POL 450C, Homework 2
Solutions

April 13th, 2017

Assigned: 4/13/2017
Due: 4/20/2017

For this homework assignment we're going to work with one survey from Franco, Grimmer, and Whang (2017). Specifically, we're going to analyze survey results from September 27, 1991, when George HW Bush made a legislative appeal to reduce nuclear weapons as part of a peace deal with the soviets. The data are in the Homework 2 folder and contain:

1) Y_i = Approval (0/1)

2) X_{i1} =Treat (interviewed pre/post speech)

3) X_{i2} =Age of respondent

4) X_{i3} = Education of respondent

5) X_{i4} =Income of respondent

We're going to model Approval as a function of treatment status, age, education, income and an intercept $\boldsymbol{X}_i = (1, X_{i1}, X_{i2}, X_{i3}, X_{i4})$

1 Logistic Regression

Suppose:

$$Y_i \sim \text{Bernoulli}(\pi_i)$$
$$\pi_i = \text{logistic}(\boldsymbol{X}_i^{'}\boldsymbol{\beta})$$

a) Write out the log-likelihood, derive the score function, and derive the Hessian

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left[\text{logistic}(\boldsymbol{X}_i\boldsymbol{\beta})^{Y_i} \left(1 - \text{logistic}(\boldsymbol{X}_i\boldsymbol{\beta})\right)^{1-Y_i} \right]$$

$$\implies \ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[Y_i \log \text{logistic}(\boldsymbol{X}_i\boldsymbol{\beta}) + (1 - Y_i) \log\left(1 - \text{logistic}(\boldsymbol{X}_i\boldsymbol{\beta})\right) \right]$$

$$= \sum_{i=1}^{n} \left[-Y_i \log\left(1 + \exp(-\boldsymbol{X}_i\boldsymbol{\beta})\right) - (1 - Y_i) \log\left(1 + \exp(\boldsymbol{X}_i\boldsymbol{\beta})\right) \right]$$

$$\implies \frac{\partial \ell(\boldsymbol{\beta} \mid Y_i)}{\partial \beta_k} = \frac{\partial}{\partial \beta_k} \left[-Y_i \log\left(1 + \exp(-\boldsymbol{X}_i\boldsymbol{\beta})\right) - (1 - Y_i) \log\left(1 + \exp(\boldsymbol{X}_i\boldsymbol{\beta})\right) \right]$$

$$= \left[Y_i \frac{\exp(-\boldsymbol{X}_i\boldsymbol{\beta})}{1 + \exp(-\boldsymbol{X}_i\boldsymbol{\beta})} - (1 - Y_i) \frac{\exp(\boldsymbol{X}_i\boldsymbol{\beta})}{1 + \exp(\boldsymbol{X}_i\boldsymbol{\beta})} \right] X_{ik}$$

$$= \left[Y_i \frac{1}{1 + \exp(\boldsymbol{X}_i\boldsymbol{\beta})} - (1 - Y_i) \frac{1}{1 + \exp(-\boldsymbol{X}_i\boldsymbol{\beta})} \right] X_{ik}$$

$$\implies \frac{\partial^2 \ell(\boldsymbol{\beta} \mid Y_i)}{\partial \beta_j \beta_k} = \frac{\partial}{\partial \beta_j} \left[Y_i \frac{1}{1 + \exp(\boldsymbol{X}_i\boldsymbol{\beta})} - (1 - Y_i) \frac{1}{1 + \exp(-\boldsymbol{X}_i\boldsymbol{\beta})} \right] X_{ik}$$

$$= - \left[Y_i \frac{\exp(\boldsymbol{X}_i\boldsymbol{\beta})}{(1 + \exp(\boldsymbol{X}_i\boldsymbol{\beta}))^2} + (1 - Y_i) \frac{\exp(-\boldsymbol{X}_i\boldsymbol{\beta})}{(1 + \exp(-\boldsymbol{X}_i\boldsymbol{\beta}))^2} \right] X_{ik} X_{ij}$$

