Econ 508B: Lecture 4

Modes of Convergence and Important Inequalities

Hongyi Liu

Washington University in St. Louis

August 6, 2018

Outline

Outline

Motivation

Recall the convergence of numbers: If one has a sequence $x_1, x_2, x_3, ... \in \mathbb{R}$ of real numbers x_n, it is unambiguous what it means for what a sequence converges to a limit $x \in \mathbb{R}$.

Definition 1.1

$\forall \epsilon > 0$, $\exists$ an N such that $|x_n - x| \leq \epsilon$ for all $n > N$.

More generally, if one has a sequence of d-dimensional vectors v_n in a real vector space $\mathbb{R}^d$, it follows with a similar definition for that sequence to converge to a limit $v \in \mathbb{R}^d$ with norm $|| \cdot ||$.

Motivation

Recall the convergence of numbers: If one has a sequence $x_1, x_2, x_3, ... \in \mathbb{R}$ of real numbers x_n, it is unambiguous what it means for what a sequence converges to a limit $x \in \mathbb{R}$.

Definition 1.1

$\forall \epsilon > 0$, $\exists$ an N such that $|x_n - x| \leq \epsilon$ for all $n > N$.

More generally, if one has a sequence of d-dimensional vectors v_n in a real vector space $\mathbb{R}^d$, it follows with a similar definition for that sequence to converge to a limit $v \in \mathbb{R}^d$ with norm $||\cdot||$.

Question: what about convergence of a sequence of random variables $\{X_1, X_2, ...\}$*, or in general, a sequence of measurable functions* $\{f_1, f_2, ...\}$

What different types of convergence are there?

Two basic *modes of convergence* shown in the undergraduate course:

Definition 1.2 (Pointwise convergence)

$\{f_n\}_{n\geq 1}$ converges to f *pointwise* if, for every $x \in X$, $\lim_{n\to\infty} f_n(x) = f(x)$. In other words, for every $\epsilon > 0$ and $x \in X$, $\exists\ N$(that depends on **both** ϵ and x) such that $|f_n(x) - f(x)| \leq \epsilon$ whenever $n \geq N$.

Definition 1.3 (Uniform convergence)

$\{f_n\}_{n\geq 1}$ converges to f *uniformly* if, for every $\epsilon > 0$, $\exists\ N$ such that for every $n > N$, $|f_n(x) - f(x)| \leq \epsilon$ for every $x \in X$. In other words, $\{f_n\}_{n\geq 1}$ converges to f *uniformly* if

$$\lim_{k\to\infty} \sup\{|f_n(x) - f(x)| : x \in X\} = 0$$

Comment

The difference between *uniform convergence* and *pointwise convergence* is that with the former one, N does not depend on x, but must instead be chosen uniformly in x.

Comment

The difference between *uniform convergence* and *pointwise convergence* is that with the former one, N does not depend on x, but must instead be chosen uniformly in x.

Uniform convergence implies *pointwise convergence*, but not conversely.

Example 1.1

Let $f_n : \mathbb{R} \to \mathbb{R}$ defined by $f_n(x) := x/n$ converge pointwise to $f(x) := 0$, but not uniformly if we pick $x = n$ and $\epsilon = 1/2$ while $|f_n(x) - f(x)| = 1 > 1/2$.

More notions

To discuss some modes of convergence that arise from measure theory, the domain space X is equipped with the triplet, measure space $(X, \mathcal{F}, \mu)$, and the functions f_n (and their limit f) are measurable with respect to this space. In this context, we have the following additional *modes of convergence*:

Definition 1.4 (Pointwise Almost Everywhere Convergence)

$\{f_n\}_{n \geq 1}$ converges to f *pointwise almost everywhere* if, for μ- almost everywhere $x \in X$, $f_n(x)$ converges to $f(x)$, denoted by $f_n \to f$ a.e. μ. In other words, if $\exists$ a set Y in $\mathcal{F}$ such that $\mu(X \setminus Y) = 0$ and $\lim_{n \to \infty} f_n(x) = f(x)$ for all $x \in Y$.

