Lecture seven: Cox Proportional Hazards Models (II)

1. **Maximum Likelihood Estimation (MLE)**

   (a) MLE

      i. The likelihood function

      Let $t_1, \ldots, t_n$ be a simple random sample (iid) from pdf $f(t, \beta)$, the likelihood function is

$$L(\beta) = \prod_{i=1}^{n} f(t_i, \beta)$$

      ii. The score equations

$$u(\hat{\beta}_j) \equiv \frac{d\, log\, L(\beta)}{d\beta_j}|_{\hat{\beta}} = 0$$

      iii. Hessian matrix $H(\beta)$: The (j,k)th element of $H(\hat{\beta})$ is the second derivative of the log-likelihood function:

$$\frac{\partial^2 log\, L(\hat{\beta})}{\partial\beta_j \partial\beta_k},$$

      iv. The matrices

      The *observed information matrix*:

$$I(\hat{\beta}) = -H(\hat{\beta})$$

      The *expected information matrix*:

$$EI(\hat{\beta}) \equiv E(I(\hat{\beta}))$$

      v. MLE $\hat{\beta}$ and $var(\hat{\beta})$

      Under some regular conditions, the asymptotic variance of MLE $\hat{\beta}$ is $EI^{-1}(\hat{\beta})$, but usually use $I^{-1}(\hat{\beta})$ instead.

vi. The tests

The likelihood ratio:

$$2\{log\, L(\hat{\beta}) - log\, L(0)\},$$

The Wald test:

$$\hat{\beta}' I(\hat{\beta})\hat{\beta},$$

The score test:

$$u'(0)I(0)^{-1}u(0),$$

Each of the three statistics has a $\chi^2(p)$ distribution under the null hypothesis that $\beta = 0$.

vii. Example: Exponential distribution: $Exp(\lambda)$

2. MLE and MPLE in survival setting

(a) Parametric model

When construct the likelihood function, we have to take censoring information (partially observed survival times) into account. The likelihood function

$$L(\beta) = \prod_{i=1}^{r} f(t_{(i)}) \prod_{j=1}^{n-r} S(t_j^*) = L(\beta) = \prod_{i=1}^{n} f^{\delta_i}(t_i) S^{1-\delta_i}(t_i)$$

(b) Semi-parametric model (Cox model): partial likelihood (PL)

i. Assume no ties at each death time (only one dies at each death time)

$$PL(\beta) = \prod_{j=1}^{r} \frac{exp(\beta' x_{(j)})}{\sum_{l \in R(t_{(j)})} exp(\beta' x_l)},$$

where $R(t_{(j)})$ is called the *risk set*: individuals who are alive and uncensored at a time just prior $t_{(j)}$; Or

$$PL(\beta) = \prod_{i=1}^{n} [\frac{exp(\beta' x_i)}{\sum_{l \in R(t_i)} exp(\beta' x_l)}]^{\delta_i}.$$

The log-likelihood function is given by

$$logPL(\beta) = \sum_{i=1}^{n} \delta_i \{\beta' x_i - log \sum_{l \in R(t_i)} exp(\beta' x_l)\}.$$

ii. Illustration

Follow the arguments at section 3.3.1: Consider following conditional probability (only one subject failed among $R(t_{(j)})$ at time $t_{(j)}$)

$$P[\textit{individual with variables } x_{(j)} \textit{ dies at } t_{(j)} \mid \textit{one death at } t_{(j)}], \ ....$$

For more rigorous mathematical reasoning, see the sections in **Fleming** and **Harrington**'s book (page 11 and page 139).

iii. treatment of ties: see section 3.3.2. In PROC PHREG, there are 4 ways to handle ties: Breslow, Exact, Discrete and Efron.

(c) Notes for the Cox models:

i. MLE is unavailable since $h_0(t)$ is unspecified.

ii. PL depends on covariates, and the ranks of survival times, not on the actual values observed.

iii. Under certain regularity conditions, the MPLE is asymptotically unbiased, consistent and normal.

iv. The efficiency of MPLE is almost comparable with that of MLE as if the baseline $h_0(t)$ were specified.

3. Newton-Raphson method

(a) Only a few of score equations have closed form solution (true for many other type of equations).

(b) The Newton-Raphson procedure is one of numeric methods. The iterative procedure is

$$\hat{\beta}_{s+1} = \hat{\beta}_s + I^{-1}(\hat{\beta}_s)u(\hat{\beta}_s),$$

for $s = 0, 1, 2, ...$

(c) There are several convergence criteria for this procedure.

(d) Illustration: For one parameter case, apply Taylor series to the following score equation

$$u(\beta) \equiv \frac{\partial log\, PL(\beta)}{\partial \beta} = 0$$

4. Delta method: General case (more than one parameters).

5. Confidence intervals and hypothesis tests for the β's.

   (a) hazard ratios

   $$\{exp(\hat{\beta}\}^2 var(\hat{\beta}),$$

   by Delta method.

   (b) Example 3.1: Breast cancer study.

      i. SAS output

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
GROUP	1	0.90801	0.50092	3.2858	0.0699

Analysis of Maximum Likelihood Estimates

Variable	Hazard Ratio	95% Hazard Ratio Confidence Limits	
GROUP	2.479	0.929	6.618

      ii. SAS program: ex31.sas

```
options ls =80 nodate;
libname fu '../../sdata';
data work;
        set fu.hpa;
proc phreg;
        model survt*censor(0)=group /covb rl ties=BRESLOW;
run;
```

(c) Example 3.2: multiple myeloma study

i. Description of the study (Table 1.3 at p9)

AGE: age of the patient, SEX: sex of the patient
BUN: Blood Urea Nitrogen, CA: serum CAlcium
HB: Serum HaemogloBin, PC: percentage of Plasma Cells
BJ: Bence-Jones protein (0: absent, 1: present)

ii. Rescale covariates: interpretation and effect

iii. SAS output

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
AGE	1	-0.01936	0.02792	0.4806	0.4882
SEX	1	-0.25090	0.40229	0.3890	0.5328
BUN	1	0.02083	0.00593	12.3397	0.0004
CA	1	0.01312	0.13244	0.0098	0.9211
HB	1	-0.13524	0.06889	3.8537	0.0496
PC	1	-0.00159	0.00658	0.0587	0.8085
BJ	1	-0.64044	0.42669	2.2529	0.1334

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

Variable	Hazard Ratio	95% Hazard Ratio Confidence Limits	
AGE	0.981	0.929	1.036
SEX	0.778	0.354	1.712
BUN	1.021	1.009	1.033
CA	1.013	0.782	1.314
HB	0.874	0.763	1.000
PC	0.998	0.986	1.011
BJ	0.527	0.228	1.216

iv. SAS program: ex32.sas

```
options ls=80 nodate;
libname fu '../../sdata';
data work;
        set fu.myeloma;
/* in Table 1.3, 1 for male and 2 for female */
        if sex=1 then sex=0; else sex=1;
proc phreg;
        model survt*censor(0)= age sex bun ca hb pc bj /covb rl;
run;
```

v. Throw all covariates into the model? Model building strategies (next lecture).