

ECON-GA 2001

First Half

Quang Vuong

New York University

Revised October 2025

- I. Probability Spaces and Random Variables
- II. Measurable Functions and Lebesgue Integration
- III. Expectations and Conditional Expectations
- IV. Densities and Distributions
- V. Modes of Convergence
- VI. M-Estimation
- VII. NLLS, ML and GMM Estimation

I. Probability Spaces and Random Variables

(Ω, \mathcal{F}, P) Probability space

→ Ω : underlying space (nonempty)

→ \mathcal{F} : σ -algebra or σ -field

→ P : Probability measure

I.1. σ -algebra. (Bierens, Sections 1.1, 1.3–1.4)

DEFINITION: An *algebra/field* is a nonempty collection of subsets of Ω satisfying

- (i) If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$ (closed under complementation)
- (ii) If $A, B \in \mathcal{F}$ then $A \cup B \in \mathcal{F}$ (closed under finite unions)

REMARKS:

- In the definition, (ii) can be replaced by

(ii') If $A, B \in \mathcal{F}$ then $A \cap B \in \mathcal{F}$ (closed under finite intersections)

- Examples: $\mathcal{F}_1 = \{\emptyset, \Omega\}$ (trivial σ -algebra), $\mathcal{F}_2 = \{\emptyset, A, A^c, \Omega\}$, $\mathcal{F}_3 = 2^\Omega$ (power set)

DEFINITION: A σ -algebra/field is a nonempty collection of subsets of Ω satisfying

- (i) If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$ (closed under complementation)
- (ii) If $A_1, A_2, \dots \in \mathcal{F}$ then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$ (closed under countable unions)

The pair (Ω, \mathcal{F}) is called a *measurable space*.

REMARKS:

- In the definition, (ii) can be replaced by

- (ii') If $A_1, A_2, \dots \in \mathcal{F}$ then $\cap_{i=1}^{\infty} A_i \in \mathcal{F}$ (closed under countable intersections)

- Examples: $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ are σ -algebras

PROPERTIES OF σ -ALGEBRA:

- (a) \mathcal{F} is a σ -algebra $\Rightarrow \mathcal{F}$ is an algebra. The reverse is not necessarily true.
- (b) \mathcal{F} is an algebra with a finite number of elements $\Rightarrow \mathcal{F}$ is a σ -algebra.

- Let $\mathcal{F}_\theta, \theta \in \Theta$ be a (possibly uncountable) collection of σ -algebras of Ω indexed by θ . Then
 - $\rightarrow \cap_{\theta \in \Theta} \mathcal{F}_\theta$ is a σ -algebra.
 - $\rightarrow \cup_{\theta \in \Theta} \mathcal{F}_\theta$ is not necessarily a σ -algebra.

DEFINITION: The smallest σ -algebra containing a collection \mathcal{C} of subsets of Ω is called the *σ -algebra generated by \mathcal{C}* . It is denoted $\sigma(\mathcal{C})$ and satisfies $\sigma(\mathcal{C}) = \cap_{\{\mathcal{F} : \mathcal{C} \subseteq \mathcal{F}\}} \mathcal{F}$.

EXAMPLES:

- The σ -algebra generated by unions $\vee_{\theta \in \Theta} \mathcal{F}_\theta \equiv \sigma(\cup_{\theta \in \Theta} \mathcal{F}_\theta)$
- Let $\Omega = \mathbb{R}^k$ and $\mathcal{C} = \{\times_{i=1}^k (a_i, b_i); a_i < b_i \in \mathbb{R}\}$. Then
 - $\mathcal{B}^k \equiv \sigma(\mathcal{C})$ is the (Euclidean) *Borel σ -algebra* and any element of \mathcal{B}^k is called a *Borel set*.

REMARKS:

- $\rightarrow \mathcal{B}^k = \sigma(\{\times_{i=1}^k ([a_i, b_i]; a_i \leq b_i \in \mathbb{R})\}) = \sigma(\{\times_{i=1}^k (-\infty, a_i]; a_i \in \mathbb{R}\})$, etc.
- $\rightarrow \mathcal{B}^k = \sigma(\{\text{Open sets of } \mathbb{R}^k\}) = \sigma(\{\text{Closed sets of } \mathbb{R}^k\})$
- $\rightarrow \mathcal{B}^k \neq 2^{\mathbb{R}^k}$ (Vitali example)

I.2. Probability Measures. (Bierens, Sections 1.1, 1.5–1.7)

DEFINITION: A *measure* μ on (Ω, \mathcal{F}) is a mapping $\mathcal{F} \rightarrow \mathbb{R} \cup \{+\infty\}$ satisfying

- (i) $\mu(A) \geq 0$ for all $A \in \mathcal{F}$ (nonnegativity)
- (ii) $\mu(\emptyset) = 0$
- (iii) If $A_1, A_2, \dots \in \mathcal{F}$ are disjoint, then $\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$ (countable additivity).

REMARKS: $(\Omega, \mathcal{F}, \mu)$ is called a *measured space*.

$\rightarrow \mu$ is *finite* if $\mu(\Omega) < \infty$

$\rightarrow \mu$ is σ -*finite* if $\Omega = \cup_{i=1}^{\infty} A_i$ with $\mu(A_i) < \infty$

EXAMPLES:

- Let $\mathcal{F} = 2^{\Omega}$ and $\mu_c(A) \equiv \#A$. Then μ_c is called the *counting measure*. It is finite if $|\Omega| < \infty$.
- Let $(\Omega, \mathcal{F}) = (\mathbb{R}^k, \mathcal{B}^k)$. The *Lebesgue measure* λ on $(\mathbb{R}^k, \mathcal{B}^k)$ is

$$\lambda(B) \equiv \inf_{B \subseteq \cup_{j=1}^{\infty} \{\times_{i=1}^k (a_{ij}, b_{ij})\}} \sum_{j=1}^{\infty} \prod_{i=1}^k (b_{ij} - a_{ij})$$

The Lebesgue measure λ is σ -finite.

DEFINITION: A *probability measure* P on a *measurable space* (Ω, \mathcal{F}) is a measure with (ii) replaced by (ii) $P(\Omega) = 1$. The triplet (Ω, \mathcal{F}, P) is called a *probability space*.

REMARK: A probability measure P on an algebra \mathcal{A} is defined similarly by restricting $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$ in (iii). It can be uniquely extended to $\sigma(\mathcal{A})$. (Carathéodory Theorem)

EXAMPLE: Let $(\Omega, \mathcal{F}) = (\{1, 2, 3, 4, 5, 6\}, 2^{\Omega})$. The *uniform probability measure* is $P_0(A) = \frac{\#A}{6}$.

PROPERTIES OF MEASURES AND PROBABILITY MEASURES:

- (a) $P(\emptyset) = 0$
- (b) $P(A^c) = 1 - P(A)$ (only for probability measures)
- (c) $A \subseteq B$ implies $P(A) \leq P(B)$
- (d) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- (e) If $A_1 \subseteq A_2 \subseteq \dots$, then $P(A_n) \uparrow P(\cup_{i=1}^{\infty} A_n)$
- (f) If $A_1 \supseteq A_2 \supseteq \dots$, then $P(A_n) \downarrow P(\cap_{i=1}^{\infty} A_n)$ (only for finite measures)
- (g) $P(\cup_{i=1}^{\infty} A_n) \leq \sum_{i=1}^{\infty} P(A_n)$

I.3. Random Variables/Vectors. (Bierens, Sections 1.8, 1.10)

DEFINITION: Let (Ω, \mathcal{F}, P) be a probability space and $(\tilde{\Omega}, \tilde{\mathcal{F}})$ be a measurable space. A mapping $X : \Omega \rightarrow \tilde{\Omega}$ is a random element iff $X(\cdot)$ is $\mathcal{F}/\tilde{\mathcal{F}}$ -measurable, i.e. iff

$$X^{-1}(\tilde{B}) \in \mathcal{F} \quad \text{for every} \quad \tilde{B} \in \tilde{\mathcal{F}}$$

where $X^{-1}(\tilde{B}) \equiv \{\omega \in \Omega : X(\omega) \in \tilde{B}\}$.

EXAMPLE: When $\tilde{\Omega} = \{f(\cdot) : \mathcal{T} \rightarrow \mathbb{R}\}$ and \mathcal{F} is the σ -algebra generated by the open sets relative to a metric on $\tilde{\Omega}$, then $X(\cdot)$ is a *random process* indexed by \mathcal{T} and denoted $X(\omega; t)$.

REMARKS:

- The collection $\mathcal{F}_X \equiv \{X^{-1}(\tilde{B}) : \tilde{B} \in \tilde{\mathcal{F}}\}$ is a σ -algebra called the *σ -algebra generated by X* .

$X(\cdot)$ is a random element iff $\mathcal{F}_X \subseteq \mathcal{F}$.

- Let $P_X : \tilde{\mathcal{F}} \rightarrow \mathbb{R} \cup \{+\infty\}$ defined by $P_X(\tilde{B}) \equiv P[X^{-1}(\tilde{B})]$ for every $\tilde{B} \in \tilde{\mathcal{F}}$. Then

$(\tilde{\Omega}, \tilde{\mathcal{F}}, P_X)$ is a probability space, and P_X is called the *probability measure induced by X* .

DEFINITION: When $(\tilde{\Omega}, \tilde{\mathcal{F}}) = (\mathbb{R}^k, \mathcal{B}^k)$, the random element $X : \Omega \rightarrow \mathbb{R}^k$ is called a *random variable/vector* (r.v.).

REMARKS:

- The $\mathcal{F}/\mathcal{B}^k$ -measurability condition $\{X^{-1}(\tilde{B}) : \tilde{B} \in \mathcal{B}^k\} \subseteq \mathcal{F}$ is equivalent to

$$\{X^{-1}((-\infty, x]) : x \in \mathbb{R}^k\} \subseteq \mathcal{F}$$

- Let $F_X : \mathbb{R}^k \rightarrow [0, 1]$ defined by $F(x) \equiv P_X[(-\infty, x]] = P\{X^{-1}[(-\infty, x]]\}$ for every $x \in \mathbb{R}^k$.

$F_X(\cdot)$ is called the *distribution function of X*.

PROPERTIES OF DISTRIBUTION FUNCTIONS:

(a) A distribution function F_X on \mathbb{R}^k corresponds to a unique probability measure P_X satisfying

$$P_X[(-\infty, x]] = F_X(x) \text{ for every } x \in \mathbb{R}^k.$$

(b) A distribution function $F_X(\cdot)$ is monotone nondecreasing, right-continuous with left-limits

(*cadlag*), at most countable discontinuities, $\lim_{x \downarrow -\infty} F_X(x) = 0$ and $\lim_{x \uparrow +\infty} F_X(x) = 1$.