Remark 1.1

In probability theory, when f_n and f are random variables, *pointwise convergence almost everywhere* is called **almost sure convergence**.

Definition 1.5 (Convergence in $L^\infty(\mu)$)

$\{f_n\}_{n\geq 1}$ converges to f *in* $L^\infty(\mu)$ if

$$\lim_{n\to\infty} ||f_n - f||_\infty = 0,$$

where for any $\mathcal{F}$-measurable function g on $(\Omega, \mathcal{F}, \mu)$,

$$||g||_\infty = \inf\{K : K \in (0, \infty), \mu(\{|g| > K\}) = 0\}.$$

Definition 1.5 (Convergence in $L^\infty(\mu)$)

$\{f_n\}_{n\geq 1}$ converges to f *in* $L^\infty(\mu)$ if

$$\lim_{n\to\infty} ||f_n - f||_\infty = 0,$$

where for any $\mathcal{F}$-measurable function g on $(\Omega, \mathcal{F}, \mu)$,

$$||g||_\infty = \inf\{K : K \in (0, \infty), \mu(\{|g| > K\}) = 0\}.$$

Definition 1.6 (almost uniformly)

$\{f_n\}_{n\geq 1}$ converges to f *almost uniformly* if, for every $\epsilon > 0$, $\exists$ a set $A \in \mathcal{F}$ such that $\mu(A) < \epsilon$ and on A^c,$f_n \to f$ uniformly, i.e., $\sup\{|f_n(x) - f(x)| : x \in A^c\} \to 0$ as $n \to \infty$.

Convergence in $L^p(\mu)$

Definition 1.7 (Convergence in $L^p(\mu)$)

Let $0 < p < \infty$. Then, $\{f_n\}_{n\geq 1}$ *converges to* f *in* $L^p(\mu)$, denoted by $f_n \xrightarrow{L^p} f$, if $\int |f_n|^p d\mu < \infty$ for all $n \geq 1$, $\int |f|^p d\mu < \infty$ and

$$\lim_{n\to\infty} \int |f_n - f|^p d\mu = 0.$$

Convergence in $L^p(\mu)$

Definition 1.7 (Convergence in $L^p(\mu)$)

Let $0 < p < \infty$. Then, $\{f_n\}_{n\geq 1}$ *converges to f in $L^p(\mu)$*, denoted by $f_n \xrightarrow{L^p} f$, if $\int |f_n|^p d\mu < \infty$ for all $n \geq 1$, $\int |f|^p d\mu < \infty$ and

$$\lim_{n\to\infty} \int |f_n - f|^p d\mu = 0.$$

Remark 1.2

if $f \in L^p$, then $(\int |f|^p)^{1/p} < \infty$; e.g. if $f \in L^1$, then $\int |f| d\mu < \infty$. *Convergence in L^p* is equivalent to $||f_n - f||_p \to 0$ as $n \to \infty$, where for any $\mathcal{F}$-measurable function g and $0 < p < \infty$, $||g||_p = (\int |g|^p d\mu)^{\min\{\frac{1}{p},1\}}$

- For $p = 1$, when f_n and f are random variables, *convergence in L^1* is called *convergence in mean.*
- For $p = 2$, this is referred to as *convergence in mean square.*

Convergence in measure

Definition 1.8 (Convergence in measure)

$\{f_n\}_{n\geq 1}$ *converges to* f *in measure*, denoted by $f_n \xrightarrow{m} f$, if for each $\epsilon > 0$,

$$\lim_{n\to\infty} \mu(\{|f_n - f| > \epsilon\}) = 0.$$

when f_n and f are random variables, *convergence in measure* is often referred to as **convergence in probability** such as $\lim_{n\to\infty} P(\{|X_n - X| > \epsilon\}) = 0.$

- Another mode of convergence is *convergence in distribution*, we will cover it later on the lecture of asymptotic theory.
- Students or readers are highly encouraged to draw a diagram that summarize the logic between the above modes of convergence.

