Modeling extreme values with a GEV mixture probability distributions

Application to a wind speed data

Pascal Alain Dkengne Sielenou

2023-10-26

```r
# library(xfun)

path <- ".."

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"))

library(readr)

vent <- xfun::in_dir(dir = path, expr = read_csv("./applications/vent.csv"))
```
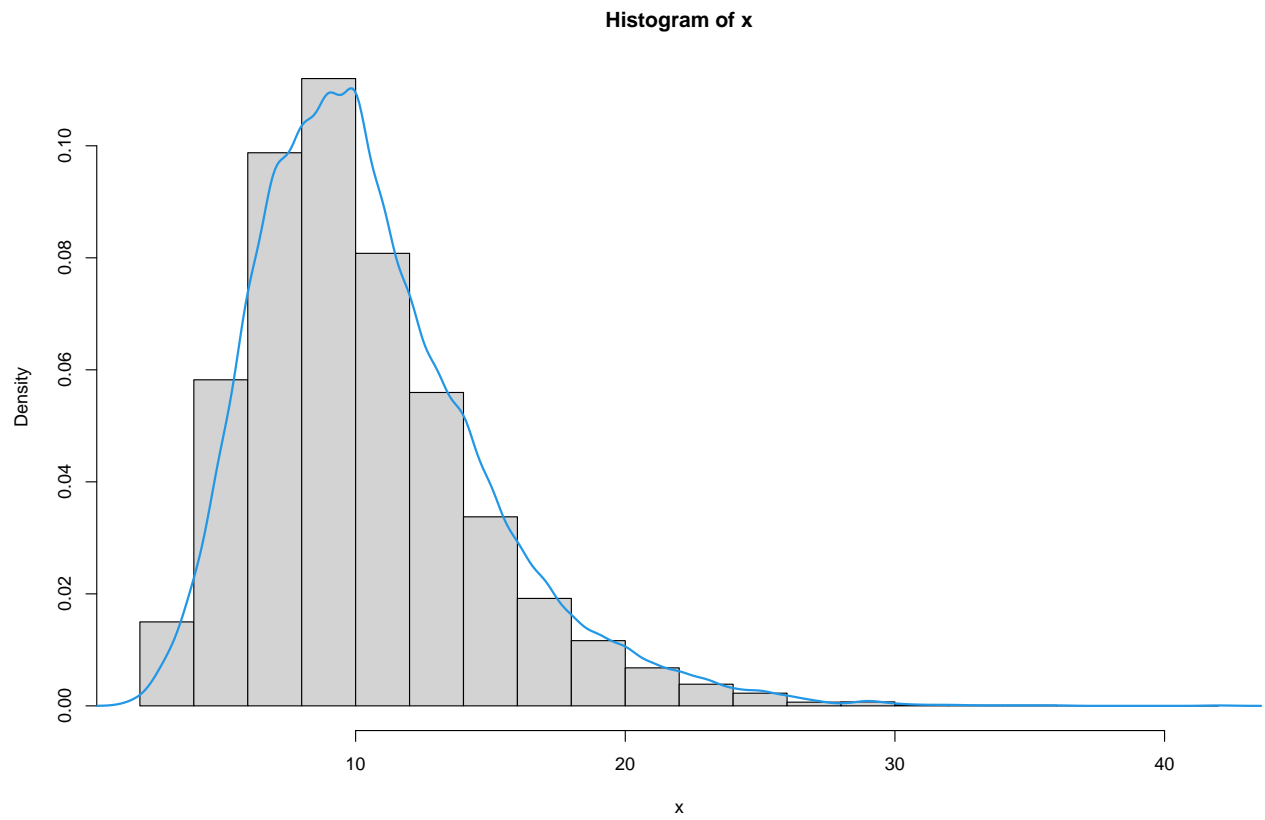
```
## Rows: 10627 Columns: 2
## -- Column specification --------------------------------------------------
## Delimiter: ","
## dbl  (1): Vent
## date (1): Date
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```
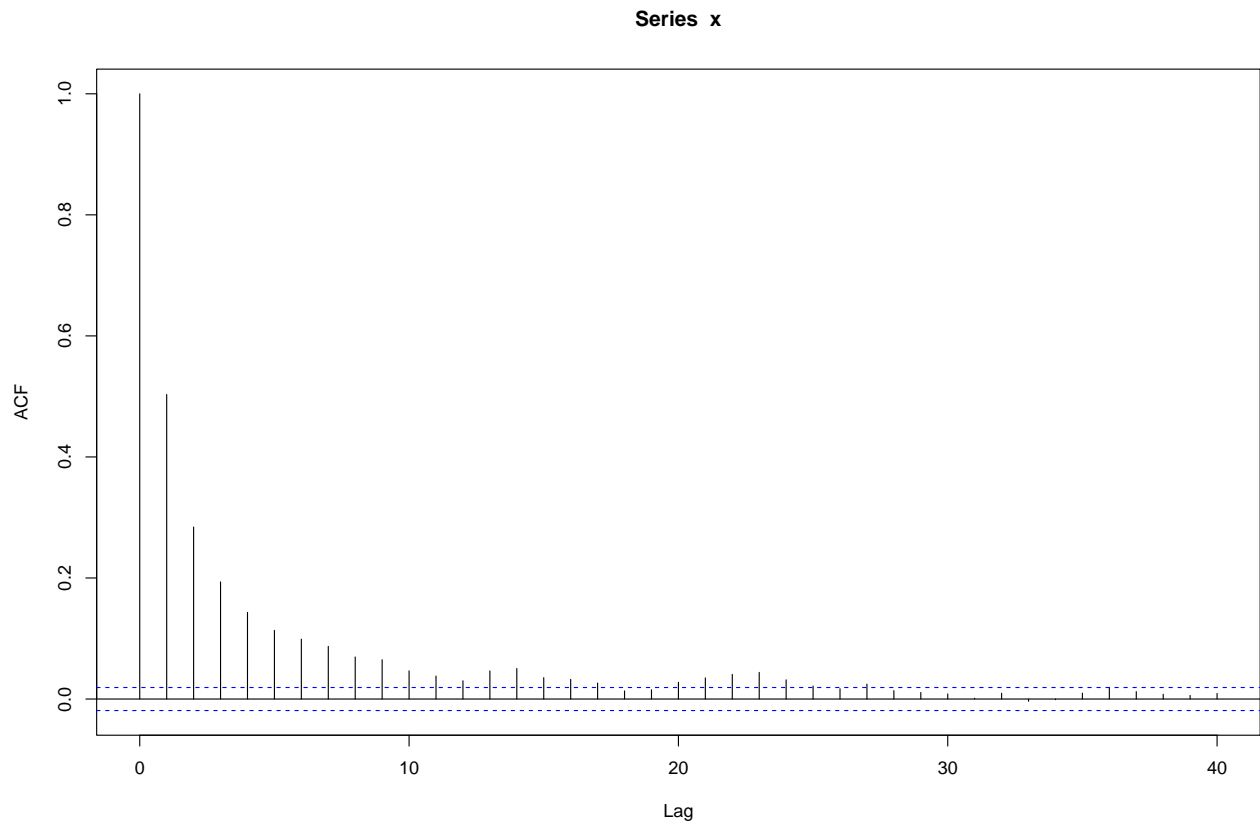
```r
x <- vent$Vent
x <- x[!is.na(x)]
n <- length(x)

