Package 'weibulltools'

January 29, 2019

Type Package

Title Statistical Methods for Life Data Analysis

Version 1.0.1

Description Contains methods for examining bench test or field data using the well-known Weibull Analysis. It includes Monte Carlo simulation for estimating the life span of products that have not failed, taking account of registering and reporting delays as stated in (Verband der Automobilindustrie e.V. (VDA), 2016, <ISSN:0943-9412>). If the products looked upon are vehicles, the covered mileage can be estimated as well.
It also provides non-parametric estimators like Median Ranks, Kaplan-Meier (Abernethy, 2006, <ISBN:978-0-9653062-3-2>), Johnson (Johnson, 1964, <ISBN:978-0444403223>), and Nelson-Aalen for failure probability estimation within samples that contain failures as well as censored data.
Methods for estimating the parameters of lifetime distributions, like Maximum Likelihood and Median-Rank Regression, (Genschel and Meeker, 2010, <DOI:10.1080/08982112.2010.503447>) as well as the computation of confidence intervals of quantiles and probabilities using the delta method related to Fisher's confidence intervals (Meeker and Escobar, 1998, <ISBN:9780471673279>) and the beta-binomial confidence bounds are also included.
If desired, the data can automatically be divided into subgroups using segmented regression. And if the number of subgroups in a Weibull Mixture Model is known, data can be analyzed using the EM-Algorithm.
Besides the calculation, methods for interactive visualization of the edited data using *plotly* are provided as well. These visualizations include the layout of a probability plot for a specified distribution, the graphical technique of probability plotting and the possibility of adding regression lines and confidence bounds to existing plots.

License GPL-2

Imports dplyr, LearnBayes, magrittr, plotly, Rcpp, sandwich, segmented, SPREDA, survival

LinkingTo Rcpp (>= 0.12.18), RcppArmadillo

Depends R (>= 3.3.0)

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

Suggests ggplot2, knitr, rmarkdown, tidyverse

VignetteBuilder knitr

NeedsCompilation yes

Author Hensel Tim-Gunnar [aut, cre]

Maintainer Hensel Tim-Gunnar <tim-gunnar.hensel@tu-berlin.de>

Repository CRAN

Date/Publication 2019-01-29 16:10:03 UTC

R topics documented:

`calculate_ranks` *Computation of Johnson Ranks*

Description

This function calculates the Johnson ranks which are used to estimate the failure probabilities in case of (multiple) right censored data.

Usage

```
calculate_ranks(f, n_out, n)
```

Arguments

`f`	a numeric vector indicating the number of failed units for a specific realization of the lifetime characteristic.
`n_out`	a numeric vector indicating the number of failed and censored units that have a shorter realization of lifetime characteristic as unit *i*.
`n`	an integer value indicating the sample size.

Value

A numeric vector containing the computed Johnson ranks.

Examples

```
defectives <- c(0, 1, 2, 0, 0, 0, 3, 0, 2, 0)
n_out <- c(0, 2, 4, 8, 9, 11, 12, 16, 20, 22)
n <- 23
johnson_ranks <- calculate_ranks(f = defectives, n_out = n_out, n = n)
```

`confint_betabinom` *Beta Binomial Confidence Bounds for Quantiles and/or Probabilities*

Description

This non-parametric approach calculates confidence bounds for quantiles and/or failure probabilities using a procedure that is similar to that used in calculating median ranks. The location-scale (and threshold) parameters estimated by rank regression are needed.

Usage

```
confint_betabinom(x, event, loc_sc_params, distribution = c("weibull",
  "lognormal", "loglogistic", "normal", "logistic", "sev", "weibull3",
  "lognormal3", "loglogistic3"), bounds = c("two_sided", "lower",
  "upper"), conf_level = 0.95, direction = c("y", "x"))
```

Arguments

`x`	a numeric vector which consists of lifetime data. `x` is used to specify the range of confidence region(s).
`event`	a vector of binary data (0 or 1) indicating whether unit i is a right censored observation (= 0) or a failure (= 1).
`loc_sc_params`	a (named) numeric vector of estimated location and scale parameters for a specified distribution. The order of elements is important. First entry needs to be the location parameter μ and the second element needs to be the scale parameter σ. If a three-parametric model is used the third element is the threshold parameter γ.
`distribution`	supposed distribution of the random variable. The value can be `"weibull"`, `"lognormal"`, `"loglogistic"`, `"normal"`, `"logistic"`, `"sev"` *(smallest extreme value)*, `"weibull3"`, `"lognormal3"` or `"loglogistic3"`. Other distributions have not been implemented yet.
`bounds`	a character string specifying the interval(s) which has/have to be computed. Must be one of "two_sided" (default), "lower" or "upper".
`conf_level`	confidence level of the interval. The default value is `conf_level = 0.95`.
`direction`	a character string specifying the direction of the computed interval(s). Must be either "y" (failure probabilities) or "x" (quantiles).

Value

A data frame containing the lifetime characteristic, interpolated ranks as a function of probabilities, the probabilities which are used to compute the ranks and computed values for the specified confidence bound(s).

References

Meeker, William Q; Escobar, Luis A., Statistical methods for reliability data, New York: Wiley series in probability and statistics, 1998

Examples

```
# Example 1: Beta-Binomial Confidence Bounds for two-parameter Weibull:
obs   <- seq(10000, 100000, 10000)
state <- c(0, 1, 1, 0, 0, 0, 1, 0, 1, 0)

df_john <- johnson_method(x = obs, event = state)

mrr <- rank_regression(x = df_john$characteristic,
                       y = df_john$prob,
                       event = df_john$status,
                       distribution = "weibull",
                       conf_level = .95)

conf_betabin <- confint_betabinom(x = df_john$characteristic,
                                  event = df_john$status,
                                  loc_sc_params = mrr$loc_sc_coefficients,
                                  distribution = "weibull",
                                  bounds = "two_sided",
                                  conf_level = 0.95,
                                  direction = "y")

# Example 2: Beta-Binomial Confidence Bounds for three-parameter Weibull:
# Alloy T7987 dataset taken from Meeker and Escobar(1998, p. 131)
cycles   <- c(300, 300, 300, 300, 300, 291, 274, 271, 269, 257, 256, 227, 226,
              224, 213, 211, 205, 203, 197, 196, 190, 189, 188, 187, 184, 180,
              180, 177, 176, 173, 172, 171, 170, 170, 169, 168, 168, 162, 159,
              159, 159, 159, 152, 152, 149, 149, 144, 143, 141, 141, 140, 139,
              139, 136, 135, 133, 131, 129, 123, 121, 121, 118, 117, 117, 114,
              112, 108, 104, 99, 99, 96, 94)
state <- c(rep(0, 5), rep(1, 67))

df_john2 <- johnson_method(x = cycles, event = state)
mrr_weib3 <- rank_regression(x = df_john2$characteristic,
                       y = df_john2$prob,
                       event = df_john2$status,
                       distribution = "weibull3",
                       conf_level = .95)

conf_betabin_weib3 <- confint_betabinom(x = df_john2$characteristic,
                                  event = df_john2$status,
                                  loc_sc_params = mrr_weib3$loc_sc_coefficients,
                                  distribution = "weibull3",
                                  bounds = "two_sided",
                                  conf_level = 0.95,
                                  direction = "y")
```

---

confint_fisher *Fisher Confidence Bounds for Quantiles and/or Probabilities*

Description

This method computes normal-approximation confidence intervals for quantiles and/or failure probabilities using the delta_method. The required (log-)location-scale (and threshold) parameters and variance-covariance matrix of these need to be estimated by Maximum Likelihood.

Usage

```
confint_fisher(x, event, loc_sc_params, loc_sc_varcov,
  distribution = c("weibull", "lognormal", "loglogistic", "normal",
  "logistic", "sev", "weibull3", "lognormal3", "loglogistic3"),
  bounds = c("two_sided", "lower", "upper"), conf_level = 0.95,
  direction = c("y", "x"))
```

Arguments

x	a numeric vector which consists of lifetime data. x is used to specify the range of confidence region(s).
event	a vector of binary data (0 or 1) indicating whether unit i is a right censored observation (= 0) or a failure (= 1).
loc_sc_params	a (named) numeric vector of estimated (by Maximum Likelihood) location and scale parameters for a specified distribution. The order of elements is important. First entry needs to be the location parameter μ and the second element needs to be the scale parameter σ. If a three-parametric model is used the third element is the threshold parameter γ.
loc_sc_varcov	a (named) numeric matrix of estimated (by Maximum Likelihood) location and scale variances and covariances for a specified distribution. The order of elements is important. First entry of the diagonal needs to be the variance of the location parameter Var(μ) and the second element of the diagonal needs to be the variance of the scale parameter Var(σ). If a three-parametric model is used the third element of the diagonal needs to be the variance of the threshold parameter Var(γ).
distribution	supposed distribution of the random variable. The value can be "weibull", "lognormal", "loglogistic", "normal", "logistic", "sev" *(smallest extreme value)*, "weibull3", "lognormal3" or "loglogistic3". Other distributions have not been implemented yet.
bounds	a character string specifying the interval(s) which has/have to be computed. Must be one of "two_sided" (default), "lower" or "upper".
conf_level	confidence level of the interval. The default value is conf_level = 0.95.
direction	a character string specifying the direction of the computed interval(s). Must be either "y" (failure probabilities) or "x" (quantiles).

Value

A data frame containing the lifetime characteristic, the probabilities, estimated standard errors by the delta method and computed values for the specified confidence bound(s).

Examples

```
obs   <- seq(10000, 100000, 10000)
state <- c(0, 1, 1, 0, 0, 0, 1, 0, 1, 0)
df_john <- johnson_method(x = obs, event = state)
mle <- ml_estimation(x = obs, event = state,
                     distribution = "weibull", conf_level = 0.95)
conf_fish <- confint_fisher(x = df_john$characteristic,
                            event = df_john$status,
                            loc_sc_params = mle$loc_sc_coefficients,
                            loc_sc_varcov = mle$loc_sc_vcov,
                            distribution = "weibull",
                            bounds = "two_sided",
                            conf_level = 0.95,
                            direction = "y")
```

`delta_method` *Delta Method for Parametric Lifetime Distributions*

Description

The Delta Method estimates the standard error for quantities that can be written as non-linear functions of ML estimators like quantiles. I.e. the (log-)location-scale (and threshold) parameters and variance-covariance matrix of these need to be estimated by Maximum Likelihood.

Usage

```
delta_method(p, loc_sc_params, loc_sc_varcov, distribution = c("weibull",
  "lognormal", "loglogistic", "normal", "logistic", "sev", "weibull3",
  "lognormal3", "loglogistic3"), direction = c("y", "x"))
```

Arguments

`p`	a numeric value of a probability or a quantile. If the standard error of quantile is of interest a specific probability needs to be supplied and if the standard error of a standardized quantile (z-value) should be calculated a specific quantile should be provided.
`loc_sc_params`	a (named) numeric vector of estimated (by Maximum Likelihood) location and scale parameters for a specified distribution. The order of elements is important. First entry needs to be the location parameter μ and the second element needs to be the scale parameter σ. If a three-parametric model is used the third element is the threshold parameter γ.
`loc_sc_varcov`	a (named) numeric matrix of estimated (by Maximum Likelihood) location and scale variances and covariances for a specified distribution. The order of elements is important. First entry of the diagonal needs to be the variance of the location parameter Var(μ) and the second element of the diagonal needs to be the variance of the scale parameter Var(σ). If a three-parametric model is used the third element of the diagonal needs to be the variance of the threshold parameter Var(γ).

distribution	supposed distribution of the random variable. The value can be "weibull", "lognormal", "loglogistic", "normal", "logistic", "sev" *(smallest extreme value)*, "weibull3", "lognormal3" or "loglogistic3". Other distributions have not been implemented yet.
direction	a character string specifying the direction of the computed standard errors. Must be either "y" (used for confidence intervals of failure probabilities in confint_fisher) or "x" (used for confidence intervals of quantiles in confint_fisher). If p is a quantile then *direction* needs to be "y" and vice versa.

