Chapter 3: Parametric Survival Analysis

Chien-Fu Jeff Lin, MD. PhD.

cflin@mail.ntpu.edu.tw

http://web.ntpu.edu.tw/~cflin

Department of Statistics
National Taipei University, Taipei, Taiwan

Spring 2007

Outline I

Important Parametric Distributions I

1. Many parametric distributions are used in survival analysis.
2. These distributions are chosen, not only because of their popularity among researchers, but also because they offer insight into the nature of the various parameters and hazard functions.
3. Some parametric families are easily written in the form

$$Y = \log(X) = \mu + \sigma W, \tag{0.1}$$

where W has a specific fixed distribution. This is useful.

Outline

Outline

1. Some Common Lifetime Distribution
   - Exponential Distribution – One Parameter
   - Weibull Distribution – Two Parameters
   - Generalized Gamma Distribution – Three Parameters
   - Log-normal Distribution
   - Log-logistic Distribution
   - Extreme Value Distribution – Gumbel Distribution
   - Standardized Extreme Value (Minimum) Distribution
2. Estimation
3. Parametric Examples

Exponential Distribution – One Parameter I

Let X be a failure time random variable and follow exponential distribution as

$$h(x) = \lambda \tag{1.1}$$

$$H(x) = \int_0^x h(t)dt = \lambda x \tag{1.2}$$

$$S(x) = \exp[-\lambda x] \tag{1.3}$$

$$f(x) = \lambda \exp[-\lambda x] \tag{1.4}$$

$$F(x) = 1 - S(x) = 1 - \exp[-\lambda x] \tag{1.5}$$

Exponential Distribution – One Parameter I

$$\mathcal{E}(x) = \frac{1}{\lambda} \quad (1.6)$$

$$\mathbf{Var}(X) = \frac{1}{\lambda^2} \quad (1.7)$$

$$\mathbf{mrl}(x) = \frac{1}{\lambda} \quad (1.8)$$

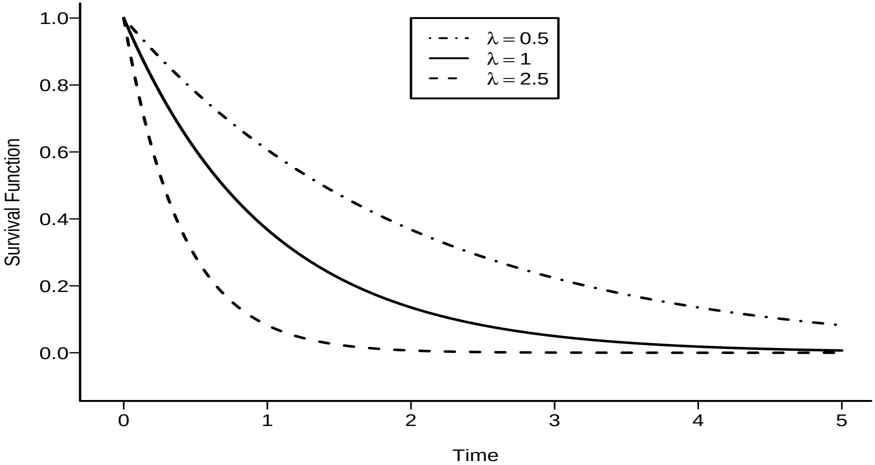

Figure 1: Exponential Survival Function with Different Hazard Rates

Exponential Distribution – One Parameter I

1. Cumulative hazard $H(x) = \int_0^x h(t)dt = \lambda x$ is linear through the origin.
2. The death rate of most lifetime does change with time.
3. So it is not a good model for long-interval time of human survival function, but it is good for short-interval time.

Exponential Distribution with Trasformation I

1. Let $Y = \log(X)$ then

$$
\begin{aligned}
S_Y &= Pr(Y > y) = Pr(\log X > y) = Pr(X > e^y) && (1.9)\\
&= S_X(e^y) && (1.10)\\
&= \exp[-\lambda e^y] && (1.11)\\
f_Y(y) &= -\frac{d}{dy} S_Y(y) && (1.12)\\
&= \lambda e^y \exp[-\lambda e^y] && (1.13)
\end{aligned}
$$

Exponential Distribution with Trasformation II

2. Let $\mu = -\log\ \lambda$, then

$$\lambda e^{y} = e^{\log\lambda}\ e^{y} = e^{y+\log\lambda} \tag{1.14}$$

$$= e^{y-\mu} \tag{1.15}$$

$$S_Y(y) = \exp[-e^{y-\mu}] \tag{1.16}$$

$$f_Y(y) = \exp[y - \mu - e^{y-\mu}] \tag{1.17}$$

Exponential Distribution with Trasformation III

3. Let $w = y - \mu$ then

$$f_W(w) = \exp[w - e^w] \tag{1.18}$$

4. This is not important in data analysis, but it is useful in regression modelling.

Exponential Distribution with Trasformation IV

5. For example, let $Y = \log(X), \mu = -\log \lambda, w = y - \mu,$

$$-\exp[w - \mu] = -\lambda x \tag{1.19}$$

$$-\exp[w] \times \exp[-\mu] = \lambda x \tag{1.20}$$

Exponential Distribution with SAS I

1. In SAS Output, μ is intercept, such that

$$\lambda = \exp[-\mu] = \exp[-\text{intercept}] \tag{1.21}$$

2. Calculate $t_{0.5}$, 50% survival time as

$$S(t_{0.5}) = \exp[-\lambda t_{0.5}] = 0.5 \tag{1.22}$$

$$t_{0.5} = -\frac{\log 0.5}{\lambda} = \frac{\log 2}{\lambda} \tag{1.23}$$

Outline

1. Some Common Lifetime Distribution
   - Exponential Distribution – One Parameter
   - Weibull Distribution – Two Parameters
   - Generalized Gamma Distribution – Three Parameters
   - Log-normal Distribution
   - Log-logistic Distribution
   - Extreme Value Distribution – Gumbel Distribution
   - Standardized Extreme Value (Minimum) Distribution
2. Estimation
3. Parametric Examples

Weibull Distribution – Two Parameters I

Let X be a failure time random variable and follow Weibull distribution and let $\theta > 0$ and $\alpha > 0$, and $\Gamma(\alpha) = (\alpha - 1)!$, or $\Gamma(k) = \int_0^\infty u^{k-1} e^{-u}\, du$, then

$$h(x) = \theta\alpha(\theta x)^{\alpha-1} \tag{1.24}$$

$$H(x) = (\theta x)^{\alpha} \tag{1.25}$$

$$S(x) = \exp[-(\theta x)^{\alpha}] \tag{1.26}$$

$$f(x) = \theta\alpha(\theta x)^{\alpha-1} \exp[-(\theta x)^{\alpha}] \tag{1.27}$$

$$F(x) = 1 - \exp[-(\theta x)^{\alpha}] \tag{1.28}$$

Weibull Distribution – Two Parameters I

$$\mathcal{E}(x) = \theta^{-1}\Gamma\left[\left(1 + \frac{1}{\alpha}\right)\right] \tag{1.29}$$

$$\textbf{Var}(x) = \theta^{-2}\left\{\Gamma\left(1 + \frac{2}{\alpha}\right) - \left[\Gamma\left(1 + \frac{1}{\alpha}\right)\right]^2\right\} \tag{1.30}$$

Weibull Distribution – Two Parameters I

α is called "shape" parameter such that

$$h'(x) > 0 \text{ , if } \alpha > 1 \quad (1.31)$$
$$h'(x) < 0 \text{ , if } \alpha < 1$$
$$h(x) = \theta \text{ , if } \alpha = 1\text{, as constant hazard rate, i.e. exponential distribution}$$

Weibull Distribution – Two Parameters I

1. Weibull distribution is a two-parameter distribution and is the most commonly used distribution.
2. Hazard can increase or decrease.
3. Limiting distribution for minimum of sample from a distribution, making it conceptually plausible as time to first of several failure mechanism.
4. Log cumulative hazard

$$\log(H(x)) = \log(-\log(S(x))) = \alpha\log(\theta) + \alpha\log(x), \quad (1.32)$$

is linear in $\log(x)$.

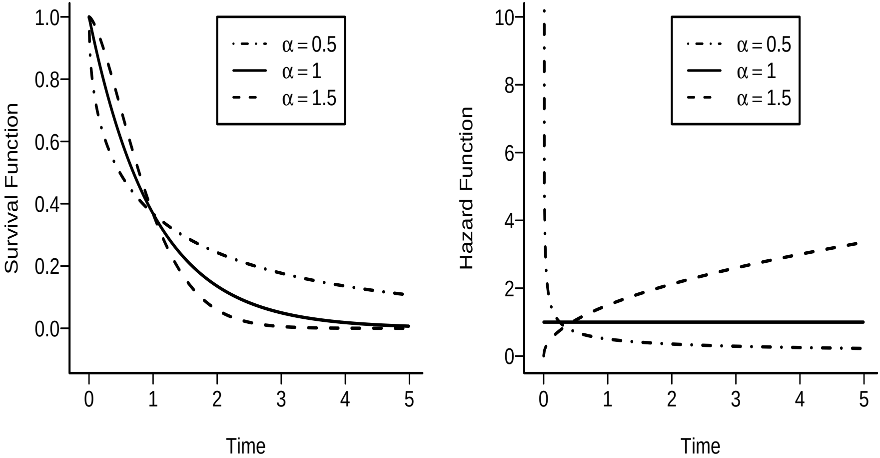

Figure 2: Weibull Survival Function and Hazard Function

Weibull Distribution: Transformation I

1. if $y = g(x)$ is one to one transformation, then

$$f_Y(y) = \left| \frac{d}{dy} g^{-1}(y) \right| f_X(g^{-1}(y)). \tag{1.33}$$

Weibull Distribution: Transformation II

2. Let $y = \log(x), \mu = -log(\theta), \sigma = \beta^{-1}, w = (y-\mu)/\sigma$, then

$$f_Y(y) = (\frac{1}{\sigma}) \exp[\frac{y-\mu}{\sigma} - e^{\frac{y-\mu}{\sigma}}] \quad (1.34)$$

$$S_Y(y) = \exp[-e^{\frac{y-\mu}{\sigma}}] \quad (1.35)$$

$$f_W(w) = \exp[w - e^w] \quad (1.36)$$

$$S_w(w) = \exp[-e^w] \quad (1.37)$$

Weibull Distribution in SAS I

1. In SAS Output, $\beta = (\text{“ Scale”})^{-1}$ and $\mu =$ “Intercept”, such that

$$\beta = \sigma^{-1} = \text{scale}^{-1} \tag{1.38}$$

$$\theta = \exp[-\mu] = \exp[\text{- intercept}] \tag{1.39}$$

Weibull Distribution in SAS II

2. Calculate $t_{0.5}$, 50% survival time as

$$S(t_{0.5}) = \exp[-(\theta t_{0.5})^{\beta}] = 0.5 \quad (1.40)$$

$$\log(-\log\ t_{0.5}) = \beta\ [\log\theta + \log t_{0.5}] \quad (1.41)$$

$$\log(-\log\ t_{0.5}) - \beta\ log\ \theta = \beta\ \log\ t_{0.5} \quad (1.42)$$

$$t_{0.5} = \exp\left[\frac{\log(\log 2)}{\beta} - \log\ \theta\right] \quad (1.43)$$

Weibull Distribution: II I

One form of parameterization of Weibull distribution as in some textbooks

$$h(x) = \alpha\lambda x^{\alpha-1} \tag{1.44}$$

$$S(x) = \exp[-\lambda x^{\alpha}] \tag{1.45}$$

$$f(x) = \alpha\lambda x^{\alpha-1}\exp[-\lambda x^{\alpha}] \tag{1.46}$$

where $\lambda = \theta^{\beta}$, and in SAS Output, $\alpha = \sigma^{-1}$ and $\lambda = \exp[-\mu\sigma]$.

