STAT 810 Probability Theory I

Chapter 6: Convergence Concepts

Dr. Dewei Wang
Associate Professor
Department of Statistics
University of South Carolina
deweiwang@stat.sc.edu

On $(\Omega, \mathcal{B}, P)$, we say that a statement about random elements holds **almost surely** (a.s./a.e./a.c./a.a.) if there exists an event $N \in \mathcal{B}$ with $P(N) = 0$ such that the statement holds if $\omega \in N^c$:

- $X = X'$ a.s. means $P(X = X') = 1$; i.e., there exists $N \in \mathcal{B}$, such that $X(\omega) = X'(\omega)$ for $\omega \in N^c$ and $P(N) = 0$.
- $X \leq X'$ a.s. means there exists $N \in \mathcal{B}$, such that $X(\omega) \leq X'(\omega)$ for $\omega \in N^c$ and $P(N) = 0$.
- $\lim_{n\to\infty} X_n$ exists a.s. means there exists $N \in \mathcal{B}$, such that $\lim_{n\to\infty} X_n(\omega)$ exists for $\omega \in N^c$ and $P(N) = 0$.

Most probabilistic properties of random variables are invariant under the relation almost sure equality. For example, if $X = X'$ a.s. then $X \in L_1$ iff $X' \in L_1$ and in this case $E(X) = E(X')$.

Example 6.1.1

Consider $([0,1], \mathcal{B}([0,1]), \lambda)$, where λ is Lebesgue measure. Define

$$X_n(s) = \begin{cases} n & \text{if } 0 \leq s \leq n^{-1}, \\ 0 & \text{if } n^{-1} < s \leq 1. \end{cases}$$

Let $N = \{0\}$, we see that for $s \notin N$, $X_n(s) \to 0$ and $\lambda(N) = 0$. Thus X_n converges to 0 almost surely. Note that N is not empty.

Proposition 6.1.1

Let $\{X_n\}$ be iid random variables with common distribution function $F(x)$. Assume $F(x) < 1$ for all x. Set $M_n = \bigvee_{i=1}^{n} X_i$. Then $M_n \uparrow \infty$ a.s.

Proof: By definition, we know $\{M_n(\omega)\}$ is monotone increasing. Let

$$
\begin{aligned}
N^c &= \{\omega : \lim_{n\to\infty} M_n(\omega) = \infty\} \\
&= \{\omega : \forall j, \exists k(\omega, j), \forall n \geq k(\omega, j), M_n(\omega) > j\} \\
&= \cap_j \left(\cup_{k\geq 1} \cap_{n\geq k} [M_n \geq j]\right) = \cap_j \liminf_{n\to\infty} [M_n > j].
\end{aligned}
$$

Thus $N = \cup_j (\limsup_{n\to\infty}[M_n \leq j])$. Because $\sum_n P[M_n \leq j] = \sum_n F^n(j) < \infty$, by Borel-Cantelli lemma, $P(\limsup_{n\to\infty}[M_n \leq j]) = 0$. Thus $P(N) = 0$.

Suppose X_n, $n \geq 1$, and X are random variables. Then $\{X_n\}$ **converges in probability** (i.p.) to X, written $X_n \xrightarrow{P} X$, if for any $\epsilon > 0$,

$$\lim_{n\to\infty} P[|X_n - X| > \epsilon] = 0.$$

Example 6.2.1

Consider $([0,1], \mathcal{B}([0,1]), \lambda)$, where λ is Lebesgue measure. Define $X_1 = I_{[0,1]}$, $X_2 = I_{[0,1/2]}$, $X_3 = I_{[1/2,1]}$, $X_4 = I_{[0,1/3]}$, $X_5 = I_{[1/3,2/3]}$, $X_6 = I_{[2/3,1]}$, $\dots$.

Then $P[|X_n - 0| > \epsilon] = 1, 1/2, 1/2, 1/3, 1/3, 1/3, 1/4, 1/4, \cdots \to 0$. Thus $X_n \xrightarrow{P} 0$. However, for any $s \in [0,1]$, $X_n(s) = 1$ or 0 for infinitely many values of n. Thus $\lim_{n\to\infty} X_n$ does not exist almost surely.

Theorem 6.2.1 (Convergence a.s. imples convergence i.p.)
Suppose $\{X_n\}$ are random variables on $(\Omega, \mathcal{B}, P)$. If $X_n \stackrel{a.s.}{\to} X$, then $X_n \stackrel{P}{\to} X$.

