Regression Methods of Estimation

BIOS 6611

CU Anschutz

Week 15

1. Methods of Estimation

2. Maximum Likelihood Estimation

Methods of Estimation

Methods of Estimation

There are three primary methods to estimation used in BIOS 6611/12 for the different regression approaches explored:

1. **Least Squares Estimation:** our primary focus this semester where we minimize the sum of square error
2. **Maximum Likelihood Estimation:** an approach that maximizes the likelihood function to derive parameter estimates and is used for linear mixed effects models to handle correlated data
3. **Generalized Linear Models:** an approach that ultimately provides us with linear, logistic, Poisson, etc. regression models, often based on maximum likelihood estimation

We will introduce these last two to begin building the framework for extending to new regression approaches next semester.

Maximum Likelihood Estimation

Maximum Likelihood Estimation for Regression Parameters

Consider the multiple linear regression model fit using a random sample of n individuals:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \epsilon$$

The observed data for the i^{th} subject is given by $(Y_i, X_{i1}, ..., X_{ip}), i = 1, 2, ..., n$.

We will assume that the Y_i are normally distributed with variance $Var(Y_i) = \sigma^2$ not varying with i and that the **X** are measured without error.

We must also assume that the n random variables $Y_1, Y_2, \dots, Y_n$ are mutually independent.

Ultimately we wish to estimate $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, ...\beta_p, \sigma^2)$.

A Note About Our Independence Assumption for Y_i

We are assuming that the n random variables $Y_1, Y_2, \ldots, Y_n$ are mutually independent.

This allows the precise description of the joint distribution of the variables (i.e., the likelihood function) solely on the basis of knowledge of the separate behavior (i.e., the so-called marginal distribution) of each variable in the set (the product of the marginal distributions).

MLE Set-Up: Y_i's PDF

- Recall, the probability density function (pdf) for the normal distribution is $f(Y|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(Y_i-\mu)^2\right)$.
- In our regression framework $E(Y) = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p$.
- Noting that $E(Y) = \mu$, we can write the probability density function for our multiple regression's normally distributed random variable Y_i as

$$f(Y_i|\beta_0,\beta_1,...,\beta_p,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(Y_i - (\beta_0 + \beta_1 X_{i1} + ... + \beta_p X_{ip})\right)^2\right)$$

- This corresponds to $Y|(X_1,...,X_p) \sim N(\beta_0 + \beta_1 X_1 + ... + \beta_p X_p, \sigma^2)$ (where we have replaced μ with our regression equation in $Y \sim N(\mu,\sigma^2)$).

MLE Set-Up: The Likelihood Function

$f(Y_i|\beta_0, \beta_1, ..., \beta_p, \sigma^2)$ is the probability density function for the i^{th} observation.

What we still need is the *joint probability* of $Y_1 = y_1, Y_2 = y_2, ..., Y_n = y_n$. This is given by our *likelihood* of the data:

$$
\begin{aligned}
L(\boldsymbol{\theta}|\mathbf{Y}) =& L(\beta_0, \beta_1, ..., \beta_p, \sigma^2|\mathbf{Y}) \\
=& \prod_{i=1}^{n} f(Y_i|\beta_0, \beta_1, ..., \beta_p, \sigma^2) \\
=& \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\left(Y_i - (\beta_0 + \beta_1 X_{i1} + ... + \beta_p X_{ip})\right)^2\right\} \\
=& \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(Y_i - (\beta_0 + \beta_1 X_{i1} + ... + \beta_p X_{ip})\right)^2\right\}
\end{aligned}
$$

MLE Set-Up: The Log-Likelihood

Here we will use two properties of the log: (1) $\log(abc) = \log(a) + \log(b) + \log(c)$ and (2) $\log(z^a) = a\log(z)$.

Taking the (natural) log of the likelihood function makes it easier to work with our complex expression.

$\log(L(\beta_0, \beta_1, ..., \beta_p, \sigma^2|\mathbf{Y}))$

$$
\begin{aligned}
&= \log\left\{\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - (\beta_0 + \beta_1 X_{i1} + ... + \beta_p X_{ip}))^2\right)\right\} \\
&= \frac{-n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - (\beta_0 + \beta_1 X_{i1} + ... + \beta_p X_{ip}))^2 \\
&= \frac{-n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{X}\beta)^T(\mathbf{Y} - \mathbf{X}\beta)
\end{aligned}
$$

MLE of β

Now that we have the log-likelihood, we can take the first derivative with respect to our parameters of interest (i.e., $\beta_0, \beta_1, ..., \beta_p, \sigma^2$), set it equal to 0, and solve for our estimates.

For $\boldsymbol{\beta}$, we will use the matrix form from the previous slide:

$$\frac{\partial LL(\boldsymbol{\beta}, \sigma^2 | \mathbf{Y})}{\partial \boldsymbol{\beta}} \propto \frac{\partial}{\partial \boldsymbol{\beta}} (\mathbf{Y} - \mathbf{X}\beta)^T (\mathbf{Y} - \mathbf{X}\beta) = \frac{\partial SS_{Error}}{\partial \boldsymbol{\beta}}$$

This is identical to our least squares estimators! Therefore, the MLE estimates for the $\boldsymbol{\beta}$'s are the same.

MLE of σ^2

However, the ML estimator $\hat{\sigma}^2$ of σ^2 results in a biased estimate:

$$\frac{\partial LL(\beta, \sigma^2|\mathbf{Y})}{\partial \sigma^{\mathbf{2}}} = -\frac{n}{2}\left(\frac{1}{\sigma^2}\right) + \frac{1}{2(\sigma^2)^2}(\mathbf{Y} - \mathbf{X}\beta)^T(\mathbf{Y} - \mathbf{X}\beta) = 0$$

Solving now for the MLE:

$\frac{n}{2}\left(\frac{1}{\hat{\sigma}^2}\right) = \frac{1}{2(\hat{\sigma}^2)^2}(\mathbf{Y} - \mathbf{X}\hat{\beta})^T(\mathbf{Y} - \mathbf{X}\hat{\beta})$

$n\hat{\sigma}^2 = (\mathbf{Y} - \mathbf{X}\hat{\beta})^T(\mathbf{Y} - \mathbf{X}\hat{\beta})$

$\hat{\sigma}^2 = \frac{(\mathbf{Y}-\mathbf{X}\hat{\beta})^T(\mathbf{Y}-\mathbf{X}\hat{\beta})}{n} = \frac{SS_{Error}}{n} = \frac{n-p-1}{n}\left(\frac{SS_{Error}}{n-p-1}\right) = \frac{n-p-1}{n}\hat{\sigma}^2_{Y|X} = \frac{n-p-1}{n}MSE$

We therefore see the maximum likelihood estimator of $\hat{\sigma}^2$ is biased:

$$E(\hat{\sigma}^2) = \left(\frac{n-p-1}{n}\right)\sigma^2_{Y|X} \neq \sigma^2_{Y|X}$$

MLE of σ^2

In ordinary least squares estimation, we had

$$\hat{\sigma}^2_{Y|X} = MS_{Error} = \frac{SS_{Error}}{n-p-1}$$

For our maximum likelihood estimate, we have

$$\hat{\sigma}^2_{MLE} = \left(\frac{n-p-1}{n}\right) \sigma^2_{Y|X}$$

As n, our sample size, increases, $\frac{n-p-1}{n}$ results in only minor differences between the OLS and MLE estimators. When sample sizes are small, there can be larger differences in the two approaches.

Generalized Linear Models

In our next lecture we will briefly introduce the concept of generalized linear models (GLMs), a flexible framework for modeling a wide variety of outcome types.