Package 'extWeibQuant'

February 19, 2015

Type Package

Title Estimate Lower Extreme Quantile with the Censored Weibull MLE and Censored Weibull Mixture

Version 1.1

Date 2014-12-03

Author Yang (Seagle) Liu

Maintainer Yang (Seagle) Liu <yang.liu@stat.ubc.ca>

Description It implements the subjectively censored Weibull MLE and censored Weibull mixture methods for the lower quantile estimation. Quantile estimates from these two methods are robust to model misspecification in the lower tail. It also includes functions to evaluation the standard error of the resulting quantile estimates. Also, the methods here can be used to fit the Weibull or Weibull mixture for the Type-I or Type-II right censored data.

License GPL (>= 2)

NeedsCompilation yes

Repository CRAN

Date/Publication 2014-12-04 01:14:47

R topics documented:

`bootstrapCenWbMix` *Bootstrap Censored Weibull Mixture for censoring threshold selection and standard error of the quantile estimates*

Description

Similarly as in `bootstrapCMLE`, We will bootstrap the data set to obtain the MSE and SE of the quantile estimate under a certain subjective censoring threshold. Warning: The computation burden of this function is extremely heavy.

Usage

```
bootstrapCenWbMix(dat, qInt = 0.05, canSet = c(0.5, 0.7, 1), B = 1000,
iniVec = NULL, randSeed = NULL, conCr = 1e-06, nIter = 1000)
```

Arguments

`dat`	The complect data set. See `bootstrapCMLE`
`qInt`	Quantile of interest. See `bootstrapCMLE`
`canSet`	Candidate set of thresholds. See `bootstrapCMLE`
`B`	Number of bootstrap replicates
`iniVec`	A vector of length 6*length(`canSet`), supplying the initial values of the EM algorithm under each censoring threshold for the original data set. Should be orginized in the following way (p, 1-p, a_1, a_2, b_1, b_2, p, 1-p, ...). If null, the initial values will be generated randomly, which is not recommended.
`randSeed`	See `bootstrapCMLE`
`conCr`	Same as in `emCenWbMix.T1`
`nIter`	Same as in `emCenWbMix.T1`

Details

The EM algorithm for Weibull mixture is extremely senstive to the initial value, so the user should always try to find proper initial values for each threshold and data set.

The initial values in the argument are for the original data set. The EM algorithm for the original data set will be first carried out from these initial values. If they succesfully converged, the EM algorithm for the bootstrap data sets will start from the estimates of the original data set in the previous step. Otherwise, the EM algorithm will start from the initial value in the arguements. Although we could not gurantee the "global" maximum can be reached for each bootstrap replicate under this scheme, our simulation shown that it works reasonably well.

Similarly as `bootstrapCMLE`, the main calculation of this function is done in C.

Value

`results`	A matrix of length(`canSet`) by four. The first column is the candidate threshold (proportion). The second column is the quantile estimate under this censoring threshold. The third column and fourth column are the bootstrap estimate of the standard error (SE) and root mean squared error (RMSE) of this quantile estimate.
`parameters`	A matrix of length(`canSet`) by six, recording the parameter estimates of the Weibull mixture under each threshold.
`bQEst`	A matrix of B-by-length(`canSet`). The quantile estimates under each censoring threshold for each bootstrap replicate.

Note

Please report the numerical problems and inconvenience when using this function to the author.

Author(s)

Yang (Seagle) Liu <yang.liu@stat.ubc.ca>

References

Liu Y. (2012). Lower Quantile Estimation of Wood Strength Data. *Master Thesis, Department of Statistics, UBC*. Downloadable here.

See Also

`bootstrapCMLE`

Examples

```
###
mmix = rbind(c(0.7, 5, 7),c(0.3, 15, 6))
vmix = c(mmix) #A vector version of the paramters
set.seed(1)
y <- sort(simWbMix(100, mmix)) #Generate Data
#Run the EM for 70% Type II censoring
mix70 <- emCenWbMix.T2(y[1:70], 100, iniParam=vmix, useC=TRUE)
#Run the EM for the complete data set
mix100<- emCenWbMix.T2(y, 100, iniParam=vmix, useC=TRUE)
#Use bootstrap with the previous estimates as initials
bootstrapCenWbMix(y, canSet=c(0.7, 1),
iniVec=c(c(mix70$estimates), c(mix100$estimates)), B=5,randSeed=1)
#Or use the "true" value as the initials.
bootstrapCenWbMix(y, canSet=c(0.7, 1), iniVec=c(vmix, vmix), B=5,randSeed=1)
#B=5 is just for illustration. It should be at least 1000.
```

`bootstrapCMLE` *Bootstrap Censored Weibull MLE for censoring threshold selection and standard error of the quantile estimates*

Description

Proposed in Chapter 5 of Liu (2012). The original data set will be bootstrapped to obtain an estimate of the mean squared error of the quantile estimates, such that quantile estimates under different subjective censoring thresholds could be compared. We could also obtain the standard error of quantile estimates via bootstrap.