Proof: Let $N^c = \{\omega : \lim_{n\to\infty} |X_n(\omega) - X(\omega)| = 0\} = \{\omega : \forall j > 0, \exists N(\epsilon), \forall n \geq N, |X_n(\omega) - X(\omega)| \leq 1/j\} = \cap_j \cup_{N\geq 1} \cap_{n \geq N} [|X_n - X| \leq j^{-1}]$. Then $N = \cup_j \limsup_{n\to\infty} [|X_n - X| > j^{-1}]$. $P(N) = 0$ since $X_n \stackrel{a.s.}{\to} X$. Thus, for any j,

$$\begin{aligned} 0 &= P(\limsup_{n\to\infty} [|X_n - X| > j^{-1}]) \\ &= \lim_{N\to\infty} P(\cup_{n\geq N} [|X_n - X| > j^{-1}]) \\ &\geq \lim_{n\to\infty} P[|X_n - X| > j^{-1}]. \end{aligned}$$

Pick j such that $\epsilon > j^{-1}$, then we have $\lim_{n\to\infty} P[|X_n - X| > \epsilon] = 0$.

6.2.1 Statistical Terminology

In statistical estimation theory, almost sure and in probability convergence have analogues as **strong** or **weak consistency**.

Given a family of probability models $(\Omega, \mathcal{B}, P)$. Suppose the statistician gets to observe random variables $X_1, \ldots, X_n$, defined on Ω and based on these observations must decide which is the correct model; that is, which is the correct value of θ. Statistical **estimation** means: select the correct model.

6.2.1 Statistical Terminology

Suppose $\Omega = \mathbb{R}^\infty$, $\mathcal{B} = \mathcal{B}(\mathbb{R}^\infty)$. Let $\omega = (x_1, x_2, \dots)$ and define $X_n(\omega) = x_n$. For each $\theta \in \mathbb{R}$, let P_θ be the product measure on $\mathbb{R}^\infty$ which makes $\{X_n\}$ iid with common $N(\theta, 1)$ distribution. Based on observing $X_1, \dots, X_n$, one estimates θ with an appropriate function of the observations

$$\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n).$$

$\hat{\theta}_n(X_1, \dots, X_n)$ is called a **statistic** and is also an **estimator** (a random element). When one actually does the experiment and observes, $X_1 = x_1, \dots, X_n = x_n$, then $\hat{\theta}_n(x_1, \dots, x_n)$ is called the **estimate** (a number of a vector). In this example, we often take $\hat{\theta}_n = \sum_{i=1}^n X_i/n$. The estimator $\hat{\theta}_n$ is **weakly consistent** (denoted by $\hat{\theta}_n \overset{P_\theta}{\to} \theta$) if for all $\theta \in \Theta$,

$$P_\theta[|\hat{\theta}_n - \theta| > \epsilon] \to 0, n \to \infty.$$

We say $\hat{\theta}_n$ is **strongly consistent** if for all $\theta \in \Theta$, $\hat{\theta}_n \to \theta$, $P_\theta - a.s.$

Theorem 6.3.1 Relations between i.p. and a.s. convergence

Suppose that $\{X_n\}$ and X are real-valued random variables

(a) **Cauchy criterion:** $\{X_n\}$ converges in probability to X iff $\{X_n\}$ is Cauchy in probability; i.e., $X_n - X_m \xrightarrow{P} 0$, as $n, m \to \infty$; i.e., for any $\epsilon > 0$, $\delta > 0$, there exists $n_0 = n_0(\epsilon, \delta)$, such that for all $r, s \geq n_0$, we have $P[|X_r - X_s| > \epsilon] < \delta$.

(b) $X_n \xrightarrow{P} X$ iff each subsequence $\{X_{n_k}\}$ contains a further subsequence $\{X_{n_{k(t)}}\}$ which converges almost surely to X.

Proof of Theorem 6.3.1: We approach (a) with 2 steps: (i) We first show that if $X_n \xrightarrow{P} X$ then $\{X_n\}$ is Cauchy i.p. This can be done easily by using the inequality $P[|X_r - X_s| > \epsilon] \leq P[|X_r - X| > \epsilon/2] + P[|X_s - X| > \epsilon/2]$, which comes from the use of the triangle inequality.

(ii) Next, we prove if $\{X_n\}$ is Cauchy i.p., then there exists a subsequence $\{X_{n_j}\}$ which converges almost surely. Call the almost sure limit X. Then it is also true that $X_n \xrightarrow{P} X$.
To prove this, we define n_j by $n_1 = 1$ and

$$n_j = \inf\{N > n_{j-1} : P[|X_r - X_s| > 2^{-j}] < 2^{-j} \text{ for all } r, s \geq N\}.$$

By Cauchy i.p, n_j always exists by setting $\epsilon = \delta = 2^{-j}$ and $n_j > n_{j-1}$. Consequently, we have $\sum_{j=1}^{\infty} P[|X_{n_{j+1}} - X_{n_j}| > 2^{-j}] < \infty$. By the Borel-Cantelli Lemma, let $N = limsup_{j\to\infty}[|X_{n_{j+1}} - X_{n_j}| > 2^{-j}]$, $P(N) = 0$.