Value

A numeric value with estimated standard errors of quantiles or standardized z values. Both are required for the computation of normal approximation confidence intervals. If standard errors of standardized z values are compueted one can calculate confidence intervals for distribution probabilities (z-procedure, which is used inside confint_fisher).

References

Meeker, William Q; Escobar, Luis A., Statistical methods for reliability data, New York: Wiley series in probability and statistics, 1998

Examples

```
obs   <- seq(10000, 100000, 10000)
state <- c(0, 1, 1, 0, 0, 0, 1, 0, 1, 0)

mle <- ml_estimation(x = obs, event = state,
                     distribution = "weibull", conf_level = 0.95)
delta_prob <- sapply(obs, delta_method,
                          loc_sc_params = mle$loc_sc_coefficients,
                          loc_sc_varcov = mle$loc_sc_vcov,
                          distribution = "weibull",
                          direction = "y")
```

---

dist_delay_register *Parameter Estimation of the Delay in Registration Distribution*

Description

This function introduces a delay random variable by calculating the time difference between the registration and production date for the sample units and afterwards estimates the parameter(s) of a supposed distribution, using MLE.

Usage

```
dist_delay_register(date_prod, date_register, distribution = "lognormal")
```

Arguments

date_prod	a vector of class "character" or "Date", in the format "yyyy-mm-dd", indicating the date of production of a unit. If no date is available use NA.
date_register	a vector of class "character" or "Date", in the format "yyyy-mm-dd", indicating the date of registration of a unit. If no date is available use NA.
distribution	supposed distribution of the random variable. The default value is "lognormal". So far no other distribution is implemented.

Value

A named vector of estimated parameters for the specified distribution.

Examples

```
date_of_production   <- c("2014-07-28", "2014-02-17", "2014-07-14",
                          "2014-06-26", "2014-03-10", "2014-05-14",
                          "2014-05-06", "2014-03-07", "2014-03-09",
                          "2014-04-13", "2014-05-20", "2014-07-07",
                          "2014-01-27", "2014-01-30", "2014-03-17",
                          "2014-02-09", "2014-04-14", "2014-04-20",
                          "2014-03-13", "2014-02-23", "2014-04-03",
                          "2014-01-08", "2014-01-08")
date_of_registration <- c(NA, "2014-03-29", "2014-12-06", "2014-09-09",
                          NA, NA, "2014-06-16", NA, "2014-05-23",
                          "2014-05-09", "2014-05-31", NA, "2014-04-13",
                          NA, NA, "2014-03-12", NA, "2014-06-02",
                          NA, "2014-03-21", "2014-06-19", NA, NA)

params_delay_regist  <- dist_delay_register(
                                  date_prod = date_of_production,
                                  date_register = date_of_registration,
                                  distribution = "lognormal")
```

---

dist_delay_report *Parameter Estimation of the Delay in Report Distribution*

Description

This function introduces a delay random variable by calculating the time difference between the report and repair date for the sample units and afterwards estimates the parameter(s) of a supposed distribution, using MLE.

Usage

```
dist_delay_report(date_repair, date_report, distribution = "lognormal")
```

Arguments

date_repair	a vector of class "character" or "Date", in the format "yyyy-mm-dd", indicating the date of repair of a failed unit. If no date is available use NA.
date_report	a vector of class "character" or "Date", in the format "yyyy-mm-dd", indicating the date of report of a failed unit. If no date is available use NA.
distribution	supposed distribution of the random variable. The default value is "lognormal". So far no other distribution is implemented.

Value

A named vector of estimated parameters for the specified distribution.

Examples

```
date_of_repair <- c(NA, "2014-09-15", "2015-07-04", "2015-04-10", NA,
                    NA, "2015-04-24", NA, "2015-04-25", "2015-04-24",
                     "2015-06-12", NA, "2015-05-04", NA, NA,
                     "2015-05-22", NA, "2015-09-17", NA, "2015-08-15",
                     "2015-11-26", NA, NA)

date_of_report <- c(NA, "2014-10-09", "2015-08-28", "2015-04-15", NA,
                     NA, "2015-05-16", NA, "2015-05-28", "2015-05-15",
                     "2015-07-11", NA, "2015-08-14", NA, NA,
                     "2015-06-05", NA, "2015-10-17", NA, "2015-08-21",
                     "2015-12-02", NA, NA)

params_delay_report  <- dist_delay_report(date_repair = date_of_repair,
                                          date_report = date_of_report,
                                          distribution = "lognormal")
```

---

dist_mileage *Parameter Estimation of the Mileage Distribution*

Description

This function introduces a random variable of annual mileage using the units in the sample that had a failure and afterwards estimates the parameter(s) of a supposed distribution, using MLE.

Usage

```
dist_mileage(x, event, mileage, distribution = "lognormal")
```

Arguments

x	a numeric vector of operating times. If not available use NA.
event	a vector of binary data (0 or 1) indicating whether unit i is a right censored observation (= 0) or a failure (= 1).

mileage	a numeric vector of driven distances. If not available use NA.
distribution	supposed distribution of the random variable. The default value is "lognormal". So far no other distribution is implemented.

Value

A named vector of estimated parameters for the specified mileage distribution.

Examples

```
date_of_registration <- c(NA, "2014-03-29", "2014-12-06", "2014-09-09", NA,
                          NA, "2014-06-16", NA, "2014-05-23", "2014-05-09",
                          "2014-05-31", NA, "2014-04-13", NA, NA, "2014-03-12",
                          NA, "2014-06-02", NA, "2014-03-21", "2014-06-19",
                          NA, NA)
date_of_repair <- c(NA, "2014-09-15", "2015-07-04", "2015-04-10", NA,
                    NA, "2015-04-24", NA, "2015-04-25", "2015-04-24",
                    "2015-06-12", NA, "2015-05-04", NA, NA, "2015-05-22",
                    NA, "2015-09-17", NA, "2015-08-15", "2015-11-26", NA, NA)

op_time <- as.numeric(difftime(as.Date(date_of_repair),
                               as.Date(date_of_registration),
                               units = "days"))
mileage <- c(NA, 15655, 13629, 18292, NA, NA, 33555, NA, 21737,
             29870, 21068, NA, 122283, NA, NA, 36088, NA, 11153,
             NA, 122842, 20349, NA, NA)
state <- c(0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0)

params_mileage_annual <- dist_mileage(x = op_time, event = state,
                                      mileage = mileage,
                                      distribution = "lognormal")
```

---

johnson_method *Estimation of Failure Probabilities using Johnson's Method*

Description

This non-parametric approach is used to estimate the failure probabilities in terms of uncensored or (multiple) right censored data. Compared to complete data the correction is done by calculating adjusted ranks which takes non-defective units into account.

Usage

```
johnson_method(x, event, id = rep("XXXXXX", length(x)))
```

Arguments

x	a numeric vector which consists of lifetime data. Lifetime data could be every characteristic influencing the reliability of a product, e.g. operating time (days/months in service), mileage (km, miles), load cycles.
event	a vector of binary data (0 or 1) indicating whether unit *i* is a right censored observation (= 0) or a failure (= 1).
id	a character vector for the identification of every unit.

Value

A data frame containing id, lifetime characteristic, status of the unit, the adjusted rank and the estimated failure probability. For right censored observations the cells of the rank and probabilty columns are filled with NA values.

Examples

```
obs   <- seq(10000, 100000, 10000)
state <- c(0, 1, 1, 0, 0, 0, 1, 0, 1, 0)
uic   <- c("3435", "1203", "958X", "XX71", "abcd", "tz46",
           "fl29", "AX23", "Uy12", "kl1a")

df_john <- johnson_method(x = obs, event = state, id = uic)
```

---

kaplan_method *Estimation of Failure Probabilities using Kaplan-Meier*

Description

Whereas the non-parametric Kaplan-Meier estimator is used to estimate the survival function *S(t)* in terms of (multiple) right censored data, the complement is an estimate of the cumulative distribution function *F(t)*. One modification is made in contrast to the orginial Kaplan-Meier estimator (based on *NIST/SEMATECH e-Handbook of Statistical Methods*, 8.2.1.5.): If the last unit (unit with highest observed lifetime) is a defective unit, the estimator is adjusted in such a way that the survival estimate for this unit is not *zero* and therefore the estimate for the failure probability is not equal to *one*. Otherwise the estimate in this context would be too pessimisitc. Since the failure probability estimation in this function is not based on *Median Ranks*, the Betabinomial confidence intervals cannot be calculated on the basis of Kaplan-Meier failure probabilities.

Usage

```
kaplan_method(x, event, id = rep("XXXXXX", length(x)))
```

Arguments

x	a numeric vector which consists of lifetime data. Lifetime data could be every characteristic influencing the reliability of a product, e.g. operating time (days/months in service), mileage (km, miles), load cycles.
event	a vector of binary data (0 or 1) indicating whether unit *i* is a right censored observation (= 0) or a failure (= 1).
id	a character vector for the identification of every unit.

Value

A data frame containing id, lifetime characteristic, status of the unit and the estimated failure probabilty. For right censored observations the cells of probability column are filled with NA.

References

NIST/SEMATECH e-Handbook of Statistical Methods, *8.2.1.5. Empirical model fitting - distribution free (Kaplan-Meier) approach*, https://www.itl.nist.gov/div898/handbook/apr/section2/apr215.htm, 30/04/2018

Examples

```
# Example 1
obs   <- seq(10000, 100000, 10000)
state <- c(0, 1, 1, 0, 0, 0, 1, 0, 1, 0)
uic   <- c("3435", "1203", "958X", "XX71", "abcd", "tz46",
           "fl29", "AX23","Uy12", "kl1a")

df_kap <- kaplan_method(x = obs, event = state, id = uic)

# Example 2
 df <- data.frame(obs = c(10000, 10000, 20000, 20000, 30000,
                          30000, 30000, 30000, 40000, 50000,
                          50000, 60000, 70000, 70000, 70000,
                          70000, 80000, 80000, 80000, 80000,
                          90000, 90000, 100000),
                  state = rep(1, 23))

df_kap2 <- kaplan_method(x = df$obs, event = df$state)
```

`loglik_function` *Log-Likelihood Function for (Log-) Location-Scale Distributions (with Threshold)*

Description

This function computes the log-likelihood value with respect to a given set of parameters. For two-parametric models the location and scale parameters are required. If a three-parametric lifetime distribution is needed an additional threshold parameter is required. In terms of *Maximum Likelihood Estimation* this function can be optimized (`optim`) to estimate the parameters and variance-covariance matrix of the parameters.

Usage

```
loglik_function(x, event, wts = rep(1, length(x)), pars,
  distribution = c("weibull", "lognormal", "loglogistic", "normal",
  "logistic", "sev", "weibull3", "lognormal3", "loglogistic3"))
```

Arguments

`x`	a numeric vector which consists of lifetime data. Lifetime data could be every characteristic influencing the reliability of a product, e.g. operating time (days/months in service), mileage (km, miles), load cycles.
`event`	a vector of binary data (0 or 1) indicating whether unit i is a right censored observation (= 0) or a failure (= 1).
`wts`	optional vector of case weights. The length of `wts` must be the same as the number of observations `x`. Default is that `wts` is a vector with all components being 1 (same weights).
`pars`	a numeric vector of parameters. The first element is the location parameter (μ), the second is the scale parameter (σ) and if a three-parametric model is used the third element is the threshold parameter (γ).
`distribution`	supposed distribution of the random variable. The value can be `"weibull"`, `"lognormal"`, `"loglogistic"`, `"normal"`, `"logistic"`, `"sev"` *(smallest extreme value)*, `"weibull3"`, `"lognormal3"` or `"loglogistic3"`. Other distributions have not been implemented yet.

