R documentation

of 'flexsurvspline.Rd'

July 25, 2014

---

flexsurvspline *Flexible survival regression using the Royston/Parmar spline model.*

Description

Flexible parametric modelling of time-to-event data using the spline model of Royston and Parmar (2002).

Usage

```
flexsurvspline(formula, data, k=0, knots=NULL, bknots=NULL,
               scale="hazard", ...)
```

Arguments

formula	A formula expression in conventional R linear modelling syntax. The response must be a survival object as returned by the Surv function, and any covariates are given on the right-hand side. For example,
	`Surv(time, dead) ~ age + sex`
	specifies a model where the log cumulative hazard (by default, see scale) is a linear function of the covariates age and sex.
	If there are no covariates, specify 1 on the right hand side, for example `Surv(time, dead) ~ 1`.
	Time-varying covariate effects can be specified using the method described in flexsurvreg for placing covariates on ancillary parameters. The ancillary parameters here are named gamma1, ..., gammar where r is the number of knots k plus one (the "degrees of freedom" as defined by Royston and Parmar). So for the default Weibull model, there is just one ancillary parameter gamma1.
	Therefore a model with one internal spline knot, where the equivalents of the Weibull shape and scale parameters, but not the higher-order term gamma2, vary with age and sex, can be specified as:
	`Surv(time, dead) ~ age + sex + gamma1(age) + gamma1(sex)`

or alternatively (and more safely, see `flexsurvreg`)

`Surv(time, dead) ~ age + sex, anc=list(gamma1=~age + sex)`

Only `Surv` objects of `type="right"` or `type="counting"`, corresponding to right-censored and/or left-truncated observations, are supported.

`data` A data frame in which to find variables supplied in `formula`. If not given, the variables should be in the working environment.

`k` Number of knots in the spline. The default `k=0` gives a Weibull, log-logistic or lognormal model, if `"scale"` is `"hazard"`, `"odds"` or `"normal"` respectively. `k` is equivalent to `df-1` in the notation of `stpm` for Stata. The knots are then chosen as equally-spaced quantiles of the log uncensored survival times, for example, at the median with one knot, or at the 33% and 67% quantiles of log time with two knots. To override this default knot placement, specify `knots` instead.

`knots` Locations of knots on the axis of log time. If not specified, knot locations are chosen as described in `k` above. Either `k` or `knots` must be specified. If both are specified, `knots` overrides `k`.

`bknots` Locations of boundary knots, on the axis of log time. If not supplied, these are are chosen as the minimum and maximum log death time.

`scale` If `"hazard"`, the log cumulative hazard is modelled as a spline function of log time.

If `"odds"`, the log cumulative odds is modelled as a spline function of log time.

If `"normal"`, $-\Phi^{-1}(S(t))$ is modelled as a spline function of log time, where $\Phi^{-1}()$ is the inverse normal distribution function `qnorm`.

`...` Any other arguments to be passed to or through `flexsurvreg`, for example, `anc`, `inits`, `fixedpars`, `weights`, `subset`, `na.action`, and any options to control optimisation. See `flexsurvreg`.

Details

This function works as a wrapper around `flexsurvreg` by dynamically constructing a custom distribution using `dsurvspline`, `psurvspline` and `unroll.function`.

In the spline-based survival model of Royston and Parmar (2002), a transformation $g(S(t,z))$ of the survival function is modelled as a natural cubic spline function of log time $x = \log(t)$ plus linear effects of covariates z.

$$g(S(t,z)) = s(x, \boldsymbol{\gamma}) + \boldsymbol{\beta}^T \mathbf{z}$$

The proportional hazards model (`scale="hazard"`) defines $g(S(t,\mathbf{z})) = \log(-\log(S(t,\mathbf{z}))) = \log(H(t,\mathbf{z}))$, the log cumulative hazard.

The proportional odds model (`scale="odds"`) defines $g(S(t,\mathbf{z})) = \log(S(t,\mathbf{z})^{-1} - 1)$, the log cumulative odds.

