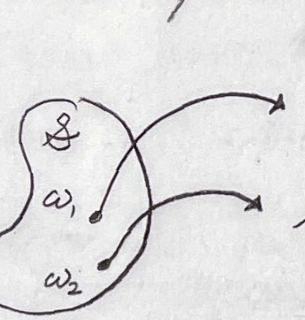
Stochastic Convergence

* Consider sequence of RVs, whether & how they converge



$X_1(\omega_1), X_2(\omega_1), \ldots$

$X_1(\omega_2), X_2(\omega_2), \ldots$

(Which ω in $\mathcal{S}$ occurs determines the entire infinite sequence in which we observe)

---

* Def.: **Random Sequence** ← Sequence of RVs defined on $(\mathcal{S}, \mathcal{F}, P)$, written as $\{X_n(\omega), n \geq 1\}$ for $\omega \in \mathcal{S}$

* Def.: **Convergence of a Sequence** (of Real Numbers $x_1, x_2 \ldots$) ← Sequence of real #s that converges to $x \in \mathbb{R}$ if $\forall \varepsilon > 0, \exists\, n_\varepsilon \in \mathbb{N} \ni |x_n - x| < \varepsilon \;\; \forall n \geq n_\varepsilon$

i.e., $\lim_{n\to\infty}(x_n) = x$ or $x_n \to x$ as $n \to \infty$

---

* for a random sequence, the dependance on $\omega \in \mathcal{S}$ makes convergence tricky

* Example:

Let $X_k = S + W_k$, where $S \in \mathbb{R}$ (sensor) & W_k (noise sequence) is a random sequence w/ $E[W_k] = 0 \;\; \forall k = 1, 2, \ldots$

(think of W_k as a noise sample when measuring ($S + W_k$) at time k)

(noise is commonly modeled as having zero mean, $E[W_k] = 0$)

* How do you find S?

(average out the n measurements)

(Y_n is sequence of averages, for n measurements)

$\Rightarrow$ Let $Y_n = \frac{1}{n}\sum_{k=1}^{n} X_k \;\therefore\; E[Y_n] = S, \forall n$ (but Y is a RV, so we cannot expect $Y_n = S \;\; \forall n$!)

* Y_n is a better estimate of S as $n \to \infty$

$E[Y_n] = S, \forall n \not\Leftrightarrow Y_n = S, \forall n$ (b/c Y_n is a random variable)

* Cannot guarentee since Y_n is not a RV, instead look at what happens as $n \to \infty$

$\Rightarrow$ Does Y_n converge to S as $n \to \infty$, & if so, in what sense?

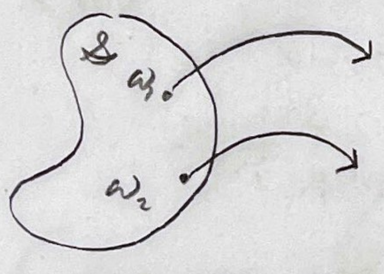
Types of Convergence

* Since $X_n(\omega)$ depends on $\omega \in \mathcal{S}$, what does it mean for X_n to converge?

* Def.: X_n Converges everywhere if $X_1(\omega), X_2(\omega), \ldots$

(1) Converges to some value $X(\omega)$, $\forall \omega \in \mathcal{S}$

(Also called "convergence surely" or "sure convergence")



$X_1(\omega_1), X_2(\omega_1), \ldots \to X(\omega_1)$

$X_1(\omega_2), X_2(\omega_2), \ldots \to X(\omega_2)$

Notation: $\boxed{X_n \xrightarrow{e} X \quad \text{OR} \quad X_n \longrightarrow X \text{ everywhere}}$ } (Shorthand for stochastic convergence)

(In practice it is not possible to know X for every ω)

* Def. : X_n Converges Almost Everywhere, or "almost surely",

(2) if $X_n(\omega) \longrightarrow X(\omega)$, $\forall \omega \in A$ for some $A \in \mathcal{F}$,

with probability of A, $P(A) = 1$

(a.k.a., convergance w/ probability 1)

Notation : $\boxed{X_n \xrightarrow{a.e.} X \text{ OR } X_n \xrightarrow{a.s.} X \text{ OR } X_n \xrightarrow{wp1} X}$

(type of convergence in the "strong law of large numbers")

(the dependence of ω makes the sequence & the limit different)

---

* Def. : X_n Converges in Mean Square to X if

(3) $\underbrace{E[|X_n - X|^2]}_{\text{(Is a sequence of Real Numbers)}} \rightarrow 0$ as $n \rightarrow \infty$

