Lecture 14

3 Coordinate descent for the elastic net Give the detailed coordinate descent algorithm for the elastic net.

$$\hat{\beta}^{enet}(\lambda,\alpha) = \arg\min_{b_0,\cdots,b_p} \left[\sum_{i=1}^{n}(y_i - b_0 - b_1 x_{i1} - \cdots - b_p x_{ip})^2\right] + \lambda \sum_{j=1}^{p}\{\alpha b_j^2 + (1-\alpha)|b_j|\}]$$

We first standardize the covariates: $n^{-1}\sum_{i=1}^{n} x_{ij} = 0$, $n^{-1}\sum_{i=1}^{n} x_{ij}^2 = 1$ $(j=1,2,\cdots,p)$

Then center the outcome: $n^{-1}\sum_{i=1}^{n} y_i = 0$

① Initialize $\hat{\beta}$

② Update $\hat{\beta}_j$ given all other coefficients.

Define: $r_{ij} := y_i - \sum_{k\neq j}\hat{\beta}_k x_{ik}$. Updating $\hat{\beta}_j$ is equivalent to minimizing: $\frac{1}{2n}\sum_{i=1}^{n}(r_{ij} - b_j x_{ij})^2 + \lambda[\alpha b_j^2 + (1-\alpha)|b_j|]$

Define: $\hat{\beta}_{j,0} = \frac{\sum_{i=1}^{n} x_{ij} r_{ij}}{\sum_{i=1}^{n} x_{ij}^2} = n^{-1}\sum_{i=1}^{n} x_{ij} r_{ij}$ as the ols fit of r_{ij}'s on x_{ij}'s

$$\frac{1}{2n}\sum_{i=1}^{n}(r_{ij} - b_j x_{ij})^2 = \frac{1}{2n}\sum_{i=1}^{n}(r_{ij} - \hat{\beta}_{j,0} x_{ij})^2 + \frac{1}{2n}\sum_{i=1}^{n} x_{ij}^2 (b_j - \hat{\beta}_{j,0})^2$$

$$= \text{constant} + \frac{1}{2}(b_j - \hat{\beta}_{j,0})^2$$

Then updating $\hat{\beta}_j$ is equivalent to minimizing:

$$r_j := \frac{1}{2}(b_j - \hat{\beta}_{j,0})^2 + \lambda\alpha b_j^2 + \lambda(1-\alpha)|b_j|, \quad \frac{\partial r_j}{\partial b_j} = b_j - \hat{\beta}_{j,0} + 2\lambda\alpha b_j + \lambda(1-\alpha)\text{sign}(b_j)$$

$$\frac{\partial^2 r_j}{\partial b_j^2} = 1 + 2\lambda\alpha > 0$$

i) $b_j > 0$, $\frac{\partial r_j}{\partial b_j} = (1+2\lambda\alpha) b_j + \lambda(1-\alpha) - \hat{\beta}_{j,0}$

$$\hat{\beta}_j = \begin{cases} \frac{\hat{\beta}_{j,0} - \lambda(1-\alpha)}{1+2\lambda\alpha}, & \text{if } \hat{\beta}_{j,0} \geq \lambda(1-\alpha) \\ 0 & \text{if } \hat{\beta}_{j,0} < \lambda(1-\alpha) \end{cases}$$

ii) $b_j < 0$, $\frac{\partial r_j}{\partial b_j} = (1+2\lambda\alpha) b_j + \lambda(\alpha-1) - \hat{\beta}_{j,0}$

$$\hat{\beta}_j = \begin{cases} \frac{\hat{\beta}_{j,0} + \lambda(1-\alpha)}{1+2\lambda\alpha}, & \text{if } \hat{\beta}_{j,0} \leq -\lambda(1-\alpha) \\ 0 & \text{if } \hat{\beta}_{j,0} > -\lambda(1-\alpha) \end{cases}$$

$$\therefore \hat{\beta}_j = \begin{cases} \dfrac{\hat{\beta}_{j,0} - \lambda(1-\alpha)}{1+2\lambda\alpha} & \text{if } \hat{\beta}_{j,0} \geq \lambda(1-\alpha) \\ 0 & \text{if } -\lambda(1-\alpha) < \hat{\beta}_{j,0} < \lambda(1-\alpha) \\ \dfrac{\hat{\beta}_{j,0} + \lambda(1-\alpha)}{1+2\lambda\alpha} & \text{if } \hat{\beta}_{j,0} > -\lambda(1-\alpha) \end{cases}$$

