Random process

- A random process or a stochastic process is a collection of random variables: $\{X_t,\ t \in T\}$
- Markov chain is an example. Here $T = \{0, 1, \cdots\}$
- We call T the index set.
- Normally, T is either (a subset of) set of integers or an interval on real line.
- We think of the index t as time
- Thus a random process can represent the time-evolution of the state of a system or a random signal
- We assume T is infinite
- The index need not necessarily represent time. It can represent, for example, space coordinates.

- A random process: $\{X_t,\ t \in T\}$
- The set T can be countable e.g., $T = \{0, 1, 2, \cdots\}$
- Or, T can be continuous e.g., $T = [0, \infty)$
- These are termed **discrete-time** or **continuous-time** processes
- The random variables, X_t, may be discrete or continuous
- These are termed **discrete-state** or **continuous-state** processes
- The Markov chain we considered is a discrete-time discrete-state process

- A random process: $\{X_t,\ t \in T\}$
- We can think of this as a mapping: $X : \Omega \times T \to \Re$
- $X(\cdot, t)$ is a random variable and
  $X(\omega, \cdot)$ is a real-valued function over T
- So, we can think of the process also as a collection of time functions.
- X can be thought of as a map that associates with each $\omega \in \Omega$ a real-valued function on T.
- These functions are called sample paths or paths of the process
- We can view the random process as a collection of random variables, or as a collection of functions
- We will denote the random variables as X_t or $X(t)$

- A finite collection of random variables is completely specified by its joint distribution
- How do we characterize a random process?
- We need to specify joint distribution of $X_{t_1}, X_{t_2}, \cdots X_{t_n}$ for all n and all $t_1, t_2, \cdots t_n \in T$.
- The specified distributions should be *consistent* with each other
- One can show this completely specifies the process.
- As we saw, for a Markov chain, π_0 and P together specify all such joint distributions

Distributions of a random process

- A random process: $\{X_t,\ t \in T\}$ or $X : \Omega \times T \to \Re$
- The first order distributions function of X is

$$F_X(x; t) = Pr[X_t \leq x] = F_{X_t}(x)$$

- The second order distribution function of X is

$$F_X(x_1, x_2; t_1, t_2) = Pr[X_{t_1} \leq x_1,\ X_{t_2} \leq x_2] = F_{X_{t_1} X_{t_2}}(x_1, x_2)$$

- This is an abuse of notation!
- First order df is a function from $\Re \times T$ to $[0, 1]$
  Second order df is a function from $\Re^2 \times T^2$ to $[0, 1]$
  We are using the same symbol, F_X for all.

Distributions of a random process

- A random process: $\{X_t,\ t \in T\}$ or $X : \Omega \times T \to \Re$
- The first order distributions:

$$F_X(x;t) = Pr[X_t \leq x] = F_{X_t}(x)$$

- The second order distributions:

$$F_X(x_1, x_2; t_1, t_2) = Pr[X_{t_1} \leq x_1,\ X_{t_2} \leq x_2]$$

- The n^{th} order distributions:

$$F_X(x_1, \cdots, x_n; t_1, \cdots t_n) = Pr[X_{t_i} \leq x_i,\ i = 1, \cdots, n]$$

- When it is a discrete-state process, all X_t would be discrete random variables
- We can specify distributions through mass functions:

$$f_X(x;t) = Pr[X_t = x] = f_{X_t}(x)$$

$$f_X(x_1, x_2; t_1, t_2) = Pr[X_{t_1} = x_1,\ X_{t_2} = x_2]$$

$$f_X(x_1, \cdots, x_n; t_1, \cdots t_n) = Pr[X_{t_i} = x_i,\ i = 1, \cdots, n]$$

- If all X_t are continuous random variables and if all distributions have density functions, then we denote joint density of $X_{t_1}, \cdots, X_{t_n}$ by $f_X(x_1, \cdots, x_n; t_1, \cdots t_n)$

- Specifying the n^{th} order distributions for all n separately is not feasible.
- Hence one needs some assumptions on the model so that these are specified implicitly.
- One example is the Markovian assumption.
- As we saw, in a Markov chain, the transition probabilities and initial state probabilities would determine all the distributions
- Another such useful assumption is what is called a process with independent increments

- A random process $\{X(t),\ t \in T\}$ is said to be a process with independent increments if

  for all $t_1 < t_2 \leq t_3 < t_4$, *the random variables* $X(t_2) - X(t_1)$ *and* $X(t_4) - X(t_3)$ *are independent*

