Package 'splinesurv'

June 27, 2008

Version 0.6-9

Title Nonparametric bayesian survival analysis

Author Emmanuel Sharef <ess28@cornell.edu>

Maintainer Emmanuel Sharef <ess28@cornell.edu>

Depends R (>= 2.5.0), MASS, survival

Suggests coda, tkrplot

Description Utilities for nonparametric Bayesian analysis of clustered survival data. The baseline hazard function and frailty density are modeled using penalized B-splines. Options include adaptive knot selection and the inclusion of a parametric component.

License GPL (>= 2)

R topics documented:

---

plot.splinesurv   *Plots for splinesurv objects*

Description

Function to plot the the baseline hazard, frailty density, and posterior density of regression parameters for objects of type splinesurv, with optional pointwise credible bands.

1

Usage

```
## S3 method for class 'splinesurv':
plot(x, which = c("hazard", "survival", "frailty", "coef", "all"),
    newdata = NULL, iter=NULL, fn = mean, plotknots = TRUE, npoints = 100,
    npost = 100, alpha = .05, legend = NULL, lty = 1, col = 2, lwd = 2,
    lty.knots = 1, col.knots = 8, lwd.knots = 1, xlab = NULL, ylab = NULL,
    main = NULL, xlim = NULL, ylim = NULL, tk = FALSE, ...)
```

Arguments

x	an object of type splinesurv.
which	a string specifying which plot to produce. Setting "hazard" plots the baseline hazard (default), "survival" plots the survivor function, "frailty" plots the frailty density, "coef" plots the posterior density of the coefficients. Setting "all" plots all plots at once.
newdata	a data frame in the same format as the original data containing data for prediction of the hazard or survival function. If it has more than one row, multiple functions will be plotted, but credible bands will not.
iter	integer, indicating which iteration of the MCMC chain to plot. If NULL (default), the posterior means are used
plotknots	boolean indicating whether frailty spline knots should be plotted as vertical lines
fn	function to be applied to the posterior samples. Defaults to mean, but can be set to median or any other summary function.
npoints	number of points to be used for drawing
npost	number of samples from the posterior to be used for estimating the curve
alpha	level for plotting pointwise confidence bands. Setting alpha=NULL omits bands.
legend	legend text, if desired
lty	line type for primary curve
col	color for primary curve
lwd	line width for primary curve
lty.knots	line type for knots
col.knots	color of the knot lines
lwd.knots	line width of the knot lines
xlab	x-axis label
ylab	y-axis label
main	main plot title
xlim	x-axis limits
ylim	y-axis limits
tk	boolean indicator whether to use the tkrplot GUI
...	additional parameters passed on to plot

---

predict.splinesurv *Prediction for splinesurv objects*

Description

Function to predict the the baseline hazard, frailty density, linear predictor, or risk for objects of type splinesurv.

Usage

```
## S3 method for class 'splinesurv':
predict(object, type = c("hazard","survival","lp","risk","frailty"),
        x = NULL, newdata = NULL, iter = NULL, fn = mean, alpha = NULL, npost =
```

Arguments

object	an object of type splinesurv.
type	a string specifying the type of prediction. Setting "hazard" predicts the baseline hazard (default), survival predicts the survivor function, "lp" predicts the linear predictor, "risk" produces risk estimates, and "frailty" predicts the frailty density.
x	a vector containing the values at which prediction is desired. If NULL, a sequence of length 100 over the data range is used.
newdata	a data frame in the same format as the original data containing data for prediction. If cluster membership is not set to NA, the frailty is included in predictions of the hazard or survival.
iter	integer, indicating which iteration of the MCMC chain to use. If NULL (default), the posterior means are used
npost	number of samples from the posterior to be used for estimating the curve
fn	function to be applied to the posterior samples. Defaults to mean, but can be set to median or any other summary function.
alpha	level of pointwise confidence bands. Setting alpha=NULL omits bands.
...	additional parameters currently ignored

---

sim.sample *Simulated survival data generator*

Description

Generates a simulated sample of clustered survival data, allowing for flexible baseline hazard and frailty distributions.