③ Iterative until convergence $\square$

Lecture 16

3 Difference-in-means with weights With a binary covariate x_i, show that the coefficient of x_i in the WLS of y_i on $(1, x_i)$ with weights w_i $(i = 1, \ldots, n)$ equals $\bar{y}_{w,1} - \bar{y}_{w,0}$, where

$$\bar{y}_{w,1} = \frac{\sum_{i=1}^n w_i x_i y_i}{\sum_{i=1}^n w_i x_i}, \quad \bar{y}_{w,0} = \frac{\sum_{i=1}^n w_i (1-x_i) y_i}{\sum_{i=1}^n w_i (1-x_i)}$$

are the weighted averages of the outcome under treatment and control, respectively.

$$\hat{\beta}_w = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y, \text{ where } \Sigma := \mathrm{diag}(w_i^{-1})_{i=1}^n, \ X := (1, x_i)$$

$$\therefore \hat{\beta}_w = \begin{pmatrix} \sum_{i=1}^n w_i & \sum_{i=1}^n w_i x_i \\ \sum_{i=1}^n w_i x_i & \sum_{i=1}^n w_i x_i \end{pmatrix}^{-1} \cdot \begin{pmatrix} \sum_{i=1}^n w_i y_i \\ \sum_{i=1}^n w_i x_i y_i \end{pmatrix} =: \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}$$

$$\begin{pmatrix} \sum_{i=1}^n w_i & \sum_{i=1}^n w_i x_i \\ \sum_{i=1}^n w_i x_i & \sum_{i=1}^n w_i x_i \end{pmatrix}^{-1} = \frac{1}{\sum_{i=1}^n w_i x_i \left(\sum_{i=1}^n w_i - \sum_{i=1}^n w_i x_i\right)} \begin{pmatrix} \sum_{i=1}^n w_i x_i & -\sum_{i=1}^n w_i x_i \\ -\sum_{i=1}^n w_i x_i & \sum_{i=1}^n w_i \end{pmatrix}$$

$$\therefore \hat{\beta}_1 = \frac{1}{\sum_{i=1}^n w_i x_i \left(\sum_{i=1}^n w_i - \sum_{i=1}^n w_i x_i\right)} \left(-\sum_{i=1}^n w_i y_i \cdot \sum_{i=1}^n w_i x_i + \sum_{i=1}^n w_i x_i y_i \cdot \sum_{i=1}^n w_i\right)$$

$$= \frac{\sum_{i=1}^n w_i x_i y_i \cancel{\left(\sum_{i=1}^n w_i - \sum_{i=1}^n w_i x_i\right)}}{\sum_{i=1}^n w_i x_i \cancel{\left(\sum_{i=1}^n w_i - \sum_{i=1}^n w_i x_i\right)}} - \frac{\cancel{\sum_{i=1}^n w_i x_i} \cdot \sum_{i=1}^n w_i (1-x_i) y_i}{\cancel{\sum_{i=1}^n w_i x_i} \cdot \sum_{i=1}^n (1-x_i) w_i}$$

$$= \frac{\sum_{i=1}^n w_i x_i y_i}{\sum_{i=1}^n w_i x_i} - \frac{\sum_{i=1}^n w_i (1-x_i) y_i}{\sum_{i=1}^n (1-x_i) w_i} = \bar{y}_{w,1} - \bar{y}_{w,0} \quad \square$$

