PROBLEM SET 3

NIMA MOHAMMADI
NIMAMO@VT.EDU

1. MLE Bernoulli

Consider the Bernoulli model (also called "Binary response model") where a given random draw $y_i = 1$ with probability π, and $y_i = 0$ with probability $1 - \pi$. Thus, the density for this model can be compactly written as $f(y_i) = \pi^{y_i}(1-\pi)^{1-y_i}$. The moments for this density are given as $E(y_i) = \pi, V(y_i) = \pi(1-\pi)$. Assume you draw a sample of n observations from this distribution.

(1) Derive $lnL(\pi), g(\pi), H(\pi), I(\pi)$.

$f(y_i) = \pi^{y_i}(1-\pi)^{1-y_i}$
$\Rightarrow l(\pi) = \pi^{y_i}(1-\pi)^{1-y_i}$
$\Rightarrow \ln l(\pi) = y_i \ln(\pi) + (1-y_i)\ln(1-\pi)$

Then, summing over all observations we have:

$$\begin{aligned}
\ln \mathrm{L}(\pi) &= \ln(\pi)\sum_{i=1}^{n} y_i + \ln(1-\pi)\sum_{i=1}^{n}(1-y_i) \\
g(\pi) &= \frac{\partial \ln L(\pi)}{\partial \pi} = \frac{1}{\pi}\sum_{i=1}^{n} y_i - \frac{1}{1-\pi}\sum_{i=1}^{n}(1-y_i) \\
H(\pi) &= \frac{\partial^2 \ln L(\pi)}{\partial \pi^2} = -\frac{1}{\pi^2}\sum_{i=1}^{n} y_i - \frac{1}{(1-\pi)^2}\sum_{i=1}^{n}(1-y_i) \\
I(\pi) &= -\mathrm{E}_y(H(\pi)) = -\mathrm{E}_y\left(-\frac{1}{\pi^2}\sum_{i=1}^{n} y_i - \frac{1}{(1-\pi)^2}\sum_{i=1}^{n}(1-y_i)\right) \\
&= \frac{n\pi}{\pi^2} + \frac{n(1-\pi)}{(1-\pi)^2} = \frac{n}{\pi(1-\pi)}
\end{aligned}$$

(2) Derive the ML estimator (call it p), and its asymptotic variance.

$$
\begin{aligned}
g(\pi) &= \frac{1}{\pi}\sum_{i=1}^{n} y_i - \frac{1}{1-\pi}\sum_{i=1}^{n}(1-y_i) = 0 \\
&\Rightarrow \frac{1}{\pi}\sum_{i=1}^{n} y_i = \frac{1}{1-\pi}\sum_{i=1}^{n}(1-y_i) \Rightarrow \frac{1}{\pi}\sum_{i=1}^{n} y_i = \frac{n}{1-\pi} - \frac{1}{1-\pi}\sum_{i=1}^{n} y_i \\
&\Rightarrow \frac{1}{\pi}\sum_{i=1}^{n} y_i + \frac{1}{1-\pi}\sum_{i=1}^{n} y_i = \frac{n}{1-\pi} \Rightarrow \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)\sum_{i=1}^{n} y_i = \frac{n}{1-\pi} \\
&\Rightarrow \sum_{i=1}^{n} y_i = n\pi \\
&\Rightarrow p = \frac{\sum_{i=1}^{n} y_i}{n}
\end{aligned}
$$

$$
Var(p) = [I(\pi)]^{-1} = [\frac{n}{\pi(1-\pi)}]^{-1} = \frac{\pi(1-\pi)}{n}
$$

(3) Given a sample of four "1"'s and one "0" (so n=5), compute $lnL\left(\pi\right)$ in terms of π and numerically.