Usage

```
bootstrapCMLE(dat, qInt = 0.05, canSet = seq(0.1, 0.5, by = 0.1),
B = 5000, randSeed = NULL, conCr = 1e-09, nIter = 1000)
```

Arguments

`dat`	A complete data set without censoring. See `Details`
`qInt`	The quantile of interest, e.g. 5% quantile or 10% quantile.
`canSet`	A vector of the candidate subjective censoring thresholds, expressed as the proportion of data smaller it, e.g. 10%, 20%, ..., 100% (non-censoring)
`B`	Number of bootstrap replicate data sets
`randSeed`	The seed for random number generation. If NULL, the random seed will not be set
`conCr`	Convergence criterion for the algorithm to calculate the censored Weibull MLE. See `cenWbMLE.T1`
`nIter`	See `cenWbMLE.T1`

Details

This function is designed to only work for a complete data set where every observation is fully observed (non-censoring). We could decide the best threshold (proportion) of subjective censoring based on bootstrap. For the advantage of subjective censoring, please see Liu (2012).

This function will call C to do all calculations. So it is recommended that the user should make sure the `cenWbMLE.T2` could work for their original data set.

Value

`results`	A matrix of length(`canSet`) by six. The first column is the candidate threshold (proportion). The second and third column is the parameter estimates of the Weibull model for the original data set. The fourth column is the quantile estimate under this censoring threshold. The fifth column and sixth column are the bootstrap estimate of the standard error (SE) and root mean squared error (RMSE) of this quantile estimate.
`bQEst`	A matrix of B-by-length(`canSet`). The quantile estimates under each censoring threshold for each bootstrap replicate.

Note

Please report the numerical problems and inconvenience when using this function to the author.

Author(s)

Yang (Seagle) Liu <yang.liu@stat.ubc.ca>

References

Liu Y. (2012). Lower Quantile Estimation of Wood Strength Data. *Master Thesis, Department of Statistics, UBC*. Downloadable here.

See Also

bootstrapCenWbMix

Examples

```
set.seed(1)
y <- sort(rweibull(100, 7, 7))
tlist <- bootstrapCMLE(y, B=1000, canSet=c(0.1, 0.5, 1), randSeed=1)
tlist$results #Usually, we only need to look at the results part.
```

---

cenWbMLE.T1 *censored Weibull MLE for Type I right-censored data*

Description

Calculate the censored Weibull (two parameter, shape and scale) MLE for the Type I censored data with the algorithm described in ASTM 5457(2004). Return the estimates of the shape and scale parameters. A comprehesive description of this algorithm can be found in Liu (2012).

Usage

```
cenWbMLE.T1(dat, Cx=NULL, useC = FALSE, conCr = 1e-09, nIter = 1000)
```

Arguments

dat	A vector of the data, should not contain any negative observations, but NA is allowed. The NA's and observations larger than the censoring threshold Cx will be censored in the calculation. The algorithm will decide the original sample size (before censoring) from the length of this vector.
Cx	The censoring threshold of Type I right-censoring. If NULL, the complete (un-censored MLE) will be calculated.
useC	Default to be false and use the R routine to calculate estimates. If true, the function will use the C routine, which is much faster than the R routine, but harder for the user to identify the numerical issues (if there is any).

conCr	In terms of the relative change in the negative log-likelihood. The algorithm is viewed as converged if the relative change is smaller than conCr
nIter	The maximum numer of iterations allowed in the function.

Value

convergence	an integer indicating why the algorithm terminated • 0, successfully converged; • 1, Numerical over/under-flow, maybe the shape parameter diverges to infinity; • 2, Iteration limit reached; • 3, Shape parameter becomes negative; • 4, Shape parameter diverges to infinity.
estimates	Shape and scale parameter estimates

Note

Please report the numerical problems and inconvenience when using this function to the author.

