Estimation of Parameter Values

Nutrition Models Workshop

Luis Moraes
The Ohio State University
06/25/2017
Outline

• Nutrition models are VERY diverse
  • Combination of empirical, mechanistic, dynamic and static models
  • Regression, linear and nonlinear mixed models, differential equations

• Today: Main approaches for estimating parameters in a variety of models
  • Some mathematical description
  • Idea is for you to understand the reasoning and challenges of different approaches

• One exercise/demonstration in the end
  • Fit model with two approaches
Introduction

• Different types of models have been used for nutrition modeling
  • Compartmental, regression, meta-analysis, nonlinear mixed models, .....

• One feature is common to almost all these models
  • Parameters are needed to describe the system
  • Quantify relationship between variables
Introduction

• Simple example: linear regression

\[ Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \]

- \( Y_i \) is the response variable for the \( i \)th observation
- \( x_i \) is the predictor variable in the \( i \)th observation
- \( \beta_0 \) is the intercept
- \( \beta_1 \) is the slope
- \( \epsilon_i \) is the error, \( E[\epsilon_i] = 0 \), \( \text{Var}[\epsilon_i] = \sigma^2 \) and \( \epsilon_i \) are independent
- \( i = 1, \ldots, n \)

In matrix notation:

\[ \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix} \]
Introduction

• In practice, parameter true values are unknown
  • Estimators from a sample

• Parameters have to be optimal in some sense
  • Least square estimators minimize squared errors
  • Maximum likelihood estimators maximize the likelihood function
Least Squares Estimators
Least Square Estimators

• The least squares estimators minimize the square errors:
\[ Q = \sum_{i=1}^{n} \left[ Y_i - (\beta_0 + \beta_1 x_i) \right]^2 \]

• How do we get them???

• We can find points of minimum and maximum of a function using derivatives.

For example for \( f(x) = 160x - 16x^2 \)

Set derivative to zero and “solve” for \( x \):

\[ 160 - 32x = 0 \]
\[ x = 5 \]

Second derivative test: \(-32\)
Least Squares Estimators

\[
\frac{\partial Q}{\partial \beta_0} = -2 \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_i)
\]

\[
\frac{\partial Q}{\partial \beta_1} = -2 \sum_{i=1}^{n} x_i (Y_i - \beta_0 - \beta_1 x_i)
\]

Setting these partial derivatives to zero, we construct the normal equations

\[
\sum_{i=1}^{n} Y_i = nb_0 + b_1 \sum_{i=1}^{n} X_i
\]

\[
\sum_{i=1}^{n} X_i Y_i = b_0 \sum_{i=1}^{n} X_i + b_1 \sum_{i=1}^{n} X_i^2
\]
Least Squares Estimators

• The least square estimators are the solutions to the normal equations

\[
b_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}
\]

\[
b_0 = \bar{Y} - b_1 \bar{x}
\]

• The concept extends to multiple regression

\[
Q = \sum_{i=1}^{n} \left[ Y_i - (\beta_0 + \beta_1 x_{i1} + \beta_{p-1} x_{i,p-1}) \right]^2
\]

• General form of the least squares estimators: \( \mathbf{b} = (\mathbf{X}^\mathbf{T} \mathbf{X})^{-1} \mathbf{X}^\mathbf{T} \mathbf{Y} \)
Least Squares Estimators

• Estimates of the uncertainty associated with these parameters

• Estimator of the error’s variance

\[
MSE = \frac{1}{n-p} \sum_{i=1}^{n} \left[ Y_i - \left( b_0 + b_1 x_{i1} + \cdots + b_{p-1} x_{i,p-1} \right) \right]^2
\]

• Estimated variance-covariance matrix of the parameters

\[
MSE \left( X^T X \right)^{-1}
\]
Nonlinear Models

• So far, we can estimate parameters in linear models

• Many phenomena in biology are nonlinear
  • For example, reaction velocity vs. substrate concentration in an enzymatic reaction
  • Before we start with nonlinear models, let’s clarify

\[ Y_i = \frac{\theta_1}{1 + \theta_2 \exp(\theta_3 x_i)} + \varepsilon_i \] is a nonlinear model

\[ Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i \] is a linear model
Nonlinear Models

Michaelis-Menten Kinetics

\[ v = \frac{V_{max} \cdot [S]}{K_M + [S]} \]
Nonlinear Regression

\[ Y_i = f(x_i, \theta) + \varepsilon_i \]

