Stratified v.s. Unstratified Analysis

Chao Cheng

September 4, 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

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

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

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

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

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

Plug this back into (2) will give us

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

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

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

We can take derivative of $\log L$ and try to solve for β by setting the derivative to 0:

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

where we use the notation that $Z_i \in \{0, 1\}$ to indicate control and treatment group and $\delta_i \in \{0, 1\}$ to indicate observed censor or event. Setting $\frac{\partial \log L}{\partial \beta} = 0$ we have the MLE:

$$\hat{\beta} = \log\left(\frac{n_{evt\,\&\,trt}}{n_{evt\,\&\,ctrl}}\cdot\frac{\sum\limits_{i\in\ ctrl} t_i}{\sum\limits_{i\in\ trt} t_i}\right) \tag{4}$$

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

$$\prod_{i=1}^{n}\left(\frac{e^{\beta Z_i}}{\sum\limits_{\{j|t_j\geq t_i\}} e^{\beta Z_j}}\right)^{\delta_i}$$

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

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

In this case, assume sample size in each stratum are $n_1, \cdots, n_K$. Then the observed time is

$$\begin{aligned}
\boldsymbol{t} &= (t_{i,k},\ i \in \{1, \cdots, n_k\}\ k \in \{1, \cdots, K\})^T \\
&= \left(\boldsymbol{t}_1^T, \boldsymbol{t}_2^T, \cdots, \boldsymbol{t}_K^T\right)^T \\
&= (t_{1,1}, \cdots, t_{n_1,1}; t_{1,2}, \cdots, t_{n_2,2}; \cdots; t_{1,K}, \cdots, t_{n_K,K})^T
\end{aligned}$$

The corresponding covariate (treatment allocation) vector is

$$\begin{aligned}
\boldsymbol{Z} &= \left(\boldsymbol{Z}_1^T, \cdots, \boldsymbol{Z}_K^T\right)^T \\
&= (z_{1,1}, \cdots, z_{n_1,1}; z_{1,2}, \cdots, z_{n_2,2}; \cdots; z_{1,K}, \cdots, z_{n_K,K})^T
\end{aligned}$$

The parameter vectors are

$$\begin{aligned}
\boldsymbol{p} &= (p_1, \cdots, p_K)^T \\
\boldsymbol{\lambda} &= (\lambda_1, \cdots, \lambda_K)^T
\end{aligned}$$

And the data likelihood is

$$\begin{aligned}
L(\boldsymbol{t}; \boldsymbol{Z}|\boldsymbol{p}, \boldsymbol{\lambda}, \beta) &= \prod_{k=1}^{K} L(\boldsymbol{t}_k; \boldsymbol{Z}_k|p_k, \lambda_k, \beta) \\
&= \prod_{k=1}^{K} \prod_{i=1}^{n_k} \left(p_k \lambda_k^{p_k} e^{\beta Z_{i,k}} t_{i,k}^{p_k - 1}\right)^{\delta_{i,k}} \exp\left(-(\lambda_k t_{i,k})^{p_k} e^{\beta Z_{i,k}}\right)
\end{aligned}$$

And the partial likelihood would be

$$\begin{aligned}
L_1(\boldsymbol{t}; \boldsymbol{Z}|\boldsymbol{p}, \boldsymbol{\lambda}, \beta) &= \prod_{k=1}^{K} L_1(\boldsymbol{t}_k; \boldsymbol{Z}_k|p_k, \lambda_k, \beta) \\
&= \prod_{k=1}^{K} \prod_{i=1}^{n_k} \left(\frac{e^{\beta Z_{i,k}}}{\sum_{\left\{j \middle| t_{j,k} \geq t_{i,k}\right\}} e^{\beta Z_{j,k}}} \right)^{\delta_{i,k}}
\end{aligned}$$

3.2 Unstratified analysis under stratified setting

We use $f_k(t)$, $F_k(t)$, $S_k(t)$ and $h_k(t)$ to denote the p.d.f, c.d.f, survival function and hazard function of stratum k. Then we have marginally, the c.d.f of event time

$$F(t) = P(T \leq t) = \sum_{k=1}^{K} P(T \leq t, \quad \text{T from stratum k}) = \sum_{k=1}^{K} \pi_k F_k(t)$$

where $\pi_k = P\left(\text{T from stratum k}\right)$ denotes the probability that the subject comes from stratum k. Therefore

$$
\begin{aligned}
f\left(t\right) &= \frac{\partial F}{\partial t} = \sum_{k=1}^{K} \pi_k f_k\left(t\right) \\
S\left(t\right) &= 1 - F\left(t\right) = 1 - \sum_{k=1}^{K} \pi_k F_k\left(t\right) = \sum_{k=1}^{K} \pi_k S_k\left(t\right) \\
h\left(t\right) &= \frac{f\left(t\right)}{S\left(t\right)} = \frac{\sum\limits_{k=1}^{K} \pi_k f_k\left(t\right)}{\sum\limits_{k=1}^{K} \pi_k S_k\left(t\right)}
\end{aligned}
$$

So for a subject from control group, the hazard function (still based on Weibull distribution) is

$$
h_0\left(t\right) = \frac{\sum\limits_{k=1}^{K} \pi_k p_k \lambda_k^{p_k} t^{p_k-1} \exp\left(-\left(\lambda_k t\right)^{p_k}\right)}{\sum\limits_{k=1}^{K} \pi_k \exp\left(-\left(\lambda_k t\right)^{p_k}\right)} \tag{5}
$$

