Lecture 18

2. From the proposition 1, we know covariance matrix:

$$\begin{pmatrix} \pi_1(1-\pi_1) & -\pi_1\pi_2 & \cdots & -\pi_1\pi_{k-1} \\ \vdots & & & \vdots \\ \vdots & - & - \; - \; - & \pi_{k-1}(1-\pi_{k-1}) \end{pmatrix} := \Sigma$$

is positive semi-definite, where any $x \in R^n$, $x^T \Sigma x \geq 0$

As for Hessian matrix H:

$$\frac{\partial^2 \log L(\beta)}{\partial \beta_k \partial \beta_k^T} = -\sum_{i=1}^{n} \pi_k(x_i,\beta)\{1-\pi_k(x_i,\beta)\} x_i x_i^T \quad k=1,\cdots,k-1$$

$$\frac{\partial^2 \log L(\beta)}{\partial \beta_k \partial \beta_{k'}^T} = \sum_{i=1}^{n} \pi_k(x_i,\beta)\pi_{k'}(x_i,\beta) x_i x_i^T \quad k\neq k',\ k'=1,\cdots,k-1$$

$$\forall z \in R^{[p\times(k-1)]\times 1} \quad z = \begin{bmatrix} z_1 \\ \vdots \\ \vdots \\ z_{k-1} \end{bmatrix} \begin{matrix} \text{-- } p \\ \\ \\ \text{-- } p \end{matrix} \qquad H = \frac{\partial^2 \log L(\beta)}{\partial \beta \partial \beta^T}$$

$$z^T H z = \sum_{m=1}^{p} z_m^T \frac{\partial^2 \log L(\beta)}{\partial \beta_m \partial \beta_m^T} z_m + \sum_{m\neq h} z_h^T \frac{\partial^2 \log L(\beta)}{\partial \beta_h \partial \beta_m^T} z_m$$

$$= -\sum_{m=1}^{p} \left(\sum_{i=1}^{n} \pi_m(x_i,\beta)\{1-\pi_m(x_i,\beta)\} z_m^T x_i x_i^T z_m \right) + \sum_{m\neq h} \left(\sum_{i=1}^{n} \pi_m(x_i,\beta)\pi_h(x_i,\beta) z_m^T x_i x_i^T z_h \right)$$

$$= \sum_{i=1}^{n} \left\{ -\sum_{m=1}^{p} \pi_m(x_i,\beta)\{1-\pi_m(x_i,\beta)\} z_m^T x_i x_i^T z_m + \sum_{m\neq h} \pi_m(x_i,\beta)\pi_h(x_i,\beta) z_m^T x_i x_i^T z_h \right\}$$

Now we define:

$$Y_i^T := (z_1^T x_i, z_2^T x_i, \cdots, z_{k-1} x_i)$$

$$\Sigma_i := \begin{pmatrix} \pi_1(x_i,\beta)\{1-\pi_1(x_i,\beta)\} & -\pi_1(x_i,\beta)\pi_2(x_i,\beta) & \cdots & \pi_1(x_i,\beta)\pi_{k-1}(x_i,\beta) \\ \vdots & \ddots & & \vdots \\ \pi_{k-1}(x_i,\beta)\{1-\pi_{k-1}(x_i,\beta)\} & - \; - \; - & & \pi_{k-1}(x_i,\beta)\{1-\pi_{k-1}(x_i,\beta)\} \end{pmatrix}$$

From proposition 1, we know Σ_i is positive semi-definite,

thus $\gamma_i^T \Sigma_i \gamma_i \geq 0$

$$\Rightarrow \sum_{m=1}^{p} \pi_m(x_i,\beta)\{1-\pi_m(x_i,\beta)\} z_m^T x_i x_i^T m -$$

$$\sum_{m \neq h} \pi_m(x_i,\beta)\pi_h(x_i,\beta) z_m^T x_i x_i^T z_h \geq 0$$

$$\Rightarrow z^T H z = -\sum_{i=1}^{n}\Big\{\sum_{m=1}^{p} \pi_m(x_i,\beta)\{1-\pi_m(x_i,\beta)\} z_m^T x_i x_i^T m -$$

$$\sum_{m \neq h} \pi_m(x_i,\beta)\pi_h(x_i,\beta) z_m^T x_i x_i^T z_h\Big\} \leq 0$$

Therefore, H, the Hessian Matrix in multinomial logit model, is negative semi-definite.