Outline

1. Convergence for measurable functions
2. Important inequalities

Markov's inequality

Theorem 2.1

Let f be a non-negative measurable functions on a measure space $(\Omega, \mathcal{F}, \mu)$. Then for any $0 < t < \infty$,

$$\mu(\{f \geq t\}) \leq \frac{\int f d\mu}{t}$$

Proof: Let f be non-negative,

$$\int f d\mu \geq \int_{(\{f \geq t\})} f d\mu \geq \int_{(\{f \geq t\})} t d\mu = t\mu(\{f \geq t\})$$

Applications

Corollary 2.1

Let X be a random variable on a probability space $(\Omega, \mathcal{F}, P)$. For $r > 0, t > 0$,

$$P(|X| \geq t) \leq \frac{E|X|^r}{t^r}$$

When $r = 2$, it is called *Chebyshev's inequality*: $P(|X - E(X)| \geq t) \leq \frac{Var(X)}{t^2}$.

Proof: $\{|X| \geq t\} = \{|X|^r \geq t^r\}$ for all $t > 0, r > 0$.

Example 2.1

Suppose $E(|X_n - c|) \to 0$ for some finite c. By *Markov's inequality*,

$$P(|X_n - c| > \epsilon) \leq \frac{E(|X_n - c|)}{\epsilon} \to 0.$$

Thus, X_n converges in probability to c.

Applications

Corollary 2.2

Let $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ be non-decreasing. For any random variable X and $0 < t < \infty$,

$$P(|X| \geq t) \leq \frac{E\phi(|X|)}{\phi(t)}$$

Proof: Use *Markov's inequality* and the fact that $P(|X| \geq t) = P(\phi(|X|) \geq \phi(t))$.

Example 2.2

For i.i.d Bernoulli trials with success probability $p = 1/2$, let T_n denote the number of times in the first n trials that a success is followed by a failure.

- Denote $I_i = I\{i\text{th trial is a success and } (i+1)\text{st trial is a failure}\}$, $T_n = \sum_{i=1}^{n-1} I_i$, then prove that $\frac{T_n}{n} \xrightarrow{p} \frac{1}{4}$. (Homework)

Jensen's inequality

Theorem 2.2

Let f be a measurable function on a probability space $(\Omega, \mathcal{F}, P)$ with $P(f \in (a,b)) = 1$ for some interval $(a,b), -\infty < b \leq \infty$ and let $\phi : (a,b) \to \mathbb{R}$ be convex. Then

$$\phi(\int f dP) \leq \int \phi(f) dP,$$

provided $\int |f| dP < \infty$ and $\int |\phi(f)| dP < \infty$.

Corollary 2.3

In terms of random variables X, we have $\phi(EX) \leq E\phi(X)$, provided $E|X| < \infty$ and $E|\phi(X)| < \infty$.

If ϕ is a concave function, then the *Jensen's inequality* just change the direction of inequality.

Application of Jensen's inequality

Let $k \geq 1$ be an integer.

(1) Let $x_1, x_2, ..., x_k$ be real and $p_1, p_2, ..., p_k$ be positive numbers such that $\sum_{i=1}^{k} p_i = 1$. Then

$$\sum_{i=1}^{k} p_i \exp(x_i) \geq \exp(\sum_{i=1}^{k} p_i x_i).$$

(2) Let $y_1, y_2, ..., y_k$ be non-negative numbers and $p_1, p_2, ..., p_k$ be as in (1). Then

$$\sum_{i=1}^{k} p_i y_i \geq \prod_{i=1}^{k} {y_i}^{p_i},$$

In particular, when $p_i = 1/k$ for $i = 1, ..., k$.

