Computer Practical 2

Statistics 2

Dom Hutchinson

Consider the random variables $Y_i \overset{\text{ind}}{\sim} \text{Bernoulli}(\sigma(\theta^T x_i))$ for $i \in [1, n]$ where x_i are observed d-dimensional real vectors of explanatory variables and σ is the standard logistic function.

Question 1

By definition we have $L(\theta; \mathbf{y}) \propto \prod_{i=1}^{n} f_{Y_i}(y_i; \theta)$.

Thus, in this case $L(\theta; \mathbf{y}) \propto \prod_{i=1}^{n} \sigma(\theta^T \mathbf{x}_i)^{y_i} (1 - \sigma(\theta^T \mathbf{x}_i))^{1-y_i}$ We have

$$
\begin{aligned}
\ell(\theta; \mathbf{y}) &:= \ln L(\theta; \mathbf{y}) \\
&= \ln \left[L(\theta; \mathbf{y}) \propto \prod_{i=1}^{n} \sigma(\theta^T \mathbf{x}_i)^{y_i} (1 - \sigma(\theta^T \mathbf{x}_i))^{1-y_i} \right] \\
&= \sum_{i=1}^{n} \ln \sigma(\theta^T \mathbf{x}_i)^{y_i} + \ln[1 - \sigma(\theta^T \mathbf{x}_i)]^{1-y_i} \\
&= \sum_{i=1}^{n} y_i \ln \sigma(\theta^T \mathbf{x}_i) + (1 - y_i) \ln[1 - \sigma(\theta^T \mathbf{x}_i)]
\end{aligned}
$$

Question 2

Consider $\frac{\partial}{\partial \theta_j} \sigma(\theta^T \mathbf{x}_i)$ we have

$$
\begin{aligned}
\frac{\partial}{\partial \theta_j} &= \left[\frac{\partial}{\partial \theta_j} \sigma(\theta^T \mathbf{x}_i) \right] \cdot \left[\frac{\partial}{\partial \theta_j} \theta^T \mathbf{x}_i \right] \quad \text{by chain rule} \\
&= \sigma(\theta^T \mathbf{x}_i)(1 - \sigma(\theta^T \mathbf{x}_i)) x_{ij}
\end{aligned}
$$

Note that $\frac{\partial}{\partial \theta_j} \theta_T = (\delta_{ij} : i \in [1, n])$, the vectors of all zeros except at position j where it is 1.
Thus $\frac{\partial}{\partial \theta_j} \theta_T \mathbf{x}_i = x_{ij}$.
Now consider $\frac{\partial}{\partial \theta_j} \ell(\theta; \mathbf{y})$. Then

$$
\begin{aligned}
\frac{\partial}{\partial \theta_j} \ell(\theta; \mathbf{y}) &= \sum_{i=1}^{n} y_i \frac{\partial}{\partial \theta_j} \ln \sigma(\theta^T \mathbf{x}_i) + (1 - y_i) \frac{\partial}{\partial \theta_j} \ln[1 - \sigma(\theta^T \mathbf{x}_i)] \\
&= \sum_{i=1}^{n} y_i \frac{x_{ij} \sigma(\theta^T \mathbf{x}_i)[1 - \sigma(\theta^T \mathbf{x}_i)]}{\sigma(\theta^T \mathbf{x}_i)} + (1 - y_i) \frac{-x_{ij} \sigma(\theta^T \mathbf{x}_i)[1 - \sigma(\theta^T \mathbf{x}_i)]}{1 - \sigma(\theta^T \mathbf{x}_i)} \\
&= \sum_{i=1}^{n} y_i x_{ij} [1 - \sigma(\theta^T \mathbf{x}_i)] - (1 - y_i) x_{ij} \sigma(\theta^T \mathbf{x}_i) \\
&= \sum_{i=1}^{n} y_i x_{ij} - y_i x_{ij} \sigma(\theta^T \mathbf{x}_i) - x_{ij} \sigma(\theta^T \mathbf{x}_i) + y_i x_{ij} \sigma(\theta^T \mathbf{x}_i) \\
&= \sum_{i=1}^{n} [y_i - \sigma(\theta^T \mathbf{x}_i)] x_{ij}
\end{aligned}
$$

