hw2

Giang Vu

2/3/2021

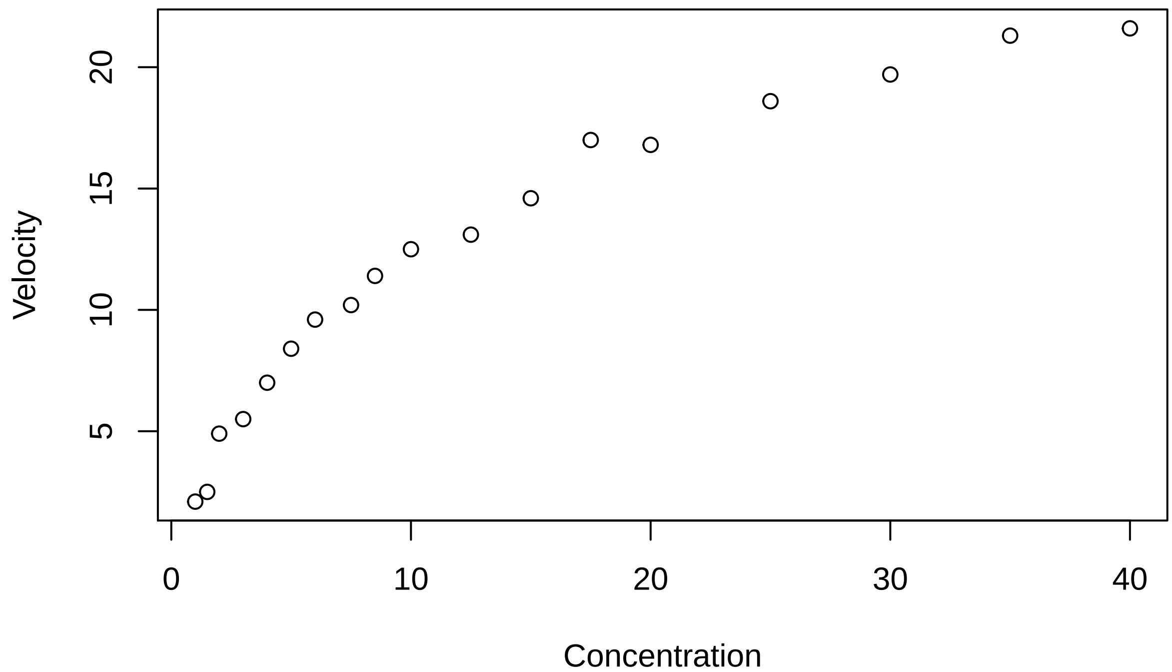
Homework 2

1.

(a) The starting points for Gauss-Newton I chose are 25 for gamma0 and 1 for gamma1.
From the equation
$1/E(Y_i) = (1/X_i)\gamma_1/\gamma_0 + 1/\gamma_0$
$\rightarrow$ As X approaches infinity, $1/X_i$ approaches 0.
$\rightarrow$ Then $1/E(Y_i)$ approaches $1/\gamma_0$, or Y_i approaches γ_0.
$\rightarrow$ γ_0 is the horizontal asymptote of this function.
From the plot of the data, the points seem to converge towards near 25, so I picked 25 as the starting point. Also, as X increases, Y also increases, so I expect γ_1 to be positive, and the magnitude of it tells us how fast Y increases as X increases, so I picked the starting value 1.

```r
#read data
enzyme <- read.delim("/Users/giangvu/Desktop/STAT 2132 - Applied Stat Method 2/HW/HW2/Enzyme.txt")

#plot data
plot(enzyme$X, enzyme$Y, xlab="Concentration", ylab="Velocity")
```



(b) Gamma0 and gamma1 are estimated using the starting points from part (a) below

