Using `glmBfp`: Cox models with test based Bayes Factors

Isaac Gravestock

September 7, 2018

The `glmBfp` package implements a new approach to model fitting and variable selection which are described in two articles, Held et al. [2015] and Held et al. [2016]. This vignette shows how to use the simplified interface for fitting Cox models with `glmBfp`. We use the SMART data described in Steyerberg [2009] and available at `http://www.clinicalpredictionmodels.org`. The package includes a processed copy of this data with the missing values imputed and variables transformed as in the example code on that site

```
library(glmBfp)
data(SMARTfull)
```

The workhorse function of the package is `glmBayesMfp()` which does all the model fitting and returns a large list of possible models, which then need to be evaluated and selected from. To make things easier, `coxTBF()` is a simplified formula based interface to `glmBayesMfp()` that fits and chooses Cox models.

We first need to define the formula, using `Surv(time,event)`. The function needs to know which variables must be included in the model and which are "uncertain" should tested for inclusion. These are wrapped in the formula with `uc()`.

```
f1 <- Surv(TEVENT, EVENT) ~ AGE.TRANS + SEX +  uc(SMOKING) +
  uc(ALCOHOL) + uc(BMI) + uc(SYSTH) + uc(HDL) + uc(DIABETES) +
  uc(HISTCAR2)
```

Using this we can fit a model. We choose `type="MAP"` to select the model with maximum posterior probability. Other possibilities are the median probability model (`MPM`) and Bayesian model averages (`BMA`).

```
f1_MAP <- coxTBF(f1, data = SMARTfull, type="MAP",
                 useOpenMP=FALSE, chainlength=200)
```

The resulting object contains the formula of selected model and the coefficient estimates.

```
f1_MAP$formula

## survival::Surv(TEVENT, EVENT) ~ BMI + DIABETES + HDL + HISTCAR2 +
##     SYSTH + AGE.TRANS + SEX
## <environment: 0x562e9cb7ec50>

f1_MAP$coefs

##          BMI     DIABETES          HDL     HISTCAR2        SYSTH
## -0.028174790  0.271684998 -0.471942464  0.402346077  0.006333977
##    AGE.TRANS          SEX
##  0.001563272 -0.195002679
```

Also included is the survivor function, so we can predict survival probabilities at specified times.

```
predict(f1_MAP, times = c(100,1000,2000,3000), newdata = SMARTfull[1:3,])

##            100      1000      2000      3000
## [1,] 0.9917292 0.9510541 0.9054819 0.8553081
## [2,] 0.9842335 0.9084385 0.8269689 0.7415086
## [3,] 0.9886579 0.9333954 0.8725212 0.8068130
```

Other parameters given to `coxTBF` are passed through to `glmBayesMfp`. This can be used to specify new g-priors and change MCMC options. We can also save the models found in the search for later investigation with `keepModelList=TRUE`. This time we select the MPM (median probability model). If this model is not one of the models found in the Monte Carlo search, then it is constructed and returned.

```
# Hyper g/n.obs
prior <- InvGammaGPrior(a=1/2, b=sum(SMARTfull$EVENT)/2)
f1_MPM <- coxTBF(f1, data = SMARTfull, type="MPM", useOpenMP=FALSE,
                 chainlength=500, nModels=50, keepModelList=TRUE,
                 priorSpecs=list(gPrior=prior, modelPrior="sparse"))

## [1] "MPM model wasn't fitted so we construct it."
```

```
f1_MPM$formula

## survival::Surv(TEVENT, EVENT) ~ DIABETES + HDL + HISTCAR2 + SYSTH +
##     AGE.TRANS + SEX
## <environment: 0x562e8c9e9fe0>

f1_MPM$coefs

##     DIABETES          HDL     HISTCAR2        SYSTH    AGE.TRANS
##  0.234646707 -0.417436550  0.411543874  0.006400702  0.001626840
##          SEX
## -0.215846893
```

References

Leonhard Held, Daniel Sabanés Bové, Isaac Gravestock, et al. Approximate Bayesian model selection with the deviance statistic. *Statistical Science*, 30(2):242–257, 2015.

Leonhard Held, Isaac Gravestock, and Daniel Sabanés Bové. Objective Bayesian model selection for Cox regression. *Statistics in Medicine*, 35(29):5376–5390, 2016.

E. W. Steyerberg. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Springer, 2009.