iwtp: Software for Analysis of Self-Selected Interval Data

Wenchao Zhou, Yuri Belyaev, and Bengt Kriström*

Version 1.0.0

2012-02-29

Contents

*Department of Forest Economics, Swedish University of Agricultural Sciences, 901 83 Umeå, Sweden; zhou.wenchao@sekon.slu.se

1 Introduction

In collecting data on willingness to pay (wtp) points, rather than asking a respondent to state an estimate of WTP - point or select one between given brackets, the respondent can freely self-select any interval of choice that contains his/her WTP - point. For the collected data, we found that presence of strong rounding is typical feature. The self-selected intervals can be considered as censoring and they may depend on the unobserved postitions of their WTP - points. Usually in the Survival Analysis it is assumed that the censoring intervals are independent of such points and cover only some of them. The considered data are lists of self-selected intervals with rounded ends. Many of them can be selected repeatly by may respondents. We suppose that respondents's WTP - points are values of independent identically distributed variables and we are interested to find consistent estimate of this distribution. We propose statistical models which admit dependency of censoring intervals on positions of corresponding censored WTP - points. The suggeted statistical models allow to imitate various possible behaviors (preferences) of respondents during selecting appropriated intervals containing their WTP - points. We have to distinguish the probability to select an interval containing WTP - point and the probabilty that the interval contains WTP - point. In obtaining likelihood function we use the division generated by the different self-selected intervals in the collected data. The divison is a set of intervals such that any self-selected interval is a union of these intervals. After that we consider different conditional probabilities to obtain self-selected intervals,given that the WTP - points are inside division's intervals. We estimate nonparametric WTP - survival functions and also find ML - estimates of parameters based on the Weibull and the mixed Weibull/Exponential parametric families of distributions. It can be viewed as an extension of popular in the Survival Analysis Turnbull's estimator Turnbull (1976). Refer to Belyaev and Kriström (2010) for details.

2 Setup

2.1 Software Requirements

`iwtp` works with the R project for Statistical Computing, and can run on any platform where R is installed (Windows, Mac, Linux). R is available free for download at the Comprehensive R Archive Network (CRAN) at http://cran.r-project.org/. The current version `iwtp` has been tested on the most recent version of R.

3 A user's Guide

We show here how to use `iwtp` by using a data set: the lax valuation data. In the following examples, we don't use the orginal data. In stead, we use a re-sampled version of the data set, and here we name it IL. For a detailed descripton of the valuation data, please refer to Belyaev, Håkansson and Kriström (2009). Briefly, in 2004 a contingent valuation study with interval questions was carried out in order to investigate the costs and benefits of changing in stream flow at the Stornorrfors hydropower plant on the Vindel River, in northern Sweden. In the questions, the respondents were asked about their WTP for increasing the number of salmon (lax in swedish) that reach their spawing grounds in the river each year. In our re-sampled data set IL, it contains 46 observations of self-selected intervals. This data set is used only for illustrative purpose. The data format is shown below.

3.1 Input Data Format

To use `iwtp`, the input data must follow a format identical to IL as shown below:

```
> require(iwtp)
> data(IL)
> IL[1:10,]

   respondent no. lower value upper value
1               1          10          50
2               2          50         100
3               3          50         100
4               4          50         100
5               5         100         200
6               6          50         100
7               7           0          50
8               8          20          50
9               9          50         100
10             10          50         100

> class(IL)

[1] "data.frame"
```

As shown in the above example, the input data is required to be of type data.frame. It should contain at least three columns: column 1 is the nr. of self-selected intervals; column 2 is the left-end value of the self-selected intervals; column 3 is the right-end value of the self-selected intervals.

