Dr. Osonde Osoba

This Week:

- Stochastic Convergence
  - Review: sequential and functional (u, e) convergence
  - Stochastic modes of convergence
    - o/a.e/a.s, p, m, d
  - Cauchy Criterion

H/W:

~~L-G~~: 7.41, 7.44, 7.45, 7.50

Gubner: 14.1, 14.2, 14.3

Summary:

$a_n \xrightarrow{e} a \iff \forall \varepsilon > 0, \exists n_0 \in \mathbb{N}, \forall n > n_0, |a_n - a| < \varepsilon$

	a_n	a	
e - everywhere	$X_n(\omega)$	$X(\omega)$	$\forall \omega \in \Omega$
m - mean square	$E[(X_n - X)^2]$	0	—
p - in probability	$P(\lvert X_n - X\rvert > \varepsilon)$	0	$\forall \varepsilon > 0$
d - in distribution	$F_{X_n}(x)$	$F_X(x)$	$\forall$ pts of continuity for $F_X(\cdot)$

o: with probability one: $P(\lim_n X_n = X) = 1$

$u \Rightarrow e \Rightarrow o \Rightarrow p \Rightarrow d \qquad m \Rightarrow p \Rightarrow d \qquad o \not\Leftrightarrow m$

0/ **Review:**

$\{a_n\}_{n\in\mathbb{N}}, \quad \lim_{n\to\infty} a_n = L$

$\iff \forall \varepsilon>0 \quad \exists n_0 \in \mathbb{N} : \forall n>n_0 \quad |a_n - L| < \varepsilon$

$\{f_n\}_{n\in\mathbb{N}}$: a sequence of functions, $f_n : \Omega \to \mathbb{R}$

Pointwise (pt) convergence of f_n (or convergence everywhere (e)) applies sequential convergence to each $\omega \in \Omega$. i.e.

$f_n \xrightarrow{e} f$

$\iff \forall \omega\in\Omega, \ \forall \varepsilon>0, \ \exists n_0 \in \mathbb{N} : \forall n>n_0, \ |f_n(\omega) - f(\omega)| < \varepsilon$

Uniform (u) Convergence is similar to pointwise convergence except the value of n_0 no longer depends on $\omega\in\Omega$. i.e.

$f_n \xrightarrow{u} f$ $\quad\quad$ $[u \Rightarrow e]$

$\iff \forall \varepsilon>0, \ \exists n_0 \in \mathbb{N} \ \forall \omega\in\Omega : \forall n>n_0, \ |f_n(\omega) - f(\omega)| < \varepsilon$

Recall that "random variables" $X_n : \Omega \to \mathbb{R}$ are just (measurable) functions. So e- and u- convergence also apply to rv-sequences $\{X_n\}_n$. But we can also develop other modes of convergence that incorporate probabilistic or measure-theoretic ideas.

I Stochastic Convergence:

(a) a.e- or o-Convergence:

e- and u-Convergence enforce a limit condition $\forall \omega \in \Omega$.

The "almost-sure"(a.s)/"almost-everywhere"(a.e)/"with-probability-one"(o) qualifier relaxes the condition by enforcing only on "most" $\omega \in \Omega$.

e.g. $X \overset{a.e.}{=} Y \iff P(\{\omega \in \Omega : X(\omega) \neq Y(\omega)\}) = 0$

i.e X and Y are equal except for an event, A,

$A = \{\omega \in \Omega : X(\omega) \neq Y(\omega)\}$ of zero measure, i.e. $P(A) = 0$.

A useful property of $\overset{a.e}{=}$: for any measurable fXn $g(\cdot)$

$$[X \overset{a.e}{=} Y] \iff E[g(X)] = E[g(Y)]$$

(proof?)

Thus Convergence almost-everywhere (a.e) or with probability-one (o):

$$\boxed{X_n \xrightarrow{o} X \iff P\left(\{\omega \in \Omega : \lim_{n \to \infty} X_n(\omega) \neq X(\omega)\}\right) = 0}$$

i.e convergence may fail pointwise, but only on a zero-measure set.

e.g.: $\Omega = [0,1]$; $P((a,b]) = b - a \quad \forall (a,b] \in [0,1]$

$X_n(\omega) = n \cdot I_{A_n}(\omega) \qquad A_n = [0, 1/\sqrt{n}]$

$$\Rightarrow \lim_{n \to \infty} X_n(\omega) = \lim(n) \cdot \lim I_{A_n}(\omega) = \begin{cases} 0 & \omega \in (0,1) \\ \infty & \omega \in \lim_n A_n = \{0\} \end{cases}$$

