Econometrics I

(Tue., 8:50-10:20)

Room # 1 (法経講義棟)

- This class is based on **Statistics** (**統計**,『コア・テキスト　統計学』大屋 幸輔 著, 新世社) and **Econometrics** (**計量経済**,『計量経済学』山本 拓 著，新世社), which are provided by Department of Economics,

  and **Basic Statistics** (**統計基礎**,『コア・テキスト　統計学』大屋 幸輔 著，新世社), provided by Graduate School of Economics.

  Thus, Statistics and Econometrics of undergraduate level are prerequisites.

- Furthermore, **Special Lectures in Economics (Statistical Analysis)** or **Statistical Analysis** (**統計解析**), provided by Graduate School of Economics, should be studied with this class.

  Or, do self-study using the lecture notes of

`https://www2.econ.osaka-u.ac.jp/~tanizaki/class/2012/econome1/index.htm`

  (The notes are written in English with Japanese translation for econometrics terms).

TA Session:

TAs: **Kaito Sakaguchi (坂口 開人)**
u190954f [at] ecs. osaka-u. ac. jp

Jukina Hatakeyama (畠山 樹輝凪):
u868710a [at] ecs. osaka-u. ac. jp

Date: **Thrs. 15:10-16:40**

Place: **Room #509 (509セミナー室)**

Contents: **Basic Statistics, Matrix Algebra, and etc.**

Ask TAs directly in the TA session
if you have questions about class, homework and etc.

- **Download the lecture notes from the following websites:**

http://www2.econ.osaka-u.ac.jp/~tanizaki/class/2025/econome2/

Contents

1 Maximum Likelihood Estimation (MLE, 最尤法) **9**

2 Qualitative Dependent Variable (質的従属変数) **36**

2.1 Discrete Choice Model (離散選択モデル) 37

2.1.1 Binary Choice Model (二値選択モデル) 37

2.2 Limited Dependent Variable Model (制限従属変数モデル) 58

2.3 Count Data Model (計数データモデル) 67

2.4 Applications 75

3 Panel Data **101**

3.1 GLS — Review 101

3.2 Panel Model Basic 105

3.2.1 Fixed Effect Model (固定効果モデル) 106
3.2.2 Random Effect Model (ランダム効果モデル) 114
3.3 Hausman's Specification Error (特定化誤差) Test 118
3.4 Choice of Fixed Effect Model or Random Effect Model 120
3.4.1 The Case where $\boldsymbol{X}$ is Correlated with $\boldsymbol{u}$ — Review 120
3.4.2 Fixed Effect Model or Random Effect Model 122
3.5 Applications 124

4 Introduction to Causal Inference (因果推論)
— Micro-econometrics — 129

5 Generalized Method of Moments (GMM, 一般化積率法) 137
5.1 Method of Moments (MM, 積率法) 137
5.2 Generalized Method of Moments (GMM, 一般化積率法) 145

5.3 Generalized Method of Moments (GMM, 一般化積率法) II — Non-linear Case — 167

6 Time Series Analysis (時系列分析) 189

6.1 Introduction 189

6.2 Autoregressive Model (自己回帰モデル or AR モデル) 194

6.3 Unit Root (単位根) Test (Dickey-Fuller (DF) Test) 208

6.4 AR(p) model: Augmented Dickey-Fuller (ADF) Test 224

7 Bayesian Estimation 226

7.1 Elements of Bayesian Inference 226

7.1.1 Bayesian Point Estimate 228

7.1.2 Bayesian Interval for Parameter 230

7.1.3 Prior Probability Density Function 231

7.2 Sampling Methods 236
7.2.1 Gibbs Sampling 236
7.2.2 Metropolis-Hastings Algorithm 246

1 Maximum Likelihood Estimation (MLE, 最尤法)

$\longrightarrow$ **Review**

1. The distribution function of $\{X_i\}_{i=1}^n$ is $f(x;\theta)$, where $x = (x_1, x_2, \cdots, x_n)$.

   θ is a vector or matrix of unknown parameters, e.g., $\theta = (\mu, \Sigma)$, where $\mu = \mathrm{E}(X_i)$ and $\Sigma = \mathrm{V}(X_i)$.

