Chapter 3. Parametric approach

In this chapter we consider a homogeneous population of individuals, each having a failure time T. A simple inference to the distribution of the failure time T can be done by assuming that it follows a particular distribution. Then, apply the technique of maximum likelihood estimation (MLE) to obtain the corresponding parameters estimate. In the next section, we will introduce some common distributions for survival time. Later, we review the methodology of MLE in Section 3.2. In Section 3.3, a sketch of model checking will be discussed. At the end of this Chapter, we introduce a useful technique, $\delta-$method, which is frequently used in maximum likelihood approach.

3.1 Some useful distributions for survival time

Typically, the failure time T is assumed to be a continuous non-negative random variable. Therefore, three common distributions, Exponential distribution, Gamma distribution and Weibull distribution, are useful in describing the failure times of a homogeneous (i.i.d.) population of subjects. In fact, Exponential distribution is a special case of both Gamma and Weibull distributions. We will give the definition and some important properties of these three distributions in the following.

(1) Exponential distribution: $T\tilde{\ }Exp(\lambda)$

(i) $h(t) = \lambda$, (ii) $f(t) = \lambda \exp\{-\lambda t\}$, (iii) $S(t) = \exp\{-\lambda t\}$, (iv) $H(t) = \lambda t$

Properties:

(i) $E(T) = 1/\lambda$, $\mathrm{Var}(T) = 1/\lambda^2$, median $= \ln 2/\lambda$

(ii) Lack of memory: $\Pr(T \geq t_0 + a|T \geq t_0) = \exp\{-\lambda a\} = \Pr(T \geq a)$

(iii) If Y is continuous r.v with cumulative hazard function $H(\cdot)$, then $T = H(Y)\tilde{\ }Exp(1)$

(2) Gamma distribution: $T\tilde{}\Gamma(\lambda, k)$

(i) $f(t) = \frac{\lambda^k t^{k-1} e^{-\lambda t}}{\Gamma(k)}$ where $\Gamma(k) = \int_0^\infty \lambda^k t^{k-1} e^{-\lambda t} dt$

Properties:

(i) $E(T) = k/\lambda$, $\text{Var}(T) = k/\lambda^2$

(ii) when $k > 1$, $h(t) \nearrow$, $h(0) = 0$, $h(\infty) = \lambda$
when $k < 1$, $h(t) \searrow$, $h(0) = \infty$, $h(\infty) = \lambda$

(iii) $\Gamma(\lambda, 1) = Exp(\lambda)$, $Exp(1/2) = \chi^2_2$

(iv) If $T_1, \cdots, T_k$ are iid $Exp(\lambda)$, then $T_1 + \cdots + T_k\tilde{}\Gamma(\lambda, k)$

(v) $T\tilde{}\Gamma(\lambda, k)$, then $cT\tilde{}\Gamma(\frac{\lambda}{c}, k)$

Note: This properties tell us that we can apply chi-square table one wants to find the critical point of gamma distribution.

(Hint: $2\lambda\Gamma(\lambda, k) = \Gamma(\frac{1}{2}, k) = \sum_{i=1}^k \Gamma(\frac{1}{2}, 1)$)

(vi) $\Gamma(\lambda, k)/k \rightarrow$ normal distribution when $k \rightarrow \infty$. (Hint: $\Gamma(\lambda, k) = \sum_{i=1}^k \Gamma(\lambda, 1)$)

(3) Weibull distribution: $T\tilde{}W(\lambda, k)$

(i) $h(t) = k\lambda t^{k-1}$, (ii) $f(t) = k\lambda t^{k-1} \exp\{-\lambda t^k\}$, (iii) $S(t) = \exp\{-\lambda t^k\}$

Properties:

(i) $E(T) = \lambda^{-1/k}\Gamma(\frac{1}{k} + 1)$, $\text{Var}(T) = \lambda^{\frac{-2}{k}}[\Gamma(1 + \frac{2}{k}) - \Gamma^2(1 + \frac{1}{k})]$

(ii) $W(\lambda, 1) = Exp(\lambda)$,

(iii) $cW(\lambda, k) = W(\frac{\lambda}{c}, k)$,

(iv) $T\tilde{}W(\lambda, k)$, then $T^k\tilde{}Exp(\lambda)$

(v) when $k > 1$, $h(t) \nearrow$, $h(0) = 0$, $h(\infty) = \lambda$
when $k < 1$, $h(t) \searrow$, $h(0) = \infty$, $h(\infty) = \lambda$

Other than these three distributions, some other distributions of survival data include:

- Log Normal Distribution: It assumes that logarithm of the failure time $log(T)$ follows normal distribution.