n
```

```
## [1] 10607
```

```r
# Histogram of all data

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

Histogram of x


```r
# Autocorrelation function of all data
acf(x)
```

Series x
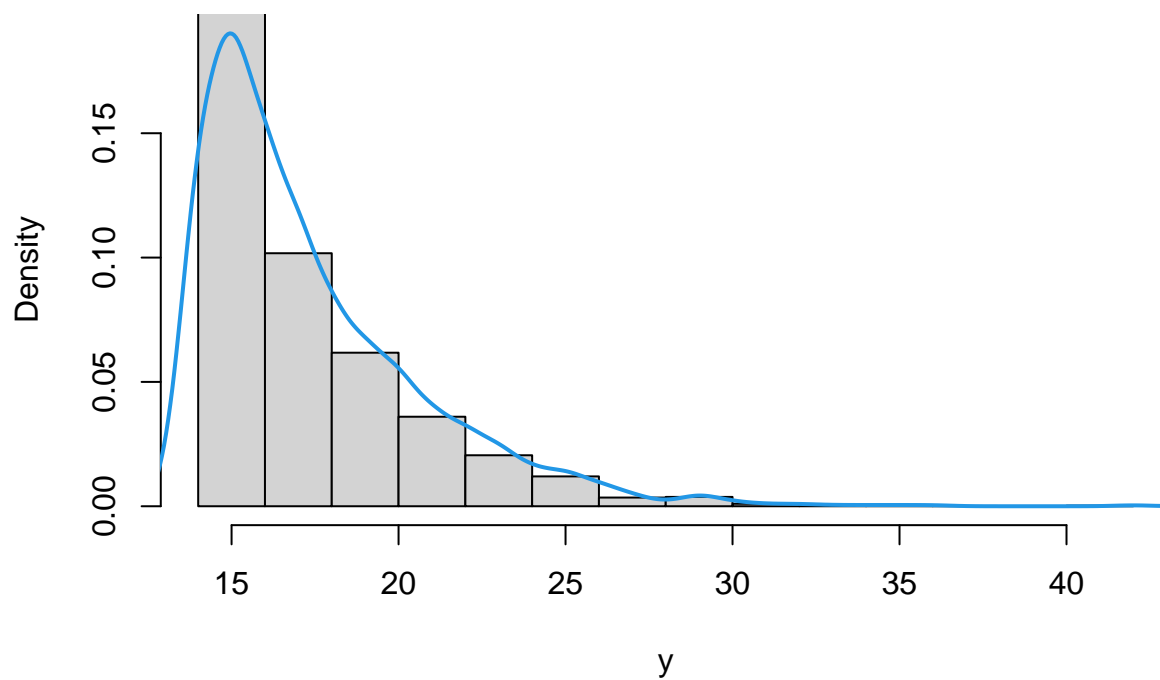


```r
# Histogram of the largest data
```

```r
nlargest <- 2000
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)
```
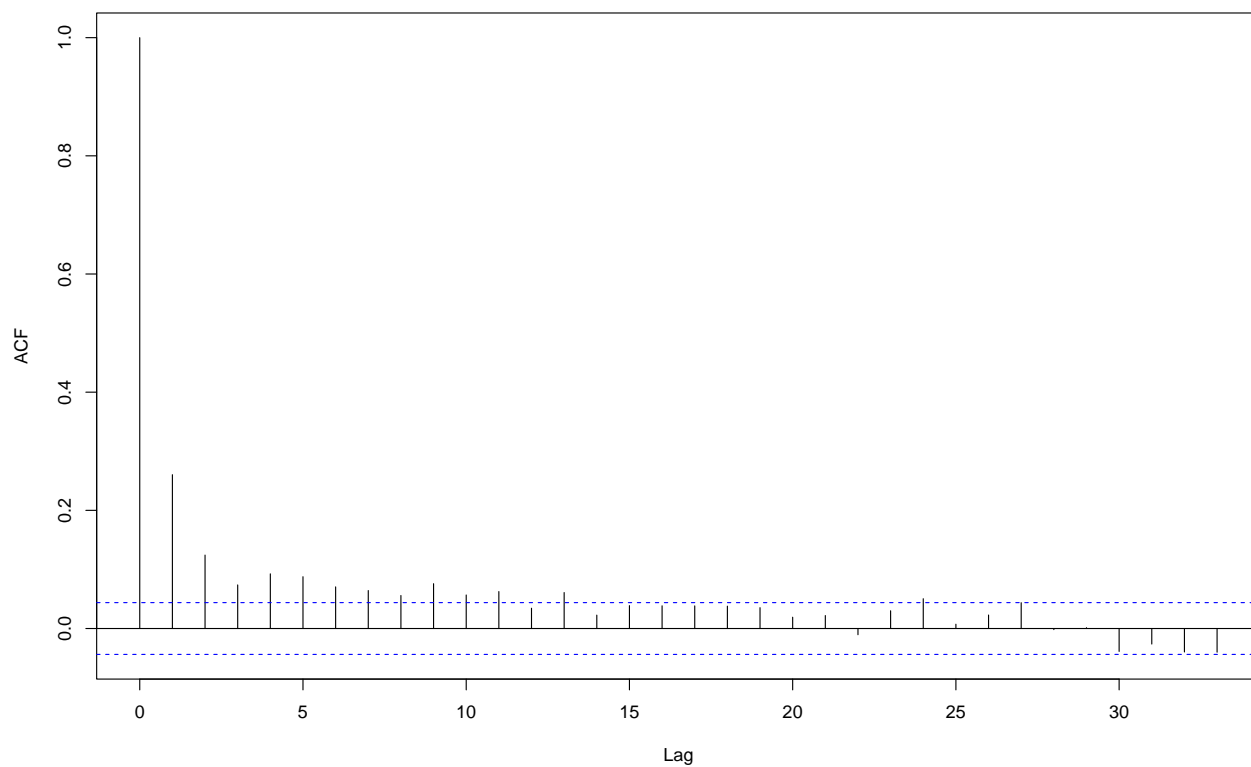
Histogram of y


```r
# Autocorrelation function of the largest data

acf(y)
```

