23/02/16 ALSM 2

Response Variable:

$y \in \{0,1\}$ Bernoulli

$y \in \{0,1,\dots,N\}$ Binomial

$y \in \mathbb{N}$ (integer) Poisson

$y \in \mathbb{R}$ Gausian

$y \in \mathbb{R}^+$ exponential / weibull → positive real numbers - like time until failure

Lukemia example - response variable y is time. Is placebo time same as drug time?

SURVIVAL ANALYSIS

- Modelling time until failure
- SA concerned with the statistical modelling of time to failure from a well defined origin of time or starting point.
  → Time of hardware to fail since bought.
  → Time of patient to die from the time disease has been diagnosed.

Distribution

- Times y are non-negative ($y \in \mathbb{R}^+$)
- Distributions are skewed with long tails
- Some subjects may survive beyond the study and their failure is not observed
  ↳ In this case, the data is said to be 'censored'

Exponential Distribution

$p(y|\theta) = \theta \exp[-\theta y] \quad y \in \mathbb{R}^+ \quad \theta \in \mathbb{R}^{+*}$ ← excludes 0.

Expectation $\mathbb{E}[y]$?

$\mathbb{E}[y] = \int_{y \in \mathbb{R}} y \, p(y|\theta) dy$

$= \int_0^\infty y \, \lambda \exp(-\lambda y) \, dy$

$= [-y\exp(-\lambda y)]_0^\infty + \int_0^\infty \exp(-\lambda y) \, dy$ (integration by parts)

$= [0-0] + [\frac{-1}{\lambda} \exp(-\lambda x)]_0^\infty$

$= 0 + 0 + 1/\lambda = 1/\lambda$ = positive number

23/02/16

Prove it is a function $\int_0^\infty f_x(x)dx = \int^\infty \lambda \exp(-\lambda y)dy$
$= [\exp[-\lambda y]]_0^\infty$
$= 0 - (-1) = 1 \Rightarrow$ Distribution ✓

Best link function to map θ to positive real number $\mathbb{R}^+$ onto $\mathbb{R}$ g should be exponential, g^{-1} should be log

Weibull Distribution
$P(y|\theta,\lambda) = \lambda\theta y^{\lambda-1} \exp[-\theta y^\lambda]$
$y \in \mathbb{R}^+ \quad \theta \in \mathbb{R}^{+*} \quad x \in \mathbb{R}^{+*}$

Exponential distribution is a special case of weibull with $\lambda = 1$

$\mathbb{E}[y] = \int_0^\infty y\lambda\theta y^{\lambda-1} \exp[\theta y^\lambda]\, dy$
$= \int^\infty \lambda\theta y^\lambda \exp[-\theta y^\lambda]\, dy$

Using substitution $y^\lambda = u^{1/\lambda}$ results in
$du = \lambda y^{\lambda-1} dy \qquad dx = \frac{du}{\lambda y^{\lambda-1}} = \frac{dy}{\lambda\mu^{\frac{\lambda-1}{\lambda}}} = \mu^{\frac{1}{\lambda}-1}\frac{du}{\lambda}$

$\mathbb{E}[x^k] = \int \lambda y^{\lambda+\theta-1} e^{y^\lambda}\, dy$
$= \int \lambda u^{(\frac{\theta}{\lambda}+1)} \frac{1}{\lambda} e^{-u} (\mu^{-1/\lambda-1})\, du_\lambda$
$= \int u^{(\theta/\lambda+1)-1} e^{-u} du$
$= \Gamma(\theta/\lambda + 1) \qquad \Gamma(u) = \int_0^\infty s^{u-1}\exp(-s)ds$

Survivor and Hazard Function
A. Probability of Failure $P(0 \le t \le y) = \int_0^y p(t|\theta)dt$
P can be exponential or weibull (area between 0 and y)
Probability of survival beyond time U
$S(y) = 1 - F(y)$ (area to right of y)

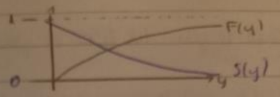
B. $F(y)$ and $S(y)$ of exponential Distribution
$F(y) = \int^y \theta \exp(-\theta t)dt$

Could be hints 15/03/16 DW

2 23/02/16 ALSM2 3

$= [-\exp(-\theta t)]_0^y$

$= 1 - \exp(-\theta y)$ = probability of failure



$S(y) = 1 - 1 + \exp(-\theta y) = \exp(-\theta y)$

- At 0.5 same probability of failure & survival
- When $F(y) = 0.5$ & $S(y) = 0.5$ called the median point of survival $F(y) = S(y) = 0.5$

C. Hazard Function

$h(y) = \frac{P(y \le t \le y + dy)}{P(y \le t)}$ chance of failure between time y and y° / Surviving beyond time y $S(y)$

$P(y|\theta)dy$

"Chances of dieing in the next minute given a survival rate and given you have not died before time y"

