05/12/23 Lecture-6 31/03/21

Ergodicity: It is similar to the Stationarity concept we have studied earlier.

→ Ergodicity is again the property our processes needs to have. The Timeline that connects the Design time with the working time. It is more sophisticated effect.
Basically, depends on the fact that Design time should be less than the working time.

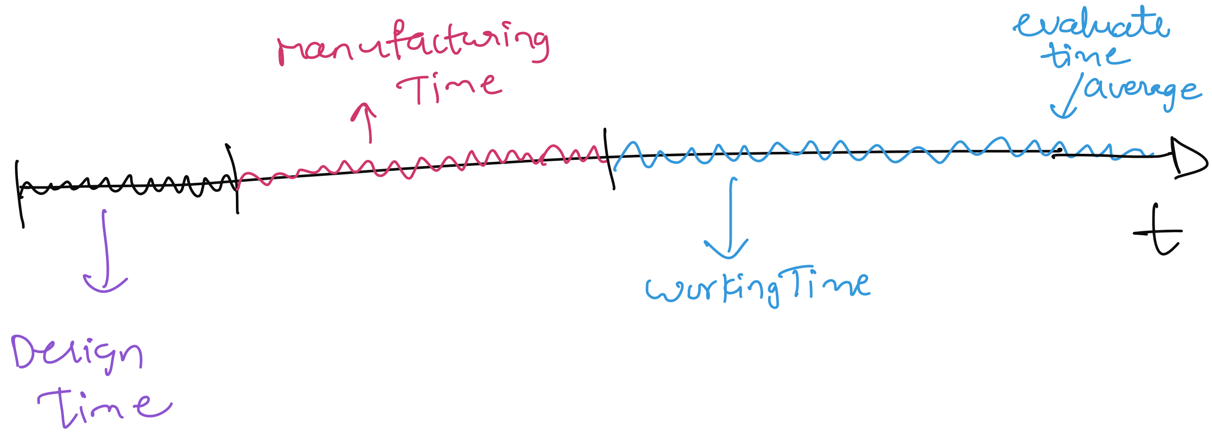
ex: If we spend 1 year Designing a Product we would like the Product to work like 5-10 years without any issues.

→ When in Design Time! The only thing we can relate to are statistical features. while during working time we are not sure that the Expectation is not exactly what we measure and performance of the Device. But we usually measure something long time.

* So we pay for is something which develops over long time.

* What we Design for is something which is Statistical characterisation.

ex: The Design and operation of a MicroProcessor



→ Since Design Time is generally shorter compared to working Time.

* If we want to measure the consumption of μp during Design time, even we could build some prototypes of my μp, since the design Time is short. I cannot test it extensively.

What we usually do, is to characterize it's Consumption is to say.