3.2 Estimation of Survival Functions

3.2.1 The Weibull survival function

In the following, we show how to estimate the Weibull survival function by using function `iwtp`

```
> require(iwtp)
> data(IL)
> wtp.wb <- iwtp(IL,dist="weibull")
> wtp.wb

Respondent no.:  500
Self-selected interval no.: 46

Division intervals: 23 in total
(0, 5], (5, 10], (10, 15], (15, 20], (20, 25],
(25, 30], (30, 40], (40, 50], (50, 60], (60, 70],
(70, 75], (75, 80], (80, 100], (100, 150], (150, 170],
(170, 200], (200, 250], (250, 300], (300, 400], (400, 500],
(500, 600], (600, 1000], (1000, 2000]
Survival function: Weibull
Behavior model: BM5
Parameter estimates:
    Scale (a): 99.502888
    Shape (b): 1.075699
    Coefficient c: 17.963289
Maximum llik: -1481.514640
Mean WTP: 96.742760

Nonparametric estimate of the mean value of the empirical survival function
(based on the right-end values of self-selected intervals): 154.77

Nonparametric estimate of the mean value based on Turnbull estimator
[see Turnbull, 1976]: 91.30338

Estimated Mean WTP based on parametric Weibull model (taken self-selected
intervals as Turnbull intervals [see Turnbull, 1976]): 90.44379
```

In the above `R` code, we use the function `iwtp`, and specify the survival function as "weibull". The output object `wtp.wb` contains many usefull information, including three components: `sf` contains the values of parameters estimated for the Weibull distribution;

`esf` contains the data list of empirical survival function based on the right-end values of self-selcted intervals. `intv` contains the informaiton about division intervals generated by the data set. A summary of the content of the object `wtp.wb` is then shown by using `print`.

The following R code will plot the empirical survival function, and it will be drawn in the form of stepwise line with blue color, as shown in Figure 1.

```
> esfPlot(wtp.wb,xlab="Willingness to pay (SEK)",col="blue")
```

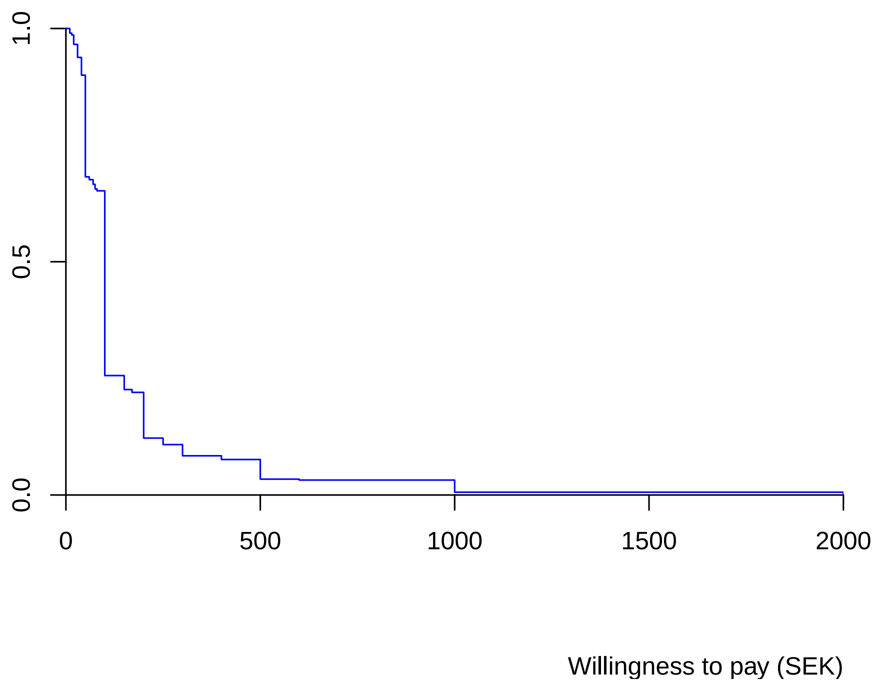

Figure 1: The empirical survival function based on the right-end values of self-selected intervals. Data: IL

In addition, we can show the estimated Weibull survival function by calling function `plot` as below:

```
> plot(wtp.wb,xlab="Willingness to pay (SEK)",col=c("blue","green"))
```

The estimated survival function (the Weibull) is shown in Figure 2.