- Piecewise Exponential Distribution: Recall that the hazard function of the exponential distribution is a positive constant. The piecewise exponential distribution assumes the hazard function is a step function. As we know, a step function is a good approximation to any unknown function. The approximation becomes better when the length of each step becomes smaller. Thus, piecewise exponential distribution can be used to be an approximated distribution while the true distribution is unknown.

3.2 Maximum likelihood technique

This section gives a simple review to MLE that includes definitions (Section 3.2.1), asymptotic properties (Section 3.2.2) and three test statistics based on likelihood function (Section 3.2.3). A more detail inference for right-censoring data (Section 3.2.4) and left-truncation data (Section 3.2.5), respectively, through MLE will also be given.

3.2.1 Review of maximum likelihood theory

Given $X_1, \cdots, X_n$ are iid from p.d.f. $f(x;\theta)$ when θ is an unknown $p \times 1$ parameter vector. Then

- Likelihood function is $L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$

- MLE $\hat{\theta}$ of θ is the value of θ that maximize $L(\theta)$ given the data.

Example: $f(x;\theta)\tilde{}N(\mu,\sigma^2) \Rightarrow \hat{\mu} = \frac{1}{n}\sum x_i$.

The following are some corresponding definitions.

- $l(\theta) = \ln L(\theta) = \sum_{i=1}^{n} \ln f(x_i;\theta) =$ log likelihood function

- $U(\theta) = \frac{\partial l(\theta)}{\partial \theta} =$ score function

- $I(\theta) = -\frac{\partial U(\theta)}{\partial \theta} = -\frac{\partial^2 U(\theta)}{\partial \theta^2} =$ sample Fisher information matrix

- $i(\theta) = E(I(\theta)) =$ expected Fisher information matrix

3.2.2 Asymptotic results:

(i) $\hat{\theta} \to \theta$ almost surely

(ii) $(\hat{\theta}-\theta)^T[i(\theta)]^{\frac{1}{2}}$ converges weakly to multivariate normal distribution with zero mean and covariance matrix $I_{p\times p}$

(iii) $(\hat{\theta}-\theta)^T[i(\theta)](\hat{\theta}-\theta)$ converges weakly to X_p^2

(iv) $\sqrt{n}(\hat{\theta}-\theta) \approx$ multivariate normal with mean 0 and covariance matrix $[i(\theta)]^{-1}$

(v) Define the likelihood ratio statistic as

$$LRT(\theta) = -2\ln[L(\theta_0)/L(\hat{\theta})] = 2[l(\hat{\theta}) - l(\theta_0)] \sim X_p^2$$

(vi) $U(\theta)[i(\theta)]^{-\frac{1}{2}}$ converges weakly to $N(0, I_{p\times p})$. $i(\theta)$ may be replaced by $I(\theta), I(\hat{\theta}), i(\hat{\theta})$.

3.2.3 Testing $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$

(i) Wald's test: $\hat{\theta} - \theta_0\tilde{\ }N(0, I(\theta_0)^{-1})$ or equivalently $(\hat{\theta}-\theta_0)^T I(\theta_0)(\hat{\theta}-\theta_0)\tilde{\ }X_p^2$

(ii) Score test: $U(\theta_0)\tilde{\ }N(0, I(\theta_0))$ or equivalently $U(\theta_0)^T[I(\theta_0)]^{-1}U(\theta_0)\tilde{\ }X_p^2$

(iii) Likelihood ratio test: $2[l(\hat{\theta}) - l(\theta_0)]\tilde{\ }X_p^2$

3.2.4 Inference for right censoring

Let T_i be r.v with survival function $S(\cdot;\theta)$ and p.d.f. $f(\cdot;\theta)$, c_i be censoring r.v. with survival function $\bar{G}$ and p.d.f. g_i. We observe (Y_i, δ_i) where $Y_i = \min(T_i, C_i)$ and $\delta_i = I(Y_i = T_i)$ for $i = 1, \cdots, n$. Then the likelihood function is

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

If $\bar{G}$ is not a function of θ, then

$$\begin{aligned} L(\theta) &\propto L_1(\theta) = \prod_{i=1}^{n}[f(y_i;\theta)]^{\delta_i}[S(y_i;\theta)]^{1-\delta_i} \\ &= \prod_{i=1}^{n}[f(y_i;\theta)/S(y_i;\theta)]^{\delta_i}S(y_i;\theta) \\ &= \prod_{i=1}^{n}[h(y_i;\theta)]^{\delta_i}S(y_i;\theta). \end{aligned}$$