$P(\lim_n A_n) = P(\{0\}) = 0$

$$\Rightarrow P(\{\omega \in \Omega : \lim X_n(\omega) \neq 0\}) = P(\{0\}) = 0$$

$$\therefore X_n \xrightarrow{a.s} 0$$

(b) **P-Convergence**:

$X_n \xrightarrow{P} X$ [X_n converges ~~to X in probability~~]

$$\Leftrightarrow \boxed{\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0 \quad \forall \varepsilon > 0}$$

This differs from $\xrightarrow{a}$ in that we are only concerned that X_n deviates from X by $>\varepsilon$ only over a set that shrinks in probability as $n \to \infty$. $\xrightarrow{P}$ does **not** require that $\lim X_n = X$ for event of non-zero probability.

(c) **m-Convergence**:

X_n converges to X in mean-square

$$\boxed{X_n \xrightarrow{m} X \Leftrightarrow \lim_{n\to\infty} E[(X_n - X)^2] = 0}$$

- This is about the convergence of the integral $\int (X_n - X)^2 dP = E[(X_n - X)^2]$ not the function $(X_n - X)$.
- It is a special case of convergence in $\mathcal{L}^P$:

$$X_n \xrightarrow{\mathcal{L}^P} X \Leftrightarrow \boxed{\lim_{n\to\infty} E[(X_n - X)^P] = 0}$$

$$m = \mathcal{L}^P \Big|_{P=2}$$

(d) <u>d-Convergence</u> :

Convergence in distribution (or weak convergence) is a property of the CDFs F_{X_n}, not the rvs $\{X_n\}_n$:

$$X_n \xrightarrow{d} X$$

$$\Longleftrightarrow \boxed{\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \quad \forall \text{ pts of continuity of } F_X}$$

Another formulation useful in higher-dim r.v : $X : \Omega \to \mathbb{R}^d$

$$X_n \xrightarrow{d} X \iff \boxed{\begin{array}{l} \lim_{n\to\infty} E_{X_n}[g(X_n)] = E_X[g(X)] \\ \forall \text{ bounded continuous } g(\cdot) \text{ on } \mathbb{R}^d \end{array}}$$

This is the weakest mode of stochastic convergence.

<u>Summary</u>: UC MOPED

$$a_n \to a \iff \forall \varepsilon > 0, \exists n_0 \in \mathbb{N} : \forall n > n_0, |a_n - a| < \varepsilon$$

	a_n	a	
e – everywhere	$X_n(\omega)$	$X(\omega)$	$\forall \omega \in \Omega$
m – mean square	$E[(X_n - X)^2]$	0	
p – in probability	$P(\lvert X_n - X\rvert > \varepsilon)$	0	$\forall \varepsilon > 0$
d – in distribution	$F_{X_n}(x)$	$F_X(x)$	$\forall$ pts of cont for $F_X(\cdot)$

Also: $[X_n \xrightarrow{o} X] \iff [P(\lim X_n = X) = 1]$

$$\boxed{\begin{tikzcd} u \arrow[r, Rightarrow] \arrow[drr, Rightarrow] & e \arrow[r, Rightarrow] \arrow[dr, "\times" description] & o \arrow[r, Rightarrow] \arrow[d, leftrightarrow, "\times" description] & p \arrow[r, Rightarrow] & d \\ & & m \arrow[ur, Rightarrow] & & \end{tikzcd}}$$

Examples:

i) $\{X_n\}_n$: $P(X_n = k) = \begin{cases} 1 - 1/n^2; & k=0 \\ 1/n^2; & k=n \\ 0; & k \notin \{0,n\} \end{cases}$

$\lim_n X_n \triangleq X_\infty$

$\Rightarrow P(X_\infty = k) = \begin{cases} 1; & k=0 \\ 0; & k \neq 0 \end{cases}$ i.e $X_\infty \overset{a.e}{=} 0$

$\Rightarrow P(\{\lim_n X_n = 0\}) = 1$

$\Leftrightarrow X_n \xrightarrow{a.e} 0$ $(o \Rightarrow p \Rightarrow d)$

$\lim_n E[(X_n - 0)^2] = \lim_n E[X_n^2] = \lim_n \left(0^2(1-1/n^2) + n^2(1/n^2)\right)$

$\Rightarrow \lim_n E[(X_n - 0)^2] = 1 \quad \therefore X_n \overset{m}{\not\to} 0$

[Counter-example illustrating $o \not\Rightarrow m$]

ii) $\{X_n\}_n$: $X_n(\omega) = 1/n \quad \omega \in [0,1] = \Omega$ (Constant rv sequence not very interesting...)