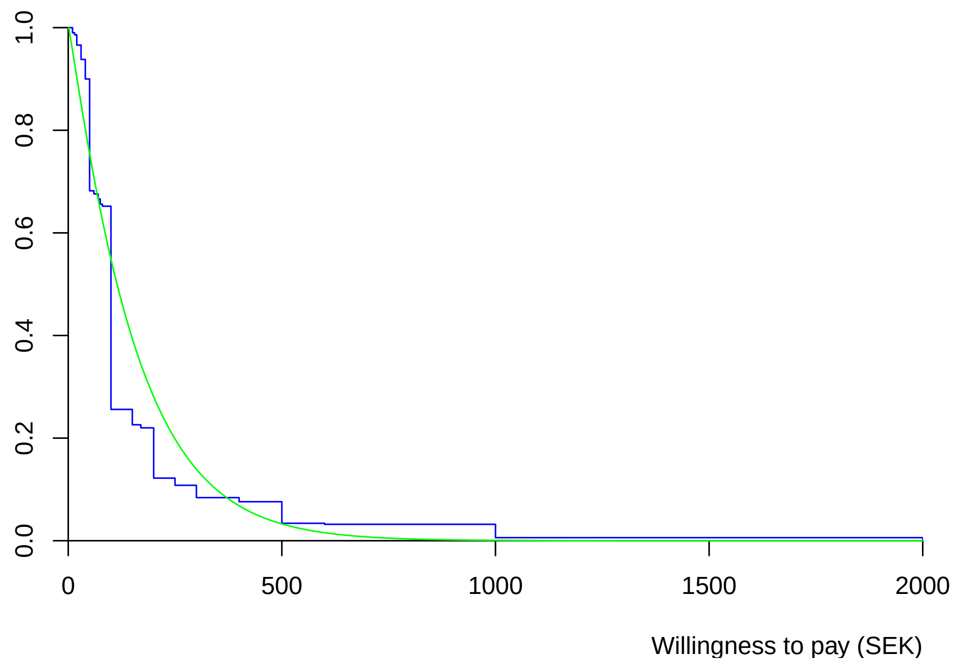

Figure 2: The estimated Weibull survival function (green), and the empirical survival function (blue). Data: IL

3.2.2 The mixed Weibull/exponential survival function

In the following, we show how to estimate the mixed Weibull and the exponential survival function by using function `iwtp`

```
> data(IL)
> wtp.wemix <- iwtp(IL,dist="wemix",plot=FALSE)
> print(wtp.wemix)

Respondent no.:  500
Self-selected interval no.: 46

Division intervals: 23 in total
(0, 5], (5, 10], (10, 15], (15, 20], (20, 25],
(25, 30], (30, 40], (40, 50], (50, 60], (60, 70],
(70, 75], (75, 80], (80, 100], (100, 150], (150, 170],
(170, 200], (200, 250], (250, 300], (300, 400], (400, 500],
(500, 600], (600, 1000], (1000, 2000]
Survival function: the mixture of Weibull/exponential
Bahavior model: BM5
Parameter estimates:
    mixing parameter p: 0.672290
    scale parameter a: 64.794383
    shape parameter b: 2.404809
    parameter m: 200.000000
    Coefficient c: 54.398761
Maximum llik: -1413.560768
Mean WTP: 104.159088

Nonparametric estimate of the mean value of the empirical survival function
(based on the right-end values of self-selected intervals): 154.77

Nonparametric estimate of the mean value based on Turnbull estimator
[see Turnbull, 1976]: 91.30338

Estimated Mean WTP based on parametric Weibull model (taken self-selected
intervals as Turnbull intervals [see Turnbull, 1976]): 90.44379

>
```

Again, we show the estimated mixed survival function by calling function `plot` as below:

```
> plot(wtp.wemix,xlab="Willingness to pay (SEK)",col=c("blue","red"))
```

The estimated survival function (the mixed Weibull) is shown in Figure 3.