Lecture 18

3 For Newton's Method, we know:

$$\beta^{new} = \beta^{old} - \left\{\frac{\partial^2 \log L(\beta^{old})}{\partial\beta\partial\beta^T}\right\}^{-1} \frac{\partial \log L(\beta^{old})}{\partial \beta}$$

$$\frac{\partial \log L(\beta)}{\partial \beta} = \begin{pmatrix} \frac{\partial \log L(\beta)}{\partial \beta_1} \\ \vdots \\ \frac{\partial \log L(\beta)}{\partial \beta_{k-1}} \end{pmatrix}, \quad \frac{\partial \log L(\beta)}{\partial \beta_k} = \sum_{i=1}^{n} x_i \{ I(y_i = k) - \pi_k(x_i, \beta) \} = X^T(1_k - \pi_k)$$

define: $1_k := \begin{pmatrix} I(y_1=k) \\ \vdots \\ I(y_n=k) \end{pmatrix}$, $\pi_k := \begin{pmatrix} \pi_k(x_1, \beta) \\ \vdots \\ \pi_k(x_n, \beta) \end{pmatrix}$, $1_y := \begin{pmatrix} 1_1 \\ \vdots \\ 1_{k-1} \end{pmatrix}$, $\pi := \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_{k-1} \end{pmatrix}$

Thus, $\frac{\partial \log L(\beta)}{\partial \beta} = (I_{k-1} \otimes X^T)(1_y - \pi) \Rightarrow \frac{\partial \log L(\beta^{old})}{\partial \beta} = (I_{k-1} \otimes X^T)(1_y - \pi^{old})$

$$\frac{\partial^2 \log L(\beta)}{\partial \beta_k \partial \beta_k^T} = -\sum_{i=1}^{n} \pi_k(x_i, \beta)\{1 - \pi_k(x_i, \beta)\} x_i x_i^T = -X^T W_{k,k} X$$

$$\frac{\partial^2 \log L(\beta)}{\partial \beta_k \partial \beta_{k'}^T} = \sum_{i=1}^{n} \pi_k(x_i, \beta)\pi_{k'}(x_i, \beta) x_i x_i^T = X^T W_{k,k'} X$$

define: $W_{k,k} = \mathrm{diag}[\pi_k(x_i, \beta)\{1 - \pi_k(x_i, \beta)\}]_{i=1}^{n}$

$W_{k,k'} = \mathrm{diag}[\pi_k(x_i, \beta)\pi_{k'}(x_i, \beta)]_{i=1}^{n} \quad k \neq k'$

$$W := \begin{pmatrix} W_{1,1} & \cdots & -W_{1,k-1} \\ \vdots & \ddots & \vdots \\ -W_{k-1,1} & \cdots & W_{k-1,k-1} \end{pmatrix}$$

$$\frac{\partial^2 \log L(\beta)}{\partial \beta \partial \beta^T} = -(I_{k-1} \otimes X^T) W (I_{k-1} \otimes X)$$

$$\Rightarrow \frac{\partial^2 \log L(\beta^{old})}{\partial \beta \partial \beta^T} = -(I_{k-1} \otimes X^T) W^{old} (I_{k-1} \otimes X) = -(I_{k-1} \otimes X)^T W^{old} (I_{k-1} \otimes X)$$

$$\therefore \beta^{new} = \beta^{old} + [(I_{k-1} \otimes X)^T W^{old} (I_{k-1} \otimes X)]^{-1} (I_{k-1} \otimes X)^T (1_y - \pi^{old})$$

$$\therefore \beta^{new} = \beta^{old} + (\hat{X}^T W^{old} \hat{X})^{-1} \hat{X}^T (1_y - \pi^{old}),$$

where $\hat{X} := (I_{k-1}) \otimes X$

Lecture 19

7. $$y_i|x_i \sim \begin{cases} 0 & \text{with prob } p_i \\ \text{poisson}(\lambda_i) & \text{with prob } 1-p_i \end{cases}$$

where $p_i = \dfrac{e^{x_i^T r}}{1+e^{x_i^T r}}$, $\lambda_i = e^{x_i^T \beta} \Rightarrow \dfrac{\partial p_i}{\partial r} = p_i(1-p_i)x_i$, $\dfrac{\partial \lambda_i}{\partial \beta} = \lambda_i x_i$

$$P(y_i = j|x_i) = \begin{cases} p_i + (1-p_i)\exp(-\lambda_i) & \text{if } j=0 \\ (1-p_i)\dfrac{\lambda_i^{y_i}\exp(-\lambda_i)}{y_i!} & \text{if } j>0 \end{cases}$$