Usage

```
sim.sample(m = 10, Ji = rep(5,10), params=NULL)
```

Arguments

`m`	an integer, number of clusters in the sample.
`Ji`	a vector of length `m` containing the number of subjects in each cluster.
`params`	a list, optionally containing the parameters that specify the covariate distribution, baseline hazard, frailty distribution, and censoring time, as follows:

beta the "true" value of the single covariate.

haz.type a string specifying the type of hazard to be used.

haz.params a list with parameters corresponding to the hazard type.

frail.type a string specifying the type of frailty to be used.

frail.params a list with parameters corresponding to the frailty type.

Z.type a string specifying the type of covariate to be used.

Z.params a list with parameters corresponding to the covariate type.

C.type a string specifying the type of censoring to be used.

C.params a list with parameters corresponding to the censoring type.

The default corresponds to standard normal covariates, a Weibull hazard, gamma frailties and a Weibull censoring time. See details.

Details

The possible values for `frail.type`, `Z.type`, `C.type` and the corresponding entries in `frail.params`, `Z.params` and `C.params` are as follows:

*.type="fixed" `*.params` is a list with component `value`, containing the fixed value.

*.type="weibull" `*.params` is a list with components `\lambda0` and `gamweib`, for components λ and γ of the Weibull distribution.

*.type="gamma", *.type="normal", *.type="lognormal" `*.params` is a list with components `mu` and `sigma2`, giving the mean and variance of the gamma, normal or lognormal distribution desired.

*.type="normmix", *.type="lognormmix" `*.params` is a list of with components `mu`, `sigma2` and `w`, all vectors of the same length, giving the mean, variance and weight of each component of the normal or lognormal mixture.

For `haz.type`, the possible values and parameters are:

haz.type="weibull" `haz.params` is a list with components `\lambda0` and `gamweib`, for components λ and γ of the Weibull distribution.

haz.type="stepfunction" `haz.params` is a list with vectors `breaks` giving the positions of jumps in the stepfunction, and `haz` giving the initial value of the hazard and its value after each jump.

haz.type="bspline" `haz.params` is a list with components `b`, a basis matrix generated by `bs` specifying the B-spline basis, and `w`, a vector of weights for each basis element.

Value

A list with three components:

`agdata`	a data.frame in with columns

    i cluster id

    j patient id within the cluster

time time of event

delta censoring indicator

Z value of the covariate

Ui   a vector of frailties used in the simulation

params   the set of parameters in the format of the input params. See details.

See Also

[splinesurv](#)

Examples

```
# Generate a simulated sample with default parameters
# (Weibull baseline and gamma frailty)
s <- sim.sample(m = 10, Ji = rep(10,10))
plot(survfit(coxph(Surv(time,delta) ~ Z + frailty(i), data = s$agdata)))

# Generate a sample with piecewise baseline hazard and
# lognormal mixture frailties
s2 <- sim.sample(m = 10, Ji = rep(10,10), params = list(
    haz.type = "stepfunction", haz.params = list(breaks = c(.25, .5),
    haz = c(1, 0 , 2)), frail.type = "lognormmix",
    frail.params = list(mu = c(.5, 2), sigma2 = c(.2, .2),
    w = c(.666, .333))))
plot(survfit(coxph(Surv(time,delta) ~ Z + frailty(i), data = s2$agdata)))
```

---

splinesurv–package *Nonparametric bayesian survival analysis*

Description

Utilities for nonparametric Bayesian analysis of clustered survival data. The baseline hazard function and frailty density are modeled using penalized B-splines, and may optionally have a parametric component.

Details

The core function is [splinesurv](#), and methods for printing and plotting are provided as well. See the [splinesurv](#) documentation for details on model specification and fitting procedure.

Package:	splinesurv
Version:	0.3-4
Depends:	R (>= 2.5.0), MASS, survival
Suggests:	coda
License:	GPL (>= 2)
Packaged:	Wed Dec 26 18:47:45 2007; esharef
Built:	R 2.6.0; ; 2007-12-26 18:47:51; unix

Index:

[plot.splinesurv](#)   Plots for splinesurv objects

[plot.splinesurv](#)	Bayesian survival analysis with B-spline baseline hazard and random effects density
[plot.splinesurv](#)	Summary method for splinesurv objects

Author(s)

Emmanuel Sharef <ess28@cornell.edu>

---

splinesurv	*Bayesian survival analysis with B-spline baseline hazard and random effects density*

Description

Allows the fitting of proportional hazards survival models to possibly clustered data using Bayesian methods. The function follows a MCMC method to sample from the posterior distribution of the regression parameters, frailties, and parameters specifying the hazard and frailty distribution.

