Lecture3 Handout

Elizabeth Colantuoni

3/30/2021

I. Objectives

Upon completion of this session, you will be able to do the following:

- Define "confounding" and approaches to assess confounding in linear and logistic regression models
- Specify the likelihood function for the logistic regression model and for the generalized linear model (GLM) family
- Use iteratively weighted least squares to maximize the log-likelihood for the generalized linear model (GLM) family

II. Assessing confounding in generalized linear models

A. Application to the NMES

In this section, we want to address the following question: Is age a confounder for the big expenditure vs. MSCD relationship?

We would consider:

1. Is age associated with MSCD?
2. Is age associated with big expenditure?
3. Is age in the causal pathway between MSCD and big expenditure?

We can answer questions 1 and 2 using statistical analyses; question 3 is not a statistical question.

The most common approach to addressing the confounding question is to fit two models, one without and with age; compare coefficients!

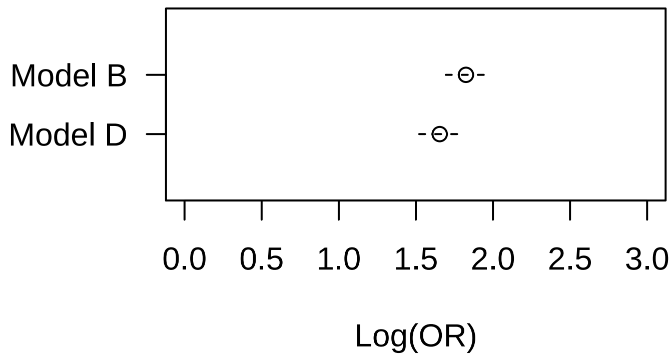
Recall,

- Model B: $logit[Pr(Y_i = 1|MSCD_i)] = \beta_{0B} + \beta_{1B} MSCD_i$
- Model D: $logit[Pr(Y_i = 1|MSCD_i, Older_i)] = \beta_{0D} + \beta_{1D} MSCD_i + \beta_{2D} Older_i$
- Fit these two models and estimate: $\frac{\hat{\beta}_{1B} - \hat{\beta}_{1D}}{se(\hat{\beta}_{1B} - \hat{\beta}_{1D})}$. NOTE: you could estimate the denominator via a bootstrap procedure.

From the analysis of the NMES data, we found:

```
##      Estimate    2.5 %   97.5 %   Chisq    Pr(>Chisq)
## mscd 1.825045 1.694177 1.955913 747.095 1.718138e-164
```

```
##      Estimate    2.5 %   97.5 %  Chisq    Pr(>Chisq)
## mscd 1.654913 1.521564 1.788262 591.65 1.096494e-130
```




```
library(boot)
```

```
##
## Attaching package: 'boot'
```

```
## The following object is masked from 'package:biostat3':
##
##     melanoma
```

```
## The following object is masked from 'package:survival':
##
##     aml
```

```
my.boot = function(dat,x) {
  d = dat[x,]
  modelB = glm(bigexp~mscd,data=d,family="binomial")
  modelD = glm(bigexp~mscd+older,data=d,family="binomial")
  modelB$coefficients[2] - modelD$coefficients[2]
}
#set.seed(8712)
#confounding.boot = boot(data1,my.boot,R=1000)
#quantile(confounding.boot$t,p=c(0.025,0.975))
```

When you run the bootstrap above (note it takes a minute or so to run), you get an estimate of $\hat{\beta}_{1B} - \hat{\beta}_{1D} =$ 0.175 and a 95% bootstrap percentile confidence interval: 0.14 to 0.20.

The unadjusted and adjusted estimates are statistically different; are they qualitatively different??

B. Caveats on assessing confounding within logistic regression models

You can use the process described above to assess for confounding for linear models (i.e. assuming normally distributed residuals and identity link function) or generalized linear models with log links.

However, you have to pay close attention when using this approach in a logistic regression model!

To be more general, let's assume we have an outcome variable Y and binary primary exposure variable $X, (X = 1 \text{ or } 0)$ and additional variable (potential confounder) Z.

If we fit the following logistic regression models:

$$\text{Model M: } logit[Pr(Y = 1|X)] = \beta_{0m} + \beta_{1m}X$$

$$\text{Model C: } logit[Pr(Y = 1|X, Z)] = \beta_{0c} + \beta_{1c}X + \beta_{2c}Z$$

These models are referred to as the "marginal" or "conditional" models, respectively.

