

# Asymptotic Distribution of Maximum Likelihood Estimator

## Overview

For today, denote the true value of the parameter by  $\theta_0$ .

### Goal: Theorem

As  $n \rightarrow \infty$ , the distribution of the maximum likelihood estimator becomes approximately

$$\hat{\Theta}^{MLE} \sim \text{Normal} \left( \theta_0, \frac{1}{nI_i(\theta_0)} \right),$$

where  $I_i(\theta_0) = E \left[ \left\{ \frac{d}{d\theta} \ell(\theta|X_i) \Big|_{\theta=\theta_0} \right\}^2 \right] = -E \left[ \frac{d^2}{d\theta^2} \ell(\theta|X_i) \Big|_{\theta=\theta_0} \right]$  is the Fisher information about the parameter  $\theta$  from a single observation  $X_i$ , evaluated at the true parameter value  $\theta_0$ .

More formally,  $\sqrt{n} \left( \hat{\Theta}^{MLE} - \theta_0 \right)$  converges in distribution to a  $\text{Normal}(0, \frac{1}{I_i(\theta_0)})$  random variable.

### Note:

- The theorem could be equally well stated in terms of:
  - the Fisher information or the observed Fisher information
  - the (observed) Fisher information for all observations instead of individual observations

$$\begin{aligned} \hat{\Theta}^{MLE} &\sim \text{Normal} \left( \theta_0, \frac{1}{nI_i(\theta_0)} \right) \\ \hat{\Theta}^{MLE} &\sim \text{Normal} \left( \theta_0, \frac{1}{I(\theta_0)} \right) \\ \hat{\Theta}^{MLE} &\sim \text{Normal} \left( \theta_0, \frac{1}{\sum_{i=1}^n J_i(\theta_0)} \right) \\ \hat{\Theta}^{MLE} &\sim \text{Normal} \left( \theta_0, \frac{1}{J(\theta_0)} \right) \end{aligned}$$

- We will just prove the main statement above, but I think you'll see the opportunities for these variations in the proof.

### Strategy for Proof:

- 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
- Do a little algebra. After rearranging, we will arrive at an expression involving the following two quantities:
  1.  $\frac{1}{n} \sum_{i=1}^n \frac{d^2}{d\theta^2} \log \{f_{X_i}(X_i|\theta)\} \Big|_{\theta=\hat{\Theta}^{MLE}}$
  2.  $\sqrt{n} \frac{1}{n} \sum_{i=1}^n \frac{d}{d\theta} \log \{f_{X_i}(X_i|\theta)\} \Big|_{\theta=\theta_0}$
- Let  $n \rightarrow \infty$ . As the sample size grows, we can apply the central limit theorem and the law of large numbers to the quantities above and get our desired result.

There are conditions for when the theorem holds; I'll list them at the end and briefly point out where they were needed in the proof then.

## Preliminary Results (ingredients for the proof)

Let  $X_1, \dots, X_n$  be independent and identically distributed with mean  $\mu$  and variance  $\sigma^2$ .

### Law of Large Numbers, Convergence in Probability

- Informal statement of the law of large numbers:

As  $n \rightarrow \infty$ , the sample mean approaches  $\mu$ .

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

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

- Formal statement of law of large numbers:

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

This is a statement that the sample mean **converges in probability** to  $\mu$ .

**Based on the law of large numbers, what does this quantity (almost number 1 in the list at the bottom of the previous page) converge to as  $n \rightarrow \infty$ ?**

$$\frac{1}{n} \sum_{i=1}^n \frac{d^2}{d\theta^2} \log \{f_{X_i}(X_i|\theta)\} |_{\theta=\theta^*}$$

## Central Limit Theorem, Convergence in Distribution

- Informal statement of the central limit theorem:

As  $n \rightarrow \infty$ , the distribution of the sample mean is approximately Normal  $\left(\mu, \frac{\sigma^2}{n}\right)$ .

- Slightly more formal (and rearranged) statement of central limit theorem:

As  $n \rightarrow \infty$ , the distribution of  $\sqrt{n}(\bar{X} - \mu)$  approaches the distribution of a random variable  $Z$  that follows a Normal( $0, \sigma^2$ ) distribution.

- Formal statement of central limit theorem:

Let  $F_{Z_n}(\bar{x})$  be the cumulative distribution function of  $Z_n = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n X_i - \mu\right)$  based on a sample of size  $n$ , and  $F_Z(z)$  be the cumulative distribution function of a Normal( $0, \sigma^2$ ) random variable.

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

This is a statement that the (centered and scaled by  $\sqrt{n}$ ) sample mean **converges in distribution** to a random variable  $Z$  that follows a Normal( $0, \sigma^2$ ) distribution.

### Warm up example

Suppose  $W_1, \dots, W_n \stackrel{\text{i.i.d.}}{\sim} \text{Poisson}(\lambda)$ . For each  $i$ ,  $E(W_i) = \lambda$  and  $\text{Var}(W_i) = \lambda$ .

For a large value of  $n$ , what is the approximate distribution of the sample mean,  $\bar{W} = \frac{1}{n} \sum_{i=1}^n W_i$ ?

For a large value of  $n$ , what is the approximate distribution of  $\sqrt{n}(\bar{W} - \lambda) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n W_i - \lambda\right)$ ?

