STA 7179 Survival Analysis
Spring 2009

Instructor Hani Doss—222 Griffin-Floyd; Office Hours: M period 5 and WF period 8. (You can also see me WF 11:40–12:00.) Phone: 273-2991.

Course Web Page `http://www.stat.ufl.edu/~doss/Courses/sa`

Course Description This course gives a theoretical development of statistical methods for analyzing life history data, including censored data and truncated data. Topics covered include the Kaplan-Meier estimator, k-sample tests, proportional hazards regression, and the asymptotic theory associated with all these. Throughout, the counting process approach to survival analysis is used.

Prerequisites Prerequisites—in substance, as opposed to by catalog number—are:

- a course in regression at the graduate level;
- a one-year sequence in theoretical statistics at the graduate level;
- a one-year sequence in probability at the graduate level.

The material in STA 6207 (Regression Analysis), STA 6326-7 (Theoretical Statistics I and II) and STA 6466-7 (Probability Theory I and II) are adequate.

Regarding the probability prerequisite, it is also possible to take this course if you have had only one semester of probability but are currently taking the second semester. It is also possible—but much more difficult—to take this course if you have not had a year of probability but have taken a course in large-sample theory. Specific material from probability theory that we will need is a set of basic facts regarding discrete time martingales, and the central limit theorem (in Lindeberg-Feller form). This material will be reviewed, but a student who is seeing this for the first time should allocate extra time in order to learn this material.

Orientation of the Course This is a Ph.D. elective in the Department of Statistics, and is a core course for the Ph.D. in Biostatistics (in the School of Public Health). A substantial theoretical component is involved. This is not a course on applied survival analysis. If what you are looking for is a course on applied survival analysis you should drop this course immediately.

Texts (i) J.D. Kalbfleish and R.L. Prentice *The Statistical Analysis of Failure Time Data* (2nd edition, 2002), Wiley.

(ii) W.N. Venables and B.D. Ripley, *Modern Applied Statistics with S* (4th edition, 2002), Springer.

We will use the statistical computing language `R` (which can be downloaded for free from `http://www.r-project.org`), and a student who is not familiar with it is strongly advised to become so as soon as possible. This could be done for example by reading

in Venables and Ripley Chapters 1–3, skipping Section 3.10 (to be done by the end of the second week of the semester), Chapter 5, Sections 1–3, and taking a look at Chapter 4 (to be done by the fourth week of the semester). If you prefer to use other statistical languages or statistical packages and do not intend to learn R, you should drop this course immediately.

The text by Kalbfleish and Prentice is very good, but we will not use it much. You may not wish to buy it, but you need to be advised that any serious doctoral student should have a reference text for any course in which he or she has invested a substantial amount of time. This text comes closest to the approach that we will use.

Grading There will be 6 or 7 homework assignments, counting for 15% of the grade; two midterms, each counting for 25% of the grade; and a final, counting for 35% of the grade.

The solutions to the homework assignments must be entirely your own (this applies also to R code).

The two midterms will be given in the evening. Tentative dates are Wednesday February 11 and Wednesday March 18, both at 5:30 pm.

Overview The counting process approach to survival analysis enables us to handle in a unified way a number of different data structures such as random censoring, random truncation, and in general life history data (which arises when individuals are monitored through a series of illness, treatment, wellness, and relapse events). This approach, which will be used in the course, is based on the theory of continuous-time martingales and stochastic integrals, together with the associated martingale central limit theorems. Familiarity with continuous-time martingales will not be assumed; the necessary material will be developed from scratch. The counting process approach involves a different way of viewing and thinking about the data, and this is reflected in the way the computations are done.

Topics

- Overview of models in survival analysis
- Counting processes, stochastic integrals, and martingales
- The one-sample problem and the Kaplan-Meier estimator
- The martingale central limit theorem, and the asymptotic distribution of the Kaplan-Meier estimator
- Comparison of k populations
- Proportional hazards regression. This will take about half the course, and we will discuss the following:
  - The proportional hazards model
  - Partial likelihood
  - Asymptotic distributions and hypothesis tests regarding the regression parameters
  - Residuals and diagnostics
  - Frailties and penalized Cox regression

There will be no class on Friday January 9.