Our estimator gives

$$
p = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{4}{5} = 0.8
$$

We have

$$
\begin{aligned}
f(y_i = 1) &= \pi \\
f(y_i = 0) &= (1-\pi)
\end{aligned}
$$

Then the likelihood function is evaluated as

$$
L(\pi) = \pi^4(1-\pi) \Rightarrow \ln L(\pi) = 4\ln(\pi) + \ln(1-\pi)
$$

Then

$$
\ln L(p) = 4\ln(.8) + \ln(1-.8) = -2.502
$$

(4) Now suppose that in the sample of 5 observations, all are "1's". Derive the numerical solution for the ML estimator for this case.

$$
\begin{aligned}
p &= \frac{\sum_{i=1}^{n} y_i}{n} = \frac{5}{5} = 1 \\
&\Rightarrow \ln L(\pi) = 5\ln(\pi) \\
&\Rightarrow \ln L(p) = 5\ln(1) = 0
\end{aligned}
$$

2. MLE Parameterized Exponential

Consider the Exponential density with parameterized mean, i.e

$$
\begin{aligned}
f\left(y_i|\mathbf{x}_i\right) &= \frac{exp\left(-y_i/\mathbf{x}_i'\boldsymbol{\beta}\right)}{\mathbf{x}_i'\boldsymbol{\beta}} \quad \text{with} \\
E\left(y_i|\mathbf{x}_i\right) &= \mathbf{x}_i'\boldsymbol{\beta}, \quad V\left(y_i|\mathbf{x}_i\right) = \left(\mathbf{x}_i'\boldsymbol{\beta}\right)^2
\end{aligned}
$$

Assume the sample size is n and $\mathbf{x}_i$ is $k \times 1$.

(1) Derive $lnL(\boldsymbol{\beta}), g(\boldsymbol{\beta}), H(\boldsymbol{\beta})$ and $I(\boldsymbol{\beta})$

$$l(\boldsymbol{\beta}) = \frac{\exp\left(-y_i/\mathbf{x}_i'\boldsymbol{\beta}\right)}{\mathbf{x}_i'\boldsymbol{\beta}}$$

$$\ln l(\boldsymbol{\beta}) = -\frac{y_i}{\mathbf{x}_i'\boldsymbol{\beta}} - \ln\left(\mathbf{x}_i'\boldsymbol{\beta}\right)$$

$$\begin{aligned}
\ln L(\boldsymbol{\beta}) &= -\sum_{i=1}^{n} y_i\left(\mathbf{x}_i'\boldsymbol{\beta}\right)^{-1} - \sum_{i=1}^{n}\ln\left(\mathbf{x}_i'\boldsymbol{\beta}\right) \\
g(\boldsymbol{\beta}) &= \frac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n}\left[y_i\left(\mathbf{x}_i'\boldsymbol{\beta}\right)^{-2}\mathbf{x}_i\right] - \sum_{i=1}^{n}\left(\mathbf{x}_i'\boldsymbol{\beta}\right)^{-1}\mathbf{x_i} \\
H(\boldsymbol{\beta}) &= \frac{\partial g(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} = -2\sum_{i=1}^{n}\left[y_i\left(\mathbf{x}_i'\boldsymbol{\beta}\right)^{-3}\mathbf{x}_i\mathbf{x}_i'\right] + \sum_{i=1}^{n}\left(\mathbf{x}_i'\boldsymbol{\beta}\right)^{-2}\mathbf{x}_i\mathbf{x}_i' \\
I(\boldsymbol{\beta}) &= -E_y[H(\boldsymbol{\beta})] = 2\sum_{i=1}^{n}E\left(y_i\right)\left(\mathbf{x}_i'\boldsymbol{\beta}\right)^{-3}\mathbf{x}_i\mathbf{x}_i' - \sum_{i=1}^{n}\left(\mathbf{x}_i'\boldsymbol{\beta}\right)^{-2}\mathbf{x}_i\mathbf{x}_i' \\
&= 2\sum_{i=1}^{n}\left(\mathbf{x}_i'\boldsymbol{\beta}\right)\left(\mathbf{x}_i'\boldsymbol{\beta}\right)^{-3}\mathbf{x}_i\mathbf{x}_i' - \sum_{i=1}^{n}\left(\mathbf{x}_i'\boldsymbol{\beta}\right)^{-2}\mathbf{x}_i\mathbf{x}_i' = 2\sum_{i=1}^{n}\left(\mathbf{x}_i'\boldsymbol{\beta}\right)^{-2}\mathbf{x}_i\mathbf{x}_i' - \sum_{i=1}^{n}\left(\mathbf{x}_i'\boldsymbol{\beta}\right)^{-2}\mathbf{x}_i\mathbf{x}_i' \\
&= \sum_{i=1}^{n}\left(\mathbf{x}_i'\boldsymbol{\beta}\right)^{-2}\mathbf{x}_i\mathbf{x}_i'
\end{aligned}$$