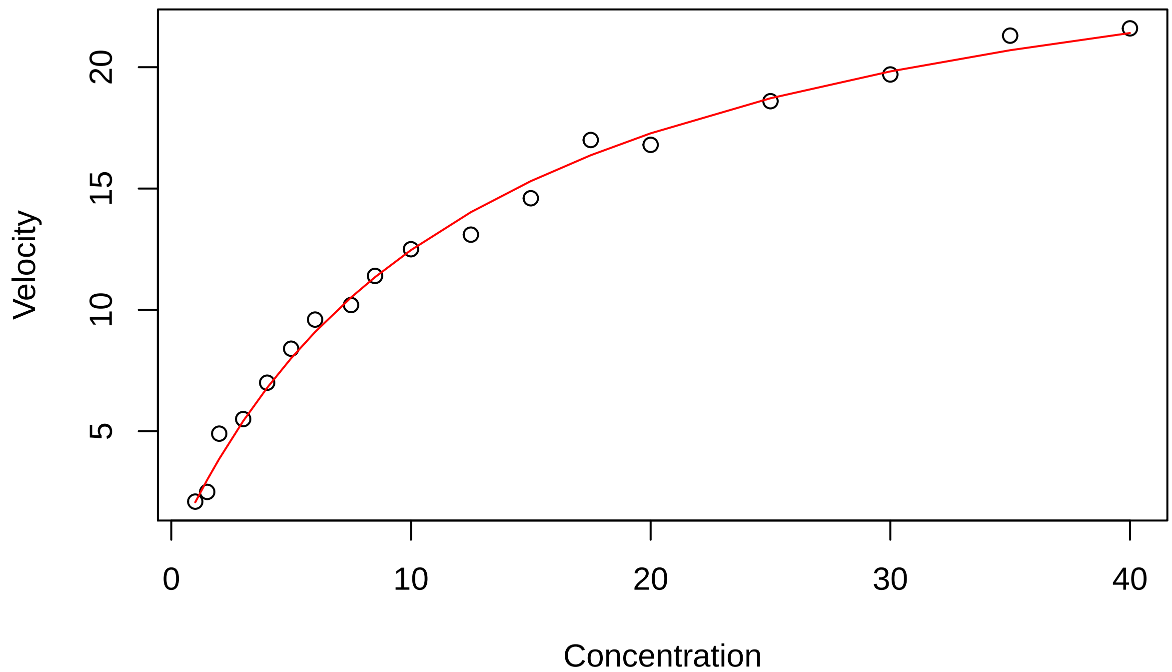
```
#nls
out.start <- nls(Y ~ (gamma0*X)/(gamma1 + X), data = enzyme, start = list(gamma0=25,gamma1=1))
#estimates for gamma's
gamma.hat <- out.start$m$getAllPars()
gamma.hat
```

```
##   gamma0   gamma1
## 28.13703 12.57443
```

2.