Lecture 17

2 Two logistic regressions Given data $(x_i, z_i, y_i)_{i=1}^n$ where x_i denotes the covariates, z_i denotes the binary treatment, and y_i denotes the binary outcome. We can fit two separate logistic regressions:

$$\text{logit}\{\text{pr}(y_i = 1 \mid z_i = 1, x_i)\} = \gamma_1 + x_i^{\mathrm{T}}\beta_1$$

and

$$\text{logit}\{\text{pr}(y_i = 1 \mid z_i = 0, x_i)\} = \gamma_0 + x_i^{\mathrm{T}}\beta_0$$

with the treated and control data, respectively. We can also fit a joint logistic regression using the pooled data:

$$\text{logit}\{\text{pr}(y_i = 1 \mid z_i, x_i)\} = \alpha_0 + \alpha_z z_i + x_i^{\mathrm{T}}\alpha_x + z_i x_i^{\mathrm{T}}\alpha_{zx}.$$

Let hats denote MLEs, for example, $\hat{\gamma}_1$ is the MLE for γ_1. Find $(\hat{\alpha}_0, \hat{\alpha}_z, \hat{\alpha}_x, \hat{\alpha}_{zx})$ in terms of $(\hat{\gamma}_1, \hat{\beta}_1, \hat{\gamma}_0, \hat{\beta}_0)$.

By joint logistic regression, we can get:

$$\log\{\Pr(y_i=1 \mid z_i=1, x_i)\} = \alpha_0 + \alpha_z + x_i^T\alpha_x + x_i^T\alpha_{zx} = \alpha_0+\alpha_z + x_i^T(\alpha_x+\alpha_{zx})$$

$$\log\{\Pr(y_i=1 \mid z_i=0, x_i)\} = \alpha_0 + x_i^T\alpha_x$$

$$L(\gamma_1, \beta_1) = \prod_{i=1}^{n_1} \{\pi(x_i, \gamma_1, \beta_1)\}^{y_i}\{1-\pi(x_i,\gamma_1,\beta_1)\}^{1-y_i}$$

$$l(\gamma_1, \beta_1) = \sum_{i=1}^{n_1}\{y_i(\gamma_1 + x_i^T\beta_1) - \log(1+e^{\gamma_1+x_i^T\beta_1})\} \quad ①$$

$$l(\gamma_0, \beta_0) = \sum_{i=1}^{n_0}\{y_i(\gamma_0 + x_i^T\beta_0) - \log(1+e^{\gamma_0+x_i^T\beta_0})\} \quad ②$$

$$l(\alpha_0, \alpha_z, \alpha_x, \alpha_{zx}) = \sum_{i=1}^{n_1}\{y_i(\underbrace{\alpha_0 + \alpha_z}_{a_1} + x_i^T\underbrace{(\alpha_x+\alpha_{zx})}_{b_1}) - \log(1+e^{\underbrace{(\alpha_0+\alpha_z)}_{a_1}+x_i^T\underbrace{(\alpha_x+\alpha_{zx})}_{b_1}})\}$$

$$+ \sum_{i=1}^{n_0}\{y_i(\underbrace{\alpha_0}_{a_0} + x_i^T\underbrace{\alpha_x}_{b_0}) - \log(1+e^{\underbrace{\alpha_0}_{a_0} + x_i^T\underbrace{\alpha_x}_{b_0}})\}$$

$$:= \sum_{i=1}^{n_1}\{y_i(a_1 + x_i^Tb_1) - \log(1+e^{a_1+b_1})\} + \sum_{i=1}^{n_0}\{y_i(a_0+x_i^Tb_0) - \log(1+e^{a_0+x_i^Tb_0})\} \quad ③$$

Therefore ①+② is equivalent to ③, because $(\hat{\gamma}_0, \hat{\beta}_0, \hat{\gamma}_1, \hat{\beta}_1)$ maximum $L(\gamma, \beta)$, then $(\hat{\alpha}_0, \hat{\alpha}_x, \hat{\alpha}_0+\hat{\alpha}_z, \hat{\alpha}_x+\hat{\alpha}_{zx})$ also maximum $L(\alpha_0, \cdots, \alpha_{zx})$

$\therefore (\hat{\alpha}_0, \hat{\alpha}_z, \hat{\alpha}_x, \hat{\alpha}_{zx}) = (\hat{\gamma}_0, \hat{\gamma}_1 - \hat{\gamma}_0, \hat{\beta}_0, \hat{\beta}_1 - \hat{\beta}_0)$ □

