HT 2007: Statistical Lifetime Models, Sheet 5

1. (a) Supppose that we have a random sample which includes right-censored data (censoring assumed non-informative). We wish to decide whether or not a Weibull distribution is appropriate. Using an estimator of the survival function how might we graphically investigate the appropriateness of the model? Given that the model appears to be appropriate how would you test whether or not the special case of an exponential model is valid?

   Suppose that the Weibull model does not appear to be appropriate what graph would you use to consider a log-logistic model?

   (b) Now suppose that there are two groups to be considered (eg smokers v. non-smokers). What graphs would be appropriate for consideration of a proportional hazards model, accelerated life model respectively?

   (c) Investigate the Gehan leukaemia data in respect of both a) and b).

Controls:	$1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23$
Treatment:	$6+, 6, 6, 6, 7, 9+, 10+, 10, 11+, 13, 16, 17+, 19+, 20+, 22, 23, 25+, 32+, 32+, 34+, 35+$

   Here + indicates censored times.

2. Describe the proportional hazards model, explaining what is meant by the partial likelihood and how this can be used to estimate resgression coefficients. How might standard errors be generated?

   Drug addicts are treated at two clinics (clinic 0 and clinic 1) on a drug replacement therapy. The response variables are the time to relapse (to re-taking drugs) and the status relapse =1 and censored =0. There are three explanatory variables, clinic (0 or 1), previous stay in prison (no=0, yes=1) and the prescribed amount of the replacement dose.

   The following results are obtained using a proportional hazards model, $h(t; x) = \mathrm{e}^{\beta . x} h_0(t)$.

Variable	Coeff	St Err	p-value
clinic	-1.009	0.215	0.000
prison	0.327	0.167	0.051
dose	-0.035	0.006	0.000

   What is the estimated hazard ratio for a subject from clinic 1 who has not been in prison as compared to a subject from clinic 0 who has been in prison, given that they are each assigned the same dose?

   Find a 95% confidence interval for the hazard ratio comparing those who have been in prison to those who have not, given that clinic and dose are the same.

3. The life span distribution of machine components of a particular type is affected by varying levels of stress. The stress level is measured by a non-negative variable x. The base level is measured by $x = 0$ and the median lifetime becomes shorter with increasing x.

   Assume that the cumulative hazard rate function for an item under stress x is of the form

$$H(t; x) = H_0(te^{\beta x})$$

   where $H_0(t)$ is the cumulative hazard rate at baseline $x = 0$. What type of model does this describe?

   (a) Show that this is equivalent to assuming that a lifetime $T(x)$ under stress x has the same distribution as $T(0)e^{-\beta x}$.

   (b) Suppose that independent data is to be collected for stress levels $x_1, x_2, \cdots, x_n$. Show that the assumed form of distribution leads to a regression model for $\ln T(x_j)$ (generally non-normal) of the form

$$\ln T(x_j) = \alpha - \beta x_j + \epsilon_j$$

   where ϵ_j are independent, identically distributed and with zero mean. What is the connection between the constant α and $\ln T(0)$?

   (c) If the baseline distribution is unknown suggest how you might estimate α and β. Give these estimators and also their variances (you will need $var(\ln T(0))$ for the latter).

   Is there any way in which you could test the validity of the model?

4. Sketch the shape of the hazard function in the following cases, paying attention to any changes of shape due to changes in value of κ where appropriate.

a) Weibull: $S(t) = e^{-(\rho t)^\kappa}$

b) Log-logistic: $S(t) = \frac{1}{1+(\rho t)^\kappa}$

c) Suppose that it is thought that an accelerated life model is valid and that the hazard function has a maximum at a non-zero time point. Which parametric models might be appropriate?

d) Suppose that $y_1, \cdots, y_n$ are observations from a lifetime distribution with respective vectors of covariates $x_1, \cdots, x_n$. It is thought that an appropriate distribution for lifetime y is Weibull with parameters ρ, κ where the link is $\log \rho = \beta' x$. In the case that there is no censoring write down the likelihood and, using maximum likelihood, give equations from which the vector of estimated regression coefficients β (and also the estimate for κ) could be found . What would be the asymptotic distribution of the vector of estimators?

How would the likelihood differ if some of the observations y_i were right censored (assuming independent censoring)?