References

Meeker, William Q; Escobar, Luis A., Statistical methods for reliability data, New York: Wiley series in probability and statistics, 1998

Examples

```
# Alloy T7987 dataset taken from Meeker and Escobar(1998, p. 131)
cycles   <- c(300, 300, 300, 300, 300, 291, 274, 271, 269, 257, 256, 227, 226,
              224, 213, 211, 205, 203, 197, 196, 190, 189, 188, 187, 184, 180,
              180, 177, 176, 173, 172, 171, 170, 170, 169, 168, 168, 162, 159,
              159, 159, 159, 152, 152, 149, 149, 144, 143, 141, 141, 140, 139,
              139, 136, 135, 133, 131, 129, 123, 121, 121, 118, 117, 117, 114,
              112, 108, 104, 99, 99, 96, 94)
state <- c(rep(0, 5), rep(1, 67))

# Example 1: Evaluating Log-Likelihood function of two-parametric weibull:
```

```
loglik_weib <- loglik_function(x = cycles, event = state, pars = c(5.29, 0.33),
                               distribution = "weibull")

# Example 2: Evaluating Log-Likelihood function of three-parametric weibull:
loglik_weib3 <- loglik_function(x = cycles, event = state,
                                pars = c(4.54, 0.76, 92.99),
                                distribution = "weibull3")
```

`loglik_profiling` *Log-Likelihood Profile Function for Log-Location-Scale Distributions with Threshold*

Description

This function evaluates the log-likelihood with respect to a given threshold parameter of a three-parametric lifetime distribution. In terms of *Maximum Likelihood Estimation* this function can be optimized (`optim`) to estimate the threshold parameter.

Usage

```
loglik_profiling(x, event, thres, distribution = c("weibull3",
  "lognormal3", "loglogistic3"))
```

Arguments

`x`	a numeric vector which consists of lifetime data. Lifetime data could be every characteristic influencing the reliability of a product, e.g. operating time (days/months in service), mileage (km, miles), load cycles.
`event`	a vector of binary data (0 or 1) indicating whether unit i is a right censored observation (= 0) or a failure (= 1).
`thres`	a numeric value of the threshold parameter.
`distribution`	supposed distribution of the random variable. The value can be `"weibull3"`, `"lognormal3"` or `"loglogistic3"`.

Value

Returns the log-likelihood value for a specified threshold value.

References

Meeker, William Q; Escobar, Luis A., Statistical methods for reliability data, New York: Wiley series in probability and statistics, 1998

Examples

```
# Alloy T7987 dataset taken from Meeker and Escobar(1998, p. 131)
cycles   <- c(300, 300, 300, 300, 300, 291, 274, 271, 269, 257, 256, 227, 226,
              224, 213, 211, 205, 203, 197, 196, 190, 189, 188, 187, 184, 180,
              180, 177, 176, 173, 172, 171, 170, 170, 169, 168, 168, 162, 159,
              159, 159, 159, 152, 152, 149, 149, 144, 143, 141, 141, 140, 139,
              139, 136, 135, 133, 131, 129, 123, 121, 121, 118, 117, 117, 114,
              112, 108, 104, 99, 99, 96, 94)
state <- c(rep(0, 5), rep(1, 67))

# Determining threshold parameter for which the log-likelihood is maximized
# subject to the condition that the threshold parameter must be smaller
# as the first failure cycle, i.e 94:
threshold <- seq(0, min(cycles[state == 1]) - 0.1, length.out = 100)
profile_logL <- sapply(threshold, loglik_profiling,
                       x = cycles,
                       event = state,
                       distribution = "weibull3")
threshold[which.max(profile_logL)]

# plot:
# plot(threshold, profile_logL, type = "l")
# abline(v = threshold[which.max(profile_logL)], h = max(profile_logL), col = "red")
```

`mcs_delays` *Adjustment of Operating Times by Delays using a Monte Carlo Approach*

Description

This function is a wrapper that combines both, the `mcs_delay_register` and `mcs_delay_report` function for adjusting the operation times of censored units.

Usage

```
mcs_delays(date_prod, date_register, date_repair, date_report, x, event,
  distribution = "lognormal", details = FALSE, seed = NULL)
```

Arguments

`date_prod`	a vector of class `"character"` or `"Date"`, in the format "yyyy-mm-dd", indicating the date of production of a unit. If no date is available use `NA`.
`date_register`	a vector of class `"character"` or `"Date"`, in the format "yyyy-mm-dd", indicating the date of registration of a unit. If no date is available use `NA`.
`date_repair`	a vector of class `"character"` or `"Date"`, in the format "yyyy-mm-dd", indicating the date of repair of a failed unit. If no date is available use `NA`.
`date_report`	a vector of class `"character"` or `"Date"`, in the format "yyyy-mm-dd", indicating the date of report of a failed unit. If no date is available use `NA`.

x	a numeric vector of operating times.
event	a vector of binary data (0 or 1) indicating whether unit *i* is a right censored observation (= 0) or a failure (= 1).
distribution	supposed distribution of the random variable. The default value is "lognormal". So far no other distribution is implemented.
details	a logical variable, where the default value is FALSE. If FALSE the output consists of a vector with corrected operating times for the censored units and the input operating times for the failed units. If TRUE the output consists of a detailed list, i.e the same vector as described before, simulated random numbers, estimated distribution parameters and a seed for reproducibility.
seed	if seed = NULL a random seed is used. Otherwise the user can specify an integer for the seed.

Value

A numerical vector of corrected operating times for the censored units and the input operating times for the failed units if details = FALSE. If details = TRUE the output is a list which consists of the following elements:

- time : Numerical vector of corrected operating times for the censored observations and input operating times for failed units.
- x_sim_regist : Simulated random numbers of specified distribution with estimated parameters for delay in registration. The length of x_sim_regist is equal to the number of censored observations.
- x_sim_report : Simulated random numbers of specified distribution with estimated parameters for delay in report. The length of x_sim_report is equal to the number of censored observations.
- coefficients_regist : Estimated coefficients of supposed distribution for delay in registration.
- coefficients_report : Estimated coefficients of supposed distribution for delay in report.
- int_seed : Integer seed number for reproducibility.

Examples

```
date_of_production   <- c("2014-07-28", "2014-02-17", "2014-07-14",
                          "2014-06-26", "2014-03-10", "2014-05-14",
                          "2014-05-06", "2014-03-07", "2014-03-09",
                          "2014-04-13", "2014-05-20", "2014-07-07",
                          "2014-01-27", "2014-01-30", "2014-03-17",
                          "2014-02-09", "2014-04-14", "2014-04-20",
                          "2014-03-13", "2014-02-23", "2014-04-03",
                          "2014-01-08", "2014-01-08")
date_of_registration <- c("2014-08-17", "2014-03-29", "2014-12-06",
                          "2014-09-09", "2014-05-14", "2014-07-01",
                          "2014-06-16", "2014-04-03", "2014-05-23",
                          "2014-05-09", "2014-05-31", "2014-08-12",
                          "2014-04-13", "2014-02-15", "2014-07-07",
                          "2014-03-12", "2014-05-27", "2014-06-02",
```

```
                          "2014-05-20", "2014-03-21", "2014-06-19",
                          "2014-02-12", "2014-03-27")
date_of_repair <- c(NA, "2014-09-15", "2015-07-04", "2015-04-10", NA,
                    NA, "2015-04-24", NA, "2015-04-25", "2015-04-24",
                     "2015-06-12", NA, "2015-05-04", NA, NA,
                     "2015-05-22", NA, "2015-09-17", NA, "2015-08-15",
                     "2015-11-26", NA, NA)

date_of_report <- c(NA, "2014-10-09", "2015-08-28", "2015-04-15", NA,
                     NA, "2015-05-16", NA, "2015-05-28", "2015-05-15",
                     "2015-07-11", NA, "2015-08-14", NA, NA,
                     "2015-06-05", NA, "2015-10-17", NA, "2015-08-21",
                     "2015-12-02", NA, NA)

op_time <- rep(1000, length(date_of_repair))
state <- c(0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0)

# Example 1 - Simplified vector output:
x_corrected <- mcs_delays(date_prod = date_of_production,
                          date_register = date_of_registration,
                          date_repair = date_of_repair,
                          date_report = date_of_report,
                          x = op_time,
                          event = state,
                          distribution = "lognormal",
                          seed = NULL,
                          details = FALSE)

# Example 2 - Detailed list output:
list_detail <- mcs_delays(date_prod = date_of_production,
                                date_register = date_of_registration,
                                date_repair = date_of_repair,
                                date_report = date_of_report,
                                x = op_time,
                                event = state,
                                distribution = "lognormal",
                                seed = NULL,
                                details = TRUE)
```

`mcs_delay_register` *Adjustment of Operating Times by Delays in Registration using a Monte Carlo Approach*

Description

In general the amount of information about units in the field, that have not failed yet, are rare. For example it is common that a supplier, who provides parts to the automotive industry does not know when a vehicle was put in service and therefore does not know the exact operating time of the supplied parts. This function uses a Monte Carlo approach for simulating the operating

times of (multiple) right censored observations, taking account of registering delays. The simulation is based on the distribution of operating times that were calculated from complete data (see dist_delay_register).

Usage

```
mcs_delay_register(date_prod, date_register, x, event,
  distribution = "lognormal", seed = NULL, details = FALSE)
```

Arguments

date_prod	a vector of class "character" or "Date", in the format "yyyy-mm-dd", indicating the date of production of a unit. If no date is available use NA.
date_register	a vector of class "character" or "Date", in the format "yyyy-mm-dd", indicating the date of registration of a unit. If no date is available use NA.
x	a numeric vector of operating times.
event	a vector of binary data (0 or 1) indicating whether unit i is a right censored observation (= 0) or a failure (= 1).
distribution	supposed distribution of the random variable. The default value is "lognormal". So far no other distribution is implemented.
seed	if seed = NULL a random seed is used. Otherwise the user can specify an integer for the seed.
details	a logical variable, where the default value is FALSE. If FALSE the output consists of a vector with corrected operating times for the censored units and the input operating times for the failed units. If TRUE the output consists of a detailed list, i.e the same vector as described before, simulated random numbers, estimated distribution parameters and a seed for reproducibility.

Value

A numeric vector of corrected operating times for the censored units and the input operating times for the failed units if details = FALSE. If details = TRUE the output is a list which consists of the following elements:

- time : Numeric vector of corrected operating times for the censored observations and input operating times for failed units.
- x_sim : Simulated random numbers of specified distribution with estimated parameters. The length of x_sim is equal to the number of censored observations.
- coefficients : Estimated coefficients of supposed distribution.
- int_seed : Integer seed number for reproducibility.

Examples

```
date_of_production   <- c("2014-07-28", "2014-02-17", "2014-07-14",
                          "2014-06-26", "2014-03-10", "2014-05-14",
                          "2014-05-06", "2014-03-07", "2014-03-09",
                          "2014-04-13", "2014-05-20", "2014-07-07",
```

```
                          "2014-01-27", "2014-01-30", "2014-03-17",
                          "2014-02-09", "2014-04-14", "2014-04-20",
                          "2014-03-13", "2014-02-23", "2014-04-03",
                          "2014-01-08", "2014-01-08")
date_of_registration <- c(NA, "2014-03-29", "2014-12-06", "2014-09-09",
                          NA, NA, "2014-06-16", NA, "2014-05-23",
                          "2014-05-09", "2014-05-31", NA, "2014-04-13",
                          NA, NA, "2014-03-12", NA, "2014-06-02",
                          NA, "2014-03-21", "2014-06-19", NA, NA)

op_time <- rep(1000, length(date_of_production))
state <- c(0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0)

# Example 1 - Simplified vector output:
x_corrected <- mcs_delay_register(date_prod = date_of_production,
                                  date_register = date_of_registration,
                                  x = op_time,
                                  event = state,
                                  distribution = "lognormal",
                                  seed = NULL,
                                  details = FALSE)

# Example 2 - Detailed list output:
list_detail <- mcs_delay_register(date_prod = date_of_production,
                                  date_register = date_of_registration,
                                  x = op_time,
                                  event = state,
                                  distribution = "lognormal",
                                  seed = NULL,
                                  details = TRUE)
```

`mcs_delay_report` *Adjustment of Operating Times by Delays in Report using a Monte Carlo Approach*

Description

The delay in report describes the time between the occurence of a damage and the registration in the warranty database. For a given date where the analysis is made there could be units which had a failure but are not registered in the database and therefore treated as censored units. To overcome this problem this function uses a Monte Carlo approach for simulating the operating times of (multiple) right censored observations, taking account of reporting delays. The simulation is based on the distribution of operating times that were calculated from complete data, i.e. failed items (see `dist_delay_report`).

