Advanced Methods Homework 1

Bohao Tang

5.1:

(a) denote $P_0 = P(y=0)$, $P_1 = P(y=1)$

Then $$P(y=1|x) = \frac{P(x|y=1)P(y=1)}{P(x|y=1)P(y=1) + P(x|y=0)P(y=0)}$$

$$= \frac{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu_1)^2}{2\sigma^2}} P(y=1)}{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu_1)^2}{2\sigma^2}} P(y=1) + \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu_0)^2}{2\sigma^2}} P(y=0)}$$

$$= \frac{e^{\left\{-\frac{[(x-\mu_1)^2-(x-\mu_0)^2]}{2\sigma^2} + \log\frac{P_1}{P_0}\right\}}}{e^{\left\{-\frac{[(x-\mu_1)^2-(x-\mu_0)^2]}{2\sigma^2} + \log\frac{P_1}{P_0}\right\}} + 1}$$

$$\Rightarrow \log\frac{P(y=1|x)}{1-P(y=1|x)} = \frac{2(\mu_1-\mu_0)}{2\sigma^2}x + \log\frac{P_1}{P_0} - \frac{\mu_1^2-\mu_0^2}{2\sigma^2}$$

$$= \frac{\mu_1-\mu_0}{\sigma^2}x + \log\frac{P_1}{P_0} + \frac{\mu_0^2-\mu_1^2}{2\sigma^2}$$

$\Rightarrow P(y=1|x)$ satisfies a logistic regression model with $\beta_1 = \frac{\mu_1-\mu_0}{\sigma^2}$

(b) It's basically the same, now

$$P(y=1|x) = \frac{e^{\left\{-\left[\frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(x-\mu_0)^2}{2\sigma_0^2}\right] + \log\frac{P_1}{P_0}\right\}}}{e^{\left\{-\left[\frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(x-\mu_0)^2}{2\sigma_0^2}\right] + \log\frac{P_1}{P_0}\right\}} + 1}$$

$$\Rightarrow \log\frac{P(y=1|x)}{1-P(y=1|x)} = \log\frac{P_1}{P_0} + \frac{\mu_0^2}{2\sigma_0^2} - \frac{\mu_1^2}{2\sigma_1^2} + \left(\frac{\mu_1}{\sigma_1^2} - \frac{\mu_0}{\sigma_0^2}\right)x + \left(\frac{1}{2\sigma_0^2} - \frac{1}{2\sigma_1^2}\right)x^2$$

which is a logistic model with a quadratic term

(C) now

$$P(y=1|x) = \frac{h(x)\, e^{\{x\theta_1 - b(\theta_1)\}} \cdot P_1}{h(x)\, e^{\{x\theta_1 - b(\theta_1)\}} \cdot P_1 + h(x)\, e^{\{x\theta_0 - b(\theta_0)\}}\, P_0}$$

$$= \frac{e^{\{x(\theta_1-\theta_0) - b(\theta_1) + b(\theta_0) + \log\frac{P_1}{P_0}\}}}{e^{\{x(\theta_1-\theta_0) - b(\theta_1) + b(\theta_0) + \log\frac{P_1}{P_0}\}} + 1}$$

$$\Rightarrow \log \frac{P(y=1|x)}{1-P(y=1|x)} = \log\frac{P_1}{P_0} + b(\theta_0) - b(\theta_1) + (\theta_1 - \theta_0)x$$

which is a logistic model with $\beta_1 = \theta_1 - \theta_0$

5.5:

In this setting we have that

$$\eta_i = \Sigma \beta_j X_{ij} \quad \Rightarrow \mu_i = F(\eta_i)$$

$$\mu_i = E\, y_i = \pi_i$$

$$var(y_i) = \cancel{E\left(\frac{n_i y_i}{n_i}\right)^2 - \pi_i^2} \quad var\left(\frac{n_i y_i}{n_i}\right) = \frac{\pi_i(1-\pi_i)}{n_i}$$

$$\Rightarrow \frac{\partial \mu_i}{\partial \eta_i} = f(\eta_i) \qquad \text{where } f(x) = \frac{dF(x)}{dx} = \text{p.d.f of } F(\cdot)$$

$$\Rightarrow w_i = \frac{f^2(\eta_i)\cdot n_i}{\pi_i(1-\pi_i)} = \frac{n_i}{\pi_i(1-\pi_i)} f^2\left(\Sigma_j \beta_j X_{ij}\right) \;;\quad \pi_i = F(\Sigma_j \beta_j X_{ij})$$

$\Rightarrow$ the asymptotic variance of $\hat\beta$ is $(X^T W X)^{-1}$ where $W = \cancel{\text{diag}}\ \text{diag}\{w_i\}$

5.6: For any fixed j, we ~~rearrg~~ rearrange X and W to let $X = [X_j, X_{-j}]$ where $X_j = \begin{pmatrix} X_{1j} \\ X_{2j} \\ \vdots \\ X_{Nj}\end{pmatrix}$ and X_{-j} is X deleted jth column

Then
$$var(\hat\beta_j) = (X^T W X)^{-1}_{11} = \begin{pmatrix} X_j^T W X_j, & X_j^T W X_{-j} \\ X_{-j}^T W X_j, & X_{-j}^T W X_{-j}\end{pmatrix}^{-1}_{11}$$

$$= \left[X_j^T W X_j - X_j^T W X_{-j} (X_{-j}^T W X_{-j})^{-1} X_{-j}^T W X_j \right]^{-1}$$

since $var(\hat{\beta}_j) \geq 0$ and $var(\hat{\beta}_j) < +\infty$

we have $X_j^T W X_j > X_j^T W X_{-j} (X_{-j}^T W X_{-j})^{-1} X_{-j}^T W X_j$

under some mild condition when we can have a sense of continuity $\widehat{\text{of } X}$

that $\exists\ 0 < \lambda < 1$ constant

$$\lambda \cdot X_j^T W X_j > X_j^T W X_j (X_{-j}^T W X_{-j})^{-1} X_{-j}^T W X_j \quad \text{holds} \quad \cancel{\text{whenever we}}$$

when n_i or N changes.

