

# Package ‘IDPSurvival’

February 19, 2015

**Version** 1.0

**Date** 2014-07-14

**Title** Imprecise Dirichlet Process for Survival Analysis

**Author** Francesca Mangili <francesca@idsia.ch>,  
Alessio Benavoli <alessio@idsia.ch>,  
Cassio P. de Campos <cassiopc@acm.org>,  
Marco Zaffalon <zaffalon@idsia.ch>

**Maintainer** Francesca Mangili <francesca@idsia.ch>

**Depends** R (>= 3.0.2), Rsolnp, gtools, survival

**Description** This package contains functions to perform robust nonparametric survival analysis with right censored data using a prior near-ignorant Dirichlet Process.

**License** GPL (>= 3) | file LICENSE

**URL** <http://ipg.idsia.ch/software/>

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2014-07-14 14:50:08

## R topics documented:

Aids2 . . . . .	2
aml . . . . .	3
isurvdiff . . . . .	4
isurvdiff.smax . . . . .	6
isurvfit . . . . .	7
isurvfit.object . . . . .	9
plot.isurvfit . . . . .	10

<b>Index</b>	<b>11</b>
--------------	-----------

---

Aids2

*Australian AIDS Survival Data*

---

**Description**

Data on patients diagnosed with AIDS in Australia before 1 July 1991.

**Usage**

Aids2

**Format**

This data frame contains 2843 rows and the following columns:

state Grouped state of origin: "NSW" includes ACT and "other" is WA, SA, NT and TAS.

sex Sex of patient.

diag (Julian) date of diagnosis.

death (Julian) date of death or end of observation.

status "A" (alive) or "D" (dead) at end of observation.

T. categ Reported transmission category.

age Age (years) at diagnosis.

**Note**

This data set has been slightly jittered as a condition of its release, to ensure patient confidentiality.

**Source**

Dr P. J. Solomon and the Australian National Centre in HIV Epidemiology and Clinical Research.

**References**

Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth edition. Springer.

---

aml

*Remission Times for Acute Myelogenous Leukaemia*

---

## Description

The `aml` data frame has 23 rows and 3 columns.

A clinical trial to evaluate the efficacy of maintenance chemotherapy for acute myelogenous leukaemia was conducted by Embury et al. (1977) at Stanford University. After reaching a stage of remission through treatment by chemotherapy, patients were randomized into two groups. The first group received maintenance chemotherapy and the second group did not. The aim of the study was to see if maintenance chemotherapy increased the length of the remission. The data here formed a preliminary analysis which was conducted in October 1974.

## Usage

`aml`

## Format

This data frame contains the following columns:

`time` The length of the complete remission (in weeks).

`cens` An indicator of right censoring. 1 indicates that the patient had a relapse and so `time` is the length of the remission. 0 indicates that the patient had left the study or was still in remission in October 1974, that is the length of remission is right-censored.

`group` The group into which the patient was randomized. Group 1 received maintenance chemotherapy, group 2 did not.

## Note

Package **boot** also has the dataset `aml`.

## Source

The data were obtained from

Miller, R.G. (1981) *Survival Analysis*. John Wiley.

## References

Davison, A.C. and Hinkley, D.V. (1997) *Bootstrap Methods and Their Application*. Cambridge University Press.

Embury, S.H., Elias, L., Heller, P.H., Hood, C.E., Greenberg, P.L. and Schrier, S.L. (1977) Remission maintenance therapy in acute myelogenous leukaemia. *Western Journal of Medicine*, **126**, 267-272.

---

isurvdiff

*Test Survival Curves Differences for two right censored data*


---

### Description

Tests if there is a difference between two survival curves based on two samples (X and Y) with right censored data. More precisely it test whether the probabiliy  $P(X<Y)$  is greater than, lower than or equal to  $1/2$ . The prior near-ignorance Dirichlet Process (IDP) rank sum test is used. It returns the result of the deicison.  $H=1$  indicates that the alternative hypothesis is true with posterior probability greater than `level`.  $H=0$  indicates the hypothesis is not true with posterior greater than `level`,  $H=2$  indicates an indeterminate instance. This means that the decision depends on the choice of the prior.