1. Non-linearity effect in logistic regression models

Assume Y is binary AND X and Z are indepenent.

You can show that $|\beta_{1c}| \geq |\beta_{1m}|$ and the difference between β_{1c} and β_{1m} depends on the association between X and Z on Y and on the variance of Z.

This difference is what is referred to as the "nonlinearity effect" (see the Janes et al Biostatistics paper) and this feature of the logistic regression model is referred to as "non-collapsability".

- When comparing β_{1m} to β_{1c} in general (i.e. in cases where we don't want to assume X and Y are independent), we expect these two coefficients to be different. What we have to tease out is if they are different due to the "non-linearity effect" or due to the fact that there is confounding.
- Recall the three types of confounding:
  - Positive Confounding: $|\beta_{1m}| > |\beta_{1c}|$
  - Negative Confounding: $|\beta_{1m}| < |\beta_{1c}|$
  - Qualitative: change of sign!
- If you find in a data example that $|\beta_{1m}| > |\beta_{1c}|$ then you have identified positive confounding despite the "non-linearity effect"
- If you find in a data example that $|\beta_{1m}| < |\beta_{1c}|$ then it is not clear if what you observe is due to the "non-linearity effect" or negative confounding is present. Same holds for qualitative confounding.

2. Why no issue with linear models?

To better understand the "non-linearity effect", let's start by thinking about the same set up in a general linear model setting; i.e. $Y \sim Normal$ AND X and Y are independent.

The marginal and conditional models are:

$$\text{Model M: } E(Y|X) = \beta_{0m} + \beta_{1m}X$$
$$\text{Model C: } E(Y|X,Z) = \beta_{0c} + \beta_{1c}X + \beta_{2c}Z$$

The interpretation of $\beta_{1m} = E(Y|X=1) - E(Y|X=0)$.

The interpretation of $\beta_{1c} = E(Y|X=1,Z) - E(Y|X=0,Z)$. That is: at every value of Z the difference in the expected value of Y comparing $X=1$ to $X=0$ is the same.

Using the law of iterated expectations we have:

$$\begin{aligned} E(Y|X=1) - E(Y|X=0) &= E_Z[E(Y|X=1,Z) - E(Y|X=1,Z)] \\ &= E_Z[(\beta_{0c} + \beta_{1c} + \beta_{2c}Z) - (\beta_{0c} + \beta_{2c}Z)] \\ &= \beta_{1c} \end{aligned}$$

3. Non-linearity effect in logistic regression

Go back to our marginal and conditional logistic regression models, then we have:

$$exp(\beta_{1m}) = \frac{exp(\beta_{0m} + \beta_{1m})}{exp(\beta_{0m})} = \frac{Pr(Y=1|X=1)/Pr(Y=0|X=1)}{Pr(Y=1|X=0)/Pr(Y=0|X=0)}$$

$$exp(\beta_{1c}) = \frac{exp(\beta_{0c} + \beta_{1c} + \beta_{2c}Z)}{exp(\beta_{0m} + \beta_{2c}Z)} = \frac{Pr(Y=1|X=1,Z)/Pr(Y=0|X=1,Z)}{Pr(Y=1|X=0,Z)/Pr(Y=0|X=0,Z)}$$

- When would these two expressions be the same?
  - $\beta_{2c} = 0$
  - Y and X are independent (β_{1m} and β_{1c} are 0)
  - $Var(Z) = 0$
- Why aren't these two expressions the same? The expection is being taken over a non-linear function of Z.

$$E(Y|X) = Pr(Y=1|X) = E_Z\left[\frac{exp(\beta_{0c} + \beta_{1c}X + \beta_{2c}Z)}{1 + exp(\beta_{0c} + \beta_{1c}X + \beta_{2c}Z)}\right]$$

4. Simulation study: non-linearity effect

Assume the following model:

$$logit[Pr(Y = 1|X, Z)] = -2 + 0.4X + Z$$

where $Z \sim N(0, 2)$.

This model says that for persons with the same value of Z, the odds ratio for Y comparing those with and without exposure X is $exp(0.4) = 1.5$, and this odds ratio is the same regardless of the value of Z.