μp

20% Internal computation → W_{IC}

30% Addressing cache → W_{AC}

10% Serving Interrupts → W_{SI}

40% Addressing External Bus → W_{AD}

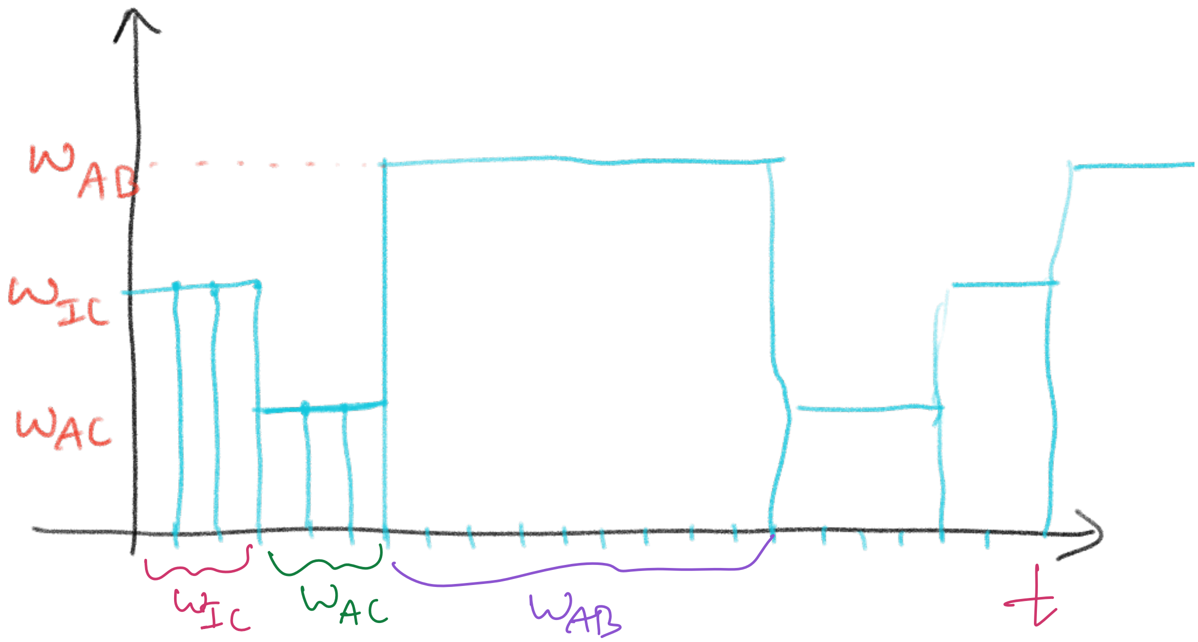
Average Consumption ← Average 'W' on Datasheet

$E[W]$ = Statistical Average

Note:
- But this won't be Average consumption over Longtime. It is because we had assumed that at certain % of time it will be one of the above four states.
- But 20% W_{IC} is a Probability of opening our μp and finding internal computation at similar to what we have mentioned above.

20% is just Probability, The reason is we cannot measure anything because our Design time is short.

→ Yet, when we deploy the μp we can have a situation like in the Graph below.



→ If we look at the Graph there is time 't' on x-axis i-e for certain no-of Time stamps (time interval) comes along with certain consumption..

So, what we really measure along time is some-thing which is Time Average

$w(t)$:

$$\frac{1}{N}\sum_{t=0}^{N-1} W(t) = \text{Time Average of consumption}$$

This is observation During The Working Time

→ E[w] we have seen above is a statistical Average
First Instance (During Design period)
because we computed it assuming that we open a μp to know what it is doing we get

20% probability of Internal computation.

Second Method: (During working period)

Here we are simply finding what our μp is doing at different Time stamps and Finding the Average w.r.to Time.

$$\therefore \text{Time Average} = \frac{1}{N} \sum_{t=0}^{N-1} W(t)$$

is different from E[W] (expectation)

Ideally,
We would like both E[w] & Time Average both to be equal. otherwise what we write on the Data sheet of μp is different from what our User experiences during Operation Time.

→ Therefore we need something that puts together Expectation E[w] which is statistical Average with Time Average. This is what is called

Birkov Ergodic Theorem

* The theorem has some assumptions that have to satisfied before I can state my results.

The Assumption is if our process of consumption is Ergodic then time Time Average is going to be equal to the Expectation. (atleast Asymptotically)

If $X(t)$ is Ergodic:

$$\boxed{E[w] = \underbrace{\lim_{N \to \infty} \frac{1}{N} \sum_{t=0}^{N-1} w(t)}_{\text{Time Average experiend by the user}}}$$

- Firstly, To be Ergodic $X(t)$ must be Stationary.

→ Before explaining what Ergodicity is we have to link it with Stationarity

For sure Ergodicity means stationarity i.e if we didn't have stationarity. If statistical feature at design time are different from Running Time, The Expectation (That we calculate) would be different from everything else! (even from Time Average)