Usage

```
mcs_delay_report(date_repair, date_report, x, event,
  distribution = "lognormal", details = FALSE, seed = NULL)
```

Arguments

date_repair	a vector of class "character" or "Date", in the format "yyyy-mm-dd", indicating the date of repair of a failed unit. If no date is available use NA.
date_report	a vector of class "character" or "Date", in the format "yyyy-mm-dd", indicating the date of report of a failed unit. If no date is available use NA.
x	a numeric vector of operating times.
event	a vector of binary data (0 or 1) indicating whether unit i is a right censored observation (= 0) or a failure (= 1).
distribution	supposed distribution of the random variable. The default value is "lognormal". So far no other distribution is implemented.
details	a logical variable, where the default value is FALSE. If FALSE the output consists of a vector with corrected operating times for the censored units and the input operating times for the failed units. If TRUE the output consists of a detailed list, i.e the same vector as described before, simulated random numbers, estimated distribution parameters and a seed for reproducibility.
seed	if seed = NULL a random seed is used. Otherwise the user can specify an integer for the seed.

Value

A numeric vector of corrected operating times for the censored units and the input operating times for the failed units if details = FALSE. If details = TRUE the output is a list which consists of the following elements:

- time : Numeric vector of corrected operating times for the censored observations and input operating times for failed units.
- x_sim : Simulated random numbers of specified distribution with estimated parameters. The length of x_sim is equal to the number of censored observations.
- coefficients : Estimated coefficients of supposed distribution.
- int_seed : Integer seed number for reproducibility.

Examples

```
date_of_repair <- c(NA, "2014-09-15", "2015-07-04", "2015-04-10", NA,
                    NA, "2015-04-24", NA, "2015-04-25", "2015-04-24",
                     "2015-06-12", NA, "2015-05-04", NA, NA,
                     "2015-05-22", NA, "2015-09-17", NA, "2015-08-15",
                     "2015-11-26", NA, NA)

date_of_report <- c(NA, "2014-10-09", "2015-08-28", "2015-04-15", NA,
                     NA, "2015-05-16", NA, "2015-05-28", "2015-05-15",
                     "2015-07-11", NA, "2015-08-14", NA, NA,
                     "2015-06-05", NA, "2015-10-17", NA, "2015-08-21",
                     "2015-12-02", NA, NA)

op_time <- rep(1000, length(date_of_repair))
state <- c(0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0)
```

```
# Example 1 - Simplified vector output:
x_corrected <- mcs_delay_report(date_repair = date_of_repair,
                                date_report = date_of_report,
                                x = op_time,
                                event = state,
                                distribution = "lognormal",
                                seed = NULL,
                                details = FALSE)

# Example 2 - Detailed list output:
list_detail <- mcs_delay_report(date_repair = date_of_repair,
                                date_report = date_of_report,
                                x = op_time,
                                event = state,
                                distribution = "lognormal",
                                seed = NULL,
                                details = TRUE)
```

`mcs_mileage` *Estimation of Driving Distances for Censored Observations using a Monte Carlo Approach*

Description

This function simulates driving distances for censored observations under the condition that the operating time of these items is known up to a certain date where analysis is made. Operating times for these units can be estimated with functions like `mcs_delay_register`, `mcs_delay_report` and `mcs_delays`. The failed observations (where the driving distances are known) are used to estimate an annual mileage distribution. If the mileage distribution is fully specified annual random driving distances are drawn from this distribution and afterwards adjusted to the operating times of the censored observations.

Usage

```
mcs_mileage(x, event, mileage, distribution = "lognormal", seed = NULL,
  details = FALSE)
```

Arguments

`x`	a numeric vector of operating times. If not available use `NA`.
`event`	a vector of binary data (0 or 1) indicating whether unit i is a right censored observation (= 0) or a failure (= 1).
`mileage`	a numeric vector of driven distances. If not available use `NA`.
`distribution`	supposed distribution of the random variable. The default value is `"lognormal"`. So far no other distribution is implemented.
`seed`	if `seed = NULL` a random seed is used. Otherwise the user can specify an integer for the seed.

`details` a logical variable, where the default value is `FALSE`. If `FALSE` the output consists of a vector with simulated driving distances for the censored units regarding to their current operating time and the input driving distances for the failed units. If `TRUE` the output consists of a detailed list, i.e the same vector as described before, simulated annual driving distances, estimated distribution parameters and a seed for reproducibility.

Value

A numerical vector of simulated driving distances for the censored units and the input driving distances for the failed units if `details = FALSE`. If `details = TRUE` the output is a list which consists of the following elements:

- `mileage` : Simulated driving distances for the censored units and the input driving distances for the failed units.
- `mileage_sim_annual` : Simulated annual driving distances of specified distribution with estimated parameters. The length of `x_sim` is equal to the number of censored observations.
- `coefficients` : Estimated coefficients of supposed distribution.
- `int_seed` : Integer seed number for reproducibility.

Examples

```
# Example 1 - Simplified vector output (complete data):
date_of_registration <- c("2014-08-17", "2014-03-29", "2014-12-06",
                          "2014-09-09", "2014-05-14", "2014-07-01",
                          "2014-06-16", "2014-04-03", "2014-05-23",
                          "2014-05-09", "2014-05-31", "2014-08-12",
                          "2014-04-13", "2014-02-15", "2014-07-07",
                          "2014-03-12", "2014-05-27", "2014-06-02",
                          "2014-05-20", "2014-03-21", "2014-06-19",
                          "2014-02-12", "2014-03-27")
date_of_repair <- c("2014-10-21", "2014-09-15", "2015-07-04", "2015-04-10",
                    "2015-02-15", "2015-04-14", "2015-04-24", "2015-02-27",
                    "2015-04-25", "2015-04-24", "2015-06-12", "2015-08-26",
                    "2015-05-04", "2015-04-04", "2015-09-06", "2015-05-22",
                    "2015-08-21", "2015-09-17", "2015-09-15", "2015-08-15",
                    "2015-11-26", "2015-08-22", "2015-10-05")

op_time <- as.numeric(difftime(as.Date(date_of_repair),
                               as.Date(date_of_registration),
                               units = "days"))
mileage <- c(5227, 15655, 13629, 18292, 24291, 34455, 33555, 21659, 21737,
             29870, 21068, 22986, 122283, 31592, 49050, 36088, 10918, 11153,
             122437, 122842, 20349, 65656, 40777)
state <- sample(c(0, 1), size = length(op_time), replace = TRUE)

mileage_corrected <- mcs_mileage(x = op_time, event = state,
                                 mileage = mileage,
                                 distribution = "lognormal", seed = NULL,
                                 details = FALSE)
```

```
# Example 2 - Detailed list output (complete data):
list_detail <- mcs_mileage(x = op_time, event = state, mileage = mileage,
                           distribution = "lognormal", seed = NULL,
                           details = TRUE)

# Example 3 - Detailed list output (realistic example):
op_time <- c(65, 170, 210, 213, 277, 287, 312, 330, 337, 350, 377, 379, 386,
             413, 426, 436, 451, 472, 483, 512, 525, 556, 557)
mileage <- c(NA, 15655, 13629, NA, 24291, 34455, NA, 21659, 21737,
             NA, 21068, 22986, NA, 31592, 49050, NA, 10918, 11153,
             NA, 122842, 20349, NA, 40777)
state <- c(0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0,
           1, 1, 0, 1)

list_detail <- mcs_mileage(x = op_time, event = state, mileage = mileage,
                           distribution = "lognormal", seed = NULL,
                           details = TRUE)
```

`mixmod_em` *Mixture Model Estimation using EM-Algorithm*

Description

This method uses the EM-Algorithm to estimate the parameters of a univariate mixture model. Until now, the mixture model can consist of k two-parametric Weibull distributions. If no mixture of k components can be estimated, the function is forced to stop and a message with instructions is given.

Usage

```
mixmod_em(x, event, post = NULL, distribution = "weibull",
  conf_level = 0.95, k = 2, method = "EM", n_iter = 100L,
  conv_limit = 1e-06, diff_loglik = 0.5)
```

Arguments

`x`	a numeric vector which consists of lifetime data. Lifetime data could be every characteristic influencing the reliability of a product, e.g. operating time (days/months in service), mileage (km, miles), load cycles.
`event`	a vector of binary data (0 or 1) indicating whether unit *i* is a right censored observation (= 0) or a failure (= 1).
`post`	a numeric matrix specifiying initial a-posteriori probabilities. If post is `NULL` (default) a-posteriori probabilities are assigned randomly using the Dirichlet distribution (`rdirichlet` from *LearnBayes* Package), which is the conjugate prior of a Multinomial distribution. This idea was taken from the blog post of Mr. Gelissen (linked under *references*).
`distribution`	supposed mixture model. Only `"weibull"` can be used. Other distributions have not been implemented yet.

conf_level	confidence level for the confidence intervals of the parameters of every component k. The default value is conf_level = 0.95.
k	integer of mixture components, default is 2.
method	default method is "EM". Other methods have not been implemented yet.
n_iter	integer defining the maximum number of iterations.
conv_limit	numeric value defining the convergence limit.
diff_loglik	numeric value defining the maximum difference between log-likelihood values, which seems permissible. The default value is 0.5. See **Details** for the usage of this argument.

Details

In mixmod_em the function mixture_em_cpp is called. The computed posterior probabilities are then used as weights inside function ml_estimation to model a weighted log-likelihood. This strategy enables the computation of confidence intervals for the parameters of the separated sub-distributions, since ml_estimation provides a variance-covariance matrix. Using this strategy, a potential problem that can occur is, that the value of the complete log-likelihood, computed by mixture_em_cpp, differs considerably from the complete log-likelihood after re-estimating parameters with ml_estimation. If so, the estimated quantities like prior and posterior probabilities, as well as the model parameters are not reliable anymore and the function is forced to stop with the message: "Parameter estimation was not successful!" But if the log-likelihood values are close to each other, the presence of the mixture is strengthened and a reasonable fit is provided. Thus, a check of the absolute differences in the log-likelihood values is made and the critical difference has to be specified in argument diff_loglik.

Value

Returns a list where the length of the list depends on the number of k subgroups. The first k lists have the same information as provided by ml_estimation, but the values logL, aic and bic are the results of a log-likelihood function, which is weighted by a-posteriori probabilities. The last list summarizes further results of the EM-Algorithm and is therefore called em_results. It contains the following elements:

- a_priori : A vector with estimated a-priori probabilities.
- a_posteriori : A matrix with estimated a-posteriori probabilities.
- groups : Numeric vector specifying the group membership of every observation.
- logL : The value of the complete log-likelihood.
- aic : Akaike Information Criterion.
- bic : Bayesian Information Criterion.