Series y

```r
# Estimation of gev mixture models

gev_mixture_model <- suppressWarnings(estimate_gev_mixture_model_parameters(x = x,
                                               block_sizes = 10:40,
                                               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
```

```
##           10           11           12           13           14           15
## 0.6691287041 0.6691287041 0.6691287041 0.7116100064 0.7116100064 0.7083622670
##           16           17           18           19           20           21
## 0.6691287041 0.7235605139 0.6691287041 0.7083622670 0.7235605139 0.7083622670
##           22           23           24           25           26           27
## 0.7235605139 0.6882408990 0.7246083284 0.7235605139 0.6379524564 0.6691287041
##           28           29           30           31           32           33
## 0.6882408990 0.7235605139 0.6882408990 0.6379524564 0.6882408990 0.6379524564
##           34           35           36           37           38           39
## 0.6882408990 0.7083622670 0.6379524564 0.6691287041 0.6379524564 0.6879635631
##           40
## 0.6379524564
```

```r
gev_mixture_model$normalized_gev_parameters_object
```

```
##         loc_star  scale_star        shape_star
## 10 13.905018483 3.229445814 -0.0134919705477
## 11 14.072305538 3.074366538  0.0004598492663
## 12 13.513064598 3.434241708 -0.0277638117638
## 13 14.047279272 3.181889444 -0.0111302750829
## 14 13.731210944 3.176589461 -0.0042148568548
## 15 13.735024480 3.270901478 -0.0197068714746
## 16 12.938440341 3.664933126 -0.0412137186479
## 17 13.757458005 2.953168946  0.0213357541128
## 18 12.530733728 3.783643557 -0.0454855670742
## 19 12.207454222 3.916602815 -0.0500647997067
## 20 13.339404221 3.386243511 -0.0222016275348
## 21 12.120283966 3.906377152 -0.0533039821013
## 22 12.528390898 3.658864734 -0.0349411538984
## 23 13.584349251 3.118116705  0.0001124426024
## 24 13.554242389 3.240016830 -0.0182563959382
## 25 12.795352420 3.669494229 -0.0379260347583
## 26 14.978565180 2.628178877  0.0304036700219
## 27 10.228845186 4.593014880 -0.0769598337587
## 28 12.414671325 3.512081028 -0.0144312981038
```

```
## 29 12.492112381 3.837265211 -0.0474754050521
## 30 13.017878862 3.465678541 -0.0271683273755
## 31 14.581985396 2.526069010  0.0524553323527
## 32 12.440046281 3.655038452 -0.0329641740013
## 33 13.900613980 2.897107860  0.0175588295455
## 34 10.571721606 4.405852497 -0.0597603934781
## 35  7.956573550 5.758776233 -0.1175170546092
## 36 14.591645767 2.681754521  0.0306909008983
## 37  7.925418866 5.304655126 -0.0924866381669
## 38 14.338779818 2.813414984  0.0174012536989
## 39 10.898293258 4.099321612 -0.0442725396631
## 40 12.046054094 3.729905045 -0.0315836341529
```

gev_mixture_model$full_normalized_gev_parameters_object

```
##          loc_star  scale_star       shape_star
## 10 12.603972281 3.246999491 -0.0134919705477
## 11 12.837204173 3.073798577  0.0004598492663
## 12 12.125534397 3.472764836 -0.0277638117638
## 13 12.962667796 3.193961468 -0.0111302750829
## 14 12.649679691 3.181147960 -0.0042148568548
## 15 12.603378460 3.293202680 -0.0197068714746
## 16 11.453688781 3.726125259 -0.0412137186479
## 17 12.805188733 2.932851563  0.0213357541128
## 18 10.996569923 3.853425868 -0.0454855670742
## 19 10.845287729 3.984799407 -0.0500647997067
## 20 12.239768642 3.410657211 -0.0222016275348
## 21 10.760912723 3.978837052 -0.0533039821013
## 22 11.337770161 3.700466397 -0.0349411538984
## 23 12.419394313 3.117985715  0.0001124426024
## 24 12.507480291 3.259126933 -0.0182563959382
## 25 11.600694862 3.714802853 -0.0379260347583
## 26 13.805256657 2.592505992  0.0304036700219
## 27  8.354642295 4.737253223 -0.0769598337587
## 28 11.098956573 3.531068500 -0.0144312981038
## 29 11.240898510 3.896667096 -0.0474754050521
## 30 11.716450747 3.501036166 -0.0271683273755
## 31 13.459820122 2.467205457  0.0524553323527
## 32 11.066020272 3.700332084 -0.0329641740013
## 33 12.603514014 2.874332303  0.0175588295455
## 34  8.907108905 4.505330407 -0.0597603934781
## 35  5.930172098 5.996912963 -0.1175170546092
## 36 13.394496360 2.645012927  0.0306909008983
## 37  5.754026888 5.505479871 -0.0924866381669
## 38 13.079106471 2.791495089  0.0174012536989
## 39  9.352302896 4.167766532 -0.0442725396631
## 40 10.357536185 3.783234577 -0.0315836341529
```