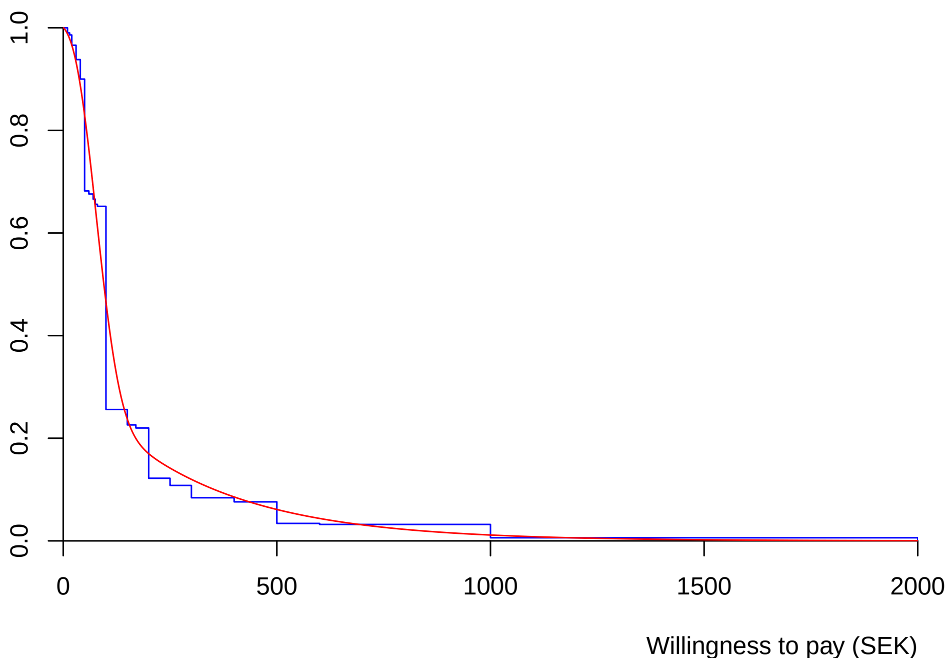

Figure 3: The estimated mixed Weibull/exponential survival function (red), and the empirical survival function (blue). Data: IL

The following code shows how to plot two estimated survival functions in one single figure, shown in Figure 4

```
> plot(wtp.wb,wtp.wemix,xlab="Willingness to pay (SEK)",
col=c("blue","green","blue","red"),lty=c(1,1,1,2))
```

3.3 Estimation Accuracy Assessment

In this section, we show how to use function `resam` for assessing the estimation accuracy. The function `resam` takes object of class 'iwtp' as input. In the follwing code, we take object wtp.wb as input to function `resam`, and specify at the resampling $size = 200$.

```
> res.wb <- resam(wtp.wb, size=10, plot=F)
> print(res.wb)
```