b) In R, use optim to obtain the maximum likelihood estimate of $\boldsymbol{\beta}$, given a set of binary responses Y and covariates $\boldsymbol{X}$.

```
> load('Gallup1991.RData')
> d = gallup[complete.cases(gallup), ]
> X = cbind(rep(1, nrow(d)), d[,2:ncol(d)])
> colnames(X)[1] <- '(Constant)'
> Y = d[,1]

## logistic function
> logistic <- function(value){
  solve <- -1/(1+exp(-value))
  return(solve1)
}

## log likelihood
>lik <- function(params, X, Y){
beta <- params
ytilde <- X%*%beta
yprob <- logistic(ytilde)
out <- (Y%*%log(yprob) + (1-Y)%*%log(1-yprob))
return(out)
}
```

```
## Optim
> start <- rnorm(5)
> out <- optim(start, lik, Y=Y, X=X, method = "BFGS", hessian = T, control = li
```

c) Using your function from (b) calculate the maximum likelihood estimates using the data. Report the coefficients and standard errors in a table.

Table 1: Estimates from optim

	estimate	SE
Intercept	1.296	0.424
treatment	0.09	0.265
age	-0.0099	0.0046
education	-0.03	0.0569
income	0.0865	0.0516

d) Using those estimates, calculate the first difference for each respondent. (Hint: calculate the expected value for each respondent, given their characteristics. Then calcuate the expected value in for the counterfactual value of treatment. Then take a difference)

```
coef <- out$par
X.treat = X.control = X
X.treat[,2] = 1
X.control[,2] = 0
fd = plogis(X.treat %*% coef) - plogis(X.control %*% coef)
mean(fd)
[1] 0.0168
```

The results seem to predcit a small positive treatment effect.

e) Interpret the differences: what do the coefficients from the logistic regression convey? What do the first differences suggest? What is the average first difference? The coefficients from the logistic regression convey that the mean approval rating is high, treatment and income are likely to increase approval rating, and age and education are likely to decrease approval rating. The average first difference is 0.017 or 1.7%, suggesting that this is not a very large treatment effect.

2 Score Function

Suppose $Y_i \sim \text{Normal}(\mu_i, \sigma^2)$ and
$\mu_i = \beta_0 + \beta_1 X_{i1}$

- Write out the likelihood and log-likelihood equation for this bivariate regression

$$L(\boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\beta_0 + \beta_1 X_i - y_i)^2}{2\sigma^2} \right) \right]$$

$$\implies \ell(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}\log 2\pi\sigma^2 - \sum_{i=1}^{n} \frac{(\beta_0 + \beta_1 x_i - y_i)^2}{2\sigma^2}$$

- Derive the score function $s(\beta, \sigma^2)$ (please describe the steps)

$$\frac{\partial}{\partial \beta_0} \ell(Y_i \mid \boldsymbol{\beta}, \sigma^2) = \frac{\partial}{\partial \beta_0} \left[-\frac{1}{2}\log 2\pi\sigma^2 - \frac{(\beta_0 + \beta_1 x_i - y_i)^2}{2\sigma^2} \right]$$

$$= -\frac{\beta_0 + \beta_1 x_i - y_i}{\sigma^2}$$

$$\frac{\partial}{\partial \beta_1} \ell(Y_i \mid \boldsymbol{\beta}, \sigma^2) = -\frac{x_i(\beta_0 + \beta_1 x_i - y_i)}{\sigma^2}$$

$$\frac{\partial}{\partial \sigma^2} \ell(Y_i \mid \boldsymbol{\beta}, \sigma^2) = \frac{\partial}{\partial \sigma^2} \left[-\frac{1}{2}\log 2\pi - \frac{1}{2}\log \sigma^2 - \frac{(\beta_0 + \beta_1 x_i - y_i)^2}{2\sigma^2} \right]$$

$$= -\frac{1}{2\sigma^2} + \frac{(\beta_0 + \beta_1 x_i - y_i)^2}{2\sigma^4}$$

$$\implies s(Y_i \mid \boldsymbol{\beta}, \sigma^2) = \left\{ -\frac{\beta_0 + \beta_1 x_i - y_i}{\sigma^2}, -\frac{x_i(\beta_0 + \beta_1 x_i - y_i)}{\sigma^2}, -\frac{1}{2\sigma^2} + \frac{(\beta_0 + \beta_1 x_i - y_i)^2}{2\sigma^4} \right\}$$

- Prove that $\mathrm{E}[s(\beta, \sigma^2)] = 0$ for the bivarate regression.

