m() function's consistency

2019-10-28

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

Roadmap

- 1. Derivation of the likelihood function
- 2. Show that the true parameter is the one that maximize the likelihood function
- 3. The consistency of the θ
- 4. The consistency of the $m()$

MLE function derivation

True θ_0 maximizes the likelihood function

Consistency of $\hat{\theta}_n$:

Consistency of $m_{\hat{\theta}_n}(t)$

- 1. the true θ is the θ that can maximize the likelihood function
- 2. normal, o means

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.

Define:

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

Let θ_0^* be the θ that can maximize the likelihood function. Let $\hat{\theta}_n$ denote the maximum likelihood estimation.

Theroem *$\sqrt{n}(\hat{\theta}_n - \theta_0^*)$ is asymptiotically normal, with $N(0, I^{-1}(\theta_0^*))$*

Proof:

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

And

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

Based on Law of Large Number (LLN),

$$\frac{1}{n}\sum_{i=1}^{n}\log\big(\frac{m_{\theta_0^*}(z_i)f_H^*(z_i)}{m_\theta(z_i)f_H(z_i)}\big) \xrightarrow{P} E(\log\big(\frac{m_{\theta_0^*}(z_i)f_H^*(z_i)}{m_\theta(z_i)f_H(z_i)}\big))$$

Since

$$E(\log\big(\frac{m_{\theta_0^*}(z_i)f_H^*(z_i)}{m_\theta(z_i)f_H(z_i)}\big)) = \int_0^\infty \log\big(\frac{m_{\theta_0^*}(z_i)f_H^*(z_i)}{m_\theta(z_i)f_H(z_i)}\big)[m_{\theta_0^*}(z_i)f_H^*(z_i)]dz_i$$

Recall Kullback–Leibler divergence,

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

Therefore,

$$E(\log\big(\frac{m_{\theta_0^*}(z_i)f_H^*(z_i)}{m_\theta(z_i)f_H(z_i)}\big)) = \int_0^\infty \log\big(\frac{m_{\theta_0^*}(z_i)f_H^*(z_i)}{m_\theta(z_i)f_H(z_i)}\big)[m_{\theta_0^*}(z_i)f_H^*(z_i)]dz_i \geq 0$$

Similiarly,

$$\frac{1}{n}\sum_{i=1}^{n}\log\big(\frac{(1-m_{\theta_0^*}(z_i))f_H(z_i)}{(1-m_\theta(z_i))f_H(z_i)}\big) \to E(\log\big(\frac{(1-m_{\theta_0^*}(z_i))f_H(z_i)}{(1-m_\theta(z_i))f_H(z_i)}\big)) \geq 0$$

Therefore, $l_{\theta_0^*} \geq l_\theta$ for any other θ that is not the true θ_0^*.

The true θ_0^* maximizes the likelihood function.

Suppose θ_0^* is the true vaue of θ. Suppose $f_H^*(z)$ is the true density. We would like to prove that

$$l_{\theta_0^*} = sup l_\theta$$

Which equivalent to

$$\begin{aligned}&\sum_{i=1}^{n}\Big[\delta_i\log(m_{\theta_0^*}(z_i)f_H(z_i)) + (1-\delta_i)\log((1-m_{\theta_0^*}(z_i))f_H(z_i))\Big]\\ &-\sum_{i=1}^{n}\Big[\delta_i\log(m_\theta(z_i)f_H(z_i)) + (1-\delta_i)\log((1-m_\theta(z_i))f_H(z_i))\Big] \geq 0\\ \to &\frac{1}{n}\sum_{i=1}^{n}\delta_i\log\big(\frac{m_{\theta_0^*}(z_i)f_H(z_i)}{m_\theta(z_i)f_H(z_i)}\big) + \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\log\big(\frac{(1-m_{\theta_0^*}(z_i))f_H(z_i)}{(1-m_\theta(z_i))f_H(z_i)}\big) \geq 0\end{aligned}$$

Based on Law of Large Number (LLN),

$$\frac{1}{n}\sum_{i=1}^{n}\log\Big(\frac{m_{\theta_0^*}(z_i)f_H^*(z_i)}{m_\theta(z_i)f_H(z_i)}\Big) \to E(\log\Big(\frac{m_{\theta_0^*}(z_i)f_H^*(z_i)}{m_\theta(z_i)f_H(z_i)}\Big))$$

Since

$$E(\log\Big(\frac{m_{\theta_0^*}(z_i)f_H^*(z_i)}{m_\theta(z_i)f_H(z_i)}\Big)) = \int_0^\infty \log\Big(\frac{m_{\theta_0^*}(z_i)f_H^*(z_i)}{m_\theta(z_i)f_H(z_i)}\Big)[m_{\theta_0^*}(z_i)f_H^*(z_i)]dz_i$$

According to Kullback–Leibler divergence,

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

Therefore,

$$E(\log\Big(\frac{m_{\theta_0^*}(z_i)f_H^*(z_i)}{m_\theta(z_i)f_H(z_i)}\Big)) \geq 0$$

Similiarly,

$$\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\log\Big(\frac{(1-m_{\theta_0^*}(z_i))f_H(z_i)}{(1-m_\theta(z_i))f_H(z_i)}\Big) \to (1-\delta_i)E(\log\Big(\frac{(1-m_{\theta_0^*}(z_i))f_H(z_i)}{(1-m_\theta(z_i))f_H(z_i)}\Big)) \geq 0$$

Therefore, $l_{\theta_0^*} \geq l_\theta$ for any other θ that is not the true θ_0^*.

The true θ_0^* maximizes the likelihood function.

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

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

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

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:

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

$I_n(\theta)$ is the Fisher information, which is the negative second derivative of log likelihood function:

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

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,

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

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

For the MLE $\hat{\theta}_n$, $u_n(\hat{\theta}_n) = 0$ Therefore,

$$(\hat{\theta}_n - \theta_0) = -\frac{u_n(\theta_0)}{u_n'(\theta_0)} = u_n(\theta_0) I_n^{-1}(\theta_0)$$