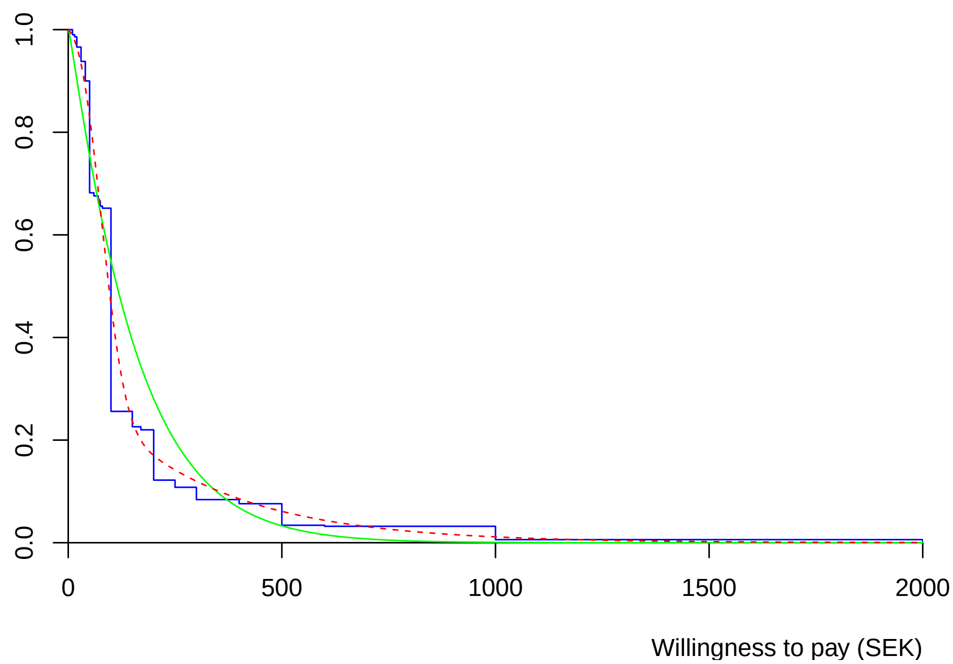

Figure 4: The estimated Weibull (green) and the mixed Weibull (red) survival functions. Data: IL

```
Survival function: Weibull

Resampling size: 10
Mean of mean WTP: 97.79
Quantile of mean WTP:
    5%    50%    95%
 85.10  97.65 110.26

Run time: 21.92919 secs

>
```

The function `resam` produces an object `res.wb`. With object `res.wb`, we can illustrate the estimation accuracy by using scatter plots as shown in 5:

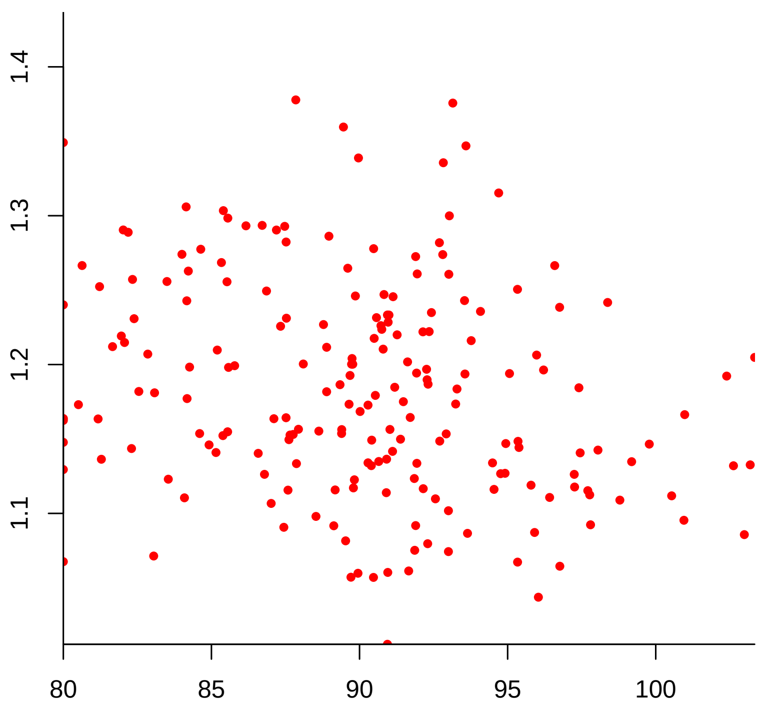

Figure 5: Scatter plot showing ML-estimates of scale α (x-axis) versus β (y-axis) parameters of the Weibull distribution based on the resmapling methods. The number of re-samplings is 200. Data: IL

In addition, we can assess deviations from the unknown true mean WTP values by using function `QQplot`, which will present a quantile-quantile(Q-Q) plot based on the results of re-samplings.

