Econometrics II TA Session #7[*]

Makoto SHIMOSHIMIZU[†]

December 3, 2019

Contents

---

[*]All comments welcome!
[†]E-mail: vgm067sm@student.econ.osaka-u.ac.jp

1 Lebesgue Stieltjes Expression

From the definition of the expectation of a random variable, we can symbolically write the expectation based on the **Lebesgue Stieltjes integral** as

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \mathrm{d}F(x) := \begin{cases} \displaystyle\sum_i x_i \mathbb{P}[X = x_i] & \text{in the case of a discrete random variable;} \\ \displaystyle\int_{-\infty}^{\infty} x f(x) \mathrm{d}x & \text{in the case of a continuous random variable.} \end{cases}$$

The mapping $\mathbb{P}\colon \Omega \to \mathbb{R}$ stands for the probability that the realized value of X becomes x_i on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Also, $f(x)$ represents the probability density function defined as the derivative of the cummurative density function $F(x)\colon \mathbb{R} \to \mathbb{R}$:

$$\frac{\mathrm{d}F(x)}{\mathrm{d}x} = f(x),$$

provided that the derivative exists.

2 Markov's Inequality and Chebyshev's Inequality

In this section, we review two useful theorems providing upper bounds on some probability. First, we provide the **Markov's inequality**.

Theorem 2.1 (Markov's Inequality). If X is a nonnegative random variable and δ is a positive constant, then

$$\mathbb{P}[X \geq \delta] \leq \frac{\mathbb{E}[X]}{\delta}. \tag{2.1}$$

Moreover, If $\phi\colon \mathbb{R} \to \mathbb{R}$ is a monotonically increasing nonnegative function for the non-negative reals, $X\colon \Omega \to \mathbb{R}$ is a random variable, $\delta \geq 0$, and $\phi(\delta) > 0$, then

$$\mathbb{P}[|X| \geq \delta] = \mathbb{P}[\phi(|X|) \geq \phi(\delta)] \leq \frac{\mathbb{E}[\phi(|X|)]}{\phi(\delta)}. \tag{2.2}$$

Proof. We prove Eq. (2.1). Since the random variable X is a nonnegative random variable,

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \mathrm{d}F(x) = \int_{0}^{\infty} x \mathrm{d}F(x).$$

From this we can derive

$$\mathbb{E}[X] = \int_{0}^{\infty} x \mathrm{d}F(x) = \int_{0}^{\delta} x \mathrm{d}F(x) + \int_{\delta}^{\infty} x \mathrm{d}F(x)$$

$$\geq \int_{\delta}^{\infty} x \mathrm{d}F(x) \geq \int_{\delta}^{\infty} \delta \mathrm{d}F(x) = \delta \int_{\delta}^{\infty} \mathrm{d}F(x) = \delta \mathbb{P}[X \geq \delta].$$

From this it is easy to see that Eq. (2.1) holds. A similar calculation yields the extended (or general) form of the Markov's inequality or Eq. (2.2). $\qquad\square$

The Markov's inequality gives an upper bound for the probability that a non–negative function of a random variable is greater than or equal to some positive constant. Next, we present the **Chebyshev's inequality**.

Theorem 2.2 (Chebyshev's Inequality). Let X be a (integrable) random variable with finite expected value μ and finite non-zero variance σ^2. Then for any real number $\varepsilon > 0$,

$$\mathbb{P}[|X - \mu| \geq \varepsilon] \leq \frac{\mathbb{E}[|X - \mu|^2]}{\varepsilon^2} = \frac{\mathbb{V}[X]}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}.$$