For an example, suppose $T_i\tilde{}Exp(\lambda)$ (Note, mean $= \mu = 1/\lambda$). If no censoring, $\hat{\mu} = \sum T_i/n$. However, if we can only observe $Y_i = \min(T_i, C_i)$ (i.e. right censoring), the average $\tilde{\mu} = \sum Y_i/n$ will underestimate the true mean. Applying the method of maximum likelihood, we find θ that maximize

$$l_1(\lambda) = \ln L_1(\lambda) = \sum_{i=1}^{n}(\delta_i \ln\lambda - \lambda Y_i).$$

It can be solved by solving

$$U(\lambda) = \frac{\partial l_1(\lambda)}{\partial\lambda} = \sum_{i=1}^{n}(\frac{\delta_i}{\lambda} - Y_i) = 0$$

This implies the MLE of λ will be

$$\hat{\lambda} = r/\sum Y_i$$

where $r = \sum \delta_i$. Therefore $\hat{\mu} = \sum Y_i/r$.

(i) If no censoring, we have $r = n$. Then

(a) $n/\hat{\lambda} = \sum Y_i = \sum T_i\tilde{}\Gamma(\lambda, n)$

(b) $2\lambda n/\hat{\lambda}\tilde{}X^2_{2n}$

(c) $\Pr(a \le 2\lambda n/\hat{\lambda} \le b) = 0.95 \Rightarrow \Pr(\frac{a\hat{\lambda}}{2n} \le \lambda \le \frac{b\hat{\lambda}}{2n}) = 0.95$.

(ii) If type II censoring:

(a) $1/\hat{\lambda} = \frac{1}{r}[T_{(1)} + \cdots + T_{(r-1)} + (n-r)T_{(r)}] = \sum_{i=0}^{r-1}[T_{(i+1)} - T_{(i)}]$ where $T_{(0)} = 0$ and $T_{(i)}$ $(i = 1, \cdots, r)$ are order statistics of T_i $(i = 0, \cdots, n)$.

(b) $2\lambda r/\hat{\lambda}\tilde{}X^2_{2r}$

(c) $\Pr(a^* \le 2\lambda r/\hat{\lambda} \le b^*) = 0.95 \Rightarrow \Pr(\frac{a^*\hat{\lambda}}{2r} \le \lambda \le \frac{b^*\hat{\lambda}}{2r}) = 0.95$.

(iii) If random censoring:

(a) $I(\lambda) = \frac{-\partial U(\lambda)}{\partial\lambda} = \frac{r}{\lambda^2}$

(b) $\sqrt{r}(\hat{\lambda} - \lambda)/\hat{\lambda}\tilde{}N(0, 1)$

(c) $\Pr(-1.96 \le \sqrt{r}(\hat{\lambda} - \lambda)/\hat{\lambda} \le 1.96) = 0.95 \Rightarrow \Pr(\hat{\lambda} - \frac{1.96\hat{\lambda}}{\sqrt{r}} \le \lambda \le \hat{\lambda} + \frac{1.96\hat{\lambda}}{\sqrt{r}}) = 0.95$

(In this case, we may obtain a 95% C.I. of λ that includes negative value. Since a function of MLE is still MLE, a simple modification is to create a 95% C.I. of $\beta = \ln\lambda$ first and to transform it back to obtain a 95% C.I. of λ. To do this, we need delta method in getting the variance estimate. A more detail discussion will be given in section 3.4.)

3.2.5 Inference for left truncation

Suppose $Y_1, Y_2, \cdots, Y_m \tilde{} F(\cdot;\theta)$ (p.d.f. $f(\cdot;\theta)$) and $K_1, K_2, \cdots, K_m \tilde{} G$ (p.d.f. g, left truncation time). We can only observe (Y_i^*, K_i^*) with $Y_i^* \geq K_i^*$ $(i = 1, \cdots, n)(n \leq m)$. Assume $Y \perp K$, then the joint p.d.f. of (Y^*, K^*) is $f(y;\theta)g(k)/c$ $\forall y \geq k$ where $c = \int\int_{k<y} f(y)g(k)dydk$. The conditional likelihood of Y_i^* given K_i^* is

$$\begin{aligned} L_c(\theta) &= \prod_{i=1}^{n} \Pr(Y_i = y_i^* | Y_i \geq k_i^*) \\ &= \prod_{i=1}^{n} \frac{f(y_i^*)}{S(y_i^*)} \frac{S(y_i^*)}{S(k_i^*)} = \prod_{i=1}^{n} h(y_i^*) \frac{S(y_i^*)}{S(k_i^*)}. \end{aligned}$$

For example, if $Y \tilde{} Exp(\lambda)$, then $L_c(\lambda) = \prod_{i=1}^{n} \lambda \exp\{-\lambda(y_i^* - k_i^*)\}$

3.3 Model Checking

In this subsection, we discuss a widely used model checking technique to check whether a data is generated from Weibull distribution.