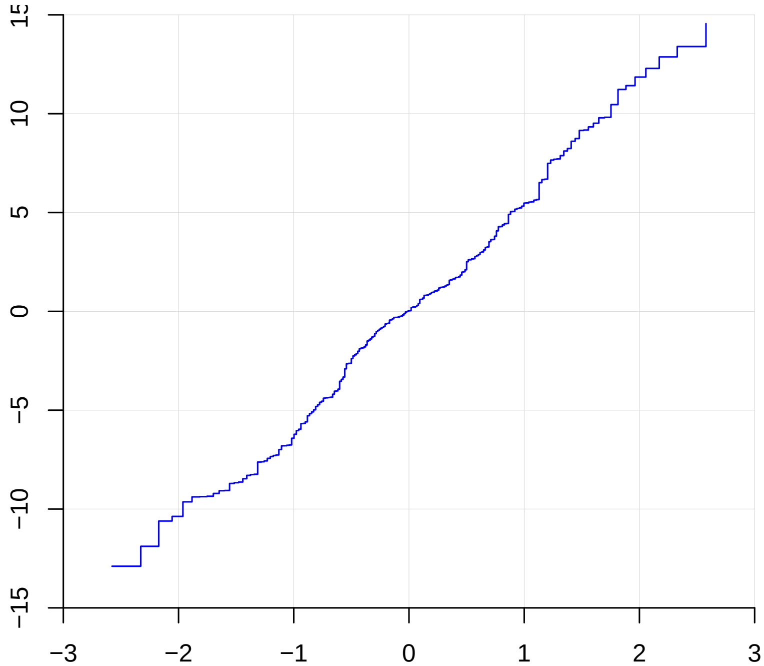

Figure 6: QQ-plot based on re-sampling methods. Data: IL

```
> res.wemix <- resam(wtp.wemix,size=10,plot=F)
```

```
> QQplot(res.wb, col="blue")
> QQplot(res.wb,res.wemix,col=c("blue","green"))
```

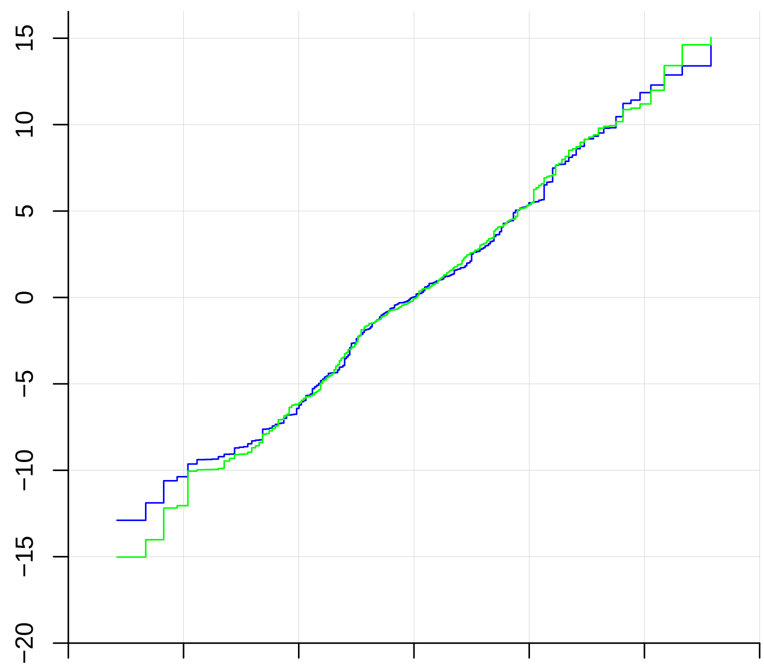

Figure 7: QQ-plots based on re-sampling methods (Blue: the Weibull; green: the mixture). Data: IL

4 Reference to iwtp's functions

4.1 `IL`: A resampled Lax valuation dataset with 500 observations

Description

This is a resampled version of the lax valuation dataset used for demonstrative purposes only.

Usage

```
data(IL)
```

Format

A data frame with 500 observations on the following variable

`respondent no.` nr. of the self-selected intervals

`lower value` the left-end value of self-selected intervals

`upper value` the right-end value of self-selected intervals

Details

This is a re-sampled version of the original CeHa (2009) data set.

Source

see references

References

Belyaev, Yuri and Bengt Kriström. 2010. Approach to Analysis of Self-Selected Interval Data. Technical report SLU, Department of Forest Economics 90183 Umeå, Sweden.

4.2 iwtp: Estimation of Survival Function (Weibull or mixed Weibull)

Description

Estimating the Weibulll or the mixed Weibull and exponential distribution function

Usage

```
iwtp(data,dist="weibull",bm.type=5,plot=FALSE,bounds=list(),limits=list())
## S3 method for class 'iwtp'
plot(x,x2=NULL,xlim=NULL,xlab="",col="red",lty=1,lwd=1,asp=0.6,...)
## S3 method for class 'iwtp'
print(x,...)
```

