Implementation Example of Diagnostics for the Cox Model

Yishu Xue, Elizabeth D. Schifano

We use a simulated dataset to demonstrate some of the diagnostic methods for the Cox model (Cox 1972). The dataset sampleDat.csv consists of 700 observations, with a censoring rate of 45.6%. There are two categorical covariates, Occupation and Gender, and two continuous covariates, Age and Initial Body Weight. The continuous covariates have been standardized before fitting the model. Survival times are measured in days.

1. Functional Form

As suggested in Keele (2010), we first attempt to identify appropriate functional forms before checking for nonproportionality. We have two continuous covariates, Age and Weight, and we use the techniques of Terry M. Therneau, Grambsch, and Fleming (1990) to choose their appropriate functional forms. The analysis is performed using the survival package (Terry M Therneau 2017). We fit two null models excluding Age and Weight one at a time, obtain their martingale residuals, and plot them against the ignored covariates, respectively. We also superimpose the pointwise loess confidence bands. While the loess fit for Age is roughly linear, that for Weight is curvy. Therefore, it is reasonable to use Age and the square of Weight in the model. This is also in accordance with our intuition that being too heavy or too thin is an indicator of negative health condition.

```r
# a model without Age
testFit1 <- coxph(Surv(survtime, status) ~ Gender + Occupation + Weight, data = sampleDat)
m.res1 <- residuals(testFit1, "martingale")
qplot(sampleDat$Age, m.res1, size = I(0.3)) + geom_smooth() + xlab("Age") +
ylab("Martingale Residual")
```

```r
# a model without Weight
testFit2 <- coxph(Surv(survtime, status) ~ Age + Gender + Occupation, data = sampleDat)
m.res2 <- residuals(testFit2, "martingale")
qplot(sampleDat$Weight, m.res2, size = I(0.3)) + geom_smooth() + xlab("Weight") +
ylab("Martingale Residual")
```
2. Proportional Hazards

We use \texttt{coxph} to fit a proportional hazards model, using \textit{Age}, \textit{Gender}, \textit{Occupation}, \textit{Weight} and \textit{Weight2}, which is the square of \textit{Weight}.

\begin{verbatim}
sampleDat$Weight2 <- (sampleDat$Weight)^2
fml <- Surv(survtime, status) ~ Age + Gender + Occupation + Weight + Weight2
trueFit <- coxph(fml, data = sampleDat)
trueFit
## Call:
## coxph(formula = fml, data = sampleDat)

## coef exp(coef) se(coef)  z  p
## Age 0.1707 1.1862 0.0488 3.50 0.00047
## GenderMale 0.0180 1.0182 0.1030 0.17 0.86117
## OccupationNon-Faculty -0.3632 0.6955 0.1077 -3.37 0.00075
## Weight -0.0515 0.9498 0.0489 -1.05 0.29154
## Weight2 0.0940 1.0986 0.0346 2.72 0.00658

## Likelihood ratio test=31.3 on 5 df, p=0.00000806
## n= 700, number of events= 381
\end{verbatim}

The function \texttt{cox.zph} enables us to test for violations of the proportional hazards assumption both globally and for each individual covariate. Since our censoring rate is 45.6%, we consider using the KM-transformation and rank transformation when performing the test as suggested by Park and Hendry (2015).

\begin{verbatim}
cox.zph(trueFit, transform = "km")

## rho  chisq  p
## Age 0.0495 0.8606 0.354
## GenderMale 0.0619 1.4543 0.228
## OccupationNon-Faculty -0.0276 0.2889 0.591
## Weight -0.0154 0.0889 0.766
## Weight2 -0.0730 2.0887 0.148
## GLOBAL  NA 4.2838 0.509
\end{verbatim}

\begin{verbatim}
cox.zph(trueFit, transform = "rank")

## rho  chisq  p
\end{verbatim}
Both results indicate that all covariates pass the individual proportionality test at the 0.05 level, and the model passes the global proportionality test at the 0.05 level.

We plot the Schoenfeld residuals against survival times to graphically assess proportionality. If the assumption is satisfied, the Schoenfeld residuals should approximately scatter around 0 (Grambsch and Therneau 1994). For all five covariates, 0 is within the smoothed loess pointwise confidence bands. In addition, the scattering pattern of the Schoenfeld residuals for the two categorical covariates do not fluctuate significantly with respect to time. The survminer package (Kassambara and Kosinski 2017) provides functions to make survival plots in ggplot2 style.

```r
ggcoxdiagnostics(trueFit, type = "schoenfeld", point.size = 0.3)
```

In addition to the Schoenfeld residuals, we can also use a plot of cumulative sums of martingale residuals against survival time/continuous covariates to assess proportionality (Lin, Wei, and Ying 1993). The cumulative sums of martingale residuals with respect to Age and Weight both fluctuate around 0.

```r
m.res <- residuals(trueFit, type = "martingale")
m.res.sorted1 <- m.res[order(sampleDat$Age)]
```
m.res.sorted2 <- m.res[order(sampleDat$Weight)]
mplot1 <- qplot(y = cumsum(m.res.sorted1), x = sampleDat$Age[order(sampleDat$Age)], geom = "path", size = I(0.3)) + xlab("Age") + ylab("Cumulative Martingale Residual")
mplot2 <- qplot(y = cumsum(m.res.sorted2), x = sampleDat$Weight[order(sampleDat$Weight)], geom = "path", size = I(0.3)) + xlab("Weight") + ylab("Cumulative Martingale Residual")
grid.arrange(mplot1, mplot2, ncol = 2)

For the categorical variables, we use the methods described in Hess (1995) to plot survival curves. With the ggfortify package (Tang, Horikoshi, and Li 2016), we are able to plot the survival curves and its different transformations.