- We take this to also imply $X(t_1)$ is independent of $X(t_2) - X(t_1)$ for all $t_1 < t_2$.
- Now suppose this is a discrete-state process.
- Then we can write n^{th} order pmf's as

$$
\begin{aligned}
& Pr[X(t_1) = x_1, X(t_2) = x_2, \cdots X(t_n) = x_n] \\
&\quad = Pr[X(t_1) = x_1, X(t_2) - X(t_1) = x_2 - x_1, \cdots] \\
&\quad = Pr[X(t_1) = x_1]\ Pr[X(t_2) - X(t_1) = x_2 - x_1] \cdots \\
&\qquad\qquad \cdots Pr[X(t_n) - X(t_{n-1}) = x_n - x_{n-1}]
\end{aligned}
$$

- We only need up to second order distributions

- Let $\{X(t),\ t \in T\}$ be a discrete-state process with independent increments
- Then we specify $f_X(x;t)$ and another function

$$g(x_1, x_2; t_1, t_2) = Pr[X(t_2) - X(t_1) = x_2 - x_1]$$

- Now we can get all distributions as

$$\begin{aligned} & f_X(x_1, \cdots, x_n; t_1, \cdots t_n) \\ & \quad = Pr[X(t_i) = x_i,\ i = 1, \cdots, n] \\ & \quad = f_X(x_1; t_1) \prod_{i=1}^{n-1} Pr[X(t_{i+1}) - X(t_i) = x_{i+1} - x_i] \\ & \quad = f_X(x_1; t_1) \prod_{i=1}^{n-1} g(x_i, x_{i+1}; t_i, t_{i+1}) \end{aligned}$$

- $\{X(t),\ t \in T\}$ is called a Gaussian process if all finite dimensional distributions are Gaussian.
- Then, $X_{t_1}, \cdots, X_{t_n}$ are jointly Gaussian for all $t_1, \cdots, t_n$
- This Gaussian density is completely specified if we know all means and covariances.
- Hence all we need are two functions:

$$\eta(t) = E[X(t)], \qquad C(t_1, t_2) = \mathsf{Cov}(X_{t_1}, X_{t_2})$$

- Another example of how all finite dimensional distributions of a process can be specified.

- Given a random process $\{X(t),\ t \in T\}$
- Its mean or mean function is defined by

$$\eta_X(t) = E[X(t)],\ \ t \in T$$

- We define the autocorrelation of the process by

$$R_X(t_1, t_2) = E[X(t_1)X(t_2)]$$

- We define the autocovariance of the process by

$$\begin{aligned} C_X(t_1, t_2) &= E\left[(X(t_1) - E[X(t_1)])(X(t_2) - E[X(t_2)])\right] \\ &= R_X(t_1, t_2) - \eta_X(t_1)\eta_X(t_2) \end{aligned}$$

Stationary Processes

- A random process $\{X(t),\ t \in T\}$ is said to be stationary if

  for all n*, for all* $t_1, \cdots, t_n$*, for all* $x_1, \cdots x_n$ *and for all* τ *we have*

$$F_X(x_1, \cdots, x_n \,;\, t_1, \cdots, t_n) = F_X(x_1, \cdots, x_n \,;\, t_1+\tau, \cdots, t_n+\tau)$$

- For a stationary process, the distributions are unaffected by translation of the time axis.
- This is a rather stringent condition and is often referred to as strict-sense stationarity

- A homogeneous Markov chain started in its stationary distribution is a stationary process
- As we know, if π_0 is the stationary distribution then π_n is same for all n.
- This, along with the Markov condition would imply that shift of time origin does not affect the distributions

$$\begin{aligned}
Pr[X_n = x_0, X_{n+1} = x_1, \cdots X_{n+m} = x_m]& \\
&= \pi_n(x_0)P(x_0, x_1) \cdots P(x_{m-1}, x_m) \\
&= \pi_0(x_0)P(x_0, x_1) \cdots P(x_{m-1}, x_m) \\
&= Pr[X_0 = x_0, X_1 = x_1, \cdots X_m = x_m]
\end{aligned}$$

- Suppose $\{X(t),\ t \in T\}$ is (strict-sense) stationary
- Then the first order distribution is independent of time

$$F_X(x;t) = F_X(x;t+\tau),\ \forall x,t,\tau \quad \Rightarrow \quad \text{e.g.,} \quad F_X(x;t) = F_X(x;0)$$

- This implies $\eta_X(t) = \eta_X$, a constant
- The second order distribution has to satisfy

$$F_X(x_1,x_2;t,t+\tau) = F_X(x_1,x_2;0,\tau),\ \forall x_1,x_2,t,\tau$$

  Hence $F_X(x_1,x_2;t_1,t_2)$ can depend only on $t_1 - t_2$
- This implies

$$R_X(t,t+\tau) = E[X(t)X(t+\tau)] = R_X(\tau)$$

  Autocorrelation depends only on the time difference

- The process $\{X(t),\ t \in T\}$ is said to be wide-sense stationary if

$$\begin{aligned} F_X(x;t) &= F_X(x;t+\tau),\ \forall x,t,\tau \\ F_X(x_1,x_2;t_1,t_2) &= F_X(x_1,x_2;t_1+\tau,t_2+\tau) \end{aligned}$$

- The process is wide-sense stationary if the first and second order distributions are invariant to translation of time origin
- This is the definition normally used in mathematics books.