$$\mathbb{E}\left[-\frac{\beta_0 + \beta_1 x_i - y_i}{\sigma^2} \right] = -\frac{\beta_0 + \beta_1 x_i - \mathbb{E}_{Y_i}[y_i]}{\sigma^2}$$

$$= -\frac{\beta_0 + \beta_1 x_i - \beta_0 - \beta_1 x_i}{\sigma^2}$$

$$= 0$$

$$\mathbb{E}\left[-\frac{x_i(\beta_0 + \beta_1 x_i - y_i)}{\sigma^2} \right] = -\frac{x_i(\beta_0 + \beta_1 x_i - \mathbb{E}_{Y_i}[y_i])}{\sigma^2}$$

$$= -\frac{x_i(\beta_0 + \beta_1 x_i - \beta_0 - \beta_1 x_i)}{\sigma^2}$$

$$= 0$$

$$\mathbb{E}\left[-\frac{1}{2\sigma^2} + \frac{(\beta_0 + \beta_1 x_i - y_i)^2}{2\sigma^4} \right] = -\frac{1}{2\sigma^2} + \frac{\mathbb{E}\left[(\beta_0 + \beta_1 x_i - y_i)^2 \right]}{2\sigma^4}$$

$$= -\frac{1}{2\sigma^2} + \frac{\sigma^2}{2\sigma^4}$$

$$= -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^2}$$

$$= 0$$

- Use simulation to show that the $E[s(\beta, \sigma^2)] = 0$. (Hint: repeat the data generating process and calculate the score for each value. The average of the score should be zero)

```
score_func = function(beta, sigma, X, Y) {
  v = matrix(NA, nrow = 3, ncol = length(X))
  v[1,] = - (beta[1] + beta[2]*X - Y) / sigma^2
  v[2,] = - (beta[1] + beta[2]*X - Y) * X / sigma^2
  v[3,] = -1/(2 * sigma^2) + (beta[1] + beta[2]*X - Y)^2 / (2 * sigma^4)
  return(v)
}

beta0 <- .7
beta1 <-.9
sigma2 <- 1.993

X.2 = runif(1e4, -2, 2)
Y.2 = beta0 + beta1 * X.2 + rnorm(1e4, 0, sigma2)
rowMeans(score_func(c(beta0, beta1), sigma2, X.2, Y.2))
## [1] 3.1e-03 -3.9e-03 -6.5e-05
```

3 Robit regression (Using a t-distribution instead of normal/logistic)

Suppose

$$\begin{align}
Y_i &\sim \text{Bernoulli}(\pi_i) \\
\pi_i &= \mathcal{T}(\boldsymbol{X}_i'\boldsymbol{\beta}, 8)
\end{align}$$

Where $\mathcal{T}(\cdot, 8)$ refers to the cumulative t-distribution 8 degrees of freedom

a) Write out the likelihood and log-likelihood

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left[\mathcal{T}(\boldsymbol{X}_i\boldsymbol{\beta})^{Y_i} (1 - \mathcal{T}(\boldsymbol{X}_i\boldsymbol{\beta}))^{1-Y_i} \right]$$

$$\implies \ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} [Y_i \log \mathcal{T}(\boldsymbol{X}_i\boldsymbol{\beta}) + (1 - Y_i) \log (1 - \mathcal{T}(\boldsymbol{X}_i\boldsymbol{\beta}))]$$

b) Using your answer from a, write a function in R using `optim` to evaluate the log-likelihood of different values of β

```
> rob_like <- function(params, X, Y, df){
+    beta <- params
+    y.tilde <- X%*%beta
```

```
+    y.prob <- pt(y.tilde, df = df)
+    out <- Y%*%log(y.prob) + (1-Y)%*%log((1-y.prob))
+    return(out)
+ }
>
> start <- rnorm(5)
>
> out2 <- optim(start, rob_like, X=X, Y=Y, df = 8, method = 'BFGS', hessian = T
>
```

c) Use that function in optim to obtain maximum likelihood estimates for $\boldsymbol{\beta}$ and the variance-covariance matrix using the data from Problem 1.

```
> coefs <- out2$par
> varcov <- round(solve(-out2$hessian), digits = 3)
> varcov
        [,1]    [,2]    [,3]    [,4]    [,5]
[1,]   0.073 -0.003 -0.001 -0.006 -0.002
[2,] -0.003  0.028  0.000  0.000  0.000
[3,] -0.001  0.000  0.000  0.000  0.000
[4,] -0.006  0.000  0.000  0.001 -0.001
[5,] -0.002  0.000  0.000 -0.001  0.001
> ses <- round(sqrt(diag(solve(-out2$hessian))), digits = 3)
> t <- cbind(round(coefs, digits = 3), ses)
> rownames(t) <- c("Intercept", "Treatment", "Age", "Education", "Income")
> colnames(t) <- c("Coeffs", "SES")
> t
            Coeffs    SES
Intercept    0.831 0.270
Treatment    0.058 0.169
Age         -0.006 0.003
Education   -0.020 0.036
Income       0.055 0.033
```

