JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

DEPARTMENT OF MATHEMATICS AND ACTUARIAL SCIENCE

PROGRAMMES

BFE 4, BST 4, BOR 4 AND BBS 4

LECTURE NOTES

ON

STA 2408: REGRESSION MODELLING II

BY

PROF. ANTHONY WAITITU

AND

THEOPHILUS ASAMOAH

.

SEPTEMBER, 2024 TO DECEMBER, 2024

Example 2.3: The Puromycin Data

Consider Bates and Watts (1988) usage of the Michaelis – Menten model for chemical kinetics to relate the initial velocity of an enzymatic reaction to the substrate concentration x. The data for the initial rate of a reaction for an enzyme treated with puromycin are shown in Table 3. The model is;

$$y = \frac{\theta_1 x}{x + \theta_2} + \epsilon \tag{1}$$

Table 1: Reaction Velocity and Substrate Concentration

i	**Substrate Concentration (ppm)**	**Velocity [(counts/min)]**
1	0.02	76
2	0.02	47
3	0.06	97
4	0.06	107
5	0.11	123
6	0.11	139
7	0.22	159
8	0.22	152
9	0.56	191
10	0.56	201
11	1.10	207
12	1.10	200

Estimate the parameters of the model using the linearisation method.

Solution to Example 2.3: The Puromycin Data

The plot of the data is shown in Figure 2.

Figure 2 shows a non-linear relationship between reaction velocity and substrate concentration. Hence, a non-linear model fitting is desirable.

Now, we apply the Gauss–Newton method to fit the Michaelis–Menten model to the puromycin data in Table 4 using the starting values;

$$\theta_{01} = 205 \tag{2}$$

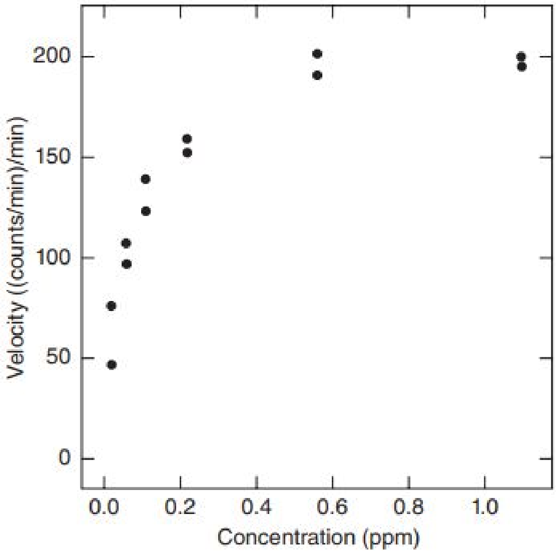

Figure 1: Reaction velocity versus substrate concentration

and

$$\theta_{02} = 0.08. \tag{3}$$

We shall discuss how these starting values were obtained later. At this starting points or values, the SSRes is;

$$\theta_0 = 3155 \tag{4}$$

Table 4 shows the data, fitted values, residuals, and derivatives evaluated at each observation.

Now, we explain the details of how the various quantities or columns in Table 4 were filled with the data given as follows:

1. For the f_i^0 column, the values are estimated as follows;

$$f_i^0 = f(x_i, \theta_0) = \frac{x_i \theta_{01}}{(x_i + \theta_{02})} \tag{5}$$

Consequently, from Table 3, for $i = 1$, we have;

$$f_1^0 = \frac{x_1 \theta_{01}}{(x_1 + \theta_{02})} = \frac{(205)(0.020)}{(0.080 + 0.020)} = 41.000 \tag{6}$$

Table 2: Data, Fitted Values, Residuals, and Derivatives at $\hat{\theta}'_0 = [205, 0.08]'$

i	x_i	y_i	f_i^0	$y_i - f_i^0$	Z_{i1}^0	Z_{i2}^0
1	0.02	76	41.00	35.00	0.2000	-410.00
2	0.02	47	41.00	6.00	0.2000	-410.00
3	0.06	97	87.86	9.14	0.4286	-627.55
4	0.06	107	87.86	19.14	0.4286	-627.55
5	0.11	123	118.68	4.32	0.5789	-624.65
6	0.11	139	118.68	20.32	0.5789	-624.65
7	0.22	159	150.33	8.67	0.7333	-501.11
8	0.22	152	150.33	1.67	0.7333	-501.11
9	0.56	191	179.38	11.62	0.8750	-280.27
10	0.56	201	179.38	21.62	0.8750	-280.27
11	1.10	207	191.10	15.90	0.9322	-161.95
12	1.10	200	191.10	8.90	0.9322	-161.95

Similarly, for $i = 12$, we have;

$$f_{12}^0 = \frac{x_{12}\theta_{01}}{(x_{12} + \theta_{02})} = \frac{(205)(1.10)}{(0.08 + 0.020)} = 200.000 \tag{7}$$

