Notes for Biostatistics 140.641.01 'Survival Analysis'

First Term, 2022

Chapter 2. Parametric modelling and estimation

Parametric models assume the knowledge of the probability density function (pdf) up to k unknown parameters. In this chapter, we only consider the case $k = 1$ or 2. Assume the survival time T is a 'regular' variable which has the density function $f(t; \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ is the unknown vector of parameters. Clearly, the pdf is completely specified if $\boldsymbol{\theta}$ is known. For simplicity, we sometimes write the pdf $f(t; \boldsymbol{\theta})$ as $f(t)$, and write the survival function $S(t; \boldsymbol{\theta})$ as $S(t)$.

2.1 Distributions for non-negative variables

Exponential distribution. Suppose survival time T follows $\exp(\theta)$ distribution, $\theta > 0$.

The pdf is $f(t) = \theta e^{-\theta t}$ for $t > 0$, and 0 for $t \leq 0$; or it can be more formally expressed as $f(t) = \theta e^{-\theta t} I(t > 0)$.

The survival function is

$$S(t) = \begin{cases} \int_0^\infty f(u)du = 1 & t \leq 0 \\ \int_t^\infty f(u)du = \int_t^\infty \theta e^{-\theta u} du = e^{-\theta t} & t > 0 \end{cases}$$

The hazard function is

$$\lambda(t) = \frac{f(t)}{S(t)} = \begin{cases} 0 & t \leq 0 \\ \theta & t > 0 \end{cases}$$

The expected value (mean) of T is $\mathrm{E}[T] = 1/\theta$.

Weibull distribution. The Weibull distribution with the parameters $\theta > 0$ and $\gamma > 0$ assumes the parameterized survival function

$$S(t) = e^{-(\theta t)^{\gamma}}, \qquad t > 0$$

The density function is, for $t > 0$,

$$f(t) = -\frac{dS(t)}{dt} = \gamma\theta(\theta t)^{\gamma-1} e^{-(\theta t)^{\gamma}}$$

The hazard function is, for $t > 0$,

$$\lambda(t) = \frac{f(t)}{S(t)} = \gamma\theta(\theta t)^{\gamma-1} \ .$$

Note that, for $t > 0$,

- if $\gamma = 1$ then $\lambda(t) = \theta$, a constant,
- if $\gamma > 1$ then $\lambda(t)$ is increasing in t and $\lambda(0) = 0$,
- if $0 < \gamma < 1$ then $\lambda(t)$ is decreasing in t and $\lim_{t\to 0^+} \lambda(t) = \infty$.

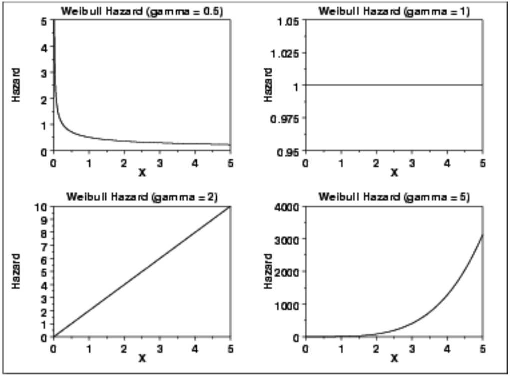

Figure: Weibull Hazard Functions (Note: $\theta = 1$; 'x' should be replaced by 't')

Log-normal distribution. A random variable T is said to have a lognormal distribution with parameters $-\infty < \mu < \infty$ and $\sigma > 0$. The probability density function of T is

$$f(t) = \frac{1}{\sigma(2\pi)^{1/2}} t^{-1} \exp\{-(log\, t - \mu)^2 / 2\sigma^2\} \ , \qquad t \geq 0 \ ,$$

from which the survival and hazard functions can be derived.

The hazard function for lognormal distribution is less interpretable as compared with the hazard function for Weibull distribution, but lognormal distribution is useful in regression analysis where $\log T$ follows the normal distribution.