$$\sum_{i=1}^{k} \frac{1}{k} y_i \geq \prod_{i=1}^{k} {y_i}^{\frac{1}{k}}$$

Example 2.3

Let $X \in S \subset \mathbb{R}$ be a discrete random variable. Let p be the p.d.f of X and q be any other p.d.f. defined on S. Consider the likelihood functions $p(X)$ and $q(X)$. Noting that these are random variables, we have that

$$E[\log \frac{p(X)}{q(X)}] \geq 0$$

Example 2.3

Let $X \in S \subset \mathbb{R}$ be a discrete random variable. Let p be the p.d.f of X and q be any other p.d.f. defined on S. Consider the likelihood functions $p(X)$ and $q(X)$. Noting that these are random variables, we have that

$$E[\log \frac{p(X)}{q(X)}] \geq 0$$

Proof: W.L.O.G. assume that $p(x) > 0, q(x) > 0$ for all $x \in S$ (why?).

$$\begin{aligned}
E[\log \frac{p(X)}{q(X)}] &= \sum_{x \in S} p(x) \log \frac{p(x)}{q(x)} = -\sum_{x \in S} p(x) \log \frac{q(x)}{p(x)} \\
&= -E[\log \frac{q(X)}{p(X)}] \geq -\log E[\frac{q(X)}{p(X)}] \qquad \text{(Jensen)} \\
&= -\log[\sum_{x \in S} q(x)] = -\log(1) = 0
\end{aligned}$$

Hölder inequality

- Hölder's inequality is an important tool relating the expected value of a product to the product of expected values.

Theorem 2.3

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let $1 < p < \infty$, $f \in L^p(\Omega, \mathcal{F}, \mu)$ and $g \in L^q(\Omega, \mathcal{F}, \mu)$, where $1/p + 1/q = 1$. Then

$$\int |fg| d\mu \leq (\int |f|^p d\mu)^{1/p} (\int |g|^q d\mu)^{1/q},$$

i.e., $||fg||_1 \leq ||f||_p ||g||_q$. If $||fg||_1 \neq 0$, then equality holds if.f. $|f|^p = c|g|^q$ for some $c \in (0, \infty)$.

Question: How to prove it? Young's inequality.

Young's inequality

Lemma 2.1

Let $a, b \geq 0$, and $1 < p, q < \infty$ be such that $1/p + 1/q = 1$. Then

$$\frac{1}{p}a^p + \frac{1}{q}b^q \geq ab.$$

with equality if.f. $a^p = b^q$.

Proof: W.L.O.G., assume that $a, b > 0$. Then

$$\begin{aligned} ab &= \exp(\log a + \log b) = \exp(\frac{1}{p}\log a^p + \frac{1}{q}\log b^q) \\ &\leq \frac{1}{p}\exp(\log a^p) + \frac{1}{q}\exp(\log b^q) && \text{(Jensen)} \\ &= \frac{1}{p}a^p + \frac{1}{q}b^q \end{aligned}$$

Proof of Hölder's inequality

Consider X, Y are random variables such that $E|X|^p, E|Y|^q < \infty$. Then the target is to prove

$$|E(X,Y)| \leq E|XY| \leq ||X||_p||Y||_q$$

Define

$$a = \frac{|X|}{||X||_p}, b = \frac{|Y|}{||Y||_q}$$

Then applying *Young's inequality* and taking expectations on both sides, we obtain

$$1 = \frac{1}{p} + \frac{1}{q} \geq \frac{E|XY|}{||X||_p||Y||_q}$$

When $p = q = 2$, it becomes *Cauchy-Schwarz's inequality*.

Minkowski's inequality

Theorem 2.4

Let $1 \leq p \leq \infty$ and $f, g \in L^p$. Then

$$(\int |f+g|^p d\mu)^{1/p} \leq (\int |f|^p d\mu)^{1/p} + (\int |g|^p d\mu)^{1/p}$$

i.e.

$$||f+g||_p \leq ||f||_p + ||g||_p.$$

Corollary 2.4

Given random variables X, Y and $p \geq 1$,

$$(E|X+Y|^p)^{1/p} \leq (E|X|^p)^{1/p} + (E|Y|^p)^{1/p}$$