STAT 6390: Analysis of Survival Data

Textbook coverage: Chapter 8

Steven Chiou

Department of Mathematical Sciences,
University of Texas at Dallas

Parametric models

- Methods described previously are *non-parametric*; no distributional assumptions were made.
- Non-parametric (and semi-parametric) methods have the flexibility to accommodate a wide range of applications.
- If the assumption of a particular probability distribution for the data is valid, inferences based on such an assumption will be more precise.
- The validity of the parametric methods depends heavily on the appropriateness of the distributional assumption.
- Parametric models are often much easier to work with.

Maximum likelihood estimation

- Suppose *actual survival times* observed for n individuals are $\{t_1, \ldots, t_n\}$.
- If the probability density function of the random variable associated with theos survival times is $f(t)$, the likelihood of the n observations is

$$\prod_{i=1}^{n} f(t_i).$$

- If a distributional assumption is made (e.g., $f(t) = \lambda e^{-t\lambda}$.), the unkonwon parameters (λ) can be estimated by maximzing the likelihood.

Maximum likelihood estimation

- Now suppose a the survival data includes (right) censored data.
- In this case, n pairs of observations are observed $(\tilde{t}_i, \Delta_i), i = 1, \ldots, n$, where $\tilde{T}_i$ is the observed survival time.
- When $\Delta_i = 0$, t_i is right-censored.
- The likelihood then takes the form

$$L = \prod_{i=1}^{n} [f_T(t_i)]^{\Delta_i} \cdot [S_T(t_i)]^{1-\Delta_i} = \prod_{i=1}^{n} [h_T(t_i)]^{\Delta_i} \cdot S_T(t_i). \tag{1}$$

- The last equation follows from the property $h(t) = f(t)/S(t)$.
- Note that the derivation of of L does not require a distributional assumption.

Maximum likelihood estimation

- A more careful derivation of the likelihood function in (1) is to assume the censoring times to be random.
- Let C_i be the random variable associated with the censoring time.
- Let $\tilde{T}_i$ be the observed survival time, $\tilde{T}_i = \min(C_i, T_i)$.
- We will consider censored and uncensored cases separately.
- For the censored observation:

$$\mathrm{P}(\tilde{T}_i = t, \Delta_i = 0) = \mathrm{P}(C_i = t, T_i > t).$$

- For the uncensored observation:

$$\mathrm{P}(\tilde{T}_i = t, \Delta_i = 1) = \mathrm{P}(T_i = t, C_i > t).$$

- The likelihood is then

$$L^* = \prod_{i=1}^{n} [\mathrm{P}(T_i = t, C_i > t)]^{\Delta_i} \cdot [\mathrm{P}(C_i = t, T_i > t)]^{1-\Delta_i}.$$

Maximum likelihood estimation

- Under the assumption that C_i and T_i are independent, L^* becomes

$$L^* = \prod_{i=1}^{n} [f_T(t_i)S_C(t_i)]^{\Delta_i} \cdot [f_C(t_i)S_T(t_i)]^{1-\Delta_i}.$$

- If the interest is in the parameter estimation in $f_T(\cdot)$, e.g., the λ in the exponential assumption, $f_C(\cdot)$ and $S_C(\cdot)$ can be considered as constant in the maximum likelihood estimation and L^* reduces to L.
- This construction shows the relevance of the assumption of independent censoring.

Exponential model

- If a random variable, T, follows an exponential distribution with rate λ, then
$$f(t) = \lambda e^{-\lambda t}, S(t) = e^{-\lambda t}, \text{ and } h(t) = \lambda.$$
- If we are willing to assumption $\{t_1, \ldots, t_n\}$ are iid samples from an exponential distribution with rate λ, then the likelihood $L(\lambda)$ is
$$L(\lambda) = \prod_{i=1}^{n} \left[\lambda e^{-\lambda t_i}\right]^{\Delta_i} \cdot \left[e^{-\lambda t_i}\right]^{1-\Delta_i} = \prod_{i=1}^{n} \lambda^{\Delta_i} \cdot e^{-\lambda t_i}.$$
- The log-likelihood is
$$\log L(\lambda) = \ell(\lambda) = \log(\lambda)\left(\sum_{i=1}^{n} \Delta_i\right) - \lambda \sum_{i=1}^{n} t_i.$$

Exponential model

- Solving for

$$\frac{\mathrm{d} \log L(\lambda)}{\mathrm{d}\lambda} = \ell'(\lambda) = 0 \text{ gives } \hat{\lambda} = \frac{\sum_{i=1}^{n} \Delta_i}{\sum_{i=1}^{n} t_i}.$$

- The maximum likelihood estimator (MLE), $\hat{\lambda}$, is the *number of deaths* divides by the total survival time (*number of person-years*).
- The MLE for the average survival time is $1/\hat{\lambda}$, which is the total survival time divides by the number of deaths.
- With $\hat{\lambda}$, other quantities like the MLE for median survival times, can be derived.