Question 3

Consider $\frac{\partial^2}{\partial\theta_j\partial\theta_k}\ell(\theta;\mathbf{y})$. Then

$$
\begin{aligned}
\frac{\partial^2}{\partial\theta_j\partial\theta_k}\ell(\theta;\mathbf{y}) &= \frac{\partial}{\partial\theta_k}\left[\frac{\partial}{\partial\theta_j}\ell(\theta;\mathbf{y})\right]\\
&= \frac{\partial}{\partial\theta_k}\left[\sum_{i=1}^{n}[y_i - \sigma(\theta^T x_i)]x_{ij}\right]\\
&= \sum_{i=1}^{n}0 - x_{ij}\frac{\partial}{\partial\theta_k}\sigma(\theta^T x_i)\\
&= -\sum_{i=1}^{n}x_{ij}\sigma(\theta^T x_i)[1 - \sigma(\theta^T x_i)]x_{ik} \text{ by result in Question 2}\\
&= -\sum_{i=1}^{n}\sigma(\theta^T x_i)[1 - \sigma(\theta^T x_i)]x_{ij}x_{ik}
\end{aligned}
$$

```r
# Standard Logistic Function
sigma <- function(v) {
  1/(1+exp(-v))
}


# Log likelihood function
ell <- function(theta, X, y) {
  p <- as.vector(sigma(X%*%theta))
  sum(y*log(p) + (1-y)*log(1-p))
}


# Score Function
score <- function(theta, X, y) {
  p <- as.vector(sigma(X%*%theta))
  as.vector(t(X)%*%(y-p))
}


# Calculate Hessian Matrix
hessian <- function(theta, X) {
  p <- as.vector(sigma(X%*%theta))
  -t(X)%*%((p*(1-p))*X)
}
```

```r
fep.eet <- read.csv("data/FEP_EET.csv") # load data
head(fep.eet)
```

```
##   MO_Emp Female Parent Age PANSS_G PANSS_P PANSS_N Y1_Emp
## 1      0      1      0  18      39      10      19      0
## 2      1      1      0  18      29      12      12      1
## 3      0      0      0  18      39      15      16      0
## 4      0      1      0  17      51      29      26      0
## 5      1      1      0  16      45      10      25      1
## 6      1      1      0  18      35      24      10      1
```

```r
X.raw <- as.matrix(fep.eet[,1:7]) # Explanatory variables (ie first 7 cols)
X <- cbind(1,X.raw) # Add column of 1s at start of X
head(X)
```

```
##        MO_Emp Female Parent Age PANSS_G PANSS_P PANSS_N
```

```
## [1,] 1      0      1      0  18      39      10      19
## [2,] 1      1      1      0  18      29      12      12
## [3,] 1      0      0      0  18      39      15      16
## [4,] 1      0      1      0  17      51      29      26
## [5,] 1      1      1      0  16      45      10      25
## [6,] 1      1      1      0  18      35      24      10
```

```r
y <- fep.eet$Y1_Emp # Response variables (ie last col)
d <- 8 # Number dimensions of parameters

# Compute Maximum Likelihood Estimate
maximize.ell <- function(ell, score, X, y, theta_0) {
  optim.out <- optim(theta_0, fn=ell, gr=score, X=X, y=y, method="BFGS",
    control=list(fnscale=-1, maxit=1000, reltol=1e-16))
  optim.out$par
}

mle <- maximize.ell(ell, score, X, y, rep(0,d))
mle
```

```
## [1]  0.34557722  3.53898725 -0.56967106  0.60441618 -0.03680464  0.02056926
## [7] -0.08936738 -0.04371038
```

Question 4

Let $\alpha \in (0,1)$ and $z_\beta := \Phi^{-1}(1-\beta)$.

We have $\dfrac{\hat{\theta}_{n,j} - \theta_j}{\sqrt{I_n(\hat{\theta}_n)_{jj}^{-1}}} \to_{\mathcal{D}(\cdot;\theta)} \text{Normal}(0,1)$. Then

$$
\begin{aligned}
\mathbb{P}\left(-z_{\alpha/2} \leq \frac{\hat{\theta}_{n,j} - \theta_j}{\sqrt{I_n(\hat{\theta}_n)_{jj}^{-1}}} \leq z_{\alpha/2}\right) &= 1 - \alpha \\
\implies \quad \mathbb{P}\left(-\hat{\theta}_{n,j} - z_{\alpha/2}\sqrt{I_n(\hat{\theta}_n)_{jj}^{-1}} \leq -\theta_j \leq -\hat{\theta}_{n,j} + z_{\alpha/2}\sqrt{I_n(\hat{\theta}_n)_{jj}^{-1}}\right) &= 1 - \alpha \\
\implies \quad \mathbb{P}\left(\hat{\theta}_{n,j} - z_{\alpha/2}\sqrt{I_n(\hat{\theta}_n)_{jj}^{-1}} \leq \theta_j \leq \hat{\theta}_{n,j} + z_{\alpha/2}\sqrt{I_n(\hat{\theta}_n)_{jj}^{-1}}\right) &= 1 - \alpha
\end{aligned}
$$