$\Rightarrow \forall \varepsilon > 0, \exists n_0 \in \mathbb{N} : |X_n(\omega) - 0| < \varepsilon \quad \forall n > n_0, \forall \omega \in \Omega$

Just pick $n_0 = \lceil 1/\varepsilon \rceil \quad \therefore X_n \xrightarrow{u,e,o} 0$

iii) Define the sequence $\{X_n(\omega)\}_{n \in \mathbb{N}}$ on $\Omega = [0,1]$ with $P((a,b]) = b-a$

$X_1(\omega) = I_{[0,1]}$

$X_2(\omega) = I_{[0,1/2]}, \qquad X_3 = I_{[1/2,1]}$

$X_4(\omega) = I_{[0,1/4]}, \; X_5(\omega) = I_{[1/4,1/2]}, X_6 = I_{[1/2,3/4]}, X_7 = I_{[3/4,1]}$

$\vdots \qquad n = 2^k + j \qquad X_n(\omega) = I_{[j/2^k, (j+1)/2^k]}$

Qu: $X_n \xrightarrow{a.e/p} 0$???

$\forall \omega \in \Omega,\ \exists \varepsilon > 0,\ \forall n_0 \in \mathbb{N}: |X_n(\omega) - 0| > \varepsilon$ for some $n > n_0$

i.e. $X_n \overset{e}{\not\to} 0$ (since the $\max_{\omega}\{X_n(\omega)\} = 1\ \forall n$)

This holds for all $\omega \in A$ with $P(A) > 0$

$\therefore\ X_n \overset{a.e}{\not\to} 0$

But $\forall \varepsilon > 0$ $\quad P(|X_n - 0| > \varepsilon) = P(X > \varepsilon)$

$$= P(A_n) \approx 2^{-\log_2 n} \to 0$$

$$\Rightarrow \lim_{n\to\infty} P(|X_n - 0| > \varepsilon) = 0$$

$$\therefore\ X_n \xrightarrow{P} 0$$

but $X_n \overset{a.e, e}{\not\to} 0$ //

iv) $\{X_i\}_{i=1}^{\infty}$: iid with $\sigma_x^2 < \infty$

$$\bar{X}_n = \tfrac{1}{n} \sum_i^n X_i$$

Claim: $\quad \bar{X}_n \xrightarrow{P} \mu_x$

Proof:

Chebyshev's Inequality

$$0 \le \lim_n P(|\bar{X}_n - \mu_x| > \varepsilon) \le \lim_n \sigma_{\bar{X}_n}^2 / \varepsilon^2 = 0$$

$$\Rightarrow \lim_n P(|\bar{X}_n - \mu_x| > \varepsilon) = 0$$

$$\therefore\ \bar{X}_n \xrightarrow{P} \mu_x \ //$$

Weak Law of Large Numbers: $\boxed{\bar{X}_n \xrightarrow{P} \mu_x}$

Strong Law of Large Numbers: $\boxed{\bar{X}_n \xrightarrow{a.e} \mu_x}$

v/ MS-LLN:

$$\bar{X}_n \xrightarrow{m} \mu_x$$

Proof:

$$E[(\bar{X}_n - \mu_x)^2] = E[\bar{X}_n^2] + \mu_x^2 - 2\mu_x E[\bar{X}_n]$$
$$= E[\bar{X}_n^2] - \mu_x^2$$
$$= \sigma_{\bar{X}_n}^2 + \mu_{\bar{X}_n}^2 - \mu_x^2$$

$$\sigma_{\bar{X}_n}^2 = \sigma_x^2/n \quad ; \quad \mu_{\bar{X}_n}^2 = \mu_x^2$$

$$\Rightarrow E[(\bar{X}_n - \mu_x)^2] = \sigma_x^2/n$$

$$\therefore \lim_{n\to\infty} E[(\bar{X}_n - \mu_x)^2] = \lim_{n\to\infty} \sigma_x^2/n = 0$$

i.e. $\bar{X}_n \xrightarrow{m} \mu_x$

Tools for Testing Stochastic Convergence:

i/ $e \Rightarrow a \Rightarrow p \Rightarrow d$; $m \Rightarrow p \Rightarrow d$

ii/ Markov and Chebyshev Inequalities especially for p-convergence.

