m() function's consistency

2019-10-29

Contents

1. MLE function derivation **1**

2. $L(\theta) \leq L(\theta_0)$ **2**

3. Asymptotic normality of $\hat{\theta}_n$ **3**

4. The consistency of the $m()$ **5**

1. MLE function derivation

To make things easy, I just consider the one dimension scenario at this time.

We denote $Y_i, i = 1, ..., N$ are the independent, identically, distributed (iid) lifetimes, whose corresponding cumulative distriubtion function (CDF) is F, probability distribution function (PDF) is f; the censoring time is defined as $C_i, i = 1, ..., N$. C_is are also iid, with CDF denoted as G and PDF denoted as g. We set the censors happen on the right and the ovserved time is $Z_i = Y_i \wedge C_i$, whose CDF is H and PDF is h. The $\delta_i = I_{[T_i \leq C_i]}$ is the status indicator, which shows whether subject i is censored ($\delta_i = 0$) or not ($\delta_i = 0$). The corresponding hazard function of lifetime is λ_F and cumulative hazard function is Λ_F.

Instead of the strong assumption of independent between Y_i and C_i, we proposed that $T \perp\!\!\!\perp C$ at a small neighborhood, where $T = C$. And define:

$$m_\theta(t) = P(\delta = 1 | Z = z) = \lambda_F(t)/\lambda_H(t)$$

Giving observed $(\delta_1, Z_1), (\delta_2, Z_2), ..., (\delta_n, Z_n)$, the likelihood function can be written as:

$$L(\theta) = \prod_{i=1}^{n} m_\theta(z_i)^{\delta_i}(1 - m_\theta(z_i))^{1-\delta_i}\lambda_H(z_i)S_H(z_i)$$

where $f_\theta(\delta_i, z_i) = \Big[m_\theta(z_i)\Big]^{\delta_i}\Big[(1 - m_\theta(z_i))\Big]^{1-\delta_i}\lambda_H(z_i)S_H(z_i)$ And

$$\log(L(\theta)) = \sum_{i=1}^{n} \Big[\delta_i \log\Big(m_\theta(z_i)\lambda_H(z_i)S_H(z_i)\Big) + (1 - \delta_i)\log((1 - m_\theta(z_i))\lambda_H(z_i)S_H(z_i))$$

2. $L(\theta) \leq L(\theta_0)$

Let θ_0 be the θ that can maximize the likelihood function $L(\theta)$. Let $\hat{\theta}_n$ denote the maximum likelihood estimation, which maximize $L_n(\theta)$.

Lemma 1: We have that for any θ,

$$L(\theta) \leq L(\theta_0)$$

Proof:

Since

$$L(\theta) \leq L(\theta_0)$$
$$\log(L(\theta)) \leq log(L(\theta_0))$$

Then

$$\begin{aligned}
l(\theta_0) - l(\theta) =& \sum_{i=1}^{n} \log \left(m_{\theta_0}(z_i)\right)^{\delta_i} \log(1 - m_{\theta_0}(z_i))^{(1-\delta_i)} \lambda_H(z_i) S_H(z_i) \\
&- \sum_{i=1}^{n} \log \left(m_{\theta}(z_i)\right)^{\delta_i} \log(1 - m_{\theta}(z_i))^{(1-\delta_i)} \lambda_H(z_i) S_H(z_i) \\
=& \sum_{i=1}^{n} \log \frac{\left(m_{\theta_0}(z_i)\right)^{\delta_i} \log(1 - m_{\theta_0}(z_i))^{(1-\delta_i)} \lambda_H(z_i) S_H(z_i)}{\left(m_{\theta}(z_i)\right)^{\delta_i} \log(1 - m_{\theta}(z_i))^{(1-\delta_i)} \lambda_H(z_i) S_H(z_i)}
\end{aligned}$$

Based on Law of Large Number (LLN),

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{n} \log \frac{\left(m_{\theta_0}(z_i)\right)^{\delta_i} \log(1 - m_{\theta_0}(z_i))^{(1-\delta_i)} \lambda_H(z_i) S_H(z_i)}{\left(m_{\theta}(z_i)\right)^{\delta_i} \log(1 - m_{\theta}(z_i))^{(1-\delta_i)} \lambda_H(z_i) S_H(z_i)} \\
&\xrightarrow{P} E(\log \frac{\left(m_{\theta_0}(z_i)\right)^{\delta_i} \log(1 - m_{\theta_0}(z_i))^{(1-\delta_i)} \lambda_H(z_i) S_H(z_i)}{\left(m_{\theta}(z_i)\right)^{\delta_i} \log(1 - m_{\theta}(z_i))^{(1-\delta_i)} \lambda_H(z_i) S_H(z_i)})
\end{aligned}$$