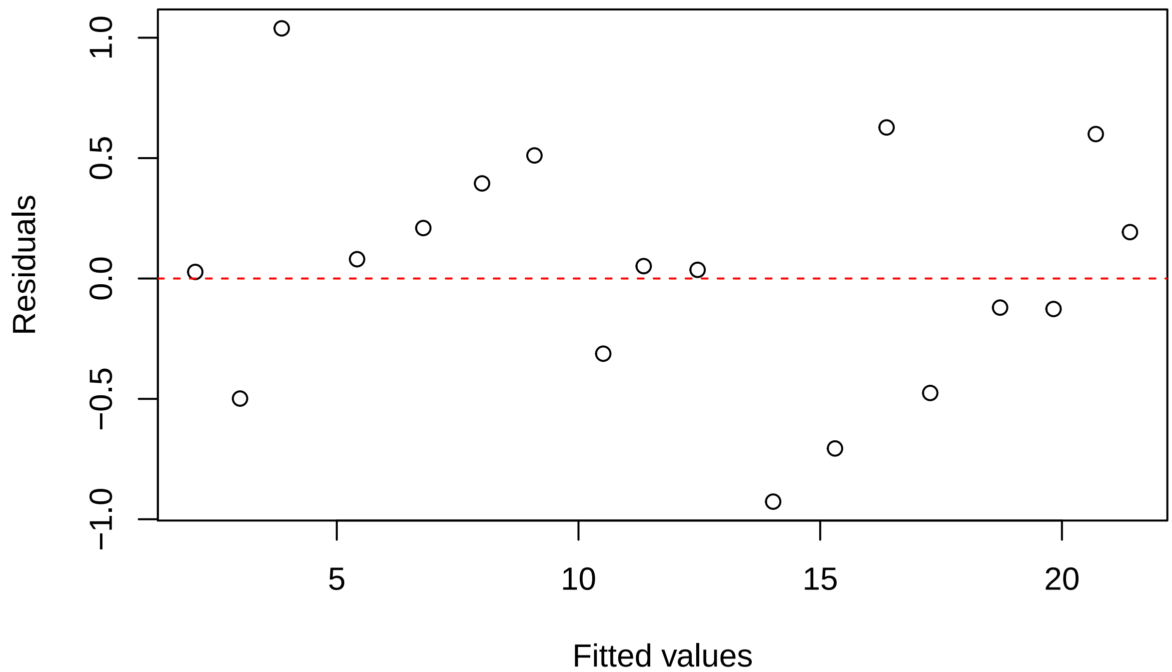
(a) The fit seems to be adequate.

```
#plot data
plot(enzyme$X, enzyme$Y, xlab="Concentration", ylab="Velocity")
#add the nonlinear regression function line
lines(enzyme$X, out.start$m$fitted(), col="red")
```