   Note that X is a vector of random variables and x is a vector of their realizations (i.e., observed data).

   Likelihood function $L(\cdot)$ is defined as $L(\theta; x) = f(x;\theta)$.

   Note that $f(x;\theta) = \prod_{i=1}^n f(x_i;\theta)$ when X_1, X_2, $\cdots$, X_n are mutually independently and identically distributed.

The maximum likelihood estimate (MLE) of θ is the θ such that:

$$\max_{\theta} \; L(\theta; x). \qquad \Longleftrightarrow \qquad \max_{\theta} \; \log L(\theta; x).$$

Thus, MLE satisfies the following two conditions:

(a) $\dfrac{\partial \log L(\theta; x)}{\partial \theta} = 0. \qquad \Longrightarrow \qquad$ Solution of θ: $\tilde{\theta} = \tilde{\theta}(x)$

(b) $\dfrac{\partial^2 \log L(\theta; x)}{\partial \theta \partial \theta'}$ is a negative definite matrix.

2. $x = (x_1, x_2, \cdots, x_n)$ are used as the observations (i.e., observed data).

   $X = (X_1, X_2, \cdots, X_n)$ denote the random variables associated with the joint distribution $f(x; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$.

3. Replacing x by X, we otain the maximum likelihood **estimator** (MLE, which is the same word as the maximum likelihood **estimate**).

   That is, MLE of θ satisfies the following two conditions:

   (a) $\dfrac{\partial \log L(\theta; X)}{\partial \theta} = 0. \quad \Longrightarrow \quad$ Solution of θ: $\tilde{\theta} = \tilde{\theta}(X)$

   (b) $\dfrac{\partial^2 \log L(\theta; X)}{\partial \theta \partial \theta'}$ is a negative definite matrix.

4. **Fisher's information matrix (フィッシャーの情報行列)** or simply **information matrix**, denoted by $I(\theta)$, is given by:

$$I(\theta) = -\mathrm{E}\Big(\frac{\partial^2 \log L(\theta; X)}{\partial \theta \partial \theta'}\Big),$$

   where we have the following equality:

$$-\mathrm{E}\Big(\frac{\partial^2 \log L(\theta; X)}{\partial \theta \partial \theta'}\Big) = \mathrm{E}\Big(\frac{\partial \log L(\theta; X)}{\partial \theta}\frac{\partial \log L(\theta; X)}{\partial \theta'}\Big) = \mathrm{V}\Big(\frac{\partial \log L(\theta; X)}{\partial \theta}\Big)$$

   Note that $\mathrm{E}(\cdot)$ and $\mathrm{V}(\cdot)$ are expected with respect to X.

Proof of the above equality:

$$\int L(\theta; x)\mathrm{d}x = 1$$

Take a derivative with respect to θ.

$$\int \frac{\partial L(\theta; x)}{\partial \theta}\mathrm{d}x = 0$$

(We assume that (i) the domain of x does not depend on θ and (ii) the derivative $\dfrac{\partial L(\theta; x)}{\partial \theta}$ exists.)

(*) **Differentiation of Composite Functions (合成関数の微分)** or **Chain rule (連鎖律)**:

$$\frac{\partial \log L(\theta; x)}{\partial \theta} = \frac{\partial \log L(\theta; x)}{\partial L(\theta; x)} \frac{\partial L(\theta; x)}{\partial \theta} = \frac{1}{L(\theta; x)} \frac{\partial L(\theta; x)}{\partial \theta}$$

i.e.,

$$\frac{\partial L(\theta; x)}{\partial \theta} = \frac{\partial \log L(\theta; x)}{\partial \theta} L(\theta; x)$$