- Consider persons with $Z = 0$, then we know:

$$Pr(Y = 1|X = 1, Z = 0) = \frac{exp(-2 + 0.4)}{1 + exp(-2 + 0.4)} = 0.17$$

$$Pr(Y = 1|X = 0, Z = 0) = \frac{exp(-2)}{1 + exp(-2)} = 0.12$$

$$OR(Y, X|Z = 0) = \frac{0.17/0.83}{0.12/0.88} = 1.5$$

- Consider persons with $Z = 2$, then we know:

$$Pr(Y = 1|X = 1, Z = 2) = \frac{exp(-2 + 0.4 + 2)}{1 + exp(-2 + 0.4 + 2)} = 0.60$$

$$Pr(Y = 1|X = 0, Z = 2) = \frac{exp(-2 + 2)}{1 + exp(-2 + 2)} = 0.5$$

$$OR(Y, X|Z = 2) = \frac{0.6/0.4}{0.5/0.5} = 1.5$$

- What about $Pr(Y = 1|X)$ for $X = 1, 0$? We would calculate these by taking:

$$Pr(Y = 1|X) = \int_z Pr(Y = 1|X, z)f(z)dz$$

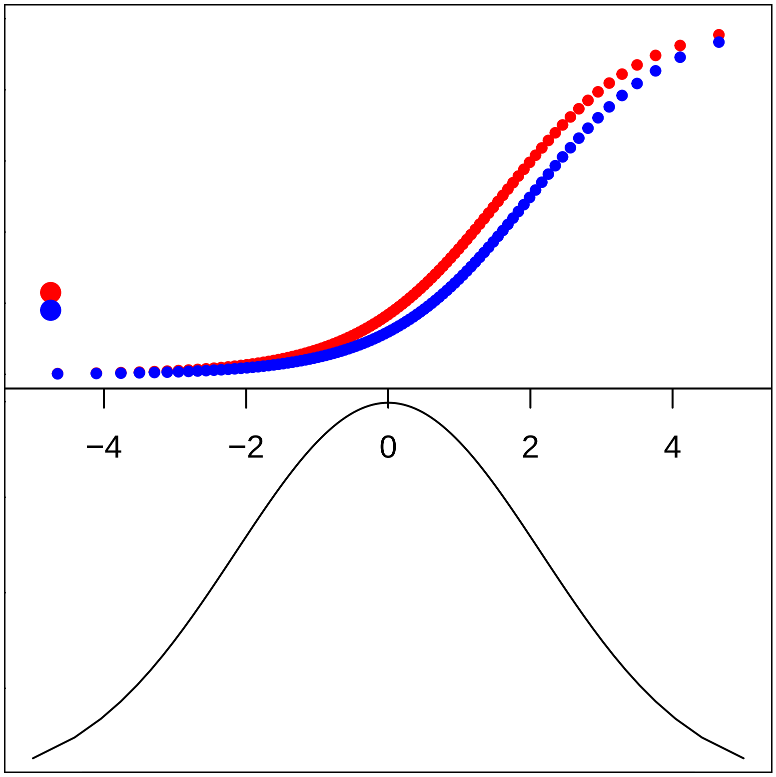
i.e. a weighted average where the weights are determined by the density function of Z (which is normal with mean 0 and standard deviation 2 in this example).

$$Pr(Y = 1|X = 1) = 0.23$$

$$Pr(Y = 1|X = 0) = 0.18$$

$$OR(Y, X) = \frac{0.23/0.77}{0.18/0.82} = 1.36$$

In this simulation example, $\beta_{1c} = 0.4$ and $\beta_{1m} = log(1.36) = 0.3$.



IMPORTANT NOTE: The “non-linearity effect” tends to behave in the same magnitude on both the $\hat{\beta}_{1c}$ and $se(\hat{\beta}_{1c})$. So in addition to comparing the estimates of β_{1c} and β_{1m}, you can compare the test statistic for β_{1c} and β_{1m}. Differences there should be due to confounding, as the “non-linearity effect” cancels out in the test statistics!

5. Simulation study: Confounding present

Here I am mimicking the simulation example from Janes et al (Biostatistics, 2010). To demonstrate a few examples where there is no confounding or there is confounding and the effect of confounding may vary.

