Advanced Methods Homework 1

Bohao Tang

5.1:

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

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

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

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

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

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

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

(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

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

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

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

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

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

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

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 of X

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

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

when n_i or N changes.

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

$\hat{\pi}_l$ 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

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

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

$\Rightarrow$ log likelihood $$\ell(y;x,\beta) = \sum_{i=1}^{N}\Big[y_i \log\Phi(\Sigma_j\beta_j x_{ij}) + \cancel{(1-y_i)\log\Phi(\Sigma_j\beta_j x_{ij})} \; (1-y_i)\log[1-\Phi(\Sigma_j\beta_j x_{ij})]\Big]$$

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

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

$$\Rightarrow \sum_{i=1}^{N} \phi(\Sigma_j\beta_j x_{ij})\, x_{ij}\left[y_i\cdot\frac{1}{\hat\pi_i} - (1-y_i)\frac{1}{1-\hat\pi_i}\right] = 0$$

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

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 , ~~corr(y, $\hat{y}$)~~
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

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

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

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.

Coding

6.

Here we have that "x1_covariate1" is just the intercept term, so we don't need to include it since we already add intercept in our algorithm. We use least square estimator of beta to initialize our algorithm.

```
myLR <- function(Y, X, it_max=25, eps=1e-8){
    D = cbind(1, X)
    p = dim(D)[2]
    beta = solve(t(D) %*% D) %*% t(D) %*% Y
    prob = exp(D %*% beta) / (1 + exp(D %*% beta))
    for(i in 1:it_max){
        WAUX = diag(array(prob * (1-prob)))
        W = solve(t(D) %*% WAUX %*% D)
        dbeta = W %*% t(D) %*% (Y - prob)
        if(norm(dbeta,"2") <= eps)
            break
        beta = beta + dbeta
        prob = exp(D %*% beta) / (1 + exp(D %*% beta))
    }
    std = sqrt(diag(W))
    result = list(beta = beta, std = std)
    return(result)
}

data = read.csv("Ex0107.txt", sep=" ")
Y = data$y_response
X2 = data$x2_covariate2

fit = glm(Y ~ X2, family = binomial)
result = myLR(Y,X2)

summary(fit)
```

```
##
## Call:
## glm(formula = Y ~ X2, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0372  -0.7635   0.3422   0.8527   1.3705
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.6172     0.5711   1.081   0.2798
## X2            1.5091     0.7773   1.941   0.0522 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
##     Null deviance: 27.526  on 19  degrees of freedom
## Residual deviance: 21.390  on 18  degrees of freedom
## AIC: 25.39
##
## Number of Fisher Scoring iterations: 5
```

```r
print(result)
```

```
## $beta
##        [,1]
##   0.6172003
## X 1.5090921
##
## $std
##                     X
## 0.5710882 0.7772833
```

And we can see here my estimator consists with R function.

7.

```r
library(magrittr)
library(dplyr)

data2 = read.csv("Ex0109.csv")
counts = aggregate(cbind(Pro05,Anti05,Pro06,Anti06,Pro07,Anti07) ~ Party, data = data2, sum)
print(counts)
```

```
##   Party Pro05 Anti05 Pro06 Anti06 Pro07 Anti07
## 1     D  2825    551  1905    332  3242    554
## 2     R   346   2957   366   1840   671   3105
```

```r
X = rbind(c(5,0),
          c(6,0),
          c(7,0),
          c(5,1),
          c(6,1),
          c(7,1))
colnames(X) = c("Year","Party")
Pro = c(2825,1905,3242,346,366,671)
Anti = c(551,332,554,2957,1840,3105)
fit = glm(cbind(Pro,Anti) ~ X, family=binomial)
summary(fit)
```

```
##
## Call:
## glm(formula = cbind(Pro, Anti) ~ X, family = binomial)
##
## Deviance Residuals:
##       1        2        3        4        5        6
##  2.0658   0.5652  -2.5970  -3.8257   2.4914   1.2503
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)  0.66794    0.14341   4.658 3.20e-06 ***
## XYear        0.17428    0.02360   7.384 1.54e-13 ***
## XParty      -3.47344    0.04118 -84.339  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 10145.221  on 5  degrees of freedom
## Residual deviance:    33.738  on 3  degrees of freedom
## AIC: 86.475
##
## Number of Fisher Scoring iterations: 4
```

We can see here the p value for Party is extremely small, so the behavior of two parties towards pro-environment voting is significant different regardless of year effect.