Suppose $T \tilde{} W(\lambda, k)$, then $S(t) = exp\{-\lambda t^k\}$. This implies

$$\ln[-\ln S(t)] = \ln \lambda + k \ln t.$$

Therefore, plots of $\ln[-\ln \hat{S}(t)]$ vs $\ln t$ gives an approximately straight line if T is generated from Weibull distribution. Here, $\hat{S}(t)$ is a non-parametric survival function estimate of $S(t)$. Typically, Kaplan-Meier (K-M) estimator of survival function is considered. We will introduce it later. In addition, since $H(t) = -\ln S(t)$, it can also be looked as plots of $\ln H(t)$ vs $\ln t$.

To check any other distribution, suppose T is an arbitrary, non-negative, continuous random variable with integrated hazard function $H(t)$. This implies that $H(T) \tilde{} Exp(1)$. Then, plots of $\ln[\hat{S}(H(t))]$ vs $H(t)$ gives an approximately straight line if the assumption to the cumulative hazard function $H(t)$ is true. Here, $\hat{S}(H(t))$ is calculated by two steps. In the first step, we specify $H(t) = H(t|\theta)$ and estimate θ, say $\hat{\theta}$, through MLE. For the second step, we set $Y_i = H(T_i|\hat{\theta})$ and find $\hat{S}(y)$ based on Y_i. Specifically, $\hat{S}(y)$ can be the K-M estimator based on Y_i. Then plot $\hat{S}(y)$ vs y.

For an example, suppose we observe a survival data with right censored mechanism, says (t_i, δ_i) $(i = 1, \cdots, n)$. The notation here is the same as we discuss in the previous section. If we want to check whether the failure time is generated from a gamma distribution, we then have the following likelihood function

$$L(\lambda, k) \propto \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}$$

where $f(\cdot)$ is the pdf of $\Gamma(\lambda, k)$ and $S(t) = \int_t^\infty f(\mu)d\mu$ is the corresponding survival function.

Suppose that $\hat{\lambda}, \hat{k}$ be the MLE of λ, k. Set $y_i = H(t_i; \hat{\lambda}, \hat{k})$. Let $\hat{S}(H(t))$ be the K-M estimate based on (y_i, δ_i) $(i = 1, \cdots, n)$. This implies that plots of $-\ln \hat{S}(y)$ vs y gives an approximately straight line if the failure time has gamma distribution.

3.4 δ-method

Suppose k is a smooth function of θ, $k^{'}(\theta) \neq \theta$ $\forall \theta$. Then

$$\sqrt{n}(\hat{\theta} - \theta)\tilde{}N(0, \sigma^2) \Rightarrow \sqrt{n}(k(\hat{\theta}) - k(\theta))\tilde{}N(0, [k^{'}(\theta)]^2\sigma^2).$$

(This can be viewed through Taylar's expansion $k(\hat{\theta}) \approx k(\theta) - k^{'}(\theta)(\hat{\theta} - \theta)$).

Example (Continue): $\sqrt{r}(\hat{\lambda} - \lambda)\tilde{}N(0, \lambda^2)$

(i) $k(\lambda) = 1/\lambda = \mu \Rightarrow k^{'}(\lambda) = -1/\lambda^2$. Then $\sqrt{r}(\hat{\mu} - \mu)\tilde{}N(0, \lambda^{-2}) = N(0, \mu^2)$. This impies 95% C.I. for μ is $\hat{\mu} \pm 1.96\frac{\hat{\mu}}{\sqrt{r}}$.

(ii) $k(\lambda) = \ln \lambda \Rightarrow k^{'}(\lambda) = 1/\lambda$. Then $\sqrt{r}(\ln \hat{\lambda} - \ln \lambda)\tilde{}N(0, 1)$. This impies 95% C.I. for λ is $\hat{\lambda} \exp\{\pm 1.96\frac{1}{\sqrt{r}}\}$.