Notation : $\boxed{X_n \xrightarrow{ms} X}$

(requires both knowing the probability distribution in X_n & X (joint distribution))

---

* Def. : X_n Converges in Probability to X if $\forall \varepsilon > 0$,

(4) $\underbrace{P(|X_n - X| > \varepsilon)}_{\text{(Is a sequence of Real Numbers)}} \rightarrow 0$ as $n \rightarrow \infty$

Notation : $\boxed{X_n \xrightarrow{P} X}$

(type of sequence in the "weak law of large numbers")

Stochastic Convergence (Part 2)

* Def. : **X_n Converges in Distribution** to X if

(5) $F_{X_n}(x) \to F_X(x)$ as $n \to \infty$ at every

$x \in \mathbb{R}$ where F_X is continuous

(for a fixed $x \in \mathbb{R}$, $F_n(x)$ is a sequence of real numbers)

Notation : $\boxed{X_n \xrightarrow{d} X}$ "weak convergence"

---

* It is not the case that if $F_{X_n}(x) \to F_X(x)$ that

$f_{X_n}(x) \to f_X(x)$ where $f_X(x) = \frac{dF_X}{dx}(x)$ (at x)

---

* Example : Let X_n have pdf $f_{X_n}(x) = 1 + \cos(2\pi n x)$

for $0 \le x \le 1$, then

$$F_{X_n}(x) = \begin{cases} 0 & , x < 0 \\ x + \frac{\sin(2\pi n x)}{2\pi n} & , 0 \le x \le 1 \\ 1 & , x > 1 \end{cases}$$
(CDF)

so $F_{X_n}(x)$ converges to this function:

$$F_{X_n}(x) \longrightarrow \begin{cases} 0 & x < 0 \\ x & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$
← converges to a uniform RV on [0,1]

[however the sequence of pdfs will ONLY converge at points $x = 0$ & $x = 1$]

$f_{X_n}(x)$ does not converge for $x \in (0,1)$

Cauchy Criterion for M.S. Convergence

It is sometimes possible to show M.S. convergence, or $X_n \xrightarrow{m.s.} X$, w/o knowing what X is

---

* **Def.**: Convergence of a Sequence of Real Numbers x_n,

if $|x_{n+m} - x_n| \to 0$ as $n \to \infty$

$\forall m > 0$, then x_n is referred to as a <u>Cauchy Sequence</u>

$\Rightarrow$ Cauchy Criterion for Convergence: It can be shown that sequence x_n converges <u>iff</u> x_n is a Cauchy Sequence

---

Cauchy Criterion for MS Convergence of X_n:

It can be shown that X_n <u>converges in MS</u>

<u>iff</u> $E[|X_{n+m} - X_n|^2] \to 0$ as $n \to \infty$, $\forall m > 0$

knowing the joint distribution for X_{n+m} & X_n (i.e., joint distribution for 2 RVs)

It can be shown that the following relationships exist among <u>types of convergence</u>:

$$\begin{tikzcd} e \arrow[r, Rightarrow] & ae \arrow[r, Rightarrow] & p \arrow[r, Rightarrow] & d \\ & & ms \arrow[u, Rightarrow] & \end{tikzcd}$$

Chebyshev Inequality

HOLDS TRUE FOR ALL RANDOM VARIABLES

Let X be a RV w/ mean μ & variance σ^2, then

$$\forall \varepsilon > 0, \quad P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

Provide an upper bound on the probability that a RV is more than ε away from it's mean $(|X-\mu| \geq \varepsilon)$

---

*Proof : Let

$$g_1(x) = I_{\{x \in \mathbb{R} : |x-\mu| \geq \varepsilon\}}(x)$$

$$g_2(x) = \frac{(x-\mu)^2}{\varepsilon^2} \quad \text{for a fixed } \varepsilon > 0$$

Indicator function
zero for all points $(x-\mu)$ less than ε, & one for all points $(x-\mu)$ greater than ε