Lecture 18

6 Case-control study and multinomial logistic model Assume that

$$\operatorname{pr}(y_i = k \mid x_i) = \frac{e^{\alpha_k + x_i^{\mathrm{T}}\beta_k}}{\sum_{k'=1}^{K} e^{\alpha_{k'} + x_i^{\mathrm{T}}\beta_{k'}}}$$

with $\alpha_K = 0$ and $\beta_K = 0$, and

$$\operatorname{pr}(s_i = 1 \mid y_i = k, x_i) = \operatorname{pr}(s_i = 1 \mid y_i = k) = p_k$$

for $k = 1, \ldots, K$. Show that

$$\operatorname{pr}(y_i = k \mid x_i, s_i = 1) = \frac{e^{\alpha_k + \log p_k + x_i^{\mathrm{T}}\beta_k}}{\sum_{k'=1}^{K} e^{\alpha_{k'} + \log p_{k'} + x_i^{\mathrm{T}}\beta_{k'}}}.$$

$$Pr(y_i=k \mid x_i, S_i=1) = \frac{Pr(y_i=k, S_i=1 \mid x_i)}{Pr(S_i=1 \mid x_i)}$$

$$= \frac{Pr(y_i=k, S_i=1 \mid x_i)}{\sum_{k=1}^{k} Pr(S_i=1 \mid y_i=k, x_i) \cdot Pr(y_i=k \mid x_i)}$$

$$= \frac{Pr(S_i=1 \mid x_i, y_i=k)\, Pr(y_i=k \mid x_i)}{\sum_{k=1}^{k} Pr(S_i=1 \mid y_i=k, x_i)\, Pr(y_i=k \mid x_i)}$$

$$= \frac{p_k \cdot \dfrac{e^{\alpha_k + x_i^T \beta_k}}{\cancel{\sum_{k'=1}^{k} e^{\alpha_{k'} + x_i^T \beta_{k'}}}}}{\sum_{k=1}^{k} p_k \cdot \dfrac{e^{\alpha_k + x_i^T \beta_k}}{\cancel{\sum_{k'=1}^{k} e^{\alpha_{k'} + x_i^T \beta_{k'}}}}} = \frac{e^{\alpha_k + \log p_k + x_i^T \beta_k}}{\sum_{k'=1}^{k} e^{\alpha_{k'} + \log p_{k'} + x_i^T \beta_{k'}}} \quad \square$$

Lecture 20

3 Negative-Binomial covariance matrices Assume that δ is known. Derive the estimated asymptotic covariance matrices of the MLE in the Negative-Binomial regression with $\mu_i = e^{x_i^{\mathrm{T}}\beta}$, one assuming a correctly specified model and the other allowing for misspefication of the model.

$y_i \mid x_i \sim NB(\mu_i, \delta)$, where $\mu_i = e^{x_i^T\beta}$

① a correctly speified model

$$\sum_{i=1}^{n} E\left(\frac{\partial l_i}{\partial \beta}\frac{\partial l_i}{\partial \beta^T}\Big| x_i\right) = \sum_{i=1}^{n} \frac{1}{\sigma_i^2}\{\mu'(x_i^T\beta)\}^2 x_i x_i^T$$

$$\mu'(x_i^T\beta) = e^{x_i^T\beta}, \quad \sigma_i^2 = \mu + \frac{\mu^2}{\delta} = \mu\left(1+\frac{\mu}{\delta}\right)$$

$$\therefore \sum_{i=1}^{n} E\left(\frac{\partial l_i}{\partial \beta}\frac{\partial l_i}{\partial \beta^T}\Big| x_i\right) = \sum_{i=1}^{n} \frac{\mu_i^{\cancel{2}}}{\cancel{\mu_i}\left(1+\frac{\mu_i}{\delta}\right)} x_i x_i^T = \sum_{i=1}^{n} \frac{e^{x_i^T\beta}\delta}{\delta + e^{x_i^T\beta}} x_i x_i^T$$

$$\therefore \hat{V} = \left(\sum_{i=1}^{n} \frac{\delta e^{x_i^T\beta}}{\delta + e^{x_i^T\beta}} x_i x_i^T\right)^{-1}$$