(2) Using the gradient for the entire sample, show that the score identity holds.

$$\begin{aligned}
E_y[g(\boldsymbol{\beta})] &= \sum_{i=1}^{n}\left[E\left(y_i|\mathbf{x}_i\right)\left(\mathbf{x}_i'\boldsymbol{\beta}\right)^{-2}\mathbf{x}_i - \left(\mathbf{x}_i'\boldsymbol{\beta}\right)^{-1}\mathbf{x_i}\right] \\
&= \sum_{i=1}^{n}\left[\left(\mathbf{x}_i'\boldsymbol{\beta}\right)^{-1}\mathbf{x}_i - \left(\mathbf{x}_i'\boldsymbol{\beta}\right)^{-1}\mathbf{x}_i\right] = 0
\end{aligned}$$

(3) Using the gradient for the entire sample, show that the information matrix identity holds.

$$\begin{aligned}
V[g(\boldsymbol{\beta})|\mathbf{X}] &= \sum_{\Sigma} V\left(y_i|\mathbf{x}_i\right)\left(\mathbf{x}_i'\boldsymbol{\beta}\right)^{-4}\mathbf{x}_i\mathbf{x}_i' \\
&= \sum_{i=1}^{n}\left(\mathbf{x}_i'\boldsymbol{\beta}\right)^{2}\left(\mathbf{x}_i'\boldsymbol{\beta}\right)^{-4}\mathbf{x}_i\mathbf{x}_i' = \sum_{i=1}^{n}\left(\mathbf{x}_i'\boldsymbol{\beta}\right)^{-2}\mathbf{x}_i\mathbf{x}_i' \\
&= I(\boldsymbol{\beta})
\end{aligned}$$

3. Hypothesis Testing in MLE

Consider the hedonic property value data set from script `mod3s2c` and the log-linear regression version of this model. Estimate this model via MLE, using analytical gradient and Hessian (as in mod2s1b). Use the inverted negative Hessian to derive standard errors for all estimates. Capture your MLE results in a nice table.

Loading data:

```
data <- read.table('/Users/nima/AAEC5126/data/hedonics.txt', sep="\t", header=FALSE)

colnames(data) <- c("price", "lnacres", "lnsqft", "age","gradeab", "pkadeq",
                    "vacant", "empden", "popden", "metro", "distair", "disthaz")

attach(data)
n <- nrow(data)
X <- cbind(rep(1, n), data[,-1])
colnames(X)[1] <- "const"
X <- as.matrix(X)
k <- ncol(X)
y <- log(data$price)
```

Function for optimization components (llf, g, H):

```
CLRMllfan <- function(x, y, X, n, k){
    bm <- x[1:k]
    sig2 <- x[k + 1]^2  #square to keep positive

    llf <- -(n/2) * log(2 * pi) - (n/2) * log(sig2) - ((1/(2 *
        sig2)) * t(y - X %*% bm) %*% (y - X %*% bm))
    # sample log-lh for the CLRM

    # Gradient
    g1 <- (t(X) %*% (y - X %*% bm))/sig2
    g2 <- -(n/(2 * sig2)) + ((t(y - X %*% bm) %*% (y -
        X %*% bm))/((2 * sig2^2)))
    g <- rbind(g1, g2)

    # Hessian
    H1 <- -(t(X) %*% X)/sig2
    H2 <- -(t(X) %*% (y - X %*% bm))/(sig2^2)
    H3 <- t(H2)
    H4 <- n/(2 * sig2^2) - (t(y - X %*% bm) %*% (y - X %*%
        bm)/sig2^3)
    H <- rbind(cbind(H1, H2), cbind(H3, H4))
    return (list(llf, g, H))
}
```