Weibull Distribution: III I

Another form of parameterization for Weibull distribution is as in other textbooks as

$$h(x) = \lambda \alpha x^{\alpha - 1} \tag{1.47}$$

$$H(x) = \lambda x^{\beta} \tag{1.48}$$

$$S(x) = \exp[-\lambda x^{\alpha}] \tag{1.49}$$

$$f(x) = \lambda \beta x^{\alpha - 1} \exp[-\lambda x^{\beta}] \tag{1.50}$$

$$F(x) = 1 - \exp[-\lambda x^{\beta}] \tag{1.51}$$

$$\mathcal{E}(X) = \left(\lambda^{1/\alpha}\right)^{-1} \Gamma\left[\left(1 + \frac{1}{\alpha}\right)\right] \tag{1.52}$$

$$\textbf{Var}(X) = (\lambda^{1/\alpha})^{-2} \left\{ \Gamma\left(1 + \frac{2}{\alpha}\right) - \left[\Gamma\left(1 + \frac{1}{\alpha}\right)\right]^2 \right\} \tag{1.53}$$

Weibull Distribution: III II

(1.54)

Where in SAS Output, $\alpha = 1/\sigma$, σ is the scale parameter, $\lambda = \exp[-(\frac{\mu}{\sigma})]$, μ is “intercept”.

Weibull Distribution – Two Parameters in R I

In R, the Weibull distribution (of the probability functions) is defined as

$$h(x) = \left(\frac{1}{b}\right) a \left(\frac{1}{b} x\right)^{a-1} \tag{1.55}$$

$$S(x) = \exp\left[- \left(\frac{1}{b}\right)^{a} (x)^{a} \right] \tag{1.56}$$

$$f(x) = \left(\frac{a}{b}\right) \left(\frac{x}{b}\right)^{(a-1)} \exp\left(- \left(\frac{x}{b}\right)^{a} \right). \tag{1.57}$$

Outline

Generalized Gamma Distribution: Three Parameters I

1. A variable X is said to follow the generalized Gamma distribution as

$$f(x) = \frac{\beta}{\Gamma(k)} \frac{x^{\beta k-1}}{\alpha^{\beta k}} \exp\left[-(\frac{x}{\alpha})^{\beta}\right] \tag{1.58}$$

2. Suppose $Y = \log X$, $u = log\ \alpha$ and $b = \beta^{-1}$, then reparameterization as log Gamma distribution as

$$W_1 = \frac{Y-u}{b} \tag{1.59}$$

$$f(w_1) = \frac{1}{\gamma(k)} \exp[kw_1 - e^{w_1}] \tag{1.60}$$

Generalized Gamma Distribution: Three Parameters II

3. Another transformation for Generalized Gamma Distribution is

$$W = \sqrt{k}(W_1 - \log k) \tag{1.61}$$

$$= \frac{Y - (u + b\log k)}{b\sqrt{k}} \tag{1.62}$$

$$= \frac{Y - \mu}{\sigma} \tag{1.63}$$

$$f(w,k) = \frac{k^{k-1/2}}{\Gamma(k)} \exp[\sqrt{k}w - ke^{w/\sqrt{k}}] \tag{1.64}$$

where

$$\sigma = \frac{b}{\sqrt{k}}, \ \mu = u + b\log k \tag{1.65}$$

Generalized Gamma Distribution: SAS I

1. In SAS Output,

$$\text{Intercept} = \mu = u + b \log k = \log \alpha + \frac{\log k}{\beta} \tag{1.66}$$

$$\text{Scale} = \sigma = b/\sqrt{k} = \frac{1}{\beta\sqrt{k}} \tag{1.67}$$

$$\text{Shape1} = \delta = \frac{1}{\sqrt{k}} \tag{1.68}$$

2. if "shape" $= 1$, then $k = 1$, $f(x)$ is Weibull distribution,
3. if "shape" $= 0$, $k = \infty$, $f(x)$ is log-normal distribution.

Generalized Gamma Distribution I

Another parameterization as in the textbook as

$$f(x) = \frac{\alpha\lambda^{\beta} x^{\alpha\beta-1} \exp[-\lambda x^{\alpha}]}{\Gamma(\beta)} \tag{1.69}$$

or

$$f(x) = \frac{\eta}{\Gamma(k)}(\theta^{k}) x^{\eta k-1} \exp[-\theta x^{\eta}] \tag{1.70}$$

Generalized Gamma Distribution I

Note: the transformation of parameters in different equations is

(1.69)	(1.70)	(1.58);	relationship (1.69) and SAS
α	η	β_*	$\text{scale} = \dfrac{1}{\alpha\sqrt{\beta}}$
β	k	k	$\text{shape1} = 1/\sqrt{\beta}$
λ	θ	α_*^{-k}	$\text{intercept} = -\dfrac{\log\lambda}{\beta} + \dfrac{\log\beta}{\alpha}$

Outline

Log-normal Distribution I

1. A variable X is said to follow the log-normal distribution if $Y = \ln X$ follows normal distribution as

$$y = \log X \quad \sim N(\mu, \sigma^2) \tag{1.71}$$

$$S(x) = 1 - \Phi\left(\frac{\ln x - \mu}{\sigma}\right) \tag{1.72}$$

2. Log-normal distribution is a very useful if there is no censoring.

Outline

1. Some Common Lifetime Distribution
   - Exponential Distribution – One Parameter
   - Weibull Distribution – Two Parameters
   - Generalized Gamma Distribution – Three Parameters
   - Log-normal Distribution
   - Log-logistic Distribution
   - Extreme Value Distribution – Gumbel Distribution
   - Standardized Extreme Value (Minimum) Distribution
2. Estimation
3. Parametric Examples

Log-logistic Distribution I

1. A variable X is said to follow the log-logistic distribution if $Y = \ln X$ follows logistic distribution as

$$f(y) = \frac{\exp[\frac{y-\mu}{\sigma}]}{\sigma[1 + \exp(\frac{y-\mu}{\sigma})]^2} \tag{1.73}$$

whereμ and σ are, respectively, the mean and scale parameter of Y.

Log-logistic Distribution II

2. The hazard rate and survival function for log logistic distribution can be written as

$$h(x) = \frac{\alpha\lambda x^{\alpha-1}}{1+\lambda x^{\alpha}} \quad (1.74)$$

$$S(x) = \frac{1}{1+\lambda x^{\alpha}} \quad (1.75)$$

where $\alpha = 1/\sigma$ and $\lambda = \exp(-\mu/\sigma)$.

Outline

1. Some Common Lifetime Distribution
   - Exponential Distribution – One Parameter
   - Weibull Distribution – Two Parameters
   - Generalized Gamma Distribution – Three Parameters
   - Log-normal Distribution
   - Log-logistic Distribution
   - Extreme Value Distribution – Gumbel Distribution
   - Standardized Extreme Value (Minimum) Distribution
2. Estimation
3. Parametric Examples

Extreme Value Distribution – Gumbel Distribution I

1. Let X be a random variable with Gumbel distribution and the probability density and survival function are

$$f(x) = b^{-1} \exp\left[\frac{x-u}{b} - \exp\left(\frac{x-u}{b}\right)\right], \; -\infty < x < \infty \quad (1.76)$$

$$S(x) = \exp\left[-\exp\left(\frac{x-u}{b}\right)\right] \quad (1.77)$$

where $b > 0$ and $u(-\infty < u < \infty)$ are parameters. u is a location and b is a scale parameter.

2. The extreme value distribution with $u = 0$ and $b = 1$ is termed the standard extreme value distribution.

Outline

Extreme Value (Minimum) Distribution I

The standardized extreme value distribution for a random variable W as defined as

$$h(w) = \exp(w) \tag{1.78}$$

$$H(w) = \exp(w) \tag{1.79}$$

$$S(w) = \exp(-e^{w}) \tag{1.80}$$

$$f(w) = \exp(w - e^{w}) \tag{1.81}$$

$$F(w) = 1 - \exp(-e^{w}). \tag{1.82}$$

Extreme Value (Minimum) Distribution I

The standardized extreme value distribution is a distribution or which there exists a distribution function G, such that if $x_1, \ldots, x_n$,, for all $x_i > 0$. are i.i.d. each with distribution G, then $\min(x_1, \ldots, x_n)$, properly normalized, converges in distribution F.

