Group 11

Nonlinear Regression

The Gauss-Newton Estimation

Yilan LIANG 1830010014
Yunzhi XIAO 1830006195
Gengchen SUN 1830021031
Kaiyang LIU 1830004016
Jiaying YIN 1830006230

CONTENTS

CONTENTS

1.Motivation

2.The Gauss-Newton Method

3.Algorithm

4.Properties of Estimation

5.Example

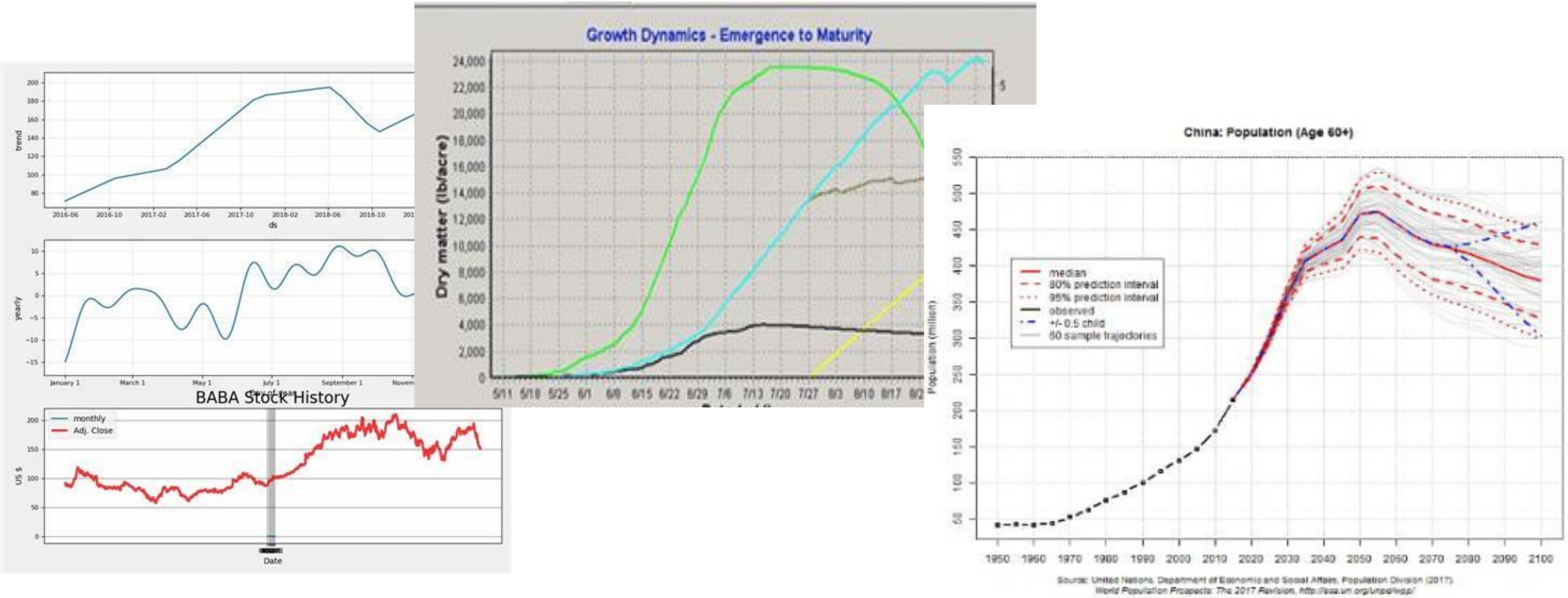
6.Discussion

Motivation

Motivation

- The problem of nonlinear least square
- The proposed of Gauss-Newton

The Nonlinear Models in the Natural World



Many models that reflecting the natural world is constructed in a non-linear way rather than a linear way