- Let $\{X(t),\ t \in T\}$ be wide-sense stationary. Then

1. $\eta_X(t) = \eta_X$, a constant
2. $R_X(t_1, t_2)$ depends only on $t_1 - t_2$

- In many engineering applications, we call a process wide-sense stationary if the above two hold.
- In this course we take the above as the definition of wide-sense stationary process
- When the process is wide-sense stationary, we write autocorrelation as

$$R_X(\tau) = E[X(t)X(t+\tau)]$$

- Let $\{X(t),\ t \in T\}$ be wide-sense stationary.
- Then $R_X(t_1, t_2) = E\left[X_{t_1} X_{t_2}\right]$ depends only on $t_1 - t_2$.
- So, we represent it as $R_X(\tau)$ where we take $\tau = t_2 - t_1$.

$$E\left[X_{t_1} X_{t_2}\right] = E\left[X_{t_2} X_{t_1}\right] \quad \Rightarrow \quad R_X(\tau) = R_X(-\tau)$$

The autocorrelation function of a wide-sense stationary process is a symmetric function of (single) time variable.

- Let $\{X(t),\ t \in T\}$ be wide-sense stationary.
- Let $E[X(t)] = \eta_x, \forall t$.
- For any t, $E[X^2(t)] = E[X(t)X(t)] = R_X(0)$.
- Hence, $\text{Var}(X(t)) = E[X^2(t)] - \eta_x^2 = R_X(0) - \eta_x^2$
- We also have
  $\text{Cov}(X(t), X(t+\tau)) = C_X(\tau) = R_X(\tau) - \eta_x^2$

$$\begin{aligned} \text{Var}(X(t) - X(t+\tau)) &= \text{Var}(X(t)) + \text{Var}(X(t+\tau)) \\ &\quad -2\text{Cov}(X(t), X(t+\tau)) \\ &= 2(R_X(0) - \eta_x^2) - 2(R_X(\tau) - \eta_x^2) \end{aligned}$$

This gives: $R_X(0) \geq R_X(\tau)$.

Power Spectral Density

- Let $\{X(t),\ t \in T\}$ be wide-sense stationary.
- Then the Fourier Transform of $R_X(\tau)$ is called the Power Spectral density

$$S_X(f) = \int_{-\infty}^{\infty} R_X(\tau)\, e^{-j2\pi f\tau}\, d\tau$$

- Since $R_X(\tau)$ is symmetric, $S_X(f)$ is real.
- It plays an important role in analysis of linear systems with random inputs

Ergodicity

- Suppose $X(n)$ is a discrete-time discrete-state process
- Suppose it is wide-sense stationary.
  Then $E[X(n)] = \eta_X$ for all n
- Ergodicity is the question of

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X(i) \stackrel{?}{=} E[X(n)] = \eta_X$$

- The question is : do 'time-averages' converge to 'ensemble-averages'
- The process is wide-sense stationary and hence all $X(i)$ have the same distribution. Hence it is similar to law of large numbers.
- But, thus is not implied by law of large numbers; $X(i)$ need not be uncorrelated (e.g., Markov chain)

- Ergodicity is a question of whether time-averages converge to ensemble-averages?

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X(i) \stackrel{?}{=} E[X(n)] = \eta_X$$

Or, more generally

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g(X(i)) \stackrel{?}{=} E[g(X(n))]$$

- We proved that this is true for an irreducible, aperiodic, positive recurrent Markov chain (with a finite state space)
- Such Markov chains are called ergodic chains.

- Ergodicity: do time-averages converge to ensemble-averages?

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X(i) \overset{?}{=} E[X(n)] = \eta_X$$

Or, more generally

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g(X(i)) \overset{?}{=} E[g(X(n))]$$

For a continuous time process we can write this as

$$\lim_{\tau\to\infty} \frac{1}{2\tau} \int_{-\tau}^{\tau} X(t)\, dt \overset{?}{=} E[X(t)] = \eta_X$$

- This may be true if correlation between $X(i)$ is 'weak'
- One sufficient condition could be that covariance between $X(t)$ and $X(t+\tau)$ decreases fast with increasing τ.

