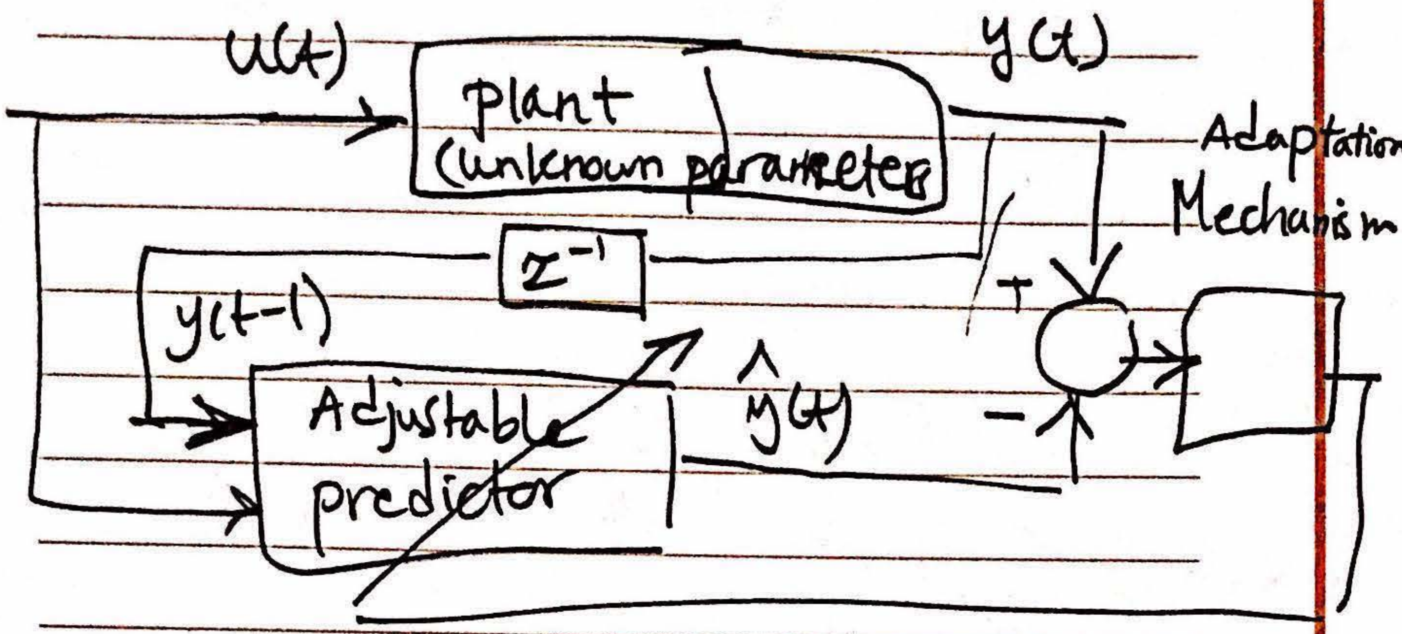
Lesson 15

Convergence of Random Variables (Stochastic Convergence)

Stochastic Convergence is concerned about convergence of sequences of random variables, and is of extreme importance in probability theory and random processes. One reason for that is because "limit theorems" in probability theory concern

themselves with asymptotic (limiting behavior of random variables. Consider the following online parameter estimation scheme in an adaptive control system



This scheme uses the error between predictions of an adjustable predictor and the true output of the system to estimate the unknown parameters of the system and "adapt" a controller for desired performance.

The parameter estimates are sequences of random variables that update as the plant operates. Of particular interest is whether the parameter estimates "converge" to the true parameters of the system in some sense.

Sequences Again!!

Recall that a sequence is a function from $\mathbb{N}$ to $\mathbb{R}$ (actually it can be a function from any countable set to any set)

Def (Limit of a Sequence Again): We say that the sequence $\{a_n\}$ converges to some $a \in \mathbb{R}$ iff $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}$ s.t $\forall n > N$,

$|a_n - a| < \varepsilon$

and we write $\lim_{n \to \infty} a_n = a$.