2. Furthermore, for the $y_i - f_i^0$ column, we have for $i = 1$;

$$y_1 - f_1^0 = 76 - 41 = 35.00 \tag{8}$$

Similarly, for $i = 12$, we have;

$$y_{12} - f_{12}^0 = 200 - 191 = 8.90 \tag{9}$$

3. For the Z_{ij}^0 columns, we note from the question that the that the partial derivatives with respect to θ_1 and θ_2 from the given model are;

$$Z_{i1}^0 = \frac{\partial f(x, \theta_1, \theta_2)}{\partial \theta_1} = \frac{x}{(\theta_2 + x)} \tag{10}$$

and

$$Z_{i2}^0 = \frac{\partial f(x, \theta_1, \theta_2)}{\partial \theta_2} = -\frac{\theta_1 x}{(\theta_2 + x)^2} \tag{11}$$

From Table 3, the first observation on x is $x_1 = 0.02$ and we have;

$$Z_{11}^0 = \left.\frac{x}{\theta_2 + x}\right|_{\theta_2=0.08} = \frac{0.020}{(0.08 + 0.020)} = 0.200 \tag{12}$$

Similarly, from Table 3, the second observation on x is $x_2 = 0.02$ and we have;

$$Z_{12}^0 = -\left.\frac{\theta_1 x}{\theta_2 + x}\right|_{\theta_1=205,\theta_2=0.08} = -\frac{(205)(0.020)}{(0.08 + 0.020)^2} = -410.000 \tag{13}$$

The process is continued and the Z_{ij}^0 are estimated at the observation on x_i.

4. Now, having completed all the columns of Table 4, the derivatives, $Z_{ij}^0 s$ are collected into the matrix Z_0, given as;

$$Z_0 = \begin{pmatrix} 0.2000 & -410.00 \\ 0.2000 & -410.00 \\ 0.4286 & -627.55 \\ 0.4286 & -627.55 \\ 0.5789 & -624.65 \\ 0.5789 & -624.65 \\ 0.7333 & -501.11 \\ 0.7333 & -501.11 \\ 0.8750 & -280.27 \\ 0.8750 & -280.27 \\ 0.9322 & -161.95 \\ 0.9322 & -161.95 \end{pmatrix} \tag{14}$$

5. The product $Z_0^T Z_0$, is;

$$Z_0^T Z_0 = \begin{pmatrix} 5.4623 & -2952.4956 \\ -2952.4956 & 2615993.8650 \end{pmatrix} \tag{15}$$

6. The matrix, Y_0 is given as;

$$Y_0 = \begin{pmatrix} 35.00 \\ 6.00 \\ 9.14 \\ 19.14 \\ 4.32 \\ 20.32 \\ 8.67 \\ 1.67 \\ 11.62 \\ 21.62 \\ 15.90 \\ 8.90 \end{pmatrix} \tag{16}$$

7. The product $Z_0^T Y_0$, is;

$$Z_0^T Y_0 = \begin{pmatrix} 94.370 \\ -68462.5022 \end{pmatrix} \tag{17}$$

and the vector of increments is;

$$\hat{\beta}_{\mathbf{0}} = \begin{pmatrix} 5.4623 & -2952.4956 \\ -2952.4956 & 2615993.8650 \end{pmatrix}^{-1} \begin{pmatrix} 94.370 \\ -68462.5022 \end{pmatrix} = \begin{bmatrix} 8.0288 \\ -0.0171 \end{bmatrix} \tag{18}$$

8. So, the revised estimates, $\hat{\theta}_1$ are;

$$\hat{\theta}_1 = \hat{\beta}_0 + \theta_0 = \begin{bmatrix} 8.03 \\ -0.017 \end{bmatrix} + \begin{bmatrix} 205.00 \\ 0.08 \end{bmatrix} = \begin{bmatrix} 213.03 \\ 0.063 \end{bmatrix} \tag{19}$$

9. So, the SSRes at these points is;

$$S(\hat{\theta}_1) = 1204.34 \tag{20}$$

which is considerably smaller than $S(\theta_0)$. Therefore, $\hat{\theta}_1$ is adopted as the revised estimate of θ, and subsequent iterations are performed following the same procedures. The SSRes at this point is estimated as presented in Table 5 as;

