Analysis of Survival Data: Task Sheet 5 for Week 11

(Released 05/12/11, see §1.1–§4.6)

1. Now that the course is complete this is a good time to consolidate the material by reading the introduction and summary sections of all the chapters, perhaps going back to particular sections if there are items in the summaries which you feel you did not understand fully at the time. The first of the two questions below is another exercise on Kaplan-Meier estimation. The other is straightforward for parts (a) and (b) following almost directly from the definitions and inter-relationships of hazard, survivor and density functions. However, parts (c) and (d) may be more tricky unless you have seen examples on finding the distribution of the minimum of n independent random variables (part (c)) and can follow the argument used for deriving the partial likelihood in a proportional hazards model on page 69. Both of these topics are **beyond the scope of course PAS361** and so these parts have been starred*****. Solutions will be provided soon and you should nevertheless check through these for all parts of the question.

2. The data given below represent survival times for lymphoma patients according to the stage of tumour (where * denotes a censored value):

Stage 3	6	20	42	43*	169*	207	253	255*		
Stage 4	4	10	20	21*	30	33*	43*	46	110	235*

i) Compute the Kaplan-Meier product limit estimates of the survivor functions for stage 3 and stage 4 separately.

ii) Provide estimates of the two cumulative hazard functions and comment on any differences.

iii) By using the log-rank test, compare the survival distributions for the two stages.

3. ★ In an accelerated-life survival model the survivor function for an individual with covariate x satisfies

$$S(t{:}x) = S_0(te^{\beta x}),$$

where $S_0(t)$ is some baseline survivor function.

a) ★ Show that the corresponding hazard function satisfies

$$h(t{:}x) = e^{\beta x}h_0(te^{\beta x})$$

where $h_0(t)$ is the baseline hazard function for $S_0(t)$.

b) ★ In a trial where n independent patients, with covariate values $x_1,x_2,\ldots,x_n$ enter at the same time, suppose that all death times are observed and that $S_0(t) = e^{-\lambda t}$ $(t>0)$. Show that the survival time T has an exponential form, and is of proportional hazard form.

c) * Show that the distribution of the time to the first death in the trial is exponential with mean

$$[\lambda\Sigma\exp(\beta x_i)]^{-1}$$

d) * Show that the probability that the j^{th} patient is the first to die is given by $\exp(\beta x_j)/\Sigma\exp(\beta x_i)$

4) * The **R** function `aftreg()` in library `eha` fits parametric accelerated failure time models. The parameter `dist` offers a choice of parametric distributions between `"weibull"`, `"gompertz"`, `"ev"`, `"loglogistic"` and `"lognormal"`.

   a) How can this be used to fit an exponential distribution?

   b) Which of these distributions also a proportional hazards model?