LET $T_n$ BE ASYMPTOTICALLY NORMAL

$$\sqrt{n}\,(T_n - \theta) \xrightarrow{d} N(0, \sigma^2(\theta))$$

TO SHOW THAT $T_n$ IS CONSISTENT

BY SLUTSKY'S LEMMA

$$T_n - \theta = \underbrace{\frac{1}{\sqrt{n}}}_{\xrightarrow{P} 0} \cdot \underbrace{\sqrt{n}\,(T_n - \theta)}_{\xrightarrow{d} N(0, \sigma^2(\theta))} \xrightarrow{d} 0 \cdot N(0, \sigma^2(\theta)) = 0$$

$$T_n - \theta \xrightarrow{d} 0 \iff T_n - \theta \xrightarrow{P} 0$$

Imperial College London

# Lecture 07: Maximum Likelihood Estimation (Asymptotic Results)

## Statistical Modelling I

Dr. Riccardo Passeggeri

# Outline

# Introduction

# Large sample properties

Let $X_1, X_2, \ldots$ be iid observations with pdf (or pmf) $f_\theta(x)$, where $\theta \in \Theta$ and $\Theta$ is an open interval. Let $\theta_0 \in \Theta$ denote the true parameter. Under regularity conditions (e.g. $\{x : f_\theta(x) > 0\}$ does not depend on $\theta$), the following holds:

(i) There exists a **consistent** sequence $(\hat{\theta}_n)_{n \in \mathbb{N}}$ of maximum likelihood estimators. [$\hat{\theta}_n$ is an MLE based on $X_1, \ldots, X_n$].

(ii) Suppose $(\hat{\theta}_n)_{n \in \mathbb{N}}$ is a consistent sequence of MLEs. Then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, (I_f(\theta_0))^{-1}),$$

where $I_f(\theta) = E_\theta[(\frac{\partial}{\partial \theta} \log f_\theta(X))^2]$ is the **Fisher Information** of a sample of size 1.

# Consistency

# Sketch of proof (1/3)

There exists a **consistent** sequence $(\hat{\theta}_n)_{n\in\mathbb{N}}$ of maximum likelihood estimators.

$$L(\theta) = L(\theta; y) = \begin{cases} P(Y=y;\theta) = P_\theta(Y=y) & \text{DISCRETE } \cancel{\text{DATA}} \text{ CASE} \\ f_Y(y;\theta) = f_\theta(y) & \text{CONTINUOUS CASE} \end{cases}$$

$X_1, X_2, \dots$ ARE iid $\boxed{X_1 \sim P_{\theta_0}}$

$$L(\theta; X) = L(\theta) = \prod_{i=1}^{n} f_\theta(X_i)$$

$$S_n(\theta) = \frac{1}{n} \log L(\theta) = \frac{1}{n} \sum_{i=1}^{n} \log f_\theta(X_i)$$

BY MONOTONICITY $\hat{\theta}_{MLE}$ MAXIMIZES $L(\theta) \Leftrightarrow \hat{\theta}_{MLE}$ MAXIMIZES $S_n(\theta)$

$$S_n(\theta) \xrightarrow{P} \underset{\theta_0}{E}[\log f_\theta(X_1)] := R(\theta)$$

## Sketch of proof (2/3)

$\theta_0$ MAXIMIZES $\cancel{R(\theta)}$ $R : \theta \to R(\theta)$. FOR EVERY $\theta \in \Theta$

$$R(\theta) - R(\theta_0) = E_{\theta_0}\left[\log f_\theta(X_1) - \log f_{\theta_0}(X_1)\right] = E_{\theta_0}\left[\log\left(\frac{f_\theta(X_1)}{f_{\theta_0}(X_1)}\right)\right]$$

$$\le E_{\theta_0}\left[\frac{f_\theta(X_1)}{f_{\theta_0}(X_1)} - 1\right]$$ BECAUSE $\forall a > 0 \quad a - 1 \ge \log(a)$

$$= \int_{\mathbb{R}} \left(\frac{f_\theta(x)}{f_{\theta_0}(x)} - 1\right) f_{\theta_0}(x)\,dx = \int_{\mathbb{R}} f_\theta(x)\,dx - \int_{\mathbb{R}} f_{\theta_0}(x)\,dx = 1 - 1 = 0$$

$$R(\theta) - R(\theta_0) \le 0 \implies R(\theta_0) \ge R(\theta) \quad \forall \theta \in \Theta$$

# Sketch of proof (3/3)

- $S_M(\theta) \xrightarrow{P} R(\theta) \quad \forall \theta \in \Theta$
- $\theta_0$ MAXIMIZES $R(\cdot)$
- $\hat{\theta}_{MLE,M}$ MAXIMIZES $S_M(\cdot)$

WE WANT TO SHOW $\hat{\theta}_{MLE} \xrightarrow{P} \theta_0$ $\Leftrightarrow$ ARGMAX OF $S_M$ CONVERGES IN PROBABILITY TO ARGMAX OF $R$

$\rightarrow$ UNDER MILD REGULARITY CONDITIONS WE HAVE THAT $\hat{\theta}_{MLE} \xrightarrow{P} \theta_0$

# Asymptotic normality

## Sketch of proof (1/4)

Suppose $(\hat{\theta}_n)_{n\in\mathbb{N}}$ is a consistent sequence of MLEs. Then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, (I_f(\theta_0))^{-1}),$$

where $I_f(\theta) = E_\theta[(\frac{\partial}{\partial\theta}\log f_\theta(X))^2]$ is the **Fisher Information** of a sample of size 1.