2.2 Maximum Likelihood Estimation for Complete Data

Suppose there is no censoring and we observe complete survival times $t_1, t_2, \ldots, t_n$.

In general, for a parametric model $T \sim f(t;\theta)$, the likelihood function on the basis of independent and identically distributed survival times $\{t_1, \ldots, t_n\}$ is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(t_i; \boldsymbol{\theta}).$$

The maximum likelihood estimate (mle), $\hat{\theta}$, is the θ which maximizes the likelihood function $L(\theta)$.

Now we consider one-dimensional θ, i.e., the case when $\boldsymbol{\theta} = \theta$ is a real number. Note that

$$\begin{aligned} \log L(\theta) &= \sum_{i=1}^{n} \log f(t_i; \theta) \\ U(\theta) &= \frac{d}{d\theta} \log L(\theta) = \sum_{i=1}^{n} \frac{d}{d\theta} \log f(t_i; \theta) \end{aligned}$$

The mle $\hat{\theta}$ satisfies $U(\hat{\theta}) = 0$.

By Taylor's expansion,

$$
\begin{aligned}
0 &= U(\hat{\theta}) \\
&= U(\theta) + U^{(1)}(\theta)(\hat{\theta} - \theta) \ + \frac{U^{(2)}(\theta^*)}{2}(\hat{\theta} - \theta)^2 \quad (*)
\end{aligned}
$$

where $U^{(1)}$ and $U^{(2)}$ are the first and second derivative of U, θ^* lies between θ and $\hat{\theta}$. With a probability argument, the second term in (*) is ignorable and

$$
\begin{aligned}
\sqrt{n}(\hat{\theta} - \theta) &\approx -\frac{1}{U^{(1)}(\theta)}\sqrt{n}U(\theta) \\
&= -\frac{1}{\frac{1}{n}U^{(1)}(\theta)} \cdot \frac{\sqrt{n}}{n} \cdot \sum_{i=1}^{n} \frac{d}{d\theta} \log f(T_i; \theta) \\
&\xrightarrow{\mathrm{d}} \text{Normal Distribution}
\end{aligned}
$$

By statistical theory (law of large number, central limit theorem), when n is large,

$$\hat{\theta} \overset{\text{approx}}{\sim} N(\theta, I^{-1}(\theta))$$

or, alternatively and equivalently,

$$\hat{\theta} - \theta \overset{\text{approx}}{\sim} N\left(0, I^{-1}(\theta)\right) \ .$$

where $I(\theta) =$ Fisher information $= \mathsf{E}\left[-\frac{d^2}{d\theta^2} \log L(\theta)\right]$.

Example. $T \sim \exp(\theta)$. The density function is $f(t;\theta) = \theta e^{-\theta t} I(t > 0)$.

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} \theta e^{-\theta t_i} \\
\log L(\theta) &= \sum_{i=1}^{n} [\log \theta - \theta t_i] \\
U(\theta) &= \frac{d}{d\theta} \log L(\theta) = \sum_{i=1}^{n} \left[\frac{1}{\theta} - t_i \right] = \frac{n}{\theta} - \sum_{i=1}^{n} t_i
\end{aligned}
$$

Thus $\hat{\theta} = n / \sum_{i=1}^{n} t_i$ is the mle.
Note that the Fisher information is $I(\theta) = \mathsf{E}\left[-\frac{d^2}{d\theta^2} \log L(\theta) \right] = n/\theta^2$. Thus

$$\hat{\theta} - \theta \overset{\text{approx}}{\sim} N\left(0, \frac{\theta^2}{n}\right) \quad \text{when } n \text{ is large}$$

or, it is equivalently represented as

$$\hat{\theta} \overset{\text{approx}}{\sim} N\left(\theta, \frac{\theta^2}{n}\right) \ .$$

