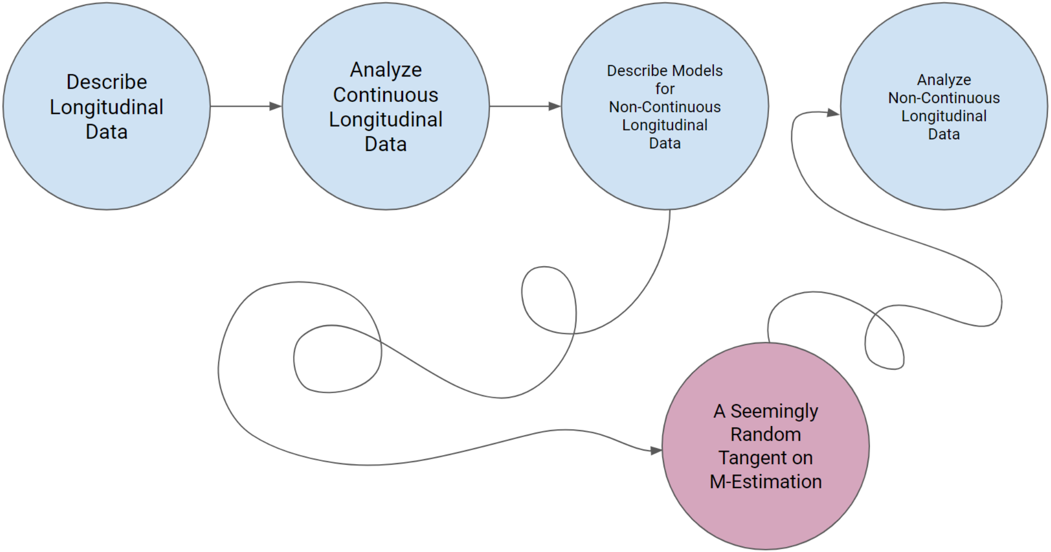
M-Estimation (A Practicing Statisticians Best Friend)

Path Through this Course



What is M-Estimation?

- M-Estimation is an **estimation technique** similar to **Maximum Likelihood Estimation** or **Least Squares Estimation**.

What is M-Estimation?

- M-Estimation is an **estimation technique** similar to **Maximum Likelihood Estimation** or **Least Squares Estimation**.
- In fact, both *MLEs* and *LSEs* are **special cases** of M-estimators.

What is M-Estimation?

- M-Estimation is an **estimation technique** similar to **Maximum Likelihood Estimation** or **Least Squares Estimation**.
- In fact, both *MLEs* and *LSEs* are **special cases** of M-estimators.
- The general theory (of unbiased estimating equations) was explored in the 1960s by various authors.

What is M-Estimation?

- M-Estimation is an **estimation technique** similar to **Maximum Likelihood Estimation** or **Least Squares Estimation**.
- In fact, both *MLEs* and *LSEs* are **special cases** of M-estimators.
- The general theory (of unbiased estimating equations) was explored in the 1960s by various authors.
- Provides an **incredibly flexible framework** for practical implementation of estimation, and **asymptotic analysis**.

What is an M-Estimator

An **M-estimator**, $\widehat{\theta}$, is an estimator for some parameter θ, which is given by the solution to a set of equations,

$$\sum_{i=1}^{n} U(Y_i; \widehat{\theta}) = 0.$$

For instance, $S(\theta) = \frac{\partial}{\partial\theta}\ell(\theta) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta}\log(f(y;\theta))$ is the **Score function**, which we solve $S(\widehat{\theta}) = 0$ for the MLE.

For **least squares estimators**, we wish to minimize $L(\theta) = \sum_{i=1}^{n}(g(Y_i) - h(\theta))^2$, for some suitable $g(\cdot), h(\theta)$, which by defining $L'(\theta) = \sum_{i=1}^{n}(g(Y_i) - h(\theta))h'(\theta)$, is given by $L'(\widehat{\theta}) = 0$.

Why do we care?

- The general theory of M-estimators is **very powerful**.

Why do we care?

- The general theory of M-estimators is **very powerful**.
- Suppose that θ_0 is a value such that, $E[U(Y_i;\theta_0)] = 0$. **Then**

Why do we care?

- The general theory of M-estimators is **very powerful**.
- Suppose that θ_0 is a value such that, $E[U(Y_i;\theta_0)] = 0$. **Then**
  - $\widehat{\theta}$ is consistent for θ_0.