Proof of Theorem 6.3.1 continued: For $\omega \in N^c = \liminf_{j\to\infty}[|X_{n_{j+1}} - X_{n_j}| \le 2^{-j}]$, we know $|X_{n_{j+1}}(\omega) - X_{n_j}(\omega)| \le 2^{-j}$ for all large j. Thus $\{X_{n_j}(\omega)\}$ is a Cauchy sequence of real numbers and consequently, $\lim_{j\to\infty} X_{n_j}(\omega)$ exists. We proved $\{X_{n_j}\}$ converges a.s. and denote the limit by X.

To show $X_n \xrightarrow{P} X$, note $P[|X_n - X| > \epsilon] \le P[|X_n - X_{n_j}| > \epsilon/2] + P[|X_{n_j} - X| > \epsilon/2]$. By the Cauchy i.p. property, for any $\delta/2 > 0$, we can find $n_0(\epsilon/2, \delta/2)$ such that when $n, n_j > n_0(\epsilon/2, \delta/2)$, $P[|X_n - X_{n_j}| > \epsilon/2] < \delta/2$.

Because $X_{n_j} \xrightarrow{P} X$ as $j \to \infty$, we can find $n_1(\epsilon/2, \delta/2)$, such that for $n_j > n_1(\epsilon/2, \delta/2)$, $P[|X_{n_j} - X| > \epsilon/2] < \delta/2$. Thus, for any $\delta > 0$, we find $n_*(\epsilon, \delta) = \max\{n_0(\epsilon/2, \delta/2), n_1(\epsilon/2, \delta/2)\}$, for $n > n_*(\epsilon, \delta)$, $P[|X_n - X| > \epsilon] < \delta$. Done (a)!

Proof of Theorem 6.3.1 continued: For (b): Suppose $X_n \stackrel{P}{\to} X$. Pick any subsequence, the subsequence also $\stackrel{P}{\to} X$. From (ii) above, we find a further subsequence converging a.s.

Conversely, Suppose every subsequence has an a.s. convergence subsequence. If X_n does not converge to X in probability. Then there exists a subsequence $\{X_{n_k}\}$ and a $\delta > 0$ and $\epsilon > 0$ such that

$$P[|X_{n_k} - X| > \epsilon] \geq \delta.$$

But we know there exists a further subsequence $\{X_{n_{k(t)}}\}$ converging a.s., hence i.p. to X. Contradiction!

Corollary 6.3.1

(i) $X_n \overset{a.s.}{\to} X$ and $g : \mathbb{R} \mapsto \mathbb{R}$ is continuous, then $g(X_n) \overset{a.s.}{\to} g(X)$.

(ii) $X_n \overset{P}{\to} X$ and $g : \mathbb{R} \mapsto \mathbb{R}$ is continuous, then $g(X_n) \overset{P}{\to} g(X)$.

Proof: (i) There exists $N \in \mathcal{B}$ with $P(N) = 0$ such that if $\omega \in N^c$, $X_n(\omega) \to X(\omega)$. By continuity, $g(X_n(\omega)) \to g(X(\omega))$ holds.
(ii) Using Theorem 6.3.1 (b).

Thus if $X_n \overset{P}{\to} X$, it is also true that $X_n^2 \overset{P}{\to} X^2$ and $\arctan X_n \overset{P}{\to} \arctan X$.

Corollary 6.3.2 (Lebesgue Dominated Convergence)

If $X_n \xrightarrow{P} X$ and if there exists a dominating random variable $\xi \in L_1$ such that $|X_n| \le \xi$, then $E(X_n) \to E(X)$.

Proof: It suffices to show every convergent subsequence of $E(X_n)$ converges to $E(X)$. Suppose $E(X_{n_k})$ converges. Then since convergence in probability is assumed, $\{X_{n_k}\}$ contains an a.s. convergent subsequence $\{X_{n_{k(t)}}\}$ such that $X_{n_{k(t)}} \xrightarrow{a.s.} X$. The Lebesgue Dominted COnvergence Theorem implies

$$E(X_{n_{k(t)}}) \to E(X).$$

Thus $E(X_{n_k}) \to E(X)$.

6.3 Connections between a.s. and i.p. Convergence

We now list several easy results related to convergence in probability (HW 6-1: prove these).