To simplify the calculation, we modify $P(y_i = j|x_i)$:

$$P^*(y_i = 1|x_i) = \begin{cases} p_i & \text{if } j=0 \\ (1-p_i)\dfrac{\lambda_i^{y_i}\exp(-\lambda_i)}{y_i!} & \text{if } j>0 \end{cases}$$

$$L(\beta, r) = \prod_{y=0} p_i \prod_{y>0}\left[(1-p_i)\frac{\lambda_i^{y_i}\exp(-\lambda_i)}{y_i!}\right]$$

$$\log L(\beta, r) = \sum_{y=0} \log p_i + \sum_{y>0}\left\{\log(1-p_i) + y_i \log(\lambda_i) - \lambda_i\right\}$$

$$\frac{\partial \log L(\beta, r)}{\partial \beta} = \sum_{y>0}\left(\frac{y_i}{\lambda_i} - 1\right)\lambda_i x_i = \sum_{y>0}\left(y_i - e^{x_i^T\beta}\right)x_i$$

$$\frac{\partial \log L(\beta, r)}{\partial r} = \sum_{y=0}\frac{1}{p_i}\cdot p_i(1-p_i)x_i + \sum_{y>0}\left(-\frac{1}{1-p_i}\right)p_i(1-p_i)x_i$$

$$= \sum_{y=0}\frac{1}{e^{x_i^T r}+1}x_i + \sum_{y>0}\frac{-e^{x_i^T r}}{e^{x_i^T r}+1}x_i$$

$$= \sum_{i=1}^{n}\left\{\frac{-e^{x_i^T r}}{e^{x_i^T r}+1} + I\{y_i=0\}\frac{1}{e^{x_i^T r}+1}\right\}x_i$$

$$\frac{\partial \log L(\beta, r)}{\partial \beta \partial r^T} = \frac{\partial \log L(\beta, r)}{\partial r \partial \beta^T} = 0$$

$$\frac{\partial \log L(\beta, r)}{\partial \beta \partial \beta^T} = \sum_{y>0} -e^{x_i^T\beta}x_i x_i^T$$

$$\frac{\partial \log L(\beta, r)}{\partial r \partial r^T} = \sum_{y>0}\frac{-e^{x_i^T r}}{(1+e^{x_i^T r})^2}x_i x_i^T + \sum_{y=0}\frac{-e^{x_i^T r}}{(e^{x_i^T r}+1)^2}x_i x_i^T$$

$$= \sum_{i=1}^{n}\frac{e^{x_i^T r}}{(1+e^{x_i^T r})^2}x_i x_i^T$$

$$\therefore \begin{pmatrix} \beta^{new} \\ \gamma^{new} \end{pmatrix} = \begin{pmatrix} \beta^{old} \\ \gamma^{old} \end{pmatrix} - \begin{pmatrix} \frac{\partial \log(\beta^{old},\gamma^{old})}{\partial\beta\partial\beta^T} & 0 \\ 0 & \frac{\partial \log(\beta^{old},\gamma^{old})}{\partial\gamma\partial\gamma^T} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial \log(\beta^{old},\gamma^{old})}{\partial\beta} \\ \frac{\partial \log(\beta^{old},\gamma^{old})}{\partial\gamma} \end{pmatrix}$$

$$\beta^{new} = \beta^{old} - \left[\frac{\partial \log L(\beta^{old},\gamma^{old})}{\partial\beta\partial\beta^T}\right]^{-1} \frac{\partial \log L(\beta^{old},\gamma^{old})}{\partial\beta}$$

$$= \beta^{old} - (X_1^T W_\beta^{old} X_1)^{-1} X_1^T (Y_1 - \Lambda_\beta^{old})$$