DEFINITION: (i) The events A_1, A_2, \dots in \mathcal{F} of a probability space (Ω, \mathcal{F}, P) are (mutually) *independent* iff for every finite collection A_{j_1}, \dots, A_{j_n} and every n

$$P(\cap_{i=1}^n A_{j_i}) = \prod_{i=1}^n P(A_{j_i})$$

(ii) The σ -algebras $\mathcal{F}_1, \mathcal{F}_2, \dots$ included in \mathcal{F} on a probability space (Ω, \mathcal{F}, P) are (mutually) *independent* iff every sequence A_1, A_2, \dots with $A_j \in \mathcal{F}_j$ is independent.

(iii) The r.v. X_1, X_2, \dots on a probability space (Ω, \mathcal{F}, P) are (mutually) *independent* iff the sequence $\mathcal{F}_{X_1}, \mathcal{F}_{X_2}, \dots$ is independent.

PROPERTY OF INDEPENDENCE:

- The r.v. X_1, X_2, \dots on a probability space (Ω, \mathcal{F}, P) are (mutually) independent iff

$$F_X(x) = \prod_{i=1}^n F_{X_{j_i}}(x_{j_i})$$

for every $x = (x_{j_1}, \dots, x_{j_n})$, (j_1, \dots, j_n) and n , where $X = (X_{j_1}, \dots, X_{j_n})$.

$F_X(\cdot)$ and $F_{X_j}(\cdot)$ are called the *joint and marginal distribution functions*, respectively.

II. Measurable Functions and Lebesgue Integration

II.1. Measurable Functions. (Bierens, Sections 2.2, 2.4)

DEFINITION: A $\mathcal{F}/\mathcal{B}^k$ -measurable function $g(\cdot)$ is a mapping from a measurable space (Ω, \mathcal{F}) to $(\mathbb{R}^k, \mathcal{B}^k)$ that satisfies $g^{-1}(B) \in \mathcal{F}$ for every $B \in \mathcal{B}^k$ where $g^{-1}(B) \equiv \{\omega : g(\omega) \in B\}$.

EXAMPLES:

- Same as a r.v. except that the latter is defined on (Ω, \mathcal{F}, P) .
- When $(\Omega, \mathcal{F}) = (\mathbb{R}^\ell, \mathcal{B}^\ell)$, then $g(\cdot)$ is simply called a *measurable function*.
 - Continuous functions from \mathbb{R}^ℓ to \mathbb{R}^k are measurable functions. The reverse is not true.

PROPERTIES:

- (a) $g(\cdot) = [g_1(\cdot), \dots, g_k(\cdot)]'$ is $\mathcal{F}/\mathcal{B}^k$ -measurable iff $g_i(\cdot)$ is \mathcal{F}/\mathcal{B} -measurable for every $i = 1, \dots, k$.
- (b) Let $X(\cdot)$ and $Y(\cdot)$ be two r.v. on (Ω, \mathcal{F}, P) . Then $\mathcal{F}_Y \subseteq \mathcal{F}_X$ iff there exists a measurable function $g(\cdot) : \mathbb{R}^{\dim X} \rightarrow \mathbb{R}^{\dim Y}$ such that $Y(\omega) = g[X(\omega)]$ for all $\omega \in \Omega$.

DEFINITION: A *simple function* $g(\cdot)$ is a \mathcal{F}/\mathcal{B} -measurable function from (Ω, \mathcal{F}) to $(\mathbb{R}, \mathcal{B})$ of the form

$$g(\omega) = \sum_{j=1}^m b_j \mathbb{I}(\omega \in A_j)$$

where $m < \infty$, $b_1, \dots, b_m \in \mathbb{R}$ and $\{A_j\}_{j=1}^m$ a partition of Ω with $A_j \in \mathcal{F}$ for $j = 1, \dots, m$.

PROPERTIES:

- (a) A function $g(\cdot) : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B})$ is \mathcal{F}/\mathcal{B} -measurable iff it is the (pointwise) limit of a sequence of simple functions $\{g_n(\cdot)\}_{n=1}^\infty$, i.e. $\lim_{n \rightarrow \infty} g_n(\omega) = g(\omega)$ for every $\omega \in \Omega$.
- (b) If $f(\cdot)$ and $g(\cdot)$ are \mathcal{F}/\mathcal{B} -measurable, then $f(\cdot) + g(\cdot)$, $f(\cdot) - g(\cdot)$, $f(\cdot) \times g(\cdot)$ and $f(\cdot)/g(\cdot)$ (with $g(\cdot) \neq 0$) are \mathcal{F}/\mathcal{B} -measurable,
- (c) Let $\{g_n(\cdot)\}_{n=1}^\infty$ be a sequence of \mathcal{F}/\mathcal{B} -measurable functions. Provided the RHS are well-defined,
 - (c.1) $\underline{f}(\cdot) \equiv \inf_{n \geq 1} g_n(\cdot)$ and $\bar{f}(\cdot) \equiv \sup_{n \geq 1} g_n(\cdot)$ are \mathcal{F}/\mathcal{B} -measurable,
 - (c.2) $\underline{h}(\cdot) \equiv \liminf_{n \rightarrow \infty} g_n(\cdot)$ and $\bar{h}(\cdot) \equiv \limsup_{n \rightarrow \infty} g_n(\cdot)$ are \mathcal{F}/\mathcal{B} -measurable,
 - (c.3) $g(\cdot) \equiv \lim_{n \rightarrow \infty} g_n(\cdot)$ is \mathcal{F}/\mathcal{B} -measurable.

II.2. Lebesgue Integration. (Bierens, Sections 2.2, 2.4, 4.3–4.4)

DEFINITION: Let $(\Omega, \mathcal{F}, \mu)$ be a measured space.

(i) The *integral* of a simple function $g(\cdot)$ w.r.t. μ is

$$\int_{\Omega} g(\omega) d\mu(\omega) \equiv \sum_{j=1}^m b_j \mu(A_j) \quad (\text{Convention: } 0 \times \infty = 0)$$

(ii) The *integral* of a $(\mu$ -a.e.) nonnegative \mathcal{F}/\mathcal{B} -measurable function $g(\cdot)$ w.r.t. μ is

$$\int_{\Omega} g(\omega) d\mu(\omega) \equiv \sup_{0 \leq g_S(\cdot) \leq g(\cdot)} \int_{\Omega} g_S(\omega) d\mu(\omega)$$

where $g_S(\cdot)$ is a simple function.

(iii) The *integral* of a \mathcal{F}/\mathcal{B} -measurable function $g(\cdot)$ w.r.t. μ is

$$\int_{\Omega} g(\omega) d\mu(\omega) \equiv \int_{\Omega} g_+(\omega) d\mu(\omega) - \int_{\Omega} g_-(\omega) d\mu(\omega)$$

where $g_+(\cdot) \equiv \max\{g(\cdot), 0\}$ and $g_-(\cdot) \equiv \max\{-g(\cdot), 0\}$ provided $\int_{\Omega} g_+(\omega) d\mu(\omega) < \infty$ or $\int_{\Omega} g_-(\omega) d\mu(\omega) < \infty$. If both are finite, then $g(\cdot)$ is μ -integrable.

REMARKS:

- $g(\cdot)$ is μ -integrable iff $|g(\cdot)|$ is μ -integrable, in which case $|\int_{\Omega} g(\omega)d\mu(\omega)| \leq \int_{\Omega} |g(\omega)|d\mu(\omega)$.
- If $g(\cdot)$ is μ -integrable, then $g(\cdot) < \infty$ μ -a.e. (allowing $g(\cdot)$ to have $\pm\infty$ values)
- Let $\int_A g(\omega)d\mu(\omega) \equiv \int_{\Omega} \mathbb{I}(\omega \in A)g(\omega)d\mu(\omega)$ for $A \in \mathcal{F}$. If $\mu(A) = 0$, then $\int_A g(\omega)d\mu(\omega) = 0$.
- When $(\Omega, \mathcal{F}) = (\mathbb{R}, \lambda)$, then $\int_A g(x)d\lambda(x)$ is called the *Lebesgue integral of $g(\cdot)$ on A* .
 - If a bounded function $g(\cdot)$ is Riemann integrable, then $g(\cdot)$ is Lebesgue integrable and $\int_a^b g(x)dx = \int_{[a,b]} g(x)d\lambda(x) < \infty$. Counterexample: $g(\cdot)$ is the Dirichlet function.
 - A bounded function $g(\cdot) : [a, b] \rightarrow \mathbb{R}$ is Riemann integrable iff its set of discontinuities is of λ -measure zero. (Lebesgue Theorem)

PROPERTIES OF LEBESGUE INTEGRATION: Let $g(\cdot)$ and $h(\cdot)$ be μ -integrable. Then

- (a) If $g(\cdot) = 0$ μ -a.e., then $\int_{\Omega} g(\omega)d\mu(\omega) = 0$.
- (b) $\int_{\Omega} [\alpha g(\omega) + \beta h(\omega)]d\mu(\omega) = \alpha \int_{\Omega} g(\omega)d\mu(\omega) + \beta \int_{\Omega} h(\omega)d\mu(\omega)$ for $\alpha, \beta \in \mathbb{R}$
- (c) If $g(\cdot) \leq h(\cdot)$ μ -a.e., then $\int_{\Omega} g(\omega)d\mu(\omega) \leq \int_{\Omega} h(\omega)d\mu(\omega)$

LIMITS AND LEBESGUE INTEGRATION:

- Let $g(\cdot)$ be μ -integrable, and $A_1, A_2, \dots \in \mathcal{F}$.

If $\lim_{n \rightarrow \infty} \mu(A_n) = 0$, then $\lim_{n \rightarrow \infty} \int_{A_n} g(\omega) d\mu(\omega) = 0$.

- *Monotone Convergence Theorem*: Let $\{g_n(\cdot)\}_{n=1}^{\infty}$ be a nondecreasing sequence of nonnegative \mathcal{F}/\mathcal{B} -measurable functions converging (pointwise) to a function $g(\cdot)$ μ -a.e. Then

$$\lim_{n \rightarrow \infty} \int_{\Omega} g_n(\omega) d\mu(\omega) = \int_{\Omega} g(\omega) d\mu(\omega)$$

- *Lebesgue Dominated Convergence Theorem*: Let $\{g_n(\cdot)\}_{n=1}^{\infty}$ be a sequence of \mathcal{F}/\mathcal{B} -measurable functions converging (pointwise) to a function $g(\cdot)$ satisfying $|g_n(\cdot)| \leq h(\cdot)$ μ -a.e. for every n where $h(\cdot)$ is μ -integrable. Then, the previous conclusion holds.