② misspeification of the model

$$\hat{B} = n^{-1}\sum_{i=1}^{n} \frac{1}{\tilde{\sigma}^2(x_i,\hat{\beta})}\frac{\partial \mu(x_i^T\hat{\beta})}{\partial \beta}\frac{\partial \mu(x_i^T\hat{\beta})}{\partial \beta^T}, \quad \tilde{\sigma}^2(x_i,\hat{\beta}) = e^{x_i^T\hat{\beta}}\left(1+\frac{e^{x_i^T\hat{\beta}}}{\delta}\right)$$

$$= n^{-1}\sum_{i=1}^{n} \frac{1}{\cancel{\mu}\left(1+\frac{\mu}{\delta}\right)}\mu^{\cancel{2}} x_i x_i^T = n^{-1}\sum_{i=1}^{n} \frac{e^{x_i^T\beta}\delta}{\delta + e^{x_i^T\beta}} x_i x_i^T$$

$$\hat{M} = n^{-1}\sum_{i=1}^{n}\left\{\frac{y_i - \mu(x_i^T\hat{\beta})}{\tilde{\sigma}^2(x_i,\hat{\beta})}\right\}^2 \frac{\partial \mu(x_i^T\hat{\beta})}{\partial \beta}\frac{\partial \mu(x_i^T\hat{\beta})}{\partial \beta^T}$$

$$= n^{-1}\sum_{i=1}^{n}\left\{\frac{y_i-\mu}{\mu\left(1+\frac{\mu}{\sigma}\right)}\right\}^2 \mu^2 x_i x_i^T = n^{-1}\sum_{i=1}^{n}\frac{\delta^2\hat{\varepsilon}_i^2}{(e^{x_i^T\hat{\beta}}+\delta)^2} x_i x_i^T.$$

where residual $\hat{\varepsilon}_i = y_i - e^{x_i^T\beta}$

$$\therefore \hat{V} = n^{-1}\hat{B}^{-1}M\hat{B}^{-1} = n^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\delta e^{x_i^T\hat{\beta}}}{\delta + e^{x_i^T\hat{\beta}}} x_i x_i^T\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\delta^2\hat{\varepsilon}_i^2}{(e^{x_i^T\hat{\beta}}+\delta)^2} x_i x_i^T\right) \cdot \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\delta e^{x_i^T\hat{\beta}}}{\delta + e^{x_i^T\hat{\beta}}} x_i x_i^T\right)^{-1}$$

$$= \left(\sum_{i=1}^{n}\frac{\delta e^{x_i^T\hat{\beta}}}{\delta + e^{x_i^T\hat{\beta}}} x_i x_i^T\right)^{-1}\left(\sum_{i=1}^{n}\frac{\delta^2\hat{\varepsilon}_i^2}{(\delta+e^{x_i^T\hat{\beta}})^2} x_i x_i^T\right)\left(\sum_{i=1}^{n}\frac{\delta e^{x_i^T\hat{\beta}}}{\delta + e^{x_i^T\hat{\beta}}} x_i x_i^T\right)^{-1} \quad \square$$

Lecture 21

4 Cluster-robust standard error for Poisson regression Similar to Sections 4.1 and 4.2, derive the cluster-robust standard error for Poisson regression.

$y_i \mid x_i \sim \text{Poisson}(\mu_i)$, while $\mu_i = e^{x_i^T \beta}$

GEE: $\sum_{i=1}^{n} \frac{\partial \mu^T(x_i,\beta)}{\partial \beta} \hat{V}^{-1}(x_i,\beta)\{Y_i - \mu(x_i,\beta)\} = 0$

$$\hat{B} = n^{-1} \sum_{i=1}^{n} D_i^T(\hat\beta) \tilde{V}^{-1}(x_i,\hat\beta) D_i(\hat\beta), \text{ where } \tilde{V}^{-1}(x_i,\hat\beta) = \begin{pmatrix} e^{x_{i1}^T\hat\beta} & & \\ & \ddots & \\ & & e^{x_{in_i}^T\hat\beta} \end{pmatrix}^{-1}$$