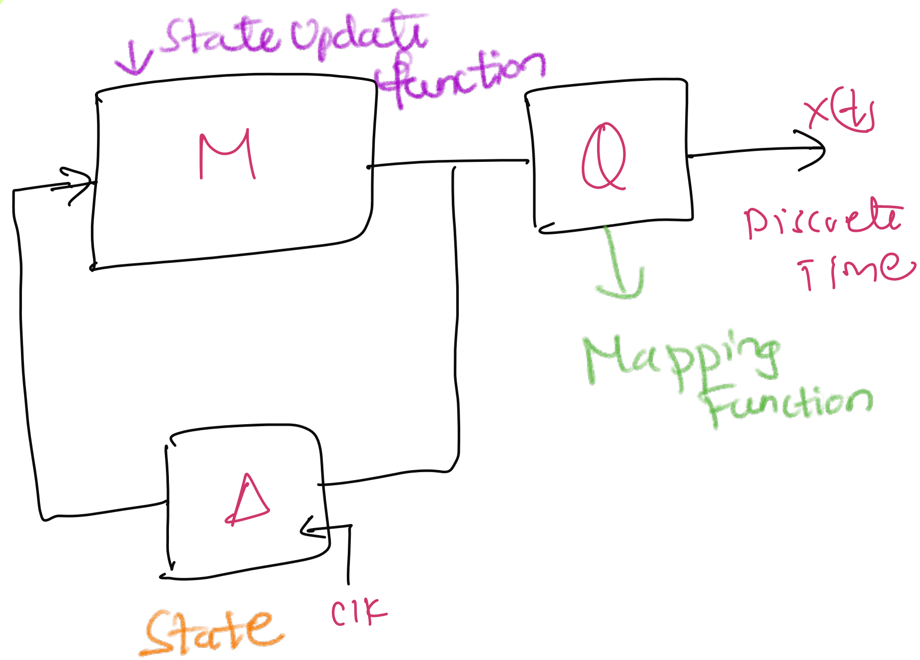
∴ Stationarity is pre-requisite of Ergodicity.

Assuming we are dealing with Discrete Time process something which is clocked i.e our consumption of I/p.

*** For clocked processes, stationarity brings wonderful structural property of Stochastic Process.

→ There is theorem which is power, which says

* For every discrete Time stochastic process which is stationary we can make a Machine which is Deterministic with nothing random in it. when initialized with Random inputs produces our process.



Ω : current state (Mathematical Entity
not a Number)

M : State update function
(define NS starting from current one)

Q : Mapping function
(Translate state to observables)

ex: waveform onto an Oscilloscope

Only thing that is Random here is the Initial state

Formally:

$M : \underbrace{\Omega}_{\text{state space}} \to \Omega$

(i) $\omega(t+1) = M(\omega(t))$

(ii) $X(t) = Q(\omega(t))$

(iii) $\omega(0)$ is Random

ex: (i) Random processes can be determined by Deterministic processes is the core of all Psuedo Random Number Generators

One of the most adopted method is Congruential Generators.

$w(0)$ Integer

$$\underbrace{w(t+1) = [a\, w(t) + b] \pmod p}_{M}$$

a, b, p are proper integers

Choosing a, b, p carefully we can choose good Psuedo Random Numbers.

② Example of Coin Tossing Stationary clocked Machine

Here we will see a Deterministic Machine producing Head/Tail similar to that of Tossing a Coin

Steps

(i) * $\Omega = [0, 1]$

State Space (collection of all possible states)

(ii) * State update function M

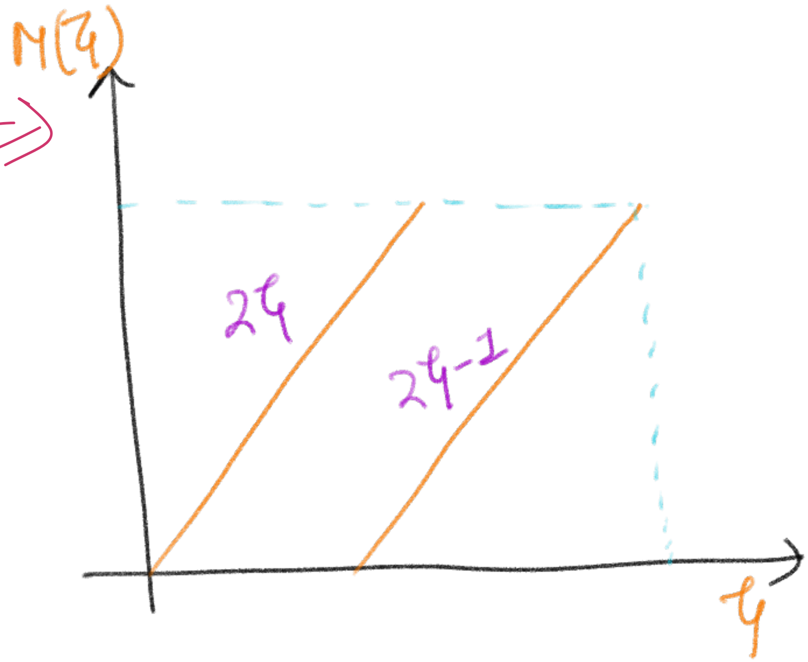
$$M(\xi) = \begin{cases} 2\xi & \xi \in [0, \frac{1}{2}[\\ 2\xi - 1 & \xi \in [\frac{1}{2}, 1] \end{cases}$$

