Survival Data Analysis

Dr Nick Fieller
Department of Probability & Statistics
University of Sheffield

visiting

16 April – 27 April 2012

Contents

Contents

- Course web page

http://nickfieller.staff.shef.ac.uk/tampere12/index.html

- Lecture notes and other course material

Purpose of Lectures

- There are 'complete' printed notes
  - These are not a textbook
    - some explanations are omitted
  - They are intended to allow you to concentrate on understanding & for me to cover some material very quickly
- Some lectures will be very close to the printed notes
  - This is **intended**
- Other lectures will fill in details & provide examples & **R** demos

- **Books**
  - **Campbell, M. J. (2001)** *Statistics at Square Two.* BMJ
  - **Collett, D. (2003)** *Modelling Survival Data in Medical Research.* (2nd Edition) Chapman & Hall
  - **Everitt, Brian & Rabe-Heskith, Sophia (2001)** *Analyzing Medical Data Using S-PLUS*. Springer. Support material at

http://web1.iop.kcl.ac.uk/IoP/Departments/BioComp/splusBook.shtml

Contents

- **Objectives**
  - statistical modelling & analysis of *lifetime data*.
  - Lifetime data arise especially in medical statistics and in reliability studies.
    - *survival time:*
      - time from diagnosis to death of a patient
      - *time to recovery* or *remission* of a patient or
      - *time to failure* of an electronic component.

- **Outline**
  - Types of survival data
  - Censoring
  - Parametric & nonparametric approaches
- **Single Sample Methods**
  - survivor & hazard functions
  - Lifetables
  - Kaplan-Meier estimators

- **Two Sample Comparisons:**
  - log rank test
  - maximum likelihood test
  - likelihood ratio test
  - proportional hazards

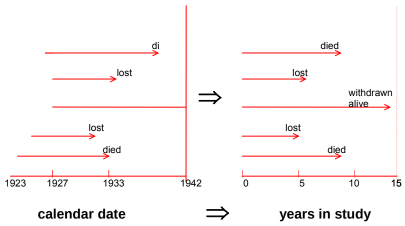
- **Regression Models**
  - parametric models
    - exponential & Weibull
  - nonparametric methods
    - proportional hazards or Cox regression
    - partial likelihood

- **Aims of survival analysis**
  - describe/model a single sample
    - making inferences on a single population
  - compare two or more groups
    - effect of treatments on survival time
  - investigate relationship with covariates
    - effects on survival time of covariates
    - adjust for covariates in comparisons

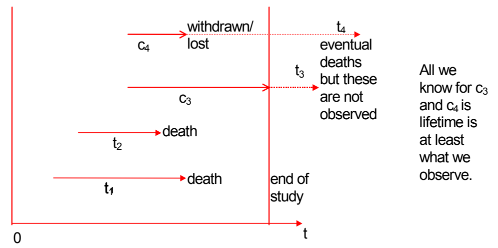
Contents

- **Censoring**
  - event of interest not yet occurred
  - observe only survival time is ***at least*** t
  - some observations t_i (i=1,…,n) are ***censored***
  - censored observations provide information on survival so cannot be ignored

Censoring



Censoring



Censoring

- ***Right censoring***
  - lifetime exceeds some value
- ***Left censoring***
  - lifetime less than some value
- ***Interval censoring***
  - failure occurred during an interval

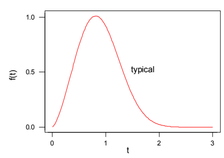
- **Aims**:
  - **estimate lifetime distributions**
    - ⇒ estimate properties of distribution
    - (median lifetime, prob of surviving > 5 years,…)
  - **Censoring**
    - nonparametric — lifetables, Kaplan-Meier
    - parametric — exponential, Weibull.