$\Rightarrow$ then

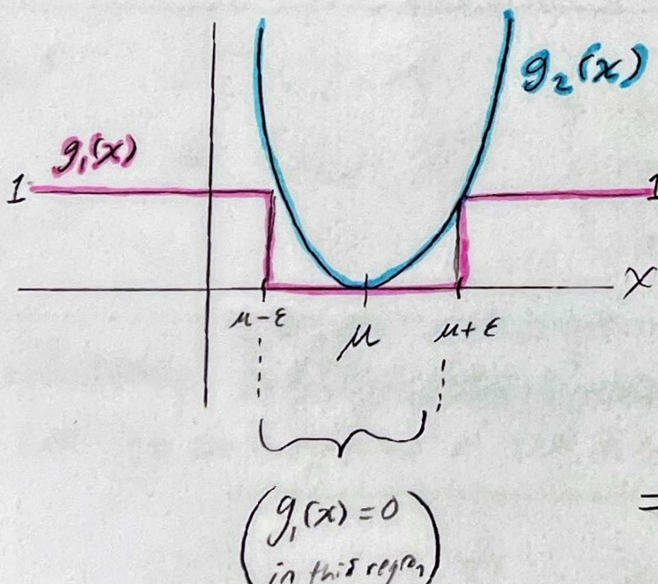
$$E[g_2(X)] = E\left[\frac{(x-\mu)^2}{\varepsilon^2}\right] = \frac{\sigma^2}{\varepsilon^2}$$

&

$$E[g_1(X)] = \int_{-\infty}^{\infty} g_1(x) f_X(x)\, dx = P(|X-\mu| \geq \varepsilon)$$

---

* Now consider : $\phi(x) = g_2(x) - g_1(x)$ ← Show that function is ≥ 0



$$\phi(x) \geq 0, \quad \forall x \in \mathbb{R}$$

$$\Rightarrow 0 \leq E[\phi(X)] = E[g_2(X)] - E[g_1(X)]$$

$$= \frac{\sigma^2}{\varepsilon} - P(|X-\mu| \geq \varepsilon)$$

$$\Rightarrow \boxed{P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}}$$

Stochastic Convergence (Part 3)

(constant Real Number signal)

* Recall Signal Example : $X_k = S + W_k$

(Sequence of Noise samples)

& $Y_n = \frac{1}{n}\sum_{k=1}^{n} X_k$ (noisy measurement of X_k's averaged)

(Expected value is S, iff samples are zero-mean)

---

* Does $Y_n \to S$ as $n \to \infty$ & if so, in what sense ?

---

* Theorem : **Weak Law of Large Numbers**

Let X_n be a sequence of i.i.d. Random Variables, with mean μ & variance σ^2

Let $Y_n = \frac{1}{n}\sum_{k=1}^{n} X_k$ ← (Sample mean of $X_1, \ldots, X_n$)

$$\Rightarrow \boxed{P(|Y_n - \mu| \geq \varepsilon) \to 0 \text{ as } n \to \infty, \quad \forall \varepsilon > 0}$$

(i.e., $Y_n \to \mu$ In Probability)

Proof : $E[Y_n] = \mu$, $Var(Y_n) = \frac{\sigma^2}{n}$ } chebyshev Inequality ⇒

(fixing n, using chebyshev for the result of 1 RV, NOT a sequence of RVs)

$$P(|Y_n - \mu| \geq \varepsilon) \leq \frac{Var(Y_n)}{\varepsilon^2} = \sigma^2/(\varepsilon^2 n) \to 0 \text{ as } n \to \infty$$

---

* Theorem : **Strong Law of Large Numbers**

Let X_n be a sequence of i.i.d. RVs w/ mean μ & variance σ^2, then :

$$\boxed{Y_n = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{a.e.} \mu \text{ as } n \to \infty}$$

* The LLNs can be used to characterize relative frequency estimates of probabilities too

→ consider a RV X & the event $\{X \in A\}$ for some $A \in \mathcal{F}$

→ estimate the probability that X is in A, $P(X \in A)$

→ draw n samples (take n measurements) of X to get RVs $X_1, X_2, \dots, X_n$ with the same distribution as X

now let

$$Y_k = \begin{cases} 1, & \text{if } X_k \in A \\ 0, & \text{if } X_k \notin A \end{cases}$$

then $E[Y_k] = P(X_k \in A) = P(X \in A)$

* since the X_k's are i.i.d., the Y_k's are also i.i.d. with mean $P(X \in A)$ & $\frac{1}{n}\sum_{k=1}^{n} Y_k$ is the relative frequency of $\{X_k \in A\}$

$\Rightarrow$ By the SLLN, $\frac{1}{n}\sum_{k=1}^{n} Y_k \xrightarrow{wp1} P(X \in A)$ as $n \to \infty$

* So wp1, relative frequency estimates converge to true probabilities defined by the measure P as $n \to \infty$