References

- Doganaksoy, N.; Hahn, G.; Meeker, W. Q., Reliability Analysis by Failure Mode, Quality Progress, 35(6), 47-52, 2002
- Blog posts by Stefan Gelissen: http://blogs2.datall-analyse.nl/2016/02/18/rcode_mixture_distribution_censored; last access on 19th January 2019

Examples

```
# Data is taken from given reference of Doganaksoy, Hahn and Meeker:
hours <- c(2, 28, 67, 119, 179, 236, 282, 317, 348, 387, 3, 31, 69, 135,
           191, 241, 284, 318, 348, 392, 5, 31, 76, 144, 203, 257, 286,
           320, 350, 412, 8, 52, 78, 157, 211, 261, 298, 327, 360, 446,
           13, 53, 104, 160, 221, 264, 303, 328, 369, 21, 64, 113, 168,
           226, 278, 314, 328, 377)
state <- c(1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1,
          1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 0,
          1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
          0, 1, 1, 1, 1, 1, 1)

mix_mod_em <- mixmod_em(x = hours,
                        event = state,
                        distribution = "weibull",
                        conf_level = 0.95,
                        k = 2,
                        method = "EM",
                        n_iter = 150)
```

`mixmod_regression` *Mixture Model Identification using Segmented Regression*

Description

This method uses piecewise linear regression to separate the data in subgroups, if appropriate. Since this happens in an automated fashion the function tends to overestimate the number of breakpoints and therefore returns too many subgroups. This problem is already stated in the documentation of the function segmented.lm, which is part of the *segmented* package. A maximum of three subgroups can be obtained.

Usage

```
mixmod_regression(x, y, event, distribution = c("weibull", "lognormal",
  "loglogistic"), conf_level = 0.95)
```

Arguments

`x`	a numeric vector which consists of lifetime data. Lifetime data could be every characteristic influencing the reliability of a product, e.g. operating time (days/months in service), mileage (km, miles), load cycles.
`y`	a numeric vector which consists of estimated failure probabilities regarding the lifetime data in `x`.
`event`	a vector of binary data (0 or 1) indicating whether unit i is a right censored observation (= 0) or a failure (= 1).

distribution	supposed distribution of the random variable. The value can be "weibull", "lognormal" or "loglogistic". Other distributions have not been implemented yet.
conf_level	confidence level of the interval. The default value is conf_level = 0.95.

Value

Returns a list where the length of the list depends on the number of identified subgroups. Each list has the same information as provided by rank_regression. Additionally each list has an element that specifies the range regarding the lifetime data for every subgroup.

References

Doganaksoy, N.; Hahn, G.; Meeker, W. Q., Reliability Analysis by Failure Mode, Quality Progress, 35(6), 47-52, 2002

Examples

```
# Data is taken from given reference:
hours <- c(2, 28, 67, 119, 179, 236, 282, 317, 348, 387, 3, 31, 69, 135,
          191, 241, 284, 318, 348, 392, 5, 31, 76, 144, 203, 257, 286,
          320, 350, 412, 8, 52, 78, 157, 211, 261, 298, 327, 360, 446,
          13, 53, 104, 160, 221, 264, 303, 328, 369, 21, 64, 113, 168,
          226, 278, 314, 328, 377)
state <- c(1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1,
          1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 0,
          1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
          0, 1, 1, 1, 1, 1, 1)
john <- johnson_method(x = hours, event = state)

mix_mod <- mixmod_regression(x = john$characteristic,
                             y = john$prob,
                             event = john$status,
                             distribution = "weibull")
```

`mixture_em_cpp` *EM-Algorithm using Newton-Raphson Method*

Description

This method uses the EM-Algorithm to estimate the parameters of a univariate mixture model. Until now, the mixture model can consist of k two-parametric Weibull distributions. The Weibull distributions are parameterized with scale η and shape β. In M-step these parameters are estimated using Newton-Raphson. This function is implemented in c++ and is called in function `mixmod_em`.

Usage

```
mixture_em_cpp(x, event, post, distribution = "weibull", k = 2L,
  method = "EM", n_iter = 100L, conv_limit = 1e-06)
```

Arguments

x	a numeric vector which consists of lifetime data. Lifetime data could be every characteristic influencing the reliability of a product, e.g. operating time (days/months in service), mileage (km, miles), load cycles.
event	a vector of binary data (0 or 1) indicating whether unit *i* is a right censored observation (= 0) or a failure (= 1).
post	a numeric matrix specifiying initial a-posteriori probabilities. The number of rows have to be in line with observations x and the number of columns must equal the mixture components k.
distribution	supposed distribution of mixture model components. The value must be "weibull". Other distributions have not been implemented yet.
k	integer of mixture components, default is 2.
method	default method is "EM". Other methods have not been implemented yet.
n_iter	integer defining the maximum number of iterations.
conv_limit	numeric value defining the convergence limit.

Value

Returns a list with the following components:

- coefficients : A matrix with estimated Weibull parameters. In the first row the estimated scale parameters η and in the second the estimated shape parameters β are provided. The first column belongs to the first mixture component and so forth.
- posteriori : A matrix with estimated a-posteriori probabilities.
- priori : A vector with estimated a-priori probabilities.
- logL : The value of the complete log-likelihood.

References

Doganaksoy, N.; Hahn, G.; Meeker, W. Q., Reliability Analysis by Failure Mode, Quality Progress, 35(6), 47-52, 2002

Examples

```
# Data is taken from given reference:
hours <- c(2, 28, 67, 119, 179, 236, 282, 317, 348, 387, 3, 31, 69, 135,
           191, 241, 284, 318, 348, 392, 5, 31, 76, 144, 203, 257, 286,
           320, 350, 412, 8, 52, 78, 157, 211, 261, 298, 327, 360, 446,
           13, 53, 104, 160, 221, 264, 303, 328, 369, 21, 64, 113, 168,
           226, 278, 314, 328, 377)
state <- c(1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1,
          1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 0,
          1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
          0, 1, 1, 1, 1, 1, 1)
post_dirichlet <- LearnBayes::rdirichlet(n = length(hours),
                                         par = rep(.1, 2))
mix_mod_em <- mixture_em_cpp(x = hours,
```

```
event = state,
post = post_dirichlet,
distribution = "weibull",
k = 2,
method = "EM",
n_iter = 150)
```

`ml_estimation` *ML Estimation for Parametric Lifetime Distributions*

Description

This method estimates the parameters and calculates normal approximation confidence intervals for a two- or three-parametric lifetime distribution in the frequently used (log-) location-scale parameterization. `ml_estimation` uses the `Lifedata.MLE` function which is defined in the *SPREDA* package. For the Weibull the estimates are transformed such that they are in line with the parameterization provided by the *stats* package like `pweibull`. The method is applicable for complete and (multiple) right censored data.

Usage

```
ml_estimation(x, event, distribution = c("weibull", "lognormal",
  "loglogistic", "normal", "logistic", "sev", "weibull3", "lognormal3",
  "loglogistic3"), wts = rep(1, length(x)), conf_level = 0.95,
  details = TRUE)
```

Arguments

`x`	a numeric vector which consists of lifetime data. Lifetime data could be every characteristic influencing the reliability of a product, e.g. operating time (days/months in service), mileage (km, miles), load cycles.
`event`	a vector of binary data (0 or 1) indicating whether unit *i* is a right censored observation (= 0) or a failure (= 1).
`distribution`	supposed distribution of the random variable. The value can be `"weibull"`, `"lognormal"`, `"loglogistic"`, `"normal"`, `"logistic"`, `"sev"` *(smallest extreme value)*, `"weibull3"`, `"lognormal3"` or `"loglogistic3"`. Other distributions have not been implemented yet.
`wts`	optional vector of case weights. The length of `wts` must be the same as the number of observations `x`. Default is that `wts` is a vector with all components being 1 (same weights).
`conf_level`	confidence level of the interval. The default value is `conf_level = 0.95`.
`details`	a logical variable, where the default value is `TRUE`. If `FALSE` the output consists of a list that only contains the estimated parameters. If `TRUE` the output is a detailed list with many more information. See below (**Value**).

Value

Returns a list with the following components (depending on `details` argument):

- `coefficients` : Provided, if `distribution` is `"weibull"`. η is the estimated scale and β the estimated shape parameter.
- `confint` : Provided, if `distribution` is `"weibull"`. Confidence interval for η and β.
- `loc_sc_coefficients` : Estimated location-scale parameters.
- `loc_sc_confint` : Confidence interval for location-scale parameters.
- `loc_sc_vcov` : Estimated Variance-Covariance matrix of the used location-scale distribution.
- `logL` : The log-likelihood value.
- `aic` : Akaike Information Criterion.
- `bic` : Bayesian Information Criterion.

References

Meeker, William Q; Escobar, Luis A., Statistical methods for reliability data, New York: Wiley series in probability and statistics, 1998

Examples

```
# Example 1: Fitting a two-parameter Weibull:
obs   <- seq(10000, 100000, 10000)
state <- c(0, 1, 1, 0, 0, 0, 1, 0, 1, 0)

mle <- ml_estimation(x = obs, event = state,
                     distribution = "weibull", conf_level = 0.90)

# Example 2: Fitting a three-parameter Weibull:
# Alloy T7987 dataset taken from Meeker and Escobar(1998, p. 131)
cycles   <- c(300, 300, 300, 300, 300, 291, 274, 271, 269, 257, 256, 227, 226,
              224, 213, 211, 205, 203, 197, 196, 190, 189, 188, 187, 184, 180,
              180, 177, 176, 173, 172, 171, 170, 170, 169, 168, 168, 162, 159,
              159, 159, 159, 152, 152, 149, 149, 144, 143, 141, 141, 140, 139,
              139, 136, 135, 133, 131, 129, 123, 121, 121, 118, 117, 117, 114,
              112, 108, 104, 99, 99, 96, 94)
state <- c(rep(0, 5), rep(1, 67))

mle_weib3 <- ml_estimation(x = cycles, event = state,
                           distribution = "weibull3", conf_level = 0.95)
```

mr_method *Estimation of Failure Probabilities using Median Ranks*

Description

This non-parametric approach (*Median Ranks*) is used to estimate the failure probabilities in terms of complete data. Two methods are available to estimate the cumulative distribution function *F(t)*:

- "benard"; Benard's approximation for Median Ranks
- "invbeta"; Exact Median Ranks using the inverse beta distribution

Usage

```
mr_method(x, event = rep(1, length(x)), id = rep("XXXXXX", length(x)),
  method = "benard")
```

Arguments

x	a numeric vector which consists of lifetime data. Lifetime data could be every characteristic influencing the reliability of a product, e.g. operating time (days/months in service), mileage (km, miles), load cycles.
event	a vector of ones indicating that every unit *i* has failed.
id	a character vector for the identification of every unit.
method	method for the estimation of the cdf. Can be "benard" (default) or "invbeta".

Value

A data frame containing id, lifetime characteristic, status of the unit, the rank and the estimated failure probability.

Examples

```
# Example 1
obs   <- seq(10000, 100000, 10000)
state <- rep(1, length(obs))
uic   <- c("3435", "1203", "958X", "XX71", "abcd", "tz46",
           "fl29", "AX23", "Uy12", "kl1a")

df_mr <- mr_method(x = obs, event = state, id = uic,
                   method = "benard")

# Example 2
df_mr_invbeta <- mr_method(x = obs, event = state, id = uic,
                           method = "invbeta")
```

`nelson_method` *Estimation of Failure Probabilities using the Nelson-Aalen Estimator*

Description

This non-parametric approach estimates the cumulative hazard rate in terms of (multiple) right censored data. By equating the definition of the hazard rate with the hazard rate according to Nelson-Aalen one can calculate the failure probabilities. Since the failure probability estimation in this function is not based on *Median Ranks*, the Betabinomial confidence intervals cannot be calculated on the basis of Nelson-Aalen failure probabilities.

Usage

```
nelson_method(x, event, id = rep("XXXXXX", length(x)))
```

Arguments

`x`	a numeric vector which consists of lifetime data. Lifetime data could be every characteristic influencing the reliability of a product, e.g. operating time (days/months in service), mileage (km, miles), load cycles.
`event`	a vector of binary data (0 or 1) indicating whether unit i is a right censored observation (= 0) or a failure (= 1).
`id`	a character vector for the identification of every unit.

Value

A data frame containing id, lifetime characteristic, status of the unit and the estimated failure probabilty. For right censored observations the cells of probability column are filled with NA.