gev_mixture_model$automatic_weights_pw_shape

```
##                10              11              12              13
##  4.235164736e-22 -4.102815838e-22  0.000000000e+00 -2.117582368e-22
##                14              15              16              17
##  3.176373552e-22  4.235164736e-22  0.000000000e+00  8.470329473e-22
##                18              19              20              21
```

6

```
##   0.000000000e+00  8.470329473e-22 -8.470329473e-22  0.000000000e+00
##              22               23               24               25
##   2.541098842e-21  1.091878409e-22  1.270549421e-21  8.470329473e-22
##              26               27               28               29
##   8.470329473e-22  0.000000000e+00  4.235164736e-22 -8.470329473e-22
##              30               31               32               33
##  -4.235164736e-22  0.000000000e+00  3.388131789e-21 -8.470329473e-22
##              34               35               36               37
##   0.000000000e+00  8.610244493e-01  0.000000000e+00  1.389755507e-01
##              38               39               40
##   1.270549421e-21  8.470329473e-22 -8.470329473e-22
```

gev_mixture_model$automatic_weights_pw_scale

```
##            10             11             12             13             14
## 0.02351943034 0.03270005652 0.01671266845 0.02618764420 0.02688240807
##            15             16             17             18             19
## 0.02121422195 0.01168295735 0.04064853955 0.01177619640 0.01356255390
##            20             21             22             23             24
## 0.01807328959 0.01347855849 0.01215204271 0.03030596471 0.02291440624
##            25             26             27             28             29
## 0.01188995873 0.06240450377 0.02961999580 0.01546023093 0.01236373523
##            30             31             32             33             34
## 0.01609282915 0.07174912152 0.01215449413 0.04408379337 0.02255001148
##            35             36             37             38             39
## 0.18980191455 0.05882746616 0.05516230168 0.04903841985 0.01615965030
##            40
## 0.01083063486
```

gev_mixture_model$automatic_weights_pw_loc

```
##              10              11              12              13              14
## 0.000000000e+00 0.000000000e+00 0.000000000e+00 0.000000000e+00 0.000000000e+00
##              15              16              17              18              19
## 0.000000000e+00 0.000000000e+00 0.000000000e+00 0.000000000e+00 0.000000000e+00
##              20              21              22              23              24
## 0.000000000e+00 0.000000000e+00 0.000000000e+00 0.000000000e+00 0.000000000e+00
##              25              26              27              28              29
## 0.000000000e+00 9.957417400e-01 0.000000000e+00 0.000000000e+00 0.000000000e+00
##              30              31              32              33              34
## 0.000000000e+00 4.236166507e-03 0.000000000e+00 0.000000000e+00 0.000000000e+00
##              35              36              37              38              39
## 0.000000000e+00 1.864267613e-05 3.388131789e-21 3.450805123e-06 0.000000000e+00
##              40
## 0.000000000e+00
```

gev_mixture_model$weighted_normalized_gev_parameters_object[3,]

```
##                      loc_star   scale_star    shape_star
## automatic_weights 14.97687578  3.801528959  -0.1140384387
```