Thus we have $\mathbb{P}\left(\theta_j \in [L(\mathbf{X})_j, U(\mathbf{X})_j]\right) = 1 - \alpha$ where

$$
L(\mathbf{X})_j = \hat{\theta}_{n,j} - z_{\alpha/2}\sqrt{I_n(\hat{\theta}_n)_{jj}^{-1}} \quad \text{and} \quad U(\mathbf{X})_j = \hat{\theta}_{n,j} + z_{\alpha/2}\sqrt{I_n(\hat{\theta}_n)_{jj}^{-1}}
$$

.

```r
compute.CI.endpoints <- function(X,y,alpha) {
  mle<-maximize.ell(ell,score,X,y,rep(0,d))

  fisher_information<-(-1)*hessian(mle,X)
  fisher_information.inverse<-solve(fisher_information)
  diagonal=diag(fisher_information.inverse)

  z_alpha=qnorm(1-alpha/2)

  # Calculate lower & upper endpoints
  lower=mle-z_alpha*sqrt(diagonal)
```

```
  upper=mle+z_alpha*sqrt(diagonal)

  return(list(lower=lower,upper=upper))
}

ci<-compute.CI.endpoints(X,y,0.05)

# Visualise confidence intervals
plot.ci <- function(mle, CI.L, CI.U, components) {
  plot(components, mle[components], pch=20, main="Observed confidence intervals",
    xlab="component", ylab="value", ylim=c(min(CI.L[components]), max(CI.U[components])))
  arrows(components, CI.L[components], components, CI.U[components], length=0.05, angle=90, code=3)
  abline(h=0.0, col="red")
  axis(side=1, at=components, labels = FALSE)
}

plot.ci(mle, ci$lower, ci$upper, 1:4)
```
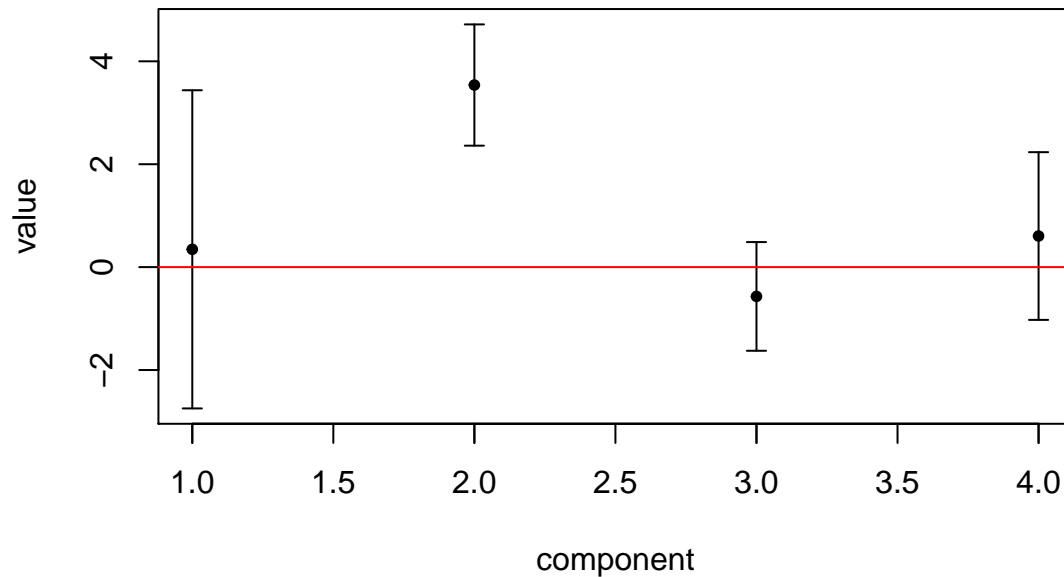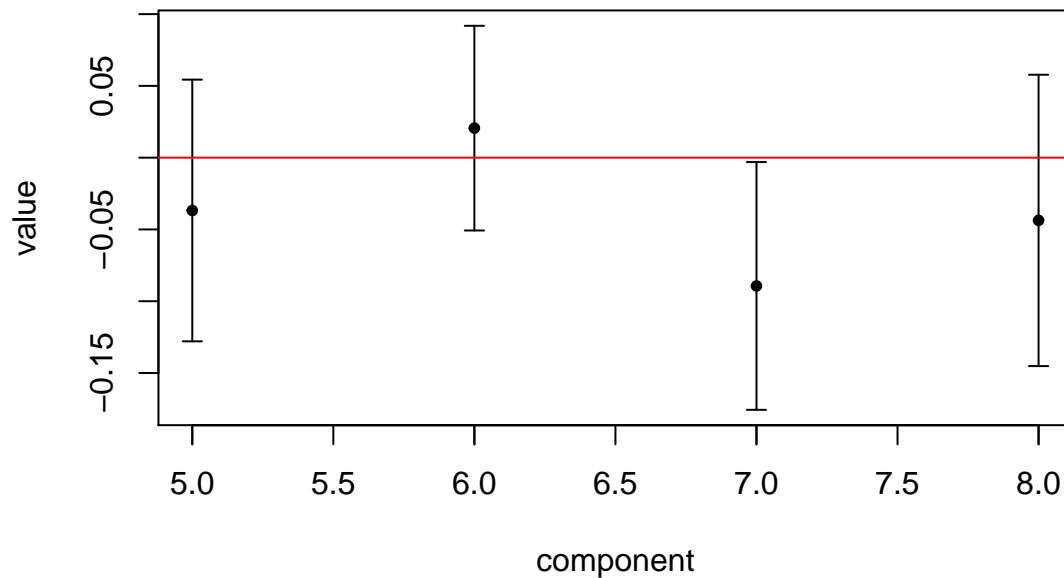
Observed confidence intervals


```
plot.ci(mle, ci$lower, ci$upper, 5:8)
```