Table 3: Estimation of the SSR with the estimated θ_1 and θ_2

x_i	y_i	$\hat{y}$	$(y-\hat{y})$	$(y-\hat{y})^2$
0.0200	76.0000	51.33253	24.66747	608.48407
0.0200	47.0000	51.33253	-4.33253	18.77082
0.0600	97.0000	103.91707	-6.91707	47.84590
0.0600	107.0000	103.91707	3.08293	9.50444
0.1100	123.0000	135.45260	-12.45260	155.06728
0.1100	139.0000	135.45260	3.54740	12.58404
0.2200	159.0000	165.60636	-6.60636	43.64400
0.2200	152.0000	165.60636	-13.60636	185.13304
0.5600	191.0000	191.48764	-0.48764	0.23779
0.5600	201.0000	191.48764	9.51236	90.48498
1.1000	207.0000	201.49011	5.50989	30.35887
1.1000	200.0000	201.49011	-1.49011	2.22043
Total				**1204.34**

10. Upon successive iterations, the algorithm converged at;

$$\hat{\theta} = [212.7, 0.0641]^T = \begin{bmatrix} 212.7 \\ 0.0641 \end{bmatrix} \tag{21}$$

with,

$$S(\hat{\theta}) = 1195 \tag{22}$$

11. Therefore, the fitted model obtained by linearization is;

$$\hat{y} = \frac{\hat{\theta}_1 x}{x + \hat{\theta}_2} = \frac{212.7x}{x + 0.0641} \tag{23}$$

Example 2.4

Consider the observations in Table 2 and write R commands that fits the following non-linear regression model;

$$y = \theta_1 exp(\theta_2 x) + \epsilon \tag{24}$$

x	y
0.5	0.68
0.5	1.58
1	0.45
1	2.66
2	2.50
2	2.04
4	6.19
4	7.85
8	56.1
8	54.2
9	89.8
9	90.2
10	147.7
10	146.3

Table 4: Data for Non-linear Regression Model

Solutions to Example 2.4

To fit the non-linear regression model in (24) to the data, we approach the problem using a non-linear least squares method as outlined;

1. Define the non-linear model as given in (24), where, θ_1 and θ_2 are the parameters we are estimating, and ϵ is the error term.

2. Using the non-linear least squares to estimates, θ_1 and θ_2, we minimize the SSRes;

$$SS(\theta_1, \theta_2) = \sum_{i=1}^{n} [(y_i - \theta_1 exp(\theta_2 x_i))]^2 \tag{25}$$

   where, y_i and x_i are the observed values from the dataset.

3. Implement it in the R software with the following commands;

```
# Data input
x <- c(0.5, 0.5, 1, 1, 2, 2, 4, 4, 8, 8, 9, 9, 10, 10)
y <- c(0.68, 1.58, 0.45, 2.66, 2.50, 2.04, 6.19, 7.85,
56.1, 54.2,89.8, 90.2, 147.7, 146.3)

# Non-linear model fitting
nls_model <- nls(y ~ theta1 * exp(theta2 *
x), start = list(theta1 = 1, theta2 = 0.1))

# Summary of the fitted model
summary(nls_model)

# Get estimated coefficients
coef(nls_model)
```

4. Output: This command above will fit the model in (65) using the starting values for θ_1 and θ_2. It will then print the summary of the fitted model and the estimated coefficients.

5. Plot the data and add the fitted curve given as;

   Plot the data

   plot(x, y, main = "Nonlinear Regression Model", xlab = "X", ylab = "Y")

   Add fitted curve

   $curve(coef(nls_{underscore}model)[1]*exp(coef(nls_{underscore}model)[2]*x), add = TRUE, col = "blue")$

0.1 Asymptotic Properties of Non-Linear Least Squares Regression Parameters

The asymptotic properties of NLLS regression parameters are crucial for understanding how well the estimators behave as the sample size increases. These properties are similar to those of OLS in linear regression, but with more complexity due to the non-linearity of the model. These properties include consistency, asymptotic normality, and asymptotic efficiency. Below are the key asymptotic properties, their definitions and proofs;

0.1.1 Consistency

Definition

A parameter estimate $\hat{\theta}_n$ is said to be consistent if it converges in probability to the true parameter value θ_0 as the sample size $n \to \infty$, i.e.,

$$\hat{\theta}_n \xrightarrow{p} \theta_0. \tag{26}$$

Proof of Consistency

Consider the non-linear regression model;

$$Y_i = f(X_i, \theta) + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2) \tag{27}$$

where, θ is a $p \times 1$ vector of unknown parameters, X_i is a vector of predictors, and ϵ_i are i.i.d. errors. The NLLS estimator $\hat{\theta}_n$ minimizes the SSR;

$$SS_n(\theta) = \sum_{i=1}^{n} \left[Y_i - f(X_i, \theta)\right]^2. \tag{28}$$

Now, we have to show that $\hat{\theta}_n \to \theta_0$ in probability. But, we first look at the following key conditions:

1. **Identifiability:** The true parameter θ_0 is the unique minimizer of the expected sum of squares function given as;

$$S(\theta) = E\left[(Y_i - f(X_i, \theta))^2\right]. \tag{29}$$

2. **Continuity:** The function $f(X_i, \theta)$ and its partial derivatives with respect to θ are continuous.