### Usage

```
isurvdiff(formula, data, groups=c(1,2), s=0.25,
          alternative = c("two.sided", "less", "greater"),
          exact=NULL, level = 0.95, display=TRUE,
          nsamples=10000, rope=0, tmax=NULL)
```

### Arguments

formula	a formula expression of the form <code>Surv(time, status) ~ predictor</code> . A single predictor is admitted.
data	an optional data frame in which to interpret the variables occurring in the formula.
groups	a vector of two element indicating which value of the predictor represents groups 1 and 2.
s	sets the value of the prior strength <code>s</code> of the Dirichlet Process.
alternative	define the direction of the test: "greater" –evaluates the hypothesis $P(X < Y) > 1/2$ , i.e., returns $H=1$ if the lower probability of the hypothesis is larger than <code>level</code> , $H=0$ if the upper probability is smaller than <code>level</code> and $H=2$ if the lower and upper probabilities encompass <code>level</code> ; "less" – evaluates the hypothesis $P(Y < X) > 1/2$ ; "two.sided" – performs a two-sided Bayesian test, i.e., returns $H=1$ if $1/2$ is not included between the left bound of the lower and the right bound of the upper <code>level</code> HPD credible intervals, $H=0$ if $1/2$ is included in both the upper and lower credible intervals, $H=2$ otherwise.
exact	computes the posterior probability if value is <code>TRUE</code> , or uses a normal approximation if value is <code>FALSE</code> . If you omit this argument, <code>isurvdiff</code> uses the exact method if at least one group has less than 100 samples and the approximate one otherwise.
level	sets the significance level $\alpha = 1 - \text{level}$ of the test.
display	determines whether the posterior distributions of $P(X < Y)$ have to be plotted ( <code>TRUE</code> ) or not ( <code>FALSE</code> ).

nsamples	if exact=TRUE, sets the number of samples used in the Monte Carlo computation of the posterior distributions. For faster but less accurate results, one can tune down this parameter. For more accurate, one might increase it.
rope	introduces a (symmetric) Region of Practical Equivalence (ROPE) around 1/2, i.e., [1/2-value, 1/2+value].
tmax	whether to consider the difference in survival up to time tmax. NULL is the default and means without limit.

## Value

a list with components:

h	The decision of the test: H=0 -> accept the null hypothesis; H=1 -> rejects the null hypothesis; H=2 -> indeterminate (a robust decision cannot be made).
prob	the probability of the alternative hypothesis $P(X < Y) > 1/2$ if alternative="greater" or $P(Y < X) > 1/2$ if alternative="less".
Lower.Cred.Int	lower HPD credible interval. Confidence level defined by level.
Upper.Cred.Int	upper HPD credible interval. Confidence level defined by level.
alternative	the direction of the test "greater", "less" or "two.sided".
strata	the number of subjects contained in each group.
exact	logical variable saying if the exact posterior distributions have been computed (TRUE) or the Gaussian approximation has been used (FALSE).

## METHOD

This function implements the IDP sum-rank test describe in Mangili and others (2014).

## References

- Benavoli, A., Mangili, F., Zaffalon, M. and Ruggeri, F. (2014). Imprecise Dirichlet process with application to the hypothesis test on the probability that  $X < Y$ . *ArXiv e-prints*, <http://adsabs.harvard.edu/abs/2014arXiv1402.2755B>.
- Mangili, F., Benavoli, A., Zaffalon, M. and de Campos, C. (2014). Imprecise Dirichlet Process for the estimate and comparison of survival functions with censored data.

## See Also

[Surv.](#)

## Examples

```
data(lung, package='survival')
test <- isurvdiff(Surv(time, status) ~ sex, lung, groups=c(1, 2),
  alternative = 'two.sided', s=0.5, nsamples=1000)
print(test)

data(Aids2)
fdata <- Surv(time, status) ~ T.categ
```

```

dataset <- Aids2
groups=c("blood","haem")
dataset["time"]<-dataset[4]-dataset[3]
dataset[5]<-as.numeric(unlist(dataset[5]))
test <-isurvdiff(fdata,dataset,groups=groups,
                 alternative = 'greater',s=0.5, nsamples=1000)
print(test)

```

---

isurvdiff.smax	<i>Maximum values of s for which the IDP test returns a determinate decision</i>
----------------	--

---

### Description

Search for the maximum values of parameter  $s$  for which the IDP test `isurvdiff(formula,...)` issues a determinate decision. The function test values of  $s$  up to the parameter `smax`. If for `smax` the IDP test is still determinate, `isurvdiff.smax` returns `list(smax,testout)`. If for  $s=0$  the test is already indeterminate, `isurvdiff.smax` returns `list(-1,testout)`, where `testout` is the last executed test.

