_model,
                                             alpha = alpha,
                                             confidence_level = 0.95,
```

```
                                            do.ci = TRUE,
                                            estimator_type = estimator_types[4])

rl_pw
```

```
## [1] 14.94246142
```

```
## Quantile from GEV mixture model with respect to distribution functions
```

```
rl_mw <- estimate_gev_mixture_model_quantile(gev_mixture_model,
                                            alpha = alpha,
                                            confidence_level = 0.95,
                                            do.ci = TRUE,
                                            estimator_type = estimator_types[1])

rl_mw
```

```
## [1] 16.61948946
```

```
## Quantiles from equivalent estimated distributions in GEV mixture model with respect to parameters
```

```
est_rl_pw <- suppressWarnings(estimate_gev_mixture_model_quantile(gev_mixture_model,
                                                                  alpha = alpha,
                                                                  confidence_level = 0.95,
                                                                  do.ci = TRUE,
                                                                  estimator_type = estimator_types[8]))

est_rl_pw
```

```
##             lower    quantile       upper
## 13   2.3468996728 13.96958275 25.59226582
## 14  -3.7474302756 16.80635608 37.36014243
## 15  -2.1419903358 15.81091531 33.76382096
## 16  -0.5952710422 14.77892345 30.15311794
## 17   2.3700405431 13.61030101 24.85056147
## 18 -23.8227373416 21.53142934 66.88559603
## 19   3.6014287009 12.45779173 21.31415476
## 20 -26.7220637170 21.63967153 70.00140678
```

```
## Comparison of estimated quantiles
```

```
est_rl_pw_range <- range(as.matrix(est_rl_pw))


## Quantiles from equivalent estimated GEV distributions in GEV mixture model respect to distribution f
```

```
est_rl_mw <- suppressWarnings(estimate_gev_mixture_model_quantile(gev_mixture_model,
                                                                  alpha = alpha,
                                                                  confidence_level = 0.95,
                                                                  do.ci = TRUE,
                                                                  estimator_type = estimator_types[7]))
est_rl_mw
```

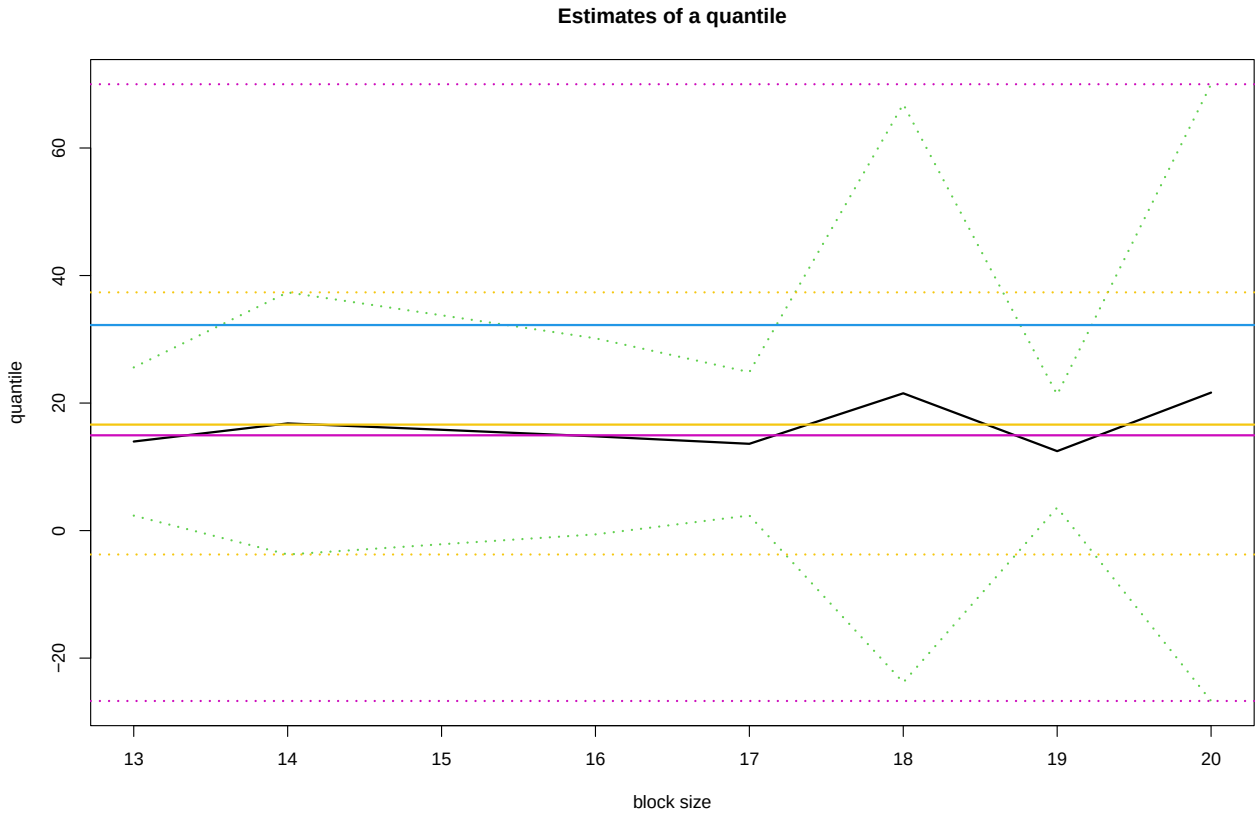
```
##            lower    quantile       upper
## 13  2.3468996728 13.96958275 25.59226582
## 14 -3.7474302756 16.80635608 37.36014243
## 15 -2.1419903358 15.81091531 33.76382096
```

```
## 16 -0.5952710422 14.77892345 30.15311794
## 17  2.3700405431 13.61030101 24.85056147
```

```
est_rl_mw_range <- range(as.matrix(est_rl_mw))

est_rl_mw_range
```

```
## [1] -3.747430276 37.360142435
```

```
matplot(x = rownames(est_rl_pw),
        y = est_rl_pw,
        xlab = "block size",
        ylab = "quantile",
        main = "Estimates of a quantile",
        ylim = range(c(est_rl_pw_range, true_rl, rl_pw)),
        cex = 1,
        cex.lab = 1,
        cex.axis = 1,
        type = "l",
        lty = c("dotted", "solid", "dotted"),
        lwd = c(2,2,2),
        col = c(3, 1, 3))

abline(h = true_rl, col = 4, lwd = 2)
abline(h = rl_mw, col = 7, lwd = 2)
abline(h = rl_pw, col = 6, lwd = 2)
abline(h = est_rl_pw_range, col = 6, lty = "dotted", lwd = 2)
abline(h = est_rl_mw_range, col = 7, lty = "dotted", lwd = 2)
```



```
# Legend:
# blue: Quantile from the true distribution
# yellow: Quantile from GEV mixture model with respect to distribution functions
# pink: Quantile from GEV mixture model with respect to parameters
```