The curves for the two levels of Gender have little discrepancy on all three plots, indicating that the proportional effect of Gender is insignificant. The curves for the two levels of Occupation, however, display clear deviations from each other, and meet the standards in Hess (1995). This also echoes the fact that Occupation is very significant in the fitted Cox model.

## for Gender
sFit1 <- survfit(Surv(survtime, status) ~ Gender, data = sampleDat)
sF1Plot1 <- ggplot2::autoplot(sFit1, censor = FALSE, conf.int = FALSE) + theme(legend.position = "top") + ylab("S(t)")
sF1Plot2 <- ggplot2::autoplot(sFit1, fun = "cumhaz", censor = FALSE, conf.int = FALSE) + theme(legend.position = "top") + ylab("Cumulative Hazard")
sF1Plot3 <- ggplot2::autoplot(sFit1, fun = "cloglog", censor = FALSE, conf.int = FALSE) + theme(legend.position = "top") + ylab("log(-log(S(t)))")
grid.arrange(sF1Plot1, sF1Plot2, sF1Plot3, ncol = 3)
3. Outliers

We plot the martingale and deviance residuals to check for possible outliers as discussed in Terry M. Therneau, Grambsch, and Fleming (1990). The scale for x-axis can be linear predictions, time, or observation ID. We choose the linear predictions here.

devPlot1 <- ggcoxdiagnostics(trueFit, type = "martingale", ox.scale = "linear.predictions", point.size = 0.3) + scale_color_grey()
devPlot2 <- ggcoxdiagnostics(trueFit, type = "deviance", ox.scale = "linear.predictions", point.size = 0.3) + scale_color_grey() + geom_hline(lty = 2, col = "red", yintercept = c(1.96, -1.96))
marrangeGrob(list(devPlot1, devPlot2), ncol = 2, nrow = 1, top = "")
The deviance residuals identified a group of potential outliers. They turned out to be younger and weigh less than the rest of observations, but they all had an event very soon.

We also use the log-odds and normal deviate residuals of Nardi and Schemper (1999) to look for potential outliers. We can see that these residuals identified approximately the same set of potential outliers as the deviance residuals.

```r
pred.lp <- predict(trueFit, type = "lp")
pred.Surv <- exp(-predict(trueFit, type = "expected"))
lgo.res <- log(pred.Surv/(1 - pred.Surv))
nd.res <- qnorm(pred.Surv)

lgo.res <- data.frame(lgo.res = lgo.res, lp = pred.lp)
nd.res <- data.frame(nd.res = nd.res, lp = pred.lp)

lgoPlot <- ggplot(lgo.res, aes(x = pred.lp, y = lgo.res)) + geom_point(size = I(0.3)) + xlab("Linear Predictions") + ylab("Log-odds Residual") + geom_hline(lty = 2, col = "red", yintercept = c(3.66, -3.66))

ndPlot <- ggplot(nd.res, aes(x = pred.lp, y = nd.res)) + geom_point(size = I(0.3)) + xlab("Linear Predictions") + ylab("Normal Deviate Residual") + geom_hline(lty = 2, col = "red", yintercept = c(1.96, -1.96))

grid.arrange(lgoPlot, ndPlot, ncol = 2)
```
4. Influential Observations

The *dfbetas* residual, which is a transformation of the score residual, enables us to check the influence of dropping any single observation on parameter estimates. We see from the figure below that no observation had any *dfbetas* residuals greater than 1, which indicates that there are no particularly influential observations. The *Weight2* covariate of the five observations (118, 143, 152, 557, 662), however, are greater than 0.2, much larger than the other observations.

```
ggcoxdiagnostics(trueFit, type = "dfbetas", point.size = 0, hline.col = "black", sline.col = "black") + geom_bar(stat = "identity")
```

We also plot the likelihood displacement (Pettitt and Daud 1989) of each observation. The five observations identified by *dfbetas* residual also have the largest likelihood displacement. The 152nd observation is significantly underweight and she experienced an event very soon. Two of the other four observations are significantly underweight, and the other two are significantly overweight, but none of them had an event.

```
s.res <- residuals(trueFit, type = "score")
ld <- diag(s.res %*% trueFit$var %*% t(s.res))
lhd <- data.frame(ObservationID = 1:nrow(sampleDat), LD = ld)
qplot(x = ObservationID, y = LD, data = lhd, geom = "point", ylab = "Likelihood Displacement", xlab = "Observation ID", size = I(0.3)) + geom_bar(stat = "identity")
```
References