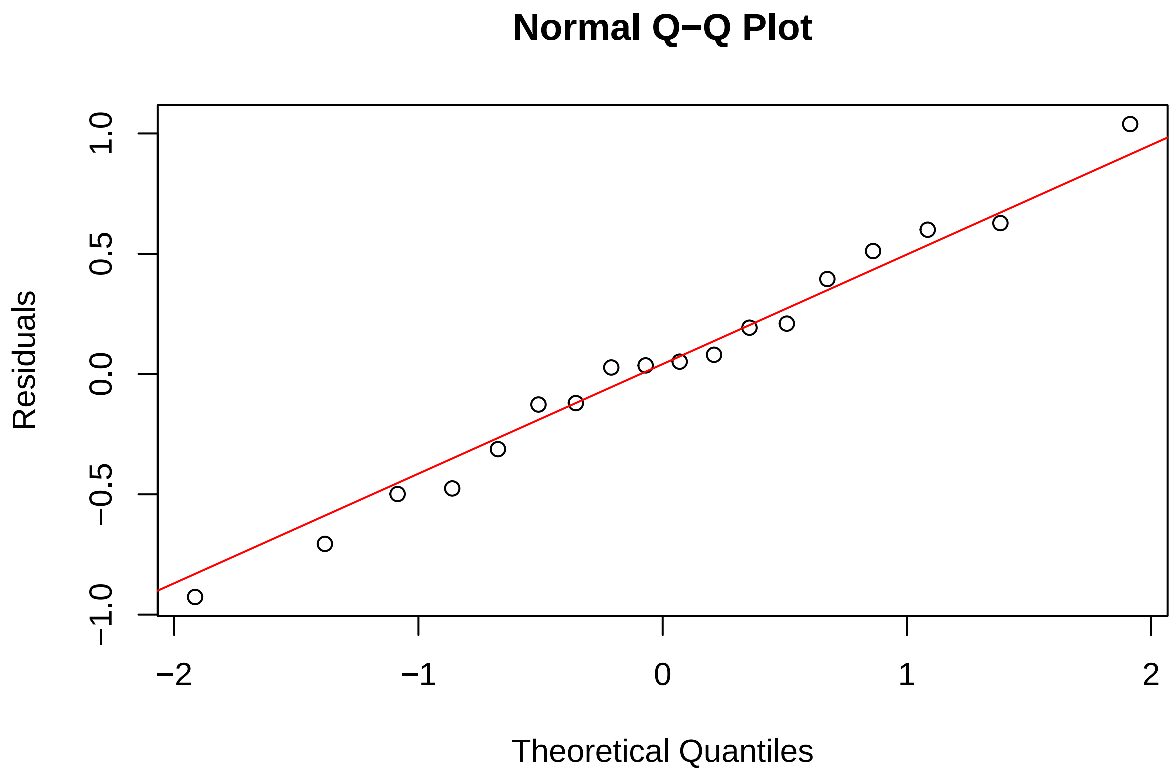


(b) Based on the residual plot and the qq-plot, the assumptions of constant variances and normality are both satisfied. The residual plots show no trend, and the tail of the qq-plot shows no skewness.

```
#residual plot - constant variances - satisfied
plot(out.start$m$fitted(), out.start$m$resid(), xlab="Fitted values", ylab="Residuals"); abline(h=0, col
```




```r
#qq plot - normality
qqnorm(out.start$m$resid(), ylab="Residuals")
qqline(out.start$m$resid(), col="red")
```



(c) Null hypothesis: Gamma1 = 20
Alternative hypothesis: Gamma1 different from 20
Assumptions: The residuals are not correlated with X, and with each other. They are also normally distributed. The test statistic is -9.73141 with 16 degrees of freedom.
The p-value is basically 0, very small
The 95% CI is (11.23,13.91)
We reject the null hypothesis, we have strong evidence to conclude that gamma1 cannot be equal to 20.

```
#test gamma1 = 20
#Jacobian
J <- out.start$m$gradient()   #The 44 x 2 gradient matrix. This acts as our design matrix.
#MSE
sigma2 <- sum(out.start$m$resid()^2)/(nrow(J)-ncol(J)) #nrow(J) = n, ncol(J) = p
#SE for inference
se.gamma0 <- sqrt(sigma2)*sqrt( solve(t(J)%*%J)[1,1] )
se.gamma1 <- sqrt(sigma2)*sqrt( solve(t(J)%*%J)[2,2] )
#t-stat for test gamma1 = 20, df = n-p = 16
t.gam1 <- (gamma.hat[2]-20)/se.gamma1

#CI
CI.gamma0 <- gamma.hat[1] + c(-1,1)*se.gamma0*qt(p = 0.95, df = nrow(J)-ncol(J))
CI.gamma1 <- gamma.hat[2] + c(-1,1)*se.gamma1*qt(p = 0.95, df = nrow(J)-ncol(J))
```

3.

After running a bootstrap with 1000 samples, the 95% percentile confidence intervals for gamma1 obtained is (11.03, 13.94). This is quite close compared to the interval based on the large sample theory we obtained in problem 2.

```r
#bootstrap with 1000 samples
set.seed(200)
enzyme_bootfcn <- function(x,i){
  d<-x[i,]
  out.boot <- nls(Y ~ (gamma0*X)/(gamma1 + X), data = d, start = list(gamma0=25,gamma1=1))
  gamma.boot <- out.boot$m$getAllPars()
  return(gamma.boot[2])
}
enzyme_bootobj <- boot(enzyme,enzyme_bootfcn,R=1000)
head(enzyme_bootobj$t)
```

```
##          [,1]
## [1,] 12.12552
## [2,] 11.60097
## [3,] 11.79980
## [4,] 11.06739
## [5,] 13.05811
## [6,] 12.56629
```

```r
#bootstrap 95% percentile CI
enzyme_bootci <- boot.ci(enzyme_bootobj)
enzyme_bootci
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 1000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = enzyme_bootobj)
##
## Intervals :
## Level      Normal              Basic
## 95%   (11.23, 14.06 )   (11.21, 14.12 )
##
## Level     Percentile            BCa
## 95%   (11.03, 13.94 )   (11.11, 14.09 )
## Calculations and Intervals on Original Scale
```