Observed confidence intervals

Question 5

```r
# generate data associated with the matrix X when theta is the true value of the parameter
generate.ys<-function(X,theta) {
  n<-dim(X)[1]
  rbinom(n,size=1,prob=sigma(X%*%theta))
}

theta_star=c(-.7,3.5,0,0,0,0,-.08,0)

prop.condition <- function(trials,X,theta,component) {
  count.coverage<-0
  count.exclude_zero<-0

  for (i in 1:trials) {
    ys<-generate.ys(X,theta)
    ci<-compute.CI.endpoints(X,ys,0.05)
    # check whether theta*_component is in interval
    if (ci$lower[component]<=theta[component] && ci$upper[component]>=theta[component]) count.coverage=c
    # check whether 0 is excluded
    if (ci$lower[component]>0 || ci$upper[component]<0) count.exclude_zero=count.exclude_zero+1
  }
  return(list(coverage=count.coverage/trials,exclude_zero=count.exclude_zero/trials))
}

seven<-prop.condition(1000,X,theta_star,7)
eight<-prop.condition(1000,X,theta_star,8)
```

```
cat("Coverage of CI for theta_7:",seven$coverage,
    "\nProbability that 0 is excluded from CI for theta_7",seven$exclude_zero,
    "\nCoverage of CI for theta_8:",eight$coverage,
    "\nProbability that 0 is excluded from CI for theta_8",eight$exclude_zero)
```

```
## Coverage of CI for theta_7: 0.936
## Probability that 0 is excluded from CI for theta_7 0.496
## Coverage of CI for theta_8: 0.945
## Probability that 0 is excluded from CI for theta_8 0.055
```

Question 6

```
big.X<-rbind(X,X)

seven<-prop.condition(1000,big.X,theta_star,7)
eight<-prop.condition(1000,big.X,theta_star,8)

cat("Coverage of CI for theta_7:",seven$coverage,
    "\nProbability that 0 is excluded from CI for theta_7",seven$exclude_zero,
    "\nCoverage of CI for theta_8:",eight$coverage,
    "\nProbability that 0 is excluded from CI for theta_8",eight$exclude_zero)
```

```
## Coverage of CI for theta_7: 0.948
## Probability that 0 is excluded from CI for theta_7 0.792
## Coverage of CI for theta_8: 0.952
## Probability that 0 is excluded from CI for theta_8 0.048
```

The only notably difference between the results in **Question 5** & those in **Question 6** are for *"The Probability that the asymptotically exact $1 - \alpha$ confidence interval for θ_7 exclude 0*, we see an increase when the data set is doubled in size. As the number of data points increases the width of the confidence interval decreases but its mean value won't change. Looking at the plots produced in **Question 4** we see that the upper bound for θ_7 is just below 0, we expect this upper bound to decrease thus moving further away from 0 meaning more samples will not include it.