Outline

Outline

1. Some Common Lifetime Distribution

2. Estimation
   - 概似函數 Likelihood Function
   - 最大概似估計式 Maximum Likelihood Estimation
   - Newton-Raphson 計算法
   - 參數假設檢定 Testing Hypothesis for Parameters
   - 參數信賴區間 Confidence Intervals for Parameters
   - Delta Method Approximation for Variance Function

3. Parametric Examples

4. Parametric Regression Models for Survival Analysis

Likelihood Function I

1. 存活分析一個重要的特別情況為右設限 (right censoring) 的隨機設限 (random censoring).
2. 令存活時間隨機變數 X 具有機率密度函數 f_X, 累積機率分配函數 F_X, 存活函數 S_C 與設限時間隨機變數 C, 具有機率密度函數 g_C, 存活函數 G_C.
3. 令 X, C 的聯合機率密度函數為 $f_{X,C}(x, c)$
4. 令觀察資料代表成對的隨機變數為 (T, δ), 使得 $T = \min(X, C)$, $\delta = I(T = X)$.
5. 則其概似函數為:

$$L(T, \delta) = \left[Pr(t = X \leq C)\right]^{\delta} \left[Pr(t = C < X)\right]^{1-\delta} \tag{2.1}$$

$$= \left[\int_t^{\infty} f_{X,C}(t, v)\, dv\right]^{\delta} \left[\int_t^{\infty} f_{X,C}(u, t)\, du\right]^{1-\delta}. \tag{2.2}$$

Likelihood Function II

6. 假設 T 與 C 為獨立, 則

$$L(T,\delta) = \left[\int_t^\infty f_X(t)f_C(v)\,dv\right]^\delta \left[\int_t^\infty f_X(u)f_C(t)\,du\right]^{1-\delta} \tag{2.3}$$
$$= \left[f_X(t)^\delta S_X(t)^{1-\delta}\right]\left[f_C(t)^{1-\delta} S_C(t)^\delta\right]. \tag{2.4}$$

7. 假設 f_X 具有參數 $\underline{\boldsymbol{\theta}}$, 且 f_C 具有參數 $\underline{\boldsymbol{\alpha}}$, 則

$$L(\underline{\boldsymbol{\theta}},\underline{\boldsymbol{\alpha}},T,\delta) = \left[f_X(t,\underline{\boldsymbol{\theta}})^\delta S_X(t,\underline{\boldsymbol{\theta}})^{1-\delta}\right]\left[f_C(t,\underline{\boldsymbol{\alpha}})^{1-\delta} S_C(t,\underline{\boldsymbol{\alpha}})^\delta\right]. \tag{2.5}$$

Likelihood Function III

8. 實際上我們對 $\underline{\boldsymbol{\theta}}$ 比較有興趣, 所以我們的對數概似函數可以寫爲:

$$\begin{aligned}
\ell &= \log\ L(\underline{\boldsymbol{\theta}}, \underline{\boldsymbol{\alpha}}, T, \delta) && (2.6)\\
&= \{\delta\ \log\left[f_X(t, \underline{\boldsymbol{\theta}})\right] + (1-\delta)\ \log\left[S_X(t, \underline{\boldsymbol{\theta}})\right]\}\\
&\quad + \{(1-\delta)\ \log\left[f_C(t, \underline{\boldsymbol{\alpha}})\right] + \delta\ \log\left[S_C(t, \underline{\boldsymbol{\alpha}})\right]\} && (2.7)\\
&= \{\text{log-likelihood}\}_X(\underline{\boldsymbol{\theta}}, t, \delta) + \{\text{log-likelihood}\}_C(\underline{\boldsymbol{\alpha}}, t, \delta) && (2.8)\\
&= \ell_X\ (\underline{\boldsymbol{\theta}}, t, \delta) + \ell_C\ (\underline{\boldsymbol{\alpha}}, t, \delta) && (2.9)
\end{aligned}$$

Likelihood Function IV

9. 若設限時間是 "無訊息的設限" (non-informative censoring), 則 $\ell_C(\underline{\boldsymbol{\alpha}}, t, \delta)$ 不包含 $\underline{\boldsymbol{\theta}}$ 參數, 因此

$$\ell = \ell_X(\underline{\boldsymbol{\theta}}, t, \delta) + c. \quad (\text{對 } \underline{\boldsymbol{\theta}} \text{ 而言, c 是一個常數}) \tag{2.10}$$

10. X 與 C 獨立是上述隨機概似函數之充分條件, 不是必要條件, 例如, 在接近死亡的情況下 C 與 X 相依.

11. 舉一個 **訊息設限 (informative censoring)** 的例子, 令 X 爲右眼失明的時間, C 爲左眼失明的時間, 則 $S_X(\underline{\boldsymbol{\theta}}, t) = S_C(\underline{\boldsymbol{\theta}}, t)$ 是合理的. 先前的概似函數 c 不再是常數, 使得 C 爲 X 的 **有訊息的設限 (informative censoring)**.

Likelihood Function V

12. 令 X_i 與 C_i 爲獨立的事件時間與設限時間之隨機變數, 且具有機率密度函數 f_X 與 g_C,
13. 令成對的隨機樣本 (T_i, δ_i), $i = 1, 2, \ldots, n$, 互相獨立且來自相同分配, 如下

$$T_i = \min(X_i, C_i) = X_i \wedge C_i \tag{2.11}$$

$$\delta_i = I(X_i \le C_i) = \begin{cases} 1 & \text{若 } X_i \le C_i\text{, 也就是 } T_i \text{ 不是設限時間,} \\ 0 & \text{若 } X_i > C_i\text{, 也就是 } T_i \text{ 是設限時間.} \end{cases} \tag{2.12}$$

Likelihood Function VI

14. 概似函數方程式 (2.17) 在計算與程式設計上爲一個良好式子.

$$
\begin{aligned}
L(\underline{\mathbf{t}}, \underline{\boldsymbol{\delta}}) &= \prod_{1}^{n} L(t_i, \delta_i) && (2.13) \\
&= \prod_{i=1}^{n} f_X(t_i)^{\delta_i} S_X(T_i)^{1-\delta_i} && (2.14) \\
&= \prod_{i=1}^{n} [h_X(t_i) S_X(t_i)]^{\delta_i} S_X(t_i)^{1-\delta_i} && (2.15) \\
&= \left[\prod_{\text{uncensoring}} f_X(t_i) \right] \left[\prod_{\text{censoring}} S_X(t_i) \right] && (2.16) \\
&= \prod_{i=1}^{n} [h_X(t_i)]^{\delta_i} S_X(t_i). && (2.17)
\end{aligned}
$$

Outline

1. Some Common Lifetime Distribution

2. Estimation
   - 概似函數 Likelihood Function
   - 最大概似估計式 Maximum Likelihood Estimation
   - Newton-Raphson 計算法
   - 參數假設檢定 Testing Hypothesis for Parameters
   - 參數信賴區間 Confidence Intervals for Parameters
   - Delta Method Approximation for Variance Function

3. Parametric Examples

4. Parametric Regression Models for Survival Analysis

計算最大概似估計式 (MLE) I

1. 令一個 $p \times 1$ 行向量的參數且概似函數爲

$$
\begin{aligned}
\underline{\boldsymbol{\theta}}_{p\times 1} &= (\underline{\boldsymbol{\theta}}, \underline{\boldsymbol{\alpha}}, \ldots, \theta_p)^T \\
\boldsymbol{L}(\underline{\boldsymbol{\theta}}) &= \prod_{i=1}^{n} L_i(\underline{\boldsymbol{\theta}}, t_i, \delta_i) \\
\boldsymbol{\ell}(\underline{\boldsymbol{\theta}}) &= \log\ L(\underline{\boldsymbol{\theta}}) && (2.18) \\
&= \sum_{i=1}^{n} \log L_i(\underline{\boldsymbol{\theta}}, t_i, \delta_i) && (2.19) \\
&= \sum_{i=1}^{n} \ell_i(\underline{\boldsymbol{\theta}}, t_i, \delta_i) && (2.20)
\end{aligned}
$$

計算最大概似估計式 (MLE) II

2. 尋找 $L(\underline{\boldsymbol{\theta}})$ 的最大概似函數值, 與尋找對數概似函數方程式同時成立的 $\widehat{\underline{\boldsymbol{\theta}}}$ 解是相同的
3. 例如令對數概似函數方程式對 $\underline{\boldsymbol{\theta}}$ 一階微分

$$0 = \frac{\partial}{\partial\theta_j}\boldsymbol{\ell}(\underline{\boldsymbol{\theta}}) \tag{2.21}$$

$$= \sum_{i=1}^{n} \frac{\partial}{\partial\theta_j}\ell_i(\underline{\boldsymbol{\theta}}, t_i, \delta_i)\ , \quad j = 1, 2, \ldots, p. \tag{2.22}$$

4. 這通常需要在電腦上使用迭代 (iteration) 的計算方法.

Outline

1. Some Common Lifetime Distribution

2. Estimation
   - 概似函數 Likelihood Function
   - 最大概似估計式 Maximum Likelihood Estimation
   - Newton-Raphson 計算法
   - 參數假設檢定 Testing Hypothesis for Parameters
   - 參數信賴區間 Confidence Intervals for Parameters
   - Delta Method Approximation for Variance Function

3. Parametric Examples

4. Parametric Regression Models for Survival Analysis

定義概似函數 I

1. 定義概似函數符號爲

$$\ell_i(\underline{\boldsymbol{\theta}}) = \log\ L_i(\underline{\boldsymbol{\theta}}, t_i, \delta_i),\ i = 1, 2, \ldots, n;$$

分數函數與資訊函數 I

1. 且定義對數概似函數方程式對 $\underline{\boldsymbol{\theta}}$ 一階微分為 **分數函數 (score function)** 與二階微分為 **資訊函數 (information matrix)** 如下

$$\frac{\partial}{\partial \theta_1}\boldsymbol{\ell}(\underline{\boldsymbol{\theta}}) = \left(\frac{\partial}{\partial \underline{\boldsymbol{\theta}}}\boldsymbol{\ell}(\underline{\boldsymbol{\theta}}), \cdots, \frac{\partial}{\partial \theta_p}\boldsymbol{\ell}(\underline{\boldsymbol{\theta}})\right)^T, \tag{2.23}$$

$$\frac{\partial^2}{\partial \underline{\boldsymbol{\theta}}^2}\boldsymbol{\ell}(\underline{\boldsymbol{\theta}}) = \begin{pmatrix} \frac{\partial^2}{\partial\theta_1\partial\theta_1}\boldsymbol{\ell}(\underline{\boldsymbol{\theta}}) & \cdots & \frac{\partial^2}{\partial\theta_1\partial\theta_p}\boldsymbol{\ell}(\underline{\boldsymbol{\theta}}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial\theta_p\partial\theta_1}\boldsymbol{\ell}(\underline{\boldsymbol{\theta}}) & \cdots & \frac{\partial^2}{\partial\theta_p\partial\theta_p}\boldsymbol{\ell}(\underline{\boldsymbol{\theta}}) \end{pmatrix} \tag{2.24}$$