The Nonlinear Least Square

Example model:

$$y = \alpha e^{\beta x}$$

$$y = \frac{\alpha}{1 + \exp(-(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k))} + \varepsilon$$

$$y = \alpha + \beta_1 x_1^{\gamma_1} + \beta_2 x_2^{\gamma_2} + \cdots + \beta_k x_k^{\gamma_k} + \varepsilon$$

$$y = a * \exp\left[-\beta_1 e^{-\beta_2 x}\right] + \varepsilon$$

The Nonlinear Least Square

$$y = \alpha e^{\beta x}$$

Sum square of residuals：

$$SS_{Res} = \sum_{i=1}^{n} (y_i - \alpha e^{\beta x_i})^2$$

Take the derivative with respect to α

$$\sum_{i=1}^{n} (y_i - \alpha e^{\beta x_i})(-e^{\beta x_i}) = 0$$

Take the derivative with respect to β

$$\sum_{i=1}^{n} (y_i - \alpha e^{\beta x_i})(-\alpha e^{\beta x_i} * x_i) = 0$$

- Cannot be computed by elementary matrix algebra
- Need to solved by iteration way

The Gauss-Newton Method

The Gauss-Newton Method

The Gauss-Newton Method

- One of the most commonly used method to solve nonlinear least square
- An iteration method
- Basic idea:
  - utilize first-order Taylor expansion to approximate the original nonlinear regression model.
  - iteration by finding the least square estimator of the first-order Taylor expansion

Algorithm

Algorithm

1. Choose an initial point

The choosing of initial point:

- Empirical analysis on the approximate parameter value.
- Choosing specific values according to the property of the model.
- Plotting a graph and choosing an approximate extremum as the initial point.

2. Conduct the Taylor expansion

The Taylor expansion:

1. Conducting the first order Taylor expansion near $\boldsymbol{\theta} = \boldsymbol{\theta_0}$, where $\boldsymbol{\theta_0}$ is the initial point.

$$f(\boldsymbol{x_i}, \boldsymbol{\theta}) \cong f(\boldsymbol{x_i}, \boldsymbol{\theta_0}) + (\theta_1, \theta_{1,0}) \left[\frac{\partial f(\boldsymbol{x_i}, \boldsymbol{\theta})}{\partial \theta_1}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} + (\theta_2, \theta_{2,0}) \left[\frac{\partial f(\boldsymbol{x_i}, \boldsymbol{\theta})}{\partial \theta_2}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}$$

$$+..+(\theta_p, \theta_{p,0}) \left[\frac{\partial f(\boldsymbol{x_i}, \boldsymbol{\theta})}{\partial \theta_p}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} \qquad (i = 1,2, \dots n)$$

2. Represent the Taylor expansion in a linear way

$$f(\boldsymbol{x_i}, \boldsymbol{\theta}) - f(\boldsymbol{x_i}, \boldsymbol{\theta_0}) \cong \gamma_1 w_{1\,i} + \gamma_2 w_{2i} + \cdots + \gamma_p w_{p\,i} \qquad (i = 1,2, \dots, n)$$

where: $\gamma_{\boldsymbol{j}} = \theta_j - \theta_{j,0}$, $w_{ji} = \left[\frac{\partial f(\boldsymbol{x_i}, \boldsymbol{\theta})}{\partial \theta_1}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}$ $\begin{matrix} i = 1,2, \dots n, \\ j = 1,2, \dots p \end{matrix}$

3. The iteration

$$y_i - f(\boldsymbol{x_i}, \boldsymbol{\theta_0}) = \gamma_1 w_{1\,i} + \gamma_2 w_{2i} + \cdots + \gamma_p w_{p\,i} + \varepsilon_i \quad (i = 1,2,\ldots,n)$$

$$\varepsilon_i = y_i - \left(f(\boldsymbol{x_i}, \boldsymbol{\theta_0}) + \gamma_1 w_{1i} + \gamma_2 w_{2i} + \cdots + \gamma_p w_{pi}\right) \quad (i = 1,2,\ldots,n)$$

Least Square Method

$$S(\boldsymbol{\theta}) = \sum_{i=1}^{n} [y_i - (f(\boldsymbol{x_i}, \boldsymbol{\theta_0}) + \sum_{j=1}^{p} w_{ji}\gamma_j)]^2$$

$$\frac{\partial S}{\partial \theta_j} = -2\sum_{i=1}^{n} [y_i - (f(\boldsymbol{x_i}, \boldsymbol{\theta_0}) + \sum_{j=1}^{p} w_{ji}\gamma_j)] * w_{ji} = 0 \quad \longrightarrow$$

$$\sum_{i=1}^{n} w_{ji} \sum_{j=1}^{p} w_{ji}\,\gamma_j = \sum_{i=1}^{n} w_{ji}[y_i - f(\boldsymbol{x_i}, \boldsymbol{\theta}_0)]$$