d) Compare the results from the t-distribution to the results from the logistic regression. What differences do you note? What happens as you increase the degrees of freedom?

```
> logitc <- logit$coefficients
> probitc <- probit$coefficients
> tab2 <- cbind(round(logitc, digits =3),
round(probitc, digits = 3), round(coefs, digits =3))
> colnames(tab2) <- c("Logit", "Probit", "Robit")
> tab2
               Logit Probit  Robit(df=8)
(Intercept)   1.297  0.796  0.831
```

```
treat          0.090  0.053  0.058
age           -0.010 -0.006 -0.006
education     -0.031 -0.019 -0.020
income         0.087  0.051  0.055


# df = 20
> max.beta.20 <- optim(start, fn = rob_like, X=X, Y=Y,
                   df = 20, method = 'BFGS',  control=list(fnscale = -1), hes

> # df = 50
> max.beta.50 <- optim(guesses, fn = rob_like, X=X, Y=Y,
                   df = 50, method = 'BFGS',  control=list(fnscale = -1), hes

> # df = 100
> max.beta.100 <- optim(guesses, fn = rob_like, X=X, Y=Y,
                    df = 100, method = 'BFGS', control=list(fnscale = -1), he

> # df = 250
> max.beta.250 <- optim(guesses, fn = rob_like, X=X, Y=Y,
                    df = 250, method = 'BFGS', control=list(fnscale = -1), he
```

	Logistic Regression	T dist, df=8	T dist, df=10	T dist, df=20	T dist, df=50	T dist, df=100	T dist, df=200
Intercept	1.30	0.83	0.82	0.81	0.80	0.80	0.80
Treatment	0.09	0.058	0.06	0.05	0.05	0.05	0.05
Age	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
Education	-0.03	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02
Income	0.09	0.07	0.06	0.05	0.05	0.05	0.05

The results are broadly similar, but with larger coecients on the logit. That is because the logistic link puts more of its density near 0, so larger cocients are required to produce the same change in magnitude of the CDF. As the degrees of freedom of the robit link are increased, it will converge to the probit link.

4 [Bonus] Proving the Relationship Between the Hessian and Information Matrix

Instructor note: this is a challenging problem. I'd love to talk about it in class and work through it together/give hints. Please start it early!

Suppose random variable Y_i has sampling density $f(y|\boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_K)$. We want to make an inference about $\boldsymbol{\theta}$ Define $L(\boldsymbol{\theta}|\boldsymbol{Y})$ to be the likelihood and $\log L(\boldsymbol{\theta}|\boldsymbol{Y}) \equiv l(\boldsymbol{\theta}|\boldsymbol{Y})$

Define entry $I_{jk}(\boldsymbol{\theta})$ of the information matrix to be:

$$I_{jk} = \mathrm{E}[\frac{\partial \ell(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \theta_j} \frac{\partial \ell(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \theta_k}] \tag{4.1}$$

We are going to show that:

$$I_{jk} = -E[\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \theta_j \partial \theta_k}] \tag{4.2}$$

For all i and j

Recall that $I_N(\boldsymbol{\theta}) = NI(\boldsymbol{\theta})$, so it will suffice to work with a single observation.

1) Calculate $\frac{\partial \ell(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \theta_j}$ and $\frac{\partial \ell(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \theta_k}$. Take their product and calculate their expected value. We're going to show that Equation 4.2) is equal to this. (Hint: You'll have an expression that involves the sampling density and partial derivatives of f)

2) Take the second derivatives of the log likelihood. You should have an expression that involves partial derivatives, the sampling density, and second derivatives.

3) Now, take the expected value of the expression in (2) and multiply by -1. It should equal the expression in (1). (Hint: take as a given that $E[\frac{\frac{\partial^2 f(Y|\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k}}{f(Y|\boldsymbol{\theta})}] = 0$)