REMARK: Let $Q(\theta) \equiv \int_{\Omega} g(\omega; \theta) d\mu(\omega)$ where $g(\cdot; \theta)$ is μ -integrable for every $\theta \in \Theta \subset \mathbb{R}^p$. Then

(i) $Q(\cdot)$ is continuous if $g(\omega, \cdot)$ is continuous and $|g(\omega, \cdot)| \leq h(\omega)$ μ -integrable for μ -almost all ω

(ii) $Q'(\theta) = \int_{\Omega} g'(\omega; \theta) d\mu(\omega)$ if $g'(\omega; \cdot)$ exists and $|g'(\omega; \cdot)| \leq h(\omega)$ μ -integrable for μ -almost all ω

See e.g. Theorem 16.8 in Billingsley (1995) for conditions.

CHANGE-OF-VARIABLE: Let $g(\cdot)$ and $h(\cdot)$ be $\mathcal{B}^k/\mathcal{B}$ and $\mathcal{B}^k/\mathcal{B}^k$ -measurable. If $h(\cdot)$ is one-to-one from $A \in \mathcal{B}^k \rightarrow B \in \mathcal{B}^k$, and continuously differentiable with Jacobian $J(x) \equiv \partial h / \partial x'$, then

$$\int_B g(y) d\lambda(y) = \int_A g[h(x)] \cdot |J(x)| d\lambda(x).$$

PRODUCT MEASURE AND FUBINI'S THEOREM. (Billingsley, Section 18)

Let $(\Omega, \mathcal{F}, \mu)$ and $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mu})$ be two measure spaces with σ -finite measures μ and $\tilde{\mu}$.

DEFINITION: The *product σ -algebra* on $\Omega \times \tilde{\Omega}$ is $\mathcal{F} \times \tilde{\mathcal{F}} \equiv \sigma(\{A \times \tilde{A} : A \in \mathcal{F}, \tilde{A} \in \tilde{\mathcal{F}}\})$.

The *product measure* ν on $\mathcal{F} \times \tilde{\mathcal{F}}$ is the unique σ -finite measure satisfying

$$\nu(A \times \tilde{A}) = \mu(A) \cdot \tilde{\mu}(\tilde{A}) \quad \text{for every } A \in \mathcal{F}, \tilde{A} \in \tilde{\mathcal{F}}.$$

The triplet $(\Omega \times \tilde{\Omega}, \mathcal{F} \times \tilde{\mathcal{F}}, \nu)$ is called the *product measure space*.

PROPERTY: (Fubini's Theorem) Let $g(\cdot) : \Omega \times \tilde{\Omega} \rightarrow \mathbb{R}$ be $(\mathcal{F} \times \tilde{\mathcal{F}})/\mathcal{B}$ -measurable. Then

$$\int_{\Omega \times \tilde{\Omega}} g(\varpi) d\nu(\varpi) = \int_{\Omega} \left[\int_{\tilde{\Omega}} g(\omega, \tilde{\omega}) d\tilde{\mu}(\tilde{\omega}) \right] d\mu(\omega) = \int_{\tilde{\Omega}} \left[\int_{\Omega} g(\omega, \tilde{\omega}) d\mu(\omega) \right] d\tilde{\mu}(\tilde{\omega})$$

where $\varpi = (\omega, \tilde{\omega})$, provided one of the integrals is finite.

III. Expectations and Conditional Expectations

III.1. Expectations. (Bierens, Sections 2.5–2.8, 5.1)

DEFINITION: Let (Ω, \mathcal{F}, P) be a probability space, $X(\cdot)$ a r.v. in $\mathbb{R}^{\dim X}$, and $g(\cdot) : \mathbb{R}^{\dim X} \rightarrow \mathbb{R}^k$ measurable. Then $E[g(X)] \equiv \int_{\Omega} g[X(\omega)]dP(\omega) \equiv \left[\int_{\Omega} g_1[X(\omega)]dP(\omega), \dots, \int_{\Omega} g_k[X(\omega)]dP(\omega) \right]'$.

EXAMPLES:

- When $\dim X = 1$, the *raw and central p-th moment* of X are $E[X^p]$ and $E[X - E(X)]^p$.
→ $E[X]$ and $E[X - E(X)]^2$ are the *mean* and *variance* of X .
- The *variance-covariance matrix* of X is $E[X - E(X)][X - E(X)]' = E[XX'] - E[X]E[X]'$.

PROPERTIES:

- (a) $E[g(X)] = \int_{\mathbb{R}^{\dim X}} g(x)dP_X(x) = \int_{\mathbb{R}^{\dim Y}} ydP_Y(y)$ where $Y \equiv g(X)$, provided one of the integrals is finite.
- (b) The r.v. X and Y are independent iff $E[g(X)h(Y)] = E[g(X)] \cdot E[h(Y)]$ for all measurable functions $g(\cdot)$ and $h(\cdot)$ on $\mathbb{R}^{\dim X}$ and $\mathbb{R}^{\dim Y}$.

SOME INEQUALITIES:

- Markov inequality $\Pr[|X| \geq \epsilon] \leq E[|X|^p]/\epsilon^p$ for every $p > 0$. (\Rightarrow Chebyshev inequality)
- Jensen inequality $E[g(X)] \leq g(E[X])$ for $g(\cdot)$ concave on \mathcal{S}_X . (\Rightarrow Liapounov inequality)
- Holder inequality $E(|XY|) \leq (E[|X|^p])^{1/p}(E[|Y|^q])^{1/q}$ for $1/p+1/q=1$. (\Rightarrow Cauchy-Schwarz in.)
- Triangle inequality $(E[|X+Y|^p])^{1/p} \leq (E[|X|^p])^{1/p} + (E[|Y|^p])^{1/p}$ for $p \geq 1$. (\Rightarrow Minkowski in.)

DEFINITION: The *moment generating function* of X is $M_X(t) \equiv E[\exp(t'X)]$ for $t \in \mathcal{D}_X \subseteq \mathbb{R}^{\dim X}$. The *characteristic function* of X is $\phi_X(t) \equiv E[\exp(it'X)]$ for $t \in \mathbb{R}^{\dim X}$.

PROPERTIES:

(a) If $(-t_o, t_o) \subseteq \mathcal{D}_X \subseteq \mathbb{R}$ with $t_o > 0$, then $M^{(p)}(0) = E[X^p] < \infty$ for all $p \in \mathbb{N}$,

$M_X(t) = \sum_{p=0}^{\infty} \frac{E[X^p]}{p!} t^p$ and $P_X(\cdot)$ is uniquely determined by $M_X(\cdot)$ (and by the moments of X).

(b) $P_X(\cdot)$ is uniquely determined by $\phi_X(\cdot)$.

When $\dim X = 1$, $\phi_X(\cdot)$ is uniformly continuous. Moreover, $\phi^{(p)}(0) = i^p E[X^p]$ if $E[X^p] < \infty$.

(c) $\phi_X(t) = M_X(it)$ and $M_X(t) = \phi_X(-it)$ whenever $M_X(\cdot)$ is well-defined.

III.2. Conditional Expectations. (Bierens, Chapter 3)

→ $E[Y|X]$ where $Y(\cdot)$ is a r.v. and $X(\cdot)$ is a random element on (Ω, \mathcal{F}, P)

EXAMPLE: $(\Omega, \mathcal{F}, P) = (\{1, 2, 3, 4, 5, 6\}, 2^\Omega, P_0)$ where P_0 is the uniform probability measure.

Let $Y(\omega) = \omega$ and $X(\omega) = \mathbb{I}[\omega \in \{2, 4, 6\}]$. Then $E[Y|X] = 3 + X$.

DEFINITION: Let Y be a random variable on (Ω, \mathcal{F}, P) with $E[Y] < \infty$. Let $\mathcal{F}_0 \subseteq \mathcal{F}$ be a σ -algebra. The *conditional expectation of Y given \mathcal{F}_0* is a mapping $E[Y|\mathcal{F}_0](\cdot) : \Omega \rightarrow \mathbb{R}$ satisfying

- (i) $E[Y|\mathcal{F}_0](\cdot)$ is $\mathcal{F}_0/\mathcal{B}$ -measurable,
- (ii) $\int_A E[Y|\mathcal{F}_0](\omega) dP(\omega) = \int_A Y(\omega) dP(\omega)$, i.e., $E\{\mathbb{I}_A E[Y|\mathcal{F}_0]\} = E[\mathbb{I}_A Y]$ for every $A \in \mathcal{F}_0$.

REMARKS:

- $E[Y|\mathcal{F}_0](\cdot)$ is uniquely defined P -a.s.
- The *conditional expectation of Y given X* is $E[Y|X](\cdot) \equiv E[Y|\mathcal{F}_X](\cdot)$.
 - When $X \in \mathbb{R}^k$, then $E[Y|X](\cdot) = g[X(\cdot)]$ for some $\mathcal{B}^k/\mathcal{B}$ -measurable function $g(\cdot)$.
 - $E[Y|X=x] \equiv g(x) = E[Y|X](\omega)$ for all $\omega \in X^{-1}(x)$.

PROPERTIES OF CONDITIONAL EXPECTATIONS:

- (a) If $Y = c$, then $E[Y|\mathcal{F}_0] = c$.
- (b) $E[\alpha X + \beta Y|\mathcal{F}_0] = \alpha E[X|\mathcal{F}_0] + \beta E[Y|\mathcal{F}_0]$.
- (c) If $X(\cdot) \leq Y(\cdot)$, then $E[X|\mathcal{F}_0] \leq E[Y|\mathcal{F}_0]$.
- (d) If $\mathcal{F}_0 = \{\emptyset, \Omega\}$, then $E[Y|\mathcal{F}_0] = E[Y]$.
- (e) If $\mathcal{F}_Y \perp \mathcal{F}_0$, then $E[Y|\mathcal{F}_0] = E[Y]$.
- (f) If $\mathcal{F}_X \subseteq \mathcal{F}_0$, then $E[X|\mathcal{F}_0] = X$ and $E[XY|\mathcal{F}_0] = XE[Y|\mathcal{F}_0]$.
- (g) *Law of Iterated Expectations:* If $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}$, then

$$E\{E[Y|\mathcal{F}_0]|\mathcal{F}_1\} = E\{E[Y|\mathcal{F}_1]|\mathcal{F}_0\} = E[Y|\mathcal{F}_0].$$

- (h) The r.v.s X and Y are *conditionally independent* given \mathcal{F}_0 iff $E[g(X)h(Y)|\mathcal{F}_0] = E[g(X)|\mathcal{F}_0] \cdot E[h(Y)|\mathcal{F}_0]$ for all measurable functions $g(\cdot)$ and $h(\cdot)$ on $\mathbb{R}^{\dim X}$ and $\mathbb{R}^{\dim Y}$.
- (i) *Best MSE:* If $E[Y^2] < \infty$, then $E[Y|X] = \arg \min_{h \in \mathcal{H}} MSE(h) \equiv E[Y - h(X)]^2$ where \mathcal{H} is the set of squared-integrable (measurable) functions of X .

