Asymptotic Distribution of Maximum Likelihood Estimator

Overview

Our goal:

As $n \to \infty$, the distribution of the maximum likelihood estimator becomes approximately $\hat{\Theta}^{MLE} \sim \text{Normal}\left(\theta, \frac{1}{I(\theta)}\right)$

Strategy:

- Take a Taylor series approximation to the first derivative of the log-likelihood
- Use the fact that the derivative of the log-likelihood function is 0 at the maximum
- Let $n \to \infty$. As the sample size grows, we can apply the central limit theorem and the law of large numbers to get our desired result.

Preliminary Results (ingredients for the proof)

Let $X_1, \ldots, X_n$ be independent and identically distributed with mean μ and variance σ^2.

Law of Large Numbers, Convergence in Probability

- Informal statement of the law of large numbers:

As $n \to \infty$, the sample mean approaches μ.

- Slightly more formal statement of the law of large numbers:

As $n \to \infty$, the probability that the sample mean is close to μ goes to 1

- Formal statement of law of large numbers:

For any $\varepsilon > 0$, $\lim_{n \to \infty} P(|\bar{X} - \mu| < \varepsilon) = 1$

This is a statement that the sample mean **converges in probability** to μ.

Central Limit Theorem, Convergence in Distribution

- Informal statement of the central limit theorem:

As $n \to \infty$, the distribution of the sample mean is approximately $\text{Normal}(\mu, \sigma^2)$.

- Slightly more statement of central limit theorem:

As $n \to \infty$, the distribution of $\frac{\sqrt{n}}{\sigma}(\bar{X} - \mu)$ approaches the distribution of a random variable Z that follows a $\text{Normal}(0, 1)$ distribution.

- Formal statement of central limit theorem:

Let $F_{Z_n}(\bar{x})$ be the cumulative distribution function of $Z_n = \frac{\sqrt{n}}{\sigma}\left(\frac{1}{n}\sum_{i=1}^n X_i - \mu\right)$ based on a sample of size n, and $F_Z(z)$ be the cumulative distribution function of a $\text{Normal}(0, 1)$ random variable.

For any point z at which F_Z is continuous, $\lim_{n \to \infty} F_{Z_n}(z) = F_Z(z)$.

This is a statement that the (centered and scaled) sample mean **converges in distribution** to a random variable Z that follows a $\text{Normal}(0, 1)$ distribution.

Slutsky's Theorem (probably new to you; stated without proof)

Suppose that $X_1, X_2, \ldots$ is a sequence of random variables that converges in distribution to a random variable X, and $Y_1, Y_2, \ldots$ is a sequence of random variables that converges in probability to a constant c.

Then $\frac{X_n}{Y_n}$ converges in distribution to $\frac{X}{c}$.

Proof of main claim

- Take a first order Taylor series approximation of the first derivative of the log-likelihood around the parameter estimate $\hat{\theta}^{MLE}$.
  - We saw last class that the first derivative is 0 (the MLE maximizes the log-likelihood)
  - So we get:

$$\begin{aligned}\ell'(\theta) &\approx \ell'(\hat{\theta}^{MLE}) + \ell''(\hat{\theta}^{MLE}|X_1, \ldots, X_n)(\theta - \hat{\theta}^{MLE}) \\ &= \ell''(\hat{\theta}^{MLE}|X_1, \ldots, X_n)(\theta - \hat{\theta}^{MLE})\end{aligned}$$

- Evaluating at the true parameter value θ_0, we get:

$$\ell'(\theta_0) \approx \ell''(\hat{\theta}^{MLE}|X_1, \ldots, X_n)(\theta_0 - \hat{\theta}^{MLE})$$

- Rearrange to get $\sqrt{n}(\hat{\theta}^{MLE} - \theta_0)$ on the left hand side. Our goal is to show that this converges in distribution to a Normal$(0, ???)$ random variable.

$$\sqrt{n}(\hat{\theta}^{MLE} - \theta_0) \approx -\frac{\sqrt{n}\ell'(\theta_0)}{\ell''(\hat{\theta}^{MLE}|X_1, \ldots, X_n)}$$

- Plug in $\ell(\theta|X_1, \ldots, X_n) = \sum_{i=1}^n \log\{f_{X_i}(x_i|\theta)\}$

$$\sqrt{n}(\hat{\theta}^{MLE} - \theta_0) \approx -\frac{\sqrt{n}\sum_{i=1}^n \frac{d}{d\theta}\log\{f_{X_i}(x_i|\theta)\}|_{\theta=\theta_0}}{\sum_{i=1}^n \frac{d^2}{d\theta^2}\log\{f_{X_i}(x_i|\theta)\}|_{\theta=\hat{\theta}^{MLE}}}$$

- Divide the numerator and denominator by $\frac{1}{n}$. We're setting up to apply the law of large numbers and the central limit theorem.

$$\sqrt{n}(\hat{\theta}^{MLE} - \theta_0) \approx -\frac{\sqrt{n}\frac{1}{n}\sum_{i=1}^n \frac{d}{d\theta}\log\{f_{X_i}(x_i|\theta)\}|_{\theta=\theta_0}}{\frac{1}{n}\sum_{i=1}^n \frac{d^2}{d\theta^2}\log\{f_{X_i}(x_i|\theta)\}|_{\theta=\hat{\theta}^{MLE}}}$$

- Let $n \to \infty$. Three things happen:

1. $\hat{\theta}^{MLE} \to \theta_0$. This means the Taylor series approximation becomes better and better, and we don't have to worry about $\approx$.
2. In the denominator, the law of large numbers applies.

As $n \to \infty$, $\frac{1}{n}\sum_{i=1}^n \frac{d^2}{d\theta^2}\log\{f_{X_i}(x_i|\theta)\}|_{\theta=\hat{\theta}^{MLE}}$ converges in probability to $E\left[\frac{d^2}{d\theta^2}\log\{f_{X_i}(x_i|\theta)\}|_{\theta=\hat{\theta}^{MLE}}\right]$. This is the definition of the Fisher information, $I(\hat{\theta}^{MLE})$.

3. In the numerator, the Central Limit Theorem applies.

A lemma is required to show that $E\left[\frac{d}{d\theta}\log\{f_X(x|\theta)\}\right]|_{\theta=\theta_0} = \frac{d}{d\theta}E\left[\log\{f_X(x|\theta)\}\right]|_{\theta=\theta_0} = 0$.

Once you have that,

$$\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^n \frac{d}{d\theta}\log\{f_{X_i}(x_i|\theta)\}|_{\theta=\theta_0} - 0\right]$$

converges in distribution to a Normal$(0, ***)$ distribution by the CLT

Above, *** is $Var\left[\frac{d}{d\theta}\log\{f_X(x|\theta)\}\right]$