As $n \to \infty$, by the Law of Large Numbers, the sample objective function $SS_n(\theta)$ converges uniformly in probability to the population objective function $S(\theta)$. Thus,:

$$\frac{1}{n} SS_n(\theta) \xrightarrow{p} S(\theta). \tag{30}$$

Since $S(\theta)$ is minimized uniquely at θ_0, the estimator $\hat{\theta}_n$ will converge to θ_0. Thus, $\hat{\theta}_n$ is consistent:

$$\hat{\theta}_n \xrightarrow{p} \theta_0. \tag{31}$$

END OF PROOF OF CONSISTENCY

0.1.2 Asymptotic Normality

Definition

The estimator $\hat{\theta}_n$ is asymptotically normal if and only if;

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \Sigma), \tag{32}$$

where, Σ is the asymptotic covariance matrix

Proof of Asymptotic Normality

We expand the first-order condition for the NLLS estimator around θ_0 using a Taylor expansion. The first-order condition for $\hat{\theta}_n$ is given by;

$$\frac{\partial}{\partial \theta} SS_n(\theta) = 0 \tag{33}$$

The gradient of the RSS is;

$$\frac{\partial SS_n(\theta)}{\partial \theta} = -2 \sum_{i=1}^{n} [Y_i - f(X_i, \theta)] \frac{\partial f(X_i, \theta)}{\partial \theta} \tag{34}$$

Expanding $f(X_i, \hat{\theta}_n)$ around θ_0 using a first-order Taylor approximation:

$$f(X_i, \hat{\theta}_n) \approx f(X_i, \theta_0) + \nabla_\theta f(X_i, \theta_0)(\hat{\theta}_n - \theta_0), \tag{35}$$

where, $\nabla_\theta f(X_i, \theta_0)$ is the gradient of f with respect to θ evaluated at θ_0. Substituting this into the first-order condition and solving for $\hat{\theta}_n - \theta_0$, we get;

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \approx \left(\frac{1}{n} \sum_{i=1}^{n} \nabla_\theta f(X_i, \theta_0)^\top \nabla_\theta f(X_i, \theta_0) \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_\theta f(X_i, \theta_0)^\top \epsilon_i. \tag{36}$$

By the Central Limit Theorem, the RHS of (77) converges in distribution to the normal distribution given as;

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_\theta f(X_i, \theta_0)^\top \epsilon_i \xrightarrow{d} N(0, \sigma^2 \Omega), \tag{37}$$

where,

$$\Omega = E\left[\nabla_\theta f(X_i, \theta_0)^\top \nabla_\theta f(X_i, \theta_0) \right]. \tag{38}$$

Thus, $\hat{\theta}_n$ is asymptotically normal with covariance matrix,

$$\Sigma = \sigma^2 \Omega^{-1} \tag{39}$$

leading to:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \sigma^2 \Omega^{-1}). \tag{40}$$

END OF PROOF OF ASYMPTOTIC NORMALITY

0.1.3 Efficiency

Definition

An estimator is asymptotically efficient if it achieves the Cramér-Rao lower bound (CRLB) in the asymptotic limit. This means it has the smallest possible variance among all consistent estimators.

Proof of Asymptotic Efficiency

The NLLS estimator is asymptotically efficient under the assumption that the errors ϵ_i are normally distributed. In the case of normally distributed errors, the NLLS estimator coincides with the MLE, which is known to be asymptotically efficient. For normally distributed errors, the likelihood function is;

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(Y_i - f(X_i, \theta))^2\right). \tag{41}$$

Maximizing the likelihood function is equivalent to minimizing the RSS, which is exactly what NLLS does. Therefore, the NLLS estimator attains the CRLB, implying that $\hat{\theta}_n$ is asymptotically efficient.

END OF PROOF OF ASYMPTOTIC EFFICIENCY

0.2 Assignment III: Estimation of Non-Linear Least Squares Regression Parameters

Consider the observations in Table 5

X	**Y**
0.02	76
0.02	47
0.06	97
0.06	107
0.11	123
0.11	139
0.22	159
0.22	152
0.56	191
0.56	201
1.10	207
1.10	200

Table 5: Data Table for X and Y values

and fit the non-linear regression model

$$y = \frac{\theta_1}{1 + \theta_2 x} + \epsilon \tag{42}$$

where;

1. y is the response variable,

2. x is the predictor variable,

3. θ_1 and θ_2 are the parameters to be estimated,

4. ϵ is the random error term.

to it using the following starting values;

$$\theta_{01} = 205 \tag{43}$$

and

$$\theta_{02} = 0.08. \tag{44}$$

NB: First, estimate the $SSRes(\theta_0)$ to guide you.