Proof. We can prove the above theorem via a direct method. Using the indicator function:

$$I_A = \begin{cases} 1 & \text{if the event } A \text{ occurs;} \\ 0 & \text{otherwise,} \end{cases}$$

we have

$$\begin{aligned}
\mathbb{P}[|X - \mu| \geq \varepsilon] &= \mathbb{E}[I_{|X-\mu|\geq\varepsilon}] \\
&= \int_{-\infty}^{\infty} I_{|x-\mu|\geq\varepsilon} \mathrm{d}F(x) \\
&= \int_{-\infty}^{\infty} I_{\left|\frac{x-\mu}{\varepsilon}\right|\geq 1} \mathrm{d}F(x) \\
&\leq \int_{-\infty}^{\infty} \left|\frac{x-\mu}{\varepsilon}\right| I_{\left|\frac{X-\mu}{\varepsilon}\right|\geq 1} \mathrm{d}F(x) \\
&\leq \int_{-\infty}^{\infty} \left|\frac{x-\mu}{\varepsilon}\right|^2 I_{\left|\frac{X-\mu}{\varepsilon}\right|\geq 1} \mathrm{d}F(x) \\
&\leq \int_{-\infty}^{\infty} \left|\frac{x-\mu}{\varepsilon}\right|^2 \mathrm{d}F(x) \\
&= \frac{1}{\varepsilon^2} \int_{-\infty}^{\infty} |x-\mu|^2 \, \mathrm{d}F(x) \\
&= \frac{\mathbb{E}[|X - \mu|^2]}{\varepsilon^2}.
\end{aligned}$$

Rewriting this yields the Chebyshev's inequality. $\qquad\square$

Chebyshev's inequality guarantees that, for a wide class of probability distributions, no more than a certain fraction of values can be more than a certain distance from the mean.

3 Law of Large Numbers

In this section, we will discuss an important result, the so-called law of large numbers (LLN), which has an important role in probability and statistics. The law of large numbers (LLN) states that the average of a large number of i.i.d. random variables converge to the expected value. There are two main versions of the law of large numbers, which are called the Weak and Strong Laws of Large Numbers. The difference between them is mostly theoretical.

3.1 Strong Law of Large Numbers

The **Strong Law of Large Numbers (SLLN)** states that the average of a large number of i.i.d. random variables converge almost surely to the expected value.

> **Theorem 3.1** (Strong Law of Large Numbers). Let $(X_n)_{n \geq 1}$ be a sequence of independent and identically distributed (i.i.d.) random variables with $\mathbb{E}\left[|X_1|^4\right] < \infty$ and $\mathbb{E}\left[X_1\right] = \mu$. Then,
>
> $$\overline{X}_n \xrightarrow{a.s.} \mu,$$
>
> where $\overline{X}_n \equiv \dfrac{1}{n} \sum_{i=1}^{n} X_i$ is the sample mean.

The proof of this theorem is a little difficult.

3.2 Weak Law of Large Numbers

The **Weak Law of Large Numbers (WLLN)** states that the average of a large number of i.i.d. random variables converge in probability to the expected value.

> **Theorem 3.2** (Weak Law of Large Numbers). Let $(X_n)_{n \geq 1}$ be a sequence of independent and identically distributed (i.i.d.) random variables with $\mathbb{E}\left[|X_1|^2\right] < \infty$. Then,
>
> $$\overline{X}_n \xrightarrow{p} \mu,$$
>
> where $\overline{X}_n \equiv \dfrac{1}{n} \sum_{i=1}^{n} X_i$ is the sample mean..

Proof. We show that for all $\varepsilon > 0$, the following equality holds:

$$\lim_{n \to \infty} \mathbb{P}\left(\{|X_n - \mu| > \varepsilon\}\right) = 0,$$

From the assumption, we have

$$\mathbb{E}\left[\overline{X}_n\right] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[X_i\right] = \frac{1}{n}n\mu = \mu;$$

$$\mathbb{V}\left[\overline{X}_n\right] = \mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n}X_i\right] = \frac{1}{n^2}\mathbb{V}\left[\sum_{i=1}^{n}X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}\left[X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}n\sigma^2 = \frac{1}{n}\sigma^2.$$

Substituting these into the Chebyshev's inequality yields

$$\mathbb{P}\left(\left|\overline{X}_n - \mathbb{E}\left[\overline{X}_n\right]\right| \geq \varepsilon\right) \leq \frac{\mathbb{V}\left[X\right]}{\varepsilon^2} \iff \mathbb{P}\left(\left|\overline{X}_n - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