Matrix Form

$$(W'W)\,\boldsymbol{\gamma}_j = W'\,[y_i - f(\boldsymbol{x_i}, \boldsymbol{\theta}_0)]$$

$$\boldsymbol{\gamma}_j = (W'W)^{-1}W'\,[y_i - f(\boldsymbol{x_i}, \boldsymbol{\theta}_0)]$$

$$\widehat{\boldsymbol{\theta}}_s = \widehat{\boldsymbol{\theta}}_{s-1} + (W'_{s-1}W_{s-1})^{-1}W'_{s-1}\left[\boldsymbol{y} - f(\widehat{\boldsymbol{\theta}}_{s-1})\right]$$

$$= \widehat{\boldsymbol{\theta}}_{s-1} + \boldsymbol{\gamma}_{s-1}$$

W is an $n \times$ p matrix whose (i, j) element is $\left[\partial f(\boldsymbol{x_i}, \boldsymbol{\theta})/\partial \theta_j\right]_{\theta=\theta_{s-1}}$

4. FRACTION INCREMENT

1. Use $\boldsymbol{\gamma}_{s-1} = (W_{s-1}' W_{s-1})^{-1} W_{s-1}' [\boldsymbol{y} - f(\boldsymbol{\theta}_{s-1})]$ to compute the standard Gauss-Newton increment vector for the s^{th} iteration (s=1,2,…).
2. Compute $\widehat{\boldsymbol{\theta}}_s = \widehat{\boldsymbol{\theta}}_{s-1} + \boldsymbol{\gamma}_{s-1}$ as the Gauss-Newton procedure suggests.
3. If $SS_{Res,s} < SS_{Res,s-1}$, continue to the next iteration using $\boldsymbol{\theta}_s$.
4. If $SS_{Res,s} > SS_{Res,s-1}$, go back to step 2; use $\boldsymbol{\gamma}_{s-1}/2$ as the vector of increments.

Flow

Determine the initial point

↓

Conduct the first order Taylor expansion

↓

Calculate the partial derivative and function value

↓

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_s &= \widehat{\boldsymbol{\theta}}_{s-1} + (W'_{s-1}W_{s-1})^{-1}W'_{s-1}\left[\boldsymbol{y} - f(\widehat{\boldsymbol{\theta}}_{s-1})\right] \\ &= \widehat{\boldsymbol{\theta}}_{s-1} + \boldsymbol{\gamma}_{s-1}\end{aligned}$$

Update the parameters

↓

If the parameters are convergence

If not

↓

Get the final parameters

Properties of Estimation

Properties of Estimation

Covariance

Important assumption : $\varepsilon \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2 \mathrm{I})$

n should be large enough to make the estimation approximately unbiased (idea from MLE)

Asymptotic variance-covariance matrix

$$Cov(\hat{\theta}) = s^2(W^T W)^{-1} \approx \begin{pmatrix} Var(\hat{\theta}_1) & \cdots & Cov(\hat{\theta}_p, \hat{\theta}_1) \\ \vdots & \ddots & \vdots \\ Cov(\hat{\theta}_1, \hat{\theta}_p) & \cdots & Var(\hat{\theta}_p) \end{pmatrix}$$

Similar as the covariance of parameters' estimators in linear regression, we have $Cov(\hat{\beta}) = \sigma^2(X^T X)^{-1}$

【Use s^2 to estimate σ^2 】

Where W is the Jacobian matrix evaluated at the least square estimates obtained from the final iteration.