$X_1 := \{x_i : y_i > 0\}$, $Y_1 := \{y_i : y_i > 0\}$

$W_\beta^{old} = \mathrm{diag}\{\exp(x_i^T\beta)\}_{i\in\{i: y_i>0\}}$, $\Lambda_\beta^{old} = \begin{pmatrix} e^{x_1^T\beta^{old}} \\ \vdots \\ e^{x_{n_1}^T\beta^{old}} \end{pmatrix}$

$$\gamma^{new} = \gamma^{old} - \left[\frac{\partial \log L(\beta^{old},\gamma^{old})}{\partial\gamma\partial\gamma^T}\right]^{-1} \frac{\partial \log L(\beta^{old},\gamma^{old})}{\partial\gamma}$$

where $\frac{\partial \log L(\beta^{old},\gamma^{old})}{\partial\gamma\partial\gamma^T} = \sum_{i=1}^{n} \frac{e^{x_i^T\gamma^{old}}}{1+e^{x_i^T\gamma^{old}}} x_i x_i^T$

$$\frac{\partial \log L(\beta^{old},\gamma^{old})}{\partial\gamma} = \sum_{i=1}^{n} \left\{ \frac{-e^{x_i^T\gamma^{old}}}{1+e^{x_i^T\gamma^{old}}} + I\{y_i=0\} \frac{1}{e^{x_i^T\gamma^{old}}+1} \right\} x_i$$

Lecture 20

1. The score equation for MLE of model (1) − (4):

$$\sum_{i=1}^{n} \frac{y_i - \mu(x_i^T\beta)}{\sigma_i^2}\mu'(x_i^T\beta)x_i = 0, \text{ define: } x_i^T = (1, z_i), \quad z_i \in \{0, 1\}$$

Model (1): $y_i | x_i \sim N(\mu_i, \sigma^2)$, with $\mu_i = x_i^T\beta$

$$\therefore \sum_{i=1}^{n} \frac{y_i - x_i^T\hat\beta}{\sigma^2} x_i = 0,$$

$$\Rightarrow \sum_{i=1}^{n} \frac{(y_i - \hat\beta_0 - \hat\beta_1 z_i)}{\sigma^2}\begin{pmatrix}1\\ z_i\end{pmatrix} = 0$$

$$\Rightarrow \sum_{i=1}^{n}(y_i - \hat\beta_0 - \hat\beta_1 z_i) = 0, \quad \sum_{i=1}^{n}(y_i - \hat\beta_0 - \hat\beta_1 z_i)z_i = 0$$

$$\therefore \sum_{i=1}^{n} z_i(y_i - \hat\beta_0 - \hat\beta_1 z_i) = \sum_{i:\{z_i=1\}}(y_i - \hat\beta_0 - \hat\beta_1) = 0,$$

$$\therefore \hat\beta_0 = \bar y_0, \quad \hat\beta_1 = \bar y_1 - \bar y_0, \text{ where } \bar y_1 = \frac{1}{n_1}\sum_{i:\{z_i=1\}} y_i, \quad \bar y_0 = \frac{1}{n_0}\sum_{\{i:z_i=0\}} y_i$$

Model (2): $y_i | x_i \sim \text{Bernoulli}(\mu_i)$, with $\mu_i = \frac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}}$

$$\therefore \sum_{i=1}^{n} \frac{y_i - \mu_i}{\cancel{\mu_i(1-\mu_i)}}\cancel{\mu_i(1-\mu_i)}x_i = \sum_{i=1}^{n}(y_i - \mu_i)x_i = 0$$

$$\therefore \sum_{i=1}^{n}\left(y_i - \frac{e^{\hat\beta_0+\hat\beta_1 z_i}}{1+e^{\hat\beta_0+\hat\beta_1 z_i}}\right) = 0, \quad \sum_{i=1}^{n}\left(y_i - \frac{e^{\hat\beta_0+\hat\beta_1 z_i}}{1+e^{\hat\beta_0+\hat\beta_1 z_i}}\right)z_i = 0$$

$$\sum_{i=1}^{n}\left(y_i - \frac{e^{\hat\beta_0+\hat\beta_1 z_i}}{1+e^{\hat\beta_0+\hat\beta_1 z_i}}\right)z_i = \sum_{\{i:z_i=1\}}\left(y_i - \frac{e^{\hat\beta_0+\hat\beta_1}}{1+e^{\hat\beta_0+\hat\beta_1}}\right) = 0$$