4) a) Using Bayes' Theorem for 3 events $Y=1$, X, and $Z=1$

$$P(Y=1|X,Z=1) = \frac{P(Z=1|Y=1,X) \cdot P(Y=1|X)}{P(Z=1|Y=1,X)P(Y=1|X) + P(Z=1|Y=0,X)P(Y=0|X)}$$

$$= \frac{\gamma_1 \cdot P(Y=1|X)}{\gamma_1 \cdot P(Y=1|X) + \gamma_0 P(Y=0|X)}$$

$$\to \text{logit}\{P(Y=1|X,Z=1)\} = \log\left(\frac{\gamma_1 \cdot P(Y=1|X) / \cancel{(\gamma_1 P(Y=1|X) + \gamma_0 P(Y=0|X))}}{\gamma_0 (P(Y=0|X)) / \cancel{(\gamma_1 P(Y=1|X) + \gamma_0 P(Y=0|X))}}\right)$$

$$= \log\left(\frac{\gamma_1 \cdot P(Y=1|X)}{\gamma_0 \cdot P(Y=0|X)}\right)$$

$$= \log\left(\frac{\gamma_1}{\gamma_0}\right) + \log\left(\frac{P(Y=1|X)}{P(Y=0|X)}\right)$$

$$= \log\left(\frac{\gamma_1}{\gamma_0}\right) + \log\left(\frac{P(Y=1|X)}{1-P(Y=0|X)}\right)$$

$$= \log\left(\frac{\gamma_1}{\gamma_0}\right) + \text{logit}(P(Y=1|X))$$

$$= \log\left(\frac{\gamma_1}{\gamma_0}\right) + \beta_0 + \beta_1 X$$

$$\Rightarrow \text{logit}\{P(Y=1|X,Z=1)\} = \beta_0^* + \beta_1 X \qquad \left(\beta_0^* = \beta_0 + \log\left(\frac{\gamma_1}{\gamma_0}\right)\right)$$

b) Since after the enriched study, β_1 does not change in the logit function, we can still use the estimated effect of X to infer the effect in the general population. This effect only depends on β_1.

c) However, β_0 is changed into β_0^* with enriched study sample, we cannot use estimated probability of $Y=1$ given $X=x_0$ to infer this for the population. This probability depends on both β_0 and β_1. And also the probability of $Y=1$ given $X=x_0$ is only based on one value x_0 of X, so it cannot be generalized for the population.

5) a) $M(\theta) = E\{e^{\theta Y}\}$

$$K(\theta) = \log\{M(\theta)\}$$

$$K'(0) = \frac{M'(\theta)}{M(\theta)}\bigg|_{\theta=0} = \frac{E(Y)}{1} = E(Y)$$

$$K''(0) = \left(\frac{M''(\theta)}{M(\theta)} - \frac{M'(\theta)^2}{M(\theta)^2}\right)\bigg|_{\theta=0}$$

$$= E(Y^2) - E(Y)^2 = Var(Y)$$

b) Call the normalizing constant α

$\Rightarrow f(y,\theta) = \alpha\, e^{\theta y} f_0(y)$, $\theta \in \mathbb{R}$

$\Rightarrow \int f(y,\theta)\, dy = 1$

$\Rightarrow \int \alpha\, e^{\theta y} f_0(y)\, dy = 1$

$\Rightarrow \alpha \int e^{\theta y} f_0(y)\, dy = 1$

$\Rightarrow \alpha \cdot M(\theta) = 1$

$\Rightarrow \alpha = \frac{1}{M(\theta)}$ is the normalizing constant

c) $l(y;\theta) = \log\{f(y;\theta)\} = \log\left\{\frac{1}{M(\theta)} e^{\theta y} f_0(y)\right\}$