Rewriting the above equation, we obtain:

$$\int \frac{\partial L(\theta; x)}{\partial \theta} \mathrm{d}x = \int \frac{\partial \log L(\theta; x)}{\partial \theta} L(\theta; x) \mathrm{d}x = 0,$$

i.e.,

$$\mathrm{E}\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right) = 0.$$

Again, differentiating the above with respect to θ, we obtain:

$$\begin{aligned}
&\int \frac{\partial^2 \log L(\theta; x)}{\partial \theta \partial \theta'} L(\theta; x) \mathrm{d}x + \int \frac{\partial \log L(\theta; x)}{\partial \theta} \frac{\partial L(\theta; x)}{\partial' \theta} \mathrm{d}x \\
&\quad = \int \frac{\partial^2 \log L(\theta; x)}{\partial \theta \partial \theta'} L(\theta; x) \mathrm{d}x + \int \frac{\partial \log L(\theta; x)}{\partial \theta} \frac{\partial \log L(\theta; x)}{\partial \theta'} L(\theta; x) \mathrm{d}x \\
&\quad = \mathrm{E}\Big(\frac{\partial^2 \log L(\theta; X)}{\partial \theta \partial \theta'}\Big) + \mathrm{E}\Big(\frac{\partial \log L(\theta; X)}{\partial \theta} \frac{\partial \log L(\theta; X)}{\partial \theta'}\Big) = 0.
\end{aligned}$$

Therefore, we can derive the following equality:

$$-\mathrm{E}\left(\frac{\partial^2 \log L(\theta; X)}{\partial \theta \partial \theta'}\right) = \mathrm{E}\left(\frac{\partial \log L(\theta; X)}{\partial \theta} \frac{\partial \log L(\theta; X)}{\partial \theta'}\right) = \mathrm{V}\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right),$$

where the second equality utilizes $\mathrm{E}\left(\dfrac{\partial \log L(\theta; X)}{\partial \theta}\right) = 0.$

5. **Cramer-Rao inequality (クラメール・ラオの不等式)** is given by:

$$\mathrm{V}(s(X)) \geq (I(\theta))^{-1},$$

where $s(X)$ denotes an unbiased estimator of θ.

$(I(\theta))^{-1}$ is called **Cramer-Rao Lower Bound (クラメール・ラオの下限)**.

Proof:

The expectation of $s(X)$ is:

$$\mathrm{E}(s(X)) = \int s(x) L(\theta; x) \mathrm{d}x.$$

Differentiating the above with respect to θ,

$$\frac{\partial \mathrm{E}(s(X))}{\partial \theta} = \int s(x) \frac{\partial L(\theta; x)}{\partial \theta} \mathrm{d}x = \int s(x) \frac{\partial \log L(\theta; x)}{\partial \theta} L(\theta; x) \mathrm{d}x$$

$$= \mathrm{Cov}\left(s(X), \frac{\partial \log L(\theta; X)}{\partial \theta}\right)$$

For simplicity, let $s(X)$ and θ be scalars.

Then,

$$\left(\frac{\partial \mathrm{E}(s(X))}{\partial \theta}\right)^2 = \left(\mathrm{Cov}\left(s(X), \frac{\partial \log L(\theta; X)}{\partial \theta}\right)\right)^2 = \rho^2 \mathrm{V}\left(s(X)\right) \mathrm{V}\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right)$$
$$\leq \mathrm{V}\left(s(X)\right) \mathrm{V}\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right),$$

where ρ denotes the correlation coefficient between $s(X)$ and $\dfrac{\partial \log L(\theta; X)}{\partial \theta}$, i.e.,

$$\rho = \frac{\mathrm{Cov}\left(s(X), \dfrac{\partial \log L(\theta; X)}{\partial \theta}\right)}{\sqrt{\mathrm{V}\left(s(X)\right)} \sqrt{\mathrm{V}\left(\dfrac{\partial \log L(\theta; X)}{\partial \theta}\right)}}.$$

Note that $|\rho| \leq 1$.

Therefore, we have the following inequality:

$$\left(\frac{\partial \mathrm{E}(s(X))}{\partial \theta}\right)^2 \leq \mathrm{V}(s(X))\ \mathrm{V}\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right),$$

i.e.,

$$\mathrm{V}(s(X)) \geq \frac{\left(\dfrac{\partial \mathrm{E}(s(X))}{\partial \theta}\right)^2}{\mathrm{V}\left(\dfrac{\partial \log L(\theta; X)}{\partial \theta}\right)}$$

Especially, when $\mathrm{E}(s(X)) = \theta$, i.e., when $s(X)$ is an unbiased estimator of θ, the numerator of the right-hand side leads to one.