gev_mixture_model$automatic_weights_mw

```
##              10              11              12              13              14
## 0.000000000e+00 0.000000000e+00 0.000000000e+00 1.000000000e+00 0.000000000e+00
##              15              16              17              18              19
## 8.673617380e-17 0.000000000e+00 0.000000000e+00 0.000000000e+00 0.000000000e+00
```

```
##                 20               21               22               23               24
## 0.000000000e+00 0.000000000e+00 0.000000000e+00 0.000000000e+00 0.000000000e+00
##                 25               26               27               28               29
## 0.000000000e+00 0.000000000e+00 0.000000000e+00 0.000000000e+00 0.000000000e+00
##                 30               31               32               33               34
## 0.000000000e+00 0.000000000e+00 0.000000000e+00 0.000000000e+00 0.000000000e+00
##                 35               36               37               38               39
## 0.000000000e+00 0.000000000e+00 1.734723476e-18 0.000000000e+00 0.000000000e+00
##                 40
## 0.000000000e+00
```
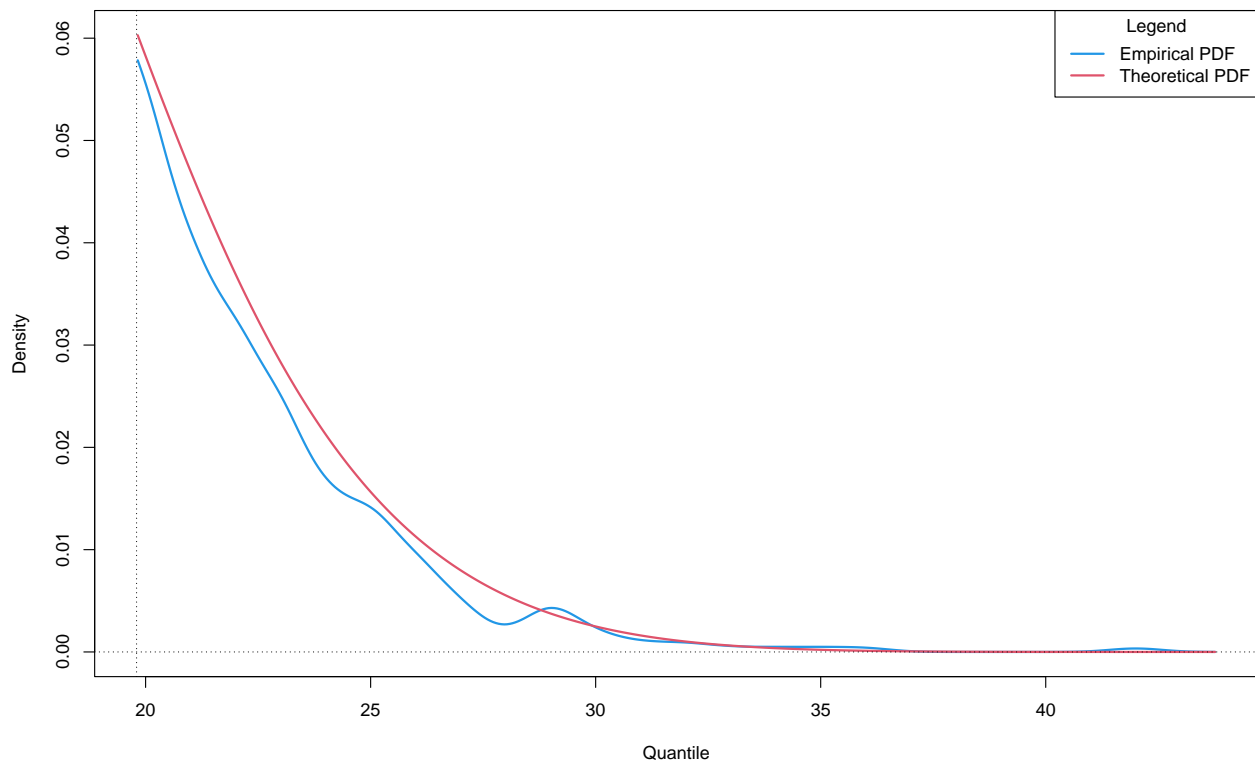
```r
# Model diagnostics
```

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

```r
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")
```

Probability Density Function (PDF) Plot : automatic_weights – model_wise = FALSE : zoom = TRUE