(iii) *

$$Q(\xi) = \begin{cases} \text{Head} & \text{if } \xi \in [0, \frac{1}{2}[\\ \text{Tail} & \text{if } \xi \in [\frac{1}{2}, 1] \end{cases}$$

$w(0) \Longrightarrow$ Random Number uniformily distributed in $[0, 1[$

$M(\xi) \Longrightarrow$



→ $w(0)$ is uniformly distributed in $[0,1[$

According to $\mathbb{Q}(q)$ we decided that if

$q \in [0, \frac{1}{2}[$ we have Heads

$q \in [\frac{1}{2}, 1]$ we have Tails

A This gives a sequence of Head/Tail

Q) How do we know this sequence of H/T is Random as tossing an Independent coin?

Ans The Trick is here to take $w(0)$ which is the only Random Component of the s/s and it will be uniformly distributed.

Binary expansion of $w(0)$ '1' excluded : 0.0110101110 . . .

$\frac{1}{4} + \frac{1}{8} + \frac{1}{32} + \frac{1}{128} + \cdots$

Since, we assumed $w(0)$ to be completely Random and all numbers are equally probable and now $w(0)$ is going to a string of bits i.e uniformly identical & Independent sequence of Random Variables

Q) What happens when we apply our Machinery to our initial condition?

i.e To apply the state update function

check first bit

Any Binary expansion: $0 \cdot 0110101110 \ldots$

$\because$ '0' $\rightarrow$ first bit $2q \longrightarrow 0 \cdot 110101110 \ldots$ "Tail"

when we multiply by "2", we will shift to the left

$\because$ '1' first bit

$2q-1 \rightarrow 0 \cdot 10101110 \ldots$ "Tail"

$2q-1 \rightarrow 0 \cdot 0101110 \ldots$ "Head"

we will get Binary digits of $w(0)$ one after the other

Ergodicity depends on M :

Q) What characterises, what says our process is made of?

Ans → It's not our mapping function ('ϕ')
But the true evolution of the process is decided by the state update function M, so Ergodicity is a process of 'M'

→ To check whether a process is $\boxed{\text{Ergodic or not}}$ since we are dealing with stochastic discrete process. We will find the machine and we put some requirements on M

Conditions on 'M' to be Ergodic:

Since we have to draw at random 'ω' we assign some probability to 'ω'

$$\omega \in \Omega$$

it is $P(\omega)$ ← Probability

$$P(\omega) \geq 0 \qquad P(\Omega) = 1$$

Q) How we should read this probability?

1) If we stop this machine with M, Δ, Q at certain point A we want to know that the current state is in the subset of the ('A') state space (Ω)

2) Then There is invariant iff $M^{-1}(A) = A$

set of counter images

$$\boxed{M^{-1}(A) = \{\omega \mid M(\omega) \in A\} = A}$$

Note: 'M' need not be invertible, but it can be invertible at set level.

** The counter images of 'A' are set of states that at next time stamp will happen to be in 'A'

3) The Nature of 'M' must be measure preserving.

'M' is measure preserving iff

$$\boxed{P(M^{-1}(A)) = P(A)}$$

Similar to "Mass preservation eqn"

Now that we have counter image concept

Q) What is the $P(M^{-1}(A))$?