Therefore, we obtain:

$$\mathrm{V}(s(X)) \geq \frac{1}{-\mathrm{E}\left(\dfrac{\partial^2 \log L(\theta; X)}{\partial \theta^2}\right)} = (I(\theta))^{-1}.$$

Even in the case where $s(X)$ is a vector, the following inequality holds.

$$\mathrm{V}(s(X)) \geq (I(\theta))^{-1},$$

where $I(\theta)$ is defined as:

$$\begin{aligned} I(\theta) &= -\mathrm{E}\left(\frac{\partial^2 \log L(\theta; X)}{\partial\theta\partial\theta'}\right) \\ &= \mathrm{E}\left(\frac{\partial \log L(\theta; X)}{\partial\theta}\frac{\partial \log L(\theta; X)}{\partial\theta'}\right) = \mathrm{V}\left(\frac{\partial \log L(\theta; X)}{\partial\theta}\right). \end{aligned}$$

The variance of any unbiased estimator of θ is larger than or equal to $(I(\theta))^{-1}$.

Thus, $(I(\theta))^{-1}$ results in the lower bound of the variance of any unbiased estimator of θ.

6. Asymptotic Normality of MLE:

Let $\tilde{\theta}$ be MLE of θ.

As n goes to infinity, we have the following result:

$$\sqrt{n}(\tilde{\theta} - \theta) \longrightarrow N\left(0, \lim_{n\to\infty}\left(\frac{I(\theta)}{n}\right)^{-1}\right),$$

where it is assumed that $\lim_{n\to\infty}\left(\dfrac{I(\theta)}{n}\right)$ converges.

$\longrightarrow$ The proof will be shown later.

That is, when n is large, $\tilde{\theta}$ is approximately distributed as follows:

$$\tilde{\theta} \sim N\left(\theta, (I(\theta))^{-1}\right).$$

Suppose that $s(X) = \tilde{\theta}$.

When n is large, $\mathrm{V}(s(X))$ is approximately equal to $(I(\theta))^{-1}$.

7. **Optimization (最適化):**

MLE of θ results in the following maximization problem:

$$\max_{\theta} \quad \log L(\theta; x).$$

We often have the case where the solution of θ is not derived in closed form.

$\Longrightarrow$ Optimization procedure

$$0 = \frac{\partial \log L(\theta; x)}{\partial \theta} \approx \frac{\partial \log L(\theta^*; x)}{\partial \theta} + \frac{\partial^2 \log L(\theta^*; x)}{\partial \theta \partial \theta'}(\theta - \theta^*).$$

Solving the above equation with respect to θ, we approximately obtain the following:

$$\theta = \theta^* - \left(\frac{\partial^2 \log L(\theta^*; x)}{\partial \theta \partial \theta'}\right)^{-1} \frac{\partial \log L(\theta^*; x)}{\partial \theta}.$$

Replace the variables as follows:

$$\theta \longrightarrow \theta^{(i+1)} \qquad\qquad \theta^* \longrightarrow \theta^{(i)}$$

Then, we have:

$$\theta^{(i+1)} = \theta^{(i)} - \left(\frac{\partial^2 \log L(\theta^{(i)}; x)}{\partial\theta\partial\theta'} \right)^{-1} \frac{\partial \log L(\theta^{(i)}; x)}{\partial\theta}.$$

$\Longrightarrow$ **Newton-Raphson method (ニュートン・ラプソン法)**

Replacing $\dfrac{\partial^2 \log L(\theta^{(i)}; x)}{\partial\theta\partial\theta'}$ by $\mathrm{E}\left(\dfrac{\partial^2 \log L(\theta^{(i)}; x)}{\partial\theta\partial\theta'} \right)$, we obtain the following optimization algorithm:

$$\begin{aligned} \theta^{(i+1)} &= \theta^{(i)} - \left(\mathrm{E}\left(\frac{\partial^2 \log L(\theta^{(i)}; x)}{\partial\theta\partial\theta'} \right) \right)^{-1} \frac{\partial \log L(\theta^{(i)}; x)}{\partial\theta} \\ &= \theta^{(i)} + \left(I(\theta^{(i)}) \right)^{-1} \frac{\partial \log L(\theta^{(i)}; x)}{\partial\theta} \end{aligned}$$

$\Longrightarrow$ **Method of Scoring (スコア法)**

Convergence speed might be improved, compared with Newton-Raphson method.