(1) If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$, then $X_n + Y_n \xrightarrow{P} X + Y$ and $X_n Y_n \xrightarrow{P} XY$.

(2) This item is a reminder that Chebychev's inequality implies the Weak Law of Large Numbers (WLLN): If $\{X_n\}$ are iid with $EX_n = \mu$ and $\mathrm{Var}(X_n) = \sigma^2$, then $\sum_{i=1}^n X_i/n \xrightarrow{P} \mu$.

(3) Bernstein's version of the Weierstrass Approximation Theorem: Let $f : [0,1] \mapsto \mathbb{R}$ and define the Bernstein polynomial of degree n by

$$B_n(x) = \sum_{k=0}^{n} f(\frac{k}{n}) C_k^n x^k (1-x)^{n-k}, 0 \le x \le 1$$

Then $B_n(x) \to f(x)$ uniformly for $x \in [0,1]$.

6.4 Quantile Estimation

Let F be a distribution function. For $0 < p < 1$, the pth order quantile of F is $F^{\leftarrow}(p) = \inf\{x : F(x) \geq p\}$. If F is unknown, we may wish to estimate a quantile.

We start with the estimation of F. Let $X_1, \ldots, X_n$ be a random sample from F ; that is, iid with common distribution function F. Define the empirical cumulative distribution function (cdf) by

$$F_n(x) = n^{-1} \sum_{i=1}^{n} I_{[X_j \leq x]},$$

which is the percentage of the sample whose value is no greater than x. It is easy to see that

$$E(F_n(x)) = F(x), \mathrm{Var}(F_n(x)) = n^{-1} F(x)(1 - F(x)), F_n(x) \xrightarrow{P} F(x),$$

for each x.

To estimate a quantile $F^{\leftarrow}(p)$, one non-parametric method uses order statistics. Rearranging $X_1, \ldots, X_n$ into $X_1^{(n)} \leq \cdots \leq X_n^{(n)}$, we define the empirical quantile function by

$$
\begin{aligned}
F_n^{\leftarrow}(p) &= \inf\{x : F_n(x) \geq p\} \\
&= \inf\{X_j^{(n)} : F_n(X_j^{(n)}) \geq p\} \\
&= \inf\{X_j^{(n)} : \frac{j}{n} \geq p\} \\
&= \inf\{X_j^{(n)} : j \geq np\} \\
&= X_{\lceil np \rceil}^{(n)},
\end{aligned}
$$

where $\lceil np \rceil$ is the ceiling of np.

Theorem 6.4.1

Suppose F is strictly increasing at $F^{\leftarrow}(p)$ which means that for all $\epsilon > 0$

$$F(F^{\leftarrow}(p)+\epsilon) > p, F(F^{\leftarrow}(p)-\epsilon) < p.$$

Then we have $X^{(n)}_{\lceil np \rceil}$ be a weakly consistent quantile estimator,

$$X^{(n)}_{\lceil np \rceil} \xrightarrow{P} F^{\leftarrow}(p).$$

Proof: It suffices to show for all $\epsilon > 0$, $P[X^{(n)}_{\lceil np \rceil} > F^{\leftarrow}(p)+\epsilon] \to 0$ and $P[X^{(n)}_{\lceil np \rceil} \leq F^{\leftarrow}(p)-\epsilon] \to 0$. We only show the latter one. Note that $X^{(n)}_{\alpha} \leq y$ iff $nF_n(y) \geq \alpha$. Thus

$$P[X^{(n)}_{\lceil np \rceil} \leq F^{\leftarrow}(p)-\epsilon] = P[nF_n(F^{\leftarrow}(p)-\epsilon) \geq \lceil np \rceil]$$

$$= P[F_n(F^{\leftarrow}(p)-\epsilon) \geq \frac{\lceil np \rceil}{n}]$$

$$= P[F_n(F^{\leftarrow}(p)-\epsilon) - F(F^{\leftarrow}(p)-\epsilon) \geq \frac{\lceil np \rceil}{n} - F(F^{\leftarrow}(p)-\epsilon)].$$

Proof continued: We also know that $\frac{\lceil np \rceil}{n} \to p$ and $2\delta \doteq p - F(F^{\leftarrow}(p) - \epsilon) > 0$. Thus, we can find $N > 0$, such that for $n \geq N$, $\frac{\lceil np \rceil}{n} - F(F^{\leftarrow}(p) - \epsilon) > \delta > 0$. Then when $n \geq N$,

$$\begin{aligned} P[F_n(F^{\leftarrow}(p) - \epsilon) - F(F^{\leftarrow}(p) - \epsilon) \geq \frac{\lceil np \rceil}{n} - F(F^{\leftarrow}(p) - \epsilon)] \\ \leq P[|F_n(F^{\leftarrow}(p) - \epsilon) - F(F^{\leftarrow}(p) - \epsilon)| \geq \delta] \to 0, \end{aligned}$$

since by the WLLN, $F_n(F^{\leftarrow}(p) - \epsilon) \xrightarrow{P} F(F^{\leftarrow}(p) - \epsilon)$.
Similarly, one can show $P[X^{(n)}_{\lceil np \rceil} > F^{\leftarrow}(p) + \epsilon] \to 0$.