Assume the following:

- We simulate 1000 samples of 250 persons, half with exposure ($X = 1$) and half without ($X = 0$).
- We generate Z as follows: $Z|X \sim N(\alpha_0 + \alpha_1 X, 1)$
- We generate Y from: $logit[Pr(Y = 1|X, Z)] = \beta_0 + \beta_1 X + \beta_2 Z$.
- We set $\alpha_0 = 0$, $\beta_0 = 0$ and $\beta_1 = log(2)$.
- We vary α_1 and β_2 and we estimate β_1 and then β_{1m} (the marginal estimate) and the corresponding Z statistics.

```r
expit = function(x) exp(x)/(1+exp(x))

my.sim = function(x,K=2000,N=250,a0,a1,b0,b1,b2){
  set.seed(x)
  out = NULL
  for(i in 1:K) {
    X = rbinom(N,size=1,prob=0.5)
    Z = a0 + a1*X + rnorm(N,mean=0,sd=1)
    logit = b0 + b1*X+b2*Z
    Y = rbinom(N,size=1,prob=expit(logit))
    fit0 = glm(Y~X,family="binomial")
    fit1 = glm(Y~X+Z,family="binomial")
    out = rbind(out,c(summary(fit0)$coeff[2,c(1,3)],summary(fit1)$coeff[2,c(1,3)]))
    }
  out = as.data.frame(out)
  names(out) = c("MB","MZ","CB","CZ")
  out$DiffB = out$MB-out$CB
  out$DiffZ = out$MZ-out$CZ
  out
}

## Use the Janes et al Biostatistics 2010 simulation set-up
## a0 = 0, b0 = 0 and b1 = log(2)
case1 = my.sim(x=1253,K=1000,N=250,a0=0,a1=0.01,b0=0,b1=log(2),b2=0.05)
case2 = my.sim(x=2543,K=1000,N=250,a0=0,a1=0.01,b0=0,b1=log(2),b2=1.5)
case3 = my.sim(x=2875,K=1000,N=250,a0=0,a1=1,b0=0,b1=log(2),b2=0.05)
case4 = my.sim(x=3252,K=1000,N=250,a0=0,a1=1,b0=0,b1=log(2),b2=1.5)

## Create a table of results:
my.results = rbind(
  c(0.01,0.05,apply(case1,2,mean)[c(1,3,5,2,4,6)]),
  c(0.01,1.5,apply(case2,2,mean)[c(1,3,5,2,4,6)]),
  c(1,0.05,apply(case3,2,mean)[c(1,3,5,2,4,6)]),
  c(1,1.5,apply(case4,2,mean)[c(1,3,5,2,4,6)]))
my.results = as.data.frame(round(my.results,3))
names(my.results) = c("a1","b1","beta1m","beta1","beta1m-beta1","Z1m","Z1","Z1m-Z1")
```

```
my.results
```

```
##     a1   b1 beta1m beta1 beta1m-beta1   Z1m    Z1 Z1m-Z1
## 1 0.01 0.05  0.701 0.704       -0.003 2.653 2.653  0.000
## 2 0.01 1.50  0.494 0.689       -0.195 1.902 2.183 -0.281
## 3 1.00 0.05  0.752 0.711        0.041 2.839 2.389  0.450
## 4 1.00 1.50  1.617 0.707        0.910 5.312 1.951  3.361
```

We see from the table above:

- No confounding ($alpha_1 \approx 0$ and $\beta_1 \approx 0$): the "non-linearity effect" is small, i.e. $\beta_{1m} - \beta_1$ and the Z statistics are the same.
- No confounding ($alpha_1 \approx 0$) but β_1 large: the "non-linearity effect" is non-zero and the Z statistics are roughly the same.
- "Small" confounding ($\alpha_1 > 0$ but $beta_1$ small): the "non-linearity effect" is rougyly zero (driven by $\beta_1 \approx 0$) and $\beta_{1m} > \beta_1$ indicating some confounding.
- Confounding (both α_1 and β_1 large!): we would anticipate a "non-linearity effect of roughly" -0.195, plus a confounding effect which yields a total difference in the coefficients of 0.91. Also a large difference in the Z statistics.

III. Maximum Likelihood Estimation for Logistic Models

In this section, we will be deriving the MLEs for β within the logistic regression model.

A. Recall: MLE for linear models

- Define the $(p+1) \times 1$ vector of covariates for subject i as $x_i = (1, x_{1i}, x_{2i}, ..., x_{pi})$.
- Define the $(p+1) \times 1$ vector of association parameters $\beta = (\beta_0, \beta_1, ..., \beta_p)$.