\( f \) is the nonlinear function describing the relationship between \( Y \) and \( x \)

Michaelis-Menten example:
\[ y_i = \frac{V_{\text{max}} x_i}{K_M + x_i} + \varepsilon_i \quad \Rightarrow \quad f(x_i, \theta) = \frac{V_{\text{max}} x_i}{K_M + x_i} \]

- \( y_i \) is the reaction rate for the \( i \)th observation
- \( x_i \) is the associated substrate concentration

\( \theta = (V_{\text{max}}, K_M)^T \) are the parameters to be estimated

- \( \varepsilon_i \) is the error, \( E[\varepsilon_i] = 0, \text{Var}[\varepsilon_i] = \sigma^2 \) and independent
- \( i = 1, \ldots, n \)
Least Squares Estimators

• For the simple linear regression model, least squares minimize

\[ Q = \sum_{i=1}^{n} \left( Y_i - (\beta_0 + \beta_1 x_i) \right)^2 \]

• For the nonlinear regression, the idea is the same: minimize

\[ Q = \sum_{i=1}^{n} \left( Y_i - f(x_i) \right)^2 \]
Least Squares Estimators

• Solution to the normal equations are often difficult to obtain analytically

• Numerical Algorithms
  • For example, Gauss-Newton
  • Require initial values to initialize numerical procedures
Gauss-Newton

• Default in PROC NLIN and nls()

• Approximate the nonlinear model with linear terms

• Taylor series expansion and least squares as for linear regression

• Denote the least squares estimates $g$ and the initial values $g^{(0)} = (g_0^{(0)}, g_1^{(0)}, \ldots, g_{p-1}^{(0)})$

• Approximation around starting values:

$$f(x_i, \theta) \approx f(x_i, g^{(0)}) + \sum_{k=0}^{p-1} \left[ \frac{\partial f(x_i, \theta)}{\partial \theta_k} \right]_{\theta=g^{(0)}} \left( \theta_k - g_k^{(0)} \right)$$
Gauss-Newton

• Model approximation

\[ Y_i - f(x_i, \mathbf{g}^{(0)}) \approx \sum_{k=0}^{p-1} \left[ \frac{\partial f(x_i, \theta)}{\partial \theta_k} \right]_{\theta=\mathbf{g}^{(0)}} (\theta_k - g_k^{(0)}) + \epsilon_i \]

• It is a linear model!

• Estimate parameters by least squares: \[ \mathbf{b}^{(0)} = \left( \mathbf{D}^{(0)T} \mathbf{D}^{(0)} \right)^{-1} \mathbf{D}^{(0)T} \mathbf{Y}^{(0)} \]

• Update: \[ \mathbf{g}^{(1)} = \mathbf{g}^{(0)} + \mathbf{b}^{(0)} \]
Gauss-Newton

• Evaluation criteria: \( SSE^{(0)} = \sum_{i=1}^{n} (Y_i - f_i^{(0)})^2 \)

• Start the process again with \( g^{(1)} \) as the initial values

• Repeat procedure until \( SSE^{(s+1)} - SSE^{(s)} \) is negligible

• Estimate of error’s variance: \( MSE = \frac{\sum_{i=1}^{n} [Y_i - f(x_i,g)]^2}{n-p} \)

• Other methods available, e.g. Nelder-Mead and Marquardt
Compartmental Models

- Traditionally used in nutritional modeling
  - Roots on pharmacokinetics and differential calculus
Compartmental Models

• Functional forms described in terms of differential equations
  • Instead of the “integrated form”

• Strategy for parameter estimation
  • Expected mean represented by a compartmental model $f$
  • If $f$ cannot be obtained analytically, it has to be solved numerically
    • Euler, Runge-Kutta4, lsoda
  • Can use nonlinear least squares but have to numerically solve $f$ at iteration
  • Modern software estimate using maximum likelihood
Maximum Likelihood Estimation

• Another strategy for parameter estimation

• For regression models with independent \( \varepsilon_i \sim N(0, \sigma^2) \), estimators coincide with least squares estimators

• Estimators maximize the likelihood function
  • Parameter values that are in best agreement with the data
Maximum Likelihood Estimation

\[ p(y | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left( -\frac{(y - \mu)^2}{2\sigma^2} \right) \]