Examples

```
obs   <- seq(10000, 100000, 10000)
state <- c(0, 1, 1, 0, 0, 0, 1, 0, 1, 0)
uic   <- c("3435", "1203", "958X", "XX71", "abcd", "tz46",
           "fl29", "AX23","Uy12", "kl1a")

df_nel <- nelson_method(x = obs, event = state, id = uic)
```

plot_conf *Add Confidence Region(s) for Quantiles or Probabilities*

Description

This function is used to add estimated confidence region(s) to an existing probability plot which also includes the estimated regression line.

Usage

```
plot_conf(p_obj, x, y, direction = c("y", "x"),
  distribution = c("weibull", "lognormal", "loglogistic", "normal",
  "logistic", "sev", "weibull3", "lognormal3", "loglogistic3"),
  title_trace = "Confidence Limit")
```

Arguments

p_obj	a plotly object provided by function plot_mod.
x	a list containing the x-coordinates of the confidence region(s). The list can be of length 1 or 2. For more information see **Details**.
y	a list containing the y-coordinates of the Confidence Region(s). The list can be of length 1 or 2. For more information see **Details**.
direction	a character string specifying the direction of the plotted interval(s). Must be either "y" (failure probabilities) or "x" (quantiles).
distribution	supposed distribution of the random variable. The value can be "weibull", "lognormal", "loglogistic", "normal", "logistic", "sev" *(smallest extreme value)*, "weibull3", "lognormal3" or "loglogistic3". Other distributions have not been implemented yet.
title_trace	a character string which is assigned to the trace shown in the legend.

Details

It is important that the length of the vectors provided as lists in x and y match with the length of the vectors x and y in the function plot_mod. For this reason the following procedure is recommended:

- Calculate confidence intervals with the function confint_betabinom or confint_fisher and store it in a data.frame. For instance call it df.
- Inside plot_mod use the output df$characteristic for x and df$prob for y of the function(s) named before.
- In **Examples** the described approach is shown with code.

Value

Returns a plotly object containing the probability plot with plotting positions, the estimated regression line and the estimated confidence region(s).

References

Meeker, William Q; Escobar, Luis A., Statistical methods for reliability data, New York: Wiley series in probability and statistics, 1998

Examples

```
# Alloy T7987 dataset taken from Meeker and Escobar(1998, p. 131)
cycles   <- c(300, 300, 300, 300, 300, 291, 274, 271, 269, 257, 256, 227, 226,
              224, 213, 211, 205, 203, 197, 196, 190, 189, 188, 187, 184, 180,
              180, 177, 176, 173, 172, 171, 170, 170, 169, 168, 168, 162, 159,
              159, 159, 159, 152, 152, 149, 149, 144, 143, 141, 141, 140, 139,
              139, 136, 135, 133, 131, 129, 123, 121, 121, 118, 117, 117, 114,
              112, 108, 104, 99, 99, 96, 94)
state <- c(rep(0, 5), rep(1, 67))
id <- 1:length(cycles)

df_john <- johnson_method(x = cycles, event = state, id = id)
# Example 1: Probability Plot, Regression Line and Confidence Bounds for Three-Parameter-Weibull:
mrr <- rank_regression(x = df_john$characteristic,
                       y = df_john$prob,
                       event = df_john$status,
                       distribution = "weibull3",
                       conf_level = .90)

conf_betabin <- confint_betabinom(x = df_john$characteristic,
                                  event = df_john$status,
                                  loc_sc_params = mrr$loc_sc_coefficients,
                                  distribution = "weibull3",
                                  bounds = "two_sided",
                                  conf_level = 0.95,
                                  direction = "y")

plot_weibull <- plot_prob(x = df_john$characteristic,
                          y = df_john$prob,
                          event = df_john$status,
                          id = df_john$id,
                          distribution = "weibull",
                          title_main = "Three-Parametric Weibull",
                          title_x = "Cycles",
                          title_y = "Probability of Failure in %",
                          title_trace = "Failed Items")

plot_reg_weibull <- plot_mod(p_obj = plot_weibull,
                             x = conf_betabin$characteristic,
                             y = conf_betabin$prob,
                             loc_sc_params = mrr$loc_sc_coefficients,
                             distribution = "weibull3",
                             title_trace = "Estimated Weibull CDF")

plot_conf_beta <- plot_conf(p_obj = plot_reg_weibull,
                            x = list(conf_betabin$characteristic),
                            y = list(conf_betabin$lower_bound,
```

```
                                  conf_betabin$upper_bound),
                           direction = "y",
                           distribution = "weibull3",
                           title_trace = "Confidence Region")

# Example 2: Probability Plot, Regression Line and Confidence Bounds for Three-Parameter-Lognormal:
mrr_ln <- rank_regression(x = df_john$characteristic,
                          y = df_john$prob,
                          event = df_john$status,
                          distribution = "lognormal3",
                          conf_level = .90)

conf_betabin_ln <- confint_betabinom(x = df_john$characteristic,
                                     event = df_john$status,
                                     loc_sc_params = mrr_ln$loc_sc_coefficients,
                                     distribution = "lognormal3",
                                     bounds = "two_sided",
                                     conf_level = 0.95,
                                     direction = "y")

plot_lognormal <- plot_prob(x = df_john$characteristic,
                            y = df_john$prob,
                            event = df_john$status,
                            id = df_john$id,
                            distribution = "lognormal",
                            title_main = "Three-Parametric Lognormal",
                            title_x = "Cycles",
                            title_y = "Probability of Failure in %",
                            title_trace = "Failed Items")

plot_reg_lognormal <- plot_mod(p_obj = plot_lognormal,
                               x = conf_betabin_ln$characteristic,
                               y = conf_betabin_ln$prob,
                               loc_sc_params = mrr_ln$loc_sc_coefficients,
                               distribution = "lognormal3",
                               title_trace = "Estimated Lognormal CDF")

plot_conf_beta_ln <- plot_conf(p_obj = plot_reg_lognormal,
                               x = list(conf_betabin_ln$characteristic),
                               y = list(conf_betabin_ln$lower_bound,
                                        conf_betabin_ln$upper_bound),
                               direction = "y",
                               distribution = "lognormal3",
                               title_trace = "Confidence Region")
```

---

plot_layout *Layout of the Probability Plot*

Description

This function is used to create the layout of a probability plot.

Usage

```
plot_layout(x, distribution = c("weibull", "lognormal", "loglogistic",
  "normal", "logistic", "sev"), title_main = "Probability Plot",
  title_x = "Characteristic", title_y = "Unreliability")
```

Arguments

x	a numeric vector which consists of lifetime data. x is used to specify the grid of the plot.
distribution	supposed distribution of the random variable. The value can be "weibull", "lognormal", "loglogistic", "normal", "logistic" or "sev" *(smallest extreme value)*. Other distributions have not been implemented yet.
title_main	a character string which is assigned to the main title of the plot.
title_x	a character string which is assigned to the title of the x axis.
title_y	a character string which is assigned to the title of the y axis.

Value

Returns a plotly object which contains the layout that is used for probability plotting.

Examples

```
# Example 1: Weibull-Grid:
x_layout <- seq(1e-5, 1e+07, length.out = 10)
grid_weibull <- plot_layout(x = x_layout,
                            distribution = "weibull",
                            title_main = "Weibull Analysis",
                            title_x = "Time to Failure",
                            title_y = "Failure Probability in %")

# Example 2: Grid of Normal Distribution:
x_layout <- seq(1, 10, length.out = 10)
grid_normal <- plot_layout(x = x_layout,
                            distribution = "normal",
                            title_main = "Normal Grid",
                            title_x = "Time to Event",
                            title_y = "Failure Probability in %")
```

---

plot_mod *Adding an Estimated Population Line to a Probability Plot*

Description

This function adds a regression line to an existing probability plot using a model estimated by rank_regression or ml_estimation.

Usage

```
plot_mod(p_obj, x, y = NULL, loc_sc_params, distribution = c("weibull",
  "lognormal", "loglogistic", "normal", "logistic", "sev", "weibull3",
  "lognormal3", "loglogistic3"), title_trace = "Fit")
```

Arguments

p_obj	a plotly object provided by function plot_prob.
x	a numeric vector containing the x-coordinates of the regression line.
y	a numeric vector containing the y-coordinates of the regression line. The default value of y is NULL. If y is set NULL the y-coordinates with respect to x are calculated by function predict_prob using estimated coefficients in loc_sc_params. If confidence interval(s) should be added to the plot y should not be set to NULL. For more information see **Details** in plot_conf.
loc_sc_params	a (named) numeric vector of estimated location and scale parameters for a specified distribution. The order of elements is important. First entry needs to be the location parameter μ and the second element needs to be the scale parameter σ. If a three-parametric model is used the third element is the threshold parameter γ.
distribution	supposed distribution of the random variable. The value can be "weibull", "lognormal", "loglogistic", "normal", "logistic", "sev" *(smallest extreme value)*, "weibull3", "lognormal3" or "loglogistic3". Other distributions have not been implemented yet.
title_trace	a character string whis is assigned to the trace shown in the legend.

Value

Returns a plotly object containing the probability plot with plotting positions and the estimated regression line.

References

Meeker, William Q; Escobar, Luis A., Statistical methods for reliability data, New York: Wiley series in probability and statistics, 1998

Examples

```
# Alloy T7987 dataset taken from Meeker and Escobar(1998, p. 131)
cycles   <- c(300, 300, 300, 300, 300, 291, 274, 271, 269, 257, 256, 227, 226,
              224, 213, 211, 205, 203, 197, 196, 190, 189, 188, 187, 184, 180,
              180, 177, 176, 173, 172, 171, 170, 170, 169, 168, 168, 162, 159,
              159, 159, 159, 152, 152, 149, 149, 144, 143, 141, 141, 140, 139,
              139, 136, 135, 133, 131, 129, 123, 121, 121, 118, 117, 117, 114,
              112, 108, 104, 99, 99, 96, 94)
state <- c(rep(0, 5), rep(1, 67))
id <- 1:length(cycles)

df_john <- johnson_method(x = cycles, event = state, id = id)
```

```
# Example 1: Probability Plot and Regression Line Three-Parameter-Weibull:
plot_weibull <- plot_prob(x = df_john$characteristic,
                          y = df_john$prob,
                          event = df_john$status,
                          id = df_john$id,
                          distribution = "weibull",
                          title_main = "Three-Parametric Weibull",
                          title_x = "Cycles",
                          title_y = "Probability of Failure in %",
                          title_trace = "Failed Items")

mrr <- rank_regression(x = df_john$characteristic,
                       y = df_john$prob,
                       event = df_john$status,
                       distribution = "weibull3",
                       conf_level = .90)

plot_reg_weibull <- plot_mod(p_obj = plot_weibull, x = cycles,
                             loc_sc_params = mrr$loc_sc_coefficients,
                             distribution = "weibull3",
                             title_trace = "Estimated Weibull CDF")



# Example 2: Probability Plot and Regression Line Three-Parameter-Lognormal:
plot_lognormal <- plot_prob(x = df_john$characteristic,
                            y = df_john$prob,
                            event = df_john$status,
                            id = df_john$id,
                            distribution = "lognormal",
                            title_main = "Three-Parametric Lognormal",
                            title_x = "Cycles",
                            title_y = "Probability of Failure in %",
                            title_trace = "Failed Items")

mrr_ln <- rank_regression(x = df_john$characteristic,
                          y = df_john$prob,
                          event = df_john$status,
                          distribution = "lognormal3",
                          conf_level = .90)

plot_reg_lognormal <- plot_mod(p_obj = plot_lognormal, x = cycles,
                               loc_sc_params = mrr_ln$loc_sc_coefficients,
                               distribution = "lognormal3",
                               title_trace = "Estimated Lognormal CDF")
```

`plot_mod_mix` *Adding Estimated Population Lines of a Separated Mixture Model to a Probability Plot*

Description

This function adds one or multiple estimated regression lines to an existing probability plot (plot_prob_mix). Depending on the output of the function mixmod_regression or mixmod_em one or multiple lines are plotted.

