R simulation exmaple in the AFT-Bayes-LASSO manuscript

Zhang, Z. and Sinha, S.
Email: zhangquake@gmail.com

August 18, 2019

Abstract

Simulation example rewritten in R for the AFT-Bayes-LASSO manuscript to compare with other methods in the literature.

Contents

Simulation example

Simulate data as in Zhang et al. [2018].

```r
rm(list=ls())
require(survival)
require(flexsurv)
require(glmnet)

set.seed(23)
n <- 500   #or fully 5000   # the size of a cohort
p <- 20 # the number of covariates
betavec <- c(0.5, 0.5, 0.35, -0.35, rep(0, 16))
z <- matrix(NA, nrow=n, ncol=p)
for(i in 1:p)  z[,i] <- rbinom(n, 1, 0.5)
mu <- z%*%betavec
epsilon_star <- rweibull(n, 0.82, 29.27)
epsilon <- rnorm(n, 0, 1) #log(epsilon_star)/sqrt(1.57)
mytime <- exp( 1 + mu + exp(-0.5*mu^2)*epsilon )
cen.time <- z[,1] + z[,2] + runif(n, 0, 200)
myv <- apply(cbind(mytime, cen.time), 1, min)
mydel <- as.numeric(mytime<cen.time)
#+++++++++ prepare your data. CAUTION: put covariate matrix z at first
dat <- data.frame(z, time=myv, status=mydel)
```

Cox-LASSO method

```r
y <- cbind(time=dat$time, status=dat$status)
x <- as.matrix(dat[,1:p])
cvfit <- cv.glmnet(x, y, family="cox") #first do 10-fold cross-validation to select lambda
m <- glmnet(x, y, family="cox", lambda=cvfit$lambda.min) #plugin the optimal lambda
cat('Cox-LASSO selected: ', paste0('X', which(m$beta!=0)), '\n')
```

```
## Cox-LASSO selected:  X1 X2 X3 X4 X6
```

step-genF method

```r
forward <- TRUE  #TRUE=forward selection, FALSE=backward elimination
if(forward){ #+++++  forward selection
  fit0 <- flexsurvreg(Surv(time, status) ~ 1, data=dat, dist='lnorm')
  coefs <- as.numeric(coef(fit0))
  init0 <- c(coefs[1],exp(coefs[2]), 0, 1e-6, coefs[-c(1:2)])
  fit <- update(fit0, inits=init0, data=dat, dist='genf')
  myfit <- fit
  len <- length(indc <- 1:p)
  q <- length(inds <- numeric(0))
  aic0 <- BIC(myfit)
  aics <- aic0 - 1
  while(q<=p){
    aic0 <- BIC(myfit)
    aics <- numeric(len)
    fits <- list(len)
    for(i in 1:len){
      inds0 <- c(inds, indc[i])
      myform0 <- as.formula(paste("Surv(time, status) ~ ",
                                  paste(names(dat)[inds0], collapse= "+")))
      fit0 <- flexsurvreg(myform0, data=dat, dist='lnorm',method="Nelder-Mead")
      coefs <- as.numeric(coef(fit0))
      init0 <- c(coefs[1],exp(coefs[2]),0, 1e-6, coefs[-c(1:2)])
      fits[[i]] <- update(fit0, inits=init0, dist='genf',method="Nelder-Mead")
      aics[i] <- BIC(fits[[i]])
    }
    optind <- which(aics == min(aics))
    if(min(aics) < aic0){
      myfit <- fits[[optind]]
      # cat('X',indc[optind],' was included.\n',sep='')
      q <- length( inds <- c(inds, indc[optind]) )
      len <- length(indc <- indc[-optind])
    }else q <- p+1 #stop
  }
}else{ #----- backward selection
```

```r
  myform0 <-  as.formula(paste("Surv(time, status) ~ ",
                               paste(names(dat)[1:p], collapse= "+")))
  fit0 <- flexsurvreg(myform0, data=dat, dist='lnorm',method = "Nelder-Mead")
  coefs <- as.numeric(coef(fit0))
  init0 <- c(coefs[1],exp(coefs[2]),0,1e-6, coefs[-c(1:2)])
  fit <- update(fit0, inits=init0, data=dat, dist='genf')
  myfit <- fit
  q <- length(inds <- 1:p)
  aic0 <- BIC(myfit)
  aics <- aic0-1
  while(q>0){
    aic0 <- BIC(myfit)
    aics <- numeric(q)
    fits <- list(q)
    for(i in 1:q){
      inds0 <- inds[-i]
      myform0 <-  as.formula(paste("Surv(time, status) ~ ",
                                   paste(names(dat)[inds0], collapse= "+")))
      fit0 <- flexsurvreg(myform0, data=dat, dist='lnorm',method = "Nelder-Mead")
      coefs <- as.numeric(coef(fit0))
      init0 <- c(coefs[1],exp(coefs[2]),0, 1e-6, coefs[-c(1:2)])
      fits[[i]] <- update(fit0, inits=init0, dist='genf')
      aics[i] <- BIC(fits[[i]])
    }
    optind <- which(aics == min(aics))
    if(min(aics) < aic0){
      myfit <- fits[[optind]]
      # cat('X',inds[optind],' was excluded.\n',sep='')
      q <- length( inds <- inds[-optind] )
    }else q <- 0 #stop
  }
}
cat('step-genF selected: ', paste0('X', sort(inds)), '\n')
```

```
## step-genF selected:  X1 X2 X3 X4
```

References

Zhen Zhang, Samiran Sinha, Tapabrata Maiti, and Eva Shipp. Bayesian variable selection in the accelerated failure time model with an application to the surveillance, epidemiology, and end results breast cancer data. *Statistical methods in medical research*, 27(4):971–990, 2018.