LIMITS AND CONDITIONAL EXPECTATIONS:

- *Monotone Convergence Theorem:* Let $\{Y_n(\cdot)\}_{n=1}^{\infty}$ be a nondecreasing sequence of nonnegative r.v.s with $E[\sup_{n \geq 1} Y_n] < \infty$. Then $\lim_{n \rightarrow \infty} E[Y_n | \mathcal{F}_0] = E[\lim_{n \rightarrow \infty} Y_n | \mathcal{F}_0]$.
- *Dominated Convergence Theorem:* Let $\{Y_n(\cdot)\}_{n=1}^{\infty}$ be a sequence of r.v.s converging to $Y(\cdot)$ P -a.s. and satisfying $|Y_n(\cdot)| \leq M(\cdot)$ with $E[M | \mathcal{F}_0] < \infty$. Then, $\lim_{n \rightarrow \infty} E[Y_n | \mathcal{F}_0] = E[Y | \mathcal{F}_0]$.
- Let $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots \subseteq \mathcal{F}$. Then $\lim_{n \rightarrow \infty} E[Y | \mathcal{F}_n] = E[Y | \mathcal{F}_{\infty}]$ where $\mathcal{F}_{\infty} \equiv \bigvee_{n=1}^{\infty} \mathcal{F}_n = \sigma(\bigcup_{n=1}^{\infty} \mathcal{F}_n)$.
 $\rightarrow \lim_{n \rightarrow \infty} E[Y_t | Y_{t-1}, \dots, Y_{t-n}] = E[Y_t | Y_{t-1}, Y_{t-2}, \dots]$

DEFINITION: Let Y be a r.v. on (Ω, \mathcal{F}, P) . The *conditional probability measure of $Y \in \mathbb{R}^{\ell}$ given $\mathcal{F}_0 \subseteq \mathcal{F}$* is the mapping $P_{Y|\mathcal{F}_0}(\cdot, \cdot) : (B, \omega) \in \mathcal{B}^{\ell} \times \Omega \rightarrow P_{Y|\mathcal{F}_0}(B, \omega) \equiv E[\mathbb{I}(Y \in B) | \mathcal{F}_0](\omega)$.

REMARKS:

- For each $\omega \in \Omega$, $(\mathbb{R}, \mathcal{B}^{\ell}, P_{Y|\mathcal{F}_0}(\cdot, \omega))$ is a probability space.
- The *cond. prob. measure of $Y \in \mathbb{R}^{\ell}$ given $X \in \mathbb{R}^k$* is $P_{Y|X}(\cdot | x) \equiv P_{Y|\mathcal{F}_X}(\cdot, \omega)$ for $\omega \in X^{-1}(x)$.
 $\rightarrow \int_A P_{Y|X}(B | x) dP_X(x) = P[X \in A, Y \in B]$ for every $(A, B) \in \mathcal{B}^k \times \mathcal{B}^{\ell}$.

IV. Densities and Distributions

IV.1. Densities. (Bierens, Sections 1.9)

DEFINITION: Let X be r.v. from (ω, \mathcal{F}, P) to $(\mathbb{R}^k, \mathcal{B}^k, P_X)$. Then P_X is *absolutely continuous w.r.t. a measure μ on $(\mathbb{R}^k, \mathcal{B}^k)$* , i.e. $P_X \ll \mu$, iff $\mu(B) = 0 \Rightarrow P_X(B) = 0$ for every $B \in \mathcal{B}^k$.

EXAMPLES: $k = 1$

- When $\mu(B) = \mu_c(B) \equiv \#(B \cap \mathbb{N})$, then P_X (or X) is *discrete* and $p_X(i) \equiv P[X = i]$.
- When $\mu(B) = \lambda(B)$, then P_X (or X) is *(absolutely) continuous*.

RADON-NIKODYM THEOREM: If $P_X \ll \mu$ with σ -finite μ , then there exists a nonnegative $\mathcal{B}^k/\mathcal{B}$ -measurable function $f_X(\cdot)$ satisfying $P_X(B) = \int_B f_X(x) d\mu(x)$ for every $B \in \mathcal{B}^k$.

→ The *(probability) density* $f_X(\cdot)$ is μ -a.e. unique.

→ $\int_B g(x) dP_X(x) = \int_B g(x) f_X(x) d\mu(x)$ and $F_X(\cdot) = \int_{(-\infty, \cdot]} f_X(x) d\mu(x)$.

→ When $\mu = \mu_c$, then $\int_B g(x) f_X(x) d\mu_c(x) = \sum_{i \in B} g(i) p_X(i)$ and $F_X(x) = \sum_{i \leq x} p_X(i)$.

→ When $\mu = \lambda$ and $F_X(\cdot)$ is continuously differentiable, then $f_X(\cdot) = F'_X(\cdot)$.

IV.2. Some Discrete and Continuous R.V.s. (Bierens, Sections 4.1, 4.5–4.6)

SOME UNIVARIATE DISCRETE R.V.s:

→ *Binomial Distribution*: $X \sim \mathcal{B}(n, p)$ where $n \in \mathbb{N}$ and $0 < p < 1$.

$$p_X(i) = \binom{n}{p} p^i (1-p)^{n-i} \text{ for } i = 0, 1, \dots, n.$$

$$E[X] = np, \text{ Var}[X] = np(1-p), \phi_X(t) = [1 - p + p \exp(it)]^n.$$

→ *Poisson Distribution*: $X \sim \mathcal{P}(\lambda)$ where $\lambda > 0$

$$p_X(i) = \frac{\lambda^i}{i!} \exp(-\lambda) \text{ for } i = 0, 1, 2, \dots$$

$$E[X] = \lambda, \text{ Var}[X] = \lambda, \phi_X(t) = \exp\{\lambda[\exp(it) - 1]\}.$$

SOME UNIVARIATE CONTINUOUS R.V.s:

→ *Normal Distribution*: $X \sim \mathcal{N}(\mu, \sigma^2)$ where $\mu \in \mathbb{R}$ and $\sigma^2 > 0$.

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}\right] \text{ for } x \in \mathbb{R}.$$

$$E[X] = \mu, \text{ Var}[X] = \sigma^2, \phi_X(t) = \exp(i\mu t - \sigma^2 t^2/2).$$

Standard Normal Distribution Z: $\mu = 0$ and $\sigma^2 = 1$.

SOME UNIVARIATE CONTINUOUS R.V.s: (continued)

→ *Chi-Square Distribution*: $X \sim \chi_k^2 = \sum_{j=1}^k Z_j^2$ where $k \in \mathbb{N}$ and Z_j are i.i.d. $\mathcal{N}(0, 1)$.

$$f_X(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} \exp(-x/2) \text{ for } x \geq 0, \text{ where } \Gamma(\alpha) \equiv \int_0^\infty x^{\alpha-1} \exp(-x) dx.$$

$$\mathbb{E}[X] = k, \text{ Var}[X] = 2k, \phi_X(t) = \frac{1}{(1-2it)^{k/2}}.$$

→ *Student-t Distribution*: $X \sim t_k = \frac{Z}{\sqrt{\chi_k^2/k}}$ where $k \in \mathbb{N}$ and $Z \sim \mathcal{N}(0, 1) \perp \chi_k^2$.

$$f_X(x) = \frac{1}{\sqrt{k}B(\frac{1}{2}, \frac{k}{2})} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}} \text{ for } x \in \mathbb{R}, \text{ where } B(\alpha, \beta) \equiv \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

$$\mathbb{E}[X] = 0 \text{ for } k \geq 2, \text{ Var}[X] = \frac{k}{k-2} \text{ for } k \geq 3, \phi_X(t) = \frac{2 \cdot \Gamma[(k+1)/2]}{\sqrt{k\pi} \cdot \Gamma(k/2)} \int_0^\infty \frac{\cos(tx)}{(1+x^2/k)^{(k+1)/2}} dx.$$

Cauchy Distribution: $k = 1$. $\mathbb{E}[X^p]$ does not exist or is infinite for any $p \in \mathbb{N}$.

→ *F Distribution*: $X \sim F(m, n) = \frac{\chi_m^2/m}{\chi_n^2/n}$ where $m, n \in \mathbb{N}$ and $\chi_m^2 \perp \chi_n^2$.

$$f_X(x) = \frac{m^{m/2} n^{n/2}}{B(m/2, n/2)} \frac{x^{m/2-1}}{(n+mx)^{(m+n)/2}} \text{ for } x \in [0, +\infty].$$

$$\mathbb{E}[X] = \frac{n}{n-2} \text{ for } n \geq 3, \text{ Var}[X] = \frac{2n^2(m+n-4)}{m(n-2)^2(n-4)} \text{ for } n \geq 5.$$

$\mathbb{E}[X] = \infty$ for $n = 1, 2$, $\text{Var}[X]$ does not exist or is infinite for $n = 1, 2, 3, 4$.

IV.3. The Multivariate Normal Distribution. (Bierens, Sections 5.2–5.3, 5.5)

DEFINITION: The r.v. X with values in \mathbb{R}^k has a *multivariate normal distribution* $\mathcal{N}(\mu, \Sigma)$ where $\mu \in \mathbb{R}^k$ and Σ is positive definite matrix iff its density w.r.t. Lebesgue measure is

$$f_X(x) = \frac{1}{(\sqrt{2\pi})^n \sqrt{\det(\Sigma)}} \exp \left[-\frac{1}{2}(x - \mu)' \Sigma^{-1}(x - \mu) \right] \quad \text{for } x \in \mathbb{R}^k.$$

PROPERTIES:

- (a) $E[X] = \mu$, $\text{Var}[X] = \Sigma$, $\phi_X(t) = \exp(it'\mu - t'\Sigma t/2)$.
- (b) If $Y = A + BX$ where $X \sim \mathcal{N}(\mu, \Sigma)$, then $Y \sim \mathcal{N}(A + B\mu, B\Sigma B')$.
- (c) If $X \sim \mathcal{N}(\mu, \Sigma)$, then its components are independent iff Σ is diagonal.
- (d) If $Y \in \mathbb{R}$ and $X \in \mathbb{R}^k$ are (jointly) normal distributed with $E[Y] = \mu_Y$, $E[X] = \mu_X$, $\text{Var}[Y] = \sigma_Y^2$, $\text{Var}[X] = \Sigma_{XX}$ nonsingular, $\text{Cov}[X, Y] = \Sigma_{XY}$, then $P_{Y|X}(\cdot|x)$ is normal with $E[Y|X] = \mu_Y + (X - \mu_X)' \Sigma_{XX}^{-1} \Sigma_{XY}$ and $\text{Var}[Y|X] = \sigma_Y^2 - \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY}$ with $\Sigma'_{YX} = \Sigma_{XY}$.
- (e) If $X \sim \mathcal{N}(0, \Sigma)$, then $X' \Sigma^{-1} X \sim \chi_k^2$.