Then we have $$\widehat{var}(\hat{\beta}_j) \leq \frac{1}{1-\lambda} [X_j^T \hat{W} X_j]^{-1}$$
$$= \frac{1}{1-\lambda} \left[\sum_{l=1}^{N} X_{lj}^2 n_l \hat{\pi}_l (1-\hat{\pi}_l) \right]^{-1}$$

$\hat{\pi}_l$ is the MLE so when can suppose it near true value and therefore (at least in probability) bounded away from 0, 1

Then we have some constant $C < +\infty$

$$\widehat{var}(\hat{\beta}_j) \leq C \left[\sum_{l=1}^{N} X_{lj}^2 n_l \right]^{-1} \quad (\text{in probability})$$

So in (a), when n_l become larger, and suppose X_{lj} don't change a lot

then $\sum_{l=1}^{N} X_{lj}^2 n_l$ will become larger and $\widehat{var}(\hat{\beta}_j)$ be smaller.

in (b), when N become larger, and suppose new X_{lj} is not always near 0, we have $\sum_{l=1}^{N} X_{lj}^2 n_l \uparrow$ and $\widehat{var}(\hat{\beta}_j)$ be smaller

Therefore we can suggest that $var(\hat{\beta}_j)$ will be smaller if we get more data but this actually depends on how the design matrix will change when new data come if X changes somehow "regularly" then $var(\hat{\beta}_j)$ will be truely decrease to 0.

5.20:

(a) likelihood for y_i given π_i is

$$L(y;\pi) = \prod_{i=1}^{N} \pi_i^{y_i}(1-\pi_i)^{1-y_i}$$

$$\Rightarrow L(y;\beta) = \prod_{i=1}^{N} \Phi(\textstyle\sum_j \beta_j x_{ij})^{y_i}\left[1-\Phi(\sum_j \beta_j x_{ij})\right]^{1-y_i}$$

$\Rightarrow$ log likelihood

$$\ell(y;x,\beta) = \sum_{i=1}^{N}\left[y_i \log\Phi(\textstyle\sum_j \beta_j x_{ij}) + \cancel{(1-y_i)\log\Phi(-\sum_j \beta_j x_{ij})} \; (1-y_i)\log[1-\Phi(\sum_j \beta_j x_{ij})]\right]$$

(b) the likelihood equations are $\dfrac{\partial \ell(y;x,\beta)}{\partial \beta_j} = 0 \qquad j=1,\cdots,p$

$$\Rightarrow \sum_{i=1}^{N} y_i \frac{\Phi'(\sum_j \beta_j x_{ij})}{\Phi(\sum_j \beta_j x_{ij})} x_{ij} + (1-y_i)\frac{-\Phi'(\sum_j \beta_j x_{ij})}{1-\Phi(\sum_j \beta_j x_{ij})} x_{ij} = 0$$

$$\Rightarrow \sum_{i=1}^{N} \phi(\textstyle\sum_j \beta_j x_{ij})\, x_{ij}\left[y_i \cdot \dfrac{1}{\hat{\pi}_i} - (1-y_i)\dfrac{1}{1-\hat{\pi}_i}\right] = 0$$

$$\Rightarrow \sum_{i=1}^{N} (y_i - \hat{\pi}_i)\frac{\phi(\sum_j \beta_j x_{ij})}{\hat{\pi}_i(1-\hat{\pi}_i)} x_{ij} = 0$$

5.38:

Use R code

```
fit <- glm( y ~ t+d, family = binomial);
summary(fit)
```

to estimate the initial model $\text{logit}(\pi) = \beta_0 + \beta_1 t + \beta_2 d$

$y|t,d \sim \text{Bernoulli}(\pi)$

we get

$\hat{\beta}_0 = -1.41734$ p-value : 0.19536

$\hat{\beta}_1 = -1.65895$ p-value : 0.07224 residual deviance : 30.138

$\hat{\beta}_2 = 0.06868$ p-value : 0.00931

we can see variable "d" is significant and "t" is somehow significant.

so we include "t" and "d" and then compare model with different link function

we have.

Link function;	AIC, ~~[illegible]~~
logit	36.138
probit	36.341
complementary log-log	37.716
log-log	35.161

so we may choose log-log as our Link function., then our model is

$$-\log[-\log(\pi_i)] = \sum_{j=1}^{p} \beta_j x_{ij}$$

$$y | x, \beta \sim \text{Bernoulli}(\pi_i)$$

we have $\hat{\beta}_0 = 0.71485$ $\hat{\beta}_1 = 1.18794$ $\hat{\beta}_2 = -0.05467$

and this means have a long duration of surgery and use laryngeal mask airway will increase the risk of a patient having sore throat on waking.