m() function's consistency

2019-10-30

Contents

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

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

Let

- $w_1(z_i; \theta) = m_\theta(z_i)\lambda_H(z_i)S_H(z_i) = f_\theta(1, z_i)$,
- $w_2(z_i; \theta) = (1 - m_\theta(z_i))\lambda_H(z_i)S_H(z_i) = f_\theta(0, z_i)$

Then

$$\log(L(\theta)) = \sum_{i=1}^{n} \delta_i \log(w_1(z_i;\theta)) + \sum_{i=1}^{n}(1-\delta_i)\log(w_2(z_i;\theta))$$

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

$$\begin{aligned}\log(L(\theta_0)) - log(L(\theta)) =& \sum_{i=1}^{n} \delta_i \log(w_1(z_i;\theta_0)) + \sum_{i=1}^{n}(1-\delta_i)\log(w_2(z_i;\theta_0)) \\ & - \sum_{i=1}^{n} \delta_i \log(w_1(z_i;\theta)) + \sum_{i=1}^{n}(1-\delta_i)\log(w_2(z_i;\theta)) \\ =& \sum_{i=1}^{n} \delta_i \log(\frac{w_1(z_i;\theta_0)}{w_1(z_i;\theta)}) + \sum_{i=1}^{n}(1-\delta_i)\log(\frac{w_2(z_i;\theta_0)}{w_2(z_i;\theta)})\end{aligned}$$

And by LLN,

$$\frac{1}{n}\sum_{i=1}^{n} \delta_i \log(\frac{w_1(z_i;\theta_0)}{w_1(z_i;\theta)}) \xrightarrow{p} \Delta E_{\theta_0}\Big(\log(\frac{w_1(z;\theta_0)}{w_1(z;\theta)})\Big)$$

where $\Delta = \frac{\sum_{i=1}^{n}\delta_i}{n}$.

Similiarly,

$$\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i) \log(\frac{w_2(z_i;\theta_0)}{w_2(z_i;\theta)}) \xrightarrow{p} (1-\Delta) E_{\theta_0}\Big(\log(\frac{w_2(z;\theta_0)}{w_2(z;\theta)})\Big)$$

And

$$E_{\theta_0}\Big(\log(\frac{w_1(z;\theta_0)}{w_1(z;\theta)})\Big) = \int \log(\frac{w_1(z;\theta_0)}{w_1(z;\theta)}) f_{w_1}(z;\theta)dz = \int \log(\frac{w_1(z;\theta_0)}{w_1(z;\theta)}) w_1(z;\theta_0)dz \tag{1}$$

Recall Kullback–Leibler divergence,

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

Therefore, equation (1) ≥ 0. Also, $(1-\Delta)E_{\theta_0}\Big(\log(\frac{w_2(z;\theta_0)}{w_2(z;\theta)})\Big) \geq 0$. Therefore, $\log(L(\theta_0)) - log(L(\theta)) \geq 0$, $L(\theta_0) \geq L(\theta)$.

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

Let

- $w_1(z_i;\theta) = \Big(m_\theta(z_i)\lambda_H(z_i)S_H(z_i)\Big)$,
- $w_2(z_i;\theta) = \Big((1-m_\theta(z_i))\lambda_H(z_i)S_H(z_i)\Big)$

Then

$$\log(L(\theta)) = \sum_{i=1}^{n}\delta_i \log w_1(z_i;\theta) + \sum_{i=1}^{n}(1-\delta_i)\log w_2(z_i;\theta)$$

. Let

- $l_1(\theta) = \Delta E_\theta(\log w_1(z;\theta)) = \Delta \int \log w_1(z;\theta) f_\theta(1,z)dz$
- $l_2(\theta) = (1-\Delta)E_\theta(\log w_2(z;\theta)) = (1-\Delta)\int \log w_2(z;\theta) f_\theta(0,z)dz$
- $l_{1,n} = \frac{1}{n}\sum_{i=1}^{n}\delta_i \log w_1(z_i;\theta)$, $l_{2,n} = \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\log w_2(z_i;\theta)$

By LLN

$$l_{1,n}(\theta) \xrightarrow{p} l_1(\theta), l_{2,n}(\theta) \xrightarrow{p} l_2(\theta)$$

The Taylor expension of $l_{i,n}(\theta)$ (i = 1, 2) at θ_0 is

$$l_{i,n}(\theta) = l_{i,n}(\theta_0) + \frac{l'_{i,n}(\theta_0)}{1!}(\theta-\theta_0) + \frac{l''_{i,n}(\theta_0)}{2!}(\theta-\theta_0)^2 + o\Big((\theta-\theta_0)^2\Big)$$

$$l_{i,n}(\theta) - l_{i,n}(\theta_0) = u_{i,n}(\theta_0)(\theta-\theta_0) + \frac{1}{2}u'_{i,n}(\theta_0)(\theta-\theta_0)^2 + R_n$$

where

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

$u_{i,n}(\theta_0)$ is the score function, which is the first derivative of the log likelihood function:

$$u_{i,n}(\theta_0) = \frac{dl_n(\theta)}{d\theta}|_{\theta=\theta_0}$$

The Taylor expension of the score fucntion $u_{i,n}(\theta)$ at θ_0 is:

$$u_{i,n}(\theta) = u_{i,n}(\theta_0) + u'_{i,n}(\theta_0)(\theta-\theta_0) + r_n$$

where,

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

$$u'_{i,n}(\theta_0) = \frac{d^2 l_{i,n}(\theta)}{d\theta^2}|_{\theta=\theta_0}$$

Besidse, we have the facts when we prove $L(\theta) \leq L(\theta_0)$:

- By defination, $\hat{\theta}_n$ is the maximizer of $l_{i,n}(\theta)$ and $u_{i,n}(\hat{\theta}_n) = 0$, $i = 1, 2$
- By defination, θ_0 is the maximizer of $l_{i,n}(\theta)$ and $u_{i,n}(\theta_0) = 0$, $i = 1, 2$

Therefore,

$$(\hat{\theta}_n - \theta_0) = -\frac{u_{i,n}(\theta_0)}{u'_{i,n}(\theta_0)} + r_{n2}$$

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \approx \sqrt{n}\frac{u_{i,n}(\theta_0)}{u'_{i,n}(\theta_0)}$$

where $r_{n2} = -\frac{r_n}{u'_{i,n}(\theta_0)} \to 0$.

Therefore, we need to look at the distributions of $u_{i,n}(\theta_0)$ and $u'_{i,n}(\theta_0)$

We can derivate it following the same way. However, does the fisher information the same for l_1 and l_2? No?