- \( p(y | \mu, \sigma^2) \) is the density function: How likely \( y \) is at each value
- The likelihood function is: \( L(\mu, \sigma^2 | y_1, \ldots, y_n) = p(y_1 | \mu, \sigma^2) \times \ldots \times p(y_n | \mu, \sigma^2) \)
  - “How likely the whole data is with that set of parameters values”
  - MLE: “maximize the likelihood of getting the observed data”
Maximum Likelihood Estimation

• Linear Regression Example: \( Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \)

\[
L(\beta_0, \beta_1, \sigma^2 \mid y_1, \ldots, y_n) = \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left[ -\frac{1}{2\sigma^2} \sum_{i} (y_i - \beta_0 - \beta_1 x_i)^2 \right]
\]

• It is easier to work with the log-likelihood

\[
\log L(\beta_0, \beta_1, \sigma^2 \mid y_1, \ldots, y_n) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i} (y_i - \beta_0 - \beta_1 x_i)^2
\]

• To find parameters that maximize the likelihood we...
  • Take the derivative with respect to each parameter and set to zero. Also need second derivative test
Mixed Models

- Modern mixed modeling relies heavily on likelihood methods
- Extension of (non)linear models with both fixed and random effects
- Probably the “type” of model you need to analyze your data or construct your nutrition model
- There’s a whole workshop on mixed models in this meeting
Nonlinear Mixed Models

• Nonlinear functional forms
  • Michaelis-Menten, logistic, exponential, Gompertz, ...

• Random effects that “enter the model” nonlinerly

• Allow you to model nonlinear clustered, longitudinal data
  • Records from the same animal, treatment means from the same study
Nonlinear Mixed Models

\[ y_{ij} = f(x_{ij}, \theta_i) + \varepsilon_{ij} \]

• \( y_{ij} \) is the \( j \)th record on the \( i \)th “subject” or cluster
• \( x_{ij} \) is the associated predictor variable
• \( \theta_i \) is the vector of subject specific parameters

\[ \theta_i = \beta + b_i \quad b_i \sim N(0, \Psi) \]

• \( \varepsilon_{ij} \) is the random error \( \sim N(0, \sigma^2) \)
Nonlinear MM Maximum Likelihood

• There is more than one source of variability
  • Between subjects and within subjects

• To represent the generative process of the data we need to take both into account
  • Joint density of the response and the random effects
Maximum Likelihood

- For the linear regression

\[ p(y \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - \mu)^2}{2\sigma^2}\right) \]

- The likelihood function is:

\[ L(\mu, \sigma^2 \mid y_1, \ldots, y_n) = p(y_1 \mid \mu, \sigma^2) \times \ldots \times p(y_n \mid \mu, \sigma^2) \]

- For the nonlinear mixed model, we need to compute the marginal density of the responses:

\[ p(y \mid \beta, \sigma^2, \Psi) = \int p(y \mid \beta, \sigma^2)p(b \mid \Psi) \, db \]
Maximum Likelihood

- Bad news from an estimation perspective

- The likelihood function to estimate the parameters requires integrating the joint density with respect to the random effects

- The integral often does not have a closed form expression

- Approximation of the likelihood function
Exercise

• Let’s go back to the compartmental model

\[
\begin{align*}
\frac{dy_1}{dt} &= -k_a y_1 \\
\frac{dy_2}{dx} &= k_a y_1 - k_e y_2
\end{align*}
\]

• \( y_1 \) is the amount of drug in the gut compartment
• \( y_2 \) is the amount of drug in the blood compartment
• \( k_a \) is the absorption rate (1/hr)
• \( k_e \) is the elimination rate (1/hr)
Exercise

• Data from Davidian and Giltinan (1995)

• 12 subjects received a single oral dose of theophylline
  • Anti-asthmatic drug
  • Single oral dose at time zero

• Measurements of blood concentrations of drug at 11 time points over a 25 hour period
## Exercise

<table>
<thead>
<tr>
<th>Time (hr)</th>
<th>Concentration (mg / L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

![Graph showing concentration over time](image-url)
Exercise

• We only observe data from the second compartment

• The differential equation for the second compartment can actually be solved analytically

• We will estimate the parameters in two different ways
  • Analytical solutions: nonlinear mixed model
  • Numerical solutions for differential equations in a mixed model framework
Exercise

• The analytical solution to the second differential equation is

\[ C(t) = \frac{Dose k_e k_a}{Cl (k_a - k_e)} \left[ \exp(-k_e t) - \exp(-k_a t) \right] \]