### Usage

```
isurvdiff.smax(formula, ..., verbose=FALSE, accuracy=0.05, smax=12)
```

### Arguments

<code>formula</code>	a formula expression of the form <code>Surv(time, status) ~ predictor</code> . A single predictor is admitted.
<code>verbose</code>	whether to display each value of $s$ that is tried
<code>accuracy</code>	to which precision $s$ should be computed
<code>smax</code>	to which maximum value $s$ should be tried
<code>...</code>	All arguments of <code>isurvdiff.smax</code> are passed to <code>isurvdiff</code> to perform the test. Refer to the help of <code>isurvdiff</code> for more details about the arguments.

### Value

A list with components:

<code>s</code>	The maximum value of $s$ for which the test returns a determinate decision ( $H=0$ or $H=1$ ).
<code>test0</code>	The value returned by <code>isurvdiff(formula,...)</code> for the last test performed. Refer to the help of <code>isurvdiff</code> for more details.

### METHOD

This function implements the IDP sum-rank test describe in Mangili and others (2014).

## References

Benavoli, A., Mangili, F., Zaffalon, M. and Ruggeri, F. (2014). Imprecise Dirichlet process with application to the hypothesis test on the probability that  $X < Y$ . *ArXiv e-prints*, <http://adsabs.harvard.edu/abs/2014arXiv1402.2755B>.

Mangili, F., Benavoli, A., Zaffalon, M. and de Campos, C. (2014). Imprecise Dirichlet Process for the estimate and comparison of survival functions with censored data.

## See Also

[Surv](#), [isurvdiff](#).

## Examples

```
data(lung,package='survival')
lung <- lung[1:40,] # reduced data set just to ensure that the
# example is very fast to run for building the package
test <- isurvdiff.smax(Surv(time,status)~sex, lung, groups=c(1,2),
  alternative = 'two.sided', nsamples=1000)
# better to use larger value of nsamples
# this small value is to run it quickly
print(test$test0)
cat("Maximum s giving the same decision: ", test$s)
```

---

isurvfit

*Create survival curves based on the IDP model*

---

## Description

This function creates survival curves from right censored data using the prior near-ignorance Dirichlet Process (IDP).

## Usage

```
isurvfit(formula, data, s=0.5, weights, subset, display=TRUE,
  conf.type=c('exact', 'approx', 'none'), nsamples=10000,
  conf.int= .95)
```

## Arguments

formula	a formula object, which must have a Surv object as the response on the left of the ~ operator and, if desired, terms separated by + operators on the right. For a single survival curve the right hand side should be ~ 1.
data	a data frame in which to interpret the variables named in the formula, subset and weights arguments.
s	sets the value of the prior strength $s$ of the Dirichlet Process.

weights	the weights must be finite and nonnegative; it is strongly recommended that they be strictly positive, since zero weights are ambiguous, compared to use of the subset argument.
subset	expression saying that only a subset of the rows of the data should be used in the fit.
display	determines whether the survival curves have to be plotted (TRUE) or not (FALSE).
conf.type	a variable saying how the credible interval should be computed: 'exact': Monte-Carlo sampling from the exact distribution, 'approx': Gaussian approximation, 'none': no credible interval is computed.
nsamples	number of samples used to approximate the credible intervals if conf.type='exact'.
conf.int	confidence level of the credible interval.

### Details

The estimates are obtained using the IDP estimator by Mangili and others (2014) based on the prior near-ignorance Dirichlet Process model by Benavoli and others (2014).

### Value

an object of class "isurvfit".

See `isurvfit.object` for details. Methods defined for `survfit` objects are `print` and `plot`.

### References

Benavoli, A., Mangili, F., Zaffalon, M. and Ruggeri, F. (2014). Imprecise Dirichlet process with application to the hypothesis test on the probability that  $X < Y$ . *ArXiv e-prints*, <http://adsabs.harvard.edu/abs/2014arXiv1402.2755B>.

Mangili, F., Benavoli, A., Zaffalon, M. and de Campos, C. (2014). Imprecise Dirichlet Process for the estimate and comparison of survival functions with censored data.

### See Also

[isurvfit.object](#), [plot.isurvfit](#), [Surv](#).