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

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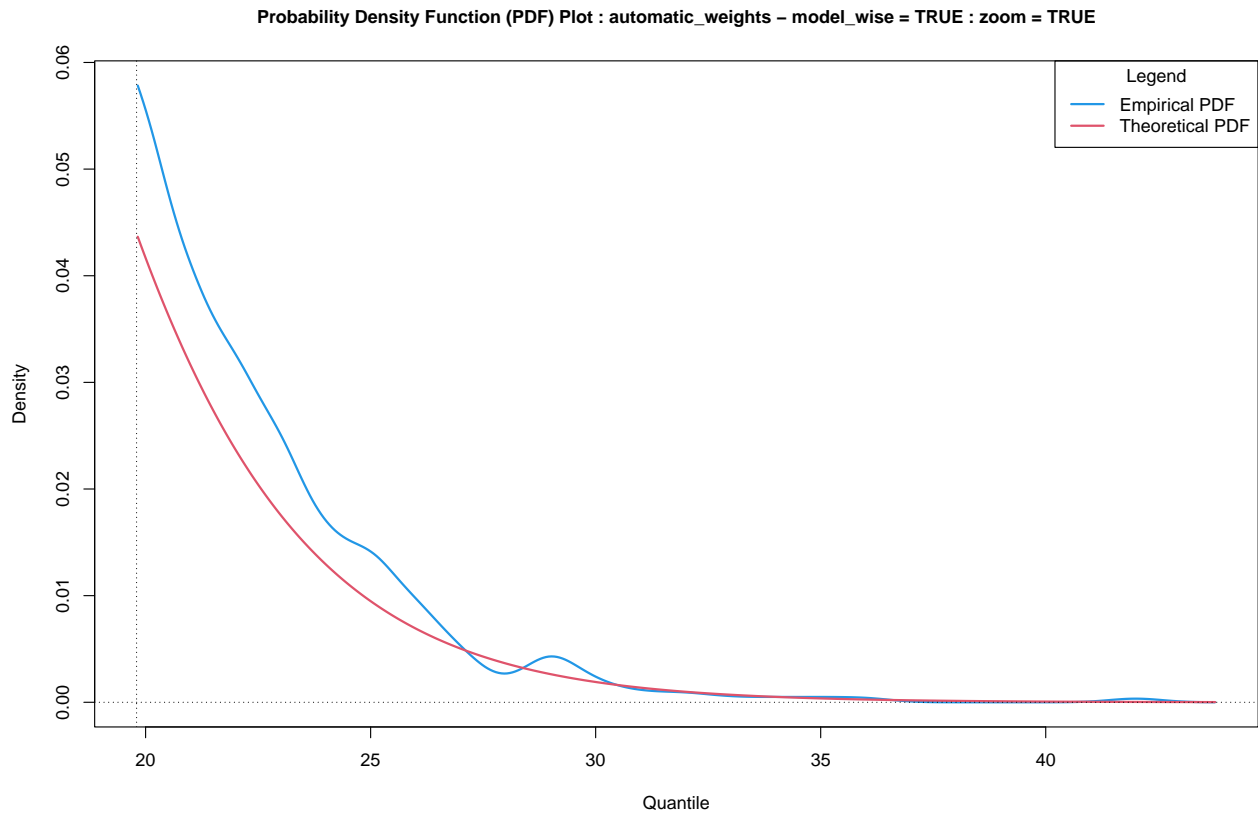
```
                         xlab = "Quantile",
                         ylab = "Density",
                         main = "Probability Density Function (PDF) Plot")
```

Probability Density Function (PDF) Plot : automatic_weights – model_wise = TRUE : zoom = TRUE

```
# 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^(-6)
```

```
## 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] 40.06544993
```

9

```
## 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] 49.53394665
```

```
## 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
## 10 29.635635427 49.26229854  68.88896166
## 11 27.714632796 50.59236926  73.47010573
## 12 28.667306226 48.18340851  67.69951079
## 13 26.561041055 49.53400705  72.50697305
## 14 25.095503649 50.62480013  76.15409661
## 15 27.444997250 48.44082926  69.43666128
## 16 27.400017896 47.28981027  67.17960264
## 17 18.038817613 53.85854943  89.67828125
## 18 26.704335199 47.18327910  67.66222299
## 19 26.213905939 47.32994974  68.44599354
## 20 22.630614267 48.81640584  75.00219742
## 21 26.160934144 46.54957695  66.93821975
## 22 22.331353595 48.21094033  74.09052707
## 23 17.447159202 50.63587483  83.82459046
## 24 23.012045160 48.27817251  73.54429986
## 25 22.662173414 47.99546811  73.32876281
## 26 14.561755001 52.27931429  89.99687358
## 27 24.656173216 45.92767292  67.19917263
## 28 13.994744870 50.74129594  87.48784702
## 29 22.543583627 47.44020746  72.33683130
## 30 19.367166012 48.19323247  77.01929893
## 31  2.697864918 55.84863537 108.99940583
## 32 17.617136907 48.35627719  79.09541747
## 33  8.797651363 51.88178074  94.96591011
## 34 16.569720352 48.03728630  79.50485224
## 35 25.113376708 44.86290580  64.61243488
## 36  7.644886259 52.72119537  97.79750448
## 37 21.284234802 46.10193289  70.91963097
## 38  8.913632336 51.18720742  93.46078251
## 39 14.304625819 48.75608944  83.20755306
## 40 12.020466288 48.78393827  85.54741026
```

```
## 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 26.56104105 49.53400705 72.50697305
## 15 27.44499725 48.44082926 69.43666128
## 37 21.28423480 46.10193289 70.91963097
```