Author(s)

Yang (Seagle) Liu <yang.liu@stat.ubc.ca>

References

ASTM (2004). Standard specfication for computing reference resistance of wood-based materials and structural connections for load and resistance factor design D5457. *American Society for Testing Materials,* Philadephia, Pa.

Liu Y. (2012). Lower Quantile Estimation of Wood Strength Data. *Master Thesis, Department of Statistics, UBC.* Downloadable here.

See Also

rweibull, cenWbMLE.T2, emCenWbMix.T1

Examples

```
set.seed(1)
y <- sort(rweibull(100, 7, 7)) ##Generate the data
cenWbMLE.T1(y) #The MLE for the complete data
cenWbMLE.T1(y, 5) #Censor the data at 5 and calculate the censored MLE
##Or
newy <- rep(NA, 100)
newy[y<=5] <- y[y<=5] #Censor the data at 5
fit <- cenWbMLE.T1(newy, 5) #Calculate the censored MLE
qweibull(0.05, fit$estimates[1], fit$estimates[2])
#Calculate the 5% quantile of the fitted distribution.
```

`cenWbMLE.T2` *censored Weibull MLE for Type II right-censored data*

Description

Calculate the censored Weibull MLE for the Type II censored data with the algorithm described in ASTM 5457(2004). Return the estimates of the shape and scale parameters. A comprehesive description of this algorithm can be found in Liu (2012).

Usage

```
cenWbMLE.T2(dat, n, useC = FALSE, conCr = 1e-09, nIter = 1000)
```

Arguments

`dat`	A vector of the observations after censoring. None NA or negative values allowed.
`n`	The original sample size, including both the censored and uncensored observations.
`useC`	See `cenWbMLE.T1`
`conCr`	See `cenWbMLE.T1`
`nIter`	See `cenWbMLE.T1`

Value

See `cenWbMLE.T1`

Note

Please report the numerical problems and inconvenience when using this function to the author.

Please notice that the ways of inputing data in `cenWbMLE.T1` and `cenWbMLE.T2` are differnt. For `cenWbMLE.T1`, the algorithm require a full "orginal" data set (with the uncensored observations as NA or a arbitary value larger than the threshold) and the original sample size is decided as the length of the input data, while `cenWbMLE.T2` requires the observed data points and the original sample size.

Author(s)

Yang (Seagle) Liu <yang.liu@stat.ubc.ca>

References

See `cenWbMLE.T1`

See Also

`rweibull`, `cenWbMLE.T1`, `emCenWbMix.T2`

Examples

```
set.seed(1)
y <- sort(rweibull(100, 7, 7)) ##Generate the data
cenWbMLE.T2(y, 100) #The MLE for the complete data
cenWbMLE.T2(y[1:10], 100) #Censor the largerst 90% of the data.
```

`emCenWbMix.T1` *EM algorithm for the Type-I right censored Weibull Mixture with two populations.*

Description

EM algorithm to estimate the parameters of a mixture of two two-parameter Weibull distributions for possibly Type-I right censored data. The PDF of this mixture distribution is

$$f(x) = p(a_1/b_1)(x/b_1)^(a_1 - 1)exp(-(x/b_1)_1^a) + (1-p)(a_2/b_2)(x/b_2)^(a_2 - 1)exp(-(x/b_2)_2^a),$$

p is the proportion of the first sub-population. a_1, a_2 are the shape parameters for the two sub-populations. b_1, b_2 are the scale parameters. More details of the mixture model and this algorithm can be found in Liu (2012).

Usage

```
emCenWbMix.T1(dat, Cx=NULL, iniParam = NULL, useC = FALSE, conCr = 1e-06, nIter = 10000)
```

Arguments

`dat`	The data vector, should not contain any negative observations, but NA is allowed. The NA's and observations larger than the censoring threshold `Cx` will be censored in the calculation. The algorithm will decide the original sample size (before censoring) from the length of this vector.
`Cx`	The censoring threshold of Type I right-censoring. If NULL, the uncensored mixture will be estimated.
`iniParam`	A vector of length six (p, 1-p, a_1, a_2, b_1, b_2). All the elements must be postive. If not provided, the algorithm will generate them randomly. See `Details`.
`useC`	See `cenWbMLE.T1`
`conCr`	See `cenWbMLE.T1`
`nIter`	See `cenWbMLE.T1`

Details

It is well known that the EM algorithm is highly influenced by the initial value of the parameters. We strongly recommend the users provide some proper/reasonable initial values for their data sets. It is usually very difficult for EM algorithm to converge from the random initials generated in this function. Also, it is better to try starting this function from different initial values to obtain a "global" maximum of the likelihood.