### Examples

```
data(aml)
fit <- isurvfit(Surv(time, cens) ~ 1, data=aml, display=TRUE, nsamples=1000)
legend('topright', c("Lower expectation",
                    "Upper expectation", "confidence intervals"), lty=c(1,1,2), lwd=c(1,2,1))
title("IDP survival curve (s=0.5) \nAcute Myelogenous Leukemia dataset")

data(Aids2)
dataset <- Aids2
dataset["time"] <- dataset[4] - dataset[3]
dataset[5] <- as.numeric(unlist(dataset[5]))
fit <- isurvfit(Surv(time, status) ~ T.categ, dataset, s=1,
              subset=(!is.na(match(T.categ, c('blood', 'haem', 'het')))))
```



```

      nsamples=1000,conf.type='none')
legend('topright',c("Heterosexual contact","Hemophilia","Blood"),
      title="Transmission category:",lty=c(1,1,1),col=c(1,2,3),pch=c(1,2,3))
title("IDP survival curve (s=1) \nAids dataset")
print(fit)

leukemia.surv <- isurvfit(Surv(time, cens) ~ group, data = aml, display=FALSE)
plot(leukemia.surv)
legend(100, .9, c("Maintenance", "No Maintenance"), lty=c(1,1),lwd=c(2,1),
      col=c('black','red'),pch=c(1,2))
title("IDP Curves\nfor AML Maintenance Study")

```

---

isurvfit.object	<i>IDP Survival Curve Object</i>
-----------------	----------------------------------

---

## Description

This class of objects is returned by the `isurvfit` class of functions to represent a IDP survival curve.

Objects of this class have methods for the functions `print`, `plot`.

## Arguments

<code>n</code>	total number of subjects in each curve.
<code>time</code>	the time points at which the curve has a step.
<code>n.risk</code>	the number of subjects at risk at $t$ .
<code>n.event</code>	the number of events that occur at time $t$ .
<code>n.censor</code>	the number of subjects who exit the risk set, without an event, at time $t$ . (This number can be computed from the successive values of the number at risk).
<code>survUP</code>	the estimate of upper expectation of the survival probability at time $t+0$ .
<code>survLOW</code>	the estimate of lower expectation of the survival probability at time $t+0$ .
<code>survLOW0</code>	the estimate of lower expectation of the survival probability at time $t=0$ . The upper expectation is always 1.
<code>std.up</code>	the standard deviation of the upper distribution of the survival probability.
<code>std.low</code>	the standard deviation of the lower distribution of the survival probability.
<code>upper</code>	upper confidence limit for the survival curve.
<code>lower</code>	lower confidence limit for the survival curve.
<code>lower0</code>	lower confidence limit for the survival curve at $t=0$ . The upper is always 1.
<code>conf.type</code>	the approximation used to compute the confidence limits.
<code>conf.int</code>	the level of the confidence limits, e.g. 90 or 95%.
<code>strata</code>	number of elements of the <code>time</code> vector corresponding to each curve. The names of the elements are labels for the curves.
<code>call</code>	an image of the call that produced the object.

**See Also**[isurvfit](#) [plot.isurvfit](#)

---

plot.isurvfit	<i>Plot method for isurvfit objects</i>
---------------	---

---

**Description**

A plot of survival curves is produced, one curve for each strata.

**Usage**

```
## S3 method for class 'isurvfit'  
plot(x, se.fit=TRUE, ...)
```

**Arguments**

x	an object of class <code>isurvfit</code> , usually returned by the <code>isurvfit</code> function.
se.fit	determines whether confidence intervals will be plotted.
...	other arguments passed to the standard plot function

**See Also**[isurvfit](#)**Examples**

```
leukemia.surv <- isurvfit(Surv(time, status) ~ x, data = aml, display=FALSE)  
plot(leukemia.surv)  
legend(100, .9, c("Maintenance", "No Maintenance"),  
      lty=c(1,1),lwd=c(2,1),col=c('black','red'),pch=c(1,2))  
title("IDP Curves\nfor AML Maintenance Study")
```

# Index

## \*Topic **IDP**

- isurvdiff, 4
- isurvdiff.smax, 6
- isurvfit, 7
- isurvfit.object, 9

## \*Topic **datasets**

- Aids2, 2
- aml, 3

## \*Topic **hplot**

- plot.isurvfit, 10

## \*Topic **survival**

- isurvdiff, 4
- isurvdiff.smax, 6
- isurvfit, 7
- isurvfit.object, 9
- plot.isurvfit, 10

Aids2, 2

aml, 3

isurvdiff, 4, 7

isurvdiff.smax, 6

isurvfit, 7, 10

isurvfit.object, 8, 9

plot.isurvfit, 8, 10, 10

print.isurvdiff (isurvdiff), 4

print.isurvfit (isurvfit), 7

Surv, 5, 7, 8