1. **Central Limit Theorem:** Let X_1, X_2, $\cdots$, X_n be mutually independently distributed random variables with mean $\mathrm{E}(X_i) = \mu$ and variance $\mathrm{V}(X_i) = \sigma^2 < \infty$ for $i = 1, 2, \cdots, n$.

   Define $\overline{X} = (1/n)\sum_{i=1}^{n} X_i$.

   Then, the central limit theorem is given by:

$$\frac{\overline{X} - \mathrm{E}(\overline{X})}{\sqrt{\mathrm{V}(\overline{X})}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \longrightarrow N(0, 1).$$

   Note that $\mathrm{E}(\overline{X}) = \mu$ and $\mathrm{V}(\overline{X}) = \sigma^2/n$.

   That is,

$$\sqrt{n}(\overline{X} - \mu) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu) \longrightarrow N(0, \sigma^2).$$

   Note that $\mathrm{E}(\overline{X}) = \mu$ and $n\mathrm{V}(\overline{X}) = \sigma^2$.

In the case where X_i is a vector of random variable with mean μ and variance $\Sigma < \infty$, the central limit theorem is given by:

$$\sqrt{n}(\overline{X} - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \longrightarrow N(0, \Sigma).$$

Note that $\mathrm{E}(\overline{X}) = \mu$ and $n\mathrm{V}(\overline{X}) = \Sigma$.

2. **Central Limit Theorem II:** Let $X_1, X_2, \cdots, X_n$ be mutually independently distributed random variables with mean $\mathrm{E}(X_i) = \mu$ and variance $\mathrm{V}(X_i) = \sigma_i^2$ for $i = 1, 2, \cdots, n$.

   Assume:

$$\sigma^2 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 < \infty.$$

   Define $\overline{X} = (1/n) \sum_{i=1}^{n} X_i$.

Then, the central limit theorem is given by:

$$\frac{\overline{X} - \mathrm{E}(\overline{X})}{\sqrt{\mathrm{V}(\overline{X})}} = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \longrightarrow N(0, 1),$$

i.e.,

$$\sqrt{n}(\overline{X} - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \longrightarrow N(0, \sigma^2).$$

Note that $\mathrm{E}(\overline{X}) = \mu$ and $n\mathrm{V}(\overline{X}) \longrightarrow \sigma^2$.

In the case where X_i is a vector of random variable with mean μ and variance Σ_i, the central limit theorem is given by:

$$\sqrt{n}(\overline{X} - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \longrightarrow N(0, \Sigma),$$

where $\Sigma = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \Sigma_i < \infty$.

Note that $\mathrm{E}(\overline{X}) = \mu$ and $n\mathrm{V}(\overline{X}) \longrightarrow \Sigma$.

[Review of Asymptotic Theories]

- **Convergence in Probability (確率収束)** $X_n \longrightarrow a$, i.e., X converges in probability to a, where a is a fixed number.

- **Convergence in Distribution (分布収束)** $X_n \longrightarrow X$, i.e., X converges in distribution to X. The distribution of X_n converges to the distribution of X as n goes to infinity.

Some Formulas

X_n and Y_n : Convergence in Probability

Z_n : Convergence in Distribution

- If $X_n \longrightarrow a$, then $f(X_n) \longrightarrow f(a)$.

- If $X_n \longrightarrow a$ and $Y_n \longrightarrow b$, then $f(X_n Y_n) \longrightarrow f(ab)$.

- If $X_n \longrightarrow a$ and $Z_n \longrightarrow Z$, then $X_n Z_n \longrightarrow aZ$, i.e., aZ is distributed with mean $\mathrm{E}(aZ) = a\mathrm{E}(Z)$ and variance $\mathrm{V}(aZ) = a^2\mathrm{V}(Z)$.

[End of Review]

3. **Weak Law of Large Numbers (大数の弱法則) — Review:**

Suppose that $X_1, X_2, \cdots, X_n$ are distributed.

As $n \longrightarrow \infty$, $\overline{X} \longrightarrow \lim_{n\to\infty} \mathrm{E}(\overline{X})$ under $\lim_{n\to\infty} n\mathrm{V}(\overline{X}) < \infty$, which is called the **weak law of large numbers**.