Recall,

$$Y_i = \mu_i + \epsilon_i, \epsilon_i \text{ iid } N(0, \sigma^2)$$

$$E(Y_i) = \mu_i(\beta) = x_i^{\prime}\beta$$

The score equation, $U(\beta) = \frac{\partial logL(\beta|y_i)}{\partial \beta} = \sum_{i=1}^{n} x_i(y_i - \mu_i(\beta))$.

Setting $U(\beta) = \sum_{i=1}^{n} x_i(y_i - \mu_i(\beta)) = 0$ and solving for β produced:

$$\hat{\beta} = (X^{\prime}X)^{-1}X^{\prime}Y$$

where X is the $n \times p$ matrix of stacked row vectors $x_i^{\prime}$ and Y is the $1 \times n$ vector of responses.

B. MLE for logistic regression

In this handout, we will walk through the derivations required to find the maximum likelihood estimates for parameters from a logistic regression model.

Assume the following model:

- $Y_i \sim Bernoulli(\mu_i)$ for $i = 1, ..., n$ independent observations.
- Define the vector of covariates for subject i as $x_i = (1, x_{1i}, x_{2i}, ..., x_{pi})$.
- Define the vector of association parameters $\beta = (\beta_0, \beta_1, ..., \beta_p)$.
- Assume the logit link such that:

$$log\left(\frac{\mu_i}{1-\mu_i}\right) = x_i^{'}\beta \rightarrow \mu_i = \frac{e^{x_i^{'}\beta}}{1+e^{x_i^{'}\beta}}$$

NOTE: We should really write $\mu_i(x_i, \beta)$ i.e. μ_i is a function of x_i and β. In this handout, I will simplify this to $\mu_i(\beta)$.

1. Logistic Regression: Likelihood and log likelihood functions

We can express the likelihood function as:

$$\begin{aligned}
L(\beta|y) &= Pr(Y_1 = y_1, Y_2 = y_2, ..., Y_n = y_n|\beta) \\
&= \prod_{i=1}^{n} Pr(Y_i = y_i|\beta) \\
&= \prod_{i=1}^{n} \mu_i(\beta)^{y_i}[1-\mu_i(\beta)]^{1-y_i}
\end{aligned}$$

The log-likelihood function is:

$$log[L(\beta|y)] = \sum_{i=1}^{n} y_i log[\mu_i(\beta)] + (1-y_i) log[1-\mu_i(\beta)]$$

To obtain the maximum likelihood estimates of β, we take the derivative of the log-likelihood function with respect to β, set that equal to 0 and solve for β.

Recall, we define the score equation $U(\beta)$ as the derivative of the log-likelihood function with respect to β.

2. Two useful calculations

Before we move on, we are going to need to calculate $\frac{\partial}{\partial\beta} log[\mu_i(\beta)]$ and $\frac{\partial}{\partial\beta} log[1-\mu_i(\beta)]$.

$$\begin{aligned}
\frac{\partial}{\partial\beta} log[\mu_i(\beta)] &= \frac{\partial}{\partial\beta} log\left(\frac{e^{x_i^{'}\beta}}{1+e^{x_i^{'}\beta}}\right) \\
&= \frac{\partial}{\partial\beta}[x_i^{'}\beta - log(1+e^{x_i^{'}\beta})] \\
&= x_i - x_i\frac{e^{x_i^{'}\beta}}{1+e^{x_i^{'}\beta}} \\
&= x_i[1-\mu_i(\beta)]
\end{aligned}$$

For the next derivation, note that:

$$\frac{\partial log[\mu_i(\beta)]}{\partial\beta} = \frac{1}{\mu_i(\beta)}\frac{\partial\mu_i(\beta)}{\partial\beta} \rightarrow \frac{\partial\mu_i(\beta)}{\partial\beta} = \mu_i \frac{\partial log[\mu_i(\beta)]}{\partial\beta}$$

$$\begin{aligned}
\tfrac{\partial}{\partial\beta} log[1-\mu_i(\beta)] &= -\tfrac{1}{1-\mu_i(\beta)}\tfrac{\partial\mu_i(\beta)}{\partial\beta} \\
&= \tfrac{-\mu_i(\beta)}{1-\mu_i(\beta)} x_i[1-\mu_i(\beta)] \\
&= -\mu_i(\beta)x_i
\end{aligned}$$

3. Logistic Regression: Score equation

The score equation, $U(\beta)$ is the derivative of the log-likelihood function with respect to β.