對數概似函數方程式 I

1. 對數概似函數方程式爲

$$0 = \sum_{i=1}^{n} \frac{\partial}{\partial \theta_j} \ell_i(\underline{\boldsymbol{\theta}}), \; j = 1, 2, \ldots, p; \tag{2.25}$$

$$\text{或是} \; \underline{\mathbf{0}} = \frac{\partial}{\partial \underline{\boldsymbol{\theta}}} \boldsymbol{\ell}(\underline{\boldsymbol{\theta}}). \tag{2.26}$$

Newton-Raphson 計算法 I

1. 假設 $\widehat{\underline{\boldsymbol{\theta}}}^0 = (\widehat{\theta}_1^0, \cdots, \widehat{\theta}_p^0)^T$ 為這個概似估計解最開始的推測值
2. 對 $\widehat{\underline{\boldsymbol{\theta}}}^0$ 作泰勒展開如下

$$\underline{\mathbf{0}} = \sum_{i=1}^n \frac{\partial}{\partial \theta_j} \ell_i(\widehat{\underline{\boldsymbol{\theta}}})$$

$$0 = \sum_{i=1}^n \frac{\partial}{\partial \theta_j} \ell_i(\widehat{\underline{\boldsymbol{\theta}}}^0) + \sum_{k=1}^p (\widehat{\theta}_k - \widehat{\theta}_k^0) \left\{ \sum_{i=1}^n \frac{\partial^2}{\partial \theta_k \partial \theta_j} \ell_i(\widehat{\underline{\boldsymbol{\theta}}}^0) \right\} + \cdots, \quad j = 1, 2$$

$$\underline{\mathbf{0}} = \frac{\partial}{\partial \underline{\boldsymbol{\theta}}} \boldsymbol{\ell}(\widehat{\underline{\boldsymbol{\theta}}})$$

$$= \frac{\partial}{\partial \underline{\boldsymbol{\theta}}} \boldsymbol{\ell}(\widehat{\underline{\boldsymbol{\theta}}}^0) + \frac{\partial^2}{\partial \underline{\boldsymbol{\theta}}^2} \boldsymbol{\ell}(\widehat{\underline{\boldsymbol{\theta}}}^0)(\widehat{\underline{\boldsymbol{\theta}}} - \widehat{\underline{\boldsymbol{\theta}}}^0) + \cdots$$

Newton-Raphson 計算法 II

3. 令 $\underline{\widehat{\boldsymbol{\theta}}}^1$ 爲忽略第二階微分與其他更高階項的對數概似函數方程式之解:

$$\underline{\widehat{\boldsymbol{\theta}}}^1 = \underline{\widehat{\boldsymbol{\theta}}}^0 + \left[-\frac{\partial^2}{\partial \underline{\boldsymbol{\theta}}^2} \ell(\underline{\widehat{\boldsymbol{\theta}}}^0) \right]^{-1} \left[\frac{\partial}{\partial \underline{\boldsymbol{\theta}}} \ell(\underline{\widehat{\boldsymbol{\theta}}}^0) \right] \tag{2.31}$$

此稱爲 Newton-Raphson **計算法** 以使用樣本的資訊矩陣爲主之計算法.

費雪分數計算法 I

1. 在此定義

$$\text{分數向量} = \mathfrak{U}(\underline{\boldsymbol{\theta}}) = \frac{\partial}{\partial \underline{\boldsymbol{\theta}}} \ell(\underline{\boldsymbol{\theta}}); \tag{2.32}$$

$$\text{樣本資訊矩陣} = \mathfrak{I}(\underline{\boldsymbol{\theta}}) = \left[-\frac{\partial^2}{\partial \underline{\boldsymbol{\theta}}^2} \ell(\underline{\boldsymbol{\theta}}) \right]; \tag{2.33}$$

$$\text{費雪資訊矩陣} = \boldsymbol{I}(\underline{\boldsymbol{\theta}}) = \mathcal{E}[\mathfrak{I}(\underline{\boldsymbol{\theta}})] = \mathcal{E}\left[-\frac{\partial^2}{\partial \underline{\boldsymbol{\theta}}^2} \ell(\underline{\boldsymbol{\theta}}) \right]. \tag{2.34}$$

費雪分數計算法 II

2. 若在式子 (2.31) 中使用 **費雪資訊矩陣 ((Fisher information matrix))** 取代樣本資訊矩陣如下

$$\underline{\widehat{\boldsymbol{\theta}}}^1 = \underline{\widehat{\boldsymbol{\theta}}}^0 + \boldsymbol{I}^{-1}(\underline{\boldsymbol{\theta}}^0)\left[\frac{\partial}{\partial \underline{\widehat{\boldsymbol{\theta}}}}\ell(\underline{\widehat{\boldsymbol{\theta}}}^0)\right] \quad (2.35)$$

$$= \underline{\widehat{\boldsymbol{\theta}}}^0 + \boldsymbol{I}^{-1}(\underline{\widehat{\boldsymbol{\theta}}}^0)\ \mathcal{U}(\underline{\widehat{\boldsymbol{\theta}}}^0) \quad (2.36)$$

3. 使用式子 (2.36) 進行迭代計算法 (iteration) 的機制則稱爲 **分數計算法 (method of scoring).**
4. 式子 (2.36) 在某些情形可能使收斂效果更好
5. 但在某些情形, 特別是有設限情形存在時, 可能不會收斂.

Outline

假設檢定 I

關於參數的假設檢定, 有三種形式的檢定問題:

1. 對所有參數值的檢定 (overall parameters),
2. 單一參數值的檢定 (test the value of a single parameter),
3. 和一群部分參數值的檢定 (test for contribution of a group parameters).

假設檢定 II

4. 在大樣本理論下, 普遍使用的假設檢定統計量, 有三種形式,
   - Wald 統計量 (**Wald statistics**),
   - 概似比檢定統計量 (**likelihood ratio test statistics**),
   - 和 分數統計量 (**score statistics**).

假設檢定 III

5. Wald 統計量是建立在參數的估計量的近似常態之下,
6. 概似比統計量是建立在對兩巢狀模式 (nested models) 的對數概似函數比值,
7. 分數統計量是建立在分數函數的近似常態分配.

分解參數向量 I

1. 一般來說, 我們通常對某些特定參數所形成相關的子集合的假設檢定有興趣,
2. 我們可以先分解參數向量爲 $\underline{\boldsymbol{\theta}}_{p\times 1} = (\underline{\boldsymbol{\theta}}_1^T, \underline{\boldsymbol{\theta}}_2^T)^T$ 其中 $\underline{\boldsymbol{\theta}}_1$ 是有興趣的 $q \times 1$ 參數向量,
3. 且 $\underline{\boldsymbol{\theta}}_2$ 是其他的 $(p - q) \times 1$ 參數向量,

分解參數向量 II

4

5 我們將資訊矩陣分解爲 $q \times q$ 和 $(p-q) \times (p-q)$ 列行資訊矩陣, 如下:

$$\mathfrak{I}(\underline{\boldsymbol{\theta}}) = [\mathfrak{I}_{rs}(\underline{\boldsymbol{\theta}})]_{(p\times p)} = \left(\frac{-\partial^2 \log L}{\partial\boldsymbol{\beta}_r \partial\boldsymbol{\beta}_s}\right)_{(p\times p)} = \begin{pmatrix} \mathfrak{I}_{11}(\underline{\boldsymbol{\theta}}) & \mathfrak{I}_{12}(\underline{\boldsymbol{\theta}}) \\ \mathfrak{I}_{21}(\underline{\boldsymbol{\theta}}) & \mathfrak{I}_{22}(\underline{\boldsymbol{\theta}}) \end{pmatrix} \quad (2.37)$$

6 其中 $\mathfrak{I}_{11}(\underline{\boldsymbol{\theta}})_{(q\times q)}$ 爲 $(q \times q)$ 部分資訊矩陣, $\mathfrak{I}_{22}(\underline{\boldsymbol{\theta}})_{(p-q)\times(p-q)}$ 爲 $(p-q) \times (p-q)$ 部分資訊矩陣.

分解參數向量 III

7. 令上式反矩陣的分解爲

$$\mathfrak{J}^{-1}(\underline{\boldsymbol{\theta}}) = \begin{pmatrix} \mathfrak{J}^{11}(\underline{\boldsymbol{\theta}}) & \mathfrak{J}^{12}(\underline{\boldsymbol{\theta}}) \\ \mathfrak{J}^{21}(\underline{\boldsymbol{\theta}}) & \mathfrak{J}^{22}(\underline{\boldsymbol{\theta}}) \end{pmatrix}. \tag{2.38}$$

分解參數向量 IV

8. 令 $\underline{\hat{\boldsymbol{\theta}}}_{(p\times 1)} = (\underline{\hat{\boldsymbol{\theta}}}_1^T, \underline{\hat{\boldsymbol{\theta}}}_2^T)^T$ 爲 $\underline{\boldsymbol{\theta}}$ 參數的最大概似估計式向量之分解,
9. 其中 $\underline{\hat{\boldsymbol{\theta}}}_1$ 爲 $q \times 1$ 向量, $\underline{\hat{\boldsymbol{\theta}}}_2$ 爲 $(p-q) \times 1$ 向量.
10. 考慮虛無假設 $H_0 : \underline{\boldsymbol{\theta}}_1 = \underline{\boldsymbol{\theta}}_{01}$ 的檢定.

(1) Wald 檢定統計量 I

1. Wald 檢定統計量 (**Wald statistic**) 形式爲

$$X_W^2 = (\hat{\underline{\theta}}_1 - \underline{\theta}_{01})^T \left[\mathfrak{I}^{11}(\hat{\underline{\theta}})\right]^{-1} (\hat{\underline{\theta}}_1 - \underline{\theta}_{01}) \tag{2.39}$$

2. 此統計量在計算可逆的部分資訊矩陣, $[\mathfrak{I}^{11}(\hat{\underline{\theta}})]^{-1}$, 須依賴所有的最大概似估計式向量, $\hat{\underline{\theta}}$.