$\longrightarrow$ Convergence in probability

$\longrightarrow$ Proved by Chebyshev's inequality

(i) Suppose that $X_1, X_2, \cdots, X_n$ are assumed to be mutually independently and identically distributed with $\mathrm{E}(X_i) = \mu$ and $\mathrm{V}(X_i) = \sigma^2 < \infty$.

Consider $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$.

Then, $\overline{X} \longrightarrow \mu$ as $n \longrightarrow \infty$.

Note that $\mathrm{E}(\overline{X}) = \mu$ and $n\mathrm{V}(\overline{X}) = \sigma^2$.

(ii) Suppoose that $X_1, X_2, \cdots, X_n$ are assumed to be mutually independently distributed with $\mathrm{E}(X_i) = \mu_i$ and $\mathrm{V}(X_i) = \sigma_i^2$.

Assume that

(a) $\mathrm{E}(\overline{X}) = \dfrac{1}{n}\sum_{i=1}^{n} \mu_i \longrightarrow \mu$, i.e., $\lim_{n\to\infty} \mathrm{E}(\overline{X}) = \mu$, and

(b) $n\mathrm{V}(\overline{X}) = \dfrac{1}{n}\sum_{i=1}^{n} \sigma_i^2 \longrightarrow \sigma^2 < \infty$, ie., $\lim_{n\to\infty} n\mathrm{V}(\overline{X}) = \sigma^2 < \infty$.

Then, $\overline{X} \longrightarrow \mu$ as $n \longrightarrow \infty$,

Note that $\mathrm{E}(\overline{X}) = \dfrac{1}{n}\sum_{i=1}^{n}\mu_i$ and $n\mathrm{V}(\overline{X}) = \dfrac{1}{n}\sum_{i=1}^{n}\sigma_i^2$.

(iii) Suppose that X_1, X_2, $\cdots$, X_n are assumed to be serially correlated with $\mathrm{E}(X_i) = \mu_i$ and $\mathrm{Cov}(X_i, X_j) = \sigma_{ij}$.

Assume that

(a) $\mathrm{E}(\overline{X}) = \dfrac{1}{n}\sum_{i=1}^{n}\mu_i \longrightarrow \mu$, i.e., $\lim_{n\to\infty}\mathrm{E}(\overline{X}) = \mu$, and

(b) $n\mathrm{V}(\overline{X}) = \dfrac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij} \longrightarrow \sigma^2 < \infty$, ie., $\lim_{n\to\infty} n\mathrm{V}(\overline{X}) = \sigma^2 < \infty$.

Then, $\overline{X} \longrightarrow \mu$ as $n \longrightarrow \infty$,

Note that $\mathrm{E}(\overline{X}) = \dfrac{1}{n}\sum_{i=1}^{n}\mu_i$ and $n\mathrm{V}(\overline{X}) = \dfrac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}$.

4. **Some Formulas of Expectaion and Variance in Multivariate Cases**

— Review:

A vector of randam variavle X: $\mathrm{E}(X) = \mu$ and $\mathrm{V}(X) \equiv \mathrm{E}((X-\mu)(X-\mu)') = \Sigma$

Then, $\mathrm{E}(AX) = A\mu$ and $\mathrm{V}(AX) = A\Sigma A'$.

Proof:

$\mathrm{E}(AX) = A\mathrm{E}(X) = A\mu$

$$
\begin{aligned}
\mathrm{V}(AX) &= \mathrm{E}((AX - A\mu)(AX - A\mu)') = \mathrm{E}(A(X-\mu)(A(X-\mu))') \\
&= \mathrm{E}(A(X-\mu)(X-\mu)'A') = A\mathrm{E}((X-\mu)(X-\mu)')A' = A\mathrm{V}(X)A' = A\Sigma A'
\end{aligned}
$$

MLE: Asymptotic Properties

1. $X_1, X_2, \cdots, X_n$ are random variables with density function $f(x;\theta)$.

   Let $\hat{\theta}_n$ be a maximum likelihood estimator of θ.