Therefore, taking a limit with respect to n results in

$$\lim_{n \to \infty} \mathbb{P}\left(\{|X_n - \mu| > \varepsilon\}\right) = 0,$$

which implies that $\overline{X}_n \xrightarrow{p} \mu$. $\qquad\square$

4   Method of Moments (MM)

In this section, we first review and derive the estimator obtained from method of moments (MM). Then, the properties of the estimator are stated with its derivation. For deriving the MM estimator for the following regression model:

$$y_i = x_i \beta + u_i,$$

where for $i \in \{1, \ldots, n\}$,

$$x_i = (x_{i1}, x_{i2}, \ldots, x_{ik}) \in \mathbb{R}^{1 \times k}; \quad \beta = (\beta_1, \beta_2, \ldots, \beta_k)' \in \mathbb{R}^{k \times 1}; \quad y_i \in \mathbb{R}; \quad u_i \in \mathbb{R},$$

we usually assume the following condition called **orthogonality condition**.

Assumption 4.1. We assume

$$\mathbb{E}[Xu] = \mathbf{0}, \tag{4.1}$$

where

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^{n \times k}; \quad u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_k \end{pmatrix} \in \mathbb{R}^{k \times 1}. \tag{4.2}$$

Then the estimator for fixed effect model is given as follows.

Theorem 4.1 (Method of Moments Estimator)**.** The estimator of MM is given by

$$\hat{\beta}_{MM} = \left(\sum_{i=1}^{n} x_i' x_i \right)^{-1} \left(\sum_{i=1}^{n} x_i' y_i \right) = (X'X)^{-1} X'y. \tag{4.3}$$

The next subsection shows how to derive the above estimator.

4.1   Estimator: Derivation

From the **Law of Large Numbers (LLN)**, we have

$$\frac{1}{n} \sum_{i=1}^{n} x_i' u_i = \frac{1}{n} \sum_{i=1}^{n} x_i'(y_i - x_i \beta) \xrightarrow{p} \mathbb{E}[x_i u_i] = \mathbf{0}.$$

Thus, the MM estimator of β, deonoted by $\hat{\beta}_{MM}$, satisfies

$$\frac{1}{n} \sum_{i=1}^{n} x_i'(y_i - x_i \beta) = \mathbf{0}.$$

Therefore, β_{MM} is given by Eq. (4.3). $\square$

Note that $\hat{\beta}_{MM}$ is equivalent to OLSE and MLE, which means that $\hat{\beta}_{MM}$ has the same properties of OLSE and MLE under some additional assumptions.

4.2 Estimator: Properties

Here we show some properties of $\hat{\beta}_{MM}$.

Theorem 4.2 (Properties of the MM Estimator). Under the assumptions we set above and some additional appropriate assumptions, the estimator $\hat{\beta}_{MM}$ obtained above has the following properties.

(i) **Unbiasedness** $\hat{\beta}_{MM}$ becomes an unbiased estimator:

$$\mathbb{E}[\hat{\beta}_{MM}] = \beta. \tag{4.4}$$

(ii) **Consistency** $\hat{\beta}_{MM} = (X'X)^{-1} X'y$ satisfies

$$\hat{\beta}_{MM} \xrightarrow{p} \beta \quad \text{or} \quad \operatorname*{plim}_{n \to \infty} \hat{\beta}_{MM} = \beta. \tag{4.5}$$

(iii) **Efficiency** The variance of the MM estimator is the minimum one in the class of linear unbiased estimator.

Proof. We can prove the above properties directly as follows.