(2) 概似比檢定統計量 LRT I

1. 概似比檢定統計量 (**likelihood ratio test statistic, LRT**) 爲

$$X_{LR}^2 = 2\ \{\ \log \boldsymbol{L}(\hat{\underline{\boldsymbol{\theta}}}) - \log \boldsymbol{L}(\underline{\boldsymbol{\theta}}_{01}, \hat{\underline{\boldsymbol{\theta}}}_2(\underline{\boldsymbol{\theta}}_{01}))\ \} \tag{2.40}$$

2. 其中 $\hat{\underline{\boldsymbol{\theta}}}_2(\underline{\boldsymbol{\theta}}_{01})$ 是指在給定 $\underline{\boldsymbol{\theta}}_1$ 固定於虛無假設的值 $\underline{\boldsymbol{\theta}}_{01}$ 之下, $\underline{\boldsymbol{\theta}}_2$ 的最大偏概似估計.

(3) Rao 分數檢定統計量 I

1. Rao 分數檢定統計量 (**Rao score test statistic**) 爲

$$X^2_{SC} = \mathfrak{U}_1[\underline{\boldsymbol{\theta}}_{01}, \hat{\underline{\boldsymbol{\theta}}}_2(\underline{\boldsymbol{\theta}}_{01})]^T \, [\mathfrak{I}^{11}[\underline{\boldsymbol{\theta}}_{01}, \hat{\underline{\boldsymbol{\theta}}}_2(\underline{\boldsymbol{\theta}}_{01})] \, \mathfrak{U}_1[\underline{\boldsymbol{\theta}}_{01}, \hat{\underline{\boldsymbol{\theta}}}_2(\underline{\boldsymbol{\theta}}_{01})] \quad (2.41)$$

2. 其中 $\mathfrak{U}_1[\underline{\boldsymbol{\theta}}_{01}, \hat{\underline{\boldsymbol{\theta}}}_2(\underline{\boldsymbol{\theta}}_{01})]$ 是對 $\underline{\boldsymbol{\theta}}_1$ 的 $(q \times 1)$ 分數函數向量, $\mathfrak{U}_1$ 之計算是在 $\underline{\boldsymbol{\theta}}_{01}$ 爲虛無假設檢定爲眞的值, 以及 $\underline{\boldsymbol{\theta}}_2$ 受限於 $\underline{\boldsymbol{\theta}}_{01}$ 之下的偏最大概似估計值計算而出.

大樣本理論 I

1. 當樣本之個體數目趨近無趣窮大
2. 以及不同死亡時間點之數目趨近無趣窮大,
3. 且當虛無假設爲眞時,
4. 此三個統計量近似卡方分配, 自由度爲 q, q 是虛無假設與對立假設兩模型間參數個數的差.

$$X^2_W,\ X^2_{LR},\ X^2_{SC} \sim \chi^2_q. \tag{2.42}$$

檢定統計量比較 I

1. Wald 是最保守的檢定, 且當樣本數很小時, 會有怪異的變化.
2. Rao 分數檢定的方法不要求 MLE, 只要在 H_0 下模型合適就好, 並且它的迭代 (iteration) 趨近於 MLE.
3. 在一個特定的參數模型下, 參數的轉換 (transformation) 的概似檢定 (LRT) 與分數檢定是不變的.
4. 在設限的狀況下, 因為 $\boldsymbol{I}(\underline{\boldsymbol{\theta}})$ 的計算方式困難且概似函數可能不會收斂, 我們也許常需要以 $\mathfrak{I}(\underline{\boldsymbol{\theta}})$ 代替 $\boldsymbol{I}(\underline{\boldsymbol{\theta}})$.

Outline

參數信賴區間 I

1. 大樣本下, 合理的近似推論爲

$$\underline{\hat{\theta}} \sim N(\,\underline{\theta}, \widehat{\mathbf{Var}}(\underline{\hat{\theta}})\,), \quad \text{或} \quad \underline{\hat{\theta}} \sim N(\,\underline{\theta}, [\,\mathfrak{I}(\underline{\hat{\theta}})\,]^{-1}\,); \qquad (2.43)$$

2. 其中 $\widehat{\mathbf{Var}}(\underline{\hat{\theta}}) = [\,\mathfrak{I}(\underline{\hat{\theta}})\,]^{-1}$ 爲 **樣本資訊矩陣.** $\mathfrak{I}(\underline{\hat{\theta}})$ 來自於牛頓演算法中, 迭代最後步驟的結過, 參數信賴區間如參數假設檢定, 有不同的計算方式, 常見的式根據上節所述的參數假設檢定.

Wald 信賴區間 I

1. 我們可藉此對單一共變數參數數進行檢定虛無假設 $H_0 : \theta_j = 0$, 以單一樣本近似 Z 檢定統計量

$$z = \frac{\hat{\theta}_j}{\text{s.e.}(\hat{\theta}_j)} \sim N(0, 1), \tag{2.44}$$

2. 其中 $s.e.(\hat{\theta}_j)$ 是 $\widehat{\mathbf{Var}}(\underline{\hat{\boldsymbol{\theta}}})$ 的第 (j, j) 元素的平方根,
3. 這是利用上節所述, 對單一共變數係數進行名爲 Wald **檢定**. 利用
4. ald 檢定可依建構單一共變數參數, θ_j, 的近似 $(1-\alpha) \times 100\%$ 信賴區間爲

$$\hat{\theta}_j \pm Z_{1-\alpha/2}\, \text{s.e.}(\hat{\theta}_j) \tag{2.45}$$

Wald 信賴區間 II

5. 對多個共變數之參數 $q \times 1$ 向量, $\underline{\boldsymbol{\theta}}_1$, 之近似 $(1-\alpha) \times 100\%$ 信賴區間, 可由上式 Wald 檢定推廣,
6. 回顧 Wald 檢定爲 $X_W^2 = (\hat{\underline{\boldsymbol{\theta}}}_1 - \underline{\boldsymbol{\theta}}_{01})^T \, [\mathfrak{I}^{11}(\hat{\underline{\boldsymbol{\theta}}})]^{-1} \, (\hat{\underline{\boldsymbol{\theta}}}_1 - \underline{\boldsymbol{\theta}}_{01})$
7. 因此 $\hat{\underline{\boldsymbol{\theta}}}_1$ 之近似 $(1-\alpha) \times 100\%$ 信賴區間, $\hat{\underline{\boldsymbol{\theta}}}_1$ 須滿足下列不等式

$$\{(\hat{\underline{\boldsymbol{\theta}}}_1 - \underline{\boldsymbol{\theta}}_{01})^T \, [\mathfrak{I}^{11}(\hat{\underline{\boldsymbol{\theta}}})]^{-1} \, (\hat{\underline{\boldsymbol{\theta}}}_1 - \underline{\boldsymbol{\theta}}_{01})\} \; < \; \chi^2_{q,(1-\alpha)}. \qquad (2.46)$$

概似比信賴區間 I

1. 回想概似比檢定統計量為 $X^2_{LR} = 2\,\{\,\log \boldsymbol{L}(\hat{\underline{\boldsymbol{\theta}}}) - \log \boldsymbol{L}(\underline{\boldsymbol{\theta}}_{01}, \hat{\underline{\boldsymbol{\theta}}}_2(\underline{\boldsymbol{\theta}}_{01}))\,\}$,
2. 其中 $\hat{\underline{\boldsymbol{\theta}}}_2(\underline{\boldsymbol{\theta}}_{01})$ 是指在 $\underline{\boldsymbol{\theta}}_1$ 固定於 $\underline{\boldsymbol{\theta}}_{01}$ 之下, $\underline{\boldsymbol{\theta}}_2$ 的最大概似估計.
3. 令 **profile likelihood (側面概似函數)** 為

$$\log \boldsymbol{L}(\underline{\boldsymbol{\theta}}_1, \hat{\underline{\boldsymbol{\theta}}}_2(\underline{\boldsymbol{\theta}}_1)), \tag{2.47}$$

4. 將上式視為 $\underline{\boldsymbol{\theta}}_1$ 的函數.
5. 對每個 $\underline{\boldsymbol{\theta}}_1$ 而言, 此函數提供限制 $\underline{\boldsymbol{\theta}} = [\underline{\boldsymbol{\theta}}_1, \hat{\underline{\boldsymbol{\theta}}}_2(\underline{\boldsymbol{\theta}}_1)]^T$ 下的個體的最大概似函數,

概似比信賴區間 II

6. 以 profile 概似比爲基礎的信賴區間對於下式而言是 $\underline{\boldsymbol{\theta}}_1$ 的集合滿足下列不等式

$$2\left\{\log \mathbf{L}(\hat{\underline{\boldsymbol{\theta}}}) - \log \mathbf{L}(\underline{\boldsymbol{\theta}}_{01}, \hat{\underline{\boldsymbol{\theta}}}_2(\underline{\boldsymbol{\theta}}_{01}))\right\} < \chi^2_{q,(1-\alpha)}. \tag{2.48}$$

7. 此包含了名義 α 大小之下概似比檢定中未拒絕虛無假設的 $\underline{\boldsymbol{\theta}}_1$.

分數信賴區間 I

1. 回想分數檢定統計量爲
$X^2_{SC} = \mathfrak{U}_1[\underline{\theta}_{01}, \hat{\underline{\theta}}_2(\underline{\theta}_{01})]^T\ [\mathfrak{I}^{11}[\underline{\theta}, \hat{\underline{\theta}}_2(\underline{\theta}_{01})]\ \mathfrak{U}_1[\underline{\theta}_{01}, \hat{\underline{\theta}}_2(\underline{\theta}_{01})];$
2. 其中 $\mathfrak{U}_1[\underline{\theta}_{01}, \hat{\underline{\theta}}_2(\underline{\theta}_{01})]$ 是對 $\underline{\theta}_1$ 的 $(q \times 1)$ 分數函數向量, $\mathfrak{U}_1$ 之計算是 $\underline{\theta}_1$ 在虛無假設檢定爲眞的值 $\underline{\theta}_{01}$, 以及 $\underline{\theta}_2$ 受限於 $\underline{\theta}_{01}$ 之下的最大概似估計值計算而出.
3. 以分數檢定爲基礎的信賴區間對於下式而言是 $\underline{\theta}_1$ 的集合, 滿足下列不等式

$$\mathfrak{U}_1[\underline{\theta}_1, \hat{\underline{\theta}}_2(\underline{\theta}_1)]^T\ [\mathfrak{I}^{11}[\underline{\theta}_1, \hat{\underline{\theta}}_2(\underline{\theta}_1)]\ \mathfrak{U}_1[\underline{\theta}_1, \hat{\underline{\theta}}_2(\underline{\theta}_1)] < \chi^2_{q,(1-\alpha)}. \quad (2.49)$$

信賴區間之比較 I

1. 以 profile 概似比與分數檢定爲基礎的的信賴區間在表達式中並非封閉的形式, 而是決定在數值方法,
2. 當 $\hat{\underline{\theta}}$ 服從常態分配時, 對數概似函數會有拋物線的形狀, 換言之, 二階多項式.