And the hazard function for treatment group is

$$
h_1\left(t\right) = \frac{\sum\limits_{k=1}^{K} \pi_k p_k \lambda_k^{p_k} t^{p_k-1} e^{\beta} \exp\left(-\left(\lambda_k t\right)^{p_k} e^{\beta}\right)}{\sum\limits_{k=1}^{K} \pi_k \exp\left(-\left(\lambda_k t\right)^{p_k} e^{\beta}\right)} \tag{6}
$$

Therefore the **Hazard Ratio** is

$$
hr\left(t\right) = \frac{h_1\left(t\right)}{h_0\left(t\right)} = \frac{\frac{\sum_{k=1}^{K} \pi_k p_k \lambda_k^{p_k} t^{p_k-1} e^{\beta} \exp\left(-\left(\lambda_k t\right)^{p_k} e^{\beta}\right)}{\sum_{k=1}^{K} \pi_k \exp\left(-\left(\lambda_k t\right)^{p_k} e^{\beta}\right)}}{\frac{\sum_{k=1}^{K} \pi_k p_k \lambda_k^{p_k} t^{p_k-1} \exp\left(-\left(\lambda_k t\right)^{p_k}\right)}{\sum_{k=1}^{K} \pi_k \exp\left(-\left(\lambda_k t\right)^{p_k}\right)}}, \tag{7}
$$

which is **NOT CONSTANT** w.r.t. t, meaning the constant hazard ratio assumption for Cox regression is violated.

Even for a simple case, where we assume $p_k = 1, k = 1, \cdots, K$, we still have

$$
hr(t) = \frac{\frac{\sum_{k=1}^{K} \pi_k \lambda_k e^{\beta} \exp\left(-\lambda_k t e^{\beta}\right)}{\sum_{k=1}^{K} \pi_k \exp\left(-\lambda_k t e^{\beta}\right)}}{\frac{\sum_{k=1}^{K} \pi_k \lambda_k \exp(-\lambda_k t)}{\sum_{k=1}^{K} \pi_k \exp(-\lambda_k t)}} = \frac{\left(\sum\limits_{k=1}^{K} \pi_k \exp\left(-\lambda_k t\right)\right)}{\left(\sum\limits_{k=1}^{K} \pi_k \lambda_k \exp\left(-\lambda_k t\right)\right)} \cdot \frac{\left(\sum\limits_{k=1}^{K} \pi_k \lambda_k \exp\left(-\lambda_k t e^{\beta}\right)\right)}{\left(\sum\limits_{k=1}^{K} \pi_k \exp\left(-\lambda_k t e^{\beta}\right)\right)} \cdot e^{\beta}, \tag{8}
$$

which again, is **NOT CONSTANT** w.r.t. t. Further more

$$
\begin{aligned}
&\frac{\partial}{\partial t}\frac{\sum_{k=1}^{K}\pi_k \exp\left(-\lambda_k t\right)}{\sum_{k=1}^{K}\pi_k\lambda_k \exp\left(-\lambda_k t\right)}\\
=&\frac{\left(\sum_{k=1}^{K}\pi_k \exp\left(-\lambda_k t\right)\left(-\lambda_k\right)\right)\sum_{k=1}^{K}\pi_k\lambda_k \exp\left(-\lambda_k t\right)-\sum_{k=1}^{K}\pi_k \exp\left(-\lambda_k t\right)\left(\sum_{k=1}^{K}\pi_k\lambda_k \exp\left(-\lambda_k t\right)\left(-\lambda_k\right)\right)}{\left(\sum_{k=1}^{K}\pi_k\lambda_k \exp\left(-\lambda_k t\right)\right)^2}\\
=&\frac{\left(\sum_{k=1}^{K}\pi_k \exp\left(-\lambda_k t\right)\right)\left(\sum_{k=1}^{K}\pi_k\lambda_k^2 \exp\left(-\lambda_k t\right)\right)-\left(\sum_{k=1}^{K}\pi_k\lambda_k \exp\left(-\lambda_k t\right)\right)^2}{\left(\sum_{k=1}^{K}\pi_k\lambda_k \exp\left(-\lambda_k t\right)\right)^2}\\
\geq& 0
\end{aligned}
$$

This is greater than or equal to 0 due to Cauchy's inequality. Therefore

$$
\begin{aligned}
\frac{\sum_{k=1}^{K}\pi_k \exp\left(-\lambda_k te^{\beta}\right)}{\sum_{k=1}^{K}\pi_k\lambda_k \exp\left(-\lambda_k te^{\beta}\right)} &\geq \frac{\sum_{k=1}^{K}\pi_k \exp\left(-\lambda_k t\right)}{\sum_{k=1}^{K}\pi_k\lambda_k \exp\left(-\lambda_k t\right)} \qquad \text{if } e^{\beta}>1\\
\frac{\sum_{k=1}^{K}\pi_k \exp\left(-\lambda_k te^{\beta}\right)}{\sum_{k=1}^{K}\pi_k\lambda_k \exp\left(-\lambda_k te^{\beta}\right)} &\leq \frac{\sum_{k=1}^{K}\pi_k \exp\left(-\lambda_k t\right)}{\sum_{k=1}^{K}\pi_k\lambda_k \exp\left(-\lambda_k t\right)} \qquad \text{if } 0<e^{\beta}<1
\end{aligned}
$$

which means for (8) we know that

$$
\begin{aligned}
hr(t) &\leq e^{\beta} \qquad \text{if } e^{\beta}>1\\
hr(t) &\geq e^{\beta} \qquad \text{if } 0<e^{\beta}<1
\end{aligned}
$$

And this implies that using unstratified analysis under a stratified setting might lead to biased estimation of hazard ratio.

References