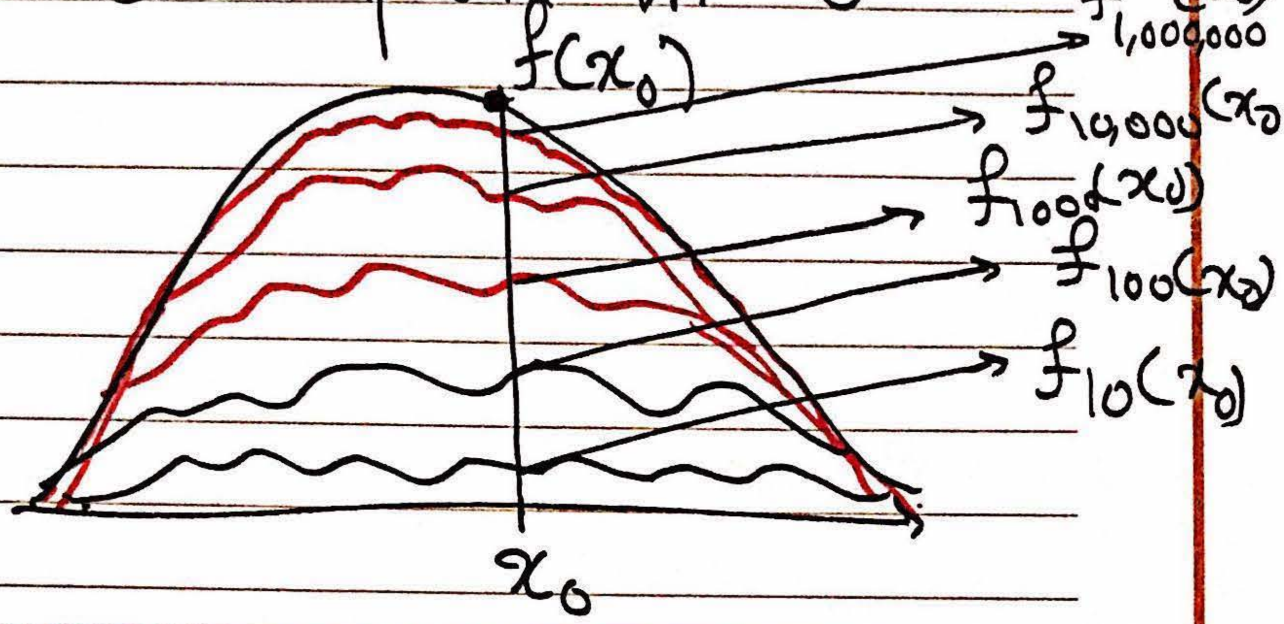
Convergence of Sequences of functions

Pointwise Convergence

(Convergence Everywhere, Sure Convergence, Convergence Surely)

Given a sequence of functions $\{f_n\}_{n\in\mathbb{N}}$: $f_n : U \rightarrow V$, pointwise convergence or convergence <u>everywhere</u>

applies sequence convergence to each point in U



$\forall x \in U \quad \lim_{n\to\infty} f_n(x) = f(x)$

i.e.

$\forall x \in U, \ \forall \varepsilon > 0 \ \exists N(x) \in \mathbb{N}$, s.t.

Uniform Convergence

Uniform convergence is *stricter than* pointwise convergence, *in the sense that* $N(x)$ is the same for all $x \in U$,

i.e. $\forall x \in U,\ N(x) = N$

Recall that sequences of r.v.'s $X_n : \Omega \to \mathbb{R}$ are just sequences of $\mathcal{F}$-measurable functions, so convergence everywhere and uniformly

apply to $\{X_n\}_{n \in \mathbb{N}}$ as well.