$$= n^{-1} \sum_{i=1}^{n} \left(e^{x_{i1}^T\hat\beta} x_{i1}, \cdots, e^{x_{in_i}^T\hat\beta} x_{in_i}\right) \begin{pmatrix} e^{x_{i1}^T\hat\beta} & & \\ & \ddots & \\ & & e^{x_{in_i}^T\hat\beta} \end{pmatrix}^{-1} \begin{pmatrix} e^{x_{i1}^T\hat\beta} x_{i1}^T \\ \vdots \\ e^{x_{in_i}^T\hat\beta} x_{in_i}^T \end{pmatrix}$$

$$= n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_i} e^{x_{ij}^T\hat\beta} x_{ij} x_{ij}^T$$

$$\hat{M} = n^{-1} \sum_{i=1}^{n} D_i^T(\beta) \hat{V}^{-1}(x_i,\beta) \hat\varepsilon_i \hat\varepsilon_i^T \tilde{V}^{-1}(x_i,\beta) D_i(\beta)\}, \text{ where } \hat\varepsilon_i = y_i - e^{x_i^T\hat\beta}$$

$$= n^{-1} \sum_{i=1}^{n} \left(e^{x_i^T\hat\beta} x_{i1}, \cdots, e^{x_{in_i}^T\hat\beta} x_{in_i}\right) \begin{pmatrix} e^{x_{i1}^T\hat\beta} & & \\ & \ddots & \\ & & e^{x_{in_i}^T\hat\beta} \end{pmatrix}^{-1} \begin{pmatrix} \hat\varepsilon_{i1} \\ \vdots \\ \hat\varepsilon_{in_i} \end{pmatrix} (\hat\varepsilon_{i1}, \cdots, \hat\varepsilon_{in_i})$$

$$\cdot \begin{pmatrix} e^{x_{i1}^T\hat\beta} & & \\ & \ddots & \\ & & e^{x_{in_i}^T\hat\beta} \end{pmatrix}^{-1} \begin{pmatrix} e^{x_i^T\hat\beta} x_{i1}^T \\ \vdots \\ e^{x_{in_i}^T\hat\beta} x_{in_i}^T \end{pmatrix}$$

$$= n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_i} x_{ij} \hat\varepsilon_{ij}^2 x_{ij}^T = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \hat\varepsilon_{ij}^2 x_{ij} x_{ij}^T$$

$$\hat{\text{cov}}(\hat\beta) = n^{-1} \hat{B}^{-1} \hat{M} \hat{B}^{-1} = \left(\sum_{i=1}^{n} \sum_{j=1}^{n_i} e^{x_{ij}^T\hat\beta} x_{ij} x_{ij}^T\right)^{-1} \left(\sum_{i=1}^{n} \sum_{j=1}^{n_i} \hat\varepsilon_{ij}^2 x_{ij} x_{ij}^T\right) \left(\sum_{i=1}^{n} \sum_{j=1}^{n_i} e^{x_{ij}^T\hat\beta} x_{ij} x_{ij}^T\right)^{-1} \square$$

Lecture 22

3 Quantile regression with a binary regressor For $i = 1, \dots, n$, the first $1/3$ observations have $x_i = 1$ and the last $2/3$ observations have $x_i = 0$; $y_i \mid x_i = 1$ follows an Exponential(1), and $y_i \mid x_i = 0$ follows an Exponential(2). Find

$$(\hat{\alpha}, \hat{\beta}) = \arg\min_{(a,b)} \sum_{i=1}^{n} \rho_{1/2}(y_i - a - bx_i).$$

and the joint asymptotic distribution.