Arguments

`data`	a data.frame
`bm.type`	numeric number, indicating the type of behavior of respondents selecting intervals. It can be 1, 2, 3, 4 and 5
`dist`	character, type of survival function. It can be "weibull" or "wemix", See Details
`plot`	logical, whether to plot the empirical survival function
`bounds`	a list, specify the lower and upper bounds and the initial values for the parameters, which will be used by the optimizer, See `optim`
`limits`	a list, specify the lower and upper bounds and the initial values for the function of modelling behavior, which will be used by the optimizer, See `optim`
`x,x2`	the output from function `iwtp`
`xlim`	numeric vector with length 2, giving the x-axis range
`xlab`	character, title of x-axis
`col`	line color, it can be a vector when more than one line are drawn, See `par`
`lty`	line type, it can be a vector when more than one line are drawn, See `par`
`lwd`	line width, it can be a vector when more than one line are drawn, See `par`
`asp`	numerical, giving the aspect ratio y/x, See `par`.

...	further graphical parameters as in `par`.

Details

Argument `data` must be a data.frame. `data` should contain at least 4 columns. Its first column is the case number, column 2 is the left-end value of stated intervals, column 2 is the right-end value of stated intervals, and column 4 is the times of the interval being stated.

Argument `dist` can be ”weibull” or ”wemix”. If ”weibull” is chosen, a Weibull survival function will be estimated; If ”wemix” is chosen, a mixed Weilbull/expoential survival function will be estimated.

Value

Function `iwtp` returns an object of class 'iwtp'. An iwtp object is a LIST with the following three components:

`sf`	parameter estimates for the assigned survival function
`esf`	a data list of the empirical survival function based on the right-end value of stated intervals
`intv`	an object of class 'iwtp.interv', containing information about division intervals

Author(s)

Wenchao Zhou

References

Belyaev, Yuri and Bengt Kriström. 2010. Approach to Analysis of Self-Selected Interval Data. Technical report SLU, Department of Forest Economics 90183 Umeå, Sweden.

See Also

`optim`

Examples

```
data(IL)

# survival function: the Weibull distribution
wtp.wb <- iwtp(IL,dist="weibull",plot=TRUE)
wtp.wb

#surival function: the mixed Weibull and exponential distribution
wtp.wemix <- iwtp(IL, dist="wemix", plot=TRUE)
wtp.wemix


# setting initial values for parameters
bounds <- list(
a=c(lower=70,upper=300,init=75),
b=c(lower=1,upper=4.5,init=1.4))
wtp.wb <- iwtp(IL, "weibull",plot=FALSE,bounds=bounds)

# plot empirical survival function
esfPlot(wtp.wb,col="red",lwd=2)

# plot empirical s.f and estimated s.fs in a single figure
plot(wtp.wb,wtp.wemix,col=c("red","green"),lty=c(1,2))
```

4.3 `iwtp.esf`: Empirical survival function

Description

This function builds the empirical survival function for a given data list

Usage

```
iwtp.esf(data)
## S3 method for class 'iwtp'
esfPlot(x,x2=NULL,xlim=NULL,xlab="",col="red",lty=1,lwd=1,asp=0.6,...)
```

Arguments

`data`	data list, it must be a matrix with at least two columns. Coumn 1 is the point estimates, and comlumn 2 is the times of that point observation being stated.
`x,x2`	output from function `iwtp`
`xlim`	numerical vectors of length 2, giving the x ranges
`xlab`	character, title for x-axis
`col`	line color, it can be a vector when more than one line are drawn, See `par`
`lty`	line type, it can be a vector when more than one line are drawn, See `par`
`lwd`	line width, it can be a vector when more than one line are drawn, See `par`
`asp`	numerical, giving the aspect ratio y/x, See `par`
`...`	further graphical parameters as in `par`.

Details

Argument data must be a matrix with at least two columns. The nonparametric mean will be calculated.

Value

Function `iwtp.esf` returns a object of class 'iwtp.esf', which is a LIST.

`esf`	data list of the empirical survival function
`mean`	the mean value of the empirical survival function

Author(s)

Wenchao Zhou

References

Belyaev, Yuri and Bengt Kriström. 2010. Approach to Analysis of Self-Selected Interval Data. Technical report SLU, Department of Forest Economics 90183 Umeå, Sweden.