Before going into details, we summarize modes of convergence in a single table for r.v.'s

Summary of convergence Modes

$$\lim_{n\to\infty} a_n = a$$

Mode	a_n	a	Comment
Everywhere (pointwise)	$X_n(\omega)$	$X(\omega)$	$\forall \omega \in \Omega$
Almost surely (with prob 1)	$X_n(\omega)$	$X(\omega)$	$\forall \omega \in A \in \mathcal{F}$, $\mathbb{P}(A)=1$
In r^{th} moment	$\mathbb{E}[\lvert X_n - X\rvert^r]$	0	—
In probability	$\mathbb{P}(\lvert X_n - X\rvert > \varepsilon)$	0	$\forall \varepsilon > 0$
In distribution	$F_{X_n}(x)$	$F_X(x)$	$\forall x$ for which $F_X(x)$ is continuous

Simple Exercise: Modify a_n's in the previous table such that all a's become 0.

Let's focus on each mode of convergence

Pointwise (everywhere/sure) Convergence

Def) Assume that $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of r.v.'s that are defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, X_n is said to Converge To X everywhere/pointwise (p.w.) iff:

Remark: Since for a fixed $\omega \in \Omega$, $\{X_n(\omega)\}_{n \in \mathbb{N}}$ is a sequence of real numbers, the meaning of convergence is clear.

Theorem: Assume that

$\forall \omega \in \Omega, \lim_{n \to \infty} X_n(\omega) = X(\omega)$,

where X_n's are r.v.'s on

$(\Omega, \mathcal{F}, \mathbb{P})$. Then X is a

r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$.

Proof: Omitted.

Remark: The above Theorem says if X is ~~a~~ the limit of X_n's, it is indeed a r.v., i.e. is an $(\mathcal{F}, \mathcal{B})$ measurable function itself.

Remark: Since pointwise convergence is for <u>all</u> members of the sample space, it is too strict for most practical applications. We can relax

it to almost sure convergence.

Def) $\{X_n\}_{n\in\mathbb{N}}$ is said to converge to X <u>almost surely</u> (a.s.) iff

$$\lim_{n\to\infty} X_n(\omega) =$$

Alternatively,

$\mathbb{P}(\{\omega \mid X_n(\omega) \longrightarrow X(\omega)\}) = 1$

Therefore, a.s. convergence is also called convergence with probability $1 \equiv$ w.p.1.

This mode of convergence gives the sequence of r.v.'s the freedom not to converge on a set of zero probability, i.e., for some $\omega \in A^c$, $X_n(\omega)$, $X_n(\omega) \not\longrightarrow X(\omega)$; therefore, it is weaker than sure/ pointwise convergence.

Remark: Almost sure convergence is quite strict itself, so a weaker mode of convergence

can be defined.

Def) (Convergence in probability (i.p.))

$\{X_n\}_{n\in\mathbb{N}}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ is said to converge in probability (i.p.) to X iff

$$\lim_{n\to\infty} \underbrace{\mathbb{P}(\qquad\qquad)}_{p_n} = 0 \quad \forall \varepsilon > 0$$

Remark: Instead of dealing with values of r.v.'s at some ω's, convergence in probability concerns itself with a sequence of probabilities. Therefore, it is different from a.s. convergence. It says the probability of the event

$\{\omega \mid |X_n(\omega) - X(\omega)| > \varepsilon\}$

(ω's for which X_n deviates from X by $\varepsilon > 0$) approaches zero when $n \to \infty$, for all positive ε's.

Def) (Convergence in the r^{th} moment / r^{th} mean / $\mathcal{L}^r$ convergence)

A sequence of r.v.'s on $(\Omega, \mathcal{F}, \mathbb{P})$ $\{X_n\}_{n \in \mathbb{N}}$ is said to converge in the r^{th} moment (mean) to X iff

$$\lim_{n \to \infty} \underbrace{\mathbb{E}\qquad\qquad}_{e_n} =$$

In particular, when $r=2$, this mode of convergence is called $\mathcal{L}^2$ convergence or convergence in the mean-squared sense.

The last mode of convergence is the weakest mode of convergence, i.e., in distribution. Essentially we look at the distributions of a sequence

of r.v.'s converging to some distribution. This mode of convergence is extremely important in understanding the Central Limit Theorem.