And

$$E(\log \frac{\left(m_{\theta_0}(z_i)\right)^{\delta_i} \log(1 - m_{\theta_0}(z_i))^{(1-\delta_i)} \lambda_H(z_i) S_H(z_i)}{\left(m_{\theta}(z_i)\right)^{\delta_i} \log(1 - m_{\theta}(z_i))^{(1-\delta_i)} \lambda_H(z_i) S_H(z_i)}) = \int_0^\infty \log(\frac{f_{\theta_0}(\delta, z)}{f_\theta(\delta, z)}) f_{\theta_0}(\delta, z) dz \quad (1)$$

Recall Kullback–Leibler divergence,

$$D_{KL}(F||G) = \int f \log(\frac{f}{g}) \geq 0$$

Therefore, equation (1) ≥ 0. Therefore, $L(\theta_0) \geq L(\theta)$

3. Asymptotic normality of $\hat{\theta}_n$

Look back to the likelihood function, let

$$l(\theta) = E_\theta\Big[\log(f_\theta(z_i, \delta_i))\Big] = \int \log(f_\theta(z_i, \delta_i)) f_\theta(z_i, \delta_i) dz$$

let

$$l_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \log(f_\theta(z_i, \delta_i))$$

By LLN, $l_n(\theta) \xrightarrow{p} l(\theta)$.

The Taylor expension of $l_n(\theta)$ at θ_0 is

$$l_n(\theta) = l_n(\theta_0) + \frac{l_n'(\theta_0)}{1!}(\theta - \theta_0) + \frac{l_n''(\theta_0)}{2!}(\theta - \theta_0)^2 + o\Big((\theta - \theta_0)^2\Big)$$

$$l_n(\theta) - l_n(\theta_0) = u_n(\theta_0)(\theta - \theta_0) + \frac{1}{2}u_n'(\theta_0)(\theta - \theta_0)^2 + R_n$$

where

$$R_n = o\Big((\theta - \theta_0)^2\Big)$$

$u_n(\theta_0)$ is the score function, which is the first derivative of the log likelihood function:

$$u_n(\theta_0) = \frac{dl_n(\theta)}{d\theta}|_{\theta=\theta_0}$$

The Taylor expension of the score fucntion $u_n(\theta)$ at θ_0 is:

$$u_n(\theta) = u_n(\theta_0) + u_n'(\theta_0)(\theta - \theta_0) + r_n$$

where,

$$r_n = o\Big(\theta - \theta_0\Big)$$

$$u_n'(\theta_0) = \frac{d^2 l_n(\theta)}{d\theta^2}|_{\theta=\theta_0}$$

Besidse, we have the facts:

- By defination, $\hat{\theta}_n$ is the maximizer of $l_n(\theta)$ and $u_n(\hat{\theta}_n) = 0$
- By defination, θ_0 is the maximizer of $l(\theta)$ and $u(\theta_0) = 0$

Therefore,

$$(\hat{\theta}_n - \theta_0) = -\frac{u_n(\theta_0)}{u_n'(\theta_0)} + r_{n2}$$

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \approx \sqrt{n}\frac{u_n(\theta_0)}{u_n'(\theta_0)}$$

where $r_{n2} = -\frac{r_n}{u_n'(\theta_0)} \to 0$.