Thus Prob$\left(\hat{\theta} - 1.96\frac{\theta}{\sqrt{n}} < \theta < \hat{\theta} + 1.96\frac{\theta}{\sqrt{n}}\right) \approx 95\%$. An asymptotic 95% confidence interval for θ is

$$\left(\hat{\theta} - 1.96\frac{\hat{\theta}}{\sqrt{n}},\ \hat{\theta} + 1.96\frac{\hat{\theta}}{\sqrt{n}}\right) \ .$$

Example. (Regression extension of exponential distribution)
Let x_i be a $p \times 1$ vector of covariates and β a $p \times 1$ vector of parameters for subject i. Assume the hazard function is $\lambda(t; x_i) = \beta' x_i$. Assume T has the pdf $(\beta' x_i) e^{-(\beta' x_i) t_i}$. Based on $(x_1, t_1), \ldots, (x_n, t_n)$, the maximum likelihood techniques can still be applied to the likelihood function

$$L(\beta) = \prod_{i=1}^{n} (\beta' x_i) e^{-(\beta' x_i) t_i}$$

In contrast, an alternative modelling approach could assume that the mean of survival time is $\beta' x_i$. In this case, the likelihood function is

$$L(\beta) = \prod_{i=1}^{n} (\beta' x_i)^{-1} e^{-(\beta' x_i)^{-1} t_i}$$

2.3 Maximum Likelihood Estimation for Censored Data

Now consider (right-)censored data. The observed data include

$$(y_1, \delta_1),\ (y_2, \delta_2), \ldots, (y_n, \delta_n)$$

Note that

$$\begin{aligned} y_i &= \min(t_i, c_i) = \begin{cases} t_i & \text{uncensored case} \\ c_i & \text{censored case} \end{cases} \\ \delta_i &= I(t_i < c_i) = \begin{cases} 1 & \text{uncensored case} \\ 0 & \text{censored case} \end{cases} \end{aligned}$$

where t_i is the survival time and c_i is the censoring time.

Let $S(t;\theta) = \text{pr}(T_i \geq t), G(c) = \text{pr}(C_i \geq c)$, and let $f(t;\theta)$ and $g(c)$ be the corresponding density functions. For ease of discussion, we use $G(c^+) = \text{pr}(C_i > c)$.

Assume T_i and C_i are independent. In this case, the censoring process is called **independent censoring**. Note that subject i's contribution to the likelihood is $f(y_i;\theta)G(y_i^+)$ if $\delta_i = 1$, and $g(y_i)S(y_i;\theta)$ if $\delta_i = 0$.
The likelihood function on the basis of $(y_1, \delta_1), \ldots, (y_n, \delta_n)$ is

$$
\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} \left\{ \left[f(y_i;\theta)G(y_i^+) \right]^{\delta_i} \left[g(y_i)S(y_i;\theta) \right]^{1-\delta_i} \right\} \\
&= \prod_{i=1}^{n} \left\{ \left[f(y_i;\theta)^{\delta_i} S(y_i;\theta)^{1-\delta_i} \right] \left[g(y_i)^{1-\delta_i} G(y_i^+)^{\delta_i} \right] \right\}
\end{aligned}
$$

or simply write

$$L = \prod_{i=1}^{n} \left[f(y_i;\theta)^{\delta_i} S(y_i;\theta)^{1-\delta_i} \right] \quad (*)$$

Note that the validity of the likelihood function in $(*)$ relies on the independence between the failure and censoring times. If T_i and C_i are not independent, we then have dependent censoring (or informative censoring) since the value of C_i could have implication on the value of T_i.

Example. Suppose $T \sim \exp(\theta)$. Assume independent censoring. Based on survival data $(0.54, 0), (2.32, 1), (1.50, 1)$, the likelihood function is

$$L = (e^{-0.54\cdot\theta}) \cdot (\theta e^{-2.32\cdot\theta}) \cdot (\theta e^{-1.50\cdot\theta})$$

Example. Suppose $T \sim \exp(\theta)$. Assume independent censoring. The density function is $f(t) = \theta e^{-\theta t} I(t > 0)$ and the survival function is $S(t) = e^{-\theta t}$, $t > 0$.