The baseline hazard and random effects density are modeled as convex combinations of a parametric component (for example, a gamma frailty, or a weibull baseline hazard) and a nonparametric component modeled as a penalized B-splines, with the penalty depending on either the integrated squared second derivative of the curve, or the sum of squared second differences in the spline parameters.

Usage

```
## S3 method for class 'formula':
splinesurv(formula, data = parent.frame(), verbose = 3,
           hazard = NULL, frailty = NULL, control = NULL, coda = FALSE,
           initial = NULL, usec = TRUE, ...)
```

Arguments

formula	a formula object, similar to [coxph](#). The response to the left of the ~ should be a survival object generated by [Surv](#). The right side may contain a cluster(x) term if x is the variable that indicates cluster membership.
data	a data.frame with columns corresponding to the terms in the formula.
verbose	an integer from 0 to 5 that determines the quantity of output printed to the screen. Setting verbose=0 is completely silent.
hazard	a list containing parameters defining the baseline hazard, with the following optional components. For any component that is not set, the default is used.

    type a string that determines whether the hazard is parametric, nonparametric, or has both components. Takes on values "spline", "parametric" or "both", with "spline" as default.

    spline.adaptive logical, determines whether adaptive knot selection should be used. Defaults to TRUE.

    spline.ord an integer greater than 1, giving the order of the spline defining the baseline hazard. Defaults to spline.ord=4, corresponding to cubic B-splines.

spline.nknots an integer giving the number of interior knots used in the baseline hazard spline. If `NULL`, the number of knots is chosen automatically, which is the default. If `spline.adaptive=TRUE`, this only serves as the initial number of knots.

spline.nknots.prior a string specifying the type of prior on the number of knots to use. Options are `"poisson"`, `"geometric"`, `"poissonmix"`, `"negbin"` and `"power"`.

spline.nknots.hyper the parameter(s) for the prior on the number of knots. For the Poisson prior, the mean, for the Geometric, the probability, for the Poisson-mixture, two means, and for the Power prior, the exponent.

spline.knotspacing string that determines the way that knots are automatically chosen, if applicable. Possible values are `"quantile"` to place knots at quantiles of observed event times, or `"equal"` to equally space knots over the range of observed times. Defaults to `"equal"`. If `spline.adaptive=TRUE`, this affects selection of candidate knot positions.

spline.knots a vector of length `spline.nknots+2*spline.ord-2` giving the positions of all the knots used in the baseline hazard spline (including boundary knots). If `NULL`, knots are chosen automatically. Defaults to `NULL`.

spline.penalty a string giving the type of penalty to be used for the baseline hazard splines. Possible values are `"none"` for a standard Gaussian prior (default), `"2diff"`, for a penalty on the second differences, or `"2deriv"` for a penalty on the integrated squared second derivative.

spline.hyper a vector of length 2, containing hyperparameters for the spline weights. These control the smoothness of the spline when `spline.adaptive=FALSE`. Defaults to `c(0.01,0.01)`.

param.dist a string giving the desired parametric form of the hazard. Currently supported are `"exponential"` and `"weibull"`.

... other parameters can be set as well, including tuning parameters, prior variances, hyperparameters, initial values for weights, etc, but this should not be attempted by someone unfamiliar with the source code.

frailty   a list containing parameters defining the frailty density, analogous to `hazard`, except for the following differences:

param.dist a string giving the desired parametric form of the frailty. Currently supported are `"gamma"` and `"lognormal"`.

control   a list containing control parameters for the MCMC and optimization, with the following optional components. For any component that is not set, the default is used.

burnin an integer giving the number of iterations discarded as burn-in. Default is `burnin=500`.

maxiter an integer giving the total number of MCMC iterations, must be greater than `burnin`. Default is `maxiter=1000`.

thin an integer giving the number of iterations to be discarded between every recorded iteration. That is, if `iter=10`, a total of `maxiter*thin` iterations will be run, but only `maxiter` will be recorded.

tun.auto logical, determines whether to adaptively calibrate the tuning parameters during the burn-in period to achieve an acceptance rate of 0.25. Defaults to `TRUE`.

        tun.int an integer giving the interval in iterations between calibration of the tuning parameters, if `tun.auto=TRUE`. Default is `100`.

initial   a list containing initial values for the chain. Not implemented and currently ignored.

coda   a logical variable indicating whether the `coda` package should be used to return the parameter history as `mcmc` objects. Defaults to `FALSE`.

usec   a logical variable, determines whether fast C code should be used. Defaults to `TRUE`, and should only be set to `FALSE` for debugging.