• As a nonlinear mixed model

\[ C_{ij} = \frac{Dose_i k_e k_{a,i}}{Cl_i (k_{a,i} - k_e)} \left[ \exp(-k_e t_{ij}) - \exp(-k_{a,i} t_{ij}) \right] + \varepsilon_{ij} \]

\[ \varepsilon_{ij} \sim N(0, \sigma^2) \quad \text{and} \quad \begin{bmatrix} Cl_i \\ k_{a,i} \end{bmatrix} \sim N\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{Cl} & 0 \\ 0 & \sigma^2_{k_a} \end{bmatrix} \right) \]
Exercise

• All three parameters must be positive

• Reparameterize model with parameters on a log scale

\[ C(t) = \frac{Dose \exp(lk_e + lk_a - lCl)}{\exp(lk_a) - \exp(lk_e)} \{\exp[-\exp(lk_e)t] - \exp[-\exp(lk_a)t]\} \]

where

\[ lk_e = \log(k_e) \quad , \quad lk_a = \log(k_a) \quad \text{and} \quad lCl = \log(Cl) \]
Exercise

• First method
  • Fit nonlinear mixed model in R
Exercise

• Second method
  • Nonlinear mixed model but solve differential equations numerically: lsoda solver

\[
\begin{align*}
\frac{dy_1}{dt} &= -k_ay_1 \\
\frac{dy_2}{dx} &= k_a y_1 - k_e y_2 \\
y_1(0) &= \text{Dose} \\
y_2(0) &= 0
\end{align*}
\]

• Observation Equation: 
\[
C(t) = \frac{y_2(t)}{Cl} k_e
\]
Bayesian Inference

• Combines prior information with new data: update of knowledge

• All parameters are treated as random variables
  • Prior distributions for parameters
  • Inference is based on the posterior distribution
  • Bayes theorem

\[ p(\theta | y) \propto g(\theta) L(\theta | y) \]

From: https://www.r-bloggers.com/the-beta-prior-likelihood-and-posterior/
Bayesian Inference

• Inference based on posterior
  • Combines prior information with the observed data

• Particularly suited for models built with many parameters that good biological knowledge is available

• Many freely software available

• Including with differential equations “solvers”

• We don’t have to have a known or tractable posterior: Markov Chain Monte Carlo (MCMC)
Example of a Multivariate Nonlinear Model


\[ F \cdot (\text{ME} - \text{ME}_M) \]

\[ (1-F) \cdot (\text{ME} - \text{ME}_M) \]

\[ \text{ME}_{PD} \]
\[ \text{ME}_M \]
\[ \text{ME}_{LD} \]

\[ k_p \]
\[ 1-k_p \]
\[ 1-k_f \]
\[ k_f \]

PD
HP
LD

Courtesy of A. B. Strathe and adapted after van Milgen and Noblet (1999)
Example for a Multivariate Nonlinear Model

Multivariate Model:

\[
PD = \frac{PD_{max}}{BW_{PD_{max}}} \cdot BW \cdot \log \left( \frac{BW_{PD_{max}} \cdot \exp(1)}{BW} \right)
\]

\[
LD = k_f \left( ME - a \cdot BW^b - PD / k_p \right)
\]

Priors from literature

<table>
<thead>
<tr>
<th>(k_f)</th>
<th>(k_p)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.72-0.88</td>
<td>0.52-0.63</td>
<td>Strathe et al. (2010a)*</td>
</tr>
<tr>
<td>0.75</td>
<td>0.56</td>
<td>van Milgen et al. (2000)*</td>
</tr>
<tr>
<td>0.77-0.82</td>
<td>0.58-0.60</td>
<td>van Milgen &amp; Noblet (1999)*</td>
</tr>
<tr>
<td>0.84</td>
<td>0.62</td>
<td>Noblet et al. (1999)†</td>
</tr>
<tr>
<td>0.76</td>
<td>0.54</td>
<td>NRC (1998)†</td>
</tr>
<tr>
<td>0.60</td>
<td>0.52</td>
<td>Tess et al. (1984)†</td>
</tr>
<tr>
<td>0.74</td>
<td>0.56</td>
<td>ARC (1981)†</td>
</tr>
</tbody>
</table>

* Partial efficiencies derived from multivariate modelling approaches

\(k_p \sim N(0.60, 0.10^2)\) and \(k_f \sim N(0.80, 0.10^2)\)

Credible Intervals for the Posterior

\(b \sim (0.60, 0.61)\), \(k_p \sim (0.59, 0.58)\), \(k_f \sim (0.78, 0.76)\)