This is an abstract Requirement.

example of paint Mixing:

To choose a colour → Then is a white paint & we add some specific colour to get our chosen colour.

A During the Mixing process (No paint was thrown out & no paint entered the Bucket)

So, if we track down the molecules of the drops of colour paints we added to white paint we can find them in the Bucket -

Q) What is the probability that we are in the subset of the state space ?
(or)
What is the probability that we are in the sub Region of the Bucket ?

we are asking that the measure is same (preserved)

i.e we don't want states to be destroyed or don't want states to be added while we are continuing in the EVOLUTION.

Now that we have INVARIANT & MEASURE PRESERVING concepts we have everything

- to define **ERGODICITY**

ERGODICITY: our process which is a DTSS is ergodic Iff

$\forall A \subseteq \Omega$ that is invariant

is either $P(A) = 1$ or $P(A) = 0$

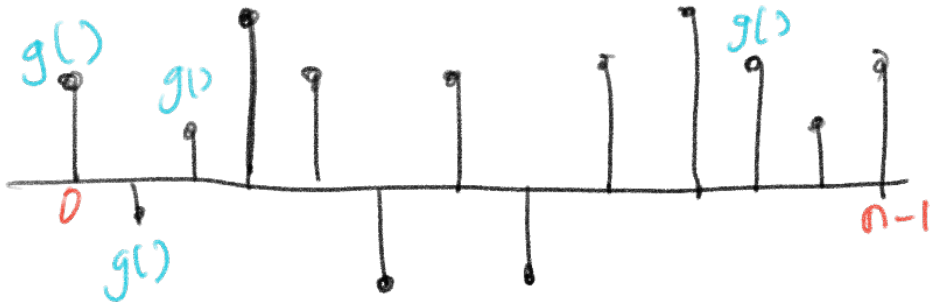
$\Downarrow$ Ergodic Theorem

Statistical Average = Time Average

$$\boxed{\lim_{n \to \infty} \frac{1}{n} \sum_{t=0}^{n-1} g(x(t)) = E[g(x(0))]}$$

Value we observe in the process which @time=t should be Measurable (that we should be put into an Integral)

- This requirement is Fundamental



∴ for any function 'g'

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} g(x(t)) = E[g(X(0))]$$

we are taking expectation of X(0) & statistics of X(0) is equal to any other process because the process is stationary.

X(0) is equally is equally representative of X(100) or X(1000) anything in

X(0) is Random X(0) = 2[W(0)]
for example

This all looks Trivial but lets look at an example

Ex: Trivial Example

let $\Omega = \{0, 1, 2, 3\}$

$\frac{1}{4} \quad \frac{1}{4} \quad \frac{1}{4} \quad \frac{1}{4}$ → Probability of each variable

$M : \Omega \to \Omega$

ω	$M(\omega)$
0	1
1	2
2	3
3	0

$Q : \Omega \to \mathbb{R}$ → $Q(q) = q$

$P : \Omega \to \mathbb{R}^+$ → $p(\omega) = \frac{1}{4}$

$g : \mathbb{R} \to \mathbb{R}$ → $g(q) = q$

let's assume at $t=0$ we have Random Initial condition.

(time stamp)

t	$w(t) = x(t)$
0	1 → Random initial condition
1	2
2	3
3	4
4	5
5	6
6	7
⋮	⋮

→ let's make Time Average

$$\frac{1}{n}\sum_{t=0}^{n-1} g(x(t)) = \frac{1}{n}\left(1+2+3+0+1+2+3+\cdots\right.$$
$$\left.---\right)$$

$$\lim_{n\to\infty} \Rightarrow \frac{6}{4}$$

Now Expectation (Statistical Average)

$$E[x(0)] = 0\cdot\frac{1}{4} + 1\cdot\frac{1}{4} + 2\cdot\frac{1}{4} + 3\cdot\frac{1}{4}$$

$$= \frac{6}{4}$$

for $n = 100$

Time Average $\Rightarrow \dfrac{(1+2+3+0)\times 25}{100} \Rightarrow \dfrac{6\times \cancel{25}}{\cancel{100}_{4}}$

for $n = 1000$

$$\Rightarrow \frac{(1+2+3+0)250}{1000} \Rightarrow \frac{6}{4}$$

$\therefore$ for $n = \infty \implies \frac{6}{4}$

we calculated $\boxed{\text{Time Average} = \text{Statistical Average}}$

Q) Does this satisfy Ergodicity?

is it Measure preserving?