Def. A Sequence of r.v.'s $\{X_n\}_{n \in \mathbb{N}}$ is said to converge to X in distribution (i.d.) iff:

$$\lim_{n \to \infty}$$

$\forall x \in \mathbb{R}$, where

In other words, the sequence of distributions must converge at all points of continuity of $F_X(\cdot)$

Important remark: Convergence in distribution is different from other modes of convergence in the sense that it does not need all of the r.v.'s to be defined on the same probability space.

Notation

Point wise Convergence $X_n \longrightarrow X$

Almost sure Convergence $X_n \longrightarrow X$

$X_n \longrightarrow X$

Convergence in probability $X_n \longrightarrow X$

Convergence in r^{th} moment $X_n \longrightarrow X$

$X_n \longrightarrow X$

$r=2$ $X_n \longrightarrow X$

Convergence in distribution $X_n \longrightarrow X$

Example: Assume that

$\Omega = [0,1]$ and $\mathbb{P}([a,b]) = b-a$

$\forall [a,b] \subseteq [0,1]$

Define

$$X_n(\omega) = \begin{cases} n & \text{if } \omega \in [0, 1/n^2] \\ 0 & \text{otherwise} \end{cases}$$

Determine if X_n converges in any sense.

Hierarchy of Modes of Convergence

An essential ~~[illegible]~~ issue in convergence of r.v.'s is whether a mode of convergence is stronger than another mode of convergence. In other words, does a mode of convergence imply another mode of convergence?

The following diagram summarizes a hierarchy of modes of convergence

$$
\begin{tikzcd}
u \arrow[r, Rightarrow] & \text{p.w.} \arrow[r, Rightarrow] & \text{a.s.} \arrow[dr, Rightarrow] & & & \\
& & & \text{i.p.} \arrow[r, Rightarrow] & d \arrow[r, Leftrightarrow] & \Phi_{X_n}(\omega) \to \Phi_X(\omega)\ \forall \omega \\
& & r^{th}\text{ moment } (r \geq 1) \arrow[ur, Rightarrow] \arrow[rr, Rightarrow] & & s^{th}\text{ moment } \arrow[ul, Rightarrow] &
\end{tikzcd}
$$

$r > s \geq 1$

Remark: It is often easier to prove convergence in r^{th} mean than convergence in probability. Also proving convergence in probability may sometimes be easier than proving convergence in distribution.

If asked to prove $\xrightarrow{P}$ check $\xrightarrow{m.s.}$ as well.

It may be easier to prove

Theorem (Skorokhod's Representation Theorem)

Assume that $\{X_n\}_{n\in\mathbb{N}}$, X are random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and $X_n \xrightarrow{d} X$. Then

there exists a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ and r.v.'s $\{Y_n\}_{n\in\mathbb{N}}$, Y on that probability space such that:

(a) $\forall n \in \mathbb{N} \quad F_{X_n}(x) = F_{Y_n}(x)$

$\forall x \in \mathbb{R}$

$$F_X(x) = F_Y(x)$$

(b) $Y_n \xrightarrow{a.s.} Y$

Proof: Omitted.

Theorem: (Continuous Mapping Theorem)

Assume that $g: \mathbb{R} \to \mathbb{R}$ is a continuous function. Then:

$$X_n \xrightarrow{a.s.} X \implies g(X_n) \xrightarrow{a.s.} g(X)$$

$$X_n \xrightarrow{P} X \implies g(X_n) \xrightarrow{P} g(X)$$

$$X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X)$$

Note: $X_n \xrightarrow{m.s.} X \not\Rightarrow g(X_n) \xrightarrow{m.s} g(X)$

Proof)

Theorem: ~~If~~ $X_n \xrightarrow{d} X$ iff for every bounded continuous function $g: \mathbb{R} \to \mathbb{R}$, we have $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$.

Proof: Uses continuous mapping theorem and is omitted.

Theorem: $X_n \xrightarrow{d} X$ iff $\varphi_{X_n}(\omega) \to \varphi_X(\omega)$ $\forall \omega \in \mathbb{R}$

Proof: The necessary condition can be proved using continuous mapping theorem. Sufficiency's proof is omitted.

