TMA4275 Life time analysis
Exercise 6, Spring 2021

Problem 1:

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

```
x=rgamma(100,2)
T=sqrt(rexp(100)*2*exp(-x))
C=rexp(100,0.5)
t=pmin(C,T)
delta=1*(T<C)
```

Write down the density functions for the variables x and C.

The triple gives a data set of 100 observations of $(\widetilde{T}, D, x)$, according to the standard notation.

Use the above R code to find an expression for the hazard rate function of the underlying lifetimes T_i.

b) Fit a Cox model to the data using for example the R command

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

Then plot the martingale residuals with a corresponding lowess smooth. To do this you can use the R commands

```
martres = cfit$residuals
plot(x,martres)
lines(lowess(x,martres))
```

Give a comment to the plot.

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

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

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

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. This you can use by the R commands

```
cfit.nox=coxph(Surv(t,delta)~1)
martres.nox = cfit.nox$residuals
```
Then make a lowess smooth and comment.

Problem 2:

In this exercise we will study data simulated from a regression model where the true "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 exponentially distributed with hazard rate 0.3. Assume that the true hazard rate is

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

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 $(\beta x + 1)^{1/(\beta x+1)}$ in the parameterization used by R.

b) Use the following R code to simulate $n = 200$ triples $(t, delta, x)$ with the given distribution for t given x when $\beta = 0.2$.

```
n=200
x=rgamma(n,2)
beta=0.2
y=beta*x+1
T=rweibull(n,y,y^(1/y))
C=rexp(n,0.3)
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 the R code

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

Read also about Schoenfeld residuals and the *cox.zph* function in Section 7.2.2 in More (2016). Note that the resulting p-value will change if you simulate a new set of 200 observations.

d) Try other choices for "transform" in the *cox.zph* function. The R-documentation of the *cox.zph* function say: 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.