B.4 Modern Survival Problem sheet 4: Nonparametric testing and semiparametric models

To be turned in by noon on 20 November, 2015

(1) The object `tongue` in the package `KMsurv` lists survival or right-censoring times in weeks after diagnosis for 80 patients with tongue tumours. The `type` random variable is 1 or 2, depending as the tumour was aneuploid or diploid respectively.

   (a) Use the log-rank test to test whether the difference in survival distributions is significant at the 0.05 level.

   (b) Repeat the above with a test that emphasises differences shortly after diagnosis.

   (c) Calculate and plot the estimated excess mortality for aneuploid compared with diploid.

(2) You may be familiar with the *Wilcoxon rank-sum test* (also called the *Mann–Whitney U test*). This is a nonparametric substitute for the T test, for comparing two samples, to test whether they came from the same distribution, without assuming that the distributions are normal. Look up the properties of this test. Show that the class of non-parametric test statistics that we have defined in section 7.4 includes the Wilcoxon rank-sum statistic, in the special case where there is no censoring or truncation. What weight function do we need to take to recover the rank-sum statistic? Derive the sampling distribution of the rank-sum statistic. *Optional: Try out the two statistics on some simulated data. They may be drawn from any distribution you like.*

(3) *(Based on problem 4.6 of [ABG08].)* Suppose we have an additive-hazards model where an individual has covariates $(X_1, \ldots, X_p)$ and the individual hazards are then

$$\alpha(t) = \beta_0(t) + \beta_1(t)X_1 + \cdots + \beta_p(t)X_p,$$

where the X_k are random variables.

   (a) Suppose the variable X_p is not observed, so is not included in the model. If the random variables X_k are all independent, show that the remaining model is still an additive-hazards model with a different baseline hazard $\beta_0(t)$.

   (b) Suppose the random variables X_k are multivariate normal (but not independent). How does the model change when X_p is dropped?

(4) In section 8.4.1 we describe fitting the Aalen additive hazards model for the special case of a single (possibly time-varying) covariate. Suppose we constrain the assumptions further, to assume that x_i is constant in time, and takes on only the values 0 and 1. Explain how this is related to the excess mortality model. Compare the results we would obtain from the methods described in this section, to those obtained from the methods of section 7.2.

(5) Refer to the AML study, which is described at length in Example 7.4.4 and analysed with the Cox model in section 10.3. Using the data described in those places, estimate the difference in cumulative hazard to 20 weeks between the two groups by

(a) The nonparametric method described in section 7.2;

(b) The semiparametric method based on the relative-risk regression.

(c) Using the proportional hazards method, suppose an individual were to switch from maintenance to non-maintenance after 10 weeks, and suppose the hazard rates change instantaneously. Estimate the difference in cumulative hazard to 20 weeks between that individual and one who had always been in the non maintenance group.