Extra exercises in STK4080 2021

Solution Exercise E4.1

a) Simulate 100 observations of the triple $(t, delta, x)$ by using the commands:

```
x=rgamma(100,2)
T=sqrt(rexp(100)*2*exp(-x))
C=rexp(100,.5)
t=pmin(C,T)
delta=1*(T<C)
```

Write down the density of the variables x and C.

Use the code to find an expression for the hazard rate function of the lifetimes T. [*Answer:* $\alpha(t|x) = te^x$]

Solution: x has density xe^{-x}. C has density $0.5e^{-0.5c}$.

We simulate random variables

$$T = \sqrt{U \cdot 2e^{-g(x)}}$$

where $f_U(u) = e^{-x}$ and $g(x) = x$. Then

$$S_T(t) = P(T > t) = P\left(U > (1/2)t^2 e^{g(x)}\right) = \exp\{-(1/2)t^2 e^{g(x)}\}$$

so the hazard rate is

$$\alpha_T(t) = \frac{-S_T'(t)}{S_T(t)} = te^{g(x)} = te^x$$

b) Fit a Cox-model $\alpha(t|x) = \alpha_0(t)e^{\beta x}$ to the data using, e.g., the command

```
cfit=coxph(Surv(t,delta)~x)          (*)
```

Then plot the martingale residuals with a corresponding lowess smooth (you may follow the setup of Slides 12). Give a comment to the plot.

Solution: Use the commands

```
cfit=coxph(Surv(t,delta)~x)
summary(cfit)
martres = cfit$residuals
plot(x,martres)
lines(lowess(x,martres))
```

c) Now let x and C have the same distributions as before, but simulate new T by

```
T=sqrt(rexp(100)*2*exp(-log(x)))
```

Write down the hazard rate of the new T and put it on the form of Cox-regression with a transformed covariate. What is the transformation of x?

Solution: Simulate using:

```
x=rgamma(200,2)
T=sqrt(rexp(200)*2*exp(-log(x)))
C=rexp(200,.5)
t=pmin(C,T)
delta=1*(T<C)
```

Let $g(x) = \log(x)$ in the solution to (a) above. From this, the hazard rate of T is now $te^{\log(x)}$. The distribution of T can now be put on Cox-form as $\alpha(t|x) = \alpha_0(t)e^{\beta \log(x)}$. The transformation is hence $\log(x)$.

d) Then fit a Cox-model using (*) with the new data, thus still assuming the hazard ratio to be $e^{\beta x}$.

Plot the new martingale residuals and the lowess smooth. Comment on the fit.

Solution: Use the commands

```
cfit=coxph(Surv(t,delta)~x)
summary(cfit)
martres = cfit$residuals
plot(x,martres)
lines(lowess(x,martres))
```

e) Finally, try to find the correct form of a transformation of x in the Cox model, i.e., try to find an $f(x)$ such that the hazard ratio is $e^{f(x)}$.

Since there is only one covariate, you should start by fitting an empty model and then look at its martingale residuals.

```
cfit.nox=coxph(Surv(t,delta)~1)
martres.nox = cfit.nox$residuals
```

Then make a lowess smooth and comment!

Solution: Use

```
cfit.nox=coxph(Surv(t,delta)~1)
summary(cfit.nox)
martres.nox = cfit.nox$residuals
plot(x,martres.nox)
lines(lowess(x,martres.nox))
```

The plot shows a tendency to a nonlinear trend, which may come from a logarithmic transformation.

We may try also the following commands:

```
splinefit = coxph(Surv(t,delta)~pspline(x))
print(splinefit)
termplot(splinefit)
```

Exercise E4.2

In this exercise we will study data simulated from a regression model where the "β" depends on time. Let there be a single covariate $x > 0$, drawn from the same distribution as in the previous exercise, and let C now be exponential with hazard rate 0.3. Assume that the true hazard rate is

$$\alpha(t|x) = e^{\beta \log(t) x} = t^{\beta x}$$

a) Is the given model a proportional hazards model?

Show that for given value of $x > 0$, T has survival function

$$S(t|x) = \exp\left\{-\frac{t^{\beta x+1}}{\beta x + 1}\right\}$$

and hence is Weibull-distributed with shape parameter $\beta x + 1$ and scale parameter $(\beta x + 1)^{1/(\beta x+1)}$ (with the parameterization used by R)

Solution: We know that

$$\alpha(t|x) = (d/dt)(-\log S(t|x)) = (d/dt)\frac{t^{\beta x+1}}{\beta x + 1} = t^{\beta x}$$

b) The following code will simulate $n = 200$ triples $(t, delta, x)$ with the given distribution for t given x:

```
n=200
x=rgamma(n,2)
beta=.2
y=beta*x+1
T=rweibull(n,y,y^(1/y))
C=rexp(n,.3)
t=pmin(C,T)
delta=1*(T<C)
```

c) Fit an ordinary Cox-model with hazard ratio $e^{\beta x}$ to the data and comment on the results. You may also look at martingale residuals.

Then do the test of proportional hazards by using the `cox.zph` function, and draw a scaled Schoenfeld residual plot. Use for example

```
cox.zph(cfit,transform="log")
plot(cox.zph(cfit,transform="log"))
```

(see Slides 12). Comment on the result. (Note that the resulting p-value will vary if you simulate new sets of 200 observations.)

Solution: Use these commands:

```
n=200
x=rgamma(n,2)
beta=.2
y=beta*x+1
T=rweibull(n,y,y^(1/y))
C=rexp(n,.3)
t=pmin(C,T)
delta=1*(T<C)

sum(delta)

cfit=coxph(Surv(t,delta)~x)
summary(cfit)

cox.zph(cfit,transform="log")
plot(cox.zph(cfit,transform="log"))
```

d) Try other choices for "transform" in the `cox.zph` function. Here is from the R-documentation: **Transform** is a character string specifying how the survival times should be transformed before the test is performed. Possible values are "km", "rank", "identity" or a function of one argument.