6.5 L_p convergence

We say $X \in L_p$ if $E(|X|^p) < \infty$. For $X, Y \in L_p$, we define the L_p metric by

$$d(X,Y) = (E|X-Y|^p)^{1/p},$$

and the induced norm on the L_p space is

$$\|X\|_p = (E|X|^p)^{1/p}.$$

A sequence $\{X_n\}$ of random variables converges in L_p to X, written

$$X_n \xrightarrow{L_p} X,$$

if

$$\|X_n - X\|_p \to 0, \text{ or } E(|X_n - X|^p) \to 0$$

as $n \to \infty$.

The most important case is when $p = 2$, in which case L_2 is a Hilbert space with the inner product of X and Y defined by the correlation of X and Y. Here are two simple examples:

1. Define $\{X_n\}$ to be a (2nd order, weakly, covariance) stationary process if $EX_n = \mu$ independent of n and $\mathrm{Corr}(X_n, X_{n+k}) = \rho(k)$ for all n. No distributional structure is specified. The **best linear predictor** of X_{n+1} based on $X_1, \ldots, X_n$ is the linear combination of $X_1, \ldots, X_n$ which achieves **minimum mean square error** (MSE). Call this predictor $\hat{X}_{n+1}$, which is of the form $\hat{X}_{n+1} = \sum_{i=1}^{n} \alpha_i X_i$ and α_i's are chosen so that

$$E(\hat{X}_{n+1} - X_{n+1})^2 = \min_{\alpha_1, \ldots, \alpha_n} E(\sum_{i=1}^{n} \alpha_i X_i - X_{n+1})^2.$$

2. Supoose $\{X_n\}$ are iid with $E(X_n) = \mu$ and $\text{Var}(X_n) = \sigma^2$. Then

$$\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n} \xrightarrow{L_2} \mu,$$

since

$$E(\frac{\sum_{i=1}^n X_i}{n} - \mu)^2 = \frac{\text{Var}(S_n)}{n^2} = \frac{\sigma^2}{n} \to 0.$$

6.5 L_p convergence

Some basic facts:

(i). L_p convergence implies convergence in probability.

For $p > 0$, if $X_n \xrightarrow{L_p} X$, then $X_n \xrightarrow{P} X$.

This follows readily from Chebychev's inequality,

$$P[|X_n - X| \geq \epsilon] \leq \frac{E|X_n - X|^p}{\epsilon^p} \to 0.$$

(ii). Convergence in probability does not imply L_p convergence.

Example: Consider $([0,1],\mathcal{B}([0,1]),\lambda)$ and set

$$X_n = 2^n I_{(0,1/n)}.$$

Then

$$P(|X_n| > \epsilon) = 1/n \to 0.$$

However

$$E(|X_n|^p) = 2^{np}/n \to \infty.$$

(iii). L_p convergence does not imply almost sure convergence.

Example: Consider $([0,1], \mathcal{B}([0,1]), \lambda)$ and define $X_1 = I_{[0,1]}$, $X_2 = I_{[0,1/2]}$, $X_3 = I_{[1/2,1]}$, $X_4 = I_{[0,1/3]}$, $X_5 = I_{[1/3,2/3]}$, $X_6 = I_{[2/3,1]}$, $\ldots$.

Then for any $p > 0$, $E(|X_n|^p) = 1/2, 1/2, 1/3, 1/3, 1/3, 1/4, \ldots$ converges to 0. Thus $X_n \xrightarrow{L_p} 0$. But $\{X_n\}$ does not coverge almost surely to 0.

6.5.1 Uniform Integrability

Deeper and more useful connections between modes of convergence depend on the notion of uniform integrability (ui). It is a property of a family of random variables which says that the first absolute moments are uniformly bounded and the distribution tails of the random variables in the family converge to 0 at a uniform rate. We give the formal definition.

Definition

A family $\{X_t : t \in T\}$ of L_1 random variables indexed by T is **uniformly integrable** (ui) if

$$\sup_{t\in T} E\left(|X_t| I_{[|X_t|>a]}\right) = \sup_{t\in T} \int_{[|X_t|>a]} |X_t| dP \to 0$$

as $a \to \infty$; that is

$$\int_{[|X_t|>a]} |X_t| dP \to 0$$

as $a \to \infty$, uniformly in $t \in T$.