Where $s^2 = \sum_{i=1}^{n} \frac{[y_i - f(x_i, \hat{\theta})]^2}{n-p} = \frac{SSE}{n-p}$

Similar as the residual mean square in linear regression, we have

$$s^2 = \hat{\sigma}^2 = \frac{Q(\hat{\beta})}{n-p} = \frac{SSE}{n-p}$$

【from the unbiased estimation of $Q(\hat{\theta})$ 】

Confidence Interval

$$Cov(\hat{\theta}) = s^2(W^TW)^{-1} \approx \begin{pmatrix} Var(\hat{\theta}_1) & \cdots & Cov(\hat{\theta}_p, \hat{\theta}_1) \\ \vdots & \ddots & \vdots \\ Cov(\hat{\theta}_1, \hat{\theta}_p) & \cdots & Var(\hat{\theta}_p) \end{pmatrix}$$

Therefore, the confidence interval for θ_i:

$$\hat{\theta}_i \pm t_{\frac{\alpha}{2}, n-p} * \sqrt{Var(\hat{\theta}_i)} = \hat{\theta}_i \pm t_{\frac{\alpha}{2}, n-p} * \sqrt{[s^2(W^TW)^{-1}]_{ii}}$$

Example

Example

Example

In a study to develop the growth behavior for protozoa colonization in a particular lake, an experiment was conducted in which 15 sponges were placed in a lake and 3 sponges at a time were gathered. Then the number of protozoa were counted at 1,3,6,15, and 21 days. In this case, the *MacArthur-Wilson Equation* was used to describe the growth mechanism.

The model is given by $y = S_{eq}(1 - e^{-g_0 t})$

Where :

$y: Total\ protozoa\ on\ the\ sponge;$

$S_{eq}: Species\ equilibrium\ constant;$

$g_0: Parameter\ that\ measures\ how\ quickly\ growth\ rises;$

$t: Time, number\ of\ days$

Observation	Day	y(Total Protozoa)
1	1	17
2	1	21
3	1	16
4	3	30
5	3	25
6	3	25
7	6	33
8	6	31
9	6	32
10	15	34
11	15	33
12	15	33
13	21	39
14	21	35
15	21	36

Goal:

1. Estimate S_{eq} and g_0 using nonlinear regression
2. Give estimated standard errors of the parameter estimates

Example

$$y = S_{eq}(1 - e^{-g_0 t})$$

$S_{eq} \approx 34$

$g_0 \approx 0.627$

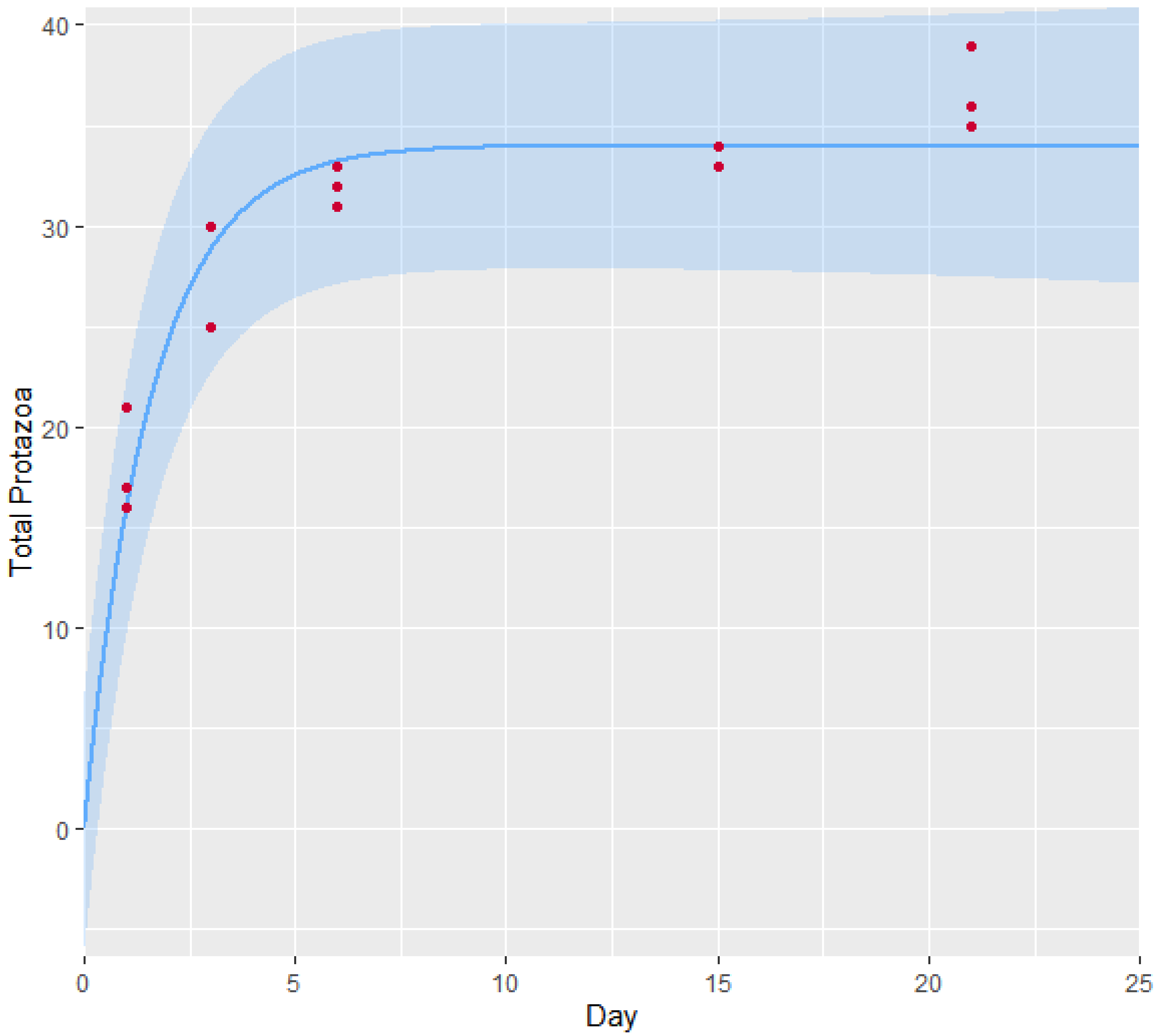
Confident interval:

$S_{eq} \in (31.93687,\ 36.06434)$
$g_0 \in (0.4601091, 0.7941507)$

Standard error of parameter

$S_{\widehat{S_{eq}}} = 0.95527080$

$S_{\widehat{g_0}} = 0.07731124$



1	3	6	15	21

Discussion

Discussion

Limitation

- Limitation of the algorithm
  - $\gamma's$ can be estimated very poorly in some problems
    - Hessian matrix $H = W(W'W)^{-1}W'$: not be positive definite
    - Stepsize: too large
    - Wrong signs may occur on the $\hat{\gamma}'s$
    - Poor starting values
  - Bad results of these problem
    - Not convergence
    - Slow convergence & large number of iteration
    - Allow the computed increments at a specific iteration to increase SS_{Res}
    - Procedure move in the wrong direction
    - Convergence to a local minimum of the SS_{Res} function

Modification to solve the above problems

- Fractional increments
  - Goal: not to allow the computed increments at a specific iteration to increase SS_{Res}
- Evaluate the Starting point (Using model $y = \alpha\left(1 - e^{\beta x}\right)$ as example)
  - Goal: Avoid convergence to a local minimum of the SS_{Res} function
  - Reasonable starting values can be determined from the data set
  - Consider the deterministic portion of the model $y = \alpha e^{\beta x}$
    - > Linearized form found by taking natural logs $\ln y = \ln \alpha + \beta x$
    - > Doing linear regression on $\ln y$ against x
    - > Starting value for α: the antilog of the intercept; Starting value of β: The slope of the regression
- Other modifications of the procedure
  - The Marquardt procedure
    - > Goal: Improve convergence
    - > Based on a revision of the iteration equation $\hat{\theta}_s = \hat{\theta}_{s-1} + \hat{\gamma}_{s-1}$

Limitation and implication of the model in our example

- Number of obeservations is small
  - Vunerable to chance
  - Difficult to identify outliers and influential points
- Laboratory conditions do not fully simulate the natural environment
  - Natural resources
  - Interspecies competition
- Practical implication of the model in our example
  - Effective conservation and use of wildlife resources
  - Maximizing the economic benefits of captive farms
  - Effective control of harmful animals

THANKS FOR LISTENLING