Usage

```
plot_mod_mix(p_obj, x, event, mix_output, distribution = c("weibull",
  "lognormal", "loglogistic"), title_trace = "Fit")
```

Arguments

p_obj	a plotly object provided by function plot_prob_mix.
x	a numeric vector containing the x-coordinates of the regression line.
event	a vector of binary data (0 or 1) indicating whether unit i is a right censored observation (= 0) or a failure (= 1).
mix_output	a list provided by mixmod_regression or mixmod_em, which consists of elements necessary to visualize the regression lines.
distribution	supposed distribution of the random variable. For output provided by mixmod_em distribution must be "weibull". Can be "weibull", "lognormal" or "loglogistic" for output provided mixmod_regression. Other distributions have not been implemented yet.
title_trace	a character string whis is assigned to the trace shown in the legend.

Details

The name of the legend entry is a combination of the title_trace and the number of determined subgroups. If title_trace = "Line" and the data could be splitted in two groups, the legend entries would be "Line 1" and "Line 2".

Value

Returns a plotly object containing the probability plot with plotting positions and estimated regression line(s).

References

Doganaksoy, N.; Hahn, G.; Meeker, W. Q., Reliability Analysis by Failure Mode, Quality Progress, 35(6), 47-52, 2002

Examples

```
# Data is taken from given reference:
hours <- c(2, 28, 67, 119, 179, 236, 282, 317, 348, 387, 3, 31, 69, 135,
           191, 241, 284, 318, 348, 392, 5, 31, 76, 144, 203, 257, 286,
           320, 350, 412, 8, 52, 78, 157, 211, 261, 298, 327, 360, 446,
           13, 53, 104, 160, 221, 264, 303, 328, 369, 21, 64, 113, 168,
           226, 278, 314, 328, 377)
```

```
state <- c(1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1,
          1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 0,
          1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
          0, 1, 1, 1, 1, 1, 1)
id <- 1:length(hours)

# Example 1 - Using result of mixmod_em in mix_output:
mix_mod_em <- mixmod_em(x = hours, event = state, distribution = "weibull",
                        conf_level = 0.95, k = 2, method = "EM", n_iter = 150)

plot_weibull_em <- plot_prob_mix(x = hours,
                                 event = state,
                                 id = id,
                                 distribution = "weibull",
                                 mix_output = mix_mod_em,
                                 title_main = "Weibull Mixture EM",
                                 title_x = "Time in Hours",
                                 title_y = "Probability of Failure",
                                 title_trace = "Subgroup")

plot_weibull_emlines <- plot_mod_mix(p_obj = plot_weibull_em,
                                   x = hours,
                                   event = state,
                                   mix_output = mix_mod_em,
                                   distribution = "weibull",
                                   title_trace = "Fitted Line")

# Example 2 - Using result of mixmod_regression in mix_output:
john <- johnson_method(x = hours, event = state)
mix_mod_reg <- mixmod_regression(x = john$characteristic,
                                 y = john$prob,
                                 event = john$status,
                                 distribution = "weibull")

plot_weibull_reg <- plot_prob_mix(x = hours,
                                  event = state,
                                  id = id,
                                  distribution = "weibull",
                                  mix_output = mix_mod_reg,
                                  title_main = "Weibull Mixture Regression",
                                  title_x = "Time in Hours",
                                  title_y = "Probability of Failure",
                                  title_trace = "Subgroup")

plot_weibull_reglines <- plot_mod_mix(p_obj = plot_weibull_reg,
                                    x = hours,
                                    event = state,
                                    mix_output = mix_mod_reg,
                                    distribution = "weibull",
                                    title_trace = "Fitted Line")
```

plot_pop *Add Population Line to an Existing Grid*

Description

This function adds a linearized CDF to an existing plotly grid.

Usage

```
plot_pop(p_obj, x, params, distribution = c("weibull", "lognormal",
  "loglogistic"), color = I("#FF0000"), title_trace = "Population")
```

Arguments

p_obj	a plotly object, which at least includes the layout provided by plot_layout.
x	a numeric vector containing the x-coordinates of the population line.
params	a (named) numeric vector, where the first entry is the location parameter μ and the second entry is the scale parameter σ of a lognormal or loglogistic distribution. If the distribution value is "weibull" the first entry must be the scale parameter η and the second entry must be the shape parameter β. Parametrization is the same used in rweibull.
distribution	supposed distribution of the random variable. The value can be "weibull", "lognormal" or "loglogistic". Other distributions have not been implemented yet.
color	the color of the population line should be added as follows: For hexadecimal codes: color = I("#3C8DBC") and for a color specified with a string: color = I("blue").
title_trace	a character string which is assigned to the trace shown in the legend.

Value

A plotly object which contains the supposed linearized population CDF. Failure probabilities must be strictly below 1 and for this very reason

Examples

```
x <- rweibull(n = 100, shape = 1, scale = 20000)
grid_weibull <- plot_layout(x = x,
                            distribution = "weibull",
                            title_main = "Weibull Analysis",
                            title_x = "Time to Failure",
                            title_y = "Failure Probability")
pop_weibull <- plot_pop(p_obj = grid_weibull,
                        x = x, params = c(20000, 1),
                        distribution = "weibull", color = I("green"),
                        title_trace = "Population")
```

plot_prob *Probability Plotting Method for Univariate Lifetime Distributions*

Description

This function is used to apply the graphical technique of probability plotting.

Usage

```
plot_prob(x, y, event, id = rep("XXXXXX", length(x)),
  distribution = c("weibull", "lognormal", "loglogistic", "normal",
  "logistic", "sev"), title_main = "Probability Plot",
  title_x = "Characteristic", title_y = "Unreliability",
  title_trace = "Sample")
```

Arguments

x	a numeric vector which consists of lifetime data. Lifetime data could be every characteristic influencing the reliability of a product, e.g. operating time (days/months in service), mileage (km, miles), load cycles.
y	a numeric vector which consists of estimated failure probabilities regarding the lifetime data in x.
event	a vector of binary data (0 or 1) indicating whether unit *i* is a right censored observation (= 0) or a failure (= 1).
id	a character vector for the identification of every unit.
distribution	supposed distribution of the random variable. The value can be "weibull", "lognormal", "loglogistic", "normal", "logistic" or "sev" *(smallest extreme value)*. Other distributions have not been implemented yet.
title_main	a character string which is assigned to the main title of the plot.
title_x	a character string which is assigned to the title of the x axis.
title_y	a character string which is assigned to the title of the y axis.
title_trace	a character string whis is assigned to the trace shown in the legend.

Details

The marker label for x is determined by the first word provided in the argument title_x, i.e. if title_x = "Mileage in km" the x label of the marker is "Mileage".

The marker label for y is determined by the string provided in the argument title_y, i.e. if title_y = "Probability in percent" the y label of the marker is "Probability".

Value

Returns a plotly object containing the layout of the probability plot provided by plot_layout and the plotting positions.

References

Meeker, William Q; Escobar, Luis A., Statistical methods for reliability data, New York: Wiley series in probability and statistics, 1998

Examples

```
# Alloy T7987 dataset taken from Meeker and Escobar(1998, p. 131)
cycles   <- c(300, 300, 300, 300, 300, 291, 274, 271, 269, 257, 256, 227, 226,
              224, 213, 211, 205, 203, 197, 196, 190, 189, 188, 187, 184, 180,
              180, 177, 176, 173, 172, 171, 170, 170, 169, 168, 168, 162, 159,
              159, 159, 159, 152, 152, 149, 149, 144, 143, 141, 141, 140, 139,
              139, 136, 135, 133, 131, 129, 123, 121, 121, 118, 117, 117, 114,
              112, 108, 104, 99, 99, 96, 94)
state <- c(rep(0, 5), rep(1, 67))

df_john <- johnson_method(x = cycles, event = state)

# Example 1: Probability Plot Weibull:
plot_weibull <- plot_prob(x = df_john$characteristic,
                          y = df_john$prob,
                          event = df_john$status,
                          id = df_john$id,
                          distribution = "weibull",
                          title_main = "Weibull Analysis",
                          title_x = "Cycles",
                          title_y = "Probability of Failure in %",
                          title_trace = "Failed Items")

# Example 2: Probability Plot Lognormal:
plot_lognormal <- plot_prob(x = df_john$characteristic,
                          y = df_john$prob,
                          event = df_john$status,
                          id = df_john$id,
                          distribution = "lognormal",
                          title_main = "Lognormal Analysis",
                          title_x = "Cycles",
                          title_y = "Probability of Failure in %",
                          title_trace = "Failed Items")
```

---

plot_prob_mix *Probability Plot for Separated Mixture Models*

Description

This function is used to apply the graphical technique of probability plotting to univariate mixture models that where separated with functions `mixmod_regression` or `mixmod_em`.

Usage

```
plot_prob_mix(x, event, id = rep("XXXXXX", length(x)),
  distribution = c("weibull", "lognormal", "loglogistic"),
  mix_output = NULL, title_main = "Probability Plot",
  title_x = "Characteristic", title_y = "Unreliability",
  title_trace = "Sample")
```

Arguments

x	a numeric vector which consists of lifetime data. Lifetime data could be every characteristic influencing the reliability of a product, e.g. operating time (days/months in service), mileage (km, miles), load cycles.
event	a vector of binary data (0 or 1) indicating whether unit *i* is a right censored observation (= 0) or a failure (= 1).
id	a character vector for the identification of every unit.
distribution	supposed distribution of the random variable. For output provided by mixmod_em distribution must be "weibull". Can be "weibull", "lognormal" or "loglogistic" for output provided mixmod_regression. Other distributions have not been implemented yet.
mix_output	a list provided by mixmod_regression or mixmod_em, which consists of values necessary to visualize the subgroups.The default value of mix_output is NULL.
title_main	a character string which is assigned to the main title of the plot.
title_x	a character string which is assigned to the title of the x axis.
title_y	a character string which is assigned to the title of the y axis.
title_trace	a character string whis is assigned to the trace shown in the legend.

Details

Depending on the separation method the function johnson_method is called in various ways. If mixmod_regression is used, johnson_method is applied to all data. If data was splitted by mixmod_em the function johnson_method is applied to subgroup-specific data. The calculated plotting positions are colored regarding the obtained split of the used splitting function. If mix_output = NULL johnson_method is applied to all data, too. The obtained plot is then equal to plot_prob. See **Examples** for all three cases.

In mixmod_regression a maximum of three subgroups can be determined and thus being plotted. The intention of this function is to give the user a hint for the existence of a mixture model. An in-depth analysis should be done afterwards.

The marker label for x is determined by the first word provided in the argument title_x, i.e. if title_x = "Mileage in km" the x label of the marker is "Mileage".

The marker label for y is determined by the string provided in the argument title_y, i.e. if title_y = "Probability in percent" the y label of the marker is "Probability".

The name of the legend entry is a combination of the title_trace and the number of determined subgroups. If title_trace = "Group" and the data could be splitted in two groups, the legend entries would be "Group 1" and "Group 2".

Value

Returns a plotly object containing the layout of the probability plot provided by plot_layout and the plotting positions.