iii/ Variance-Bias Decomposition (for m)

$$X_n \xrightarrow{m} d^{\leftarrow \text{constant}} \iff E[(X_n - d)^2] \longrightarrow 0$$

$$MSE(X_n; d) \triangleq E[|X_n - d|^2]$$

$$\therefore X_n \xrightarrow{m} d \iff MSE(X_n; d) \longrightarrow 0$$

Claim: $MSE(X;d) = Var(X) + (Bias(X;d))^2$

$[\ M = V + B^2$; (Variance-Bias Decomposition)$]$

where $Bias(X;d) = |E[X]-d|$

Proof: $MSE(X;d) = E[(X-d)^2]$

$= E[X^2] + d^2 - 2d \cdot E[X]$

$= \sigma_X^2 + \mu_X^2 + d^2 - 2d \cdot \mu_X$

$= \sigma_X^2 + (\mu_X - d)^2$

$= \sigma_X^2 + |E[X]-d|^2$

$\therefore MSE(X;d) = Var(X) + (Bias(X;d))^2$ //

e.g.: Claim: $[X_n \xrightarrow{m} X,\ Y_n \xrightarrow{m} Y] \Rightarrow [(X_n+Y_n) \xrightarrow{m} X+Y]$

Proof: let $Z_n = X_n + Y_n$, $Z = X+Y$

$E[(Z_n-Z)^2] = E[((X_n+Y_n)-(X+Y))^2]$

$= E[((X_n-X)+(Y_n-Y))^2]$

$= E[(X_n-X)^2] + E[(Y_n-Y)^2] + 2E[(X_n-X)(Y_n-Y)]$

But:

(a) $[X_n, Y_n \xrightarrow{m} X, Y] \Leftrightarrow \lim_n E[(X_n-X)^2] = \lim_n E[(Y_n-Y)^2] = 0$

(b) Generalized C-S ineq. : $(E[S \cdot T])^2 \leq E[S^2] \cdot E[T^2]$ (*)

$\Rightarrow E[(X_n-X)(Y_n-Y)] \leq (E[(X_n-X)^2] \cdot E[(Y_n-Y)^2])^{1/2}$

$\Rightarrow 0 \leq \lim_n E[(Z_n-Z)^2] \leq \lim_n (E[(X_n-X)^2] \cdot E[(Y_n-Y)^2])^{1/2} = 0$

$\Rightarrow \lim_n E[(Z_n-Z)^2] = 0 \quad \therefore Z_n \xrightarrow{m} Z$ //

(*) Prove using $E[(S-cT)^2] \triangleq g(c) \geq 0$

In general:

$$X_n \xrightarrow{o/p/m} X \quad \text{and} \quad Y_n \xrightarrow{o/p/m} Y$$

$$\Rightarrow \quad X_n + Y_n \xrightarrow{o/p/m} X + Y$$

Does not Hold for $\xrightarrow{d}$ in general

<u>Continuity</u> . <u>Theorem</u>

Suppose g is continuous

$$\left\{\begin{array}{l} X_n \xrightarrow{o} X \\ X_n \xrightarrow{p} X \\ X_n \xrightarrow{d} X \end{array}\right\} \Rightarrow \left\{\begin{array}{l} g(X_n) \xrightarrow{o} g(X) \\ g(X_n) \xrightarrow{p} g(X) \\ g(X_n) \xrightarrow{d} g(X) \end{array}\right.$$

Does not hold for $\xrightarrow{m}$ in general.

ex: [d example]

$$X_n \sim \text{Cauchy}(0, 1/n)$$

$$X_n \xrightarrow{d} ?$$

Ans: $X_n \sim \text{Cauchy}(0, 1/n)$

$$\Rightarrow F_{X_n}(x) = 1/2 + 1/\pi \arctan\left(\frac{x-m}{d}\right)$$

$$= 1/2 + 1/\pi \arctan(nx)$$

$$\lim_n F_{X_n}(x) = 1/2 + 1/\pi \lim_n \arctan(nx)$$

$$= 1/2 + 1/\pi \cdot \pi/2 \, \text{sgn}(x) = 1/2(1 + \text{sgn}(x))$$

$$= u(x) \triangleq F_X(x)$$

$$\Rightarrow X_n \xrightarrow{d} X = 0 \quad \text{i.e} \quad f_X(x) = \delta(x)$$

Central Limit Theorem: (d-convergence example)