Some criteria:

(1) **Singleton**. If $T = \{1\}$ consists of one element, then

$$\int_{[|X_1|>a]} |X_1| dP \to 0, a \to \infty$$

as a consequence of $X_1 \in L_1$.

(2) **Dominated families**. If there exists a dominating random variables $Y \in L_1$, such that $|X_t| \leq Y$ for all $t \in T$. Then $\{X_t\}$ is ui.

$$\sup_{t \in T} \int_{[|X_t|>a]} |X_t| dP \leq \int_{[|Y|>a]} |Y| dp \to 0, a \to \infty$$

Some criteria:

(3) **Finite family**. Suppose $T = \{1, 2, \ldots, n\}$ is finite. Then $\{X_t : t \in T\}$ is ui. This is because

$$|X_t| \leq \sum_{i=1}^{n} |X_i| \in L_1,$$

then apply (2).

(4) **More domination**. Suppose for each $t \in T$, $X_t \in L_1$ and $Y_t \in L_1$,

$$|X_t| \leq |Y_t|$$

for all $t \in T$. Then if $\{Y_t\}$ is ui so is $\{X_t\}$ ui.

Some criteria:

(5) **Crystal Ball Condition**. For $p > 0$, the family $\{|X_n|^p\}$ is ui, if

$$\sup_n E(|X_n|^{p+\delta}) < \infty,$$

for some $\delta > 0$.

$$\begin{aligned}
\sup_n \int_{[|X_n|^p > a]} |X_n|^p dP &= \sup_n \int_{[|X_n|/a^{1/p} > 1]} |X_n|^p dP \\
&\leq \int_{[|X_n|^\delta / a^{\delta/p} > 1]} |X_n|^p dP \\
&\leq \int |X_n|^p \frac{|X_n|^\delta}{a^{\delta/p}} dP \\
&\leq a^{-\delta/p} \sup_n E(|X_n|^{p+\delta}) \to 0.
\end{aligned}$$

Theorem 6.5.1

Let $\{X_t : t \in T\}$ be L_1 random variables. This family is ui iff

(A) **Uniform absolute continuity:** For all $\epsilon > 0$, there exists $\delta = \delta(\epsilon)$, such that

$$\forall A \in \mathcal{B} : \sup_{t \in T} \int_A |X_t| dP < \epsilon \text{ if } P(A) < \delta,$$

and

(B) **Uniform bounded first absolute moments:**

$$\sup_{t \in T} E(|X_t|) < \infty.$$

Proof: Suppose $\{X_t\}$ is ui. For any $X \in L_1$, $A \in \mathcal{B}$, $a > 0$,

$$\int_A |X| dP = \int_{A\cap[|X|\le a]} |X| dP + \int_{A\cap[|X|>a]} |X| dP$$
$$\le aP(A) + \int_{[|X|>a]} |X| dP$$

Thus

$$\sup_{t\in T} \int_A |X_t| dP \le aP(A) + \sup_{t\in T} \int_{[|X_t|>a]} |X_t| dP.$$

Letting $A = \Omega$ proves (B). To prove (A), we know $\sup_{t\in T} \int_{[|X_t|>a]} |X_t| dP \to 0$ as $a \to \infty$. Thus, for $\epsilon > 0$, we can find large a such that $\sup_{t\in T} \int_{[|X_t|>a]} |X_t| dP \le \epsilon/2$. Then picking $\delta = \epsilon/(2a)$ completes (A).

Proof continued: Conversely, Suppose (A) and (B) holds, by Chebychev's inequality and (B),

$$\sup_{t\in T} P[|X_t| > a] \leq \frac{\sup_{t\in T} E(|X_t|)}{a} = \frac{\text{a finite constant}}{a}.$$

Using (A), for $\epsilon > 0$, there exists δ such that whenever $P(A) < \delta$, $\int_A |X_t| dP < \epsilon$ for all t. We then pick large a such that $P[|X_t| > a] \leq \sup_{t\in T} P[|X_t| > a] \leq \delta$. Then for all t,

$$\sup_{t\in T} \int_{[|X_t|>a]} |X_t| dp \leq \epsilon,$$

which is the ui property.

