Contents

Preliminaries
0: Introduction
1: Background & Basic Concepts
2: Single Sample Methods
2.1 Non-parametric Methods
2.6 Parametric Models
3: Two-Sample Comparisons
4: Regression Models
4.2 Parametric Regression Models
4.4 Proportional Hazards models

- **Parametric Models**
  - Estimate unknown parameters by maximum likelihood
- **Exponential**
  - $f(t)=\lambda e^{-t}$ $(t>0)$
  - $S(t)=e^{-\lambda t}$ $(=1-F(t))=1-(1-e^{-\lambda t})$
  - $h(t)=\lambda$

- **Uncensored data**
  - Observe $t_1,t_2,\ldots,t_n$
  - $\text{Lik}(\lambda;t_1,t_2,\ldots,t_n)=L(\lambda)=\Pi f(t_i)=\lambda^n e^{-\lambda\sum_1^n t_i}$
  - $\text{Log}_e(L)=\ell(\lambda)=n\log(\lambda)-\lambda\Sigma t_i$

$$\Rightarrow \quad \hat{\lambda}=\frac{n}{\sum_1^n t_i}$$

- Confidence interval from result that

$$\sum_1^n T_i \sim \Gamma(n,\lambda)$$

- (note chi-squared is special case of gamma)

$$\left(\frac{\chi^2_{2n;\alpha/2}}{2\sum t_i},\ \frac{\chi^2_{2n;1-\alpha/2}}{2\sum t_i}\right)$$

- Also have MLEs of any other function, e.g. survival function

$$\hat{S}(t)=e^{-\hat{\lambda}t}$$

e.g. median survival time $\tilde{t}$

$$S(\tilde{t})=0.5 \Rightarrow \tilde{t}=-\lambda^{-1}\log(0.5)$$

$$\hat{\tilde{t}}=-\hat{\lambda}^{-1}\log(0.5)$$

- **Censored Data**
  - 'Time' censored, n patients,
    - potential lifetimes $T_i \sim Ex(\lambda)$)
    - observe **either** complete lifetime t_i **or** that $T_i > c_i$, $(i=1,2,\ldots,n)$.
  - Contributions to likelihood:

"$P[T_i=t_i]$" = $\lambda e^{-\lambda t_i}$ (if $t_i \le c_i$)

"$P[T_i>c_i]$" = $e^{-\lambda c_i}$ (if $t_i>c_i$)

Define δ_i=1 if $t_i \leq c_i$ (uncensored)

δ_i=0 if $t_i > c_i$ (censored)

Then likelihood $L(\lambda) = \prod_{i=1}^{n} [\lambda e^{-\lambda t_i}]^{\delta_i} [e^{-\lambda c_i}]^{1-\delta_i}$

$$\Rightarrow \hat{\lambda} = \frac{\sum_1^n \delta_i}{\sum_1^n \{\delta_i t_i + (1-\delta_i) c_i\}}$$

$$\hat{\lambda} = \frac{\text{total number of deaths observed}}{\text{total time alive of all patients in the study}}$$

holds for censored and uncensored cases

- Distribution of $\hat{\lambda}$ not easy
  - (not based on simple sum of exponentials)
  - Use general theory of MLE for approximate formula :

$$\text{var}(\hat{\lambda}) \approx \frac{\hat{\lambda}^2}{\sum_1^n (1 - e^{-\hat{\lambda} c_i})}$$

$$\text{var}(\hat{\lambda}) \approx \frac{\hat{\lambda}^2}{\sum_1^n \delta_i}$$ numerically same [almost]

- 100(1– α)% Confidence Interval for λ is

$$\hat{\lambda} \pm z_{1-\frac{1}{2}\alpha} \times \text{s.e.}(\hat{\lambda})$$

Where s.e.($\hat{\lambda}$) = $\sqrt{\text{var}(\hat{\lambda})}$

- Note use of formula

$$\text{var}\{g(\hat{\lambda})\} \approx \left[[g'(\lambda)]^2 \, \text{var}(\hat{\lambda}) \right]_{\lambda=\hat{\lambda}}$$

for finding CIs of functions g(λ) of λ

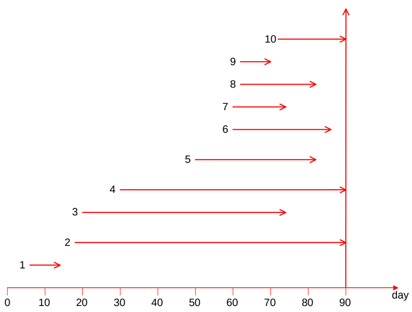
e.g. $\mu = \lambda^{-1}$ = E[T] the mean lifetime

$g(\lambda) = \lambda^{-1}$ so $g'(\lambda) = -\lambda^{-2} = \mu^2$ so

$$\text{var}(\hat{\mu}) \approx \frac{\hat{\mu}^2}{\sum_1^n (1 - e^{-c_i/\hat{\mu}})} \quad \text{or} \quad \text{var}(\hat{\mu}) \approx \frac{\hat{\mu}^2}{\sum_1^n \delta_i}$$

- Example
  - 10 patients, 3 censored (2nd,4th & 10th)

Patient no.	1	2	3	4	5	6	7	8	9	10
Entry time	9	18	20	30	49	59	59	60	61	69
Survival time t_i	2	.	51	.	33	27	14	24	4	.
max possible c_i	81	72	70	60	41	31	31	30	29	21
δ_i	1	0	1	0	1	1	1	1	1	0



- Thus $\Sigma\delta_i = 7$ deaths on study.
  $\Sigma\delta_i t_i = 155$, $\Sigma(1-\delta_i)c_i = 153$

  $\hat{\lambda} = \frac{7}{308} = 0.023$ per day

  $\hat{\mu} = \frac{308}{7} = 44.0$ days

  confidence intervals etc from formula

- data $\{(t_i,)\ ;\ i=1,\ldots,n\}$
  - δ_i=1 if death occurred, 0 if censored

$$L(\theta) = \prod_{\text{deaths}} f(t_i) \prod_{\text{censored}} S(t_i)$$

$$= \prod_{\text{deaths}} h(t_i)S(t_i) \prod_{\text{censored}} S(t_i)$$

$$= \prod_{i=1}^{n} [h(t_i)]^{\delta_i} S(t_i)$$

- This general formula holds for all distributions (lognormal, Weibull, gamma, Gumbel,…..
- All require numerical estimation
  - Minitab, SPSS, S-PLUS

- **Summary**
  - Life tables
    - Grouped data
    - Allow for censoring by adjusting # at risk
  - Kaplan-Meier
    - Individual data
    - Uncensored case = 1 – empirical CDF
    - Express as a product of terms $[1 - d_i/r_i]$
    - Censored case by adjusting # at risk

- Kaplan-Meier estimates are the key exploratory tool for censored data
  - Cannot draw histograms of censored data
    - Don't know which bin to put a censored in
- Kaplan-Meier Plot is always step 1 in analysing censored data

- Parametric models
  - Estimate parameters by MLE
  - Uncensored observations contribute $f(t_i)$ &
  - Censored contribute $S(t_i)$ to likelihood
  - Use MLE theory for standard errors
  - Plug in MLEs for other functions of θ
  - Use formula for s.e.[$g(\theta)$]