Exercises in the prognostics course*

27 februari 2015

Exercise 1 (M). Exercise 1.1 in Cox and Oakes.

Exercise 2 (M). Exercise 1.2 in Cox and Oakes.

Exercise 3 (M). Exercise 2.1 in Cox and Oakes.

Exercise 4. Exercise 2.4 in Cox and Oakes.

Exercise 5. Exercise 2.11 in Cox and Oakes.

Exercise 6 (M). The expression

$$I(x; a_{j-1}, a_j) = \begin{cases} 0 & (x < a_{j-1}) \\ x - a_{j-1} & (a_{j-1} \leq x < a_j) \\ a_j & (x \geq a_j) \end{cases}$$

on page 54 in Cox and Oakes is incorrect. Find the correct one.

Exercise 7 (M). Exercise 4.1 in Cox and Oakes.

Exercise 8 (M). Use R to do basic survival analysis on the operational data used in [3] and recreate Figures 2 and 3a,b.

There are two different ways of estimating the reliability function, either using the Kaplan-Meir (limit product) estimator or using the Nelson-Aalen estimate of the cummulative hazard function.

> *"The Kaplan-Meir estimator is not a great estimator due to poor properties in the tails. The Nelson-Aalen is a good estimator."*
> - Hemant Ishwaran (personal correspondance)

Therefore,

- Compute the Kaplan-Meir estimate of the reliability function, see [1].
- Compute the Nelson-Aalen estimator of the reliability function. Not described in [1], search and you will find.

In [1] there are two different ways of computing an estimate for the variance of the reliability function estimate, equations (4.5) and (4.6). Computre variance estimates for the reliability functions above and comment on the results.

Hint for R to load a suitable library and load the data (can be found on hte course web-site):

*Exercises with (M) denotes mandatory exercises.

```
library(survival)
vorData <- read.csv('vor_edit_R.csv', header=T, sep=",")
```

Exercise 9. Consider the simple case where each sample consists of three auxiliary variables $x = (\texttt{var1}, \texttt{var2}, and \texttt{noise1})$. The hazard for each sample is

$$h(x) = \begin{cases} 2 & \texttt{var1} = 0 \\ 0.5 & \texttt{var1} = 1 \\ 1 & \texttt{var1} = 2 \end{cases}$$

Since the hazard is constant, the failure times will be exponentially distributed.

In file `simdata.csv`, 5000 samples of the above model has been simulated. The data has a censoring rate of approximately 50%. The file `simpred.csv` consists of 9 samples that should be used to predict reliability functions.

a) Build an RSF model to model the reliability function. Motivate choice of parameters for the algorithm.

b) Plot the error-rate as a function of number of trees. Comment on the limit-value.

c) Compute a lower limit for the error-rate.

d) Predict reliability functions for the prediction samples and compare with the theoretical reliability function.

e) Verify if an accelerated-life model fits the data or not.

Exercise 10. Estimate an RSF model based on the data in `'vor_edit_R.csv`. Plot variable importance and error-rate. Predict the reliability and battery prognostics functions ([3]) for a few vehicles with age ≈ 1 year and a functioning battery.

Exercise 11. Assume a reliability function $R(t)$ and T is the random variable for the failure time

a) Prove that

$$E(T) = \int_0^\infty R(t)\, dt$$

b) Assume a functioning vehicle, at $t = t_0$, with an estimated reliability function $\hat{R}(t)$. Use the result from the a)-exercise to formulate an estimate of the expected remaining life, i.e.,

$$E(T - t_0 | T \geq t_0)$$

Exercise 12. Assume that we want to apply the method in [2] to the starter battery prognostics problem defined by Scania. It is okay to assume that battery voltage and battery current are measured. Propose one or several performance constraints C_i that define T_{EOL}. Which variables are included in each C_i? What would be the damage variables and wear parameters? In the prediction step the future inputs are needed. Which are the input signals in this case? Suggest a way to select future inputs for the RUL prediction. We have studied two different ways of predicting RUL, i.e., by using RSF [3] and as in [2]. Could these method be combined to produce an even better RUL estimation that takes both fleet data and on-board vehicle data into consideration, and if so how? Note that the idea with this exercise is not to provide explicit equations to any of the questions just to conceptually describe variable dependancies qualitatively.

Referenser

[1] David Roxbee Cox and David Oakes. *Analysis of survival data*, volume 21. CRC Press, 1984.

[2] M.J. Daigle and K. Goebel. Model-based prognostics with concurrent damage progression processes. *IEEE Transactions on Systems, Man, and Cybernetics*, 43(3):535–546, May 2013. doi: 10.1109/TSMCA.2012.2207109.

[3] Erik Frisk, Mattias Krysander, and Emil Larsson. Data-driven lead-acide battery prognostics using random survival forests. In *Proceedings of the Annual Conference of The Prognostics and Health Management Society*, Fort Worth, Texas, USA, 2014.