See Also

`iwtp`

Examples

```
data(IL)
esf <- iwtp.esf(IL)
# show mean wtp
esf$mean

# plot the empirical survival function
esfPlot(esf, col="blue",lty=1,lwd=2)
```

4.4 `resam`: Resampling methods for estimation accuracy of statistical inferecnes

Description

Function `resam` carries out re-samplings for a given number times. For each copy of the re-sampled data, the survival function is estimated. Deviations of ML-esimators from mean WTP are calculated. The obtained deviations can be used to draw normal Quantile-Quantile (Q-Q) plot.

Usage

```
resam(obj,size=10,seed=1234,plot=FALSE,
bounds=list(),limits=list())
## S3 method for class 'iwtp.resam'
QQplot(x,x2=NULL,xlim=NULL,ylim=NULL,
xlab="",ylab="",col="red",lty=1,lwd=1,asp=0.8,...)
## S3 method for class 'iwtp.resam.wb'
plot(x,xlab="",ylab="",...)
## S3 method for class 'iwtp.resam.wb'
print(x,...)
## S3 method for class 'iwtp.resam.wemix'
print(x,...)
```

Arguments

`obj`	object of class 'iwtp'
`size`	integer, sampling size
`seed`	seed for random number
`plot`	logical, whether to present a scatter plot for the estimatetd parameters: a, b for the case of Weibull distribution
`bounds`	a list, specify the lower and upper bounds and the initial values for the parameters, See function `iwtp`
`limits`	a list, specify the lower and upper bounds and the initial values for the parameters of the function for modeling behaviour, See function `iwtp`
`x,x2`	output from function `resam`
`xlim,ylim`	numerical vectors of length 2, giving the x and y coordinates ranges
`xlab`	character, title of x-axis
`ylab`	character, title of y-axis

`col`	line color, it can be a vector when more than one line are drawn, See function `iwtp`
`lty`	line type, it can be a vector when more than one line are drawn, See function `iwtp`
`lwd`	line width, it can be a vector when more than one line are drawn, See function `iwtp`
`asp`	numerical, giving the aspect ratio y/x, See function `asp`
`...`	further graphical parameters as in `par`

Details

Argument `obj` must be a object of class 'iwtp', it is an output of function `iwpt`. Argument `size` gives the number of resamplings. Freq.redo tells the function whether to re-estimate the probability associated with each of the division intervals.

Value

Function `resam` returns a object of class 'iwtp.resam.wb' or 'iwtp.resam.wemix', depending on the obj. The object is a LIST, which consists of:

`par`	a matrix with the estimated parameters for each copy of resamplings.
`sf`	survival function, a copy from argument obj
`size`	sample size, a copy of argument size
`dev.mwtp`	a numerical vector, the deviations from estimated mean WTP
`run.time`	the run time taken by the resampling

Author(s)

Wenchao Zhou

References

Belyaev, Yuri and Bengt Kriström. 2010. Approach to Analysis of Self-Selected Interval Data. Technical report SLU, Department of Forest Economics 90183 Umeå, Sweden.

See Also

`iwtp`

Examples

```
data(IL)

## the Weibull survival function
wtp.wb <- iwtp(IL,"weibull")

## Resample without re-estimation of frequencies
res.wb <- resam(wtp.wb, size=10,plot=FALSE)

## scatter plot
plot(res.wb)

## show QQ plot
QQplot(res.wb)

## the mixed Weibull/exponential survival function
wtp.wemix <- iwtp(IL, dist="wemix")
##
res.wemix <- resam(wtp.wemix, size=10, plot=FALSE)
QQplot(res.wemix)
QQplot(res.wb,res.wemix,col=c("red","blue"))
```

References

Belyaev, Yuri and Bengt Kriström. 2010. Approach to Analysis of Self-Selected Interval Data. Technical report SLU, Department of Forest Economics 90183 Umeå, Sweden: .

Belyaev, Yuri, Cecilia Håkansson and Bengt Kriström. 2009. Rouding it up: Interval and Open Ended Valuation Questions. Technical report SLU, Department of Forest Economics 90183 Umeå, Sweden: .

Turnbull, B.W. 1976. "The empirical distribution function with arbitrarily grouped, censored and truncated data." *J. Roy. Statist. Soc. Ser. B* 38:290–295.