信賴區間之比較 II

3. 對於類別資料的小樣本而言, $\underline{\hat{\theta}}$ 可能會遠離常態分配, 以及對數概似函數可能會偏離對稱拋物線形狀的曲線.
4. 當模型包含很多參數時, 這也可能發生在適當之大樣本上,
5. 在此狀況中, 建立在 $\underline{\hat{\theta}}$ 的近似常態基礎下的推論也許會有不適當地呈現.
6. Wald 和概似比檢定的推論結果有顯著地發散時, 這指出 $\underline{\hat{\theta}}$ 的分配也許並不接近常態,
7. 在很多的案例中, 推論以利用精確小樣本分配或 "高階" 近似的方法替代, 這可改善樣本常態性的問題 (Pierce 與 Peters, 1992).

信賴區間之比較 III

8. 實際上常用的是 Wald 信賴區間, 因為使用 *ML* 估計量架構簡單, 且統計軟體會算出標準差.
9. 以概似比為基礎的信賴區間, 現今已出現在較多之應用統計軟體上, 並且在類別資料中, 或較少的 n 樣本數, 概似比為基礎的信賴區間更適合.
10. 對於最常見的統計模型, 迴歸模型若具常態分配的參數, 此三種推論必須提供相同的結果.

Outline

1. Some Common Lifetime Distribution

2. Estimation
   - 概似函數 Likelihood Function
   - 最大概似估計式 Maximum Likelihood Estimation
   - Newton-Raphson 計算法
   - 參數假設檢定 Testing Hypothesis for Parameters
   - 參數信賴區間 Confidence Intervals for Parameters
   - Delta Method Approximation for Variance Function

3. Parametric Examples

4. Parametric Regression Models for Survival Analysis

Delta Method Approximation for Variance I

1. Suppose that random variable X has mean μ and variance σ^2, and suppose we want the mean and variance of some function $g(X)$.
2. We can expand $g(X)$ about μ

$$g(X) = g(\mu) + (X - \mu)g'(\mu) + \cdots \tag{2.50}$$

and ignore higher order terms to get

$$\mathcal{E}[g(X)] = \mathcal{E}[g(\mu)] \approx g(\mu) \tag{2.51}$$

$$\begin{aligned} \textbf{Var}[g(X)] &\approx \mathcal{E}\Big[g(X) - \mathcal{E}[g(u)]\Big]^2 \\ &\approx \textbf{Var}\Big[(X-\mu)g'(\mu)\Big] \\ &\approx \sigma^2\Big[g'(\mu)^2\Big] \end{aligned} \tag{2.52}$$

Delta Method Approximation for Variance II

3. For multivariate version, suppose X and Y are multivariate random variables with mean μ_x, μ_y and variance σ_x^2, σ_y^2, covariance σ_{xy}, then

$$\begin{aligned} g(X, Y) &= g(\mu_x, \mu_y) \\ &\quad + (X - \mu_x)\frac{\partial}{\partial x} g(\mu_x, \mu_y) + (Y - \mu_y)\frac{\partial}{\partial y} g(\mu_x, \mu_y) + \cdots \end{aligned} \tag{2.53}$$

4. Mean of $g(X, Y)$:

$$\mathcal{E}[g(X, Y)] \approx g(\mu_x, \mu_y) \tag{2.54}$$

Delta Method Approximation for Variance III

5. Variance of $g(X, Y)$

$$\begin{aligned}
\textbf{Var}[g(X,Y)] &\approx \mathcal{E}\Big[g(X,Y) - g(\mu_x,\mu_y)\Big]^2 \\
&\approx \mathcal{E}\Big[(X-\mu_x)\frac{\partial}{\partial x}g(\mu_x,\mu_y) + (Y-\mu_y)\frac{\partial}{\partial y}g(\mu_x,\mu_y)\Big]^2 \\
&\approx \quad \sigma_x^2\Big[\frac{\partial}{\partial x}g(\mu_x,\mu_y)\Big]^2 \\
&\quad + \quad 2 \quad \sigma_{xy}\Big[\frac{\partial}{\partial x}g(\mu_x,\mu_y)\frac{\partial}{\partial y}g(\mu_x,\mu_y)\Big] \\
&\quad + \quad \sigma_y^2\Big[\frac{\partial}{\partial y}g(\mu_x,\mu_y)\Big]^2 \qquad (2.55)
\end{aligned}$$

Outline

Example 3.1

Rat Survival with Carcinogen DMBA

A data set gives times from insult with the carcinogen DMBA to mortality from cancer in rats. Two groups were distinguished by a pretreatment regimen. (kalbfleisch and Prentice 2nd ed.)

Outline

Example: Exponential Distribution I

1. Under random censoring, random sample of pairs (T_i, δ_i), $T_i = min(X_i, C_i)$, $\delta_i = \mathrm{I}(T_i = X_i)$, $i = 1, 2, \ldots, n$, X_i are i.i.d. with exponential distribution of lifetime.

$$
\begin{aligned}
L(T, \delta) &= \prod_{i=1}^{n} [h(t_i)]^{\delta_i} S(T_i) && (3.1)\\
&= \prod_{i=1}^{n} \lambda^{\delta_i} \exp[-\lambda_i T_i] && (3.2)\\
\ell(T, \delta) &= \log\ L(T, \delta_i) && (3.3)\\
&= \sum_{1}^{n} (\delta_i \log\ \lambda - \lambda T_i) && (3.4)\\
&= \log\ \sum \delta_i - \lambda \sum T_i && (3.5)
\end{aligned}
$$

Example: Exponential Distribution II

2. Find the MLE of λ, let

$$\begin{aligned} 0 &= \frac{\partial}{\partial\lambda}\ell(T,\delta) && (3.6) \\ &= \frac{1}{\lambda}\sum\delta_i - \sum T_i && (3.7) \\ \widehat{\lambda}_{\text{MLE}} &= \frac{\sum\delta_i}{\sum T_i} && (3.8) \\ \widehat{\lambda}_{\text{MLE}} &= \frac{\text{Total \# of events}}{\text{Total time on study}} && (3.9) \end{aligned}$$

Example: Exponential Distribution III

3. Find the variance of $\widehat{\lambda}$ as

$$\frac{\partial^2}{\partial^2\lambda}\ell(T,\delta) = -\frac{\sum\delta_i}{\lambda^2} \tag{3.10}$$

$$\mathfrak{I}(\widehat{\lambda}) = \frac{\sum\delta_i}{\lambda^2} \tag{3.11}$$

$$\widehat{\lambda} \sim \text{asym } N(\lambda, \mathfrak{I}^{-1}(\lambda)) \tag{3.12}$$

$$\widehat{\lambda} \sim \text{asym } N\Big(\lambda, \frac{\lambda^2}{\sum\delta_I}\Big) \tag{3.13}$$

$$\frac{\widehat{\lambda}-\lambda}{\sqrt{\mathfrak{I}^{-1}}} \sim \text{asym } N(0,1) \tag{3.14}$$

Example: Exponential Distribution IV

4. Find $(1-\alpha) \times 100\%$ C. I. for λ as

$$\widehat{\lambda} \pm Z_{1-\alpha/2} s.e.(\widehat{\lambda}) \tag{3.15}$$

$$= \widehat{\lambda} \pm Z_{1-\alpha/2} \sqrt{\frac{\widehat{\lambda}^2}{\sum \delta_i}} \tag{3.16}$$

Example: Exponential Distribution V

5. Approximation with $Z = \log(\lambda)$ with

$$z = g(\lambda) \tag{3.17}$$

$$g' = \frac{1}{\lambda} \tag{3.18}$$

$$\sigma_z^2 = \sigma^2(\widehat{\lambda})[g'(\widehat{\lambda})]^2 \tag{3.19}$$

$$= \frac{\widehat{\lambda}^2}{\sum \delta_i} \frac{1}{\widehat{\lambda}^2} \tag{3.20}$$

Example: Exponential Distribution VI

6. Inference about $z = \log(\widehat{\lambda})$ and C.I. for $\log(\widehat{\lambda})$ as

$$\log(\widehat{\lambda}) \pm Z_{1-2/\alpha} s.e.(\log(\widehat{\lambda})) \tag{3.21}$$

$$= log(\widehat{\lambda}) \pm Z_{1-2/\alpha} \sqrt{\frac{1}{\sum \delta_i}} \tag{3.22}$$

$1 - \alpha$ C.I. for λ is

$$\exp\left[\log(\hat{\lambda}) \pm Z_{1-2/\alpha} \sqrt{\frac{1}{\sum \delta_i}}\right] \tag{3.23}$$

$$\left[\widehat{\lambda} e^{-Z_{1-2/\alpha} \sqrt{\frac{1}{\sum \delta_i}}}, \quad \widehat{\lambda} e^{+Z_{1-2/\alpha} \sqrt{\frac{1}{\sum \delta_i}}}\right] \tag{3.24}$$

Example: Exponential Distribution VII

7. Compare (3.16), (3.24), which one is more accuracy?
8. Consider repeating sampling, if repeating sampling does not depend on true value, more accuracy of inference in terms of size of hypothesis testing.
9. Variance of $\log \lambda$ does not depend on the unknown parameter λ.
10. It is an empirical fact that transforming an estimate to remove the dependence of the variance on the unknown tends to improve the convergence to normality by reducing the skewness.

Example SAS: Exponential Distribution I

1. Note: for random variable X exponential distribution as $h(x) = \lambda$ in previous section 5, as in SAS Output, μ is the intercept of exponential distribution, such that

$$
\begin{aligned}
Y &= \log(X) && (3.25)\\
\mu &= \log \lambda && (3.26)\\
\lambda &= \exp[-\mu] = \exp[-\text{intercept}] && (3.27)
\end{aligned}
$$

Example SAS: Exponential Distribution II

2. Calculate $t_{0.5}$, 50% survival time as

$$S(t_{0.5}) = \exp[-\lambda t_{0.5}] = 0.5 \tag{3.28}$$

$$t_{0.5} = -\frac{\log 0.5}{\lambda} = \frac{\log 2}{\lambda} \tag{3.29}$$

Example 3.2

DMBA Data with Exponential Distribution

For DMBA data, we first fit a exponential model, the results are shown in Table **??**.