**In the context of our proof:**

We want to apply the central limit theorem to say something about the distribution of  $\sqrt{n} \frac{1}{n} \sum_{i=1}^n \frac{d}{d\theta} \log \{f_{X_i}(X_i|\theta)\} |_{\theta=\theta_0}$  as  $n \rightarrow \infty$ .

Define  $W_i = \frac{d}{d\theta} \log \{f_{X_i}(X_i|\theta)\} |_{\theta=\theta_0}$ .

How can we write the above in terms of  $W_i$ ?

**Claim:**  $E[W_i] = 0$

- Intuitively, what does  $E[W_i] = E \left[ \frac{d}{d\theta} \log \{f_{X_i}(X_i|\theta)\} |_{\theta=\theta_0} \right]$  represent?

- Why would we expect that  $E[W_i] = 0$ ?

- Let's prove the claim together:

**Claim:**  $Var[W_i] = I(\theta_0)$

To prove this, we'll use the identity  $Var(W_i) = E(W_i^2) - [E(W_i)]^2$ .

Also, remember that  $I(\theta^*) = E \left[ \left( \frac{d}{d\theta} \log \{f_{X_i}(X_i|\theta)\} \Big|_{\theta=\theta^*} \right)^2 \right]$

Now we're ready to apply the central limit theorem to our problem. We have the following:

- We will want to apply the central limit theorem to  $\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n W_i \right) = \sqrt{n} \frac{1}{n} \sum_{i=1}^n \frac{d}{d\theta} \log \{f_{X_i}(X_i|\theta)\} \Big|_{\theta=\theta_0}$
- We have shown that  $E(W_i) = 0$  and  $Var(W_i) = I(\theta_0)$

As  $n \rightarrow \infty$ , what is the approximate distribution of  $\sqrt{n} \frac{1}{n} \sum_{i=1}^n \frac{d}{d\theta} \log \{f_{X_i}(X_i|\theta)\} \Big|_{\theta=\theta_0}$ ?

## Proof of Main Theorem

Remember that our goal is to show that as  $n \rightarrow \infty$  it is approximately true that:

- $\hat{\Theta}^{MLE} \sim \text{Normal}\left(\theta_0, \frac{1}{nI_i(\theta_0)}\right)$
- Or, rearranging,  $\left(\hat{\Theta}^{MLE} - \theta_0\right) \sim \text{Normal}\left(0, \frac{1}{nI_i(\theta_0)}\right)$
- Or  $\sqrt{n}\left(\hat{\Theta}^{MLE} - \theta_0\right) \sim \text{Normal}\left(0, \frac{1}{I_i(\theta_0)}\right)$

### Steps in proof:

Take a first order Taylor series approximation of the first derivative of the log-likelihood around the estimator  $\hat{\Theta}^{MLE}$ .

How can we simplify the above?

Let's evaluate this Taylor approximation at the true parameter value  $\theta_0$ :

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  $\text{Normal}(0, I_i(\theta_0))$  random variable.

Plug in the definition of the log-likelihood function to make clearer what's going on:

We started off by analyzing the following terms. How can we modify the result of the previous step to get these terms to show up?

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

**For the next step, we need Slutsky's theorem (probably new to you; stated without proof)**

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

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

So far, we have shown that:

- $\sqrt{n} \frac{1}{n} \sum_{i=1}^n \frac{d}{d\theta} \log \{f_{X_i}(X_i|\theta)\} |_{\theta=\theta_0}$  converges in distribution to a  $\text{Normal}(0, I_i(\theta_0))$  random variable (from the Central Limit Theorem).
- $\frac{1}{n} \sum_{i=1}^n \frac{d^2}{d\theta^2} \log \{f_{X_i}(X_i|\theta)\} |_{\theta=\hat{\theta}^{MLE}}$  converges in probability to  $I_i(\theta_0)$  (from the law of large numbers).

**Last Step!! Let  $n \rightarrow \infty$ .**

Let's think about what happens in two pieces:

1.  $\hat{\theta}^{MLE} \rightarrow \theta_0$  (really this requires proof; for today, let's accept this as OK). This means the Taylor series approximation becomes better and better, and we don't have to worry about the  $\approx$ .

2. The central limit theorem and law of large numbers apply (using Slutsky's theorem).

## Summing Up

That was a lot of steps. What just happened?

### Reminder of 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
- Do a little algebra. After rearranging, we will arrive at an expression involving the following two quantities:
  1.  $\frac{1}{n} \sum_{i=1}^n \frac{d^2}{d\theta^2} \log \{f_{X_i}(X_i|\theta)\} |_{\theta=\hat{\theta}^{MLE}}$
  2.  $\sqrt{n} \frac{1}{n} \sum_{i=1}^n \frac{d}{d\theta} \log \{f_{X_i}(X_i|\theta)\} |_{\theta=\theta_0}$
- Let  $n \rightarrow \infty$ . As the sample size grows, we can apply the central limit theorem and the law of large numbers to the quantities above and get our desired result.

**Why are we working with the log-likelihood function instead of the likelihood function?**

**What information did we use about the first derivative of the log-likelihood?**

**Why does Fisher information appear? Do we have intuition for why a variance inversely proportional to Fisher information makes sense?**

**What theorem basically drives the result? Where did this come from?**