Class Notes for Chapter 8 – April 15-17

Reminder: Don't forget Homework 7 due next Friday!

- **Survival data** (i.e., 'time until' data) is typically **skewed** and **censored**. Censored data means right (most common), left, and/or interval. We'll focus mostly on right-censored data analysis, although the nonlinear model in Section 8.6 addresses interval-censored data.
- **Notation** – needed to get at an understanding of the survival and hazard functions. Survival function gives the probability of "survival" beyond some time point. Hazard function gives the hazard of 'dying' in the next instance given survival up to time t.
- The Weibull distribution is considered and illustrated here since it is useful in applications (especially reliability and engineering) and it contains the Exponential distribution as a special case.
- **Equation (8.6)** is a key result – it gives the all-important LL!
- **Example 8.1 (Carcinoma)** – fit the Weibull distribution to these data; two measurements are (right) censored. MLE's are given in Output 8.1a. Non-parametric (i.e., without the Weibull distribution assumption) estimate of the survival function; this estimate is called the ***Kaplan-Meier estimate***; we can use SAS' **LIFETEST** procedure to do so. Using both parametric and non-parametric methods, we obtain point and interval estimates of the median survival time (like LD_{50}) – see bottom of p.5.
- **Example 8.2 (IUD)** – first graph is of N2LL for Exponential distribution, and second graph is obtained by profiling out the nuisance parameter (γ) and gives the N2PLL for the median time.
- **Cox's Proportional Hazards (PH) model** is used to relate hazard functions which vary with **x** to a baseline hazard (h_0) – the covariate vector (**x**) enters as in Equation 8.7. Equation 8.8 is an equivalent manner to write Equation 8.7. When there are two treatments (like drug A and drug B), x is just a dummy variable.

- Another model for survival data is the **Accelerated Failure Time (AFT) model** in Equation 8.11; it too brings covariate(s) into the model. Equation 8.11 is just for a dummy x, but can easily be extended to other covariates (age, gender, blood pressure, etc.).
- **Example 8.3 (Breast Cancer)** – the estimated survival curves on p.7 (obtained by KM estimation) appear to differ for the positively and negatively stained tumors: is this difference statistically significant or just an artifact?

  (1) SAS **LIFETEST** procedure provides nonparametric estimates of quartiles and medians for the two groups, and, in Output 8.2c, 3 tests of this question (null = no difference between the two curves). Most use the **log-rank test**.

  (2) SAS **PHREG** procedure fits the Cox PH model, gives an estimate of b in Output 8.2d – note that the likelihood test gives marginal significance here. The result is given in Equation 8.10: positive survival curve is shifted in from the negative survival curve.

  (3) SAS **LIFEREG** (with **Weibull** distribution) fits the ***parametric*** Cox PH model – here the likelihood test that $\beta = 0$ gives $p = 0.0418$. ***Note the doubling in the TS!!***

  (4) SAS' **LIFEREG** (with **Log-logistic** distribution) fits the (parametric) AFT model – results appear in Equation 8.13.

- Parametric, non-Normal SLR with censored data – use the LL given above to handle the censored measurements. For the Weibull distribution, median is more natural, so we use it in Equation 8.15 instead of mean in Equation 8.14.
- **Example 8.4** – Y = time until death (no censored data here) and X is dose: wish to connect by a line. Intercept (β_0), slope (β_1) and variance-type parameter (γ) are estimated in Output 8.3a; the test of no significant relationship (H_0: $\beta_1 = 0$) is simple. Also simple is the test of LOF (lack of fit) of the assumed line since we could

fit the full model: this is either the one-way ANOVA with 4 levels or a cubic polynomial. The ANOVA results are in Output 8.3b, and the line is fine here ($\chi_2^2 = 148.0 - 147.8 = 0.2$, NS).

- **Interval-censored data** are addressed in **Example 8.5**, so the binary logistic modelling ($\chi_1^2 = 31.1 - 26.4 = 4.7$, $p = 0.0302$) in Appendix 8.8.1. Using the nonlinear model of Farrington (1996) and Collett (2003:286) in conjunction with the log-logistic distribution, we obtain Appendix 8.8.1 (p.18). Here the results are NS ($\chi_1^2 = 128.4 - 126.7 = 1.7$, $p = 0.1932$). We conclude there is no difference between drugs A and B.
- Next issues:
  - nonlinear modelling with censored and skewed data
  - detecting synergy with censored and skewed data.