V. Modes of Convergence (Bierens, Sections 6.2–6.3, 6.5–6.7)

DEFINITION: Let X and $\{X_n\}_{n=1}^{\infty}$ be random variables on (Ω, \mathcal{F}, P) .

- (i) X_n converges to X almost surely, i.e. $X_n \xrightarrow{a.s.} X$, iff $P[\lim_{n \rightarrow \infty} X_n = X] = 1$,
- (ii) X_n converges to X in probability, i.e. $X_n \xrightarrow{p} X$, iff $\forall \epsilon > 0 \lim_{n \rightarrow \infty} P[|X_n - X| \leq \epsilon] = 1$,
- (iii) X_n converges to X in distribution, i.e. $X_n \xrightarrow{d} X$, iff $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$
for all continuity points x of $F_X(\cdot)$.

REMARKS:

- Alternative definitions:

- (i) $X_n \xrightarrow{a.s.} X$ iff $\forall \epsilon > 0 \lim_{n \rightarrow \infty} P[\sup_{m \geq n} |X_m - X| \leq \epsilon] = 1$,
- (iii) $X_n \xrightarrow{d} X$ iff $\lim_{n \rightarrow \infty} E[g(X_n)] = E[g(X)]$ for all bounded continuous $g(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$.

- The multivariate case: $X \in \mathbb{R}^k$ and $X_n \in \mathbb{R}^k$

- (i)-(ii) $X_n \xrightarrow{a.s./p} X$ iff $X_n \xrightarrow{a.s./p} X$ componentwise,
- (iii) $X_n \xrightarrow{d} X$ iff $\lambda' X_n \xrightarrow{d} \lambda' X$ for all $\lambda \in \mathbb{R}^k$ iff $\lim_{n \rightarrow \infty} \phi_{X_n}(t) = \phi_X(t)$ for all $t \in \mathbb{R}^k$.

PROPERTIES:

(a) $X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$.

The reverse is not true except: $X_n \xrightarrow{p} c \Leftrightarrow X_n \xrightarrow{d} c$ where c is a constant.

(b) *Continuous Mapping Theorem*: Let $g(\cdot) : \mathbb{R}^k \rightarrow \mathbb{R}^\ell$ be continuous (P_X -a.s.).

If $X_n \xrightarrow{a.s./p/d} X$, then $g(X_n) \xrightarrow{a.s./p/d} g(X)$.

Examples: $g(X_n) = X_{1n} + X_{2n}$, $X_{1n}X_{2n}$, X_{1n}/X_{2n} when $P[X_2=0]=0$.

(c) *Slutsky Theorem*: If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$, then $g(X_n, Y_n) \xrightarrow{d} g(X, c)$ provided ...

BASIC LLN AND CLT: Let $X_i, i = 1, 2, \dots$ be i.i.d. and $\bar{X}_n \equiv \frac{1}{n} \sum_{i=1}^n X_i$.

- Strong LLN (Kolmogorov): If $E[X_i] \equiv \mu < \infty$, then $\bar{X}_n \xrightarrow{a.s.} \mu$.

- CLT: If $E[X_i] \equiv \mu < \infty$ and $\text{Var}[X_i] \equiv \Sigma < \infty$ nonsingular, then $\sqrt{n} \Sigma^{-1/2}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, I)$.

REMARKS: Extensions to non i.i.d X_i . See Bierens, Sections 7.2-7.3, 7.5. Let $\mu_i \equiv E[X_i]$.

- SLLNs: $\frac{1}{n} \sum_{i=1}^n (X_i - \mu_i) \xrightarrow{a.s.} 0 \Rightarrow$ WLLNs: $\frac{1}{n} \sum_{i=1}^n (X_i - \mu_i) \xrightarrow{p} 0 \left(\Leftrightarrow \frac{1}{n} \sum_{i=1}^n (X_i - \mu_i) \xrightarrow{d} 0 \right)$.

- CLTs: When $\dim X_i = 1$, $\frac{\sum_{i=1}^n (X_i - E[X_i])}{\sqrt{\text{Var}(\sum_{i=1}^n X_i)}} \xrightarrow{d} \mathcal{N}(0, 1)$.

THE LINEAR REGRESSION MODEL (Bierens, Section 5.7)

- Data: $(Y_i, X_i) \in \mathbb{R} \times \mathbb{R}^k, i = 1, 2, \dots$ i.i.d. on $(\Omega, \mathcal{F}, P_0)$.
- Model: $E[Y|X] = X'\beta_0$ or $Y_i = X_i'\beta_0 + \epsilon_i$ where $E[\epsilon_i|X_i] = 0$.
- *OLS estimator*: $\hat{\theta}_{OLS} = \arg \min_{\beta} (1/n) \sum_{i=1}^n (Y_i - X_i'\beta)^2$. Provided \mathbf{X} is *full-column rank*

$$\hat{\beta}_{OLS} = \left(\frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n X_i Y_i \right) = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$$

- Properties of $\hat{\beta}_{OLS}$: Assume $E[XX'] \equiv \Sigma_{XX}$ is p.d., and $0 < \sigma_{\epsilon}^2(X) \equiv E[\epsilon^2|X] < \infty$
 - (a) Finite sample property: $E[\hat{\beta}_{OLS}] = \beta_0$, i.e. $\hat{\beta}_{OLS}$ *unbiased estimator* of β_0 , etc.
 - (b) Asymptotic properties: $\hat{\beta}_{OLS} \xrightarrow{a.s.} \beta_0$ and $\sqrt{n}(\hat{\beta}_{OLS} - \beta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma_{XX}^{-1} E[\sigma_{\epsilon}^2(X) XX'] \Sigma_{XX}^{-1})$.
 - (c) Approximation: $\hat{\beta}_{OLS} \approx \mathcal{N}(\beta_0, (\mathbf{X}'\mathbf{X})^{-1} [\sum_{i=1}^n e_i^2 X_i X_i'] (\mathbf{X}'\mathbf{X})^{-1})$ where $e_i \equiv Y_i - X_i'\hat{\beta}_{OLS}$.
 $\rightarrow (\hat{\beta}_j - \beta_{j0})/\hat{\sigma}_j \approx \mathcal{N}(0, 1)$ where $\hat{\sigma}_j^2 \equiv [(\mathbf{X}'\mathbf{X})^{-1} [\sum_{i=1}^n e_i^2 X_i X_i'] (\mathbf{X}'\mathbf{X})^{-1}]_{jj}$ for $j = 1, \dots, k$.
 $\hat{\sigma}_j$ is called White (1980) *Heteroscedasticity Robust Standard Error*.
 - (d) Homoscedastic case $\sigma_{\epsilon}^2(\cdot) = \sigma_{\epsilon 0}^2$: $\hat{\beta}_{OLS} \approx \mathcal{N}(\beta_0, \tilde{\sigma}_{\epsilon}^2 (\mathbf{X}'\mathbf{X})^{-1})$ where $\tilde{\sigma}_{\epsilon}^2 \equiv \frac{1}{n-k} \sum_{i=1}^n e_i^2$.
 \rightarrow When $\epsilon_i \sim \mathcal{N}(0, \sigma_{\epsilon 0}^2)$, then $(\hat{\beta}_j - \beta_{j0})/\tilde{\sigma}_j \sim t_{n-k}$ where $\tilde{\sigma}_j^2 \equiv \tilde{\sigma}_{\epsilon}^2 [(\mathbf{X}'\mathbf{X})^{-1}]_{jj}$.

VI. M-Estimation (Bierens, Sections 6.4, 6.9)

DEFINITION: An *extremum* or *M-estimator* $\hat{\theta}_n \in \Theta \subseteq \mathbb{R}^k$ is a solution of $\max_{\theta \in \Theta} Q_n(\theta)$ or $\min_{\theta \in \Theta} \tilde{Q}_n(\theta)$, where $Q_n(\theta) = -\tilde{Q}_n(\theta)$ is a (measurable) function of the data.

EXAMPLES:

- OLS estimator $\hat{\beta}_{OLS} \equiv \arg \min_{\beta \in \mathbb{R}^k} \tilde{Q}_n(\beta) = \frac{1}{n} \sum_{i=1}^n (Y_i - X_i' \beta)^2$.
- NLLS estimator $\hat{\theta}_{NLLS} \equiv \arg \min_{\theta \in \mathbb{R}^k} \tilde{Q}_n(\beta) = \frac{1}{n} \sum_{i=1}^n [Y_i - m(X_i; \theta)]^2$.
- IV estimator $\hat{\beta}_{IV} \equiv \arg \min_{\beta \in \mathbb{R}^k} \tilde{Q}_n(\beta) = \left\| \frac{1}{n} \sum_{i=1}^n Z_i (Y_i - X_i' \beta) \right\|_{W_n}$.
- GMM estimator $\hat{\theta}_{GMM} \equiv \arg \min_{\theta \in \Theta} \tilde{Q}_n(\beta) = \left\| \frac{1}{n} \sum_{i=1}^n h(Y_i, X_i, Z_i; \theta) \right\|_{W_n}$.
- ML estimator $\hat{\theta}_{ML} \equiv \arg \max_{\theta \in \Theta} Q_n(\beta) = \frac{1}{n} \sum_{i=1}^n \log f(Y_i | X_i; \theta)$.

Master Consistency Theorem: For a nonstochastic function $Q_\infty(\cdot)$ on Θ and $\theta_* \in \Theta$. If

- (i) *Separation:* $\sup_{\theta \in \Theta \cap \mathcal{N}^c} Q_\infty(\theta) < Q_\infty(\theta_*)$, for any neighborhood \mathcal{N} of θ_* ,
- (ii) *Uniform convergence:* $\sup_{\theta \in \Theta} |Q_n(\theta) - Q_\infty(\theta)| \xrightarrow{a.s./p} 0$,

then $\hat{\theta}_n \xrightarrow{a.s./p} \theta_*$ and θ_* is called the *pseudo-true parameter value*.

Proof of Strong Consistency: Consider only the ω for which $\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} |Q_n(\theta) - Q_\infty(\theta)| = 0$.