Value

convergence	An integer indicating why the algorithm terminated • 0, successfully converged; • 1-4, Error in the maximization step. Same error code as in cenWbMLE.T1; • 5, Numerically under/over flow, maybe the current parameters do not fit the data; • 6, EM Iteration limit reached.
nllh	The negative log-likelihood
estimates	A 2 by 3 matrix of the parameter estimates: p, a_1, b_1; 1-p, a_2, b_2;
iniParam	A copy of the iniParam

Note

Please report the numerical problems and inconvenience when using this function to the author.

Author(s)

Yang (Seagle) Liu <yang.liu@stat.ubc.ca>

References

Liu Y. (2012). Lower Quantile Estimation of Wood Strength Data. *Master Thesis, Department of Statistics, UBC*. Downloadable here.

See Also

rweibull, quanWbMix, simWbMix, cenWbMLE.T1, emCenWbMix.T2

Examples

```
mmix = rbind(c(0.7, 5, 7),c(0.3, 15, 6))
vmix = c(mmix) #A vector version of the paramters
set.seed(1)
y <- sort(simWbMix(300, mmix)) #Generate the data
#The uncensored mixture
emCenWbMix.T1(y, iniParam=vmix, useC=TRUE)
#The mixture if we censor the data around 9.
emCenWbMix.T1(y, Cx=9, iniParam=vmix, useC=TRUE)
```

emCenWbMix.T2 *EM algorithm for the Type-II right censored Weibull Mixture with two populations.*

Description

This is the interface for the EM algorithm of censored Weibull mixture for Type II censored data. The description of the Weibull mixture model can be found in emCenWbMix.T1.

Usage

```
emCenWbMix.T2(dat, n, iniParam = NULL, useC = FALSE, conCr = 1e-06, nIter = 10000)
```

Arguments

dat	A vector of the observations after censoring. None NA or negative values allowed.
n	The original sample size, including both the censored and uncensored observations. See cenWbMLE.T2.
iniParam	See emCenWbMix.T1
useC	See emCenWbMix.T1
conCr	See emCenWbMix.T1
nIter	See emCenWbMix.T1

Value

See emCenWbMix.T1

References

See emCenWbMix.T1

See Also

rweibull, quanWbMix, simWbMix, cenWbMLE.T2, emCenWbMix.T1

Examples

```
mmix = rbind(c(0.7, 5, 7),c(0.3, 15, 6))
vmix = c(mmix) #A vector version of the paramters
set.seed(1)
y <- sort(simWbMix(300, mmix)) #Generate the data
#The uncensored EM
emCenWbMix.T2(y, 300, iniParam=vmix, useC=TRUE)
#Type II censoring with the largest half of the data censored.
emCenWbMix.T2(y[1:150], 300, iniParam=vmix, useC=TRUE)
```

quanWbMix *Calculate the quantiles of a mixture of two Weibull distributions.*

Description

This is a function based on the uniroot to calculate the quantiles of a Weibull mixture model.

Usage

```
quanWbMix(intProb, mixParm)
```

Arguments

intProb	A vector of the probability below the quantile of interest, e.g. 5%
mixParm	A two-by-three matrix describing the Weibull mixture. See simWbMix.

Value

A two-row matrix: the first row is intProb and the second row is the corresponding quantiles.

See Also

qweibull, simWbMix, emCenWbMix.T1, emCenWbMix.T2

Examples

```
mmix = rbind(c(0.7, 5, 7),c(0.3, 15, 6))
quanWbMix(c(0.1, 0.5, 0.7), mmix)
quanWbMix(0.1, mmix)
```

simWbMix *Simulate data from a mixture of two Weibull distributions*

Description

Same as title.

Usage

```
simWbMix(n, mixParm)
```

Arguments

n	Sample size
mixParm	A two-by-three matrix describing the Weibull mixture p, a_1, b_1; 1-p, a_2, b_2

Value

A vector of length n

See Also

rweibull, quanWbMix, emCenWbMix.T1, emCenWbMix.T2

Examples

```
mmix = rbind(c(0.7, 5, 7),c(0.3, 15, 6))
set.seed(1)
y <- sort(simWbMix(300, mmix)) #Generate the data
hist(y)
```

Index