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} (\theta e^{-\theta y_i})^{\delta_i} (e^{-\theta y_i})^{1-\delta_i} \\
&= \prod_{i=1}^{n} \{\theta^{\delta_i} e^{-\theta y_i}\} \\
\log L(\theta) &= \sum_{i=1}^{n} [\delta_i \log\theta - \theta y_i] \\
U(\theta) &= \frac{d}{d\theta} \log L(\theta) = \sum_{i=1}^{n} \left[\frac{\delta_i}{\theta} - y_i\right] = \frac{\sum_{i=1}^{n} \delta_i}{\theta} - \sum_{i=1}^{n} y_i
\end{aligned}
$$

Thus $\hat{\theta} = \sum_{i=1}^{n} \delta_i / \sum_{i=1}^{n} y_i$ is the mle of θ, and $\sum_{i=1}^{n} y_i / \sum_{i=1}^{n} \delta_i$ is the mle for θ^{-1} (mean of T).

Regression for censored data:

Replace θ by a meaningful link function $\phi(x_i, \beta)$, say, $\phi(x_i, \beta) = \beta' x_i$. Maximize the likelihood function

$$L = \prod_{i=1}^{n} \left[f(y_i; \phi(x_i, \beta))^{\delta_i} S(y_i; \phi(x_i, \beta))^{1-\delta_i} \right] \qquad (*)$$

to derive the mle of β.

Example. Suppose $T \sim \exp(\beta' x_i)$. Assume that conditioning on x_i, T is independent of the censoring variable C.

$$L(\theta) = \prod_{i=1}^{n} ((\beta' x_i) e^{-(\beta' x_i) y_i})^{\delta_i} (e^{-(\beta' x_i) y_i})^{1-\delta_i}$$

*Appendix

Gamma distribution. The Gamma distribution with the parameters $\lambda > 0$ and $r > 0$ is a continuous distribution with the density function

$$f(t) = \frac{\lambda^r}{\Gamma(r)} t^{r-1} e^{-\lambda t} \ ,$$

for $t \geq 0$, where $\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} dx$. The survival and hazard functions can be derived from the density function. The mean of the Gamma distribution is r/λ and the variance is r/λ^2.

Pareto distribution. The Pareto distribution assumes the form of the hazard function:

$$\lambda(t) = \frac{\alpha}{t} \cdot I(c \leq t < \infty), \quad \alpha > 0, \quad c > 0$$

This implies

$$S(t) = \left(\frac{c}{t}\right)^{\alpha} \cdot I(c \leq t < \infty).$$

The hazard function of the Pareto distribution allows for a delayed time of risk and is decreasing over time.

Log-logistic distribution. The Log-logistic distribution with the parameters $\alpha > 0$ and $-\infty < \theta < \infty$ is a continuous distribution and has the hazard function

$$\lambda(t) = \frac{e^{\theta}\alpha t^{\alpha-1}}{1 + e^{\theta}t^{\alpha}} .$$

The hazard function decreases monotonically if $0 < \alpha \leq 1$. The hazard function has a single mode if $\alpha > 1$. The Log-logistic distribution is used in survival analysis as a parametric model for events whose rate increases initially and decreases later, for example mortality from cancer following diagnosis or treatment. The survival function is

$$S(t) = [1 + e^{\theta}t^{\alpha}]^{-1}$$

and the density function is

$$f(t) = \frac{e^{\theta}\alpha t^{\alpha-1}}{(1+e^{\theta}t^{\alpha})^2}$$

It is called the log-logistic distribution because $logT$ has a logistic distribution (a symmetric distribution with density function similar to the normal density function).