Thus for $\epsilon > 0$ and n sufficiently large, $Q_\infty(\theta) - \epsilon/3 < Q_n(\theta) < Q_\infty(\theta) + \epsilon/3$ for all $\theta \in \Theta$.

Hence, (a) $Q_\infty(\theta_*) - \epsilon/3 < Q_n(\theta_*)$ and (b) $Q_n(\hat{\theta}_n) < Q_\infty(\hat{\theta}_n) + \epsilon/3$.

Because $Q_n(\hat{\theta}_n) \geq Q_n(\theta)$ for all $\theta \in \Theta$, then $Q_n(\hat{\theta}_n) > Q_\infty(\theta_*) - \epsilon/3$ by (a). Hence, by (b)

$$Q_\infty(\hat{\theta}_n) > Q_n(\hat{\theta}_n) - \epsilon/3 > Q_\infty(\theta_*) - 2\epsilon/3 > Q_\infty(\theta_*) - \epsilon.$$

Let \mathcal{N} be an arbitrary neighborhood of θ_* and set $\epsilon = Q_\infty(\theta_*) - \sup_{\theta \in \Theta \cap \mathcal{N}^c} Q_\infty(\theta)$. Thus,

$$Q_\infty(\hat{\theta}_n) > \sup_{\theta \in \Theta \cap \mathcal{N}^c} Q_\infty(\theta) \Rightarrow \hat{\theta}_n \in \Theta \cap \mathcal{N}.$$

REMARKS:

- Condition (i) $\Rightarrow \theta_*$ is the unique maximizer of $Q_\infty(\cdot)$ on Θ . See Lemma below for the reverse.
- Condition (ii) is often proved using a uniform SLLN. See below.
- $\theta_* \neq \theta_0$. If condition (i) is satisfied by θ_0 , then $\theta_* = \theta_0$.
- Neither $Q_n(\cdot)$ nor $Q_\infty(\cdot)$ need to be continuous on Θ .
- Generalization to $\Theta \subseteq \mathcal{G}$ with a metric d on \mathcal{G} .

LEMMA: Suppose (i) Θ is compact, (ii) $Q_\infty(\cdot)$ is continuous on Θ , and (iii) *Identification*: θ_* is the unique maximizer of $Q_\infty(\cdot)$ on Θ . Then condition (i) in the basic consistency theorem holds.

Proof of Lemma: By (i) $\Theta \cap \mathcal{N}^c$ is compact for any neighborhood \mathcal{N} of θ_* . Thus, by (ii) $\sup_{\theta \in \Theta \cap \mathcal{N}^c} Q_\infty(\theta) = Q_\infty(\bar{\theta})$ for some $\bar{\theta} \in \Theta$. By (iii) $Q_\infty(\bar{\theta}) < Q_\infty(\theta_*)$.

EXAMPLE: $\tilde{Q}_\infty(\beta) \equiv E[Y - X'\beta]^2$ and $\beta_* \equiv \arg \min_{\beta \in B} \tilde{Q}_\infty(\beta)$. If $E[Y^2] < \infty$ and $E[XX'] < \infty$ nonsingular, then $\tilde{Q}_\infty(\cdot)$ is continuous on B and $\beta_* = \arg \min_{\beta \in B} E[E(Y|X) - X'\beta]^2$.

$\rightarrow \beta_* = (E[XX'])^{-1}E[XY]$ if the latter $\in B \Rightarrow$ If $E[Y|X] = X'\beta_0$ with $\beta_0 \in B$, then $\beta_* = \beta_0$.

BASIC UNIFORM SLLN: Let X_1, X_2, \dots be i.i.d. r.v.s in \mathbb{R}^p and $g(\cdot; \cdot): \mathbb{R}^p \times \Theta \rightarrow \mathbb{R}$ be \mathbb{R}^p/\mathbb{R} -measurable and continuous on compact $\Theta \subset \mathbb{R}^k$. If $\sup_{\theta \in \Theta} |g(\cdot; \theta)| \leq M(\cdot)$ with $E[M(X_i)] < \infty$, then $\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^n \{g(X_i; \theta) - E[g(X_i; \theta)]\} \right| \xrightarrow{a.s.} 0$ and $E[g(X_i; \cdot)]$ is continuous on Θ .

REMARKS:

- Extensions to non i.i.d X_i . See Bierens, Section 7.4.
- Extensions to *empirical processes*: $\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{i=1}^n \{g(X_i) - E[g(X_i)]\} \right| \xrightarrow{a.s.} 0$.

EXAMPLE (CONTINUED): $\tilde{Q}_n(\beta) \equiv \frac{1}{n} \sum_{i=1}^n (Y_i - X_i' \beta)^2$ with B compact. We have $\tilde{g}(y, x; \beta) \equiv (y - x' \beta)^2 \leq 2[y^2 + (x' \beta)^2] \leq 2(y^2 + \|x\|^2 \|\beta\|^2) \leq M(y, x) \equiv 2(y^2 + M_o^2 \|x\|^2)$. If $E[Y_i^2] < \infty$ and $E[X_i X_i'] < \infty$, then $E[M(Y_i, X_i)] < \infty$. Hence, $\sup_{\beta \in B} |\tilde{Q}_n(\beta) - \tilde{Q}_\infty(\beta)| \xrightarrow{a.s./p} 0$.

Master Asymptotic Normality Theorem: Let $\hat{\theta}_n \xrightarrow{a.s./p} \theta_*$. If

- (i) $\theta_* \in \text{int}(\Theta)$, and $Q_n(\cdot)$ is twice continuously differentiable in a neighborhood \mathcal{N} of θ_* ,
 - (ii) $\sup_{\theta \in \mathcal{N}} \|\partial^2 Q_n(\theta) / \partial \theta \partial \theta' - H(\theta)\| \xrightarrow{a.s./p} 0$ where $H(\cdot)$ is continuous and nonsingular at θ_* ,
 - (iii) $\sqrt{n} \partial Q_n(\theta_*) / \partial \theta \xrightarrow{d} \mathcal{N}(0, \Sigma)$ for Σ nonsingular,
- then $\sqrt{n}(\hat{\theta}_n - \theta_*) \xrightarrow{d} \mathcal{N}(0, H^{-1} \Sigma H^{-1})$ where $H \equiv H(\theta_*)$.

Proof of Asymptotic Normality: By (i) $\partial Q_n(\hat{\theta}_n) / \partial \theta = 0$ where $\hat{\theta}_n \in \mathcal{N}$ with probability one or approaching one. Taking a second-order Taylor expansion at θ_* on the neighborhood \mathcal{N} gives

$$0 = \sqrt{n} \frac{\partial Q_n(\theta_*)}{\partial \theta} + \frac{\partial^2 Q_n(\bar{\theta}_n)}{\partial \theta \partial \theta'} \sqrt{n}(\hat{\theta}_n - \theta_*) \quad \text{where } \bar{\theta}_n \in [\theta_*, \hat{\theta}_n] \text{ (componentwise)}$$

By (ii) $\partial^2 Q_n(\bar{\theta}_n) / \partial \theta \partial \theta' \xrightarrow{a.s./p} H = H(\theta_*)$ since $\bar{\theta}_n \xrightarrow{a.s./p} \theta_*$. By (iii) and the continuous mapping theorem, the result follows since H is nonsingular.

REMARK: Condition (ii) by USLLN while condition (iii) from CLT at θ_* . See NLLS and MLE.

EXAMPLE (CONTINUED): $\frac{\partial^2 \tilde{Q}_n(\beta)}{\partial \beta \partial \beta'} = \frac{2}{n} \sum_{i=1}^n X_i X_i' \xrightarrow{a.s./p} 2\Sigma_{XX}$, $\sqrt{n} \frac{\partial \tilde{Q}_n(\beta_*)}{\partial \beta} = \frac{-2}{\sqrt{n}} \sum_{i=1}^n X_i (Y_i - X_i' \beta_*) \xrightarrow{d}$

$\mathcal{N}(0, 4E[\sigma_{\epsilon_*}^2(X)XX'])$, where $\sigma_{\epsilon_*}(X) \equiv E[\epsilon_*^2|X]$ and $\epsilon_* \equiv Y - X'\beta_*$. Thus,

$\sqrt{n}(\hat{\beta}_{OLS} - \beta_*) \xrightarrow{d} \mathcal{N}(0, \Sigma_{XX}^{-1} E[\sigma_{\epsilon_*}^2(X)XX'] \Sigma_{XX}^{-1})$ same result as before with β_* instead of β_0 .

→ Approximation: $\hat{\beta}_{OLS} \approx \mathcal{N}(\beta_*, (\mathbf{X}'\mathbf{X})^{-1} [\sum_{i=1}^n e_i^2 X_i X_i'] (\mathbf{X}'\mathbf{X})^{-1})$ where $e_i \equiv Y_i - X_i' \hat{\beta}_{OLS}$.

THEOREM (DELTA METHOD): Let $\sqrt{n}(X_n - c) \xrightarrow{d} \mathcal{N}_p(0, \Omega)$ and $g(\cdot) : \mathbb{R}^p \rightarrow \mathbb{R}^m$ be continuously differentiable on a neighborhood of $c \in \mathbb{R}^p$. Then $\sqrt{n}[g(X_n) - g(c)] \xrightarrow{d} \mathcal{N}_m\left(0, \frac{\partial g(c)}{\partial c'} \Omega \frac{\partial g(c)'}{\partial c}\right)$.

REMARKS:

- Different from continuous mapping theorem.

- $n[g(X_n) - g(c)]' \left[\frac{\partial g(X_n)}{\partial c'} \hat{\Omega} \frac{\partial g(X_n)'}{\partial c} \right]^{-1} [g(X_n) - g(c)] \xrightarrow{d} \chi_p^2$ where $\hat{\Omega} \xrightarrow{a.s./p} \Omega$ if $\frac{\partial g(c)}{\partial c'} \Omega \frac{\partial g(c)'}{\partial c}$ nonsingular.

→ Application to testing $H_0 : g(\theta_*) = 0$ vs. $H_A : g(\theta_*) \neq 0$. If $\sqrt{n}(\hat{\theta}_n - \theta_*) \xrightarrow{d} \mathcal{N}_k(0, \Omega)$,

then $n g(\hat{\theta}_n)' \left[\frac{\partial g(\hat{\theta}_n)}{\partial \theta'} \hat{\Omega} \frac{\partial g(\hat{\theta}_n)'}{\partial \theta} \right]^{-1} g(\hat{\theta}_n) \xrightarrow{d} \chi_p^2$ under H_0 and $\xrightarrow{a.s./p} +\infty$ under H_1

where $\hat{\Omega} \equiv \hat{H}^{-1} \hat{\Sigma} \hat{H}^{-1} \xrightarrow{a.s./p} \Omega = H^{-1} \Sigma H^{-1}$ and $\hat{H} = \frac{\partial^2 Q_n(\hat{\theta}_n)}{\partial \theta \partial \theta'}$. See below for $\hat{\Sigma}$.