Let $\{X_i\}_i$ be random samples (i.e. iid) with $\sigma_X^2 < \infty$

and let $Z_n = STD(\bar{X}_n) \equiv \dfrac{\bar{X}_n - \mu_X}{\sigma_X/\sqrt{n}}$

$$\Rightarrow \quad Z_n \xrightarrow{d} Z \sim N(0,1)$$

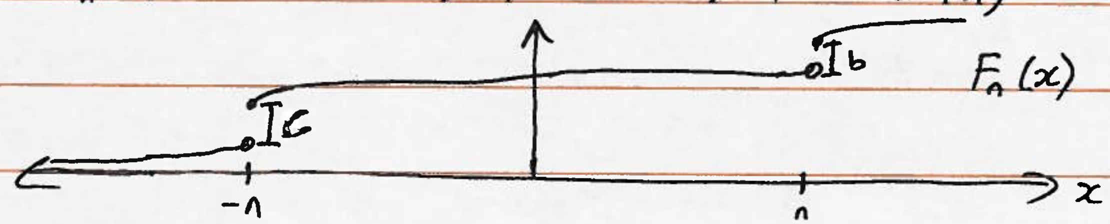
Note that $Z_n = STD(\bar{X}_n) = STD\left(\sum_{i=1}^{n} X_i\right)$

Ex: [d-Convergence can be problematic]

Suppose $0 \le a, b, c \le 1$, $a+b+c=1$, and $G(x)$ is a continuous CDF

Let $\{X_n\}_{n \ge 0}$ have CDF:

$$F_n(x) = a\,G(x) + b \cdot U(x-n) + c \cdot U(x+n)$$



$\Rightarrow \lim_{n\to\infty} F_n(x) = c + a\,G(x) \in [c, a+c] \quad \forall x$

$\therefore \lim_n F_n(x)$ is not a CDF.

i.e. The limit of a sequence of CDFs is not necessarily a CDF. So always check $\lim F_n$!

In particular, this CDF sequence $\{F_n\}_{n \ge 0}$ is not "tight".

Cauchy Criterion (c):

A sequence $\{a_n\}_n$ is a Cauchy sequence :

$$\iff \boxed{\lim_{n\to\infty} \lim_{m\to\infty} |a_n - a_m| = 0}$$

$\iff \lim_{n\to\infty} \lim_{m\to\infty} d(a_n, a_m) = 0$ where d is the absolute value metric $|a_n - a_m|$

Suppose $a_n \in S$ $\forall n$. Then S is a complete metric space iff every Cauchy sequence has its limit in S i.e. $\lim_n a_n \in S$.

Cauchy's Criterion states that if $\{b_n\} \in S$ is a Cauchy sequence in a complete metric space S, then $\lim_{n\to\infty} b_n$ exists, is unique, and $\boxed{\lim_{n\to\infty} b_n \in S}$

e.g. a: $S = (0,1)$; $a_n = 1/n$ $\quad \lim_{n\to\infty} a_n = 0 \notin S$

$\lim_n \lim_m |(a_n - a_m)| = \lim_n \lim_m |(\frac{1}{n} - 1/m)| = 0 \Rightarrow \{a_n\}_n$ is Cauchy

but S is not complete since $\lim a_n \notin S$

if $\bar{S} = [0,1]$, then $\bar{S}$ is complete and $\lim_{n\to\infty} a_n = 0 \in \bar{S}$

We can apply this idea to Stochastic Convergence because convergent r.v. sequences in v, m-, or p- form complete metric spaces (with different metrics). Thus if we can show that $\{X_n\}_{n\geq 0}$ is Cauchy in m or p then we know that $X_n \xrightarrow{m \text{ or } p}$. We just don't know the limit.

$\{X_n\}_n$ is Cauchy in mean-square (m)

$$\Leftrightarrow \boxed{\lim_{n\to\infty} \lim_{m\to\infty} E\left[(X_n - X_m)^2\right] = 0}$$

$\{X_n\}_n$ is Cauchy in probability (p)

$$\Leftrightarrow \boxed{\lim_{n\to\infty} \lim_{m\to\infty} P\left(|X_n - X_m| > \varepsilon\right) = 0 \quad \forall \varepsilon > 0}$$

e.g: $X_n(\omega) = I_{[0,1/n]}$ $\quad \{X_n\}$ independent

Does $X_n \xrightarrow{m}$?

Ans: Assume $m > n$ WLOG

$$E\left[(X_n - X_m)^2\right] = \int_{1/m}^{1/n} 1 \, d\omega = 1/n - 1/m$$

$$\left[\because I_{[0,1/n]} - I_{[0,1/m]} = I_{[1/m,1/n]}\right]$$

$$\Rightarrow \lim_m \lim_n E\left[(X_n - X_m)^2\right] = \lim_n \lim_m \left(1/n - 1/m\right) = 0$$

$\Rightarrow \{X_n\}$ is Cauchy in mean-square

$\therefore$ X_n converges in mean-square.

Note: Cauchy criterion does not say where X_n is going.