Example 6.5.1

Let $\{X_n\}$ be a sequence of random variables with

$$P[X_n = 0] = 1 - 1/n, P[X_n = n] = 1/n$$

Then $E(|X_n|) = 1$ for all n. Thus $\{X_n\}$ has uniform bounded first absolute moments. However it is not a ui family, because

$$\int_{[|X_n|>a]} |X_n| dP = I(a \leq n); i.e., \sup_n \int_{[|X_n|>a]} |X_n| dp = 1.$$

Schwartz Ineq: Suppose $X, Y \in L_2$, then

$$|E(XY)| \leq E(|XY|) \leq \sqrt{E(X^2)E(Y^2)}.$$

Hölder's ineq: Suppose $X \in L_p$ and $Y \in L_q$, where p, q satisfy

$$p > 1, q > 1, \frac{1}{p} + \frac{1}{q} = 1$$

Then

$$|E(XY)| \leq E(|XY|) \leq (E|X|^p)^{1/p}(E|X|^q)^{1/q}.$$

Or $\|XY\|_1 \leq \|X\|_p \|Y\|_q$. Schwartz ineq is a special case when $p = q = 2$.

6.5.2 Interlude: A Review of Inequalities

Minkowski Ineq: For $1 \le p < \infty$, suppose $X, Y \in L_p$. Then $X + Y \in L_p$ and

$$\|X + Y\|_p \le \|X\|_p + \|Y\|_p.$$

Jensen's ineq: Suppose $u : \mathbb{R} \mapsto \mathbb{R}$ is convex and $E(|X|) < \infty$ and $E(|u(X)|) < \infty$. Then

$$E(u(X)) \ge u(E(X)).$$

Example 6.5.2

If $X \in L_\beta$, then $X \in L_\alpha$ provided $0 < \alpha < \beta$. Furthermore

$$\|X\|_t = (E|X_t|^t)^{1/t}$$

is non-decreasing in t. Consequently, if $X_n \stackrel{L_p}{\to} X$ and $p' < p$, then

$$X_n \stackrel{L_{p'}}{\to} X.$$

Proof: Set $p = \beta/\alpha > 1$ and then let $q = \beta/(\beta - \alpha) > 1$. We have $1/r + 1/s = 1$. Then let $Z = |X|^\alpha$, $Y = 1$. Using Hölder's ineq,

$$E(|X|^\alpha) = E(|ZY|) \leq \|Z\|_p \|Y\|_q = \|Z|_p = E(|Z|^p)^{1/p}$$
$$= E(|X|^{\alpha p})^{\alpha/\beta} = E(|X|^\beta)^{\alpha/\beta}.$$

Thus $\|X\|_\alpha \leq \|X\|_\beta$.

6.6 More on L_p Convergence

We work up to an answer to the question: If random variables converge, when do their moments converge? Assume $\{X_n\}$ and X are defined on $(\Omega, \mathcal{B}, P)$.

(1) A form of Scheffé's lemma:

$$X_n \stackrel{L_1}{\to} X \iff \sup_{A \in \mathcal{B}} \left| \int_A X_n dP - \int_A X dP \right| \to 0.$$

Letting $A = \Omega$, then L_1 convergence implies $E(X_n) \to E(X)$.

(2) If $X_n \stackrel{L_p}{\to} X$, then $E(|X_n|^p) \to E(|X|^p)$, or $\|X_n\|_p \to \|X\|_p$

Proof: First show (2):$|\|X_n\|_p - \|X\|_p| \le \|X_n - X\|_p$.

Now for the $\Leftarrow$ of (1):

$$\begin{aligned} E|X_n - X| &= \int_{[X_n > X]} (X_n - X) dP + \int_{[X_n \le X]} (X - X_n) dP \\ &\le 2 \sup_A \left| \int_A X_n dP - \int_A X dP \right|. \end{aligned}$$

Proof: Now for the $\Rightarrow$ of (1):

$$\begin{aligned}\sup_A \left| \int_A X_n dP - \int_A X dP \right| &\le \sup_A \int_A |X_n - X| dP \\ &\le \int |X_n - X| dP \\ &= E(|X_n - X|) \to 0.\end{aligned}$$

Theorem 6.6.1

Suppose $X_n \in L_1$ for $n \geq 1$. The following statements are equivalent:

(a) $\{X_n\}$ is L_1-convergent.

(b) $\{X_n\}$ is L_1-Cauchy; that is, $E|X_n - X_m| \to 0$ as $n, m \to \infty$.

(c) $\{X_n\}$ is ui and $\{X_n\}$ converges in probability.

So if $X_n \stackrel{a.s}{\to} X$ or $X_n \stackrel{P}{\to} X$ (later, or $X_n \stackrel{D}{\to} X$) and $\{X_n\}$ is ui, then $X_n \stackrel{L_1}{\to} X$ and $E(X_n) \to E(X)$.