Invariant?

Ans

w	$M(w)$
0	1
1	2
2	3
3	0

$M^{-1}(A) \{ 0, 1$; $\{1, 2\} A$

Measure preserving $\implies A \subseteq \Omega$

$A = \{1, 2\}$, $M^{-1}(A) = \{0, 1\}$

$P(\{1, 2\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

$P(M^{-1}(A)) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

$P(\{0, 1\})$

They are EQUAL

Invariant property

look for $A \subseteq \Omega$

such that $M^{-1}(A) = A$

only two invariants $A' = \{0,1,2,3\}$ — whole setspace

$A'' = \{\}$ — empty set

only these two sets are invariant

→ Counter example

ω	$M(\omega)$
0	1
1	0
2	3
3	2

No invariant property

$A = \{0, 1\}$ $\quad M^{-1}\{(0,1)\} = \{0,1\}$

Here $P(A) = P(\{0,1\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

This is not equal to

$\boxed{1 \text{ (or) } 0}$

$\therefore$ The above process is NOT ERGODIC!

Interestingly

Time stamps	$X(t) = w(t)$
0	1 ← Random Initial condition
1	0
2	1
3	0
4	1
5	0
6	1
7	0
8	1
	0
	⋮

Here we can't see 2,3 which are part of the state space

The Statistical Average won't change

$$= E[g(X(t))]$$

$$= 0\cdot\frac{1}{4} + 1\cdot\frac{1}{4} + 2\cdot\frac{1}{4} + 3\cdot\frac{1}{4}$$

$$\Rightarrow \frac{6}{4}$$

But Time Average

$$= \lim_{n\to\infty} \frac{1}{n}\sum_{n=0}^{n-1} g(x(t))$$

$$\Rightarrow \lim_{n\to\infty} \frac{1}{n}\left(0+1+0+1+0+1+0+1\dots\right)$$

$$\Rightarrow \frac{1}{2}$$

$$\boxed{\text{Here, Statistical Average} \neq \text{Time Average}}$$

Average on corner cases

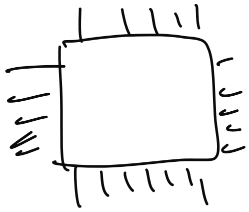
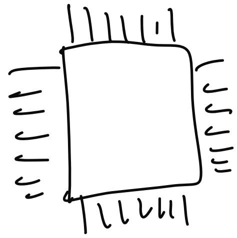
Design → Statistical Averages → low

Datasheet

If the things are not Ergodic, we can face situations like this

Fabrication Facility 1

Fabrication Facility 2

* we are designing with Statistical Averages. Here it is done with Simulating the corners where the (1) Fastest & Slowest transistors that's going to be build (2) what is the lowest & highest Threshold found in a chip

• All the above parameters impact the Power consumption of the processor.

→ So we need to do to make a design based on all the statistical features of the o/p. and we determine with these statistical what is the power consumed.

Then we take the design and send to Fabrication facilities, but they may not be similar to 100% in Fabricating the same node chip.

Fab 1	Fab 2
1) has Lower Threshold Chips for same node Technology	1) Higher Threshold Chips for same node Technology

more power is consumed for Lower V_{TH}

less power is consumed for Higher V_{TH}

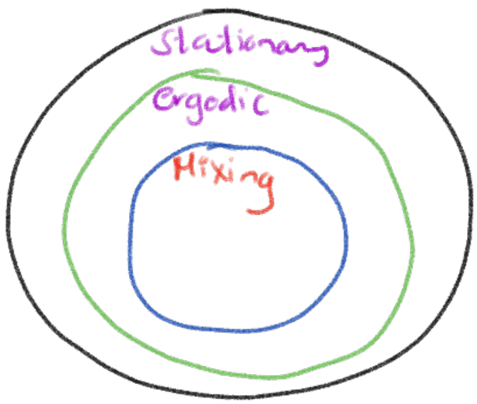
→ So the client buying the chips is not satisfied with low V_{TH} chips & satisfied with High V_{TH} Chips

* Here in the Statistical Average calculation we calculate implicitly make average b/w chips of Fab1 & Fab2

* But when we analyze in Real world we can analyze only one chip it can be from any of the Fabs.