Why do we care?

- The general theory of M-estimators is **very powerful**.
- Suppose that θ_0 is a value such that, $E[U(Y_i;\theta_0)] = 0$. **Then**
  - $\widehat{\theta}$ is consistent for θ_0.
  - $\widehat{\theta}$ will have an asymptotic distribution of

$$N(\theta_0, \mathbf{A}(\theta_0)^{-1}\mathbf{B}(\theta_0)\mathbf{A}(\theta_0)^{-1'}).$$

Why do we care?

- The general theory of M-estimators is **very powerful**.
- Suppose that θ_0 is a value such that, $E[U(Y_i;\theta_0)] = 0$. **Then**
  - $\widehat{\theta}$ is consistent for θ_0.
  - $\widehat{\theta}$ will have an asymptotic distribution of

$$N(\theta_0, \mathbf{A}(\theta_0)^{-1}\mathbf{B}(\theta_0)\mathbf{A}(\theta_0)^{-1'}).$$

  - Here we have

$$\mathbf{A}(\theta_0) = E\left[-\frac{\partial}{\partial\theta}U(Y_i;\theta)\bigg|_{\theta=\theta_0}\right],$$

    and

$$\mathbf{B}(\theta_0) = E\left[U(Y_i;\theta_0)U(Y_i;\theta_0)'\right].$$

Why do we care?

- The general theory of M-estimators is **very powerful**.
- Suppose that θ_0 is a value such that, $E[U(Y_i;\theta_0)] = 0$. **Then**
  - $\widehat{\theta}$ is consistent for θ_0.
  - $\widehat{\theta}$ will have an asymptotic distribution of

$$N(\theta_0, \mathbf{A}(\theta_0)^{-1}\mathbf{B}(\theta_0)\mathbf{A}(\theta_0)^{-1'}).$$

  - Here we have

$$\mathbf{A}(\theta_0) = E\left[-\frac{\partial}{\partial\theta}U(Y_i;\theta)\bigg|_{\theta=\theta_0}\right],$$

    and

$$\mathbf{B}(\theta_0) = E\left[U(Y_i;\theta_0)U(Y_i;\theta_0)'\right].$$

- That is, if the **estimating equation is unbiased**, then we *automatically* know the asymptotic distribution of the estimator!

Example 1

We can take $U(\mathbf{Y}, \theta) = \sum_{i=1}^{n} Y_i - \theta$.

The estimator $\widehat{\theta}$ will be the **sample mean!**

Example 2

We can take

$$U(\mathbf{Y}, \theta) = \sum_{i=1}^{n} \begin{pmatrix} Y_i - \theta_1 \\ (Y_i - \theta_1)^2 - \theta_2 \end{pmatrix}.$$

These will be **consistent moment estimators**, for $E[Y_i]$ and $\text{var}(Y_i)$!

Example 3

We have seen that $E[S(\theta)] = 0$ for the score function.

Let's consider what the **asymptotic distribution of MLEs** are then!

Example 4

When reviewing **GLMs** we discussed *quasi-likelihood estimation*, in which we defined

$$U(Y_i, \mu_i) = \frac{Y_i - \mu_i}{\phi V(\mu_i)},$$

and then solved

$$\sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta} U(Y_i, \mu_i) = 0,$$

for β.

Why Bother?

- In this course we will see M-estimators a few more times!

Why Bother?

- In this course we will see M-estimators a few more times!
- If you go on in Statistics, they will come-up time and time again.

Why Bother?

- In this course we will see M-estimators a few more times!
- If you go on in Statistics, they will come-up time and time again.
- It is *quite rare* to see them covered, despite their prevalence.

Why Bother?

- In this course we will see M-estimators a few more times!
- If you go on in Statistics, they will come-up time and time again.
- It is *quite rare* to see them covered, despite their prevalence.
- They also provide an interesting tool to solve problems: **if** you can frame an estimation problem through estimating equations, the theory is easy to derive!

In Practice

In R we can implement these with any **root finding** package!

For specific problems, the type of interest to us, we will almost always use **specially designed** tools (like `glm` or `lm`)!

A Note on Theory

There are further restrictions on U, other than unbiasedness. These *regularity conditions* are typically ignored, but are critically important for the asymptotic results to hold!