- Define
$$\eta_\tau = \frac{1}{2\tau} \int_{-\tau}^{\tau} X(t)\, dt \quad (\tau > 0)$$
- For each τ, η_τ is a rv. We write η for η_X.
- We say the process is mean-ergodic if
$$\eta_\tau \xrightarrow{P} \eta, \quad \text{as} \;\; \tau \to \infty$$
- That is, if
$$\lim_{\tau \to \infty} Pr\left[|\eta_\tau - \eta| > \epsilon \right] = 0, \;\; \forall \epsilon > 0$$
- Note that $E[\eta_\tau] = \eta, \; \forall \tau$.
- Hence it is enough if we show
$$\sigma_\tau^2 \triangleq E\left[(\eta_\tau - \eta)^2 \right] \; \to \; 0, \quad \text{as} \quad \tau \to \infty$$

- Let $C_X(t_1, t_2)$ be the autocovariance of the process

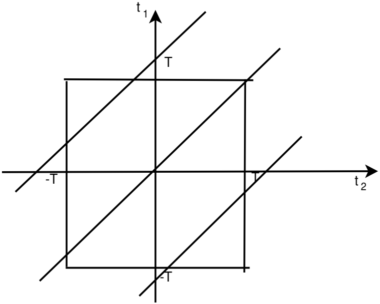
$$C_X(t_1, t_2) = E[(X(t_1) - \eta)(X(t_2) - \eta)]$$

- Assuming wide-sense stationarity, $C_X(t_1, t_2) = C_X(t_1 - t_2)$
- We can get σ_τ^2 as

$$\begin{aligned}
\sigma_\tau^2 &= E\left[(\eta_\tau - \eta)^2\right] \\
&= E\left[\frac{1}{2\tau}\int_{-\tau}^{\tau} (X(t) - \eta)\; dt \; \frac{1}{2\tau}\int_{-\tau}^{\tau} (X(t') - \eta)\; dt'\right] \\
&= \frac{1}{4\tau^2}\int_{-\tau}^{\tau}\int_{-\tau}^{\tau} E[(X(t) - \eta)(X(t') - \eta)]\; dt\; dt' \\
&= \frac{1}{4\tau^2}\int_{-\tau}^{\tau}\int_{-\tau}^{\tau} C_X(t - t')\; dt\; dt'
\end{aligned}$$

$$\text{Let } \; I = \int_{-\tau}^{\tau} \int_{-\tau}^{\tau} C_X(t_1 - t_2)\, dt_2\, dt_1$$

- Let $z = t_1 - t_2$. We want to change the integration to be over t_2 and z



- Easy to see z goes from -2τ to 2τ
  When $z \geq 0$, for a given z, t_2 goes from $-\tau$ to $\tau - z$
  When $z < 0$, for a given z, t_2 goes from $-\tau - z$ to τ

- Now we get

$$
\begin{aligned}
I &= \int_{-\tau}^{\tau} \int_{-\tau}^{\tau} C_X(t_1 - t_2)\ dt_2\ dt_1 \\
&= \int_{-2\tau}^{0} \int_{-\tau-z}^{\tau} C_X(z)\ dt_2\ dz\ +\ \int_{0}^{2\tau} \int_{-\tau}^{\tau-z} C_X(z)\ dt_2\ dz
\end{aligned}
$$

$$
= \int_{-2\tau}^{0} C_X(z)\ (\tau - (-\tau - z))\ dz\ +\ \int_{0}^{2\tau} C_X(z)\ (\tau - z - (-\tau))\ dz
$$

$$
\begin{aligned}
&= \int_{-2\tau}^{0} C_X(z)\ (2\tau + z)\ dz\ +\ \int_{0}^{2\tau} C_X(z)\ (2\tau - z)\ dz \\
&= \int_{-2\tau}^{2\tau} C_X(z)\ (2\tau - |z|)\ dz
\end{aligned}
$$

- Now we get σ_τ^2 as

$$\begin{aligned}
\sigma_\tau^2 &= \frac{1}{4\tau^2}\int_{-\tau}^{\tau}\int_{-\tau}^{\tau} C_X(t-t')\,dt\,dt' \\
&= \frac{1}{4\tau^2}\int_{-2\tau}^{2\tau} C_X(z)\,(2\tau-|z|)\,dz \\
&= \frac{1}{2\tau}\int_{-2\tau}^{2\tau} C_X(z)\,\left(1-\frac{|z|}{2\tau}\right)\,dz
\end{aligned}$$

- Hence, a sufficient condition for $\sigma_\tau^2 \to 0$ is

$$\int_{-\infty}^{\infty} |C_X(z)|\,dz \;<\; \infty$$

- This is a sufficient condition for the process being mean-ergodic

- The process is mean-ergodic if

$$\frac{1}{2\tau}\int_{-\tau}^{\tau} X(t)\, dt \xrightarrow{P} E[X(t)] = \eta, \quad \text{as } \tau \to \infty$$

- A sufficient condition for this is that the covariance between $X(t)$ and $X(t+t')$ should go to zero sufficiently fast with increasing t':

$$\int_{-\infty}^{\infty} |C_X(z)|\, dz \; < \; \infty$$

- Similar relation holds for discrete time processes also.