Then the process will not be Ergodic.

** → we want our process to be



"But Ergodicity is NOT Enough due to Engineering Considerations."

$$\boxed{\lim_{n \to \infty} \frac{1}{n} \sum_{t=0}^{n-1} g(x(t)) = E\left[g(x(0))\right]}$$

Ideally Engineering doesn't like Infinity

$$\boxed{\left| \underbrace{\frac{1}{n} \sum_{t=0}^{n-1} g(x(t))}_{\text{Approximate for finite value of 'n' what we can measure}} - \underbrace{E\left[g(x(0))\right]}_{\text{we can compute This}} \right. = \underbrace{\alpha(n)}_{\text{Error}}}$$

→ We know the Error for $n \to \infty$ will be zero

Q) But How fast does it happen?

Ans Ideally, the error $\Rightarrow 0$ for small value of 'n'

* otherwise the clients of the device will experience a performance which is equal to the efficient Average after a very long time This is something we would like to avoid.

There is a technique to achieve Good performance very fast it is called Mixing.

→ Mixing:

Are the process which are stationary & Ergodicity

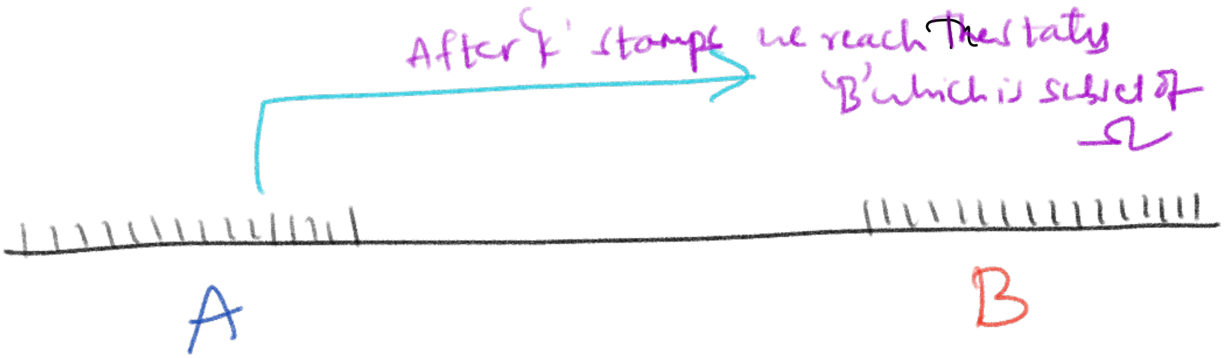
A process is Mixing if $\forall\ A, B \subseteq \Omega$

$$\boxed{\lim_{K \to \infty} P\left(A \cap M^{-K}(B)\right) = P(A)\,P(B)}$$

This implies we

have independence b/w A,B

$$M^{-k}(B) = M^{-1}(M^{-1}(M^{-1}(M^{-1}(\dots\dots M^{-1}(B)))))$$



$$A, B \subseteq \Omega$$

→ The dependencies b/w two events that are distant in time tends to vanish when the amount of time increases

ex: It is much easier to predict temperature of Tomorrow from temperature of Today.

But calculating Temperature after 6 months would be an independent from Today's temperature.

* As Time passes Dependencies
Vanish → 0

ex: As in Mixing of different set of paint colour to white paint to get our desired colour.

* It doesn't matter, if we add the colour Drops to the Borders of Buckets or the centre of the Bucket but

* Once we are done with the Mixing problem we will see an Uniform colour of our choice because Mixing has destroyed our initial condition of where we have put our colour drops.

For Mixing processes $\alpha(n)$ vanishes Exponentially Fast!

6