References

Doganaksoy, N.; Hahn, G.; Meeker, W. Q., Reliability Analysis by Failure Mode, Quality Progress, 35(6), 47-52, 2002

Examples

```
# Data is taken from given reference:
hours <- c(2, 28, 67, 119, 179, 236, 282, 317, 348, 387, 3, 31, 69, 135,
          191, 241, 284, 318, 348, 392, 5, 31, 76, 144, 203, 257, 286,
          320, 350, 412, 8, 52, 78, 157, 211, 261, 298, 327, 360, 446,
          13, 53, 104, 160, 221, 264, 303, 328, 369, 21, 64, 113, 168,
          226, 278, 314, 328, 377)
state <- c(1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1,
          1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 0,
          1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
          0, 1, 1, 1, 1, 1, 1)
id <- 1:length(hours)

# Example 1 - mix_output = NULL:
plot_weibull <- plot_prob_mix(x = hours,
                              event = state,
                              id = id,
                              distribution = "weibull",
                              mix_output = NULL,
                              title_main = "Weibull Probability Plot",
                              title_x = "Time in Hours",
                              title_y = "Probability of Failure",
                              title_trace = "Failed Items")

# Example 2 - Using result of mixmod_em in mix_output:
mix_mod_em <- mixmod_em(x = hours, event = state, distribution = "weibull",
                        conf_level = 0.95, k = 2, method = "EM", n_iter = 150)

plot_weibull_em <- plot_prob_mix(x = hours,
                                 event = state,
                                 id = id,
                                 distribution = "weibull",
                                 mix_output = mix_mod_em,
                                 title_main = "Weibull Mixture EM",
                                 title_x = "Time in Hours",
                                 title_y = "Probability of Failure",
                                 title_trace = "Subgroup")

# Example 3 - Using result of mixmod_regression in mix_output:
john <- johnson_method(x = hours, event = state)
mix_mod_reg <- mixmod_regression(x = john$characteristic,
                                 y = john$prob,
```

```
                                  event = john$status,
                                  distribution = "weibull")

plot_weibull_reg <- plot_prob_mix(x = hours,
                                  event = state,
                                  id = id,
                                  distribution = "weibull",
                                  mix_output = mix_mod_reg,
                                  title_main = "Weibull Mixture Regression",
                                  title_x = "Time in Hours",
                                  title_y = "Probability of Failure",
                                  title_trace = "Subgroup")
```

`predict_prob` *Estimation of Failure Probabilities for Parametric Lifetime Distributions*

Description

This function estimates the failure probabilities for a given set of estimated location-scale (and threshold) parameters and specified quantiles.

Usage

```
predict_prob(q, loc_sc_params, distribution = c("weibull", "lognormal",
  "loglogistic", "normal", "logistic", "sev", "weibull3", "lognormal3",
  "loglogistic3"))
```

Arguments

`q`	a numeric vector which consists of lifetime data.
`loc_sc_params`	a (named) numeric vector of estimated location and scale parameters for a specified distribution. The order of elements is important. First entry needs to be the location parameter μ and the second element needs to be the scale parameter σ. If a three-parametric model is used the third element is the threshold parameter γ.
`distribution`	supposed distribution of the random variable. The value can be `"weibull"`, `"lognormal"`, `"loglogistic"`, `"normal"`, `"logistic"`, `"sev"` *(smallest extreme value)*, `"weibull3"`, `"lognormal3"` or `"loglogistic3"`. Other distributions have not been implemented yet.

Value

A vector containing the estimated failure probabilities for a given set of quantiles and estimated parameters.

Examples

```
# Example 1: Predicted probabilities for two-parameter Weibull:
probs <- predict_prob(q = c(15, 48, 124), loc_sc_params = c(5, 0.5),
                          distribution = "weibull")

# Example 2: Predicted probabilities for three-parameter Weibull:
probs_weib3 <- predict_prob(q = c(25, 58, 134), loc_sc_params = c(5, 0.5, 10),
                          distribution = "weibull3")
```

`predict_quantile` *Estimation of Quantiles for Parametric Lifetime Distributions*

Description

This function estimates the quantiles for a given set of estimated location-scale (and threshold) parameters and specified failure probabilities.

Usage

```
predict_quantile(p, loc_sc_params, distribution = c("weibull",
  "lognormal", "loglogistic", "normal", "logistic", "sev", "weibull3",
  "lognormal3", "loglogistic3"))
```

Arguments

`p`	a numeric vector which consists of failure probabilities regarding the lifetime data.
`loc_sc_params`	a (named) numeric vector of estimated location and scale parameters for a specified distribution. The order of elements is important. First entry needs to be the location parameter μ and the second element needs to be the scale parameter σ. If a three-parametric model is used the third element is the threshold parameter γ.
`distribution`	supposed distribution of the random variable. The value can be `"weibull"`, `"lognormal"`, `"loglogistic"`, `"normal"`, `"logistic"`, `"sev"` *(smallest extreme value)*, `"weibull3"`, `"lognormal3"` or `"loglogistic3"`. Other distributions have not been implemented yet.

Value

A vector containing the estimated quantiles for a given set of failure probabilities and estimated parameters.

Examples

```
# Example 1: Predicted quantiles for two-parameter Weibull:
quants <- predict_quantile(p = c(0.01, 0.1, 0.5), loc_sc_params = c(5, 0.5),
                           distribution = "weibull")

# Example 2: Predicted quantiles for three-parameter Weibull:
quants_weib3 <- predict_quantile(p = c(0.01, 0.1, 0.5), loc_sc_params = c(5, 0.5, 10),
                           distribution = "weibull3")
```

`rank_regression` *Rank Regression for Parametric Lifetime Distributions*

Description

This method fits an **x on y** regression to a linearized two- or three-parameter cdf and is applicable for complete and (multiple) right censored data. The parameters are estimated in the frequently used (log-) location-scale parameterization. For the Weibull, estimates are transformed such that they are in line with the parameterization provided by the *stats* package like `pweibull`.

Usage

```
rank_regression(x, y, event, distribution = c("weibull", "lognormal",
  "loglogistic", "normal", "logistic", "sev", "weibull3", "lognormal3",
  "loglogistic3"), conf_level = 0.95, details = TRUE)
```

Arguments

`x`	a numeric vector which consists of lifetime data. Lifetime data could be every characteristic influencing the reliability of a product, e.g. operating time (days/months in service), mileage (km, miles), load cycles.
`y`	a numeric vector which consists of estimated failure probabilities regarding the lifetime data in `x`.
`event`	a vector of binary data (0 or 1) indicating whether unit *i* is a right censored observation (= 0) or a failure (= 1).
`distribution`	supposed distribution of the random variable. The value can be `"weibull"`, `"lognormal"`, `"loglogistic"`, `"normal"`, `"logistic"`, `"sev"` *(smallest extreme value)*, `"weibull3"`, `"lognormal3"` or `"loglogistic3"`. Other distributions have not been implemented yet.
`conf_level`	confidence level of the interval. The default value is `conf_level = 0.95`.
`details`	a logical variable, where the default value is `TRUE`. If `FALSE` the output consists of a list that only contains the estimated parameters. If `TRUE` the output is a detailed list with many more information. See below (**Value**).

Details

When using this method, the approximated confidence intervals for the Weibull parameters (based on p. 51 of Ralf Mock) can only be estimated for the following confidence levels:

- `conf_level` = 0.90,
- `conf_level` = 0.95,
- `conf_level` = 0.99.

If the distribution is not the Weibull, the confidence intervals of the parameters are calculated using a heteroscedasticity-consistent covariance matrix. Here it should be said that there is no statistical foundation to calculate the standard errors for the parameters using *Least Squares* in context of *Median Rank Regression*. For an accepted statistical method use MLE (`ml_estimation`).

Value

Returns a list with the following components (depending on `details` argument):

- `coefficients` : Provided, if `distribution` is `"weibull"` or `"weibull3"`. η is the estimated scale and β the estimated shape parameter. The estimated threshold parameter γ is available if the three-parametric weibull is used.
- `confint` : Provided, if `distribution` is `"weibull"` or `"weibull3"`. Confidence intervals for η and β (and γ if the three-parametric weibull is used).
- `loc_sc_coefficients` : Estimated location-scale parameters. Threshold parameter is provided for `"weibull3"`, `"lognormal3"` and `"loglogistic3"`.
- `loc_sc_confint` : Confidence intervals for location-scale parameters. If distribution is `"lognormal3"` or `"loglogistic3"` a confidence interval for the threshold is not computed.
- `loc_sc_vcov` : Provided, if `distribution` is not `"weibull"` or `"weibull3"`. Estimated heteroscedasticity-consistent Variance-Covariance matrix of the used location-scale distribution.
- `r_squared` : Coefficient of determination.

References

- Mock, R., Methoden zur Datenhandhabung in Zuverlässigkeitsanalysen, vdf Hochschulverlag AG an der ETH Zürich, 1995
- Meeker, William Q; Escobar, Luis A., Statistical methods for reliability data, New York: Wiley series in probability and statistics, 1998

Examples

```
# Example 1: Fitting a two-parameter Weibull:
obs   <- seq(10000, 100000, 10000)
state <- c(0, 1, 1, 0, 0, 0, 1, 0, 1, 0)

df_john <- johnson_method(x = obs, event = state)
mrr <- rank_regression(x = df_john$characteristic,
                       y = df_john$prob,
                       event = df_john$status,
```

```
                      distribution = "weibull",
                      conf_level = .90)

# Example 2: Fitting a three-parameter Weibull:
# Alloy T7987 dataset taken from Meeker and Escobar(1998, p. 131)
cycles   <- c(300, 300, 300, 300, 300, 291, 274, 271, 269, 257, 256, 227, 226,
              224, 213, 211, 205, 203, 197, 196, 190, 189, 188, 187, 184, 180,
              180, 177, 176, 173, 172, 171, 170, 170, 169, 168, 168, 162, 159,
              159, 159, 159, 152, 152, 149, 149, 144, 143, 141, 141, 140, 139,
              139, 136, 135, 133, 131, 129, 123, 121, 121, 118, 117, 117, 114,
              112, 108, 104, 99, 99, 96, 94)
state <- c(rep(0, 5), rep(1, 67))

df_john <- johnson_method(x = cycles, event = state)
mrr <- rank_regression(x = df_john$characteristic,
                      y = df_john$prob,
                      event = df_john$status,
                      distribution = "weibull3",
                      conf_level = .90)
```

`r_squared_profiling` *R-Profile Function for Log-Location-Scale Distributions with Threshold*

Description

This function evaluates the coefficient of determination with respect to a given threshold parameter of a three-parametric lifetime distribution. In terms of *Median Rank Regression* this function can be optimized (`optim`) to estimate the threshold parameter.

Usage

```
r_squared_profiling(x, y, thres, distribution = c("weibull3",
  "lognormal3", "loglogistic3"))
```

Arguments

`x`	a numeric vector which consists of lifetime data. Lifetime data could be every characteristic influencing the reliability of a product, e.g. operating time (days/months in service), mileage (km, miles), load cycles.
`y`	a numeric vector which consists of estimated failure probabilities regarding the lifetime data in `x`.
`thres`	a numeric value of the threshold parameter.
`distribution`	supposed distribution of the random variable. The value can be `"weibull3"`, `"lognormal3"` or `"loglogistic3"`.

Value

Returns the coefficient of determination for a specified threshold value.

References

Meeker, William Q; Escobar, Luis A., Statistical methods for reliability data, New York: Wiley series in probability and statistics, 1998

Examples

```
# Alloy T7987 dataset taken from Meeker and Escobar(1998, p. 131)
cycles   <- c(300, 300, 300, 300, 300, 291, 274, 271, 269, 257, 256, 227, 226,
              224, 213, 211, 205, 203, 197, 196, 190, 189, 188, 187, 184, 180,
              180, 177, 176, 173, 172, 171, 170, 170, 169, 168, 168, 162, 159,
              159, 159, 159, 152, 152, 149, 149, 144, 143, 141, 141, 140, 139,
              139, 136, 135, 133, 131, 129, 123, 121, 121, 118, 117, 117, 114,
              112, 108, 104, 99, 99, 96, 94)
state <- c(rep(0, 5), rep(1, 67))

df_john <- johnson_method(x = cycles, event = state)

# Determining threshold parameter for which the coefficient of determination is
# maximized subject to the condition that the threshold parameter must be smaller
# as the first failure cycle, i.e 94:
threshold <- seq(0, min(cycles[state == 1]) - 0.1, length.out = 100)
profile_r2 <- sapply(threshold, r_squared_profiling,
                     x = df_john$characteristic[df_john$status == 1],
                     y = df_john$prob[df_john$status == 1],
                     distribution = "weibull3")
threshold[which.max(profile_r2)]

# plot:
# plot(threshold, profile_r2, type = "l")
# abline(v = threshold[which.max(profile_r2)], h = max(profile_r2), col = "red")
```

`weibulltools` *weibulltools*

Description

The weibulltools package contains methods for examining bench test or field data using the well-known weibull analysis.

Index