   Then, under some **regularity conditions**. $\hat{\theta}_n$ is a consistent estimator of θ and the asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta)$ is given by:

$$\sqrt{n}(\hat{\theta} - \theta) \longrightarrow N\left(0, \lim_{n\to\infty}\left(\frac{I(\theta)}{n}\right)^{-1}\right)$$

2. **Regularity Conditions:**

   (a) The domain of X_i does not depend on θ.

   (b) There exists at least third-order derivative of $f(x;\theta)$ with respect to θ, and their derivatives are finite.

3. Thus, MLE is

(i) consistent,

(ii) asymptotically normal, and

(iii) asymptotically efficient.

Proof: The log-likelihood function is given by:

$$\log L(\theta) = \log \prod_{i=1}^{n} f(X_i; \theta) = \sum_{i=1}^{n} \log f(X_i; \theta)$$

X_i is a random variable.

Consider the distribution of

$$\frac{1}{n} \frac{\partial \log L(\theta)}{\partial \theta} = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \log f(X_i; \theta)}{\partial \theta}.$$

We have to obtain mean and variance of $\dfrac{\partial \log f(X_i; \theta)}{\partial \theta}$.

Suppose that X_i is a continuous type of random variable.

$f(x_i;\theta)$ denotes the density function.

Therefore, we have:

$$\int f(x_i;\theta)\mathrm{d}x_i = 1$$

Taking the derivative with respect to θ on both sides, we obtain:

$$0 = \int \frac{\partial f(x_i;\theta)}{\partial\theta}\mathrm{d}x_i = \int \frac{\partial \log f(x_i;\theta)}{\partial\theta} f(x_i;\theta)\mathrm{d}x_i = \mathrm{E}\Big(\frac{\partial \log f(X_i;\theta)}{\partial\theta}\Big)$$

Again, take the derivative with respect to θ on both sides as follows:

$$\begin{aligned}
0 &= \int \frac{\partial^2 \log f(x_i;\theta)}{\partial\theta\partial\theta'} f(x_i;\theta) + \frac{\partial \log f(x_i;\theta)}{\partial\theta}\frac{\partial f(x_i;\theta)}{\partial\theta'}\mathrm{d}x_i \\
&= \int \frac{\partial^2 \log f(x_i;\theta)}{\partial\theta\partial\theta'} f(x_i;\theta)\mathrm{d}x_i + \int \frac{\partial \log f(x_i;\theta)}{\partial\theta}\frac{\partial \log f(x_i;\theta)}{\partial\theta'} f(x_i;\theta)\mathrm{d}x_i \\
&= \mathrm{E}\Big(\frac{\partial^2 \log f(X_i;\theta)'}{\partial\theta\partial\theta}\Big) + \mathrm{E}\Big(\frac{\partial \log f(X_i;\theta)}{\partial\theta}\frac{\partial \log f(X_i;\theta)}{\partial\theta'}\Big),
\end{aligned}$$

i.e.,

$$-\mathrm{E}\Big(\frac{\partial^2 \log f(X_i;\theta)'}{\partial\theta\partial\theta}\Big) = \mathrm{E}\Big(\frac{\partial \log f(X_i;\theta)}{\partial\theta}\frac{\partial \log f(X_i;\theta)}{\partial\theta'}\Big) = \mathrm{V}\Big(\frac{\partial \log f(X_i;\theta)}{\partial\theta}\Big) = \Sigma_i$$

Thus, $\dfrac{\partial \log f(X_i;\theta)}{\partial\theta}$ is distributed with mean 0 and variance Σ_i.

Note as follows:

$$I(\theta) = -\mathrm{E}\Big(\frac{\partial^2 \log L(\theta)}{\partial\theta\partial\theta'}\Big) = -\sum_{i=1}^{n}\mathrm{E}\Big(\frac{\partial^2 \log f(X_i;\theta)}{\partial\theta\partial\theta'}\Big) = \sum_{i=1}^{n}\Sigma_i.$$

Using the central limit theorem (generalization) shown above, asymptotically we obtain the following distribution:

$$\frac{1}{\sqrt{n}}\frac{\partial \log L(\theta)}{\partial\theta} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \log f(X_i;\theta)}{\partial\theta} \longrightarrow N(0,\Sigma),$$

where $\Sigma = \lim_{n\to\infty}\Big(\dfrac{1}{n}I(\theta)\Big)$.