...   additional parameters (currently ignored).

Value

An object of class `splinesurv`, with the following components

call   the original call to the model-fitting function

posterior.mean

         a list containing the posterior means of all parameters, with the following components

        coefficients regression coefficients

        frailty frailty estimates

        hazard.spline.par parameters defining the spline component (if applicable)

        hazard.param.par parameters defining the parametric component (if applicable)

        hazard.weight weight of the spline component (if applicable)

        frailty.spline.par parameters defining the spline component (if applicable)

        frailty.param.par parameters defining the parametric component (if applicable)

        frailty.weight weight of the spline component (if applicable)

history   a list containing the parameter history of the MCMC iterations, with the following components, either as matrices or vectors of length `maxiter`:

        frailty matrix of frailty estimates at each iteration.

        coefficients matrix of regression parameter estimates at each iteration.

        hazard.spline.par matrix of parameters defining the baseline hazard spline.

        hazard.spline.knots matrix of knots defining the baseline hazard spline.

        frailty.spline.par matrix of parameters defining the frailty density spline.

        frailty.spline.knots matrix of knots defining the frailty density spline.

        frailty.spline.fvar vector of estimated variances of the frailty density spline.

        priorvar matrix of prior variance parameters for the regression parameters, hazard spline, frailty density spline, hazard parameters, frailty parameters, hazard weight, frailty weight, in that order.

        accept binary vectors of Metropolis-Hastings acceptance indicators for the regression parameters, hazard spline, frailty density spline, hazard parameters, frailty parameters, hazard weight, frailty weight, and frailties respectively.

         If in the input, `coda=TRUE`, then the returned objects are of class `mcmc` instead of `matrix`.

frailty	a list analogous to `frailty` in the input, additionally containing the number and placement of knots, if not specified in the input
hazard	a list analogous to `hazard` in the input, additionally containing the number and placement of knots, if not specified in the input
control	a list analogous to `control` in the input, containing the control parameters used in the procedure, even those not set explicitly in the call.

References

E. Sharef, D. Ruppert and R. Strawderman. "A Bayesian approach to clustered survival analysis with nonparametric baseline hazard and frailty distributions", *in preparation*.

See Also

summary.splinesurv, plot.splinesurv, coxph

Examples

```
## Generate a small survival data set:
s <- sim.sample(m = 10, Ji = rep(10, 10))
agdata <- s$agdata

## Run a (very) short MCMC chain
fit <- splinesurv(Surv(time, delta) ~ Z + cluster(i), data = agdata, control = list(m

## Run another chain,  with a Weibull hazard and linear B-spline frailties
# with fixed knots.
fit2 <- splinesurv(Surv(time, delta) ~ Z + cluster(i), data = agdata, control = list(
## View summaries and plots of the fits
summary(fit)
plot(fit, "all")
summary(fit2)
plot(fit2, "all")
```

---

summary.splinesurv *Summary method for splinesurv objects*

Description

Prints a summary of a fit from splinesurv, including posterior means and quantiles of the posterior distribution for the regression parameters.

Usage

```
## S3 method for class 'splinesurv':
summary(object,quantiles=c(.025,.975),...)
```

Arguments

object	an object of type splinesurv.
quantiles	a vector of quantiles at which the posterior distribution should be printed. By default prints the boundary points of a two-sided 95% credible interval.
...	additional parameters for print.

Value

An item of type `summary.splinesurv` with components `call, coef, iter, burnin,`
`hazard, frailty, posterior.mean` as documented in `splinesurv`, as well as:

quantiles   matrix with the posterior mean and desired quantiles of the regression parame-
                 ters.
quantiles.fvar
                 quantiles of the variance of the frailty density
quantiles.fvar2
                 quantiles of the variance of the estimated frailties

See Also

`splinesurv`

Index