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

est_rl_mw_range
```

```
## [1] 21.28423480 72.50697305
```

```
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, 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 = 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)
```

Estimates of a quantile
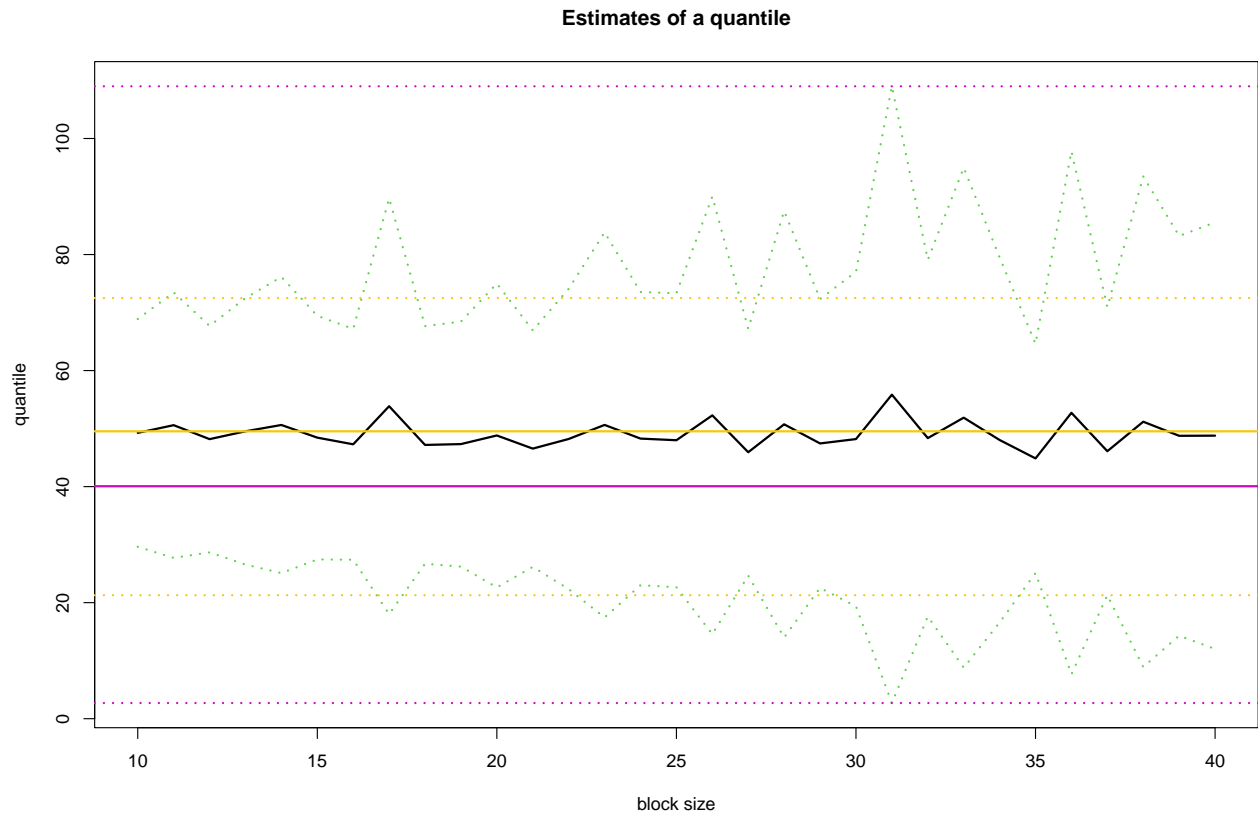

```
# Legend:
# yellow: Quantile from GEV mixture model with respect to distribution functions
# pink: Quantile from GEV mixture model with respect to parameters
```