$\hat{\theta}_{MLE,n}$ SATISFIES $\frac{\partial}{\partial} \log L(\theta)\Big|_{\theta=\hat{\theta}_{MLE,n}} = 0$

$$0 = \frac{1}{\sqrt{n}} \frac{\partial}{\partial\theta} \log L(\theta)\Big|_{\theta=\hat{\theta}_{MLE,n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f(X_i, \theta) =: g_n(\hat{\theta}_{MLE,n})$$

BY INTERMIDIATE VALUE THEOREM (OR TAYLOR)

$g_n(\hat{\theta}_{MLE,n}) = g_n(\theta_0) + g_n'(\tilde{\theta}_n)(\hat{\theta}_{MLE,n} - \theta_0)$ , $\tilde{\theta}_n$ LIES BETWEEN $\hat{\theta}_{MLE,n}$ AND $\theta_0$

$$g_n(\hat{\theta}_{MLE,n}) = g_n(\theta_0) + \frac{1}{\sqrt{n}} g_n'(\tilde{\theta}_n)\sqrt{n}(\hat{\theta}_{MLE} - \theta_0) \Leftrightarrow \sqrt{n}(\hat{\theta}_{MLE,n} - \theta_0) = \frac{g_n(\theta_0)}{-\frac{1}{\sqrt{n}} g_n'(\tilde{\theta})}$$

FIRST WE SHOW THAT $-\frac{1}{\sqrt{n}} g_n'(\tilde{\theta}_n) \xrightarrow{P} I_f(\theta_0)$, SECOND $g_n(\theta_0) \xrightarrow{d} N(0, I_f(\theta_0))$

# Sketch of proof (2/4)

$$-\frac{1}{\sqrt{n}} g_n'(\theta) = -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial}{\partial\theta}\right)^2 \log f(X_i,\theta) \xrightarrow[\text{BY WLLN}]{P} -E\left[\left(\frac{\partial}{\partial\theta}\right)^2 \log f(X_1,\theta)\right] = I_f(\theta)$$

SO $-\frac{1}{\sqrt{n}} g_n'(\theta) \xrightarrow{P} I_f(\theta)$. WE KNOW THAT $\hat{\theta}_{MLE,n} \xrightarrow{P} \theta_0 \Rightarrow \tilde{\theta}_n \xrightarrow{P} \theta_0$

THEN, UNDER MILD REGULARITY CONDITIONS WE HAVE $-\frac{1}{\sqrt{n}} g_n'(\tilde{\theta}_n) \xrightarrow{P} I_f(\theta_0)$.

$$g_n(\theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i, \quad \text{WHERE } V_i = \frac{\partial}{\partial\theta}\log f(X_i,\theta)\Big|_{\theta=\theta_0}$$

$$E_{\theta_0}[V_1] = E_{\theta_0}\left[\frac{\partial}{\partial\theta}\log f(X_1,\theta_0)\right] = E_{\theta_0}\left[\frac{f'(X_1,\theta_0)}{f(X_1,\theta_0)}\right] = \int_{\mathbb{R}} \frac{f'(x,\theta_0)}{f(x,\theta_0)} f(x,\theta_0)\,dx$$

$$= \int_{\mathbb{R}} f'(x,\theta)\,dx = \frac{\partial}{\partial\theta}\int_{\mathbb{R}} f(x,\theta)\,dx = \frac{\partial}{\partial\theta} 1 = 0$$

$$VAR_{\theta_0}(V_1) = E_{\theta_0}[V_1^2] = E_{\theta_0}\left[\left(\frac{\partial}{\partial\theta}\log f(X_1,\theta)\right)^2\right] = I_f(\theta_0)$$

## Sketch of proof (3/4)

$E_{\theta_0}[V_1] = 0 \quad , \quad VAR_{\theta_0}[V_1] = I_f(\theta_0)$

$g_n(\theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i \xrightarrow[\text{BY CLT}]{d} N(0, I_f(\theta_0))$

- $g_n(\theta_0) \xrightarrow{d} N(0, I_f(\theta_0))$
- $-\frac{1}{\sqrt{n}} g_n'(\tilde{\theta}_n) \xrightarrow{P} I_f(\theta_0)$

$$\sqrt{n}\,(\hat{\theta}_{MLE,n} - \theta_0) = \frac{g_n(\theta_0)}{-\frac{1}{\sqrt{n}} g'(\tilde{\theta})} \xrightarrow[\text{BY SLUTSKY'S LEMMA}]{d\ ''} N\left(0, \frac{I_f(\theta_0)}{(I_f(\theta_0))^2}\right) = N\left(0, \frac{1}{I_f(\theta_0)}\right)$$

## Sketch of proof (4/4)

## Summary

These and similar arguments are frequently used in asymptotic statistics.

Asymptotic normality will be used in the next lecture to derive large sample confidence intervals.