Remark: This Theorem indicates that unlike other modes of convergence that are hierarchical, convergence in distribution implies pointwise convergence of characteristic functions and vice versa.

Recall that we say φ_{X_n} converges to φ_X pointwise when $n \to \infty$, if $\forall \omega \in \mathbb{R}$,

$$\lim_{n \to \infty} \varphi_{X_n}(\omega) = \varphi_X(\omega).$$

Theorem (Lévy's continuity/convergence Theorem): Assume that $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of r.v.'s associated with a sequence of characteristic functions $\{\varphi_{X_n}(\omega)\}_{n\in\mathbb{N}}$ and $\lim_{n\to\infty} \varphi_{X_n}(\omega) = \varphi(\omega)\ \forall\omega$

(i.e. $\varphi_{X_n}(\omega) \xrightarrow{\text{everywhere}} \varphi(\omega)$). Then:

(i) If $\varphi(\cdot)$ is continuous at $\omega=0$, φ is a valid Characteristic function of some random variable X, and $X_n \xrightarrow{d} X$

(ii) If $\varphi(\cdot)$ is not continuous at $\omega = 0$, X_n does not converge in distribution.

Remark: A characteristic function is continuous at $\omega = 0$, and that is the only condition that is needed to be checked in this theorem.

Example: Assume that

$X_n \sim \mathcal{N}(0, \sigma_n^2)$, and

$\sigma_n = \dfrac{n\sigma - 1}{n+1}$, in which $\sigma \in \mathbb{R}^+$.

Does $\{X_n\}$ converges in distribution?

Example: Assume that

$X_n \sim N(0, \sigma_n)$, where $\sigma_n = n+1$.

Does $\{X_n\}$ converge in distribution?

So, far, we saw that there is a hierarchy in convergence of r.v.'s. An interesting question is if there is a special case in which a weaker convergence mode implies a stronger convergence mode. Yes, there is.

Theorem: $X_n \xrightarrow{d} c \Rightarrow X_n \xrightarrow{i.p.} c$, where $c \in \mathbb{R}$.

Convergence of Functions of Multiple Converging Random Variables

Theorem: If $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ are sequences of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and X, Y are r.v.'s on $(\Omega, \mathcal{F}, \mathbb{P})$:

a) $\begin{cases} X_n \xrightarrow{a.s.} X \\ Y_n \xrightarrow{a.s.} Y \end{cases} \Longrightarrow X_n + Y_n \xrightarrow{a.s.} X + Y$

b) $\begin{cases} X_n \xrightarrow{i.p.} X \\ Y_n \xrightarrow{i.p.} Y \end{cases} \Longrightarrow X_n + Y_n \xrightarrow{i.p.} X + Y$

c) $\begin{cases} X_n \xrightarrow{r} X \\ Y_n \xrightarrow{r} Y \end{cases} \Longrightarrow X_n + Y_n \xrightarrow{r} X + Y$

Note: Even if $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$, in general, one can <u>NOT</u> conclude that $X_n + Y_n \xrightarrow{d} X + Y$.

Proof)

Slutsky's Theorem

Theorem (Slutsky's):

Let $\{X_n\}_{n\in\mathbb{N}}$, $\{Y_n\}_{n\in\mathbb{N}}$ be sequences of r.v.'s and

$X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$, $c\in\mathbb{R}$

then

a) $X_n + Y_n \xrightarrow{d} X + c$

b) $X_n Y_n \xrightarrow{d} cX$

c) $X_n / Y_n \xrightarrow{d} X/c$, if $c \neq 0$

Remark 1: In this theorem,

$Y_n \xrightarrow{d} c$ can be replaced with

$Y_n \xrightarrow{i.p.} c$.

Remark 2: The requirement that

Y_n converges to a "degenerate"

random variable is essential.

It doesn't work if Y_n converges

to a non-degenerate r.v.

Remark 3: The theorem remains

valid if all convergences in

distribution are replaced with convergence in probability.

Proof) Omitted.