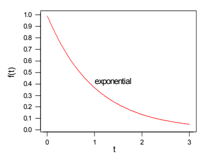
- survival time is a random variable T
- T>0, continuous variable
  - **p.d.f.** f(t) (t>0)
  - **d.f.** F(t)=P[T≤t]
  - (so $f(t) = F'(t)$ and $F(t) = \int_0^t f(u)du$)

- **Survivor function**
  - $S(t) = P[T > t] = 1 - F(t)$
  - (so $S'(t) = -f(t)$, $S(t) = \int_t^\infty f(u)du$)
- **Hazard function**
  - $h(t) = \lim_{\delta t \to 0} \left[\frac{P[t \le T < t + \delta t \mid T \ge t]}{\delta t} \right]$
  - "P[die at time T given survived until T]"

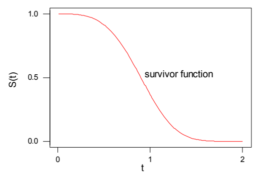
- **Cumulative hazard function**
  - $H(t) = \int_0^t h(u)du = -\log_e S(t)$
  - f(t), S(t), H(t) and h(t) are equivalent characterizations of a survival distribution
  - all interrelated

- **Typical patterns**

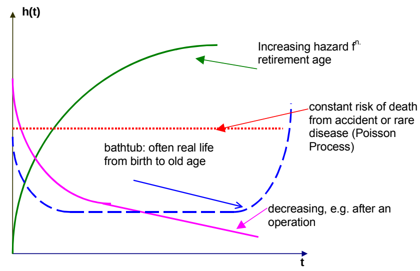




- **Survivor function**



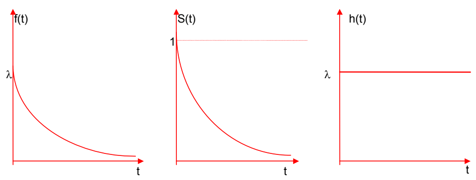
Hazard functions:



- Choose appropriate family of models by recognizing form of hazard function
  - practical situation
  - initial investigation of the data
- then can estimate parameters in model

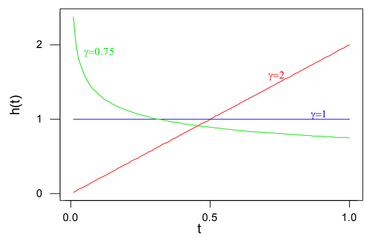
- **Example:** exponential
  - $f(t) = \lambda e^{-\lambda t}$
  - $S(t) = e^{-\lambda t}$
  - $h(t) = \lambda$ (**NB** constant)
    - [= f(t)/S(t)]
- ***only*** constant for exponential survival distribution

- exponential



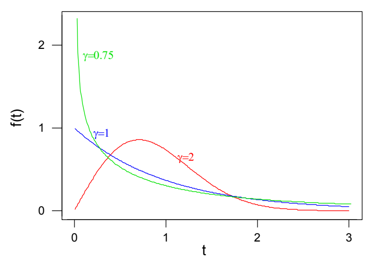
- **Example:**Weibull
  - $f(t) = \lambda\gamma t^{\gamma-1}\exp(-\lambda t^{\gamma})$
  - $S(t) = \exp(-\lambda t^{\gamma})$
  - $h(t) = \lambda\gamma t^{\gamma-1}$
    - $\gamma > 1$: increasing
    - $\gamma = 1$: constant (exponential)
    - $\gamma < 1$: decreasing

- e.g. $\lambda = 1$; Weibull hazard functions



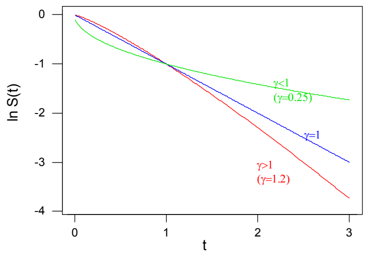
Weibull hazard functions

- e.g. $\lambda = 1$; Weibull density functions



Weibull Density Functions

- e.g. λ = 1; Weibull logsurvivor functions



Weibull log survivor functions

- Weibull family
  - very flexible models
  - allows both increasing (γ>1) & decreasing (γ<1) hazard functions
- Difficult to estimate if γ close to 1
  - (numerical instability)