$= -\log(M(\theta)) + \log(e^{\theta y}) + \log(f_0(y))$

$= -K(\theta) + \theta y + \log(f_0(y))$

$\log(f_0(y))$ only depends on y, we can call it $h(y)$

$\Rightarrow l(y;\theta) = -K(\theta) + \theta y + h(y)$, $\theta \in \mathbb{R}$

d) $K_\theta(t) = \log E\{e^{tY}\}$

$= \log E\{e^{(\theta+t-\theta)Y}\}$

$= \log E\{e^{(\theta+t)Y}\} + \log E\{e^{-\theta y}\}$

$= K(\theta+t) - K(\theta)$

5d) (continued)

We have $K_\theta(t) = K(\theta + t) - K(\theta)$

But we already know $K_\theta'(0) = E(Y)$

$$\Rightarrow \left[K'(\theta+t) - K'(\theta)\right]\Big|_{t=0} = E(Y)$$

(take the derivative w.r.t t)

$$\Rightarrow \frac{d \log(M(\theta+t))}{dt} - 0 \Big|_{t=0} = E(Y)$$

$$\Rightarrow \frac{M'(\theta+t)}{M(\theta+t)}\Big|_{t=0} = E(Y)$$

$$\Rightarrow \frac{M'(\theta)}{M(\theta)} = E(Y)$$

$$\Rightarrow \frac{d \log(M(\theta))}{d\theta} = E(Y)$$

$$\Rightarrow K'(\theta) = E(Y)$$

Also $K_\theta''(0) = Var(Y)$

$$\Rightarrow \left[K''(\theta+t) - K''(\theta)\right]\Big|_{t=0} = Var(Y)$$

$$\Rightarrow \left[\left(\frac{M''(\theta+t)}{M(\theta+t)} - \frac{M'(\theta+t)^2}{M(\theta+t)^2}\right) - 0\right]\Big|_{t=0} = Var(Y)$$

$$\Rightarrow \frac{M''(\theta)}{M(\theta)} - \frac{M'(\theta)^2}{M(\theta)^2} = Var(Y)$$

$$\Rightarrow K''(\theta) = Var(Y)$$

5e) (i) $g(\beta) = \sum_{i=1}^{n} \ell(Y_i, x_i^T \beta)$

$$= \sum_{i=1}^{n} \{ h(Y_i) + x_i^T \beta Y_i - K(x_i^T \beta) \}$$

(1) (2) (3)

When we take the second derivative of $g(\beta)$, the terms (1) and (2) will disappear (with respect to β)

The only term left will be $- \nabla_\beta^2 K(x_i^T \beta)$

But we know that $K''(\theta) = Var(Y) \geq 0$

$$\Rightarrow \quad - \nabla_\beta^2 K(x_i^T \beta) \leq 0$$

$\Rightarrow$ The second derivative of $g(\beta)$ wrt β is negative semidefinite

$\Rightarrow$ $g(\beta)$ is concave in β

(ii) The MLE $\hat{\beta}$ is the value of β that makes the derivative of $g(\beta)$ wrt β equal 0

$$\Rightarrow \quad \left. \frac{\partial g(\beta)}{\partial \beta} \right|_{\beta = \hat{\beta}} = 0$$

$$\Rightarrow \quad \left. \frac{\partial}{\partial \beta} \sum_{i=1}^{n} \{ h(Y_i) + x_i^T \beta Y_i - K(x_i^T \beta) \} \right|_{\beta = \hat{\beta}} = 0$$

$$\Rightarrow \quad \sum_{i=1}^{n} x_i^T Y_i - \sum_{i=1}^{n} x_i^T K'(x_i^T \hat{\beta}) = 0$$

$$\Rightarrow \quad X^T Y - X^T E_{\hat{\beta}}(Y) = 0$$

$$\Rightarrow \quad X^T (Y - E_{\hat{\beta}}(Y)) = 0$$

$$\left(X = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix} ; \; Y = (Y_1, \dots Y_n)^T \right)$$

Scanned with CamScanner