$(\hat{\alpha}, \hat{\beta}) = \arg\min_{a,b} \sum_{i=1}^{n} \rho_{\frac{1}{2}}(y_i - a - bx_i) = \arg\min_{a,b} \left(\sum_{x_i=1} |y_i - a - b| + \sum_{x_i=0} |y_i - a| \right)$

$= \arg\min_{a} \sum_{x_i=0} |y_i - a| + \arg\min_{a+b} \sum_{x_i=1} |y_i - (a+b)|$, define $a+b := c$, $b = c - a$

$= \arg\min_{a} \sum_{x_i=0} |y_i - a| + \arg\min_{c} \sum_{x_i=1} |y_i - c|$

Because the first $1/3$ observations have $x_i = 1$ and last $2/3$ have $x_i = 0$

then we know: $\hat{\alpha} = \text{medium}(y_{\frac{n}{3}+1}, \cdots, y_n)$

$\hat{c} = \hat{\alpha} + \hat{\beta} = \text{medium}(y_1, \cdots, y_{\frac{n}{3}})$

$\therefore \hat{\alpha} = \text{medium}(y_{\frac{n}{3}+1}, \cdots, y_n)$,

$\hat{\beta} = \text{medium}(y_1, \cdots, y_{\frac{n}{3}}) - \text{medium}(y_{\frac{n}{3}+1}, \cdots, y_n)$

$\sqrt{n}(\hat{\beta}_\tau - \beta_\tau) \xrightarrow{d} N(0, \hat{B}^{-1}\hat{M}\hat{B}^{-1})$, where

$\hat{M} = n^{-1} \sum_{i=1}^{n} \{\tau - I(y_i - x_i^T \hat{\beta}(\tau) \le 0)\}^2 x_i x_i^T$, $x_i := \begin{pmatrix} 1 \\ x_i \end{pmatrix}$, where x_i is binary

$\hat{B} = n^{-1} \sum_{i=1}^{n} \{ f_{y_i|x_i}\{x_i^T \hat{\beta}(\tau)\} x_i x_i^T \} = n^{-1} \left(\sum_{i=1}^{\frac{n}{3}} f_{y_i|x_i=1}(\hat{\alpha} + \hat{\beta}) x_i x_i^T + \sum_{i=\frac{n}{3}+1}^{n} f_{y_i|x_i}(\hat{\alpha}) x_i x_i^T \right)$

$= n^{-1} \left(\sum_{i=1}^{n/3} e^{-(\hat{\alpha}+\hat{\beta})} x_i x_i^T + \sum_{i=\frac{n}{3}+1}^{n} 2e^{-2\hat{\alpha}} x_i x_i^T \right)$

$\hat{M} = n^{-1} \sum_{i=1}^{n} \{ \frac{1}{2} - I(y_i - \hat{\alpha} - \hat{\beta} x_i) \le 0 \}^2 x_i x_i^T$

$= n^{-1} \left(\sum_{x_i=0} \{ \frac{1}{2} - I(y_i - \underline{\hat{\alpha}}) \}^2 x_i x_i^T + \sum_{x_i=1} \{ \frac{1}{2} - I(y_i - \underline{(\hat{\alpha} + \hat{\beta})}) \}^2 x_i x_i^T \right)$

medium

medium

$= n^{-1} \left(\sum_{x_i=0} \frac{1}{4} x_i x_i^T + \sum_{x_i=1} \frac{1}{4} x_i x_i^T \right) = \frac{1}{4} n^{-1} \sum_{i=1}^{n} x_i x_i^T$

$\hat{B}^{-1}\hat{M}\hat{B}^{-1} = \left[n^{-1} \left(\sum_{i=1}^{n/3} e^{-(\hat{\alpha}+\hat{\beta})} x_i x_i^T + \sum_{i=\frac{n}{3}+1}^{n} 2e^{-2\hat{\alpha}} x_i x_i^T \right) \right]^{-1} \left(n^{-1} \frac{1}{4} \sum_{i=1}^{n} x_i x_i^T \right) [\cdots]^{-1}$

$$\sum_{i=1}^{n/3} e^{-(\hat\alpha+\hat\beta)} \underbrace{x_i x_i^T}_{\begin{pmatrix}1&1\\1&1\end{pmatrix}} + \sum_{i=\frac{n}{3}+1}^{n} 2e^{-2\hat\alpha} \underbrace{x_i x_i^T}_{\begin{pmatrix}1&0\\0&0\end{pmatrix}} = \begin{pmatrix} \frac{n}{3}e^{-(\hat\alpha+\hat\beta)}+\frac{4}{3}ne^{-2\hat\alpha} & \frac{n}{3}e^{-(\hat\alpha+\hat\beta)} \\ \frac{n}{3}e^{-(\hat\alpha+\hat\beta)} & \frac{n}{3}e^{-(\hat\alpha+\hat\beta)} \end{pmatrix} = \frac{n}{3}e^{-(\hat\alpha+\hat\beta)} \begin{pmatrix} 4e^{\hat\beta-\hat\alpha}+1 & 1 \\ 1 & 1 \end{pmatrix}$$