Initial values:

```
bols <- solve((t(X)) %*% X) %*% (t(X) %*% y)  # compute OLS estimator
e <- y - X %*% bols  # Get residuals.
```

```r
SSR <- (t(e) %*% e)  #sum of squared residuals - should be minimized
s2 <- (t(e) %*% e)/(n - k)  #get the regression error (estimated variance of 'eps').
Vb <- s2[1, 1] * solve((t(X)) %*% X)
# get the estimated variance-covariance matrix of bols
se = sqrt(diag(Vb))  # get the standard erros for your coefficients;
tval = bols/se  # get your t-values.
x0 <- 0.7 * c(bols, s2)
```

Choose the following tuner settings for the optimization algorithm:

```r
cri <- 10  #initial setting for convergence criterion
cri1 <- 1e-04  #convergence criterion
# (here for the sum of the absolute values of the elements in
# the gradient)
maxiter <- 2000  #max. number of allowed iterations
stsz <- 0.1  #step size, here it seems necessary to keep it on the small side

b <- x0
jj <- 0

while ((cri>cri1) & (jj<maxiter))   {
    jj = jj + 1

    int <- CLRMllfan(b, y, X, n, k)
    llf <- int[[1]]
    g <- int[[2]]
    H <- int[[3]]

    cri<-sum(abs(g)) #evaluate convergence criterion
    db=solve(-H) %*% g; #get directional vector
    b<- b+stsz*db; #update b

    iter <- c(jj, llf, cri)
    print(iter)  #send iteration results to R's command window

    if (jj == maxiter) {
        "Maximum number of iterations reached"
        break
    }
} #end of "while"-loop
```

Reporting the output:

```r
bm <- b  #this includes sigma
sig2 <- bm[k + 1]^2
sem <- sqrt(diag(solve(-H)))  #note here we need the negative H
tm <- bm/sem
```

```r
ttmle <- data.frame(col1 = c("constant", "lnacres", "lnsqft",
    "age", "gradeab", "pkadeq", "vacant", "empden", "popden",
    "metro", "distair", "disthaz", "sigma"), col2 = bm, col3 = sem,
    col4 = tm)
colnames(ttmle) <- c("variable", "estimate", "s.e.", "t")
```

Table 1. MLE output

variable	estimate	s.e.	t
constant	9.905	0.326	30.342
lnacres	0.372	0.074	5.011
lnsqft	0.595	0.076	7.839
age	0.002	0.003	0.774
gradeab	0.716	0.234	3.059
pkadeq	0.025	0.130	0.193
vacant	-0.004	0.005	-0.811
empden	0.015	0.004	4.176
popden	-0.003	0.012	-0.223
metro	0.488	0.114	4.297
distair	0.108	0.018	6.044
disthaz	0.033	0.013	2.590
sigma	0.849	0.051	16.554

The estimated error variance for the MLE model is 0.721.
The value of the log-likelihood function at convergence is -495.795

(1) Which coefficients are significant at the 1% or 5% level?

The variables `lnacres`, `lnsqft`, `gradeab`, `empden`, `metro`, `distair`, and `disthaz` are significant at the 1% and 5% levels.

(2) Interpret the marginal effects of "lnacres" , "gradeab", and "disthaz" on the dependent variable.

"lnacres": When acreage of property is increased by 1 unite, sales price increases by 0.372 unite.

"gradeab": exp(.716)-1 = 1.046, so sales praice is 104.6% more if property received the highest score from tax assessor rather than if it did not receive such a score.

"disthaz": If the distance to hazardous waste site gets increased by 1 unit (miles), sales price will increase by 3.3%.

(3) Derive an estimate for the *variance* of the error term, along with a 95% confidence interval.

```r
sig <- b[k + 1]
H1 <- solve(-H)
Vsig <- H1[k + 1, k + 1]
```

```
sig2 <- sig^2
Vsig2 <- 4 * sig^2 * Vsig
sesig2 <- sqrt(Vsig2)

# Confidence interval:
lo <- sig2 - 1.96 * sesig2
hi <- sig2 + 1.96 * sesig2

ttDELTA <- data.frame(col1 = "error variance", col2 = sig2, col3 = sesig2,
    col4 = lo, col5 = hi)
colnames(ttDELTA) <- c("variable", "estimate", "s.e.", "lower",
    "upper")
```