$h(y) = P(y|\theta)/S(y)$ derivative/function weibull or exponential

Cumulative hazard function $H(y) = -\log[S(y)]$

D Link Function

Use Log

Exponential Distribution

$f(t) = \lambda e^{-\lambda t}$ $\lambda > 0$

$F(t) = 1 - e^{-\lambda t}$ $S(t) = e^{-\lambda t}$

$h(t) = \lambda$ constant hazard function

$H(t) = \lambda t$

Weibull Distribution

$f(t) = p\lambda^p t^{p-1} e^{-(\lambda t)^p}$

$F(t) = 1 - e^{-(\lambda t)^p}$ $S(t) = e^{(-\lambda t)^p}$

$h(t) = p\lambda^p t^{p-1}$ $t > 0$

$H(t) = (\lambda t)^p$ $\lambda > 0$ scale $p > 0$ shape

Drug Example Continued

$L(\beta) = \prod_{i=1}^{n} p(y_i|\theta)$ weibull or exponential

- This is the definition of the likelihood when y_i is uncensored.
- When y_i is censored, use $S(y_i)$ instead of pdf $p(y_i|\theta)$

- δ_i an indicator variable, 0 if y_i is censored, 1 if it is not $\delta_i \begin{cases} 0 & \text{censored} \\ 1 & \text{observed} \end{cases}$

$\hookrightarrow L(\beta) = \prod_{i=1}^{n} p(y_i|\theta_i)^{\delta_i} S(y_i)^{1-\delta_i}$ take either term depending on censored or not

- Censored: if you lose track of patient or survives duration of study

- Example: Exponential Distribution $L(\theta_1, \dots, \theta_{n2}) = \prod_{i=1}^{n2} [\theta_i \exp(-\theta_i y_i)]^{\delta_i} [\exp(-\theta_i y_i)]^{1-\delta_i}$

  $L(\beta) = \prod_{i=1}^{n2} [\exp(x_i^T \beta)]$

- GLM not defined for use of exponential/weibull use "survival" package

Is the drug better than the placebo effect?

- look at median survival time, $y_{med} = \frac{\log(2)}{\theta}$ for exponential dis

$= \frac{\log(2)}{\exp(\beta^T x)}$ $\begin{cases} x=0 & \text{placebo} \\ x=1 & \text{drug} \end{cases}$ $(\beta_0 + \beta x_i)$

$\frac{y_{med}^{x=1}}{y_{med}^{x=0}} = \frac{1}{\exp \hat{\beta}_1}$ $\exp[-1.39] = 1/4$ look at ratio between drug and placebo $\frac{y_{med}^{x=0}}{4} = y_{med}^{x=1}$

4 times higher than treatment group.

Probability is 4 times higher placebo group - 4 times more likely to fail - placebo group

R gives β's estimates and corresponding S.E's

$2.72 \pm 2(0.04)$ 95% CI for $\beta_1 \Rightarrow \beta_1$ is not 0

↳ relationship between time and drug exists

Exponential Distribution $\mathbb{E}[y] = 1/\theta$

$\log[\mathbb{E}[y]] = \log(\frac{1}{\theta}) = \beta_0 + \beta_1 x_i$

Maps $\mathbb{R}^+$ onto $\mathbb{R}$ $\quad \log \theta = \beta_0 + \beta_1 x_i$

$\theta = \exp[\beta_0 + \beta_1 x_i]$

- Look at hazard function or median to see if drug is working.

Weibull Distribution

1. Show it is a distribution

y positive, θ positive, exp is positive $\Rightarrow$ positive function

Must show it integrates to 1.

Could be Hand.

15/03/16 DW

5

3/02/16 ALSM 2

$\int_0^\infty \lambda\theta y^{\lambda-1} \exp[-\theta y^\lambda]\, dy$

$= [-\exp[-\theta y^\lambda]]_0^\infty$ at 0=1 at ∞=0

$1+0 = 1$

Probability of failure: $1 - \exp[-\theta y^\lambda]$

Probability of survival: $1 - F(y)$

Hazard function $h(y) = \frac{p(y|\lambda\theta)}{S(y)} = \lambda\theta y^{\lambda-1}$ (accelerated failure time, function of y)

Exponential case $\lambda = 1$ $h(y)$ is $\theta \Rightarrow$ not a function of y

↳ Independent of age "Chance of survival same at 100 as at 1000"

↳ lack of memory → y not important, may be a limitation

2. Show it is a member of exponential family of distributions

$= \lambda\theta y^{\lambda-1} \exp[-\theta y^\lambda]$

$= \exp[-\theta y^\lambda] \exp[\log(\lambda\theta y^{\lambda-1})]$

$= \exp[-\theta y^\lambda \log(\lambda) + \log(\theta) + \log(y^{\lambda-1})]$

$b(\theta)\ a(y) \quad c(\theta) \quad d(y)$

3. Expectation

$\mathbb{E}[y] = \int_0^\infty y\lambda\theta y^{\lambda-1} \exp[-\theta y^\lambda]\, dy$

Show $\mathbb{E}[y] = (\frac{1}{\theta})^{1/\lambda}\, \Gamma(1 + 1/\lambda)$ with $\Gamma(u) = \int_0^\infty s^{u-1} \exp(-s)\, ds$

Substitution: $\mu = \theta y^\lambda \quad d\mu = \theta\lambda y^{\lambda-1} dy$

$= \int_0^\infty y\lambda\mu \exp(-\mu)\, d\mu$

$\mathbb{E}[y] = \int_0^\infty (\frac{u}{\theta})^{1/\lambda} \exp(-\mu)$

$\log(\theta) = -3.07 - 1.731x$ x is indicator for drug

The log rescales the β values: 2.48 + 1.267 (R output numbers)