Proof: (a) to (b): If $X_n \stackrel{L_1}{\to} X$, then $E|X_n - X_m| \leq E|X_n - X| + E|X_m - X| \to 0$ as $n, m \to \infty$.

Proof continued: (b) to (c): We first show ui using Theorem 6.5.1. Because of (b), for $\epsilon > 0$, there exists N_ϵ such that for $n, m \geq N_\epsilon$, then

$$\int |X_n - X_m| dP < \epsilon/2.$$

For any $A \in \mathcal{B}$ and $n \geq N_\epsilon$,

$$\int_A |X_n| dP \leq \int_A |X_{N_\epsilon}| dP + \int |X_n - X_{N_\epsilon}| dP \leq \int_A |X_{N_\epsilon}| dP + \epsilon/2.$$

That is $\sup_{n \geq N_\epsilon} \int_A |X_n| dP \leq \int_A |X_{N_\epsilon}| dP + \epsilon/2$ and thus

$$\begin{aligned} \sup_n \int_A |X_n| dP \leq \max(\sup_{m < N_\epsilon} \int_A |X_m| dP, \int_A |X_{N_\epsilon}| dP + \epsilon/2) \\ \leq \sup_{m \leq N_\epsilon} \int_A |X_m| dP + \epsilon/2. \end{aligned}$$

Take $A = \Omega$, $\sup_n E(|X_n|) \leq \sup_{m \leq N_\epsilon} E(|X_m|) + \epsilon/2 < \infty$.

Proof continued: (b) to (c) continued: Further more, since $\{X_m : m \leq N_\epsilon\}$ is a finite family which is ui. We can find a $\delta > 0$, such that if $P(A) \leq \delta$, then

$$\sup_{m \leq N_\epsilon} \int_A |X_m| dP < \epsilon/2.$$

Finally, we conclude that for $\epsilon > 0$, we find a δ, such that if $P(A) \leq \delta$, then

$$\sup_n \int_A |X_n| dP < \epsilon/2 + \epsilon/2 = \epsilon.$$

Hence $\{X_n\}$ is ui. To check $\{X_n\}$ converges in probability, we have $P[|X_n - X_m| > \epsilon] \leq E(|X_n - X_m|)/\epsilon \to \infty$. Thus $\{X_n\}$ is Cauchy i.p.

Proof continued: (c) to (a): If $X_n \stackrel{P}{\to} X$, then there exists a subsequence $\{n_k\}$ such that $X_{n_k} \stackrel{a.s.}{\to} X$. By Fatou's lemma

$$E(|X|) = E(\liminf_{n_k\to\infty} |X_{n_k}|) \leq \liminf_{n_k\to\infty} E(|X_{n_k}|) \leq sup_n E(|X_n|) < \infty$$

since $\{X_n\}$ is ui. So $X \in L_1$. Also, for any $\epsilon > 0$,

$$\int |X_n - X|dP \leq \int_{[|X_n-X|\leq\epsilon]} |X_n - X|dP + \int_{[|X_n-X|>\epsilon]} |X_n - X|dP$$

$$\leq \epsilon + \int_{[|X_n-X|>\epsilon]} |X_n|dP + \int_{[|X_n-X|>\epsilon]} |X|dP \doteq \epsilon + A_n + B_n.$$

Because $P[|X_n - X| > \epsilon] \to 0$ and $X \in L_1$ and $\{X_n\}$ is ui, we have $A_n, B_n \to 0$.

Example

Suppose X_1 and X_2 are iid $N(0,1)$ and define $Y = X_1/|X_2|$ which has a Cauchy distribution with density $f(y) = 1/\{\pi(1+y^2)\}$, for $y \in \mathbb{R}$. Define $Y_n = X_1/(|X_2| + n^{-1})$.

Then $Y_n \to Y$. But $\{Y_n\}$ is NOT ui.

Because if it is, then $E(Y_n) = 0 \to E(Y)$ in which $E(Y)$ does not exist (contradiction).

Theorem 6.6.2

Suppose $p \geq 1$, $X_n \in L_p$ for $n \geq 1$. The following statements are equivalent:

(a) $\{X_n\}$ is L_p-convergent.

(b) $\{X_n\}$ is L_p-Cauchy; that is, $E|X_n - X_m|_p \to 0$ as $n, m \to \infty$.

(c) $\{|X_n|^p\}$ is ui and $\{X_n\}$ converges in probability.

This also states that L_p is a complete metric space; that is every Cauchy sequence has a limit.

Proof is similar and left as HW 6-2.

Other HW 6 problems: Section 6.7, Q1-Q2, Q4-6, Q9, Q13, Q15-16, Q19-Q20, Q23-26, Q31, Q33