Table 2. MLE results for error variance using DELTA method

variable	estimate	s.e.	lower	upper
error variance	0.721	0.087	0.550	0.891

(4) Perform Wald tests for the following hypotheses. For each hypothesis obtain the test statistic and the corresponding p-value, and state your decision (use script `mod3s2b` for guidance on Wald tests in an MLE context):

(a) The marginal effects of `age` (β_4), `pkadeq` (β_6), `vacant` (β_7) and `popden` (β_9) are jointly zero.
$\mathcal{H}_0$:β_4=0, β_6=0, β_7=0, β_9=0

```
invH <- solve(-H)
Vb <- invH[1:k, 1:k]  #we can use any of the ususal estimators for Vb for the Wald test;
b <- bm[1:k]  # here we'll stick to the inverted Hessian
Rmat1 <- matrix(c(0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0), nrow = 1)
Rmat2 <- matrix(c(0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0), nrow = 1)
Rmat3 <- matrix(c(0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0), nrow = 1)
Rmat4 <- matrix(c(0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0), nrow = 1)
Rmat <- rbind(Rmat1, Rmat2, Rmat3, Rmat4)
q <- matrix(c(0, 0, 0, 0), nrow = 4)
J <- nrow(Rmat)

W <- t(Rmat %*% b - q) %*% solve(Rmat %*% Vb %*% t(Rmat)) %*%
    (Rmat %*% b - q)
pval = 1 - pchisq(W, J)
```

The Wald-statistic for this test is 1.639.
The corresponding p-value is 0.802.

Therefore we can not reject our null-hypothesis (which is marginal effects of `age` (β_4), `pkadeq` (β_6), `vacant` (β_7) and `popden` (β_9) are jointly zero) at 1% or 5% level of significance.

(b) The effect of an additional mile away from the airport on the value of a property is LESS OR EQUAL TO 3 times that of an additional mile from the hazardous waste site.
(*Hint: There are multiple ways to set up the null hypothesis. Choose the one that has a zero on the right hand side of the equation.*)
$\mathcal{H}_0$:β_{11}-3*β_{12}=0

```
invH <- solve(-H)
Vb <- invH[1:k, 1:k]
b <- bm[1:k]

Rmat1 <- matrix(c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, -3), nrow = 1)
Rmat <- rbind(Rmat1)
q <- matrix(c(0), nrow = 1)
J <- nrow(Rmat)

W <- t(Rmat %*% b - q) %*% solve(Rmat %*% Vb %*% t(Rmat)) %*%
    (Rmat %*% b - q)
pval = 1 - pchisq(W, J)
```

The Wald-statistic for this test is 0.046.
The corresponding p-value is 0.831.

This value greater than .05 (5% level of significance); so, we cannot reject our hypothesis ($\mathcal{H}_0$: $\beta_{11} - 3\beta_{12} \geq 0$). Hence it can be said that an additional mile away from the airport increases the value of a property by less than 3 times an additional mile from the hazardous waste site.

(c) The ratio of (additional mile from airport / additional mile from haz. waste site) is NO SMALLER THAN the squared effect of "log of square footage".
(*Hint: There are multiple ways to set up the null hypothesis. Choose the one that has a zero on the right hand side of the equation.*)
$\mathcal{H}_0$: $\frac{\beta_{11}}{\beta_{12}} - (\beta_3)^2 = 0$

```
b3 <- b[3]
b11 = b[11]
b12 = b[12]
cb = (b11/b12) - (b3)^(2)
q = 0
J = 1
delb3 = -(2 * b3)
delb11 = 1/b12
delb12 = -b11/(b12)^(2)
```

```
C <- matrix(c(0, 0, delb3, 0, 0, 0, 0, 0, 0, 0, delb11, delb12),
    nrow = 1)
Vcb = C %*% Vb %*% t(C)
W = t(cb - q) %*% solve(Vcb) %*% (cb - q)
pval = 1 - pchisq(W, J)
```

The Wald-statistic for this test is 5.017.
The corresponding p-value is 0.025.

Based on the p-value, we can not reject the null hypothesis. Then we can say that ratio of (additional mile from airport / additional mile from hazardous waste site) is larger than the squared effect of "log of square footage.