VII. NLLS, ML and GMM Estimation (Bierens, Sections 6.4, 8.1-8.2, 8.4)

VII.1. Non-Linear Least Squares: $\hat{\theta}_{NLLS} \equiv \arg \min_{\theta \in \Theta} \tilde{Q}_n(\theta) = \frac{1}{n} \sum_{i=1}^n [Y_i - m(X_i; \theta)]^2$

THEOREM (NLLS): Let $(Y_1, X_1), (Y_2, X_2), \dots$ be i.i.d. r.v.s in $\mathbb{R} \times \mathbb{R}^p$ and $m(\cdot; \cdot) : \mathbb{R}^p \times \Theta \rightarrow \mathbb{R}$ be measurable on \mathbb{R}^p and continuous on $\Theta \subset \mathbb{R}^k$. Suppose that $E[Y^2] < \infty$ and θ_* uniquely $\min_{\theta \in \Theta} Q_\infty = E[Y - m(X; \theta)]^2$.

(i) If Θ compact and $\sup_{\theta \in \Theta} |m(\cdot; \theta)| \leq M(\cdot)$ with $E[M^2(X)] < \infty$, then $\hat{\theta}_{NLLS} \xrightarrow{a.s./p} \theta_*$.

(ii) If in addition $\theta_* \in \text{int}(\Theta)$ with $m(\cdot, \cdot)$ twice cont. differentiable on a neighborhood \mathcal{N} of θ_* ,
 $\sup_{\theta \in \mathcal{N}} \|\frac{\partial m(\cdot; \theta)}{\partial \theta}\| \leq M_1(\cdot)$ with $E[M_1^2(X)] < \infty$, $\sup_{\theta \in \mathcal{N}} \|\frac{\partial^2 m(\cdot; \theta)}{\partial \theta \partial \theta'}\| \leq M_2(\cdot)$ with $E[M_2^2(X)] < \infty$,
then $\sqrt{n}(\hat{\theta}_{NLLS} - \theta_*) \xrightarrow{d} \mathcal{N}(0, H^{-1} \Sigma H^{-1})$ provided $H = 2E \left[\epsilon_* \frac{\partial^2 m(X; \theta_*)}{\partial \theta \partial \theta'} - \frac{\partial m(X; \theta_*)}{\partial \theta} \frac{\partial m(X; \theta_*)}{\partial \theta'} \right]$
is nonsingular and $\Sigma = 4E \left[\sigma_{\epsilon_*}^2(X) \frac{\partial m(X; \theta_*)}{\partial \theta} \frac{\partial m(X; \theta_*)}{\partial \theta'} \right]$ with $\sigma_{\epsilon_*}^2(X) = E[\epsilon_*^2 | X]$ and $\epsilon_* = Y - m(X; \theta_*)$.

REMARKS: Let $e_i \equiv Y_i - m(X_i; \hat{\theta}_{NLLS})$. Then $\hat{\Sigma} = \frac{4}{n} \sum_{i=1}^n e_i^2 \frac{\partial m(X_i; \hat{\theta}_{NLLS})}{\partial \theta} \frac{\partial m(X_i; \hat{\theta}_{NLLS})}{\partial \theta'} \xrightarrow{a.s./p} \Sigma$

and $\hat{H} = \frac{2}{n} \sum_{i=1}^n \left[e_i \frac{\partial^2 m(X; \hat{\theta}_{NLLS})}{\partial \theta \partial \theta'} - \frac{\partial m(X; \hat{\theta}_{NLLS})}{\partial \theta} \frac{\partial m(X; \hat{\theta}_{NLLS})}{\partial \theta'} \right] \xrightarrow{a.s./p} H$.

- If $E[Y|X] = m(X; \theta_0)$ for $\theta_0 \in \Theta$, then $\theta_* = \theta_0$. If, in addition $\sigma_{\epsilon_*}^2(X) = \sigma_0^2$, then

$$H = -2E \left[\frac{\partial m(X; \theta_0)}{\partial \theta} \frac{\partial m(X; \theta_0)}{\partial \theta'} \right], \quad \Sigma = 4\sigma_0^2 E \left[\frac{\partial m(X; \theta_0)}{\partial \theta} \frac{\partial m(X; \theta_0)}{\partial \theta'} \right], \quad H^{-1} \Sigma H^{-1} = \sigma_0^2 \left(E \left[\frac{\partial m(X; \theta_0)}{\partial \theta} \frac{\partial m(X; \theta_0)}{\partial \theta'} \right] \right)^{-1}$$

VII.2. Maximum Likelihood: $\hat{\theta}_{ML} \equiv \arg \max_{\theta \in \Theta} Q_n(\theta) = \frac{1}{n} \sum_{i=1}^n \log f(Y_i|X_i; \theta)$

THEOREM (ML): Let $(Y_1, X_1), (Y_2, X_2), \dots$ be i.i.d. r.v.s in $\mathbb{R}^m \times \mathbb{R}^p$. Let the *model* be a family $\{f(\cdot|\cdot; \theta) : \theta \in \Theta\}$ of conditional densities w.r.t. a measure μ on \mathbb{R}^m and continuous on $\Theta \subset \mathbb{R}^k$. Suppose that θ_* uniquely $\max_{\theta \in \Theta} Q_\infty(\theta) = E[\log f(Y|X; \theta)]$.

- (i) If Θ compact and $\sup_{\theta \in \Theta} |\log f(\cdot|\cdot; \theta)| \leq M(\cdot, \cdot)$ with $E[M(Y, X)] < \infty$, then $\hat{\theta}_{ML} \xrightarrow{a.s./p} \theta_*$.
- (ii) If in addition $\theta_* \in \text{int}(\Theta)$ with $f(\cdot|\cdot; \cdot)$ twice continuously differentiable on a neighborhood \mathcal{N} of θ_* , $\sup_{\theta \in \mathcal{N}} \|\frac{\partial^2 \log f(\cdot|\cdot; \theta)}{\partial \theta \partial \theta'}\| \leq M_2(\cdot, \cdot)$, $E[M_2(Y, X)] < \infty$ and $E\left[\frac{\partial^2 \log f(Y|X; \theta_*)}{\partial \theta \partial \theta'}\right]$ nonsingular, then

$$\sqrt{n}(\hat{\theta}_{ML} - \theta_*) \xrightarrow{d} \mathcal{N}(0, A^{-1}BA^{-1})$$

$$\text{where } A = A(\theta_*) \equiv E\left[\frac{\partial^2 \log f(Y|X; \theta_*)}{\partial \theta \partial \theta'}\right] \text{ and } B = B(\theta_*) \equiv E\left[\frac{\partial \log f(Y|X; \theta_*)}{\partial \theta} \frac{\partial \log f(Y|X; \theta_*)}{\partial \theta'}\right].$$

REMARKS:

- If $\sup_{\theta \in \mathcal{N}} \|\frac{\partial \log f(\cdot|\cdot; \theta)}{\partial \theta}\| \leq M_1(\cdot)$ with $E[M_1(X)] < \infty$, then

$$\hat{A} = \frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \log f(Y_i|X_i; \hat{\theta}_{ML})}{\partial \theta \partial \theta'} \xrightarrow{a.s./p} A \text{ and } \hat{B} = \frac{1}{n} \sum_{i=1}^n \frac{\partial \log f(Y_i|X_i; \hat{\theta}_{ML})}{\partial \theta} \frac{\partial \log f(Y_i|X_i; \hat{\theta}_{ML})}{\partial \theta'} \xrightarrow{a.s./p} B.$$

- *Kullback-Leibler Divergence:* $\theta_* = \arg \min_{\theta \in \Theta} KLIC[f_{Y|X}(\cdot|\cdot); f(\cdot|\cdot; \theta)] \equiv E\left[\log \frac{f_{Y|X}(Y|X)}{f(Y|X; \theta)}\right] \geq 0.$

DEFINITION: $\{f(\cdot|\cdot; \theta) : \theta \in \Theta\}$ is *correctly specified* iff $f_{Y|X}(\cdot|\cdot) = f(\cdot|\cdot; \theta_0)$ a.s. for $\theta_0 \in \Theta$.

It is *one-to-one parameterized* iff $f(\cdot|\cdot; \theta_1) = f(\cdot|\cdot; \theta_2)$ a.s. $\Leftrightarrow \theta_1 = \theta_2$ for any $\theta_1, \theta_2 \in \Theta$.

LEMMA: If $\{f(\cdot|\cdot; \theta) : \theta \in \Theta\}$ is correctly specified and one-to-one parameterized, then θ_0 uniquely solves $\max_{\theta \in \Theta} E[\log f(Y|X; \theta)]$.

Proof of Lemma: If $f(y|x; \theta_0) > 0$, then $\log \frac{f(y|x; \theta)}{f(y|x; \theta_0)} \leq \frac{f(y|x; \theta)}{f(y|x; \theta_0)} - 1$ with equality iff $\frac{f(y|x; \theta)}{f(y|x; \theta_0)} = 1$. Thus $E \left[\log \frac{f(y|x; \theta)}{f(y|x; \theta_0)} \right] \leq E \left[\frac{f(y|x; \theta)}{f(y|x; \theta_0)} \right] - 1 \leq \int_{\{f(y|x; \theta_0) > 0\}} f(y|x; \theta) d\mu(y) dF_X(x) - 1 \leq 0$. Hence, $\theta_0 = \arg \max_{\theta \in \Theta} E[\log f(Y|X; \theta)]$. Moreover, $E \left[\log \frac{f(y|x; \theta)}{f(y|x; \theta_0)} \right] = 0 \Rightarrow E \left[\log \frac{f(y|x; \theta)}{f(y|x; \theta_0)} - \frac{f(y|x; \theta)}{f(y|x; \theta_0)} + 1 \right] = 0 \Rightarrow \log \frac{f(y|x; \theta)}{f(y|x; \theta_0)} = \frac{f(y|x; \theta)}{f(y|x; \theta_0)} - 1$ a.s. $\Rightarrow \frac{f(y|x; \theta)}{f(y|x; \theta_0)} = 1$ a.s. $\Rightarrow \theta = \theta_0$, i.e. θ_0 is the unique maximizer.

LEMMA (INFORMATION MATRIX EQUALITY): Assume that $\{f(\cdot|\cdot; \theta) : \theta \in \Theta\}$ is correctly specified. If one can switch integration and differentiation, then $A(\theta_0) + B(\theta_0) = 0$.