$$\begin{aligned}
U(\beta) &= \tfrac{\partial log[L(\beta|y)]}{\partial\beta} \\
&= \sum_{i=1}^{n} y_i \frac{\partial log[\mu_i(\beta)]}{\partial\beta} + (1-y_i)\frac{\partial log[1-\mu_i(\beta)]}{\partial\beta} \\
&= \sum_{i=1}^{n} y_i \left(x_i[1-\mu_i(\beta)]\right) + (1-y_i)[-\mu_i(\beta)x_i] \\
&= \sum_{i=1}^{n} x_i \left(y_i - y_i\mu_i(\beta) + (-\mu_i(\beta)) + y_i\mu_i(\beta)\right) \\
&= \sum_{i=1}^{n} x_i(y_i - \mu_i(\beta)) \\
&= X'(Y-\mu(\beta))
\end{aligned}$$

NOTE: We will also need to know $U'(\beta) = \frac{\partial U(\beta)}{\partial\beta}$

$$\begin{aligned}
U'(\beta) &= \tfrac{\partial U(\beta)}{\partial\beta} \\
&= \tfrac{\partial}{\partial\beta} X'(Y-\mu(\beta)) \\
&= \text{-}\ X' \tfrac{\partial\mu_i(\beta)}{\partial\beta} \\
&= \text{-}\ X'VX
\end{aligned}$$

where we already showed that:

$$\frac{\partial\mu_i(\beta)}{\partial\beta} = \mu_i(\beta)\frac{\partial log[\mu_i(\beta)]}{\partial\beta} = \mu_i(\beta)(1-\mu_i(\beta))x_i$$

and $V_{n\times n} = diag(\mu_i(\beta)[1-\mu_i(\beta)])$.

4. Finding the root via the Newton-Raphson method

We will review the Newton-Raphson method assuming β is scalar (i.e. a logistic regression model with intercept only).

See the figure below that displays $U(\beta)$ as a function of β. The goal is to find the value of β that sets $U(\beta) = 0$.

- Step 0: Pick an initial starting value for β, call this $\hat{\beta}^{(k)}$.
- Step 1: Compute the slope of $U(\beta)$ at $\hat{\beta}^{(k)}$, i.e. compute $U'(\hat{\beta}^{(k)})$.
- Step 2: Construct the tangent line, which is a line that passes through the points $(\hat{\beta}^{(k)}, U(\hat{\beta}^{(k)}))$ and $(\hat{\beta}^{(k+1)}, 0)$ and has slope $U'(\hat{\beta}^{(k)})$.
- Step 3: Solve the following for $\hat{\beta}^{(k+1)}$:

$$\begin{aligned}
U'(\hat{\beta}^{(k)}) &= \tfrac{U(\hat{\beta}^{(k)})-0}{\hat{\beta}^{(k)}-\hat{\beta}^{(k+1)}} \\
[\hat{\beta}^{(k)} - \hat{\beta}^{(k+1)}]U'(\hat{\beta}^{(k)}) &= U(\hat{\beta}^{(k)}) \\
\hat{\beta}^{(k)} - \hat{\beta}^{(k+1)} &= U'(\hat{\beta}^{(k)})^{-1}U(\hat{\beta}^{(k)}) \\
\hat{\beta}^{(k+1)} &= \hat{\beta}^{(k)} - U'(\hat{\beta}^{(k)})^{-1}U(\hat{\beta}^{(k)}) \\
&= U'(\hat{\beta}^{(k)})^{-1}\left(U'(\hat{\beta}^{(k)})\hat{\beta}^{(k)} - U(\hat{\beta}^{(k)})\right)
\end{aligned}$$

- Step 4: Stop if $|\hat{\beta}^{(k+1)} - \hat{\beta}^{(k)}|$ is small. If not, let $k = k + 1$ and repeat Steps 2 through 4.