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	5.5345	0.1667	5.2078	5.8611	1102.70	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		
Weibull Scale	1	253.2778	42.2130	182.6965	351.1268		
Weibull Shape	0	1.0000	0.0000	1.0000	1.0000		

DMBA: Exponential Distribution I

1. The intercept is 5.5345 and the standard error is 0.1667.
2. The hazard rate is $\hat{\lambda} = \exp[-\hat{\mu}] = \exp[-\text{intercept}] = \exp[-5.5345] = 0.003984$.
3. The mean survival time is $\mathcal{E}(T) = \frac{1}{\hat{\lambda}} = \frac{1}{0.003984} = 253$ days.
4. The medial survival time is $t_{0.5} = -\frac{\log(0.5)}{\hat{\lambda}} = 176$, and the one year survival rate is $S(365) = \exp[-0.003984 \times 365] = 0.24$.

DMBA: Exponential Distribution II

5. For DMBA data, we have $\sum \delta_i = 36$, $\sqrt{\frac{1}{\sum \delta_i}} = 0.16666$.
6. Figure 3 compares the kaplan-Meier estiamtes and exponential fit, it shows the unreasonable fit of exponential model.

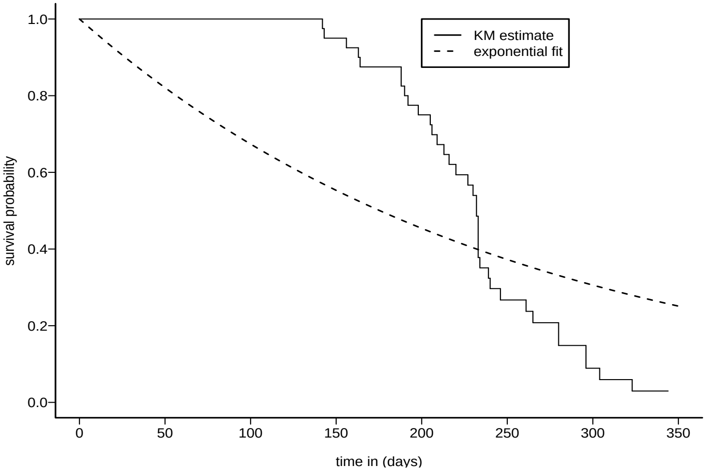

Figure 3: Kaplan-Meier Survival Curve and Exponential Survival Curve for DMBA Data

Outline

Example: Weibull Distribution I

1. For random variable X with Weibull distribution as $h(x) = \theta\beta(\theta x)^{\beta-1}$ in previous 5 section,
2. In SAS Output, $\beta =$ " Scale"$^{-1}$, "Intercept" = μ such that

$$\beta = \sigma^{-1} = \text{scale}^{-1} \tag{3.30}$$

$$\theta = \exp[-\mu] = \exp[\text{- intercept}] \tag{3.31}$$

Example: Weibull Distribution II

3. Calculate $x_{0.5}$, 50% survival time as

$$
\begin{aligned}
S(x_{0.5}) &= \exp[-(\theta x_{0.5})^{\beta}] = 0.5 && (3.32)\\
\log(-\log\ x_{0.5}) &= \beta\ [\log\theta + \log x_{0.5}] && (3.33)\\
\log(-\log\ x_{0.5}) - \beta\ log\ \theta &= \beta\ \log\ x_{0.5} && (3.34)\\
x_{0.5} &= \exp\left[\frac{\log(\log 2)}{\beta} - \log\ \theta\right] && (3.35)
\end{aligned}
$$

Example: Weibull Distribution III

4. C.I. for x_p, pth percentile of failure time from Weibull distribution.

$$p = Pr(X < x_p) \tag{3.36}$$

$$1 - p = S_X(x_p) \tag{3.37}$$

$$Y = \log(X) \tag{3.38}$$

$$1 - p = Pr(X > x_p) = Pr[\log(X) > \log(x_p)] \tag{3.39}$$

$$= Pr(y > y_p) \tag{3.40}$$

Example: Weibull Distribution IV

$$\text{Let } y = \mu + \sigma w, \text{ where } S_W(w) = \exp[-\exp(w)] \quad (3.41)$$

$$1 - p = S_W(w_p) \quad (3.42)$$

$$w_p = \log[\log(1 - p)] \quad (3.43)$$

$$1 - p = Pr(W > w_p) \quad (3.44)$$

$$= Pr(\frac{y - \mu}{\sigma} > w_p) \quad (3.45)$$

$$= Pr(y > \mu + \sigma w_p) \quad (3.46)$$

$$y_p = \mu + \sigma w_p \quad (3.47)$$

Example: Weibull Distribution V

$$\text{MLE: } \widehat{y}_p = \widehat{\mu} + \hat{\sigma} w_p, \quad \text{where} \quad w_p = \log[\log(1-p)] \tag{3.48}$$

$$\textbf{Var}(\widehat{y}_p) = \textbf{Var}(\widehat{\mu}) + 2\, w_p \,\text{Cov}\,(\widehat{\mu}, \widehat{\sigma}) + \; w_p^2 \textbf{Var}\,(\widehat{\sigma}) \tag{3.49}$$

Example: Weibull Distribution VI

5. We now are interested in x_p, using Delta method

$$\mathbf{Var}(\widehat{x}_p) = \mathbf{Var}(\widehat{y}_p)[e^{\widehat{y}_p}]^2 \tag{3.50}$$

$$(1-\alpha)\times 100\% \text{ C.I. of } x_p: \qquad \widehat{x}_p \pm Z_{1-\alpha/2}\sqrt{\mathbf{Var}(\widehat{x}_p)} \tag{3.51}$$

6. Another (better) way to get $(1-\alpha)\times 100\%$ C.I. of $\widehat{x}_p$ is

$$\exp[\widehat{y}_p \pm Z_{1-\alpha/2} s.e.(\widehat{y}_p)] \tag{3.52}$$

7. Consider $\widehat{S}(x)$ and $\mathbf{Var}(\widehat{S}(x))$ for exercises.

Example 3.3

DMBA data with Weibull Distribution

For DMBA data, we then fit a exponential model, the results are shown in Table **??**.

Analysis of Parameter Estimates

Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	5.5267	0.0340	5.4601	5.5932	26490.6	<.0001
Scale	1	0.1973	0.0244	0.1548	0.2515		
Weibull Scale	1	251.3079	8.5335	235.1271	268.6023		
Weibull Shape	1	5.0683	0.6274	3.9764	6.4599		

DMBA: Weibull Distribution I

1. Compute

$$\hat{\beta} = \frac{1}{\hat{\sigma}} = \frac{1}{0.1973} = 5.068.$$
$$\hat{\lambda} = \exp[-\hat{\mu}] = \exp[-5.5526] = 0.00398,$$
$$S(365) = \exp\left[-(\lambda t)^{\beta}\right] = 0.001,$$
$$S(t_{0.5}) = \exp\left[-(\lambda t_{0.5})^{\beta}\right] = 0.5,$$
$$t_{0.5} = 234.$$

DMBA: Weibull Distribution II

2. To compute the 95% C.I for $t_{0.5}$, we need $\hat{\textbf{Var}}(\hat{\sigma}) = 0.000597$, $\hat{\textbf{Var}}(\hat{\mu}) = 0.001153$ and $\hat{\textbf{Cov}}(\hat{\sigma}, \hat{\mu}) = -0.000207$.

DMBA: Weibull Distribution III

3. Now,

$$\begin{aligned}
p &= 0.5, w_{0.5} = \log(-\log 0.5) = -0.36651, \\
\hat{y}_{0.5} &= 5.454363, \\
t_{0.5} &= \exp[\hat{y}] = 233.776. \\
\hat{\textbf{Var}}(\hat{y}_{0.5}) &= (0.36651)^2(0.000597) + 2(-0.36651)(-0.000207) \\
&\quad +0.001153 = 0.00138493 \\
\text{s.e.}(\hat{y}_{0.5}) &= 0.037215.
\end{aligned}$$

4. Then 95% C.I. for
$\hat{y}_{0.5} = 55454363 \pm (1.96)(0.037215) = (5.38, 5.53)$, the 95% C.I. for $t_{0.5} = (e^{5.38}, e^{5.53}) = (217.3, 251.5)$.

Outline

1. Some Common Lifetime Distribution
2. Estimation
3. Parametric Examples
4. Parametric Regression Models for Survival Analysis

Regression Models with Covariates I

1. The distribution of lifetimes, X, may vary with covariates $\underline{\mathbf{Z}}' = (Z_1, Z_2, \ldots, Z_p)$.
2. More generally, the covariates can change with time, $\underline{\mathbf{Z}}'(x) = (Z_1(x), Z_2(x), \ldots, Z_p(x))$.
3. The change can differ for each member of the population, which really makes things interesting.

Two Most commonly Used Models I

There are two commonly used models for the relationship between survival times and covariates:

1. Accelerated failure time models (log linear in time)
2. Conditional hazard rate as a function of covariates
    1. Multiplicative hazard rate models (Proportional hazard models (log linear in the hazard)).
    2. Additive hazard rate models

Outline

1. Some Common Lifetime Distribution
2. Estimation
3. Parametric Examples
4. Parametric Regression Models for Survival Analysis
    - Proportional (Multiplicative) Hazard Model
    - Additive Hazard Model
    - Accelerated Failure Time (ATF) Model
    - ATF Model and Hazard Function
    - Weibull Distribution and proportional Hazard Model
    - Example: ATF Model with Weibull Distribution

Proportional Hazards Models I

1. We call $h_0(t)$ the baseline hazard function.
2. One particular useful type of proportional hazards model is **log-linear in the hazard function**:

$$g(\underline{\mathbf{Z}}) = \exp[\underline{\boldsymbol{\beta}}'\underline{\mathbf{Z}}] \tag{4.1}$$

$$\text{where } \underline{\boldsymbol{\beta}}'\underline{\mathbf{Z}} = \sum_{j=1}^{p} \beta_j z_j \tag{4.2}$$

3. e^{β_j} in a proportional hazards regression model is interpretable as a **relative event rate** (**relative risk**, or **risk ratio**) for covariate Z_j.

