Stratified v.s. Unstratified Analysis

Chao Cheng

September 3, 2025

Contents

1 Introduction

In this note we will talk about the Cox's proportional hazards (Cox's PH) model. And more specifically, what will happen when unstratified analysis is used for a data where stratified analysis is the true model.

2 A simple parametric model

Consider the Weibull distribution, denote $T \sim W(p, \lambda)$. Then

$$
\begin{aligned}
&f(t) = p\lambda^p t^{p-1} \exp\left(-(\lambda t)^p\right) \\
&F(t) = 1 - \exp\left(-(\lambda t)^p\right) \qquad S(t) = \exp\left(-(\lambda t)^p\right) \\
&h(t) = p\lambda^p t^{p-1} \\
&H(t) = (\lambda t)^p \\
&\mathrm{E}(T) = \frac{1}{\lambda} \cdot \Gamma\left(1 + \frac{1}{p}\right) \qquad \mathrm{Var}(T) = \frac{1}{\lambda^2}\left(\Gamma\left(1 + \frac{2}{p}\right) - \Gamma\left(1 + \frac{1}{p}\right)\right) \\
&\mathrm{E}(T^m) = \frac{1}{\lambda^m}\Gamma\left(1 + \frac{m}{p}\right)
\end{aligned}
$$

Then the likelihood is

$$
\begin{aligned}
L(t_1, \cdots, t_n | p, \lambda_1, \cdots, \lambda_n) &= \prod_{i=1}^{n} \left(f(t_i)^{\delta_i} \left(1 - F(t_i)\right)^{1-\delta_i}\right) = \prod_{i=1}^{n} \left(h(t_i)^{\delta_i} S(t_i)\right) \\
&= \prod_{i=1}^{n} \left(p\lambda_i^p t_i^{p-1}\right)^{\delta_i} \exp\left(-(\lambda_i t_i)^p\right)
\end{aligned} \tag{1}
$$

where $\delta_i = 1$ means an event is observed for i. Otherwise $\delta_i = 0$ represents censor is observed. Note that in (1), we assume all subjects share the same p in the Weibull distribution, but their λs can be different.

3 Cox model

For a Cox model, the key assumption is constant hazard ratio, that is

$$h(t|Z) = h_0(t) \cdot e^{\beta Z}$$

And if we plug-in the Weibull distribution, that is $h_0(t) = p\lambda_0^p t^{p-1}$. Then

$$h(t|Z) = h_0(t) \cdot e^{\beta Z} = p\lambda_0^p t^{p-1} \cdot e^{\beta Z} = p\left(\lambda_0 e^{\beta Z/p}\right)^p t^{p-1}$$

Here for our purpose, we let $Z_i \in \{0,1\}$ denote the treatment(1) or control(0) group. And in this case, the data likelihood (1) becomes

$$\begin{aligned} L(t_1, \cdots, t_n | p, \lambda, \beta) &= \prod_{i=1}^{n} \left(p\left(\lambda e^{\beta Z_i/p}\right)^p t_i^{p-1}\right)^{\delta_i} \exp\left(-\left(\lambda e^{\beta Z_i/p} t_i\right)^p\right) \\ &= \prod_{i=1}^{n} \left(p\lambda^p e^{\beta Z_i} t_i^{p-1}\right)^{\delta_i} \exp\left(-\left(\lambda t_i\right)^p e^{\beta Z_i}\right) \end{aligned}$$

And the loglikelihood is

$$\begin{aligned} \log L &= \sum_{i=1}^{n} \delta_i \left(\log p + p\log\lambda + \beta Z_i + (p-1)\, t_i\right) - (\lambda t_i)^p e^{\beta Z_i} \\ &= n_{evt}\left(\log p + p\log\lambda\right) + \sum_{i=1}^{n} \delta_i \left(\beta Z_i + (p-1)\, t_i\right) - \lambda^p \sum_{i=1}^{n} t_i^p e^{\beta Z_i} \end{aligned} \tag{2}$$

Use the profile likelihood method, first we fix β and p to maximize $\log L$ w.r.t λ:

$$\frac{\partial \log L}{\partial \lambda} = \frac{n_{evt} p}{\lambda} - p\lambda^{p-1} \sum_{i=1}^{n} t_i^p e^{\beta Z_i}$$

Set this to 0 we have

$$\hat{\lambda} = \left(\frac{n_{evt}}{\sum_{i=1}^{n} t_i^p e^{\beta Z_i}}\right)^{1/p}$$

Plug this back into (2) will give us

$$\begin{aligned} \log L &= n_{evt}\left(\log p + \log n_{evt} - \log\left(\sum_{i=1}^{n} t_i^p e^{\beta Z_i}\right)\right) + \sum_{i=1}^{n} \delta_i \left(\beta Z_i + (p-1)\, t_i\right) - \frac{n_{evt}}{\sum_{i=1}^{n} t_i^p e^{\beta Z_i}} \cdot \sum_{i=1}^{n} t_i^p e^{\beta Z_i} \\ &= n_{evt}\left(\log p + \log n_{evt} - \log\left(\sum_{i=1}^{n} t_i^p e^{\beta Z_i}\right)\right) + \sum_{i=1}^{n} \delta_i \left(\beta Z_i + (p-1)\, t_i\right) - n_{evt} \end{aligned}$$

Unfortunately, there's no analytical solution to p even when we fixed β. So let's consider a simpler case where we fix $p = 1$, i.e. Exponential distribution. Then this loglikelihood becomes

$$\begin{aligned}
\log L &= n_{evt}\left(\log n_{evt} - \log\left(\sum_{i=1}^{n} t_i e^{\beta Z_i}\right)\right) + \sum_{i=1}^{n}\delta_i \beta Z_i - n_{evt} \\
&\overset{w.r.t\ \beta}{\propto} -n_{evt}\log\left(\sum_{i=1}^{n} t_i e^{\beta Z_i}\right) + \sum_{i=1}^{n}\delta_i \beta Z_i \\
&= \sum_{i=1}^{n}\delta_i\left(\beta Z_i - \log\left(\sum_{i=1}^{n} t_i e^{\beta Z_i}\right)\right) \\
&= \sum_{i=1}^{n}\delta_i \log\frac{e^{\beta Z_i}}{\sum\limits_{i=1}^{n} t_i e^{\beta Z_i}}
\end{aligned}$$

Therefore to maximize $\log L$ with respect to β, is equivalent to maximize the following term

$$\prod_{i=1}\left(\frac{e^{\beta Z_i}}{\sum\limits_{i=1}^{n} t_i e^{\beta Z_i}}\right)^{\delta_i} \tag{3}$$

Note: (3) can be seen as objective function for β's MLE and it is **different** from the partial likelihood used in Cox regression.

3.1 Stratified setting

Now let's consider the stratified setting, with K strata. Then for each stratum $k \in \{1, \cdots, K\}$, the Weibull distribution for control group is $W(p_k, \lambda_k)$, which means the hazard is

$$h_{0,k}(t) = p_k \lambda_k^{p_k} t^{p_k - 1}.$$

Assume the constant hazard ratio is e^{β}. Then the hazard for treatment group is $h_{1,k}(t) = h_{0,k}(t)\, e^{\beta}$. Therefore

$$h_k(t|Z) = h_{0,k}(t)\, e^{\beta Z}, \quad Z \in \{0, 1\}.$$

References