Proof of Lemma: Differentiating twice $\int f(y|x; \theta) d\mu(y) = 1$ w.r.t. θ gives

$$\int \frac{\partial^2 \log f(y|x; \theta)}{\partial \theta \partial \theta'} f(y|x; \theta) d\mu(y) + \int \frac{\partial \log f(y|x; \theta)}{\partial \theta} \frac{\partial \log f(y|x; \theta)}{\partial \theta'} f(y|x; \theta) d\mu(y) = 0 \text{ for every } \theta \in \Theta.$$

Letting $\theta = \theta_0$ in $A(\theta)$ and $B(\theta)$ gives the result because $f_{Y|X}(\cdot|\cdot) = f(\cdot|\cdot; \theta_0)$.

REMARKS:

- *Information Matrix Test*: White (1982) $H_0 : A(\theta_*) + B(\theta_*) = 0$ vs. $H_A : A(\theta_*) + B(\theta_*) \neq 0$.
- If $\{f(\cdot|\cdot; \theta) : \theta \in \Theta\}$ is correctly specified and one-to-one parameterized, then
 - $\rightarrow \sqrt{n}(\hat{\theta}_{ML} - \theta_0) \xrightarrow{d} \mathcal{N}(0, B^{-1}) = \mathcal{N}(0, -A^{-1}) \Rightarrow \hat{\theta}_{ML} \approx \mathcal{N}\left(\theta_0, -\left[\sum_{i=1}^n \frac{\partial^2 \log f(Y_i|X_i; \hat{\theta}_{ML})}{\partial \theta \partial \theta'}\right]^{-1}\right)$.
 - $\rightarrow \hat{\theta}_{ML}$ is *asymptotically efficient* among consistent estimators $\hat{\theta}_n$, i.e., $\lim_{n \rightarrow \infty} \text{Var}[\sqrt{n}\hat{\theta}_n] \geq B^{-1}$.
 - $B = E\left[\frac{\partial \log f(Y|X; \theta_0)}{\partial \theta} \frac{\partial \log f(Y|X; \theta_0)}{\partial \theta'}\right]$ is the *Fisher Information Matrix for one observation*.

THEOREM (CRAMER-RAO): If $\hat{\theta}_n$ is an unbiased estimator of θ_0 , then $\text{Var}[\sqrt{n}\hat{\theta}_n] \geq B^{-1}$.

Proof of Cramer-Rao Bound: We have $\int g_n(y_1, x_1, \dots, y_n, x_n) \Pi_{i=1}^n f(y_i|x_i; \theta_0) d\mu(y_i) dF_X(x_i) = \theta_0$ for any $\theta_0 \in \Theta$, where $\hat{\theta}_n = g_n(Y_1, X_1, \dots, Y_n, X_n)$. Differentiating w.r.t. θ_0 gives $\int g_n(y_1, x_1, \dots, y_n, x_n) \left[\sum_{i=1}^n \frac{\partial \log f(y_i|x_i; \theta_0)}{\partial \theta'}\right] \Pi_{i=1}^n f(y_i|x_i; \theta_0) d\mu(y_i) dF_X(x_i) = I$, i.e., $E[\hat{\theta}_n S'_n] = I$ where $S_n \equiv \sum_{i=1}^n \frac{\partial \log f(Y_i|X_i; \theta_0)}{\partial \theta}$. But $E[S_n] = 0$ by differentiating $\int f(y_i|x_i; \theta_0) d\mu(y_i) dF_X(x_i) = 1$ w.r.t. θ_0 . Hence $\text{Cov}[\hat{\theta}_n, S_n] = I$ and $\text{Var}[S_n] = nB$. Now, $\text{Var}[\hat{\theta}_n] = \text{Var}[(nB)^{-1}S_n + \hat{\theta}_n - (nB)^{-1}S_n] = (nB)^{-1} + \text{Var}[\hat{\theta}_n - (nB)^{-1}S_n] \geq (nB)^{-1}$ since $\text{Cov}[(nB)^{-1}S_n, \hat{\theta}_n - (nB)^{-1}S_n] = 0$.

VII.3. GMM: $\hat{\theta}_{GMM} \equiv \arg \min_{\theta \in \Theta} \tilde{Q}_n(\theta) = \|\frac{1}{n} \sum_{i=1}^n h(Y_i, X_i, Z_i; \theta)\|_{W_n}$ where $\dim h = \ell$

EXAMPLE (demand and supply): For $j = d, s$, $Q = \alpha_j P + X_j' \beta_j + \epsilon_j$ where $E[Z \epsilon_j] = 0$.

REMARK: GMM estimation includes NLLS and ML estimation. Z are the *instruments* (IV).

THEOREM (GMM): Let $(Y_1, X_1, Z_1), (Y_2, X_2, Z_2), \dots$ be i.i.d. r.v.s in $\mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^q$. Suppose $h(\cdot, \cdot, \cdot, \theta)$ is C^0 on Θ , θ_* uniquely $\min_{\theta \in \Theta} Q_\infty(\theta) = \|E[h(Y, X, Z; \theta)]\|_W$, and $W_n \xrightarrow{a.s./p} W$ p.d.

(ii) If Θ compact and $\sup_{\theta \in \Theta} |h(\cdot, \cdot, \cdot; \theta)| \leq M(\cdot, \cdot, \cdot)$ with $E[M(Y, X, Z)] < \infty$, then $\hat{\theta}_{GMM} \xrightarrow{a.s./p} \theta_*$.

(ii) If in addition $\theta_* \in \text{int}(\Theta)$ with $h(\cdot, \cdot, \cdot; \theta)$ cont. differentiable on a neighborhood \mathcal{N} of θ_* ,

$\sup_{\theta \in \mathcal{N}} \|\frac{\partial h(\cdot, \cdot, \cdot; \theta)}{\partial \theta'}\| \leq M_2(\cdot, \cdot, \cdot)$, $E[M_2(Y, X)] < \infty$ and $G \equiv E[\frac{\partial h(Y, X, Z; \theta_*)}{\partial \theta'}]$ full-column rank, then

$\sqrt{n}(\hat{\theta}_{GMM} - \theta_*) \xrightarrow{d} \mathcal{N}(0, (G'WG)^{-1}G'W\Omega WG(G'WG)^{-1})$ provided $E[h(Y, X, Z; \theta_*)] = 0$

where $\Omega \equiv E[h(Y, X, Z; \theta_*)h(Y, X, Z; \theta_*)']$ is assumed to be finite and nonsingular.

REMARKS:

- $\theta^* = \theta_0$ if $E[h(Y, X, Z; \theta_0)] = 0$ for $\theta_0 \in \Theta$ (correct specification of the moments)
- *Order Condition:* $\ell \geq k$ (otherwise θ_* cannot be unique)

- $\hat{G} \equiv \frac{1}{n} \sum_{i=1}^n \frac{\partial h(Y_i, X_i, Z_i, \hat{\theta}_{GMM})}{\partial \theta'} \xrightarrow{a.s./p} G$ and $\hat{\Omega} \equiv \frac{1}{n} \sum_{i=1}^n h(Y_i, X_i, Z_i, \hat{\theta}_{GMM}) h(Y_i, X_i, Z_i, \hat{\theta}_{GMM})' \xrightarrow{a.s./p} \Omega$
- $W^* = \Omega^{-1}$ gives optimal GMM estimator $\sqrt{n}(\hat{\theta}_{GMM}^* - \theta_*) \xrightarrow{d} \mathcal{N}(0, (G' \Omega^{-1} G)^{-1})$

Proof: $(G'WG)^{-1}G'W\Omega WG(G'WG)^{-1} - (G'\Omega^{-1}G)^{-1} = D'(I - E)D = D'(I - E)^2D$ psd

where $D' \equiv (G'WG)^{-1}G'W\Omega^{1/2}$ and $E \equiv \Omega^{-1/2}G(G'\Omega^{-1}G)^{-1}G'\Omega^{-1/2}$ idempotent

- Implementation: (i) Two-Step GMM, (ii) Iterated GMM, (iii) Continuous updating GMM

$$\hat{\theta}_{UCGMM} \equiv \arg \min_{\theta \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^n h(Y_i, X_i, Z_i; \theta) \right\|_{\hat{\Omega}(\theta)} \text{ where } \hat{\Omega}(\theta) \text{ is from } \hat{\Omega} = \hat{\Omega}(\hat{\theta}_{GMM})$$

- *Overidentification* or *J-Test*: $H_0 : E[h(Y, X, Z; \theta_*)] = 0$ vs. $H_1 : E[h(Y, X, Z; \theta)] \neq 0, \forall \theta \in \Theta$

$$\text{Under } H_0 : J \equiv n \cdot \left\| \frac{1}{n} \sum_{i=1}^n h(Y_i, X_i, Z_i, \hat{\theta}_{GMM}^*) \right\|_{\hat{\Omega}^{-1}} \xrightarrow{d} \chi_{\ell-k}^2$$

$$\text{Under } H_1 : J \xrightarrow{a.s./p} +\infty \text{ (consistent test)}$$

Proof: (sketch) $H_1 : E[h(Y, X, Z; \theta_*)] \neq 0$ (by contradiction). Let $h_i(\theta) \equiv h(Y_i, X_i, Z_i; \theta)$.

Taylor expansions of $\frac{1}{n} \sum_{i=1}^n h_i(\hat{\theta})$ and FOC of $\hat{\theta} = \hat{\theta}_{GMM}$ around θ_* with $\bar{\theta}, \tilde{\theta} \in (\theta_*, \hat{\theta})$ give

$$\frac{1}{n} \sum_{i=1}^n h_i(\hat{\theta}) = \left\{ I_\ell - \left(\frac{1}{n} \sum_{i=1}^n \frac{\partial h_i(\bar{\theta})}{\partial \theta'} \right) \left[\left(\frac{1}{n} \sum_{i=1}^n \frac{\partial h_i(\hat{\theta})'}{\partial \theta} \right) \hat{\Omega}^{-1} \left(\frac{1}{n} \sum_{i=1}^n \frac{\partial h_i(\tilde{\theta})}{\partial \theta'} \right) \right]^{-1} \left(\frac{1}{n} \sum_{i=1}^n \frac{\partial h_i(\hat{\theta})'}{\partial \theta} \right) \hat{\Omega}^{-1} \right\} \frac{1}{n} \sum_{i=1}^n h_i(\theta_*)$$

Under H_0 : $\frac{1}{\sqrt{n}} \sum_{i=1}^n h_i(\hat{\theta}) \xrightarrow{d} \mathcal{N}(0, V)$, $V = \Omega - G(G'\Omega^{-1}G)^{-1}G'$, $\text{Rk}(V) = \ell - \text{Tr}(E) = \ell - k$, $\Omega^{-1} = V^-$.

That's All Folks !