Exponential model

- The second derivative of $\ell(\lambda)$ gives the *information*.
- In the exponential model, we have

$$\ell''(\lambda) = -\frac{1}{\lambda^2}\sum_{i=1}^{n}\Delta_i.$$

- The standard MLE theory implies

$$\mathrm{Var}(\hat{\lambda}) \approx \frac{\hat{\lambda}^2}{\sum_{i=1}^{n}\Delta_i}.$$

- The $100(1-\alpha)\%$ confidence interval can be constructed accordingly.
- The Delta method can be applied to obtain standard errors for $g(\lambda)$, e.g., average survival time, median survival time, etc.

Weibull model

- The simplicity of the exponential distribution makes it attractive for some specialized applications.
- A more flexibility alternative is modeling with the Weibull distribution.
- If T follows a Weibull distribution with scale parameters λ and shape parameter γ, then

$$f(t) = \lambda\gamma t^{\gamma-1} e^{-\lambda t^{\gamma}}, S(t) = e^{-\lambda t^{\gamma}}, \text{ and } h(t) = \lambda\gamma t^{\gamma-1}.$$

- It is easy to see that when $\gamma = 1$, Weibull reduces to an exponential distribution with rate λ.

Weibull model

- Following the similar procedure as before, the likelihood $L(\lambda, \gamma)$ is

$$\prod_{i=1}^{n} \left\{ \lambda \gamma t_i^{\gamma - 1} \right\}^{\Delta_i} e^{-\lambda t_i^{\gamma}}.$$

- Let $\ell(\lambda, \gamma) = \log L(\lambda, \gamma)$, the MLE for λ turns out to be

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} \Delta_i}{\sum_{i=1}^{n} t_i^{\hat{\gamma}}},$$

but there is no close-form solution for $\hat{\gamma}$.

Weibull model

- The MLE $\hat{\theta} \equiv (\hat{\lambda}, \hat{\gamma})$ can be obtained directly implementing the likelihood and optimized with `optim`.
- Numerical method like the Newton-Raphson procedure can also be used.
- The basic idea of the Newton-Raphson procedure iterates

$$\hat{\theta}_{n+1} = \hat{\theta}_n + \left(-\frac{\mathrm{d}^2 \ell(\hat{\theta}_n)}{\mathrm{d}\theta^2}\right)^{-1} \cdot \frac{\mathrm{d}\ell(\hat{\theta}_n)}{\mathrm{d}\theta}.$$

- The variance-covariance matrix comes as a by-product.

Weibull model

- Since parametric models are sensitive to the distributional assumption, it is important to have a diagnostic tool.
- A diagnostic tool for Weibull model is derived from its survival curve.
- The log-log transformation of the Weibull survival function gives

$$\log[-\log S(t)] = \log(\lambda) - \gamma \log(t).$$

- This suggest that if $\log[-\log S(t)]$ is plotted against $\log(t)$, we would expect to see a straight line if the Weibull assumption is valid.
- $S(t)$ can be replaced with $\widehat{S}_{KM}(t)$ or $\widehat{S}_{NA}(t)$.

Weibull model

- An alternative way is to select λ and γ to match the survival data at two specified time points.
- This approach is motivated by the linear relationship between $\log\left[-\log S(t)\right]$ and $\log(t)$.
- Suppose we have (t_1, s_1), and (t_2, s_2) that are two time points on a estimated survival curve (e.g., set $s_i = \widehat{S}_{KM}(t_i)$ for $i = 1, 2$).
- Then $\hat{\lambda}$ and $\hat{\gamma}$ can be obtained by solving the system of equation

$$\begin{cases} \log\left(-\log s_1\right) = \log(\lambda) - \gamma \log(t_1) \\ \log\left(-\log s_2\right) = \log(\lambda) - \gamma \log(t_2) \end{cases},$$

for λ and γ.

Weibull regression

-