Therefore, we need to look at the distributions of $u_n(\theta_0)$ and $u_n'(\theta_0)$

Recall the Centrol limit theroem

$$\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n} X_i - E(X)) \xrightarrow{d} N(0, Var(X))$$

And we have $u_n(\theta_0) = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\theta}\log(f_\theta(\delta_i, z_i))|_{\theta=\theta_0}$

$$E(\frac{\partial}{\partial\theta}\log(f_\theta(\delta, z))|_{\theta=\theta_0}) = \int \frac{f_\theta'(\delta, z)}{f_\theta(\delta, z)}|_{\theta=\theta_0} f_{\theta_0}(\delta, z)dz = \frac{\partial}{\partial\theta}\int f_{\theta_0}(\delta, z)dz = 0$$

Therefore,

$$\sqrt{n}u_n(\theta_0) \xrightarrow{d} N(0, \text{Var}(\frac{\partial}{\partial\theta}\log(f_\theta(\delta, z)|_{\theta=\theta_0}))$$

For $u_n'(\theta_0)$, by LLN,

$$u_n'(\theta_0) = \frac{1}{n}\sum_{i=1}^{n}\Big[\log f_\theta(Z_i, \delta_i)\Big]''_{\theta=\theta_0} \to E_{\theta_0}\Big[\log f_\theta(Z, \delta)\Big]''_{\theta=\theta_0}\Big]$$

where,

$$\begin{aligned}
E_{\theta_0}\Big[\log f_\theta(Z, \delta)\Big]''_{\theta=\theta_0}\Big] &= \int \frac{f_{\theta_0}''(\delta, z)f_{\theta_0}(\delta, z) - f_{\theta_0}'(\delta, z)f_{\theta_0}'(\delta, z)}{(f_{\theta_0}(\delta, z))^2} f_{\theta_0}(\delta, z)dz \\
&= \int f_{\theta_0}''(\delta, z)dz - \int (\frac{f_{\theta_0}'(\delta, z)}{f_{\theta_0}(\delta, z)})^2 f_{\theta_0}(\delta, z)dz \\
&= 0 - E_{\theta_0}\Big[(\frac{\partial}{\partial\theta}\log(f_{\theta_0}(\delta, z))|_{\theta=\theta_0})^2\Big]
\end{aligned}$$

Recall the definiation of Fisher information: Fisher information is the variance of score function, which is

$$I(\theta_0) = E_{\theta_0}\Big[(\frac{\partial}{\partial\theta}\log(f_{\theta_0}(\delta, z))|_{\theta=\theta_0})^2$$

Therefore,

$$u_n'(\theta_0) = \frac{1}{n}\sum_{i=1}^{n}\Big[\log f_\theta(Z_i, \delta_i)\Big]''_{\theta=\theta_0} \to E_{\theta_0}\Big[\log f_\theta(Z, \delta)\Big]''_{\theta=\theta_0}\Big] = -I(\theta_0)$$

Therefore,

- $\sqrt{n}u_n(\theta_0) \xrightarrow{d} N(0, \text{Var}(\frac{\partial}{\partial\theta}\log(f_\theta(\delta, z)|_{\theta=\theta_0}))$
- $u_n'(\theta_0) = -I(\theta_0)$, which is a fixed value

Then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \approx \sqrt{n}\frac{u_n(\theta_0)}{u_n'(\theta_0)} \xrightarrow{d} N(0, \frac{\text{Var}(\frac{\partial}{\partial\theta}\log(f_\theta(\delta, z)|_{\theta=\theta_0})}{I^2(\theta_0)})$$

And $\text{Var}(\frac{\partial}{\partial\theta}\log(f_\theta(\delta, z)|_{\theta=\theta_0}) = E((\frac{\partial}{\partial\theta}\log(f_\theta(\delta, z)|_{\theta=\theta_0})^2) - [E(\frac{\partial}{\partial\theta}\log(f_\theta(\delta, z)|_{\theta=\theta_0})]^2 = E((\frac{\partial}{\partial\theta}\log(f_\theta(\delta, z)|_{\theta=\theta_0})^2) - 0 = I(\theta_0)$

Therefore,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \approx \sqrt{n}\frac{u_n(\theta_0)}{u_n'(\theta_0)} \xrightarrow{d} N(0, \frac{1}{I(\theta_0)})$$

4. The consistency of the $m()$

Since $\hat{\theta}_n \xrightarrow{d} \theta_0$, according to Delta method,

$$m(\hat{\theta}_n) \xrightarrow{d} m(\theta_0)$$

$$\sqrt{n}(m(\hat{\theta}_n) - m(\theta_0)) \xrightarrow{d} N(0, \frac{m'(\theta_0))\mathrm{Var}(\frac{\partial}{\partial\theta}\log(f_\theta(\delta, z)|_{\theta=\theta_0})}{I^2(\theta_0)})$$