$$\sum_{i=1}^{n}\left(y_i - \frac{e^{\hat\beta_0+\hat\beta_1 z_i}}{1+e^{\hat\beta_0+\hat\beta_1 z_i}}\right) = \sum_{\{i:z_i=0\}}\left(y_i - \frac{e^{\hat\beta_0}}{1+e^{\hat\beta_0}}\right) + \sum_{\{i:z_i=1\}}\left(y_i - \frac{e^{\hat\beta_0+\hat\beta_1}}{1+e^{\hat\beta_0+\hat\beta_1}}\right) = 0$$

$$\Rightarrow \sum_{\{i:z_i=0\}}\left(y_i - \frac{e^{\hat\beta_0}}{1+e^{\hat\beta_0}}\right) = 0$$

$$\therefore \begin{cases}\hat\beta_0 = \log\frac{\bar y_0}{1-\bar y_0}, \text{ where } \bar y_0 = \frac{1}{n_0}\sum_{\{i:z_i=0\}} y_i \\ \hat\beta_1 = \log\frac{\bar y_1}{1-\bar y_1} - \log\frac{\bar y_0}{1-\bar y_0}, \text{ where } \bar y_1 = \frac{1}{n_1}\sum_{\{i:z_i=1\}} y_i\end{cases}$$

Model (3): $y_i | x_i \sim \text{Poisson}(\mu_i)$, with $\mu_i = e^{x_i^T \beta}$

$$\sum_{i=1}^{n} \frac{y_i - \mu_i(x_i^T\beta)}{\cancel{\mu_i(x_i^T\beta)}} \cancel{\mu_i(x_i^T\beta)} \cdot x_i = \sum_{i=1}^{n} (y_i - \mu_i(x_i^T\beta)) x_i = 0$$

$$\therefore \sum_{i=1}^{n} (y_i - e^{\beta_0 + \beta_1 z_i}) = 0, \quad \sum_{\{i: z_i = 1\}} (y_i - e^{\beta_0 + \beta_1}) = 0$$

$$\Rightarrow \sum_{i=1}^{n} (y_i - e^{\beta_0 + \beta_1 z_i}) = \sum_{\{i: z_i = 0\}} (y_i - e^{\beta_0}) + \sum_{\{i: z_i = 1\}} (y_i - e^{\beta_0 + \beta_1}) = 0$$

$$\Rightarrow \sum_{\{i: z_i = 0\}} (y_i - e^{\beta_0}) = 0 \quad \therefore \hat{\beta}_0 = \log \bar{y}_0, \text{ where } \bar{y}_0 = \frac{1}{n_0} \sum_{\{i: z_i = 0\}} y_i$$

$$\hat{\beta}_1 = \log \bar{y}_1 - \log \bar{y}_0, \text{ where } \bar{y}_1 = \frac{1}{n_1} \sum_{\{i: z_i = 1\}} y_i$$

Model (4): $y_i | x_i \sim NB(\mu_i, \delta)$, with $\mu_i = e^{x_i^T \beta}$

$$\sum_{i=1}^{n} \frac{y_i - \mu_i(x_i^T\beta)}{\mu_i + \frac{\mu_i^2}{\delta}} \mu_i(x_i^T\beta) x_i = \sum_{i=1}^{n} \frac{y_i - \mu_i(x_i^T\beta)}{1 + \frac{\mu_i}{\delta}} x_i = 0$$

$$\therefore \begin{cases} \sum_{i=1}^{n} \frac{y_i - e^{\beta_0 + \beta_1 z_i}}{1 + \frac{1}{\delta} e^{\beta_0 + \beta_1 z_i}} = 0 \Rightarrow \sum_{\{i: z_i = 0\}} \frac{y_i - e^{\beta_0}}{1 + \frac{1}{\delta} e^{\beta_0}} = 0 \\ \sum_{\{i: z_i = 1\}} \frac{y_i - e^{\beta_0 + \beta_1}}{1 + \frac{1}{\delta} e^{\beta_0 + \beta_1}} = 0 \end{cases}$$

$$\therefore \hat{\beta}_0 = \log \bar{y}_0, \ \hat{\beta}_1 = \log \bar{y}_1 - \log \bar{y}_0, \text{ where } \bar{y}_1 = \frac{1}{n_1} \sum_{\{i: z_i = 1\}} y_i, \ \bar{y}_0 = \frac{1}{n_0} \sum_{\{i: z_i = 0\}} y_i$$

Lecture 21

2.