Proportional Hazards Models II

4. Then, the hazard ratio for covariates $\underline{\mathbf{Z}} = \underline{\mathbf{Z}}_1$ and $\underline{\mathbf{Z}} = \underline{\mathbf{Z}}_2$ is

$$\frac{h(x \mid Z_1)}{h(x \mid Z_2)} = \frac{g(\underline{\beta}'\underline{\mathbf{Z}}_1)}{g(\underline{\beta}'\underline{\mathbf{Z}}_2)} \tag{4.3}$$

is independent of time.

Proportional Hazards Models III

5. Proportional hazards models are easily interpreted in terms of the survival function since

$$S(t \mid \underline{\mathbf{Z}}) = \exp[-\int_0^t h(u \mid \underline{\mathbf{Z}}) du] \quad (4.4)$$

$$\text{implies} \quad S(t \mid \underline{\mathbf{Z}}) = S_0(t)^{g(\underline{\mathbf{Z}})} \quad (4.5)$$

$$S_0(t) = \exp[-\int_0^t h_0(u) du], \quad (4.6)$$

where S_0 is the baseline survival function.

Outline

Additive Hazard Models I

1. Additive Hazard models is

$$h(x \mid Z) = h_0(x) + \underline{\boldsymbol{\beta}}'\underline{\mathbf{Z}}. \tag{4.7}$$

2. We must assure $h(x \mid Z) \geq 0$.
3. Estimation for additive models is typically made by nonparametric (weighted) least-squares methods.

Outline

Accelerated Failure Time Model I

Accelerated failure time models are just location-scale models for the log survival time.

1. There can be written in the form

$$Y = \log(T) = \mu(\underline{\mathbf{z}}) + \sigma\epsilon \qquad (4.8)$$

2. One particularly useful type of accelerated failure time model is

$$\mu(\underline{\mathbf{z}}) = \underline{\boldsymbol{\beta}}'\underline{\mathbf{z}} \qquad (4.9)$$

$$\underline{\boldsymbol{\beta}}'\underline{\mathbf{z}} = \sum_{j=1}^{p} \beta_j z_j \qquad (4.10)$$

3. SAS yield estimates of β and σ for this special cases.

Accelerated Failure Time Model II

4. The survival function for T is

$$S(t \mid \underline{\mathbf{z}}) = G_\epsilon\left(\frac{\log t - \mu(\underline{\mathbf{z}})}{\sigma}\right) \quad (4.11)$$

where $G_\epsilon(\cdot)$ is the survival function for ϵ.

Accelerated Failure Time Model III

5. This can be written as

$$S(t \mid \underline{\mathbf{z}}) = S_1\left[\left(\frac{t}{\alpha(\underline{\mathbf{z}})}\right)^{\delta}\right] \tag{4.12}$$

$$\alpha(\underline{\mathbf{z}}) = \exp[\mu(\underline{\mathbf{z}})] \tag{4.13}$$

$$\delta = \frac{1}{\sigma} \tag{4.14}$$

$$S_1(w) = G_\epsilon(\log w) \tag{4.15}$$

$$S(t \mid \underline{\mathbf{z}}) = S_0\left(\frac{t}{\alpha(\underline{\mathbf{z}})}\right) \tag{4.16}$$

$$S_0 = S_1(w^{\delta}) \tag{4.17}$$

Outline

ATF Model and Hazard Function I

1. Accelerated failure time model is log-linear model.

$$Y = \mu + \underline{\boldsymbol{\gamma}}'\underline{\mathbf{Z}} + \sigma W \tag{4.18}$$

   where $\underline{\boldsymbol{\gamma}}' = (\gamma_1, \ldots, \gamma_p)'$ is a vector of regression coefficients.

2. Let $S_0(x)$ denote the survival function of $X = e^Y$ when $\underline{\mathbf{Z}}$ is zero, that is, $S_0(x)$ is the survival function of $\exp[\mu + \sigma W]$.

ATF Model and Hazard Function II

3. Now,

$$
\begin{aligned}
& Pr[\, X > x \mid \underline{\mathbf{Z}} \,] \\
= \; & Pr[\, Y > \log x \mid \underline{\mathbf{Z}} \,] && (4.19) \\
= \; & Pr[\, \mu + \sigma W > \log x - \underline{\gamma}'\underline{\mathbf{Z}} \mid \underline{\mathbf{Z}} \,] && (4.20) \\
= \; & Pr[\, \exp(\, \mu + \sigma W \,) > x \exp(\, -\underline{\gamma}'\underline{\mathbf{Z}} \,) \mid \underline{\mathbf{Z}} \,] && (4.21) \\
= \; & S_0[\, x \;\exp(\, -\underline{\gamma}'\underline{\mathbf{Z}} \,) \,] && (4.22)
\end{aligned}
$$

ATF Model and Hazard Function III

4. The hazard rate of an individual with covariate value $\underline{\mathbf{Z}}$ for this class of models is related to a baseline hazard rate $h_0(x)$ by

$$h(x \mid \underline{\mathbf{Z}}) = h_0[x \ \exp(-\underline{\boldsymbol{\gamma}}'\underline{\mathbf{Z}})\ ] \ \ \exp(-\underline{\boldsymbol{\gamma}}'\underline{\mathbf{Z}}) \tag{4.23}$$

ATF Model and Hazard Function IV

5. Thus $\underline{\mathbf{z}}$ has a multiplicative effect on survival times in accelerated failure models.
6. Notice, $T_{\alpha(\underline{\mathbf{z}})} = \alpha(\underline{\mathbf{z}})T_0$.
7. If T_0 has distribution $S_0(t)$, then the distribution of $\alpha(\underline{\mathbf{z}})T_0$ is $S_0(\frac{t}{\alpha(\underline{\mathbf{z}})})$.
8. If $\log(\alpha(\underline{\mathbf{z}})) < 0$, then $T_{\alpha(\underline{\mathbf{z}})}$ is shorter than T_0, and we say that covariate accelerates the time to failure.

ATF Model and Hazard Function V

9. Important interpretation the result of β as

$$T(\underline{\mathbf{z}}) \propto T_0 \times \exp[\underline{\boldsymbol{\beta}}'\underline{\mathbf{z}}] \tag{4.24}$$

$$\exp[\beta_1] = \frac{\exp[\beta_1(z_{11} = 1)]}{\exp[\beta_1(z_{12} = 0)]} \tag{4.25}$$

$$= \frac{\exp[\beta_1(z_{11} = 1) + \beta_2 z_2 + \cdots + \beta_p z_p]}{\exp[\beta_1(z_{10} = 0) + \beta_2 z_2 + \cdots + \beta_p z_p]} \tag{4.26}$$

10. In group z_1, if z_1 increases one level from $z_{01} = 0$ to $z_{11} = 1$, it increase lifetime by factor $\exp[\beta_1]$.
11. So it is better to do on log scale, and transfer back to T scale. $\log T$ distribution is more symmetrical.

Outline

Weibull Distribution: ATF and Proportional Hazard Model I

Weibull distribution worked well in both accelerated failure time models and proportional hazard models.

Weibull: ATF and PH Model I

$$h_0(x) = \theta\beta(\theta x)^{\beta-1} \quad (4.27)$$

$$H_0(x) = (\theta x)^{\beta} \quad (4.28)$$

$$S_0(x) = \exp[-(\theta x)^{\beta}] \quad (4.29)$$

$$h_{\underline{\mathbf{z}}}(x \mid \underline{\mathbf{z}}) = h_0(x)\exp[\underline{\mathbf{z}}\underline{\rho}] \quad \text{proportional hazards model} \quad (4.30)$$

$$S_{\underline{\mathbf{z}}}(x \mid \underline{\mathbf{z}}) = \exp\left[\left[(\theta x)^{\beta}\right]^{\exp(\underline{\mathbf{z}}\underline{\rho})}\right] \quad (4.31)$$

$$= \exp\left[\left[(\theta x)^{\beta}\right] \quad [\exp(\underline{\mathbf{z}}\underline{\rho})]\right] \quad (4.32)$$

$$= \exp\left[(\theta)^{\beta} \quad [x\exp(\frac{\underline{\mathbf{z}}\underline{\rho}}{\beta})]^{\beta}\right] \quad (4.33)$$

which is a accelerated failure time model in terms of survival function.

Outline

Example 4.1

VA Data: Parametric Regression Example

After preliminary investigations, a Weibull regression model was fitted to these data with eight regressor variables; the results are summarized in Table **??**.

Type III Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
kps	1	38.7889	<.0001
age	1	0.5085	0.4758
diagtime	1	0.0031	0.9553
prior	1	0.0428	0.8362
cellcode	3	22.0296	<.0001
treat	1	1.4959	0.2213

Analysis of Parameter Estimates

Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	2.6358	0.7089	1.2464	4.0252	13.82	0.0002
kps		1	0.0301	0.0048	0.0206	0.0395	38.79	<.0001
age		1	0.0061	0.0086	-0.0107	0.0229	0.51	0.4758
diagtime		1	-0.0005	0.0084	-0.0169	0.0159	0.00	0.9553
prior		1	-0.0044	0.0212	-0.0460	0.0372	0.04	0.8362
Cell								
squamous	1	1	0.3977	0.2547	-0.1016	0.8970	2.44	0.1185
small	2	1	-0.4285	0.2433	-0.9054	0.0484	3.10	0.0782
adeno	3	1	-0.7350	0.2741	-1.2722	-0.1979	7.19	0.0073
large	4							
treat=Stan	1	1	0.2285	0.1868	-0.1377	0.5947	1.50	0.2213
Scale		1	0.9281	0.0615	0.8150	1.0569		
Weibull Shape		1	1.0775	0.0714	0.9461	1.2270		

VA Data: Parametric Regression Example I

1. From Table **??**, a strong prognostic effect of initial performance status is indicated as is a difference among survival times in the different cell type groups.
2. This analysis would indicate, however, that patient survival does not differ significantly between treatment groups after taking account of the prognostic effect of other variables.
3. There is, as well no apparent dependence of survival time on age or disease duration prior to entry to the clinical trial.

VA Data: Parametric Regression Example II

4. Even in a randomized study such as this, it is instructive to conduct an analysis that takes account of prognostic factors. treatment comparisons that does not control for such factors, however, are also valid and typically have a useful population-averaged interpretation.

VA Data: Parametric Regression Example III

5. The maximum likelihood estimate of σ under the Weibull model is $\hat{\sigma} = 0.928$ with an estimated standard error 0.062.
6. A test of the hypothesis $\sigma = 1$ provides no evidence against the exponential model relative to the encompassing Weibull model.