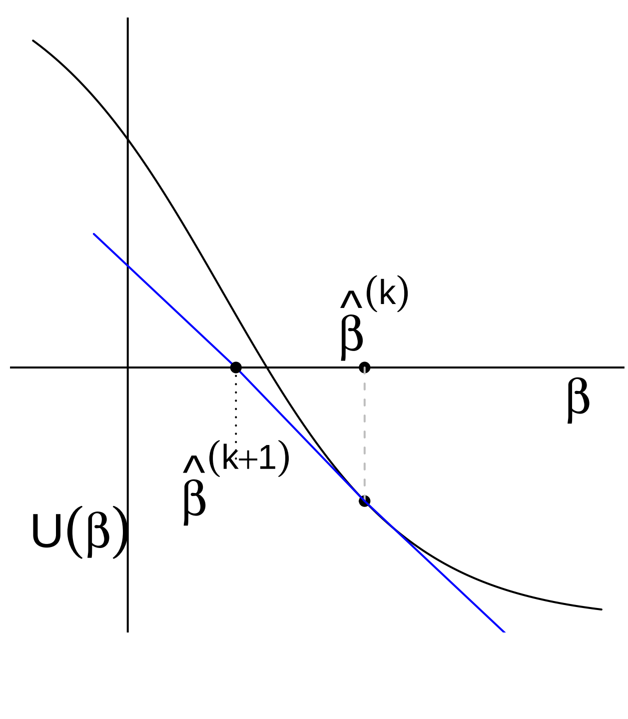


Figure 1: First Newton-Raphson Iteration

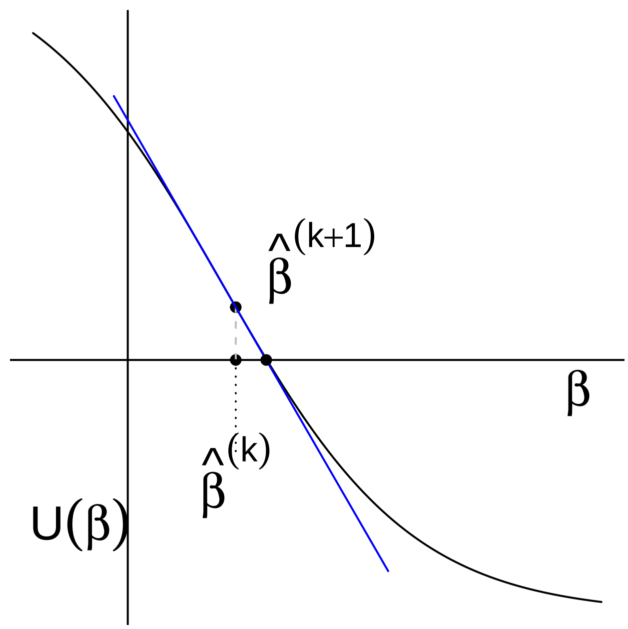


Figure 2: Second Newton-Raphson Iteration

In general, i.e. when β is a vector, we have:

$$U(\beta) = X'(Y - \mu(\beta))$$

$$U'(\beta) = -X'VX$$

And the Newton-Raphon method is expressed as:

$$\begin{aligned}
\hat{\beta}^{(k+1)} &= U'(\hat{\beta}^{(k)})^{-1}\left(U'(\hat{\beta}^{(k)})\hat{\beta}^{(k)} - U(\hat{\beta}^{(k)})\right) \\
&= -(X'V^{(k)}X)^{-1}\left[-(X'V^{(k)}X)\hat{\beta}^{(k)} - X'(Y - \mu(\hat{\beta}^{(k)}))\right] \\
&= (X'V^{(k)}X)^{-1}\left[X'V^{(k)}\left(X\hat{\beta}^{(k)} + V^{-1(k)}(Y - \mu(\hat{\beta}^{(k)}))\right)\right] \\
&= (X'V^{(k)}X)^{-1}(X'V^{(k)}Z^{(k)})
\end{aligned}$$

where $Z^{(k)} = X\hat{\beta}^{(k)} + V^{-1(k)}(Y - \mu(\hat{\beta}^{(k)}))$ = a surrogate response.

5. Iteratively Re-weighted Least Squares (IRLS)

The general procedure is:

- Step 0: Set an initial value for $\hat{\beta}^{(k)}$, $k = 0$.
- Step 1: Calculate: $V^{(k)}$, $\hat{\mu}(\hat{\beta}^{(k)})$, $Z^{(k)}$.
- Step 2: Update $\hat{\beta}^{(k+1)} = (X'V^{(k)}X)^{-1}(X'V^{(k)}Z^{(k)})$
- Step 3: Stop if $\displaystyle\sum_{j=1}^{p+1} \left(\hat{\beta}_j^{(k+1)} - \hat{\beta}_j^{(k)}\right)^2 < \epsilon$; if not, let $k = k + 1$ and repeat Steps 2 and 3.