The probit model (`scale="normal"`) defines $g(S(t,\mathbf{z})) = -\Phi^{-1}(S(t,\mathbf{z}))$, where $\Phi^{-1}()$ is the inverse normal distribution function `qnorm`.

With no knots, the spline reduces to a linear function, and these models are equivalent to Weibull, log-logistic and lognormal models respectively.

The spline coefficients $\gamma_j : j = 1, 2 \ldots$, which are called the "ancillary parameters" above, may also be modelled as linear functions of covariates $\mathbf{z}$, as

$$\gamma_j(\mathbf{z}) = \gamma_{j0} + \gamma_{j1} z_1 + \gamma_{j2} z_2 + ...$$

giving a model where the effect of covariates are arbitrarily flexible functions of time: a non-proportional hazards or odds model.

Natural cubic splines are cubic splines constrained to be linear beyond boundary knots k_{min}, k_{max}. The spline function is defined as

$$s(x, \boldsymbol{\gamma}) = \gamma_0 + \gamma_1 x + \gamma_2 v_1(x) + \ldots + \gamma_{m+1} v_m(x)$$

where $v_j(x)$ is the jth basis function

$$v_j(x) = (x - k_j)_+^3 - \lambda_j (x - k_{min})_+^3 - (1 - \lambda_j)(x - k_{max})_+^3$$

$$\lambda_j = \frac{k_{max} - k_j}{k_{max} - k_{min}}$$

and $(x - a)_+ = max(0, x - a)$.

Value

A list of class `"flexsurvreg"` with the same elements as described in `flexsurvreg`, and including extra components describing the spline model. See in particular:

`k`	Number of knots.
`knots`	Location of knots on the log time axis.
`scale`	The `scale` of the model, hazard, odds or normal.
`res`	Matrix of maximum likelihood estimates and confidence limits. Spline coefficients are labelled `"gamma..."`, and covariate effects are labelled with the names of the covariates. Coefficients `gamma1, gamma2, ...` here are the equivalent of `s0, s1, ...` in Stata `streg`, and `gamma0` is the equivalent of the `xb` constant term. To reproduce results, use the `noorthog` option in Stata, since no orthogonalisation is performed on the spline basis here. In the Weibull model, for example, `gamma0, gamma1` are `-shape log(scale), shape` respectively in `dweibull` or `flexsurvreg` notation, or `(-Intercept/scale, 1/scale)` in `survreg` notation. In the log-logistic model with shape `a` and scale `b` (as in `dllogis` from the **eha** package), `1/b^a` is equivalent to `exp(gamma0)`, and `a` is equivalent to `gamma1`. In the log-normal model with log-scale mean `mu` and standard deviation `sigma`, `-mu/sigma` is equivalent to `gamma0` and `1/sigma` is equivalent to `gamma1`.
`loglik`	The maximised log-likelihood. This will differ from Stata, where the sum of the log uncensored survival times is added to the log-likelihood in survival models, to remove dependency on the time scale.

Author(s)

Christopher Jackson <chris.jackson@mrc-bsu.cam.ac.uk>

References

Royston, P. and Parmar, M. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Statistics in Medicine 21(1):2175-2197.

See Also

`flexsurvreg` for flexible survival modelling using general parametric distributions.

`plot.flexsurvreg` and `lines.flexsurvreg` to plot fitted survival, hazards and cumulative hazards from models fitted by `flexsurvspline` and `flexsurvreg`.

Examples

```
## Best-fitting model to breast cancer data from Royston and Parmar (2002)
## One internal knot (2 df) and cumulative odds scale
spl <- flexsurvspline(Surv(recyrs, censrec) ~ group, data=bc, k=1, scale="odds")

## Fitted survival
plot(spl, ci=TRUE, lwd=3, lwd.ci=1, col.ci="gray")

## Simple Weibull model fits much less well
splw <- flexsurvspline(Surv(recyrs, censrec) ~ group, data=bc, k=0, scale="hazard")
lines(splw, col="blue")

## Alternative way of fitting the Weibull
splw2 <- flexsurvreg(Surv(recyrs, censrec) ~ group, data=bc, dist="weibull")
```

Index