Central Limit Theorem

Let X_n be a sequence of iid RVs w/ finite mean μ & finite variance σ^2

Then if $Z_n = \dfrac{\sum_{j=1}^{n} X_j - n\mu}{\sigma\sqrt{n}}$ then $Z_n \xrightarrow{d} Z$ as $n \to \infty$

where Z is $N(0,1)$

$\Rightarrow$ This means that $F_{Z_n}(z) \longrightarrow \Phi(z) \equiv \displaystyle\int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx$

Proof: Using the following lemma: Let Z_n be a sequence of RVs w/ CDFs F_{Z_n} & MGFs $\Phi_{Z_n}(s)$, & Z is a RV w/ CDF F_Z & MGF Φ_Z

$\Rightarrow$ then if $\Phi_{Z_n}(s) \longrightarrow \Phi_Z(s)$, $\forall s \in \mathbb{R}$

Then $F_{Z_n}(z) \longrightarrow F_Z(z)$, $\forall z \in \mathbb{R}$

So it is sufficient to show that $\Phi_{Z_n}(s) \longrightarrow e^{s^2/2}$, $\forall s \in \mathbb{R}$

Consider MGF $\Phi_X\left(\frac{s}{\sqrt{n}}\right) = E\left[e^{sX_k/\sqrt{n}}\right], \forall k$

First look at the case $\mu = 0, \sigma^2 = 1$

Then $\Phi_{Z_n}(s) = E\left[e^{s\sum_{k=1}^{n} X_k/\sqrt{n}}\right] = E\left[\prod_{i=1}^{n} e^{sX_k/\sqrt{n}}\right]$

$$= \prod_{k=1}^{n} E\left[e^{sX_k/\sqrt{n}}\right] = \prod_{k=1}^{n} \Phi_X\left(\frac{s}{\sqrt{n}}\right)$$

$$= \left(\Phi_X\left(\frac{s}{\sqrt{n}}\right)\right)^n$$

* look at log of $\Phi_{Z_n}(s)$
* Let $L(s) = \log \Phi_X(s)$

remember convergence of distribution → fix rate of X

Then, to show that

$\Phi_{Z_n}(s) \longrightarrow e^{s^2/2}$, can show that $n \cdot L\left(\frac{s}{\sqrt{n}}\right)$

$$n \cdot L\left(\frac{s}{\sqrt{n}}\right) \equiv \log\left(\Phi_X\left(\frac{s}{\sqrt{n}}\right)\right)^n \longrightarrow \frac{s^2}{2}$$

use $\left.\begin{array}{l} L(0) = 0 \\ L'(0) = \mu = 0 \\ L''(0) = E[X^2] = 1 \end{array}\right\}$ Now $\lim_{n\to\infty}\left(nL\left(\frac{s}{\sqrt{n}}\right)\right)$

(from the moment theorem)

(then applying l'hopital's theorem)

for general μ, σ^2, use the result $\mu = 0, \sigma^2 = 1$ with $Y_k = \frac{X_k - \mu}{\sigma}$

Will need:

$$L(0) = 0$$

$$L'(0) = \mu = 0$$

$$L''(0) = E[X^2] = 1$$

Now

$$\lim_{n\to\infty} nL\left(\frac{s}{\sqrt{n}}\right) = \lim_{n\to\infty} \frac{L\left(\frac{s}{\sqrt{n}}\right)}{n^{-1}}$$

$$= \lim_{n\to\infty} \frac{-L'\left(\frac{s}{\sqrt{n}}\right) n^{-3/2} s}{-2n^{-2}} \quad \text{(l'hopital)}$$

$$= \lim_{n\to\infty} \frac{L'\left(\frac{s}{\sqrt{n}}\right) s}{2n^{-1/2}}$$

$$= \lim_{n\to\infty} \frac{-L''\left(\frac{s}{\sqrt{n}}\right) n^{-3/2} s^2}{-2n^{-3/2}} \quad \text{(l'hopital again)}$$

$$= \boxed{\lim_{n\to\infty} \frac{L''\left(\frac{s}{\sqrt{n}}\right) s^2}{2} = \frac{s^2}{2} \quad \therefore Z_n \xrightarrow{d} Z \text{ in Distribution}}$$

For general μ, σ^2, use result for $\mu = 0, \sigma^2 = 1$ with $Y_k = \frac{X_k - \mu}{\sigma}$

* Basic idea of TCLT is that summation of n i.i.d. RVs become gaussian as n gets large (with finite mean and variances)