(i) **Unbiasedness** From the fact:

$$\hat{\beta}_{MM} = (X'X)^{-1} X'y = \beta + (X'X)^{-1} X'u, \tag{4.6}$$

under the assumption that $\underline{\mathbb{E}[u|X] = \mathbf{0}}$,

$$\begin{aligned}
\mathbb{E}[\hat{\beta}_{MM}|X] &= \mathbb{E}[\beta + (X'X)^{-1} Xu|X] \\
&= \beta + (X'X)^{-1} X \underbrace{\mathbb{E}[u|X]}_{=\mathbf{0}} \\
&= \beta,
\end{aligned}$$

which yields

$$\mathbb{E}[\hat{\beta}_{MM}] = \mathbb{E}[\mathbb{E}[\hat{\beta}_{MM}|X]] = \beta. \tag{4.7}$$

(ii) **Consistency** From Eq. (4.6), we have

$$\hat{\beta}_{MM} = \beta + \left(\frac{1}{n} \sum_{i=1}^{n} x_i' x_i \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} x_i' u_i \right).$$

where

$$X := \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^{n \times k}; \quad u := \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \in \mathbb{R}^{k \times 1}, \tag{4.8}$$

and $x_i \in \mathbb{R}^{1 \times k}$ for all $i \in \{1, \dots Tn\}$. By taking the probability limit on both sides, we have

$$\operatorname*{plim}_{n \to \infty} \beta_{MM} = \beta + \operatorname*{plim}_{n \to \infty} \left(\frac{1}{n} \sum_{i=1}^{n} x_i' x_i \right)^{-1} \operatorname*{plim}_{n \to \infty} \left(\frac{1}{n} \sum_{i=1}^{n} x_i' u_i \right). \tag{4.9}$$

Here we apply the **convergence of the product of random variables in probability**. If we assume $\mathbb{E}\left[x_i' x_i\right] < \infty$ for all $i \in \{1, \dots, n\}$, from the **weak law of large numbers (WLLN)**,

$$\frac{1}{n}\sum_{i=1}^{n} x_i' x_i \xrightarrow{p} \mathbb{E}\left[x_i' x_i\right] < \infty; \tag{4.10}$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i' u_i \xrightarrow{p} \mathbb{E}\left[x_i' u_i\right] = \mathbf{0}(\in \mathbb{R}^{k \times 1}). \tag{4.11}$$

Eq. (4.11) holds from the **orthogonal condition** with respect to X and u: $\mathbb{E}[u|X] = \mathbf{0}$. In addition,

$$\operatorname*{plim}_{n \to \infty} \left(\frac{1}{n}\sum_{i=1}^{n} x_i' x_i\right)^{-1} = \mathbb{E}\left[x_i' x_i\right]^{-1} \tag{4.12}$$

holds from the **continuous mapping theorem**. Thus, substituting Eq. (4.10) and Eq. (4.11) into Eq. (4.9) results in

$$\operatorname*{plim}_{n \to \infty} \hat{\beta}_{MM} = \beta + \mathbb{E}\left[x_i' x_i\right]^{-1} \mathbf{0} = \beta,$$

which indicates that $\hat{\beta}_{MM} \xrightarrow{p} \beta$.

(iii) **Efficiency** As for the efficiency of the MM (or OLS) estimator, the **Gauss–Markov theorem** for a multiple regression model supports the efficiency.

$\square$

Appendix

A Riemann Integral

Cauchy, A. (1821) firstly introduced the notion of "continuity of a function." He subsequently defines the definite integral in several years later (1823) as follows.

Definition A.1 (Riemann Integral). Consider the following summation of a continuous function f defined over $I := [a, b] \subset \mathbb{R}$:

$$\sum_{j=1}^{n} f(x_{j-1})(x_j - x_{j-1}),$$

Then, taking the limit with respect to the $\max_{j=1}^{n}(x_j - x_{j-1})$, the above equation converges to a value, which is defined by $\int_a^b f(x)\mathrm{d}x$.

Here we give you the (more detailed) definition of the upper and lower Riemann integral.