$$B^{-1} = \frac{1}{\frac{4}{3}e^{-2\hat\alpha}} \begin{pmatrix} 1 & -1 \\ -1 & 4e^{\hat\beta-\hat\alpha}+1 \end{pmatrix}$$

$$\sum_{i=1}^{n} x_i x_i^T = \begin{pmatrix} \frac{n}{3} & \frac{n}{3} \\ \frac{n}{3} & \frac{n}{3} \end{pmatrix} + \begin{pmatrix} \frac{2}{3}n & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} n & \frac{n}{3} \\ \frac{n}{3} & \frac{n}{3} \end{pmatrix}$$

$$\therefore B^{-1}MB^{-1} = \frac{3}{4}e^{2\hat\alpha}\begin{pmatrix} 1 & -1 \\ -1 & 4e^{\hat\beta-\hat\alpha}+1 \end{pmatrix} \cdot \frac{1}{4}\begin{pmatrix} 1 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} \end{pmatrix} \cdot \frac{3}{4}e^{2\hat\alpha}\begin{pmatrix} 1 & -1 \\ -1 & 4e^{\hat\beta-\hat\alpha}+1 \end{pmatrix}$$

$$= \frac{3^2}{4^3}e^{4\hat\alpha}\begin{pmatrix} \frac{2}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3}+\frac{16}{3}e^{2(\hat\beta-\hat\alpha)} \end{pmatrix} = \frac{3}{32}e^{4\hat\alpha}\begin{pmatrix} 1 & -1 \\ -1 & 1+8e^{2(\hat\beta-\hat\alpha)} \end{pmatrix}$$

$$\therefore \sqrt{n}\left(\begin{pmatrix}\hat\alpha\\ \hat\beta\end{pmatrix} - \begin{pmatrix}\alpha\\ \beta\end{pmatrix}\right) \xrightarrow{d} N\left(0,\ \frac{3}{32}e^{4\hat\alpha}\begin{pmatrix} 1 & -1 \\ -1 & 1+8e^{2(\hat\beta-\hat\alpha)} \end{pmatrix}\right) \square$$

Lecture 23

5 Invariance of the proportional hazards model Assume that $T \mid x$ follows a proportional hazards model, show that any non-negative and strictly increasing transformation $g(T) \mid x$ also follows a proportional hazards model.

$$\lambda(t|x) = \lim_{\Delta t \downarrow 0} Pr(t \le T < t+\Delta t \mid T \ge t, x)/\Delta t = \lambda_0(t)\exp(x^T\beta)$$

Because $g(\cdot)$ is non-negative and strictly increasing,

$g(T) := G$, then $T := g^{-1}(G)$, $\frac{\partial g^{-1}(G)}{\partial G}$ exists

$\lambda(G|X) = \frac{f(G|X)}{S(G|X)}$, where $S(G|X) = Pr(g(T) > g(t) \mid x) = Pr(T > t|X) = S(t|X)$

$f(G|X) = f(g^{-1}(G)|X)\frac{\partial g^{-1}(G)}{\partial G}$

$\therefore \lambda(G|X) = \lambda(g^{-1}(G)|X) \cdot \frac{\partial g^{-1}(G)}{\partial G} = \lambda_0(g^{-1}(G_0))\exp(X^T\beta)\frac{\partial g^{-1}(G)}{\partial G}$

define: $\frac{\partial g^{-1}(G)}{\partial G}\lambda_0(g^{-1}(G)) = \lambda_0^*(G)$

$\Rightarrow \lambda(G|X) = \lambda_0^*(G)\exp(X^T\beta)$

$\therefore$ $g(T)|X$ also follows a proportional hazard model $\square$