Definition A.2 (Upper and Lower Riemann Integral). Let $[a, b] \subset \mathbb{R}$ be a given interval. By a **partition** P of $[a, b]$, we mean a finite set of points $x_0, x_1, \ldots, x_n$, where

$$a = x_0 \leq x_1 \leq \cdots \leq x_{n-1} \leq x_n = b.$$

Remind that we write

$$\Delta x_i := x_i - x_{i-1}.$$

Now suppose that f is a bounded real function defined on $[a, b]$. Let us denote the set of all partitions by Π. Corresponding ot each partition P of $[a, b]$ we put

$$M_i := \sup f(x) \quad (x_{i-1} \leq x \leq x_i),$$
$$m_i := \inf f(x) \quad (x_{i-1} \leq x \leq x_i),$$
$$U(P, f) := \sum_{i=1}^{n} M_i \Delta x_i;$$
$$L(P, f) := \sum_{i=1}^{n} m_i \Delta x_i,$$

and finally

$$\overline{\int_a^b} f\mathrm{d}x := \inf_{P \in \Pi} U(P, f), \tag{A.1}$$

$$\underline{\int_a^b} f\mathrm{d}x := \sup_{P \in \Pi} L(P, f), \tag{A.2}$$

where the inf and sup are again taken over all partitions P of $[a, b]$. The left members of Eq. (B.1) and (B.2) called the **upper and lower Riemann integral** of f over $[a, b]$, respectively.

If the upper and lower integrals are equal, i.e.,

$$\overline{\int_a^b} f \mathrm{d}x := \inf_{P \in \Pi} U(P, f) = \underline{\int_a^b} f \mathrm{d}x := \sup_{P \in \Pi} L(P, f),$$

then we say that f is **Riemann integrable** on $[a, b]$ and denote the common value of Eq. (B.1) and (B.2) by

$$\int_a^b f \mathrm{d}x, \tag{A.3}$$

or

$$\int_a^b f(x) \mathrm{d}x. \tag{A.4}$$

This is the **Riemann integral** of f over $[a, b]$.

B   Stieltjes Integral

A more general version of the (upper and lower) Riemann integral is stated as Stieltjes Integral described below.

Definition B.1 (Stieltjes Integral). Let α be a monotonically increasing function on $[a, b] \in \mathbb{R}$ (since $\alpha(a)$ and $\alpha(b)$ are finite, it follows that α is bounded on $[a, b]$.). Corresponding to each **partition** P of $[a, b]$, we write

$$\Delta \alpha_i := \alpha(x_i) - \alpha(x_{i-1}).$$

It is clear that $\Delta \alpha_i \geq 0$. For any real function f which is bounded on $[a, b]$, we put

$$M_i := \sup f(x) \quad (x_{i-1} \leq x \leq x_i),$$
$$m_i := \inf f(x) \quad (x_{i-1} \leq x \leq x_i),$$
$$U(P, f) := \sum_{i=1}^n M_i \Delta \alpha_i;$$
$$L(P, f) := \sum_{i=1}^n m_i \Delta \alpha_i,$$

and we define

$$\overline{\int_a^b} f \mathrm{d}x := \inf_{P \in \Pi} U(P, f), \tag{B.1}$$

$$\underline{\int_a^b} f \mathrm{d}x := \sup_{P \in \Pi} L(P, f), \tag{B.2}$$

where the inf and sup are taken over all partitions P of $[a, b]$.

If the upper and lower integrals are equal, i.e.,

$$\overline{\int_a^b} f \mathrm{d}x := \inf_{P \in \Pi} U(P, f) = \underline{\int_a^b} f \mathrm{d}x := \sup_{P \in \Pi} L(P, f),$$

then we denote their common value of Eq. (B.1) and (B.2) by

$$\int_a^b f \mathrm{d}\alpha, \tag{B.3}$$

or somtimes by

$$\int_a^b f(x)\mathrm{d}\alpha(x). \tag{B.4}$$

This is the **(Riemann–) Stieltjes integral** of f with respect to α over $[a, b]$.

    If Eq. (B.3) exists, i.e., if Eq. (B.1) and (B.2) are equal, we say that f is integrable with respect to α in the Riemann sense.

References

[1] Greene, W. H., *Econometric analysis Seventh Edition*. 2012, Pearson.

[2] Rudin, W., *Principles of mathematical analysis THIRD EDITION*, 1976, McGraw-hill New York.